JOURNAL OF BACTERIOLOGY, Jan. 1988, p. 301–307
0021-9193/88/010301-07$02.00/0

Vol. 170, No. 1

Structure and Transcription Analysis of the Gene Encoding a Cellobiase from *Agrobacterium* sp. Strain ATCC 21400

W. W. WAKARCHUK, N. M. GREENBERG, D. G. KILBURN, R. C. MILLER, JR., AND R. A. J. WARREN*

Department of Microbiology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

Received 5 August 1987/Accepted 21 October 1987

The DNA sequence was determined for the cloned *Agrobacterium* sp. strain ATCC 21400 β-glucosidase gene, *abg*. High-resolution nuclease S1 protection studies were used to map the *abg* mRNA 5′ and 3′ termini. A putative *abg* promoter was identified whose sequence shows similarities to the consensus promoter of *Escherichia coli* and with the *nif* promoter regions of *Klebsiella*. The *abg* coding sequence was 1,374 nucleotides long. The molecular weight of the enzyme, based on the predicted amino acid sequence, was 51,000. The observed M_r was 50,000 to 52,000. A region of deduced protein sequence was homologous to a region from two other β-glucosidase sequences. This region of homology contained a putative active site by analogy with the active site of hen egg white lysozyme.

β-Glucosidase is one of the components required for the efficient enzymatic conversion of cellulose to glucose. The study of the enzymes involved in cellulose hydrolysis has been facilitated by the use of recombinant DNA technology (3). A number of glucanase genes, including several for β-glucosidases, have been cloned and expressed in both *Escherichia coli* and *Saccharomyces cerevisiae* (3, 15, 21, 33).

The nucleotide sequences of several glucanases have been determined (3). To date, only one β-glucosidase gene has been sequenced (12), but its transcription has not been analyzed. We report here on the sequencing and transcriptional analysis of the β-glucosidase gene (*abg*) of *Agrobacterium* sp. strain ATCC 21400. The *abg* gene encodes a β-glucosidase (Abg) with a very high specific activity and a high affinity for cellobiose (4, 30). This enzyme is interesting because β-glucosidase can be a limiting factor in cellulose hydrolysis. The availability of large amounts of very active enzyme would be very useful for the large-scale degradation of lignocellulosic materials. The *abg* gene has been expressed in *E. coli* to levels significantly higher than those observed in *Agrobacterium* spp. (31). Knowing the sequence of the gene should allow the development of improved forms of the enzyme.

The active sites of several β-1,4-glucanases were postulated to be analogous to that of egg white lysozyme after comparison of the positions of potential active site residues in the deduced or determined protein sequences of the glucanases with those in the active site of lysozyme (20, 23, 31). A similar arrangement of potential active site residues occurs in a sequence within the deduced amino acid sequence of Abg which is partially homologous to sequences within two other β-glucosidases.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages. Strain ATCC 21400 served as the source of the β-glucosidase gene. This organism was originally identified as *Alcaligenes faecalis* (9); however, the American Type Culture Collection (ATCC) has reclassified this isolate as an *Agrobacterium* species. The *E. coli* strains JM101, JM109, and JM83 were as described previously (32). The plasmids pUC13 and pUC18 were maintained in JM83 (18). The plasmids pTZ18U and pTZ19U were maintained in JM101 and have been described previously by J. Vieira (U.S. Biochemical company technical literature accompanying the vectors pTZ18/19, 1985). The phage vectors M13mp10/11 were maintained as phage preparations (18). JM101 and JM109 were maintained on M9 minimal medium plates (19). Plasmid-containing strains were grown in Luria broth (LB) (19) containing 100 μg of ampicillin per ml.

The plasmid pUC13::Δ9R5 was derived from pABG4F, which was as described previously (30). Plasmid pUC13::Δ9R5 contains 80 nucleotides (nt) of the *abg* coding sequence and 320 nt of 5′-flanking sequence and was used in

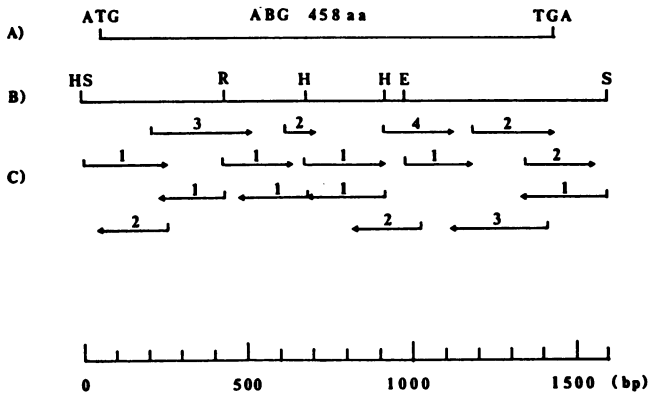

FIG. 1. Sequencing strategy for the *abg* gene. (A) Schematic showing the location of the *abg* coding region within the sequenced fragment. The initiation codon and putative termination codon are indicated. (B) Linear representation of the DNA insert in pABG5 showing the restriction endonuclease sites used in the cloning of specific fragments for sequencing. Abbreviations: E, *Eco*RI; H, *Hin*dIII; HS, hybrid *Sma*I site generated from ligating a blunt-ended exonuclease III deletion fragment into the *Sma*I site of pUC18; R, *Rsa*I; S, *Sal*I. (C) Summary of the sequences which were determined. The arrows denote the direction and lengths of each sequence. The numbers indicate the following: 1, sequence obtained with cloned restriction fragments; 2, sequence obtained with specific oligonucleotide primers; 3, sequence derived from deletion clones by the enzymatic method; 4, sequence derived by the chemical method.

* Corresponding author.

the S1 mapping experiments. Plasmid pABG5 contains the entire *abg* gene and was used for sequencing the gene.

Protein and peptide sequencing. The purification of Abg and the generation and purification of peptide fragments have been described (30). Amino acid sequences were determined with an Applied Biosystems model 470A gas-phase sequenator, utilizing the resident sequencing program. Amino acid compositions were determined with a Beckman 6300 amino acid analyzer.

DNA sequencing. DNA was sequenced by the enzymatic (27) and the chemical (17) methods. The vectors used to obtain single-stranded DNA were either M13mp10 and M13 mp11 or pTZ18U and pTZ19U. Subclones for sequencing were generated either by deletion with exonuclease III and mung bean nuclease (8) or by cloning of specific restriction fragments. Single-stranded M13 DNA template was prepared as described previously (18).

Single-stranded pTZ18/19 DNA was prepared as follows. A 2-ml culture of the recombinant clone was grown to an A_{650} of 0.5 to 0.9 at 37°C in LB medium with 200 μg of ampicillin per ml. After helper phage M13KO7 (J. Vieira, was added at multiplicity of infection of 10, the culture was diluted into 10 ml of fresh LB medium containing 70 μg of kanamycin per ml and allowed to grow for 16 to 24 h with vigorous shaking. The supernatants were then processed as described previously (18).

For chemical sequencing, restriction fragments were labeled with [α-^{32}P]dATP using the Klenow fragment of DNA polymerase I (16). Uniquely end-labeled restriction fragments were generated by a second restriction endonuclease digestion and were recovered from preparative agarose gels using DEAE paper as described previously (P. Lizardi, Binding and recovery of DNA and RNA using NA-45 DEAE membrane, Schleicher & Schuell application update, no. 364). Samples were analyzed on 6, 7, or 8% acrylamide gels (acrylamide-bisacrylamide [29:1]) containing 7 M urea and run as previously described (27). Autoradiography was done without intensifying screens for 4 to 24 h at room temperature with Kodak XRP-1 film. Sequences were analyzed with the SEQNCE program developed by Delaney Software

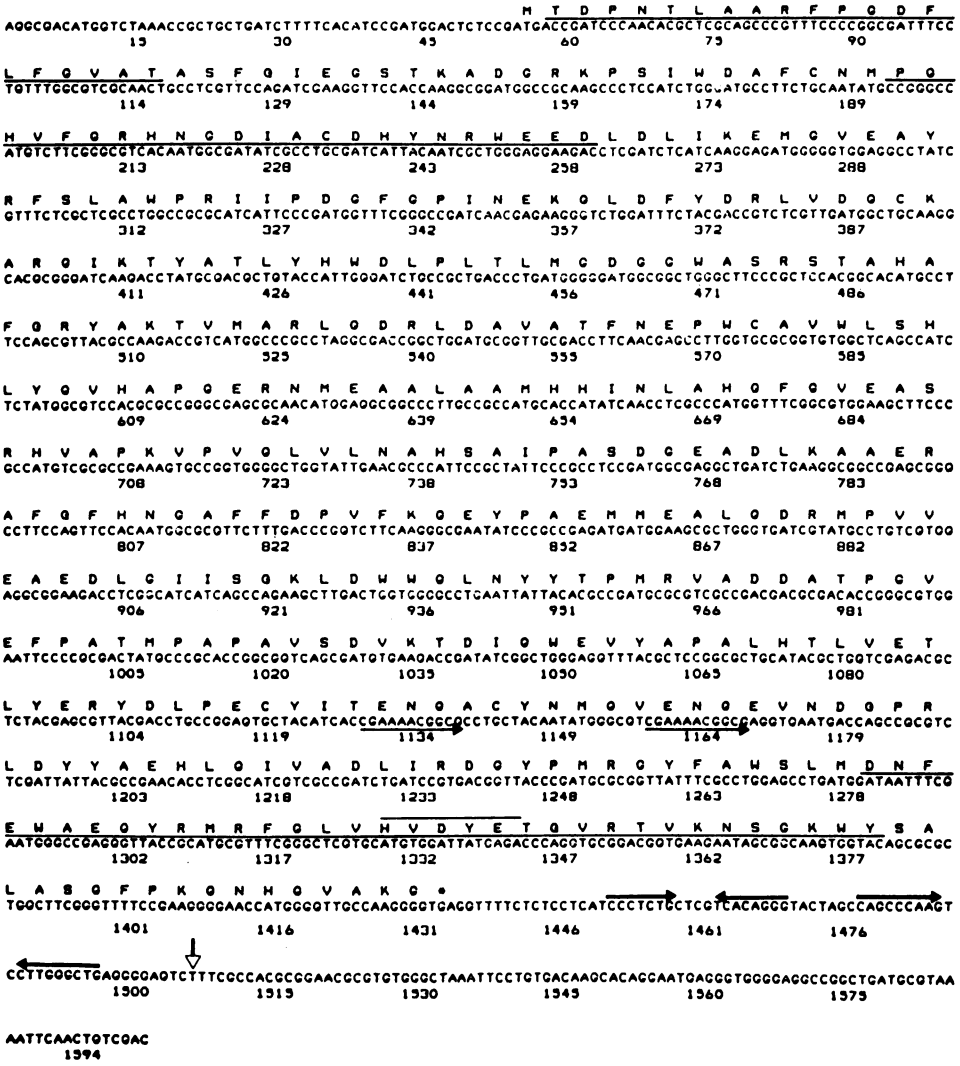

FIG. 2. Nucleic acid sequence and deduced amino acid sequence of the *abg* gene. Numbering starts at the first nonvector nucleotide after the hybrid *Sma*I site of pABG5. Peptide sequences obtained by amino acid sequencing are underlined. The oligonucleotide probe site used in the initial cloning is underlined. The 3′-terminal nucleotide of the transcript is indicated with the vertical arrow. Nucleotide sequences that may form stem loop structures are shown as opposing arrows. An 11-base-pair, directly repeated sequence is also indicated by arrows.

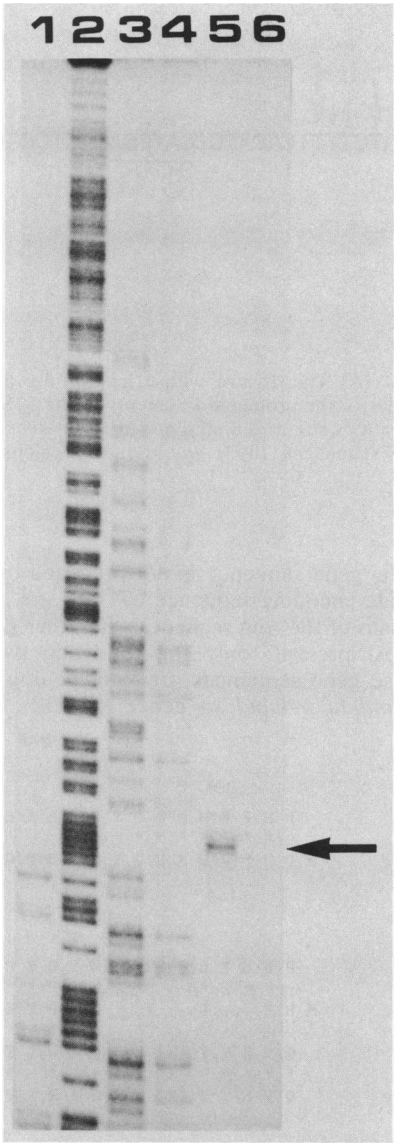

FIG. 3. Mapping the 5′ end of *abg* mRNA. After hybridization with RNA (lane 5, *Agrobacterium* RNA; lane 6, yeast tRNA), *abg*-specific ^{32}P-labeled *Sty*I-*Eco*RI probe (labeled at the 5′ *Sty*I end) was treated with S1 nuclease and analyzed on an 8% polyacrylamide–7 M urea sequencing gel alongside probe sequenced by the base-specific chemical cleavage method (17). Lanes 1 through 4 contain the sequence ladders G>A, G+A, T+C, and C>T, respectively. The arrow indicates the major protected species.

(Vancouver, British Columbia, Canada) or with the DNA Inspector II program (Textco, West Lebanon, N.H.).

RNA isolation and nuclease S1 protection studies. RNA isolation, Northern (RNA) blot analysis, and nuclease S1 protection studies were performed as previously described (6).

RESULTS AND DISCUSSION

Determination of the sequence. The strategy for the sequence determination is shown in Fig. 1. The sequence was determined by the enzymatic method. The chemical method was used to eliminate ambiguities in the region from nt 920 to 1070 and in S1 mapping experiments (see below). All restriction sites used in the subcloning of specific fragments of the insert in pABG5 were verified by sequencing through each of them. The 1,599-base-pair sequence of the *Agrobacterium* fragment from pABG5 is shown in Fig. 2. It starts at the first base after the 5′ half of the *Sma*I site of pUC18 and ends at the last nucleotide of the *Sal*I site.

Mapping the *abg* transcript initiation and termination sites with Sl nuclease. The coding region of the *abg* gene is 1,377 nt long. An intragenic fragment (nt 684 to 925 of the sequence (shown in Fig. 2) was used as a probe in a Northern blot analysis of *Agrobacterium* RNA to determine that the size of the in vivo *abg* transcript was approximately 1,500 nt (data not shown). Therefore, the *abg* mRNA is monocistronic.

Transcription initiation and termination sites were determined by high-resolution S1 nuclease mapping. A *Sty*I-*Eco*RI fragment labeled at the 5′ *Sty*I site was used in the 5′ analysis, whereas a *Sal*I-*Nco*I fragment labeled at the 3′ *Nco*I site was used in the 3′ analysis. Transcription initiated at several closely spaced sites (Fig. 3, lane 5). The major initiation site was a T residue 22 nt upstream of the translation initiation codon. There were two more major initiation sites at T residues 26 and 29 nt upstream of the ATG. A similar clustering of transcription initiation sites is seen in two cellulase-encoding genes of *Cellulomonas fimi* (6). Transcription terminated within a sequence of T residues, 71 nt downstream of the translational stop codon (data not shown). Therefore, the initiation and termination sites were about 1,480 nt apart, a finding in good agreement with the size of the mRNA (see above).

There are hexanucleotide sequences resembling the consensus *E. coli* promoter sequences (10) at appropriate distances upstream of the transcription initiation sites (Fig. 4B). The *abg* gene is expressed in *E. coli* regardless of its orientation with respect to the vector. This suggests very strongly that the hexanucleotide sequences do define the *abg* promoter. It will be interesting to see whether other *Agro-*

TABLE 1. Comparison of the amino acid composition of the β-glucosidase protein with the composition deduced from the *abg* gene sequence[a]

Amino acid	Composition deduced (no. of residues)	
	Protein[b]	DNA
Asx	54	51
Thr	19	19
Ser	19	16
Glx	40	34
Pro	29	26
Gly	48	43
Ala	54	53
Val	32	30
Met	15	15
Ile	14	16
Leu	36	34
Tyr	21	22
Phe	21	22
His	17	17
Lys	19	16
Arg	23	24
Trp	ND	13
Cys	ND	6

[a] Total residues were 460 (protein) and 458 (DNA). M_r values were 50,000 to 52,000 (protein) and 50,983 (DNA). Residues were calculated assuming an M_r of 50,000; M_r values were estimated from sodium dodecyl sulfate-polyacrylamide gel electrophoresis analyses.

[b] Average values from two determinations. ND, Not determined.

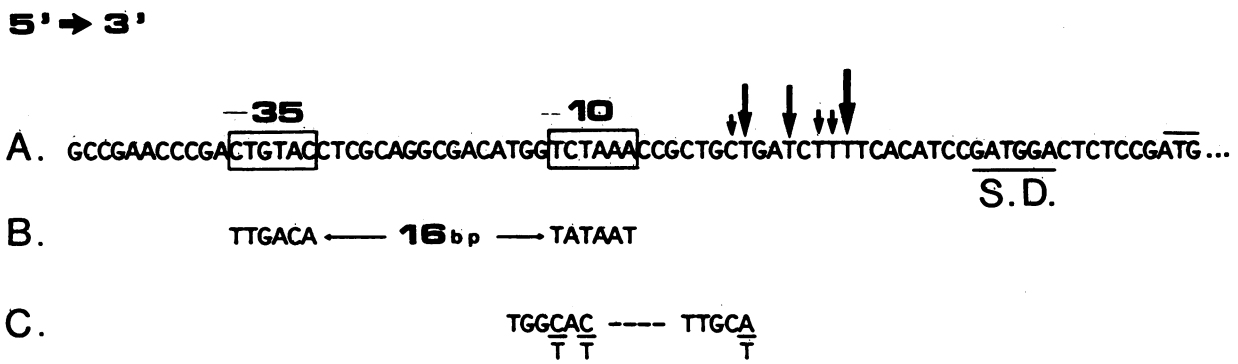

FIG. 4. DNA sequence corresponding to the 5'-terminal region of *abg* mRNA. (A) The arrows whose lengths are approximately proportional to the intensities of the bands in the gel in Fig. 3 denote the 3' nucleotides of the protected fragments of the 5' S1 probe. The ATG initiation codon is overlined. A putative Shine-Dalgarno-type ribosomal binding site (S.D.) (5) is underlined. The putative abg promoter −35 and −10 hexanucleotide sequences are boxed. (B) Consensus *E. coli* −35 and −10 sequences (10). (C) *nif* promoter regions of *Klebsiella* spp. (11).

bacterium promoters also function in *E. coli*. The putative *abg* promoter also resembles the *nif* promoters of *Klebsiella* spp. (11) (Fig. 4C). The influence of medium composition, especially of the nitrogen source, on Abg synthesis by *Agrobacterium* spp. is being determined. The transcription termination site is preceded by a sequence containing two inverted repeats which could form stem-loop structures (see Fig. 2) and which closely resemble the rho-independent transcription termination signals of *E. coli* (26).

Structure of the *abg* gene product. The deduced amino acid sequence is given in the one-letter amino acid code above the DNA sequence (Fig. 2). The translational start site of the gene was localized to the ATG codon at nt 54 by alignment with the amino terminal sequence of Abg. A protein of 458 amino acids (M_r, 50,983) could be produced by using the translation stop codon at nt 1431. This agrees with the observed M_r of 50,000 to 52,000 for the native and recombinant proteins (30).

The amino-terminal sequences of the protein and of the three CNBr peptides are underlined. The agreement of the determined peptide sequences with the deduced sequences is shown in Fig. 5. The differences between these sequences may be due to amino acid sequencing artifacts because of the small amounts of peptide material used for sequencing (R. Olaffson, personal communication). The peptide CNBr1 was only 10 residues long as deduced from the DNA sequence; however, the determined sequence was 20 residues. It is possible that the preparation contained a small quantity of an incomplete cleavage product containing the adjacent (downstream) CNBr peptide. This would have produced the correct sequence for the first 10 residues and then a very low level of the next amino acids, which then were incorrectly assigned because of high background signals.

The amino acid composition deduced for the protein and that determined for the native protein are shown in Table 1. The differences in the values of Ser, Glx, Gly, and Lys were quite high, perhaps because minor protein contaminants contributed significantly and disproportionately to the estimation of the amounts of the amino acids (D. McKay, personal communication).

It has been proposed that β-glucosidase in some bacteria and yeasts is a periplasmic protein (1, 12, 28). One characteristic structure associated with many periplasmic proteins and other secreted proteins is a leader sequence, which targets a protein for transport across membranes (24). However, the *abg* gene sequence is not preceded by a putative leader peptide-encoding sequence.

Comparisons of the Abg sequence with other β-glucosidase sequences. At present, only the following two complete β-glucosidase gene sequences are known: *abg* (this work) and the *Candida pelliculosa* gene, *cpb* (12). Part of the

```
CNBr1
aa sequence      D N F E W A E G Y R A L A L Q T Y A K T
                 * * * * * * * * * *
DNA sequence     M D N F E W A E G Y R M R F G L V H V Y D

CNBr2
aa sequence      P F G Y L(F) V(A) H V D Y E T
                 * * *  *    *    * * * * * *
DNA sequence     M R F G - L   V   H V D Y E T
aa sequence      Q V S T V [W] N S G - W L Y
                 * *   * *     * * *   *   *
DNA sequence     Q V R T V  K  N S G K W - Y

CNBr3
aa sequence      S G H V F G R [H] [N] G D I
                   * * * * * *  *   *  * * *
DNA sequence     M P G H V F G R  H   N  G D I
aa sequence      A P D H Y [ ] N Q/W [ ] E D
                 *   * * *       *      * *
DNA sequence     A C D H Y  N  R  W   E  E D

N-terminal
aa sequence      [T] D P N T L A A [A] F P G
                  *  * * * * * * *     * * *
DNA sequence      T  D P N T L A A  R  F P G
aa sequence      D F L F G V A T
                 * * * * * * * *
DNA sequence     D F L F G V A T
```

FIG. 5. Comparison of the amino acid sequences determined for the amino-terminal portion and three CNBr peptides from Abg and their sequences predicted from the DNA sequence. Symbols: ∗, identical residues; (X), residues obtained by amino acid sequencing but not by DNA sequencing; [X], residues which were uncertain by amino acid sequencing; [ ], a gap in the amino acid sequence; −, a residue missing from one of the sequences when both were compared.

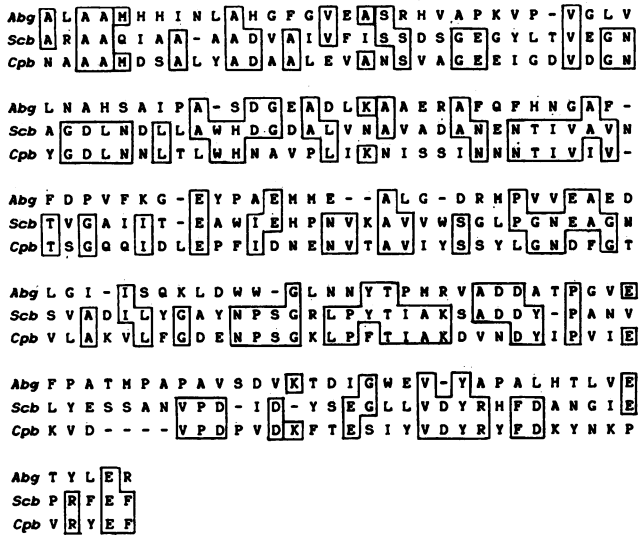

FIG. 6. Homologies in the deduced amino acid sequences of the β-glucosidases from *Agrobacterium* sp. (Abg), *S. commune* (Scb), and *C. pelliculosa* (Cpb). Homologous residues are boxed. The positions of the starting amino acids for each sequence are as follows: Abg, 195; Scb, 1; and Cpb, 499. Spaces have been introduced in the sequences to maximize the homologies.

sequence of the *Schizophyllum commune* β-glucosidase gene (*scb*) is also known (20). A comparison of these DNA sequences did not reveal any significant homologies (data not shown).

The deduced amino acid sequence of the *C. pelliculosa* β-glucosidase (Cpb) contains a sequence with 43% homology to the partial sequence deduced from the *S. commune* β-glucosidase (Scb) (21). A comparison was made between the homologous sequences from Cpb and Scb and a region from Abg (Fig. 6). The homology between the Abg and Cpb sequences was 11% (21 of 192 amino acids), and there were 10 conservative changes in the 192 amino acids compared. There was no significant homology between Abg and Cpb outside this region.

A putative active site for Spb was proposed by analogy with the active site of hen egg white lysozyme (21). Similar arrangements of putative active site residues occur in the predicted amino acid sequences of Cpb and Abg (Fig. 7). These putative active sites are in the homologous sequences described above (Fig. 6).

This arrangement of lysozymelike putative active site residues occurs in several glucanases: endoglucanases I and II from *S. commune* and cellobiohydrolase I from *Trichoderma reesei* (23), and the major endoglucanase (CenA) and an exoglucanase (Cex) from *Cellulomonas fimi* (31).

The active sites of two other β-glucosidases were investigated using the covalent inhibitor conduritol B epoxide (2, 14). Peptides which bound the inhibitor were isolated and sequenced. Those from the active sites of the *Aspergillus wentii* A_3 β-glucosidase (Awb) and the bitter almond β-glucosidase A (Bab) had little homology, although they both contained similar types of amino acids (Fig. 7). However, there was a sequence in Cpb, distinct from the sequence which contained the putative lysozymelike active site, that was almost perfectly homologous to the active site sequence from Awb (Fig. 7). This raises the interesting possibility that the Cpb protein may have more than one active site. There were no sequences in the predicted Abg or partial Scb sequences which were homologous to these Awb and Bab sequences.

The aspartic acid residue in Awb which bound conduritol B epoxide occurs in the sequence Ser/Thr-Asp-Try/Phe/Trp, which is similar to the sequence around the essential aspartic acid residue of hen egg white lysozyme. It is unfortunate that more of the amino acid sequences of Awb and Bab are not

Enzyme	AA	Active site sequence	Reference
HEWL	34	F K S N F N T Q A T N R N T D G S T D Y	i
Abg	310	V K F P A T M P A P - A V S D V K T D I	ii
Cex	33	S K F N L V V A E - N A M K W D A T E P	iii
EGI	32	N K S C A E F G N Q N I P G V K N T D Y	i
CBHI	64	N K T C A K N C C L D G A A Y A S T Y G	i
Scb	159	Y K S S A Q V P D I D Y S E G L L V D Y	iv
CenA	307	F K Y A V G F A L - N T S N Y Q T T A D	iii
Cpb	606	I K K V D V P D P V D K F T E S I T V D	iv
Awb		A Z L G F Z G F V M S D W	v
Cpb	259	Y L L N Y L L K E L G F Q G F V M T D W	iv
Bab		I T Z Z G V F G D S	vi

FIG. 7. Homology of putative active site sequences of various β-1,4-glucanases with the active site sequence of lysozyme. Abbreviations: AA, starting position in the amino acid sequence; HEWL, hen egg white lysozyme; Abg, *Agrobacterium* β-glucosidase; Cex, *Cellulomonas fimi* exoglucanase; EGI, *S. commune* endoglucanase I; CBHI, *T. reesei* cellobiohydrolase I; Scb, *S. commune* β-glucosidase; CenA, *Cellulomonas fimi* endoglucanase A; Cpb, *C. pelliculosa* β-glucosidase (sequence was aligned with lysozyme); Awb, *A. wentii* β-glucosidase A_3; Bab, bitter almond β-glucosidase A. The critical residues in the lysozyme sequence are circled. Spaces have been introduced into Abg, Exg, and EngA to allow alignment of the initial glutamic acid residue. Homologous residues are boxed; residues with conservative changes from the lysozyme sequence are underlined. Amino acids for Awb and Bab were determined by protein sequencing of a peptide identified as being involved in the active site of these proteins. Refer to the text for details. The sequence for Cpb was aligned with the *Aspergillus wentii* peptide. References: i, Paice et al. (23); ii, this study; iii, Warren et al. (31); iv, Morenelli et al. (20); v, Bause and Legler (2); vi, Legler and Harder (14).

TABLE 2. Codon usage in the *abg* gene[a]

AA	Codon	No.	AA	Codon	No.	AA	Codon	No.	AA	Codon	No.
Phe	UUU	3	Ser	UCU	0	Tyr	UAU	8	Cys	UGU	0
	UCU	19		UCC	7		UAC	14		UGC	6
Leu	UUA	0		UCA	0	Och	UAA	0	Opl	UGA	1
	UUG	1		UCG	3	Amb	UAG	0	Trp	UGG	13
	CUU	2	Pro	CCU	2	His	CAU	12	Arg	CGU	10
	CUC	13		CCC	8		CAC	5		CGC	11
	CUA	1		CCA	0	Gln	CAA	0		CGA	0
	CUG	17		CCG	16		CAG	6		CGG	3
Ile	AUU	2	Thr	ACU	2	Asn	AAU	9	Ser	AGU	0
	AUC	14		ACC	9		AAC	9		AGC	6
	AUA	0		ACA	1	Lys	AAA	1	Arg	AGA	0
Met	AUG	15		ACG	7		AAG	15		AGG	0
Val	GUU	4	Ala	GCU	6	Asp	GAU	22	Gly	GGU	9
	GUC	12		GCC	25		GAC	11		GGC	24
	GUA	1		GCA	5	Glu	GAA	11		GCA	0
	GUG	14		GCG	17		GAG	18		GGG	10

[a] AA, Amino acid; No., number of occurrences.

known so that a better comparison could be made with the lysozymelike active site sequence.

It is not surprising that these enzymes may have similar active sites since they all hydrolyze β-1,4-glycosidic linkages. The investigation of these active sites should now be facilitated by a combination of the ability to overexpress the cloned gene products, the identification of essential amino acids through the use of site-specific mutagenesis of the cloned genes, and the use of active site inhibitors.

Codon usage in the *abg* gene. Codon usage in a gene could be involved in the regulation of its expression (7, 25, 29). If efficient translation requires codon usage to be consistent with the cognate tRNA levels in the host, codon usage in a heterologous gene might affect its expression in *E. coli*.

Only 47 of the 61 codons are used in the *abg* gene (Table 2). There is a 79% bias to G or C in the third position of the codons used. The G+C contents of the first and second positions of the codons are 64.4 and 43.4%, respectively. The overall G+C content of the genus *Agrobacterium* is 59.6 to 62.8% (13), and the G+C content of the *abg* gene is 60.4%. The bias for G or C in the third position is higher than the 53% observed in *E. coli*, but is not as extreme as the bias observed in genes from organisms with higher G+C contents (22). Despite the bias for G or C in the third position of the codons, codon usage in *abg* is not very different from that seen in *E. coli* genes (data not shown).

ACKNOWLEDGMENTS

We thank D. McKay for the amino acid sequencing and the amino acid analysis.

This work was supported by a Program for Industrial Laboratory Projects grant from the Natural Sciences and Engineering Research Council of Canada to R.C.M., D.G.K., R.A.J.W., and Allelix, Inc.

LITERATURE CITED

1. **Ait, N., N. Creuzet, and J. Catteneo.** 1982. Properties of β-glucosidase purified from *Clostridium thermocellum*. J. Gen. Microbiol. **128**:569–577.
2. **Bause, E., and G. Legler.** 1980. Isolation and structure of a tryptic glycopeptide from the active site of β-glucosidase A_3 from *Aspergillus wentii*. Biochim. Biophys. Acta **626**:459–465.
3. **Beguin, P., N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., G. O'Neill, and R. A. J. Warren.** 1987. Cloning of cellulase genes. Crit. Rev. Biotechnol. **9**:129–162.
4. **Day, A. G., and S. G. Withers.** 1986. The purification and characterization of a β-glucosidase from *Alcaligenes faecalis*. Can. J. Biochem. Cell Biol. **64**:914–922.
5. **Gold, L., D. Pribnow, T. Schneider, S. Schinedling, P. Swebelius-Singer, and G. Storma.** 1981. Translation initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
6. **Greenberg, N. M., R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1987. Regulation, initiation, and termination of the *cenA* and *cex* transcripts of *Cellulomonas fimi*. J. Bacteriol. **169**: 646–653.
7. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon anti-codon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
8. **Guo, L.-H., and R. Wu.** 1983. Exonuclease III: use for DNA sequence analysis and in specific deletions of nucleotides. Methods Enzymol. **100**:468–500.
9. **Han, Y. W., and V. R. Srinivasan.** 1969. The purification and characterization of β-glucosidase of *Alcaligenes faecalis*. J. Bacteriol. **100**:1355–1363.
10. **Hawley, D. K., and W. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
11. **Kim, Y.-M., K.-J. Ahn, T. Beppu, and T. Vozumi.** 1986. Nucleotide sequence of the *nifLA* operon of *Klebsiella oxytoca* NG13 and characterization of the gene products. Mol. Gen. Genet. **205**:253–259.
12. **Kohchi, C., and A. Toh-e.** 1985. Nucleotide sequence of *Candida pelliculosa* β-glucosidase gene. Nucleic Acids Res. **13**: 6273–6282.
13. **Krieg, N. R.** 1974. Genus *Agrobacterium* Conn, p. 90–91. *In* R. E. Buchanan and N. E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore.
14. **Legler, G., and A. Harder.** 1978. Amino acid sequence at the active site of β-glucosidase A from bitter almonds. Biochim. Biophys. Acta **524**:102–108.
15. **Love, D. R., and M. B. Streiff.** 1987. Molecular cloning of a β-glucosidase gene from an extremely thermophilic anaerobe in *Escherichia coli* and *Bacillus subtilis*. Biotechnology **5**:384–387.
16. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
18. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101C**:20–78.
19. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Moranelli, F., J. R. Barbier, J. J. Dove, R. M. MacKay, V. L.**

Seligy, M. Yaguchi, and G. E. Willick. 1986. A clone coding for *Schizophyllum commune* β-glucosidase: homology with a yeast β-glucosidase. Biochem. Int. **12**:905–912.

21. **Nakamura, K., N. Misawa, and K. Kitamura.** 1986. Cellulase genes of *Cellulomonas uda* CB4. I. Cloning and expression of β-glucosidase genes in *Escherichia coli*. J. Biotechnol. **3**:239–246.
22. **O'Neill, G. P., S. H. Goh, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene **44**:325–330.
23. **Paice, M. G., M. Desrochers, D. Rho, L. Jurasek, C. Roy, C. F. Rollin, E. D. Miguel, and M. Yaguchi.** 1984. Two forms of endoglucanase from the basidiomycete *Schizophyllum commune* and their relationship to other β-1,4-glycoside hydrolases. Bio/Technology **2**:535–539.
24. **Pugsley, A. P., and M. Schwartz.** 1985. Export and secretion of proteins by bacteria. FEMS Microbiol. Rev. **32**:3–38.
25. **Robinson, M., R. Lilley, J. S. Emtage, G. Yarranton, P. Stephens, A. Millican, M. Eaton, and G. Humphreys.** 1984. Codon usage can affect efficiency of translation in *Escherichia coli*. Nucleic Acids Res. **12**:6663–6671.
26. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Stoppok, W., P. Rapp, and F. Wagner.** 1982. Formation, location, and regulation of endo-1,4-β-glucanases and β-glucosidases from *Cellulomonas uda*. Appl. Environ. Microbiol. **44**:44–53.
29. **Varrene, S., J. Buc, R. Lloubes, and C. Lazdunski.** 1984. Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. **180**:549–576.
30. **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.
31. **Warren, R. A. J., C. G. Beck, N. R. Gilkes, D. G. Kilburn, M. L. Langsford, R. C. Miller, Jr., G. P. O'Neill, and W. K. R. Wong.** 1987. Sequence conservation and region shuffling in two cellulases from *Cellulomonas fimi*. Proteins **1**:335–341.
32. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
33. **Zappe, H., D. T. Jones, and D. R. Woods.** 1986. Cloning and expression of *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *Escherichia coli*. J. Gen. Microbiol. **132**:1376–1372.