Cloning and Sequencing of an *Agrobacterium tumefaciens* β-Glucosidase Gene Involved in Modifying a *vir*-Inducing Plant Signal Molecule

LINDA A. CASTLE,†* KRISTIN D. SMITH, AND ROY O. MORRIS

Biochemistry Department, University of Missouri—Columbia, Columbia, Missouri 65211

Induction of *Agrobacterium tumefaciens* virulence genes by plant phenolic compounds is essential for successful T-DNA transfer to a host plant. In Douglas fir needles, the major virulence region inducer is the glycoside coniferin (J. W. Morris and R. O. Morris, Proc. Natl. Acad. Sci. USA 87:3612–3618, 1990). *Agrobacterium* strains with high β-glucosidase activity respond to coniferin and infect Douglas fir seedlings, whereas most strains with low β-glucosidase activity fail to respond to coniferin and are avirulent on this host. We have cloned two β-glucosidase genes from *A. tumefaciens* B3/73 and sequenced one of them, *cbg1*. It appears to be part of a polycistronic unit and shows a high bias for GC-rich codons. When expressed in *Escherichia coli*, Cbg1 β-glucosidase hydrolyzes coniferin but not cellobiose. The 88-kDa predicted product of *cbg1* is highly similar to one other bacterial β-glucosidase and several fungal β-glucosidases. There is little homology between Cbg1 and other bacterial β-glucosidases, including an *Agrobacterium* cellobiase.

Communication between *Agrobacterium tumefaciens* and host plant cells is essential for successful tumorigenesis. During infection, Ti plasmid genes of the virulence (*vir*) region control transfer of the T-DNA from the bacterial cell to the plant cell. Genes of the virulence region, first defined by transposon mutagenesis (8), were found to be transcriptionally activated by plant exudates and suspension-cultured cells (15, 43). Two of the virulence products, VirA and VirG, are essential for induction of the other loci (42). These two proteins share conserved domains found in many bacterial two-component environmental responsive systems (29, 51); VirA is similar to other sensor proteins, and VirG shows homology to other transcriptional activators.

The first native *vir* region-inducing compounds were isolated from tobacco root exudates (41). Acetosyringone and hydroxy-acetosyringone were identified. Spencer and Towers (40) and Melchers et al. (25) systematically tested a large number of synthetic phenolic compounds to determine the structural requirements of active inducers. They found that most active compounds were composed of substituted benzene rings with free 4'-hydroxyl groups, 3'-methoxy and often 5'-methoxy groups, and usually an electron-withdrawing group at the 1' position. Flavonoid compounds purified from petunia pollen and stigma tissues have also been identified as *vir* region inducers (54). Coniferin, purified from Douglas fir (*Pseudotsuga menziesii*) needle extracts, was identified as the major native *vir* region inducer in that species (28). It differs from other inducers in that it is a glucoside in which the 4'-hydroxyl group is blocked. Hess et al. (12) have shown that the 4'-hydroxyl group is essential and probably directly interacts with VirA or a receptor protein. The aglycone, coniferyl alcohol, had previously been identified as an active inducer (25). Unlike coniferyl alcohol, however, coniferin induced only a particular subset of *A. tumefaciens* strains (28).

Morris and Morris (28) have shown that hydrolysis of coniferin to coniferyl alcohol may play a direct role in determining the virulence of *A. tumefaciens* on gymnosperms. Most *A. tumefaciens* strains that were highly virulent on conifer hosts were able to utilize coniferin for *vir* gene induction, whereas strains that were weakly tumorigenic could not. Both types of strains, however, recognized the aglycone, coniferyl alcohol, as an inducer. β-Glucosidase activity of the most virulent strains was found to be much higher than that of the weakly virulent strains. It appears that conversion of coniferin to coniferyl alcohol via bacterially encoded β-glucosidase affects virulence region induction and thus virulence. If indeed β-glucosidase is a necessary component of the pathway, this will be the first reported evidence of plant signal compound processing by *A. tumefaciens.*

We now report the isolation and characterization of a β-glucosidase gene from *A. tumefaciens* B3/73. The encoded enzyme catalyzes hydrolysis of coniferin but not cellobiose. Significant sequence homology to one bacterial and several fungal β-glucosidases was found, but there was little homology to other bacterial β-glucosidases.

MATERIALS AND METHODS

Strains, vectors, and growth conditions. The wild-type *A. tumefaciens* isolate B3/73 was obtained from L. Moore (USDA Ornamentals Laboratory, Corvallis, Oreg.) and was maintained on 523 medium (16) at 28°C. *Escherichia coli* DH5α was purchased from Bethesda Research Laboratories and was used to maintain all pGA482 (1), pBR322, pUC18, and pUC19 (53) constructs. pGA482 library clones were maintained on LB medium (34) with 15 μg of tetracycline per ml at 37°C. Subclones in pBR322, pUC18, and pUC19 were maintained on LB medium with 50 μg of ampicillin per ml. Plasmids were introduced into DH5α by electroporation with a Bio-Rad Gene Pulser at 2.5 kV, 25 μF, and 200 Ω for 4 to 5 ms as recommended by the manufacturer. *E. coli* JM109 (53) was maintained on M9 minimal medium (23). Sequencing vectors M13mp18 and M13mp19 (53) were trans-

---

* Corresponding author.
† Present address: Department of Botany, Oklahoma State University, Stillwater, OK 74078.

formed into $CaCl_2$-treated JM109 (34) and maintained on LB medium.

Construction and screening of genomic library. Genomic DNA was prepared from *A. tumefaciens* B3/73 by the cetyltrimethyl ammonium bromide precipitation method of Wilson (50) except that instead of CsCl purification, the DNA was extracted with chloroform several times, treated with RNases A and T1 (Sigma Chemical Co.), ethanol precipitated, spooled twice, and sheared by vortexing and passage through an 18-gauge needle. The DNA was partially digested with *Hind*III, and the 30- to 40-kb fragments were isolated by electroelution from an agarose gel. pGA482 cosmid DNA was digested to completion with *Hind*III and treated with calf intestine alkaline phosphatase (Bethesda Research Laboratories). To favor concatemer formation, a fourfold excess of linearized pGA482 over B3/73 *Hind*III fragments was used in the ligation mix. The library was packaged with GigaPack Plus (Stratagene Inc.) according to the manufacturer's instructions and used to infect DH5α cells. Appropriate dilutions were plated on nitrocellulose filters placed on LB medium-tetracycline plates. Colonies were screened for β-glucosidase activity by replica plating and exposing the replicated colonies to top agar containing 0.1 mg of 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-glucose; Research Organics, Inc.) per ml. Positive (blue) colonies were restreaked from the master plates and tested with X-glucose again to confirm activity.

Coniferin assays. Conversion of coniferin to coniferyl alcohol was assayed with Fast Blue RR salt (Sigma Chemical Co.), which interacts with free coniferyl alcohol to give an azo dye which is red in basic solution. *E. coli* library strains to be screened were grown at 37°C in LB (plus 10 μg of tetracycline per ml) from a 2% inoculum of a stationary-phase culture. B3/73 was also grown in LB medium but at 28°C and with a 4% inoculum. Coniferin (200 μM) was added to individual cultures at time zero, 2 h, or 4 h. Cultures were assayed at 6 h. Culture densities were determined (A_{590}), cultures were centrifuged, 110 μl of supernatant of each culture was transferred to single wells of a 96-well microtiter plate, 20 μl of 0.1 M Na_2HCO_3 was added, and the A_{490} was recorded (blank reading). Immediately after addition of 20 μl of 5-mg/ml Fast Blue RR salt, the A_{490} was measured. β-Glucosidase activity was expressed as coniferyl alcohol concentration per unit of cell density (micromolar per A_{590}). Absorbance values were read with a microplate reader (Molecular Devices Corp.).

Cellobiose growth determination. The ability of selected *E. coli* library clones to use cellobiose (Sigma Chemical Co.) as a carbon source for sustained growth was determined by comparing growth in M9 medium, pH 7.0 (23), containing 1% glucose or cellobiose as a carbon source. Overnight cultures grown in M9 medium-glucose were centrifuged, resuspended in M9 medium (no carbon source) and inoculated to an A_{590} of 0.03 into medium with glucose, cellobiose, or no carbon source. Absorbance measurements were taken every 2 to 4 h for 68 h.

DNA sequencing. Single-stranded dideoxy DNA sequencing (35) with M13mp18 and M13mp19 in JM109 was carried out by using U.S. Biochemicals Sequenase 2.0 with $[\alpha^{-35}S]$ dATP (NEN). Large-scale replicative-form DNA was prepared by alkaline lysis and single-stranded DNA was prepared by polyethylene glycol precipitation, both as described by Sambrook et al. (34).

Nested *Exo*III deletion derivatives were prepared by the method of Henikoff (11) as described by Sambrook et al. (34). The *Bam*HI-*Eco*RI fragment in M13mp18 was digested

with *Pst*I to protect the primer site and *Xba*I to create a substrate for *Exo*III. The same procedure with the *Bgl*II-*Pst*I fragment in M13mp19 was carried out with a filled *Kpn*I fragment substrate end and an *Sst*I-protected end. Double-stranded sequencing (5) was carried out with plasmid DNA purified by anion-exchange chromatography on Qiagen pack-500 columns. Purified DNA (3 μg) was mixed with 10 ng of primer in a 20-μl volume. The DNA mixture was adjusted to 0.2 M NaOH and heated to 85°C for 5 min before precipitating with ethanol. The DNA pellet was resuspended in Sequenase reaction buffer, and annealing and sequencing reactions were the same as with single-stranded DNA.

Both strands of the coding region were sequenced with primers supplied in the Sequenase kit (−40 17-mer) or internal primers synthesized by the DNA Core Facility at the University of Missouri—Columbia. Both dGTP and dITP reactions were carried out for all templates, since G-C compressions were a constant problem. Computer-aided GenBank data base searching and sequence analyses were carried out on a VAX Station 3100 computer using the University of Wisconsin Genetics Computer Group package (7).

Hybridization analysis. An agarose gel of *Hind*III-digested library clone DNAs was transferred to GeneScreen Plus (NEN Research Products) membranes with a vacublot system. DNA was cross-linked to the membranes by exposure to long-wave UV light for 10 min. A probe for the coding region of the β-glucosidase gene was prepared by polymerase chain reaction with synthetic oligonucleotide primers spanning a 1.6-kb internal region. The polymerase chain reaction was carried out with a Perkin-Elmer Cetus Gene-Amp DNA Amplification Kit with *Taq* polymerase in a Perkin-Elmer Cetus DNA thermal cycler. $[\alpha^{-32}P]dCTP$ (NEN) was substituted for cold dCTP in a 20-μl polymerase chain reaction with 25 cycles of 1.5 min of denaturation at 94°C, 2 min of annealing at 37°C, and 3 min of chain elongation at 72°C. Hybridization was carried out in a formamide buffer with Denhardt's reagent (6) and dextran sulfate (45) as described in the GeneScreen instruction manual. Exposure to X-ray film was for 6 h at −70°C with an intensifying screen.

Nucleotide sequence accession number. The GenBank nucleotide sequence accession number of the complete DNA sequence is M59852.

RESULTS

Cloning of a fragment with β-glucosidase activity. Clones from an *A. tumefaciens* B3/73 total genomic cosmid library were screened for β-glucosidase activity by overlaying colonies with top agar containing X-glucose. Of 3,200 clones screened, 29 had strongly positive reactions within 2 h and an additional 21 displayed slight reactions within 24 h. Activity was confirmed in 46 of the clones by restreaking and assaying with X-glucose. It appeared from these preliminary results that two categories of active clones were present: those that reacted strongly with X-glucose and those that were slower to react. In order to confirm that two categories were indeed present, DNA from individual clones was digested with *Hind*III and restriction fragment patterns were compared. Two types of patterns were seen, as shown in Fig. 1A. Type 1 clones all contained 5.7- and 12-kb *Hind*III fragments, and type 2 clones contained 7.0- and 8.8-kb fragments.

Substrate utilization and clone differences. Because there appeared to be two genes encoding β-glucosidase activity,

A

B
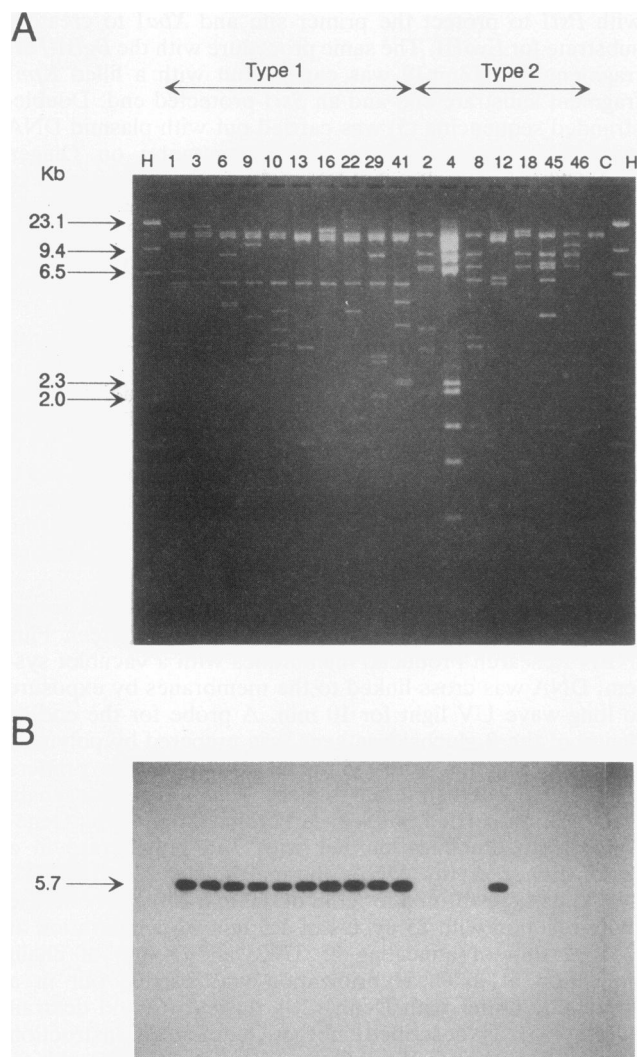

FIG. 1. Characterization of type 1 and type 2 β-glucosidase clones. (A) HindIII restriction digest patterns of library clones. Lanes 1, 3, 6, 9, 10, 13, 16, 22, 29, 41, and 12 (misplaced), type 1 clones with 5.7- and 12-kb bands; lanes 2, 4, 8, 18, 45, and 46, type 2 clones with 7.0- and 8.8-kb bands; lanes H, HindIII-digested lambda size markers; lane C, cosmid vector only. (B) Hybridization of cbg1 coding region probe to blot from gel in panel A.

FIG. 2. Utilization of coniferin as a substrate by β-glucosidase library clones. Type 1 clones 13 (□) and 22 (▽); type 2 clones 8 (△), 18 (◇), and 45 (○); and B3/73 (●) were exposed to coniferin for 0, 2, 4, or 6 h in a 6-h growth period. Coniferin hydrolysis is expressed as coniferyl alcohol concentration per unit of cell density.

the relative ability of each to cleave coniferin was assessed. As seen in Fig. 2, type 1 clones (and *A. tumefaciens* B3/73) rapidly and completely hydrolyzed coniferin to coniferyl alcohol. Over the same time period, type 2 clones were completely inactive.

The different substrate specificities of type 1 and 2 clones were also evident from their ability to grow on cellobiose (data not shown). *A. tumefaciens* B3/73 was able to use cellobiose as the sole carbon source. *E. coli* DH5α and type 1 clones were not able to grow on cellobiose. Type 2 clones
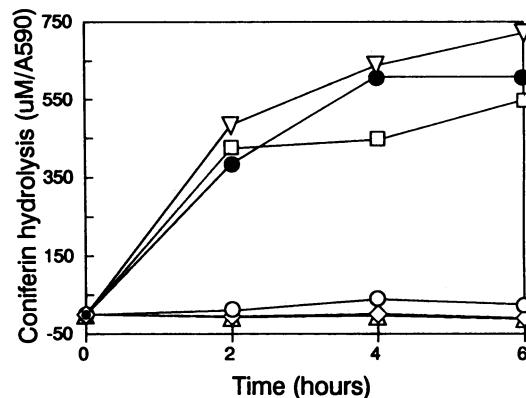
were able to utilize cellobiose but grew very slowly, suggesting that type 2 clones encode a cellobiase.

On the basis of the differences in restriction enzyme digest patterns, ability to hydrolyze coniferin, and growth on cellobiose, it was concluded that *A. tumefaciens* B3/73 has at least two β-glucosidases. The gene represented by the type 1 class was designated *cbg1* (for coniferin-specific β-glucosidase) and selected for further study.

Sequence of cbg1. The 5.7-kb HindIII fragment common to all type 1 clones was purified from a digest of clone 22 and ligated into pBR322. Clones with inserts in either orientation were able to cleave X-glucose, indicating that the entire β-glucosidase gene was probably located within this insert. An EcoRI, BamHI, BglII, and PstI restriction map of the insert was prepared, and appropriate fragments were subcloned and screened for activity on X-glucose. A 3.5-kb BamHI-PstI fragment with an internal PstI site was found to cleave X-glucose when cloned into pUC19. When cloned into pUC18, no β-glucosidase activity was expressed, indicating that this fragment probably lacked its native promoter.

The restriction map of the initial HindIII 5.7-kb insert is shown in Fig. 3A. Four fragments making up the 3.5-kb subclone, BamHI-EcoRI (1.25 kb), EcoRI-PstI (0.58 kb), PstI-BglII (1.03 kb), and BglII-PstI (0.67 kb), were cloned into the sequencing vectors, M13mp18 and M13mp19, and the ends were sequenced. XmaIII-XmaIII and EcoRV-SmaI fragments overlapping the 5' PstI site were also isolated, cloned into M13, and sequenced from both directions. Nested ExoIII deletion fragments were obtained for the top strand for most of the BamHI-EcoRI fragment. Only one nested deletion fragment was obtained for the bottom strand of the PstI-BglII fragment. Synthetic oligonucleotides derived from the sequences obtained from the subfragments were used to complete the sequence of both strands. The sequence surrounding the EcoRI site in the pUC19:3.5-kb

---

FIG. 3. Restriction map and sequence of cbg1. (A) The β-glucosidase activity of clone 22 was localized to a 5.7-kb HindIII fragment. The coding region (arrow) begins just downstream of the BamHI site and terminates 400 bp upstream of the BglII site. Abbreviations: Bg, BglII; Bm, BamHI; H, HindIII; P, PstI; R, EcoRI; R5, EcoRV; Sm, SmaI; Ss, SstI; X, XmaIII. (B) The putative ribosome binding site is underlined, upstream tandem terminators are double underlined, and a potential 3' termination GC-rich dyad symmetry at positions 2477 to 2507 is underlined. The predicted amino acid sequence is given below the nucleotide sequence.

A
1 Kb
cbg1

H P   Bm Ss   R R5 P   Sm   Bg   P H
X X
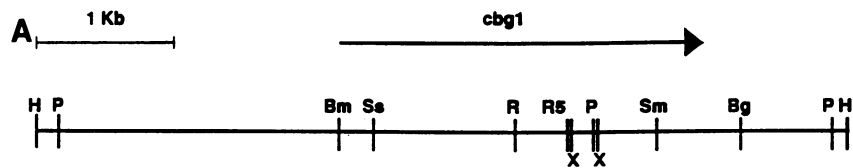
B
-180 CGATCGGCTC GGGACACGGA CAGAATTACT CACTGCTCTA TGCCTTGGCG GCGGTCAGTG CCTTTTTGGG CGCGCTGGCA

-100 ATCGTGCCGG TACGTGGCGC ACGTTGATAA GATCAATATT TCGGGCGGGA ACGCCTGCCT AATGGTCCCC ATCGGACCGA

 -20 TTAACATGGG AGGATCCCCT ATGATCGACG ATATTCTCGA TAAGATGACA CTGGAGGAGC AGGTGTCGCT GCTCTCGGGC
                        M   I   D   D   I   L   D   K   M   T   L   E   E   Q   V   S   L   L   S   G

  60 GCGGATTTCT GGACGACCGT CGCGATCGAG CGGCTCGGCG TGCCGAAGAT CAAGGTTACC GACGGCCCCA ATGGCGCACG
      A   D   F   W   T   T   V   A   I   E   R   L   G   V   P   K   I   K   V   T   D   G   P   N   G   A   R

 140 CGGCGGCGGG TCGCTGGTCG GCGGCGTCAA GTCCGCCTGC TTCCCGGTGG CAATCGCGCT TGGAGCGACG TGGGACCCGG
       G   G   G   S   L   V   G   G   V   K   S   A   C   F   P   V   A   I   A   L   G   A   T   W   D   P   E

 220 AGCTCATCGA GCGCGCCGGC GTGGCGCTGG GAGGACAAGC CAAGAGCAAG GGCGCGTCGG TGCTTCTGGC GCCGACCGTC
       L   I   E   R   A   G   V   A   L   G   G   Q   A   K   S   K   G   A   S   V   L   L   A   P   T   V

 300 AACATTCACC GCTCCGGCCT CAATGGCCGC AACTTCGAAT GCTATTCGGA AGACCCGGCG CTGACCGCCG CCTGCGCCGT
       N   I   H   R   S   G   L   N   G   R   N   F   E   C   Y   S   E   D   P   A   L   T   A   A   C   A   V

 380 CGCCTATATC AATGGCGTGC AGAGCCAGGG TGTGGCCGCC ACGATCAAGC ACTTCGTCGC CAACGAGTCC GAGATCGAGC
       A   Y   I   N   G   V   Q   S   Q   G   V   A   A   T   I   K   H   F   V   A   N   E   S   E   I   E   R

 460 GGCAGACCAT GTCCTCCGAT GTCGATGAGC GGACGCTGCG CGAAATCTAT CTGCCGCCTT TCGAGGAGGC GGTGAAGAAG
       Q   T   M   S   S   D   V   D   E   R   T   L   R   E   I   Y   L   P   P   F   E   E   A   V   K   K

 540 GCCGGCGTGA AGGCCGTCAT GTCCTCCTAC AACAAGCTCA ACGGCACCTA TACGAGCGAA AATCCCTGGC TGCTGACGAA
       A   G   V   K   A   V   M   S   S   Y   N   K   L   N   G   T   Y   T   S   E   N   P   W   L   L   T   K

 620 AGTCCTGCGC GAGGAATGGG GCTTCGACGG CGTGGTCATG TCCGACTGGT TCGGCTCGCA CTCGACGGCT GAAACCATCA
       V   L   R   E   E   W   G   F   D   G   V   V   M   S   D   W   F   G   S   H   S   T   A   E   T   I   N

 700 ATGCCGGGCT CGATCTGGAG ATGCCGGGGC CTTGGCGGGA TCGCGGCGAA AAGCTGGTCG CCGCAGTCCG GGAAGGCAAG
        A   G   L   D   L   E   M   P   G   P   W   R   D   R   G   E   K   L   V   A   A   V   R   E   G   K

 780 GTAAAGGCCG AGACCGTGCG CGCTTCGGCA CGGCGTATTC TCCTTCTGCT CGAACGCGTC GGCGCCTTTG AAAAGGCGCC
       V   K   A   E   T   V   R   A   S   A   R   R   I   L   L   L   L   E   R   V   G   A   F   E   K   A   P

 860 TGATCTCGCC GAACACGCGC TTGATCTGCC GGAAGATCGT GCGCTCATCC GCCAACTCGG TGCGGAGGGT GCTGTACTCC
       D   L   A   E   H   A   L   D   L   P   E   D   R   A   L   I   R   Q   L   G   A   E   G   A   V   L   L

 940 TGAAGAATGA CGGAGTGCTG CCGCTTGCCA AGTCGTCCTT CGACCAGATC GCCGTCATCG GCCCCAATGC GGCTTCCGCA
       K   N   D   G   V   L   P   L   A   K   S   S   F   D   Q   I   A   V   I   G   P   N   A   A   S   A

1020 CGCGTCATGG GCGGAGGAAG CGCGCGGATT GCCGCGCATT ATACGGTGAG CCCACTTGAG GGCATTCGCG CGGCGCTGTC
       R   V   M   G   G   G   S   A   R   I   A   A   H   Y   T   V   S   P   L   E   G   I   R   A   A   L   S

1100 CAACGCCAAC AGCCTCCGCC ATGCGGTCGG CTGTAACAAC AACCGGCTCA TCGACGTCTT CAGCGGCGAG ATGACGGTGG
       N   A   N   S   L   R   H   A   V   G   C   N   N   N   R   L   I   D   V   F   S   G   E   M   T   V   E

1180 AATACTTCAA GGGACGCGGC TTCGAGAGCC GTCCGGTCCA TGTCGAGACC GTCGAAAAGG GCGAATTCTT CTGGTTCGAT
       Y   F   K   G   R   G   F   E   S   R   P   V   H   V   E   T   V   E   K   G   E   F   F   W   F   D

1481

1260   CTTCCGTCCG GCGACCTTGA TCTCGCCGAT TTTTCGGCGC GCATGACGGC GACCTTCGTG CCGCAGGAAA CCGGTGAACA
        L P S G   D L D   L A D   F S A R   M T A   T F V   P Q E T   G E H

1340   CATCTTCGGC ATGACCAATG CTGGGCTTGC TCGGCTGTTC GTGGACGGCG AACTGGTGGT CGATGGCTAT GACGGTTGGA
        I F G   M T N A G L A   R L F   V D G E   L V V   D G Y   D G W T

1420   CGAAGGGTGA GAACTTTTTT GGAACCGCGA ACAGCGAGCA GCGTCGGGCG GTAACGCTTG GGGCCGCACG CCGCTACCGG
         K G E   N F F   G T A N   S E Q   R R A   V T L G   A A R   R Y R

1500   GTTGTGGTCG AATATGAGGC GCCGAAGGCC AGCCTGGACG GCATCAACAT ATGTGCGCTC CGCTTCGGTG TCGAAAAGCC
        V V V E   Y E A   P K A   S L D G   I N I   C A L   R F G V   E K P

1580   GCTCGGCGAT GCCGGGATTG CGGAGGCGGT CGAAACCGCC CGCAAGTCCG ATATCGTACT GCTCCTCGTC GGCCGTGAGG
         L G D   A G I A   E A V   E T A R   K S D   I V L   L L V   G R E G

1660   GCGAGTGGGA CACCGAAGGT CTGGATCTGC CCGACATGCG CCTGCCGGGT CGCCAGGAGG AGCTGATCGA GGCGGTCGCC
         E W D   T E G   L D L P   D M R   L P G   R Q E E   L I E   A V A

1740   GAAACCAATC CCAACGTGGT CGTGGTACTG CAAACGGGTG GTCCCATCGA GATGCCATGG CTCGGCAAGG TGCGTGCGGT
        E T N P   N V V   V V L   Q T G G   P I E   M P W   L G K V   R A V

1820   GCTGCAGATG TGGTATCCCG GCCAGGAACT TGGCAATGCG CTTGCGGACG TTCTCTTTGG TGATGTCGAG CCTGCCGGCC
         L Q M   W Y P G   Q E L   G N A   L A D V   L F G   D V E   P A G R

1900   GCTTGCCACA GACCTTCCCG AAGGCGCTCA CGGATAATTC CGCCATTACC GACGATCCGT CGATCTATCC TGGCCAGGAC
         L P Q   T F P   K A L T   D N S   A I T   D D P S   I Y P   G Q D

1980   GGCCATGTGC GCTACGCGGA AGGGATCTTC GTCGGCTATC GCCATCACGA TACAAGAGAG ATCGAACCAC TCTTCCCCTT
        G H V R   Y A E   G I F   V G Y R   H H D   T R E   I E P L   F P F

2060   CGGCTTCGGT CTTGGCTACA CCCGCTTTAC CTGGGGTGCC CCGCAACTAT CGGGAACGGA AATGGGGGCG GATGGTCTTA
         G F G   L G Y T   R F T   W G A   P Q L S   G T E   M G A   D G L T

2140   CGGTGACGGT CGATGTCACC AATATAGGCG ACAGGGCGGG ATCGGACGTG GTGCAGCTCT ATGTTCACTC TCCCAATGCC
         V T V   D V T   N I G D   R A G   S D V   V Q L Y   V H S   P N A

2220   AGGGTCGAGC GGCCGTTCAA GGAGCTGCGT GCCTTTGCGA AGCTCAAGCT GGCCCCGGGC GCGACCGGTA CGGCGGTGCT
         R V E R   P F K   E L R   A F A K   L K L   A P G   A T G T   A V L

2300   GAAGATCGCT CCTCGCGACT TGGCTTACTT CGATGTCGAG GCCGGTCGTT TCCGGGCTGA TGCGGGCAAG TACGAGCTGA
         K I A   P R D L   A Y F   D V E   A G R F   R A D   A G K   Y E L I

2380   TCGTGGCGGC CAGCGCCATC GATATCCGGG CAAGCGTAAG TATTCACTTG CCGGTCGATC ATGTGATGGA GCCGTAGCGC
         V A A   S A I   D I R A   S V S   I H L   P V D H   V M E   P *

2460   AAGCATTCAA TATGAGCCGC GCCGCCTATC TTAGAAGGAG GCGGCGCGAT CCGCGAAGTG AGTGGGACAG GGCCTGGCGA

2540   TCGCTCTCTT AATGTACATA CCTGGGATGT ATATTTCCAG GCCGATGCGC TATGATACAT CCATGTACAC ACATGGGAGC

2620   ATCCCGCTAT GCAGATGACC AAGGCATTCA AGAACGGCAA TTCCCAAGCC GTACGCATCC CAGCAGAGAT TGCCTATGAG

2700   CGCACGGACA TGGAGCTGGA GATTGAGCGG GTAGGCGATG AAATCCGCAT TCGTCCTGTC CGCCAGTCGT TGGCCGGCGT

FIG. 3—*Continued.*

```
2780  CATGGATGTT TTCGCACGCT TTTCCCCCAA TTTCATGAGC GAGGGCAGGG GTGACCAAGA GCAGGACGAG CGCGAGAGAA

2860  GATCTGATGC CGCGCTATAT GCTCGATACG AATATGTGCA TCTATCTGAT GAAGAACCAG CCCGAGCAAG TCGCTCGTCG

2940  TTTTGCCGCT TGCTACGTCG GTGACGTGGT GATGTCGGCC ATTACCTTCG CGGAGCTGGA TTACGGTGTT GCCGCGTCTG

3020  ATGCCCCGGA ACGGGAGCGC GGCAACCTTG ATGCCTTGGT GACCCTCATT GCGTAGAGCC GTTCGATATC GCCGCTGCCC

3100  ACGCTTATGG CCCAGTACGC ATGGCAACGC GAGAGCAAGA AAGACCATCT GACAGCTGAT CGCATCGCGC TCATTGCCTT

3180  GATACGTGCT GTACCACAAC ACACGAGATT CATGCTATCC ACTCAATCGA ACTGCTGATA CTGAGGCTGC GTTCGTGCTT

3260  ACGGCAGCGC CGACTAAAGG AACCATCGTC CCTTAGTCAT CTCAGATCCA TGCCTTCATC GGCATTGCGA TCATGACACA

3340  CGGCTCATTT CAGACTTGGA AACAGCCTAG AACAACCGCT CCAGCCAGCC GTGCTTGTCG GGCGCACGTC CGTTTTGAAT

3420  GTCGAGGAGT GCCGCCTTCA GGCGCGTGGT CACCGGACCG GCGCCGCCGT CGCCAATGGT GAAATTGTGC TTGCGGCCCT

3500  TGAGCTTGCC GATCGGAGTC ACCACTGCAG
```

FIG. 3—*Continued.*

*Bam*HI-*Pst*I clone and the sequence upstream of the *Bam*HI site in the pBR322:5.7-kb *Hin*dIII plasmid were obtained by double-stranded plasmid sequencing with synthetic oligonucleotide primers. Sequence data (3,710 nucleotides) were obtained for 176 nucleotides upstream of the *Bam*HI site through the downstream *Pst*I site.

The complete nucleotide sequence and derived amino acid sequence are shown in Fig. 3B. A single open reading frame of 2,457 nucleotides was present as shown. There were several possible ATG start codons in frame. Only one was preceded by a consensus ribosome binding site (39), shown underlined in Fig. 3B. No consensus *E. coli* −10 or −35 sequences (10) were found in the region upstream of the coding region. A tandem termination codon (double underlined in Fig. 3B) preceded by an open reading frame was found 70 bp upstream of the proposed start codon. It is possible, therefore, that *cbg1* is part of a polycistronic unit. No *vir* box (44) sequences were found.

Further evidence of two distinct genes cloned from *A. tumefaciens* B3/73 was obtained by probing a blot of the agarose gel shown in Fig. 1A with a polymerase chain reaction-generated probe from an internal region of *cbg1*. Only the 5.7-kb band of type 1 clones hybridized to the probe, as shown in Fig. 1B. Type 2 clone DNA did not hybridize to the *cbg1* probe.

A very good 12-bp GC-rich dyad symmetry (underlined in Fig. 3B, nucleotides 2477 to 2507) with a 7-bp loop downstream of the termination codon may serve as a transcription termination signal for *cbg1*. Consensus transcription termination sites generally have uridine-rich tracks following the hairpin structure in mRNA, although some good terminators do not (31). A track of T residues was not found downstream of any dyad symmetries near *cbg1*. The polycistronic unit does not appear to continue past *cbg1*, since no significant open reading frames exist in over 1 kb downstream of *cbg1*.

The GC content of the DNA sequenced was comparable to that reported for *A. tumefaciens*. The overall GC content of the sequenced region was 60.7%. That of the coding region was 62.6%. The GC content of *A. tumefaciens* is 60 to 62%, whereas *E. coli* has approximately 50% GC (30). It was

expected, therefore, that there would be a bias for codons with G and C in the third position compared with *E. coli* genes (18). However, there was much more GC third-position bias in *cbg1* than in *A. tumefaciens vir* genes (48). The most dramatic difference was seen for lysine codons. In *cbg1*, 97% of the 31 lysine residues were coded for by AAG rather than AAA. The *vir* genes use almost equal frequencies of both, with a bias for AAA. Threonine also showed a strong bias, with only 5% of the 24 codons coded by ACA and none by ACT, whereas the *vir* genes used these codons 40% of the time. Several other amino acids also showed a bias in codon usage.

Similarity of Cbg1 to other β-glucosidases. There are 818 amino acid residues in the translated sequence of Cbg1, and the predicted molecular mass is 88,289 Da. The protein is generally hydrophilic, and the N-terminal region does not contain a signal leader peptide for export (49).

A computer search of translated GenBank sequences revealed several β-glucosidase proteins with high homology to Cbg1. These comparisons are illustrated in Table 1. The greatest homology was to a β-glucosidase, BglB, from the

TABLE 1. Comparison of Cbg1 with fungal and bacterial β-glucosidases[a]

Protein	Species	% Identity	% Similarity
BglB	*C. thermocellum*	43.1	64.3
Glvk	*K. fragilis*	39.8	61.3
Bgl1	*T. reesei*	36.2	58.6
Bgl2	*S. fibuligera*	32.1	58.5
Bgl1	*S. fibuligera*	29.6	55.0
Clhq	*C. pelliculosa*	28.6	54.4
Abg	*Agrobacterium* sp.	23.9	47.7
BglB	*E. coli*	20.0	48.0
BglA	*C. saccharolyticum*	18.4	46.5
BglA	*B. fibrisolvens*	17.2	44.2

[a] Amino acid sequences were compared by using the GAP program from the Genetics Computer Group package (7).

```
A3       A. wentii               aZLGFZGFVMSDWaAhhagvsgAlaGLBMgsMPGb
Cbg1     A. tumefaciens      203 wLLtkvLreEwGFdGvVMSDWfgshstaetinaGLDl.eMPGp
BglB     C. thermocellum     212 YLLtevLKnEwmhdGFVvSDWgAvndrvsgkdaGLDl.eMPts
Glvk     K. fragilis         207 kLLidiLrdEwkwdGmlMSDWfgtyttaaaiknGLDi.efPGp
Bgl1     T. reesei           217 YtLqtvLKdqLGFpGyVMtDWnAqhttvqsAnsGLDM.sMPGt
Bgl1     S. fibuligera       275 YmmnhlLKeELGFqGFVvSDWgAqlsgvysAisGLDM.sMPGe
Bgl2     S. fibuligera       279 YminhlLKeELGFqGFVvSDWaAqmsgaysAisGLDM.sMPGe
Clhq     C. pelliculosa      280 YLLnylLKeELGFqGFVMtDWgAlysgidaAnaGLDM.dMPce


Consensus                        YLL---LK-ELGF-GFVMSDW-A-------A--GLDM--MPG-
```

FIG. 4. Active site homology. The sequences homologous to the identified active site sequence in A$_3$ of *Aspergillus wentii* were compared. Capital letters indicate residues identical to the consensus. The substrate-binding aspartic acid residue is in boldface. The amino acid positions of the fragments within each protein are indicated.

cellulolytic thermophilic bacterium *Clostridium thermocellum* (9). Cbg1 and BglB were 43% identical and 64% similar. The similarity extended throughout the proteins, with much of the difference accounted for by an absence of about 80 residues in the center of BglB. No consensus promoter region was found upstream of *bglB*. However, an incomplete open reading frame was found, indicating that this gene, like *cbg1*, may be part of a polycistronic unit. The other proteins showing homology to Cbg1 were from yeasts and filamentous fungi. Glvk, from the fungus *Kluyveromyces fragilis* (32), was 40% identical and 61% similar to Cbg1. Homology was found throughout the proteins, with no large gaps. The extracellular β-glucosidase Bgl1 from the cellulolytic fungus *Trichoderma reesei* (2) was also very similar to Cbg1. Bgl1 has 45 N-terminal amino acids not found in Cbg1 which probably code for a signal peptide. Two β-glucosidases, Bgl1 and Bgl2, from the yeast *Saccharomycopsis fibuligera* (22) were 29 and 32% identical and 55 and 59% similar to Cbg1, respectively. These proteins contain probable leader signal peptides for membrane transport not found in Cbg1. Both enzymes are aryl-β-glucosidases, but only Bgl1 is able to hydrolyze cellobiose. Clhq from *Candida pelliculosa* (17) and Cbg1 are 29% identical and 54% similar in sequence. Clhq also has a signal peptide for transport into the periplasmic space. A partial sequence of a β-glucosidase from the basidiomycete *Schizophyllum commune* shows 42% identity to the carboxy-terminal 222 amino acid residues of the *C. pelliculosa* β-glucosidase and probably belongs among this class of proteins (26).

Although the BglA β-glucosidase from the bacterium *Butyrivibrio fibrisolvens* (20) is similar in size (830 amino acids, 92 kDa) to Cbg1 and the related enzymes, it shows only 17% identity and 44% similarity to Cbg1. Three smaller bacterial β-glucosidases, BglB (471 residues, 53 kDa) from *E. coli* (36), Abg (458 residues, 51 kDa) from *Agrobacterium* strain ATCC 21400 (46), and BglA (453 residues, 54 kDa) from the extreme thermophile *Caldocellum saccharolyticum* (21), are also less related to Cbg1 (Table 1).

The active site of β-glucosidase A$_3$ from *Aspergillus wentii* has been identified by labeling with the tritiated substrate analog conduritol B epoxide (3). An aspartic acid residue bound the analog. Two nearby asparagine residues in the consensus Asn-X-Thr/Ser sites were glycosylated. A multiple alignment of the regions homologous to the A$_3$ active site is shown in Fig. 4. Cbg1 and each of the similar enzymes mentioned previously all contain regions homologous to the A$_3$ active site. Each of the enzymes has a tryptophan residue immediately following the conserved aspartic acid and either a serine or threonine before it. The A$_3$-type enzymes also have glutamic acid residues five and nine residues upstream of the conserved aspartic acid residue and a second con-

served aspartic acid motif, Gly-Leu-Asp-Met, which may be important for activity (2). Cbg1 has two eukaryotic consensus glycosylation sites at asparagine residues 148 and 194, although these would not be expected to be glycosylated in *A. tumefaciens*. The bacterial β-glucosidases from *E. coli*, *Agrobacterium* strain ATCC 21400, *B. fibrisolvens*, and *C. saccharolyticum* do not contain regions homologous to the A$_3$ active site.

DISCUSSION

Evidence for two β-glucosidases in *A. tumefaciens*. There appear to be at least two constitutive β-glucosidase genes in *A. tumefaciens* B3/73. Those represented by type 1 clones were able to hydrolyze X-glucose and coniferin but not cellobiose. On the other hand, those represented by type 2 clones had lower activity on X-glucose and no activity on coniferin and were able to utilize cellobiose for growth. In addition, type 2 clone inserts did not hybridize with a probe from the type 1, *cbg1*, coding region. Preliminary evidence indicates that *cbg1* hybridizes to chromosomal DNA and is not located on the Ti plasmid.

Many fungi and bacteria have two separate β-glucosidase activities (52). *A. tumefaciens* strains are characterized as being able to hydrolyze cellobiose (13). *abg*, from *Agrobacterium* strain ATCC 21400, was isolated on the basis of cellobiase activity (46, 47). It is possible, therefore, that the β-glucosidase of the type 2 clones from *A. tumefaciens* B3/73 is a cellobiase homologous to Abg.

Relationship to other β-glucosidases. β-Glucosidase enzymes are found in organisms as diverse as mammals, plants, fungi, and bacteria. There appears to be a high degree of similarity among the fungal and bacterial proteins with homology to Cbg1. Although there is a high degree of similarity surrounding the glucose-binding catalytic site in the aryl-β-glucosidases of *S. fibuligera*, the substrate specificities differ (22). Thus, other regions must be involved in specific substrate recognition and affinity. Very little information is known about aryl versus alkyl or pure carbohydrate substrate recognition. A small region of similarity to Cbg1 was found in several human tyrosine-3-hydroxylases. The sequence WFXXXXXXLD found at amino acids 418 to 427 in Cbg1 was present in the hydroxylases and may be involved in recognition of the aromatic ring of coniferin.

Potential localization of Cbg1 and relation to activity. In order for Cbg1 to act on coniferin, the enzyme and substrate must come in contact. The β-glucosidase activity of Cbg1 in *A. tumefaciens* and *E. coli* was found with intact or disrupted cells, but no enzymatic activity was seen in culture supernatants after removal of cells. Although proteins located in the periplasmic space and in the outer membrane

have N-terminal leader sequences, most proteins located in bacterial inner membranes do not have cleaved leader sequences (49). This suggests that Cbg1 is intracellular or in the inner membrane and that the substrates and products traverse the bacterial membrane by diffusion or transport. The *K. fragilis* enzyme, Glvk, which is similar to Cbg1 throughout the sequence, also lacks a leader peptide sequence and was reported to be an intracellular enzyme (33).

Because Cbg1 may be involved in phenolic signal processing, comparisons were made with the VirA and NodD proteins, which are known to interact with the same types of phenolics. VirA from *A. tumefaciens* A6 (19) is close to the same size as Cbg1 and was found to be 17% identical and 41% similar to Cbg1. However, no regions of high homology were found. *Rhizobium leguminosarum* NodD (37), which is only 303 amino acids in length, was 23% identical and 46% similar to Cbg1. As with VirA, no regions of high homology were found.

It is not known how *A. tumefaciens vir* region inducers interact with VirA to promote the T-DNA transfer cascade. The interaction has been proposed to occur in the inner membrane (24). Diffusion or transport of phenolic inducers through the outer membrane would make them available to inner membrane proteins. Recently, Huang et al. (14) have identified a chromosomal locus, *chvE*, that encodes a periplasmic receptor protein involved in uptake of sugars. It has been shown that monosaccharides act synergistically with phenolic inducers to activate the *vir* genes (4, 38). Mutations in *chvE* resulted in attenuated *vir* region induction in the presence of acetosyringone (14). Glycosylated phenolic inducers that can be transported across the outer membrane and deglycosylated by β-glucosidase would present both a glucose moiety and a phenolic moiety to maximally induce the virulence cascade in strains such as B3/73. The open reading frame upstream of *cbg1* may encode a transmembrane transport protein similar to ChvE for uptake of coniferin and other glycosides. In *E. coli*, the *bglS* gene upstream of *bglB* encodes a transport protein that mediates the intracellular accumulation of aryl-β-glucoside-phosphate substrates for BglB (36).

Role of Cbg1 in virulence. The ultimate consequence of *cbg1* expression is the most important property yet to be determined. Hydrolysis of coniferin to coniferyl alcohol clearly plays an important role in presenting the VirA protein or phenolic binding receptor with an active virulence region inducer from Douglas fir during infection (28). Determination of the cellular location of Cbg1 will help to clarify its relationship to induction of the virulence cascade. A large number of *Agrobacterium* strains have been tested for virulence on Douglas fir, with a range of infectivities evident (27). How much of the host range variation on Douglas fir can be accounted for by *cbg1*? Experiments are in progress to delete *cbg1* from virulent strains and to introduce the gene into strains not virulent on Douglas fir to determine the effect on virulence. It would be of interest to determine whether other substrate-specific inducer-modifying enzymes play analogous roles in other plant-agrobacterium interactions and perhaps in other plant-microbe systems such as legume-rhizobium interactions. Currently, the substrate specificity of Cbg1 is being determined, and potentially homologous plant enzymes and genes are being isolated.

REFERENCES

1. **An, G.** 1986. Development of plant promoter expression vectors and their use for analysis of differential activity of nopaline synthase promoter in transformed tobacco cells. Plant Physiol. **81:**86–91.
2. **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.
3. **Bause, E., and G. Legler.** 1980. Isolation and structure of a tryptic glycopeptide from the active site of β-glucosidase A₃ from *Aspergillus wentii*. Biochim. Biophys. Acta **626:**459–465.
4. **Cangelosi, G. A., R. G. Ankenbauer, and E. W. Nester.** 1990. Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. Proc. Natl. Acad. Sci. USA **87:**6708–6712.
5. **Chen, E. Y., and P. H. Seeberg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4:**165–170.
6. **Denhardt, D. T.** 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. **23:**641–646.
7. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
8. **Garfinkel, D. J., and E. W. Nester.** 1980. *Agrobacterium tumefaciens* mutants affected in crown gall tumorigenesis and octopine catabolism. J. Bacteriol. **144:**732–743.
9. **Gräbnitz, F., K. P. Rücknagel, M. Seiß, and W. L. Staudenbauer.** 1989. Nucleotide sequence of the *Clostridium thermocellum bglB* gene encoding thermostable β-glucosidase B: homology to fungal β-glucosidases. Mol. Gen. Genet. **217:**70–76.
10. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter sequences. Nucleic Acids Res. **8:**2237–2255.
11. **Henikoff, S.** 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155:**156–165.
12. **Hess, K. M., M. W. Dudley, D. G. Lynn, R. D. Joerger, and A. N. Binns.** 1991. Mechanism of phenolic activation of *Agrobacterium* virulence genes: development of a specific inhibitor of bacterial sensor/response systems. Proc. Natl. Acad. Sci. USA **88:**7854–7858.
13. **Holmes, B., and P. Roberts.** 1981. The classification, identification and nomenclature of agrobacteria. Appl. Bacteriol. **50:**443–467.
14. **Huang, M. W., G. A. Cangelosi, W. Halperin, and E. W. Nester.** 1990. A chromosomal *Agrobacterium tumefaciens* gene required for effective plant signal transduction. J. Bacteriol. **172:**1814–1822.
15. **Janssens, A., C. Genetello, M. Van Montagu, and P. Zambryski.** 1986. Plant cells induce transcription of the *Agrobacterium tumefaciens* nopaline pTiC58 virulence region. Plant Sci. **47:**185–193.
16. **Kado, C. I., and M. G. Heskett.** 1970. Selective media for isolation of *Agrobacterium*, *Corynebacterium*, *Erwinia*, *Pseudomonas* and *Xanthomonas*. Phytopathology **60:**969–976.
17. **Kohchi, C., and A. Toh-e.** 1985. Nucleotide sequence of *Candida pelliculosa* β-glucosidase gene. Nucleic Acids Res. **17:**6273–6282.
18. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80:**687–691.
19. **Leroux, B., M. F. Yanofsky, S. C. Winans, J. E. Ward, S. F. Ziegler, and E. W. Nester.** 1987. Characterization of the *virA* locus of *Agrobacterium tumefaciens*: a transcriptional regulator and host range determinant. EMBO J. **6:**849–856.

20. Lin, L. L., E. Rumbak, H. Zappe, J. A. Thomson, and D. R. Woods. 1990. Cloning, sequencing and analysis of expression of a *Butyrivibrio fibrisolvens* gene encoding a β-glucosidase. Gen. Microbiol. **136:**1567–1576.

21. Love, D. R., R. Fisher, and P. L. Bergquist. 1988. Sequence structure and expression of a cloned β-glucosidase gene from an extreme thermophile. Mol. Gen. Genet. **213:**84–92.

22. 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.

23. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual, p. 68. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

24. Melchers, L. S., T. J. G. Regensburg-Tuïnk, R. B. Bourret, N. J. A. Sedee, R. A. Schilperoort, and P. J. J. Hooykaas. 1989. Membrane topology and functional analysis of the sensory protein VirA of *Agrobacterium tumefaciens*. EMBO J. **8:**1919–1925.

25. Melchers, L. S., T. J. G. Regensburg-Tuïnk, R. A. Schilperoort, and P. J. J. Hooykaas. 1989. Specificity of signal molecules in the activation of agrobacterium virulence gene expression. Mol. Microbiol. **3:**969–977.

26. Moranelli, F., J. R. Barbier, M. J. Dove, R. M. MacKay, V. L. Seligy, M. Yaguchi, and G. E. Willick. 1986. A clone for *Schizophyllum commune* β-glucosidase: homology with a yeast β-glucosidase. Biochem. Int. **12:**905–912.

27. Morris, J. W., L. A. Castle, and R. O. Morris. 1989. Efficacy of different *Agrobacterium tumefaciens* strains in transformation of pinaceous gymnosperms. Physiol. Mol. Plant Pathol. **34:**451–461.

28. Morris, J. W., and R. O. Morris. 1990. Identification of an *Agrobacterium tumefaciens* virulence gene inducer from the pinaceous gymnosperm *Pseudostuga menziesii*. Proc. Natl. Acad. Sci. USA **87:**3614–3618.

29. Nixon, B. T., C. W. Ronson, and F. M. Ausubel. 1986. Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. Proc. Natl. Acad. Sci. USA **83:**7850–7854.

30. Normone, W. M., and J. R. Brown. 1970. Guanine plus cytosine (G+C) composition of bacteria, p. H24–H74. *In* H. A. Sober (ed.), CRC handbook of biochemistry: selected data for molecular biology, 2nd ed. The Chemical Rubber Co., Cleveland.

31. Platt, T. 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55:**339–372.

32. Raynal, A., C. Gerbaud, M. C. Francingues, and M. Guerineau. 1987. Sequence and transcription of the β-glucosidase gene of *Kluyveromyces fragilis* cloned in *Saccharomyces cerevisiae*. Curr. Genet. **12:**175–184.

33. 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.

34. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

35. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

36. Schnetz, K., C. Toloczyki, and B. Rak. 1987. β-Glucoside (*bgl*) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. J. Bacteriol. **169:**2579–2590.

37. Shearman, C. A., L. Rossen, A. W. B. Johnston, and J. A.

38. Shimoda, N., A. Toyoda-Yamamoto, J. Nagamine, S. Usami, M. Katayama, Y. Sakagami, and Y. Machida. 1990. Control of expression of *Agrobacterium vir* genes by synergistic actions of phenolic signal molecules and monosaccharides. Proc. Natl. Acad. Sci. USA **87:**6684–6688.

39. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.

40. Spencer, P. A., and G. H. N. Towers. 1988. Specificity of signal compounds detected by *Agrobacterium tumefaciens*. Phytochemistry **27:**2781–2785.

41. Stachel, S. E., E. Messens, M. Van Montagu, and P. Zambryski. 1985. Identification of the signal molecules produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. Nature (London) **318:**624–629.

42. Stachel, S. E., E. W. Nester, and P. C. Zambryski. 1986. A plant cell factor induces *Agrobacterium tumefaciens vir* gene expression. Proc. Natl. Acad. Sci. USA **83:**379–383.

43. Stachel, S. E., and P. C. Zambryski. 1986. *virA* and *virG* control the plant-induced activation of the T-DNA transfer process of *A. tumefaciens*. Cell **46:**325–333.

44. Steck, T. R., P. Morel, and C. I. Kado. 1988. Vir box sequences in *Agrobacterium tumefaciens* pTiC58 and A6. Nucleic Acids Res. **16:**8736.

45. Wahl, G. M., M. Stern, and G. R. Stark. 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76:**3683–3687.

46. Wakarchuk, W. W., N. M. Greenberg, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren. 1988. Structure and transcription analysis of the gene encoding a cellobiase from *Agrobacterium* sp. strain ATCC 21400. J. Bacteriol. **170:**301–307.

47. Wakarchuk, W. W., D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren. 1986. The molecular cloning and expression of a cellobiase gene from an *Agrobacterium* in *Escherichia coli*. Mol. Gen. Genet. **205:**146–152.

48. Ward, J. E., D. A. Akiyoshi, D. A. Regier, A. Datta, M. P. Gordon, and E. W. Nester. 1988. Characterization of the *virB* operon from an *Agrobacterium tumefaciens* Ti plasmid. Biol. Chem. **263:**5804–5814.

49. Wickner, W. T., and H. F. Lodish. 1985. Multiple mechanisms of protein insertion into and across membranes. Science **230:**400–407.

50. Wilson, K. 1990. Preparation of genomic DNA from bacteria, p. 2.4.1–2.4.5. *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. Wiley Interscience, New York.

51. Winans, S. C., P. R. Ebert, S. E. Stachel, M. P. Gordon, and E. W. Nester. 1986. A gene essential for *Agrobacterium* virulence is homologous to a family of positive regulatory loci. Proc. Natl. Acad. Sci. USA **83:**8278–8282.

52. Woodward, J., and A. Wiseman. 1982. Fungal and other β-D-glucosidases—their properties and applications. Enzyme Microb. Technol. **4:**73–79.

53. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.

54. Zerback, R., K. Dressler, and D. Hess. 1989. Flavonoid compounds from pollen and stigma of *Petunia hybrida*: inducers of the *vir* region of the *Agrobacterium tumefaciens* Ti plasmid. Plant Sci. **62:**83–91.