JOURNAL OF BACTERIOLOGY, Feb. 1992, p. 765–777
0021-9193/92/030765-13$02.00/0
Copyright © 1992, American Society for Microbiology

Vol. 174, No. 3

Nucleotide Sequences of the *arb* Genes, Which Control β-Glucoside Utilization in *Erwinia chrysanthemi*: Comparison with the *Escherichia coli bgl* Operon and Evidence for a New β-Glycohydrolase Family Including Enzymes from Eubacteria, Archeabacteria, and Humans

MOHAMMED EL HASSOUNI,[1] BERNARD HENRISSAT,[2] MARC CHIPPAUX,[1] AND FREDERIC BARRAS[1]*

Laboratoire de Chimie Bactérienne, Centre National de la Recherche Scientifique, 31 Chemin Joseph Aiguier, BP71, 13274, Marseille, Cedex 9,[1] and Centre de Recherches sur les Macromolécules Végétales, Centre National de la Recherche Scientifique, F-38041, Grenoble,[2] France

Received 6 August 1991/Accepted 22 November 1991

The phytopathogenic bacterium *Erwinia chrysanthemi*, unlike other members of the family *Enterobacteriaceae*, is able to metabolize the β-glucosides, arbutin, and salicin. A previous genetic analysis of the *E. chrysanthemi arb* genes, which mediate β-glucoside metabolism, suggested that they were homologous to the *Escherichia coli* K-12 *bgl* genes. We have now determined the nucleotide sequence of a 5,065-bp DNA fragment containing three genes, *arbG*, *arbF*, and *arbB*. Deletion analysis, expression in minicell systems, and comparison with sequences of other proteins suggest that *arbF* and *arbB* encode a β-glucoside-specific phosphotransferase system-dependent permease and a phospho-β-glucosidase, respectively. The ArbF amino acid sequence shares 55% identity with that of the *E. coli* BglF permease and contains most residues thought to be important for a phosphotransferase. One change, however, was noted, since BglF Arg-625, presumably involved in phosphoryl transfer, was replaced by a Cys residue in ArbF. An analysis of the ArbB sequence led to the definition of a protein family which contained enzymes classified as phospho-β-glucosidases, phospho-β-galactosidases, β-glucosidases, and β-galactosidases and originating from gram-positive and gram-negative bacteria, archebacteria, and mammals, including humans. An analysis of this family allowed us (i) to speculate on the ways that these enzymes evolved, (ii) to identify a glutamate residue likely to be a key amino acid in the catalytic activity of each protein, and (iii) to predict that domain II of the human lactate-phlorizin hydrolase, which is involved in lactose intolerance, is catalytically nonactive. A comparison between the untranslated regions of the *E. chrysanthemi arb* cluster and the *E. coli bgl* operon revealed the conservation of two regions which, in the latter, are known to terminate transcription under noninducing conditions and be the target of the BglG transcriptional antiterminator under inducing conditions. ArbG was found to share a high level of similarity with the BglG antiterminator as well as with *Bacillus subtilis* SacT and SacY antiterminators, suggesting that ArbG functions as an antiterminator in regulating the expression of the *E. chrysanthemi arb* genes.

Members of the family *Enterobacteriaceae* differ in their capacity to ferment the natural β-glucosides, i.e., cellobiose, arbutin, and salicin, as carbon sources. For instance, *Erwinia chrysanthemi*, a phytopathogenic bacterium (7), is capable of fermenting all three, while *Escherichia coli* K-12 strains ferment none (4, 44, 45). However, *E. coli* K-12 possesses a battery of silent genes which, upon mutagenesis, can be activated and enable the use of these sugars (15, 24, 33, 34, 36, 37, 44). Hence, a comparative analysis of the *E. coli* and *E. chrysanthemi* systems might provide an opportunity for describing evolutionary constraints exerted upon cryptic and expressed genes.

Our previous analysis revealed the existence of at least two unrelated β-glucoside assimilatory pathways in *E. chrysanthemi* (4, 16). One, referred to as the *clb* pathway, enables the bacterium to use cellobiose, arbutin, and salicin, while another, referred to as the *arb* pathway, allows growth on arbutin and salicin only. It is unknown whether the *clb* genes resemble the *E. coli* cryptic *cel* genes, which also control a cellobiose assimilatory pathway (34). Results from a genetic analysis of the *E. chrysanthemi arb* system led us to propose that it is homologous to the *E. coli* cryptic *bgl* operon (16).

E. coli bgl operon expression is regulated by antitermination (1, 2, 20, 26, 27, 46). Three genes, *bglG*, *bglF*, and *bglB*, which encode an antiterminator, a β-glucoside-specific phosphotransferase system-dependent permease, and a phospho-β-glucosidase, respectively, are present (48). The *bglG* gene is bracketed by two terminator structures which are essential to the regulation of *bgl* expression. In the absence of exogenous β-glucoside, most of the transcripts terminate at the promoter-proximal terminator (t_1), the distal one (t_2) aborting those transcripts that continued through the former. Under these conditions, the BglG protein is phosphorylated by the BglF permease and, as such, is not active. In the presence of β-glucosides, the BglF permease phosphorylates the sugar and dephosphorylates BglG, which then acts at the t_1 and t_2 sites, antagonizing the termination process (1, 2).

Previous genetic characterization of the *E. chrysanthemi arb* system suggested that (i) it contained three genes, *arbG*, *arbF*, and *arbB*, all transcribed in the same direction, and (ii) the *arbF* and *arbB* genes were equivalent to the *E. coli bglF* and *bglB* genes, i.e., they coded, in a substrate-inducible

* Corresponding author.

manner, for a PTS-dependent permease and a phospho-β-glucosidase, respectively. In this paper, we pursue the characterization of the Arb system at the nucleotide and amino acid sequence levels and compare this system with the *E. coli* cryptic *bgl* operon.

MATERIALS AND METHODS

Bacteria and plasmids. The *E. coli* K-12 derivative strains were as follows: LCB320 (C600 *thr leu rpsL*; from our collection); TG1 [Δ(*lac-pro*) *supE thi hsdR hsdM*/F′ *lacI lacZ*Δ*M15 proAB*]; AE 10 (Δ*lacX74 thi bglR11 tsx tna*::Tn*10*; from A. Wright); AE 304-3 (same as AE 10 but *bglB*; from A. Wright); and AR 1062 (*thr leu thi ara azi fhuA lacY tsx minA rpsL gal mtl xyl hsdR*; from our collection). The plasmids used were as follows: pEM31 (ColE1 Ap^r $arbG^+$ $arbF^+$ $arbB^+$) was the parental plasmid (see Fig. 1); Bluescript phagemid pKS+ (Ap^r; Stratagene) was used to subclone the $PstI_2$-*Sal*I restriction fragment from pEM31, giving rise to the pBS11 (Ap^r $arbF^+$ $arbB^+$) and pBS23 (Ap^r $arbF^+$ $arbB^+$) plasmids; plasmid pBSΔCI ($arbF^+$ Ap^r) was derived from pBS11 by deletion of the $ClaI_1$-$ClaI_2$ restriction fragment; pBS10.3 (Ap^r) was a pBS11 deletion derivative obtained during the generation of exonuclease III-mediated deletions; and pBST3 and pBST7 plasmids were pKS+ and pKS− derivatives, respectively, carrying the $BamHI_1$-*Pvu*II restriction fragment from pEM31 (see Fig. 1).

Media and chemicals. All conditions used for growing, scoring, and selecting bacteria were as already described (16).

DNA techniques. Plasmid DNA preparation, electrophoresis, and DNA fragment isolation were performed by standard procedures (3, 28). Enzymatic treatments of DNA molecules were carried out as recommended by the manufacturers. Bacterial transformation was performed as described by Chung and Miller (8).

Expression in minicell systems and electrophoretic analysis of plasmid-encoded proteins. The plasmids of interest were introduced into the minicells producing *E. coli* AR 1062. Minicells were purified by the method of Meagher (30). Strain AR 1062 carrying the analyzed plasmid was grown in minimal medium containing glycerol (0.2%), Casamino Acids (0.2%), and ampicillin (50 μg/ml) at 37°C overnight. Minicells were separated by centrifugation (4°C, 5,000 rpm, 5 min), subsequently pelleted by a second centrifugation (4°C, 10,000 rpm, 20 min), and finally resuspended in 1 ml of M9 medium (31). This minicell suspension was loaded on a continuous (0 to 23%) glycerol gradient and centrifuged (4°C, 4,500 rpm, 30 min); the minicells forming a dense band midway in the tube were removed, centrifuged, and resuspended in M9 medium as described above. These steps were done as many times as required to obtain pure minicells, as detected by microscopic observation. Minicells were concentrated to 100 μl and either used for labelling or stored in M9 medium with glycerol (50%) at −80°C. For labelling, 50 μl of the minicell suspension was complemented with 4 μl of vitamin B_1 (0.5 mg/ml), 4 μl of methionine assay medium, which contains all amino acids but methionine, and glycerol (0.2%). The suspension was incubated at 37°C for 20 min, supplemented with 5 μl of ^{35}S-methionine (1 μCi/ml), and incubated for an additional 30 min at 37°C. After centrifugation and two washings with M9 medium, the pellets were suspended in loading buffer (Tris-HCl, 100 mM, pH 8.8; EDTA, 2.5 mM; sucrose, 0.5 M; bromophenol blue, 0.005%; sodium dodecyl sulfate [SDS], 18%; dithiothreitol, 0.3 M; β-mercaptoethanol, 5%), heated to 95°C for 10 min, returned to ice, and stored at −20°C. Proteins were subsequently separated by SDS-polyacrylamide gel electrophoresis with a 12% polyacrylamide gradient gel (Pharmacia PhastSystem). Protein bands were visualized by autoradiography with Kodak XAR-5 film.

Cloning and nested deletions. Nested deletions were made in the Bluescript phagemid derivatives carrying DNA to be sequenced, namely, pBS11, pBS23, pBST7, and pBST3. Overlapping deletions were generated with exonuclease III (10 U/μg of DNA, 26°C), which was allowed to act for various periods of time. Treatment with exonuclease VII (1.2 U/μg of DNA, 45 min, 37°C) was performed to remove the protruding ends. Finally, the DNA fragment extremities were treated with the Klenow enzyme (2 U/μg of DNA) and ligated with T4 DNA ligase, and the resulting plasmids were introduced into *E. coli* TG1. The sizes of the recovered plasmids were analyzed by electrophoresis on agarose gels (1%).

Determination of the nucleotide sequence. DNA sequencing was carried out by the dideoxy chain termination method of Sanger et al. (42) with Sequenase (Genofit or USB) and [α-^{35}S]dATP. The reaction products were separated at 65°C on a gradient gel (0.2 to 0.4 mm) made with 4% polyacrylamide, 7 M urea, and Tris-borate buffer. Gels were autoradiographed by exposure to Kodak XAR-5 film. The location of Mu-*lacZ* insertions was determined by direct sequencing of the pBH324 and pBH363 plasmids (16) with an oligonucleotide (5′-TTTTTCGTGCGCCGCTT-3′) derived from the S extremity of phage Mu.

Nucleotide sequence accession number. The GenBank nucleotide sequence accession number for the *E. chrysanthemi arb* genes is M81772.

RESULTS AND DISCUSSION

Cloning and characterization of the *arb* genes. Insertion mutagenesis of the pEM31 plasmid had suggested that the *arb* genes were located between the *Pvu*II and $ClaI_2$ restriction sites, with transcription reading from left to right (Fig. 1). The slightly larger $PstI_2$-*Sal*I restriction fragment was inserted at the *Sma*I restriction site of the pKS+ plasmid, and the resulting plasmids were used to transform *E. coli* LCB320. The transformants yielded three colony types on MacConkey plates containing arbutin and ampicillin: white colonies most likely carrying the recircularized vector; dark red colonies carrying a plasmid, referred to as pBS11, in which the *arb* genes were inserted in the same orientation as the *lac* promoter present on the vector; and pink colonies containing a plasmid, referred to as pBS23, in which the *arb* genes were inserted in the opposite orientation. Nested deletions were made from the 3′ end of the inserts carried by the pBS11 and pBS23 plasmids. The various pBS11 deletion derivatives were analyzed for the phenotype that they conferred to *E. coli* TG1 (Fig. 1). All deletions ending around the $ClaI_1$ restriction site conferred an Arb^+ Sal^- phenotype. Thus, these plasmids carry a functional permease gene which allows the uptake of β-glucosides but lack the phospho-β-glucosidase gene. Phosphoarbutin is hydrolyzed in these cells by the product of the *E. coli bglA* gene (16, 43). All deletions extending upstream of the $ClaI_1$ restriction site conferred an Arb^- Sal^- phenotype.

The proteins encoded by two pBS11 deletion derivatives, pBS11ΔCI and pBS10.3, were analyzed in an *E. coli* minicell system (Fig. 2). Compared with the Bluescript phagemid, the pBS11 plasmid was found to encode two additional proteins with approximate masses of 52 and 67 kDa. Elimination of

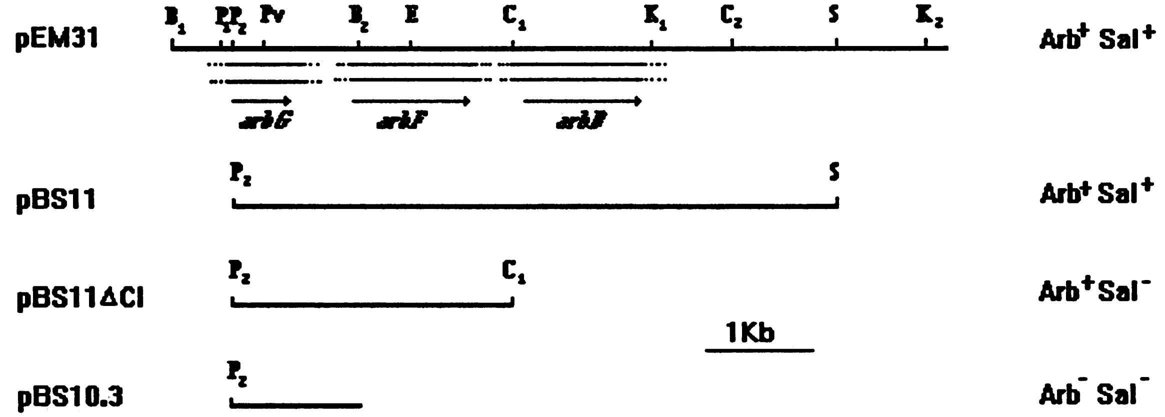


FIG. 1. Restriction map of the insert carried by the pEM31 plasmid and derivatives. On the right are shown the phenotypes conferred by the plasmids to a wild-type strain of *E. coli* K-12. The names, order, and direction of transcription of the *arb* genes were determined in a previous analysis (16). The pBS10.3 plasmid was obtained during the generation of nested deletions, while the pBS11ΔCI plasmid resulted from $ClaI_1$-$ClaI_2$ deletion. The vectors are not represented. B, C, E, K, P, Pv, and S stand for *Bam*HI, *Cla*I, *Eco*RI, *Kpn*I, *Pst*I, *Pvu*II, and *Sal*I, respectively.

the DNA region downstream of the *Cla*I$_1$ restriction site, e.g., in plasmid pBS11ΔCI, apparently led to the loss of the gene encoding the 52-kDa protein, while larger deletions led to the disappearance of the 67-kDa protein (Fig. 2). Taken together with the phenotypic characterization mentioned above, these data suggested that the proteins of 67 and 52 kDa are the permease and the phospho-β-glucosidase, respectively.

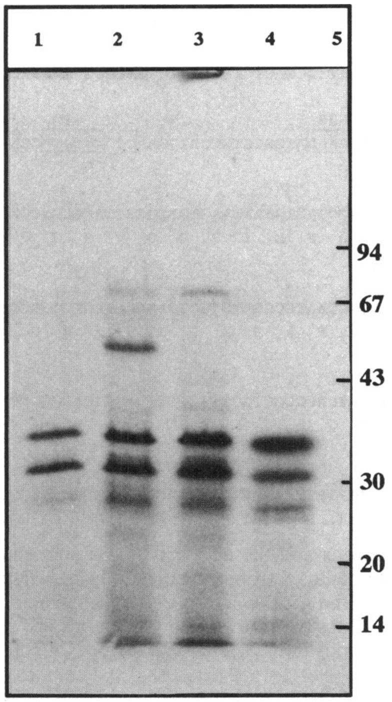


FIG. 2. Identification of two proteins encoded by the pBS11 plasmid and its deletion derivatives in an *E. coli* minicell system. Plasmids carried by strain AR 1062 and directing the synthesis of the ^{35}S-labelled proteins analyzed were as follows: lane 1, Bluescript; lane 2, pBS11; lane 3, pBS11ΔCI; and lane 4, pBS10.3. Lane 5 contains size markers (in kilodaltons). Details about plasmid constructions are given in the legend to Fig. 1.

Sequence analysis of the *arb* genes. As more precisely described below, previous results from mini-Mu-mediated mutagenesis experiments led to inaccurate localization of the upstream limit of the first *arb* gene. Therefore, the determination of the complete nucleotide sequence required the construction of a second pair of plasmids: the *Bam*HI$_1$-*Pvu*II restriction fragment was subcloned into both pKS+ and pKS− phagemids, giving rise to pBST7 and pBST3, respectively (Fig. 1). Hence, the 5,065-bp nucleotide sequence given in Fig. 3 starts at the *Bam*HI$_1$ site and ends 435 bp downstream of the *Kpn*I site (Fig. 1). Three large open reading frames (ORF), presumably corresponding to the three genes previously referred to as *arbG*, *arbF*, and *arbB*, were identified. The first ORF starts at the ATG at position 201, is preceded by a motif (AGGGGT) resembling the Shine-Dalgarno (SD) sequence located 6 bp upstream, and ends at bp 1049, thus containing 283 codons, corresponding to a protein with an M_r of 31,130. For the second ORF, there are three potential ATG initiator codons at positions 1161, 1269, and 1347. Only the last is preceded at a distance of 5 bp by a sequence, AGGATGT, resembling the consensus SD sequence. Therefore, we assume that the ATG at position 1347 is the correct initiator codon. A subsequent sequence comparison with other permeases yielded results consistent with such a hypothesis (see below). There is a stop codon at position 3239. This ORF is 631 codons long and could encode a protein with an M_r of 69,410. For the third ORF, there are two possible ATG initiator codons at positions 3272 and 3275, but there is a single consensus SD sequence, AGGAGAT, upstream at distances of 4 and 7 bp, respectively. This ORF ends at position 4669, contains 465 (or 466) codons, and could encode a protein with an M_r of 53,000. The beginning of a fourth ORF, starting with the ATG at position 4805, was identified. It is preceded by an SD-like sequence, GGAACGT, located at a distance of 4 bp. We collected evidence neither for the presence of a protein that could correspond to this ORF nor for a role of this region in the functioning of the Arb system. However, we should mention that some Lac$^+$ fusions whose preliminary mapping was compatible with an insertion in this ORF were obtained. The beginning of a similar apparently "dispensable" ORF was also identified in the *E. coli bgl* operon (48). Since the N-terminal sequences of both ORF share some sequence

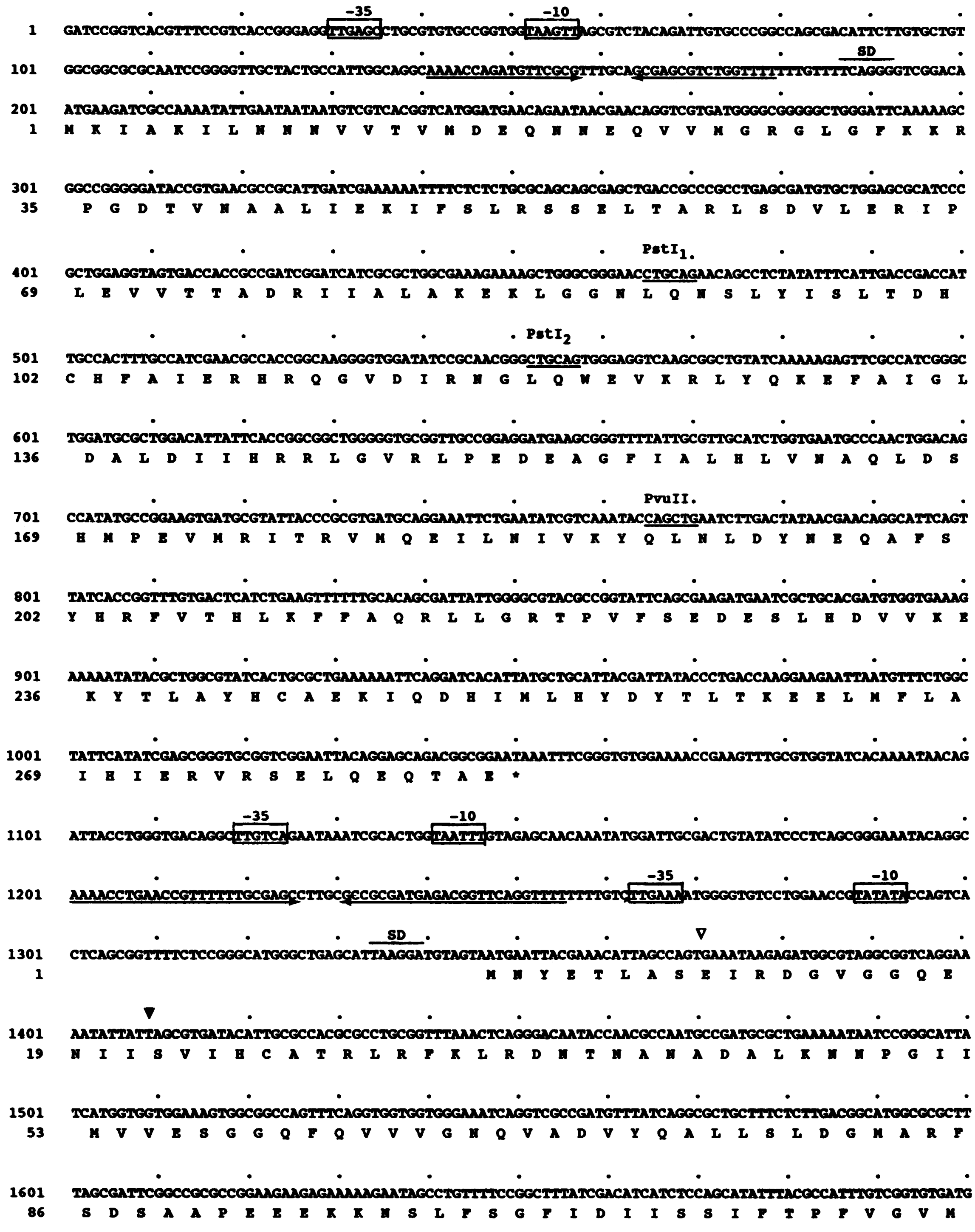

-35
-10
SD
PstI1
PstI2
PvuII
1 M K I A K I L N N N V V T V M D E Q N N E Q V V M G R G L G F K K R
35 P G D T V N A A L I E K I F S L R S S E L T A R L S D V L E R I P
69 L E V V T T A D R I I A L A K E K L G G N L Q N S L Y I S L T D H
102 C H F A I E R H R Q G V D I R N G L Q W E V K R L Y Q K E F A I G L
136 D A L D I I H R R L G V R L P E D E A G F I A L H L V N A Q L D S
169 H M P E V M R I T R V M Q E I L N I V K Y Q L N L D Y N E Q A F S
202 Y H R F V T H L K F F A Q R L L G R T P V F S E D E S L H D V V K E
236 K Y T L A Y H C A E K I Q D H I M L H Y D Y T L T K E E L M F L A
269 I H I E R V R S E L Q E Q T A E *
-35
-10
-35
-10
SD
1 M N Y E T L A S E I R D G V G G Q E
19 N I I S V I H C A T R L R F K L R D N T N A N A D A L K N N P G I I
53 M V V E S G G Q F Q V V V G N Q V A D V Y Q A L L S L D G M A R F
86 S D S A A P E E E K K N S L F S G F I D I I S S I F T P P V G V M

FIG. 3. Nucleotide sequence of the *E. chrysanthemi arb* genes and the flanking regions and the deduced amino acid sequences of the ArbG, ArbF, and ArbB proteins and of the beginning of an unknown ORF. Amino acids deduced from the nucleotide sequence are specified by one-letter representation, and each amino acid is presented under the first nucleotide of the corresponding codon. The nucleotide sequence was determined on both strands. Putative σ^{70}-recognized promoter regions are boxed. Putative ribosome binding sites are indicated (SD). Candidates for rho-independent terminator structures are shown by inverted arrows. The positions of Mu-*lacZ* fusion 324 and 363 insertions in the *arbF* gene are shown by filled and open triangles, respectively (16).

similarity (data not shown), new investigations are required to evaluate the role, if any, of these genes.

Comparison of the ArbF amino acid sequence with those of other EII permeases. Saier et al. (39, 40) compared PTS-dependent EII permeases and identified several typical motifs, such as two phosphorylation sites, each containing a histidyl residue, the presence of an amphipathic helical segment located at the N terminus of the protein, and the

```
                .BamHI2   .         .         .         .         .         .         .         .
1701  GCGGCAACGGGGATCCTGAAAGGTTTTCTGGCGCTAGGCGTCGCCACCCATGTGATATCGGAAAGCAGCGGCACCTATAAATTGCTGTTCGCCGCCAGCG
 119   A  A  T  G  I  L  K  G  F  L  A  L  G  V  A  T  H  V  I  S  E  S  S  G  T  Y  K  L  L  F  A  A  S  D

               .         .         .         .         .         .         .         .         .         .
1801  ACGCGCTGTTCTATTTCTTCCCCATTGTGCTGGGCTATACGGCCGGCAAGAAGTTTGGCGGCAACCCATTTACCACGCTGGTGATTGGCGCCACGCTGGT
 153     A  L  F  Y  F  F  P  I  V  L  G  Y  T  A  G  K  K  F  G  G  N  P  F  T  T  L  V  I  G  A  T  L  V

               .         .         .         .         .         .         .         .         .         .
1901  GCATCCGAGCATGATCGCCGCTTTCAACGCCATGCAGGCGCCGGATCACTCAACGCTGCATTTTCTGGGTATTCCAATTACTTTTATCAATTACAGCTCC
 186    H  P  S  M  I  A  A  F  N  A  M  Q  A  P  D  H  S  T  L  H  F  L  G  I  P  I  T  F  I  N  Y  S  S

               .         .         .         .         .         .         .         .         .         .
2001  TCGGTCATTCCGATTCTGTTTGCCAGTTGGGTATCCTGCAAACTGGAAAAACCGCTGAATCGCTGGCTGCACGCCAATATCCGCAATTTCTTCACGCCGC
 219   S  V  I  P  I  L  F  A  S  W  V  S  C  K  L  E  K  P  L  N  R  W  L  H  A  N  I  R  N  F  F  T  P  L

               .         .         .         .         .         .         .         .         .         .
2101  TGCTGTGTATTGTTATTAGCGTTCCGTTGACCTTTCTGCTGATTGGGCCGAGCGCTACCTGGCTGAGCCAGATGCTGGCGGGCGGATACCAGTGGCTGTA
 253     L  C  I  V  I  S  V  P  L  T  F  L  L  I  G  P  S  A  T  W  L  S  Q  M  L  A  G  G  Y  Q  W  L  Y

           EcoRI         .         .         .         .         .         .         .         .         .
2201  CGGGTTGAATTCATTGCTGGCTGGCGCCGTGATGGGCGCGTTGTGGCAGGTATGCGTGATTTTCGGGTTGCACTGGGGCTTTGTGCCGCTGATGCTGAAT
 286    G  L  N  S  L  L  A  G  A  V  M  G  A  L  W  Q  V  C  V  I  F  G  L  H  W  G  F  V  P  L  M  L  N

               .         .         .         .         .         .         .         .         .         .
2301  AATTTCAGTGTGATCGGACACGATACACTGCTGCCGCTGCTGGTGCCGGCGGTGCTGGGGCAGGCCGGCGCCACGCTGGGCGTGCTGTTGCGTACCCAGG
 319   N  F  S  V  I  G  H  D  T  L  L  P  L  L  V  P  A  V  L  G  Q  A  G  A  T  L  G  V  L  L  R  T  Q  D

               .         .         .         .         .         .         .         .         .         .
2401  ACCTGAAGCGCAAGGGGATTGCCGGGTCGGCGTTTTCCGCGGCGATTTTCGGGATTACCGAGCCAGCGGTATACGGCGTGACGCTGCCGCTGCGTCGTCC
 353     L  K  R  K  G  I  A  G  S  A  F  S  A  A  I  F  G  I  T  E  P  A  V  Y  G  V  T  L  P  L  R  R  P

               .         .         .         .         .         .         .         .         .         .
2501  CTTTATCTTCGGCTGTATCGGCGGGGCGCTGGGCGCGGCGGTGATGGGATATGCTCATACCACCATGTATTCGTTCGGTTTTCCCAGCATTTTCTCCTTT
 386    F  I  F  G  C  I  G  G  A  L  G  A  A  V  M  G  Y  A  H  T  T  M  Y  S  F  G  F  P  S  I  F  S  F

               .         .         .         .         .         .         .         .         .         .
2601  ACCCAGGTGATTCCGCCGACCGGCGTGGACAGCAGCGTTTGGGCGGCGGTGATCGGCACGCTGTTGGCCTTTGCGTTTGCTGCGTTGACCAGTTGGTCGT
 419   T  Q  V  I  P  P  T  G  V  D  S  S  V  W  A  A  V  I  G  T  L  L  A  F  A  F  A  A  L  T  S  W  S  F

               .         .         .         .         .         .         .         .         .         .
2701  TTGGCGTGCCGAAAGATGAAACGCAACCGGCAGCGGCGGATAGTCCGGCGGTACTGGCGGAAACACAGGCTAACGCTGGCGCTGTTCGTGATGAGACGTT
 453     G  V  P  K  D  E  T  Q  P  A  A  A  D  S  P  A  V  L  A  E  T  Q  A  N  A  G  A  V  R  D  E  T  L

               .         .         .         .         .         .         .         .         .         .
2801  GTTCAGTCCGCTGGCCGGTGAGGTACTGCTGCTGGAGCAGGTGGCCGATCGTACCTTTGCCAGCGGCGTGATGGGCAAAGGGATCGCCATTCGACCTACG
 486    F  S  P  L  A  G  E  V  L  L  L  E  Q  V  A  D  R  T  F  A  S  G  V  M  G  K  G  I  A  I  R  P  T

               .         .         .         .         .         .         .         .         .         .
2901  CAGGGGCGGCTGTACGCGCCGGTAGACGGCACCGTGGCGTCGCTGTTTAAAACCCATCATGCCATTGGCCTGGCATCGCGAGGCGGAGCGGAGGTGCTGA
 519   Q  G  R  L  Y  A  P  V  D  G  T  V  A  S  L  F  K  T  H  H  A  I  G  L  A  S  R  G  G  A  E  V  L  I

               .         .         .         .         .         .         .         .         .         .
3001  TACACGTCGGCATCGACACCGTCCGGCTGGATGGCCGTTATTTTACCCCGCACGTGCGTGTCGGCGATGTGGTGCGCCAGGGCGACCTGCTGCTGGAATT
 553     H  V  G  I  D  T  V  R  L  D  G  R  Y  F  T  P  H  V  R  V  G  D  V  V  R  Q  G  D  L  L  L  E  F

               .         .         .         .         .         .         .         .         .         .
3101  TGATGGCCCGGCCATTGAGGCGGCAGGCTATGACCTCACCACGCCGATTGTGATTACCAACAGCGAAGACTATCGCGGGGTTGAACCCGTCGCCAGCGGC
 586    D  G  P  A  I  E  A  A  G  Y  D  L  T  T  P  I  V  I  T  N  S  E  D  Y  R  G  V  E  P  V  A  S  G

               .         .         .         .         .         .     SD  .         .         .         .
3201  AAGGTGGACGCCAATGCGCCGCTGACGCAACTGGTGTGCTGAATGATTATTGAACAGAAAAAAGGAGATGAATGATGAGCAACCCTTTCCCGGCGCATTT
 619   K  V  D  A  N  A  P  L  T  Q  L  V  C  *  631                             M  S  N  P  F  P  A  H  F

               .         .         .         .         .         .         .         .         .         .
3301  TTTATGGGGCGGCGCGATTGCCGCCAATCAGGTGGAAGGCGCCTATCTGACCGATGGTAAAGGACTTTCCACGTCGGATTTACAGCCTCAGGGCATCTTC
  10    L  W  G  G  A  I  A  A  N  Q  V  E  G  A  Y  L  T  D  G  K  G  L  S  T  S  D  L  Q  P  Q  G  I  F
```

FIG. 3—*Continued.*

presence of charged residues at the C terminus. Below is given an analysis of the ArbF amino acid sequence with these criteria. (i) The two potential phosphorylation sites were identified at residues 309 and 553 (Fig. 4); both contained histidyl residues presumably involved in phosphorylation. The phosphorylation site closer to the C terminus (amino acid 553) was found to be highly similar to that in BglF, including the presence of a positively charged amino

```
                 .         .         .         .         .         .         .         .         .         .
3401  GGCGAAATTGTGACGCGCCAGCCGGGCGACAGCGGCATCAAAGATGTGGCGATTGATTTTTATCACCGTTACCCGCAGGACATCGCGCTGTTTGCGGAAA
  43  G  E  I  V  T  R  Q  P  G  D  S  G  I  K  D  V  A  I  D  F  Y  H  R  Y  P  Q  D  I  A  L  F  A  E  M

                 .         .  ClaI1  .         .         .         .         .         .         .         .
3501  TGGGGTTTACCTGCCTGCGCATATCGATAGCCTGGACGCGCATTTTCCCGCAGGGGGACGAGGCAGAACCGAACGAAGCGGGGCTGGCGTTTTACGATCG
  77    G  F  T  C  L  R  I  S  I  A  W  T  R  I  F  P  Q  G  D  E  A  E  P  N  E  A  G  L  A  F  Y  D  R

                 .         .         .         .         .         .         .         .         .         .
3601  GCTGTTTGATGAGCTGGCAAAGTACGGTATTCAGCCATTAGTGACGTTGTCACACTATGAAATGCCGTATGGACTGGTGGAAAAACACGGCGGCTGGGGC
 110   L  F  D  E  L  A  K  Y  G  I  Q  P  L  V  T  L  S  H  Y  E  M  P  Y  G  L  V  E  K  H  G  G  W  G

                 .         .         .         .         .         .         .         .         .         .
3701  AACCGCTTGACTATTGATTGTTTTGAGCGCTACGCCCGCACGGTGTTTGCGCGCTATCGCCACAAGGTTAAGCGTTGGCTGACGTTCAACGAGATCAATA
 143  N  R  L  T  I  D  C  F  E  R  Y  A  R  T  V  F  A  R  Y  R  H  K  V  K  R  W  L  T  F  N  E  I  N  M

                 .         .         .         .         .         .         .         .         .         .
3801  TGTCGCTGCATGCGCCCTTTACCGGCGTCGGGCTGCCGCCGGACAGCGATAAGGCGGCGATTTATCAGGCGATCCATCATCAACTGGTCGCCAGCGCCAG
 177    S  L  H  A  P  F  T  G  V  G  L  P  P  D  S  D  L  A  A  I  K  Q  A  I  H  H  Q  L  V  A  S  A  R

                 .         .         .         .         .         .         .         .         .         .
3901  AGCGGTAAAAGCCTGTCACGACATGATTCCGGACGCGCAAATCGGCAACATGCTGTTGGGGGCGATGCTCTACCCGCTGACCAGCAAGCCGGAAGATGTA
 210   A  V  K  A  C  H  D  M  I  P  D  A  Q  I  G  N  M  L  L  G  A  M  L  Y  P  L  T  S  K  P  E  D  V

                 .         .         .         .         .         .         .         .         .         .
4001  ATGGAAAGCCTGCATCAGAACCGGGAATGGCTGTTCTTCGGCGATGTGCAGGTGCGCGGCGCGTATCCGGGGTATATGCACCGGTATTTCCGGGAGCAGG
 243  M  E  S  L  H  Q  N  R  E  W  L  F  F  G  D  V  Q  V  R  G  A  Y  P  G  Y  M  H  R  Y  F  R  E  Q  G

                 .         .         .         .         .         .         .         .         .         .
4101  GTATTACGCTGAATATTACGGCGCAGGATAAGCAGGACCTGAAAGCCACCGTTGATTTTATTTCCTTCAGCTATTACATGACCGGTTGCGTCACGACCGA
 277    I  T  L  N  I  T  A  Q  D  K  Q  D  L  K  A  T  V  D  F  I  S  F  S  Y  Y  M  T  G  C  V  T  T  D

                 .         .         .         .         .         .         .         .         .         .
4201  TGAAGCGCAACTCGAAAAAACGCGCGGCAATATTTTGAATATGGTGCCGAACCCGTATCTGGAAAGCTCTGAATGGGGATGGCAGATTGATCCGCTGGGG
 310   E  A  Q  L  E  K  T  R  G  N  I  L  N  M  V  P  N  P  Y  L  E  S  S  E  W  G  W  Q  I  D  P  L  G

                 .         .         .         .         .         .         .         .         .         .
4301  CTGCGTTATTTGCTGAATTTCCTGTACGACCGTTACCAAAAGCCGTTGTTTATTGTCGAGAACGGTCTGGGTGCGAAAGATAAGATTGAAGAGAATGGCG
 343  L  R  Y  L  L  N  F  L  Y  D  R  Y  Q  K  P  L  F  I  V  E  N  G  L  G  A  K  D  K  I  E  E  N  G  D

                 .         .         .         .         .         .         .         .         .         .
4401  ACATTTATGACGATTATCGTATCCGCTACCTGAATGATCATCTGGTGCAGGTCGGCGAAGCTATCGACGACGGCGTTGAGGTGCTGGGATATACCTGCTG
 377    I  Y  D  D  Y  R  I  R  Y  L  N  D  H  L  V  Q  V  G  E  A  I  D  D  G  V  E  V  L  G  Y  T  C  W

                 .         .         .         .         .         .         .         .         .         .
4501  GGGACCGATTGACTTGGTCAGCGCCTCGAAGGCGGAAATGTCCAAACGTTATGGCTTTATTTATGTCGATCGTGATGATGCCGGTCATGGCTCGCTGGAG
 410   G  P  I  D  L  V  S  A  S  K  A  E  M  S  K  R  Y  G  F  I  Y  V  D  R  D  D  A  G  H  G  S  L  E

                 .         .   KpnI1         .         .         .         .         .         .         .
4601  CGGAGACGTAAAAAGAGTTTTTACTGGTACCAATCCGTTATTGCCAGTCACGGAAAAACGCTGACGCGATAATAAATATATCTGAAATCAACCTGCGGAT
 443  R  R  R  K  K  S  F  Y  W  Y  Q  S  V  I  A  S  H  G  K  T  L  T  R  *

                 .         .         .         .         .         .         .         .         .         .
4701  AGCGATATCCGCAGGCTGTTGACGCTCGCCGTTAACCGGCCGGCGTGAATTGTTGTTTTTCTTTATCCGCTTATTTATCCGATTATCCGACGGGGAACGT

                 .         .         .         .         .         .         .         .         .         .
4801  GAAGATGAAAATAAAGAATAGCTATCTGGTGATAGCCAGCCTGCTTTATCCCATATCATTTATTTCCACTGCGGCGTCATTAACCGTTGAACAACGGTTG
          M  K  I  K  N  S  Y  L  V  I  A  S  L  L  Y  P  I  S  F  I  S  T  A  A  S  L  T  V  E  Q  R  L

                 .         .         .         .         .         .         .         .         .         .
4901  GCGGCGTTGGAAAATGATTTACAGGAAACCAAACAGGAGTTGCAGCGTTATAAAGAACAGGAGAAGAAAAACAAAGCTATTACCCTTGTCAGGGTAAATT
      A  A  L  E  N  D  L  Q  E  T  K  Q  E  L  Q  R  Y  K  E  Q  E  K  K  N  K  A  I  T  L  V  R  V  N  S

                 .         .         .         .         .         .
5001  CCGCCGCCGACGCGGATAATAAAAGTAATGCCTTCAACGTCGCGAAAACCGCTACGCCGATCGCG
        A  A  D  A  D  N  K  S  N  A  F  N  V  A  K  T  A  T  P  I  A
```

FIG. 3—*Continued.*

acid 7 residues after the histidyl residue, a feature unique among the EII enzymes. The other phosphorylation site, at amino acid 309, was found to be less conserved and to have, in particular, a cysteinyl residue at position 303, a unique feature among the EII enzymes analyzed to date, and a phenylalanyl residue at position 312, as has also been found in the *E. coli* EII Mtl and *B. subtilis* SacP permeases (11, 13). (ii) Axial projection of a potential α-helical conformation of

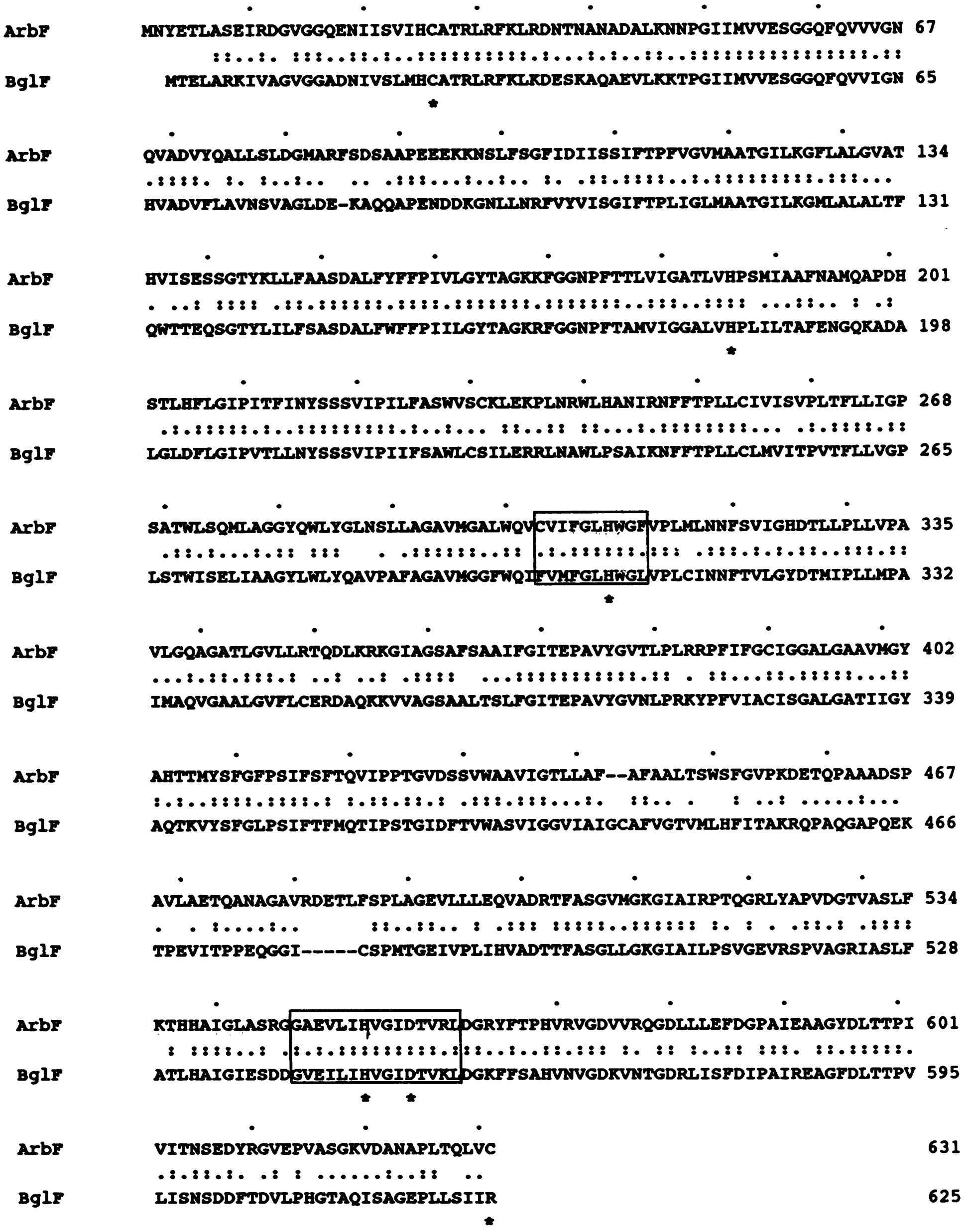
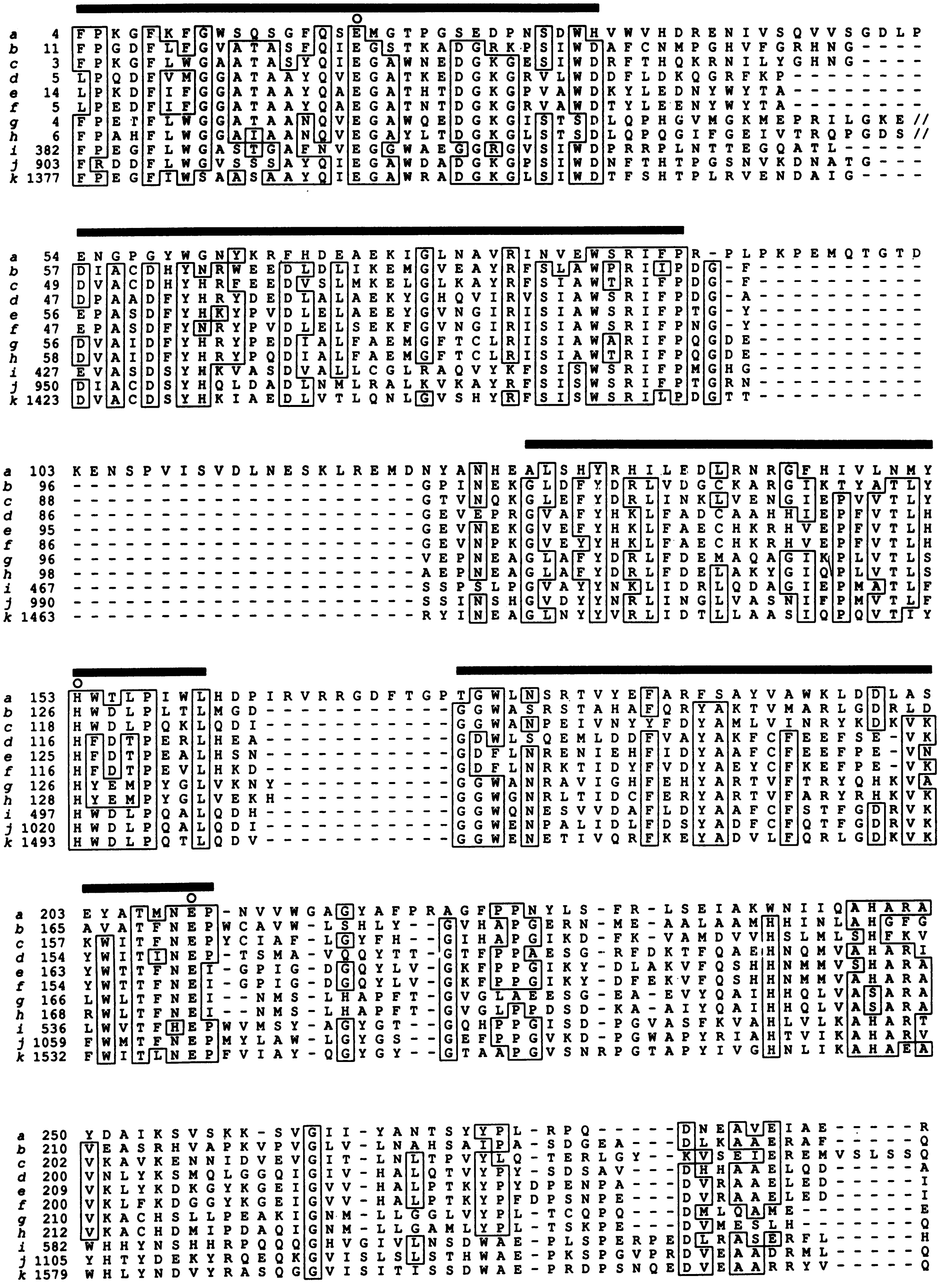


FIG. 4. Pairwise alignment of the deduced amino acid sequences of the *E. chrysanthemi* ArbF protein and the *E. coli* K-12 BglF β-glucoside permease. Colons indicate identical residues; dots indicate conservative changes. The two potential phosphorylation sites, as proposed by Saier et al. (39, 40), are boxed (see the text). Amino acids Cys-24, His-183, His-306, His-547, Asp-551, and Arg-625, shown to be important for BglF activity, are marked with asterisks below the sequences (47).

the first 14 residues revealed a picture closely related to that reported for various sucrose EII enzymes (39). In particular, one side of the helix contained only charged hydrophilic residues, while the other contained only hydrophobic ones, giving rise to a typical amphipathic α-helical segment of 14 residues terminated by a pair of glycyl residues (data not shown). It was proposed that such short N-terminal amphipathic sequences are embedded in the membrane and assist in the correct targeting of the permeases (39; see below). (iii) The C terminus of ArbF did not exhibit the signature proposed by Saier et al. (39, 40), i.e., one hydrophobic residue followed by a histidyl or an arginyl, since a cysteinyl residue occurred instead (see below).

A site-directed mutagenesis study of BglF underlined the importance of His-547, Cys-24, His-306, His-183, Asp-501, and Arg-625 residues (47). All were found to be conserved in ArbF, except for Arg-625, which was replaced by a Cys residue (Fig. 4). This result was somewhat unexpected since, in BglF, Arg-625 was proposed to function together with His-547 and Asp-551. No Arg, His, or Lys residues which could be candidates for playing the role of Arg-625 in BglF occurred at the C-terminal end of ArbF. Since there is a single base difference between the Arg-625 codon (CGC) in BglF and the Cys codon (CGA) in ArbF, the nucleotide sequence of this region was determined twice, starting from two different plasmids in each case, on both strands, but no



```
a    4  FPKGFKFGWSQSGFQSEMGTPGSEDPNSDWHVWVHDRENIVSQVVSGDLP
b   11  FPGDFLFGVATASFQIEGSTKADGRKPSIWDAFCNMPGHVFGRHNG----
c    3  FPKGFLWGAATASYQIEGAWNEDGKGESIWDRFTHQKRNILYGHNG----
d    5  LPQDFVMGGATAAYQVEGATKEDGKGRVLWDDFLDKQGRFKP--------
e   14  LPKDFIFGGATAAYQAEGATHTDGKGPVAWDKYLEDNYWYTA--------
f    5  LPEDFIFGGATAAYQAEGATNTDGKGRVAWDTYLEENYWYTA--------
g    4  FPETFLWGGATAANQVEGAWQEDGKGISTSDLQPHGVMGKMEPRILGKE//
h    6  FPAHFLWGGAIAANQVEGAYLTDGKGLSTSDLQPQGIFGEIVTRQPGDS//
i  382  FPEGFLWGASTGAFNVEGGWAEGGRGVSIWDPRRPLNTTEGQATL-----
j  903  FRDDFLWGVSSSAYQIEGAWDADGKGPSIWDNFTHTPGSNVKDNATG---
k 1377  FPEGFIWSAASAAYQIEGAWRADGKGLSIWDTFSHTPLRVENDAIG----

a   54  ENGPGYWGNYKRFHDEAEKIGLNAVRINVEWSRIFPR-PLPKPEMQTGTD
b   57  DIACDHYNRWEEDLDLIKEMGVEAYRFSLAWPRIIPDG-F----------
c   49  DVACDHYHRFEEDVSLMKELGLKAYRFSIAWTRIFPDG-F----------
d   47  DPAADFYHRYDEDLALAEKYGHQVIRVSIAWSRIFPDG-A----------
e   56  EPASDFYHKYPVDLELAEEYGVNGIRISIAWSRIFPTG-Y----------
f   47  EPASDFYNRYPVDLELSEKFGVNGIRISIAWSRIFPNG-Y----------
g   56  DVAIDFYHRYPEDIALFAEMGFTCLRISIAWARIFPQGDE----------
h   58  DVAIDFYHRYPQDIALFAEMGFTCLRISIAWTRIFPQGDE----------
i  427  EVASDSYHKVASDVALLCGLRAQVYKFSISWSRIFPMGHG----------
j  950  DIACDSYHQLDADLNMLRALKVKAYRFSISWSRIFPTGRN----------
k 1423  DVACDSYHKIAEDLVTLQNLGVSHYRFSISWSRILPDGTT----------

a  103  KENSPVISVDLNESKLREMDNYANHEALSHYRHILEDLRNRGFHIVLNMY
b   96  --------------------GPINEKGLDFYDRLVDGCKARGIKTYATLY
c   88  --------------------GTVNQKGLEFYDRLINKLVENGIEPVVTLY
d   86  --------------------GEVEPRGVAFYHKLFADCAAHHIEPFVTLH
e   95  --------------------GEVNEKGVEFYHKLFAECHKRHVEPFVTLH
f   86  --------------------GEVNPKGVEYYHKLFAECHKRHVEPFVTLH
g   96  --------------------VEPNEAGLAFYDRLFDEMAQAGIKPLVTLS
h   98  --------------------AEPNEAGLAFYDRLFDELAKYGIQPLVTLS
i  467  --------------------SSPSLPGVAYYNKLIDRLQDAGIEPMATLF
j  990  --------------------SSINSHGVDYYNRLINGLVASNIFPMVTLF
k 1463  --------------------RYINEAGLNYYVRLIDTLLAASIQPQVTIY

a  153  HWTLPIWLHDPIRVRRGDFTGPTGWLNSRTVYEFARFSAYVAWKLDDLAS
b  126  HWDLPLTLMGD-----------GGWASRSTAHAFQRYAKTVMARLGDRLD
c  118  HWDLPQKLQDI-----------GGWANPEIVNYYFDYAMLVINRYKDKVK
d  116  HFDTPERLHEA-----------GDWLSQEMLDDFVAYAKFCFEEFSE-VK
e  125  HFDTPEALHSN-----------GDFLNRENIEHFIDYAAFCFEEFPE-VN
f  116  HFDTPEVLHKD-----------GDFLNRKTIDYFVDYAEYCFKEFPE-VK
g  126  HYEMPYGLVKNY----------GGWANRAVIGHFEHYARTVFTRYQHKVA
h  128  HYEMPYGLVEKH----------GGWGNRLTIDCFERYARTVFARYRHKVK
i  497  HWDLPQALQDH-----------GGWQNESVVDAFLDYAAFCFSTFGDRVK
j 1020  HWDLPQALQDI-----------GGWENPALIDLFDSYADFCFQTFGDRVK
k 1493  HWDLPQTLQDV-----------GGWENETIVQRFKEYADVLFQRLGDKVK

a  203  EYATMNEP-NVVWGAGYAFPRAGFPPNYLS-FR-LSEIAKWNIIQAHARA
b  165  AVATFNEPWCAVW-LSHLY--GVHAPGERN-ME-AALAAMHHINLAHGFG
c  157  KWITFNEPYCIAF-LGYFH--GIHAPGIKD-FK-VAMDVVHSLMLSHFKV
d  154  YWITINEP-TSMA-VQQYTT-GTFPPAESG-RFDKTFQAEHNQMVAHARI
e  163  YWTTFNEI-GPIG-DGQYLV-GKFPPGIKY-DLAKVFQSHHNMMVSHARA
f  154  YWTTFNEI-GPIG-DGQYLV-GKFPPGIKY-DFEKVFQSHHNMMVAHARA
g  166  LWLTFNEI--NMS-LHAPFT-GVGLAEESG-EA-EVYQAIHHQLVASARA
h  168  RWLTFNEI--NMS-LHAPFT-GVGLPPDSD-KA-AIYQAIHHQLVASARA
i  536  LWVTFHEPWVMSY-AGYGT--GQHPPGISD-PGVASFKVAHLVLKAHART
j 1059  FWMTFNEPMYLAW-LGYGS--GEFPPGVKD-PGWAPYRIAHTVIKAHARV
k 1532  FWITLNEPFVIAY-QGYGY--GTAAPGVSNRPGTAPYIVGHNLIKAHAEA

a  250  YDAIKSVSKK-SVGII-YANTSYYPL-RPQ-----DNEAVEIAE-----R
b  210  VEASRHVAPKVPVGLV-LNAHSAIPA-SDGEA---DLKAAERAF-----Q
c  202  VKAVKENNIDVEVGIT-LNLTPVYLQ-TERLGY--KVSEIEREMVSLSSQ
d  200  VNLYKSMQLGGQIGIV-HALQTVYPY-SDSAV---DHHAAELQD-----A
e  209  VKLYKDKGYKGEIGVV-HALPTKYPYDPENPA---DVRAAELED-----I
f  200  VKLFKDGGYKGEIGVV-HALPTKYPFDPSNPE---DVRAAELED-----I
g  210  VKACHSLLPEAKIGNM-LLGGLVYPL-TCQPQ---DMLQAME-------E
h  212  VKACHDMIPDAQIGNM-LLGAMLYPL-TSKPE---DVMESLH-------Q
i  582  WHHYNSHHRPQQQGHVGIVLNSDWAE-PLSPERPEDLRASERFL-----H
j 1105  YHTYDEKYRQEQKGVISLSLSTHWAE-PKSPGVPRDVEAADRML-----Q
k 1579  WHLYNDVYRASQGGVISITISSDWAE-PRDPSNQEDVEAARRYV-----Q
```

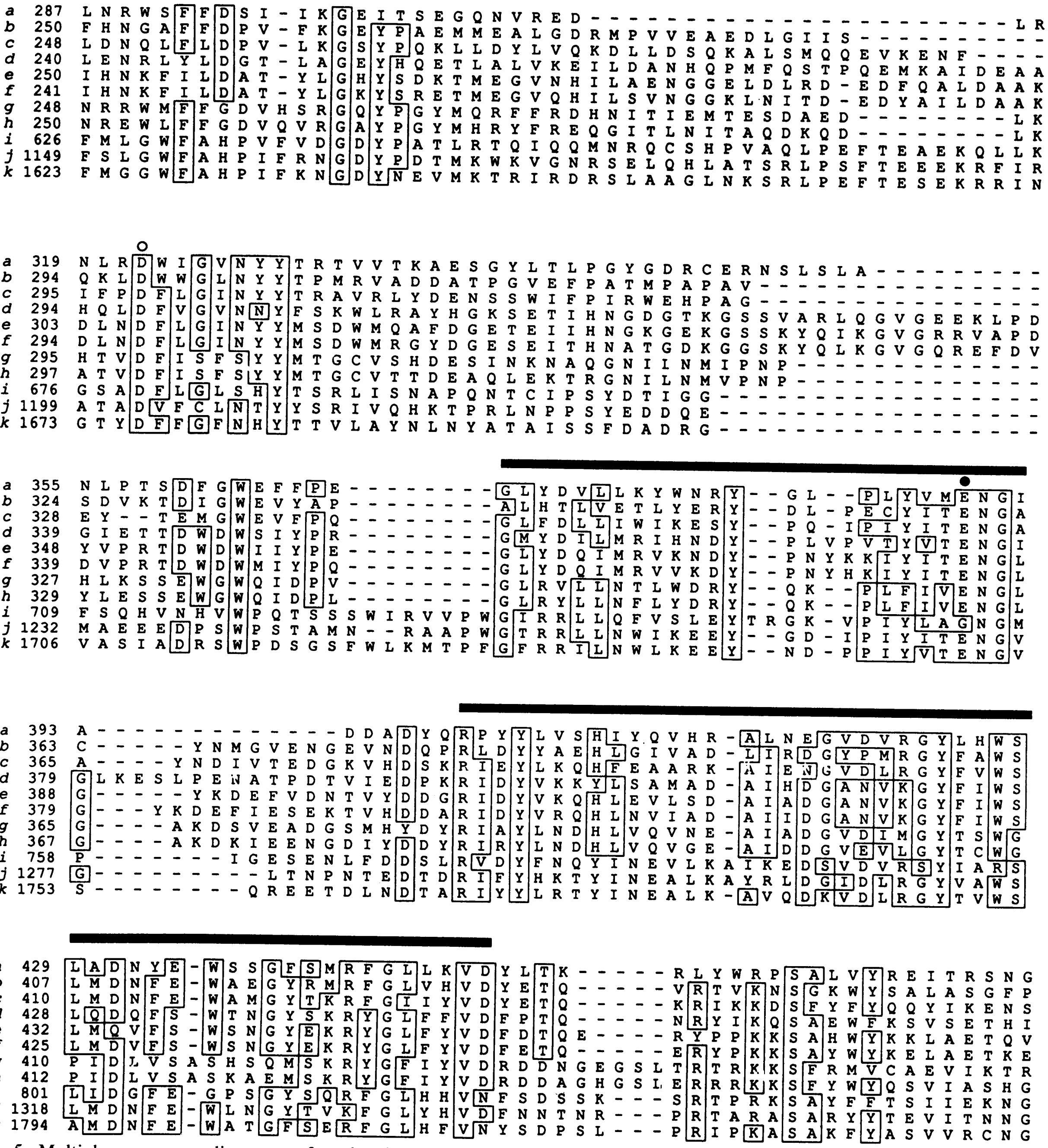

a 287 LNRWSFFDSI-IKGEITSEGQNVRED----------------------LR
b 250 FHNGAFFDPV-FKGEYPAEMMEALGDRMPVVEAEDLGIIS----------
c 248 LDNQLFLDPV-LKGSYPQKLLDYLVQKDLLDSQKALSMQQEVKENF----
d 240 LENRLYLDGT-LAGEYHQETLALVKEILDANHQPMFQSTPQEMKAIDEAA
e 250 IHNKFILDAT-YLGHYSDKTMEGVNHILAENGGELDLRD-EDFQALDAAK
f 241 IHNKFILDAT-YLGKYSRETMEGVQHILSVNGGKLNITD-EDYAILDAAK
g 248 NRRWMFFGDVHSRGQYPGYMQRFFRDHNITIEMTESDAED--------LK
h 250 NREWLFFGDVQVRGAYPGYMHRYFREQGITLNITAQDKQD--------LK
i 626 FMLGWFAHPVFVDGDYPATLRTQIQQMNRQCSHPVAQLPEFTEAEKQLLK
j 1149 FSLGWFAHPIFRNGDYPDTMKWKVGNRSELQHLATSRLPSFTEEEKRFIR
k 1623 FMGGWFAHPIFKNGDYNEVMKTRIRDRSLAAGLNKSRLPEFTESEKRRIN
○
a 319 NLRDWIGVNYYTRTVVTKAESGYLTLPGYGDRCERNSLSLA---------
b 294 QKLDWWGLNYYTPMRVADDATPGVEFPATMPAPAV---------------
c 295 IFPDFLGINYYTRAVRLYDENSSWIFPIRWEHPAG---------------
d 294 HQLDFVGVNNYFSKWLRAYHGKSETIHNGDGTKGSSVARLQGVGEEKLPD
e 303 DLNDFLGINYYMSDWMQAFDGETEIIHNGKGEKGSSKYQIKGVGRRVAPD
f 294 DLNDFLGINYYMSDWMRGYDGESEITHNATGDKGGSKYQLKGVGQREFDV
g 295 HTVDFISFSYYMTGCVSHDESINKNAQGNIINMIPNP-------------
h 297 ATVDFISFSYYMTGCVTTDEAQLEKTRGNILNMVPNP-------------
i 676 GSADFLGLSHYTSRLISNAPQNTCIPSYDTIGG-----------------
j 1199 ATADVFCLNTYYSRIVQHKTPRLNPPSYEDDQE-----------------
k 1673 GTYDFFGFNHYTTVLAYNLNYATAISSFDADRG-----------------
●
a 355 NLPTSDFGWEFFPE--------GLYDVLLKYWNRY--GL--PLYVMENGI
b 324 SDVKTDIGWEVYAP--------ALHTLVETLYERY--DL-PECYITENGA
c 328 EY--TEMGWEVFPQ--------GLFDLLIWIKESY--PQ-IPIYITENGA
d 339 GIETTDWDWSIYPR--------GMYDILMRIHNDY--PLVPVTYVTENGI
e 348 YVPRTDWDWIIYPE--------GLYDQIMRVKNDY--PNYKKIYITENGL
f 339 DVPRTDWDWMIYPQ--------GLYDQIMRVVKDY--PNYHKIYITENGL
g 327 HLKSSEWGWQIDPV--------GLRVLLNTLWDRY--QK--PLFIVENGL
h 329 YLESSEWGWQIDPL--------GLRYLLNFLYDRY--QK--PLFIVENGL
i 709 FSQHVNHVWPQTSSSWIRVVPWGIRRLLQFVSLEYTRGK-VPIYLAGNGM
j 1232 MAEEEDPSWPSTAMN--RAAPWGTRRLLNWIKEEY--GD-IPIYITENGV
k 1706 VASIADRSWPDSGSFWLKMTPFGFRRILNWLKEEY--ND-PPIYVTENGV
a 393 A-------------DDADYQRPYYLVSHIYQVHR-ALNEGVDVRGYLHWS
b 363 C-----YNMGVENGEVNDQPRLDYYAEHLGIVAD-LIRDGYPMRGYFAWS
c 365 A----YNDIVTEDGKVHDSKRIEYLKQHFEAARK-AIENGVDLRGYFVWS
d 379 GLKESLPENATPDTVIEDPKRIDYVKKYLSAMAD-AIHDGANVKGYFIWS
e 388 G-----YKDEFVDNTVYDDGRIDYVKQHLEVLSD-AIADGANVKGYFIWS
f 379 G---YKDEFIESEKTVHDDARIDYVRQHLNVIAD-AIIDGANVKGYFIWS
g 365 G----AKDSVEADGSMHYDYRIAYLNDHLVQVNE-AIADGVDIMGYTSWG
h 367 G----AKDKIEENGDIYDDYRIRYLNDHLVQVGE-AIDDGVEVLGYTCWG
i 758 P-------IGESENLFDDSLRVDYFNQYINEVLKAIKEDSVDVRSYIARS
j 1277 G---------LTNPNTEDTDRIFYHKTYINEALKAYRLDGIDLRGYVAWS
k 1753 S--------QREETDLNDTARIYYLRTYINEALK-AVQDKVDLRGYTVWS
a 429 LADNYE-WSSGFSMRFGLLKVDYLTK-----RLYWRPSALVYREITRSNG
b 407 LMDNFE-WAEGYRMRFGLVHVDYETQ-----VRTVKNSGKWYSALASGFP
c 410 LMDNFE-WAMGYTKRFGIIYVDYETQ-----KRIKKDSFYFYQQYIKENS
d 428 LQDQFS-WTNGYSKRYGLFFVDFPTQ-----NRYIKQSAEWFKSVSETHI
e 432 LMQVFS-WSNGYEKRYGLFYVDFDTQE----RYPPKKSAHWYKKLAETQV
f 425 LMDVFS-WSNGYEKRYGLFYVDFETQ-----ERYPKKSAYWYKELAETKE
g 410 PIDLVSASHSQMSKRYGFIYVDRDDNGEGSLTRTRKKSFRMVCAEVIKTR
h 412 PIDLVSASKAEMSKRYGFIYVDRDDAGHGSLERRRKKSFYWYQSVIASHG
i 801 LIDGFE-GPSGYSQRFGLHHVNFSDSSK---SRTPRKSAYFFTSIIEKNG
j 1318 LMDNFE-WLNGYTVKFGLYHVDFNNTNR---PRTARASARYYTEVITNNG
k 1794 AMDNFE-WATGFSERFGLHFVNYSDPSL---PRIPKASAKFYASVVRCNG

FIG. 5. Multiple sequence alignment of β-glycohydrolases: β-galactosidase from *S. solfataricus* (23) (a), β-glucosidases from an *Agrobacterium* sp. (51) (b) and *C. saccharolyticum* (25) (c), phospho-β-galactosidases from *Lactobacillus casei* (35) (d), *Streptococcus lactis* (5) (e), and *Staphylococcus aureus* (6) (f), phospho-β-glucosidases from *E. coli* (48) (g) and *E. chrysanthemi* (this work) (h), and human LPH (29) domains II (i), III (j), and IV (k). Human LPH domain I was omitted because its similarity to the other domains is only partial and does not encompass the active-site region (29). Similarly, the rat LPH and recently published *S. solfataricus* MT-4 β-galactosidase sequences were not included, since they exhibit a very high level of homology with those of human LPH and the other *S. solfataricus* enzyme, respectively (10, 29). The alignment was produced with the Clustal program (19) and verified by hydrophobic cluster analysis (21). Various amino acid matrix similarities were used (12, 38); all gave essentially the same type of alignment. Residues conserved in at least 6 of the 11 sequences are boxed. Symbols: ●, active-site Glu residue found to act as a nucleophile in the *Agrobacterium* enzyme (52); ○, invariant Asp, Glu, or His residues. The areas marked with a black bar are those that were used to construct the tree in Fig. 6.

ambiguity was ever found. Whether Cys can, in this context, substitute for Arg, i.e., function together with the His-547 residue in phosphate transfer, remains to be determined.

Comparison of the ArbB amino acid sequence with those of various β-glycohydrolases from eubacteria, archeabacteria, and mammals, including humans. The ArbB amino acid sequence was used to scan a data bank specific for β-glycohydrolases (18) by use of the Clustal program or hydrophobic cluster analysis (19, 21). As expected, we found a high level of identity between the *E. chrysanthemi* ArbB and *E.*

coli BglB proteins, amounting to approximately 70%. As already pointed out by others, both of these gram-negative bacterial phospho-β-glucosidases were found to share a high level of identity with phospho-β-galactosidases originating from gram-positive bacteria, such as *Streptococcus*, *Staphylococcus*, and *Lactobacillus* spp. (35). We identified six additional homologous sequences: β-glucosidases from a gram-negative *Agrobacterium* sp. and from the extreme thermophile *Caldocellum saccharolyticum*, two β-galactosidases from the archeabacterium *Sulfolobus solfataricus*, and lactate-phlorizin hydrolases (LPH) from rats and humans (Fig. 5). Hence, this family includes four types of enzymes, as defined by the International Union of Biochemistry classification: phospho-β-glucosidase (EC 3.2.1.86), phospho-β-galactosidase (EC 3.2.1.85), β-glucosidase (EC 3.2.1.21), and β-galactosidase (EC 3.2.1.23). The alignment of the sequences revealed that (i) the N- and C-terminal regions are better conserved than the middle regions, (ii) optimal alignment of the archeabacterium enzyme requires the introduction of long gaps into the other sequences, and (iii) 27 amino acids are found to be invariant (Fig. 5). On the basis of a pairwise comparison of the largest regions common to all sequences, we suggest an unrooted evolution tree (Fig. 6). The most related sequences were those of proteins occurring in taxonomically related bacteria. However, the enzymes could also be separated into phospho-β-glycosidases and β-glycosidases, i.e., depending on the presence or absence of a phosphoryl group at the C-6 position of the substrate. This result suggested that this position might have acted as a functional constraint in the evolution of the enzymes analyzed. The human LPH molecule might provide us with an example of evolution from one type of β-glycosidase to another. LPH bears two catalytic activities: β-galactosidase (domain III) and β-glucosidase (domain IV) (29) (Fig. 5). If one assumes that both domains arose by genetic duplication, structural differences between them should reflect, above all, the constraints imposed by the substrates.

A most interesting outcome from building protein families is the ability to predict functionally important residues. β-Glycosidases act by a general acid catalysis mechanism in which two acidic residues participate in a single or double displacement reaction, resulting in the inversion or retention of the configuration, respectively, at the anomeric carbon (50). It was recently shown that, in the *Agrobacterium* enzyme, the Glu-358 residue is directly involved in the glycosidic bond cleavage by acting as a nucleophile (52). This residue is invariant in 10 of the 11 members of the family. LPH domain II is the exception (Fig. 5). This result is of special interest since (i) LPH, like the *Agrobacterium* enzyme, catalyzes the glycosidic bond cleavage with the retention of the configuration at C-1 of the product (49), and (ii) LPH domain II is removed from pro-LPH by proteolytic processing in vivo (29). Therefore, we predict that domain III Glu-1273 and Glu-1749 of human LPH are directly involved in catalysis and that LPH domain II is a noncatalytic polypeptide. Because general acid catalysis also requires the participation of a proton donor residue, other appropriate invariant residues (Glu, Asp, and His) were sought and are proposed as potential candidates for site-directed mutagenesis studies (Fig. 5). While this work was in progress, the sequences of two β-glucosidases from *B. polymyxa* were reported by Gonzàles-Candelas et al. (17) along with a similar sequence analysis. These authors concluded that there were two types of β-glucosidases, the *B. polymyxa* enzymes being part of the family also containing β-galactosidases, phospho-β-galactosidases, and phospho-β-glucosidases, as we have concluded. However, their analysis did not include the archeabacterial and human enzymes in the proposed family and did not take advantage of the knowledge about the *Agrobacterium* enzyme to predict a catalytic glutamate in all members of the family.

Comparison of the ArbG amino acid sequence with those of *E. coli* and *B. subtilis* transcriptional antiterminators. The *E. coli* BglG protein functions as a transcriptional antiterminator and exhibits RNA binding capacity (20). The *E. chrysanthemi* ArbG protein exhibits 61% identity with BglG (Fig. 7). The *B. subtilis* SacT and SacY proteins presumably function as antiterminators as well (9, 13). Alignment of the ArbG

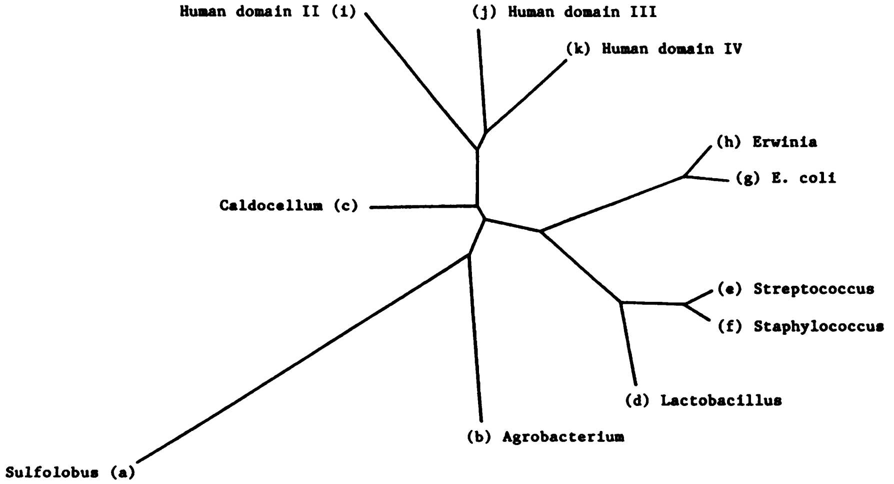


FIG. 6. Distance tree calculated from a pairwise comparison of β-glycohydrolases from various origins. The tree was constructed from pairwise comparison scores of the largest conserved domains shown in Fig. 5. Other trees were constructed with the entire sequences and were found to be similar in topology as well as relative branch length (data not shown). The tree was constructed by submitting the edited sequences to the multiple sequence alignment program of Lipman et al. (22), with the matrix of Dayhoff (12); the use of the matrix of Rissler et al. (38) did not produce a very different alignment. Thus, for each pair of sequences, the program computed the sum of the alignment costs, and these values were used to build a tree with the neighbor-joining method of Saitou and Nei (41).

```
ArbG     MK--IAKILNNNVVTVMDEQNNEQVVMGRGLGFKKRPGDTVNAALIEKIFSLRSSELTAR
BglG     MNMQITKILNNNVVVVIDDQQREKVVMGRGIGFQKRAGERINSSGIEKEYALSSHELNGR
SacT     MK--IYKVLNNNAA-LIKEDDQEKIVMGPGIAFQKKKNDLIPMNKVEKIFVVRDENE--K
SacY     MK--IKRILNHNAI-VVKDQNEEKILLGAGIAFNKKKNDIVDPSKIEKTFIRKDTPDYKQ
         *.  * ..**.*.  ...... *....* *..*.*. .. .  . .** .  ..     .

ArbG     LSDVLERIPLEVVTTADRIIALAKEKLGGNLQNSLYISLTDHCHFAIERHRQGVDIRNGL
BglG     LSELLSHIPLEVMATCDRIISLAQERLG-KLQDSIYISLTDHCQFAIKRFQQNVLLPNPL
SacT     FKQILQTLPEEHIEIAEDIISYAEGELAAPLSDHIHIALSDHLSFAIERIQNGLLVQNKL
SacY     FEEILETLPEDHIQISEQIISHAEKELNIKINERIHVAFSDHLSFAIERLSNGNVIKNPL
         ....*  .* . . . . **. *.  *.  . . ......**  ***.*  ..  . * *

ArbG     QWE-KRLYQKEFAIGLDALDIIHRRLGVRLPEDEAGFIALHLVNAQLD-SHMPEVMRITR
BglG     LWDIQRLYPKEFQLGEEALTIIDKRLGVQLPKDEVGFIAMHLVSAQMS-GNMEDVAGVTQ
SacT     LHEIKALYKKEYEIGLWAIGHVKETLGVSLPEDEAGYIALHIHTAKMDAESMYSALKHTT
SacY     LNEIKVLYPKEFQIGLWARALIKDKLGIHIPDDEIGNIAMHIHTARNNAGDMTQTLDITT
           . . ** **...*  * . ..  **. .*.** * **.*. .*. . . *  .   *

ArbG     VMQEILNIVKYQLNLDYNEQAFSYHRFVTHLKFFAQRLLGRTPVFSEDESLHDVVKEKYT
BglG     LMREMLQLIKFQFSLNYQEESLSYQRLVTHLKFLSWRILEHASINDSDESLQQAVKQNYP
SacT     MIKEMIEKIKQYFNRKVDENSISYQRLVTHLRYAVSRLESNEALHRMDEEMLYFIQKKYS
SacY     MIRDIIEIIEIQLSINIVEDTISYERLVTHLRFAIQHIKAGESIYELDAEMIDIIKEKFK
         ....... ..  .. .  *...**.*.****..   .. . ...   *...   .....

ArbG     LAYHCAEKIQDHIMLHYDYTLTKEELMFLAIHIERVRSELQEQTAE
BglG     QAWQCAERIAIFIGLQYQRKISPAEIMFLAINIERVRKE------H
SacT     FAYQCALELAEFLKNEYQLHLPESEAGYITLHVQRL-----QDLSE
SacY     DAFLCALSIGTFVKKEYGFEFPEKELCYIAMHIQRF---YQRSVAR
          *. **  .   .  .*   ..  *  .......*
```

FIG. 7. Comparison of the amino acid sequences of the *E. chrysanthemi* ArbG protein with those of the *E. coli* K-12 BglG and *B. subtilis* SacT and SacY antiterminators. An asterisk below the aligned sequences indicates a residue identical in four sequences, while a dot indicates a conservative change. Aspartate and histidine residues shown to be important for BglG and SacT functions are boxed. The Clustal program (19) was used to produce the alignment shown.

amino acid sequence with those of the *E. coli* and *B. subtilis* proteins (Fig. 7) revealed the presence of 58 identical residues among the four proteins. Among them were the Asp-His residues located at position 100 and shown to be important for the activity of BglG and SacT (9, 13) (Fig. 7). This result strongly supports the idea that the ArbG protein is an antiterminator, while it renders the observation that ArbG could not substitute for BglG more puzzling (16).

Location of *lacZ* fusion insertion points. The determination of the *arbG* nucleotide sequence challenged some conclusions derived from our previous analysis. In particular, it had been concluded that the upstream limit of *arbG* was near the *Pvu*II restriction site, since Mu insertions upstream of this site did not interfere with the ability to use arbutin, while downstream insertions led to a negative phenotype (16). It was therefore disturbing to find that the *Pvu*II restriction site lies within the central part of the *arbG* nucleotide sequence (Fig. 3). This result prompted a reinvestigation of the location of some *lacZ* fusions studied in our previous work (16). Sequencing of the insertion points of two *arb-lacZ* fusions, 324 and 363, revealed that the Mu element was inserted within the N-proximal region of *arbF* after the 9th and 21st codons, respectively (Fig. 3). Hence, none of the fusions previously used mapped within the *arbG* gene, and the conclusion that *arbG* expression was not substrate inducible must be considered erroneous. Such a reassessment necessitates additional comment, since the expression of fusion 344 located in *arbF* was induced threefold by the substrates, while the expression of fusion 363, now shown to be in *arbF* as well, was found to be noninducible (16). The basis for this result might reside in the fact that fusion 363 lies in the region of the permease that is integrated in the membrane, while inducible fusion 344 is within the cytoplasmic tail (see above).

Search for transcriptional signals. We sought to identify sequences related to the promoters recognized by the σ^{70}-containing RNA polymerase. Three putative promoters were found and are referred to as P0, lying upstream of the *arbG* gene at bp 30, and P1 and P2, located within the *arbG-arbF* intercistronic region at bp 1119 and 1263, respectively (Fig.

```
                             "Box A"                                "Box B"
                         ________________                      _____________________
STB1-arb        :        GGGTTGCTACTGCCAT        TG            GCAGGCAAAACCAGATGTTC
T1-bgl          :        GGATTGTTACTGCATT        C             GCAGGCAAAACCTGACATAA
BSbgl-IR        :        GGATTGTTACTGATAA        A             GCAGGCAAAACCTAAATTGC
BSsacR/sacB     :        GGTTTGTTACTGATAA        A             GCAGGCAAGACCTAAAATGT
BSsacR/sacP     :        GGATTGTGACTGGTAA        A             GCAGGCAAGACCTAAAATTT
T2-bgl          :        GGATTGTTACCGCACT        AA            GCGGGCAAAACCTGAAAAAA
STB2-arb        :   1)   GGATTGCGACTGTATA     TCCCTCA          GCGGGAAATACAGGCAAAAC
                    2)   GGATTGCGACTGTATA     TCCCTCAGCGGGAAAT ACAGGCAAAACCTGAACCGT
                         ________________                      _____________________
```

FIG. 8. Alignment of BoxA and BoxB terminator motifs found in the *E. chrysanthemi arb*, *E. coli bgl*, and *B. subtilis bgl*, *sacB*, and *sacP* genes. Part of BoxB is thought to be engaged in forming the RNA hairpin structure terminating transcription (see the text). Two possibilities are given for STB2 BoxB to reach the best alignment with other BoxB's. Identical nucleotides in all seven motifs are shown in boldface type. The STB1-*arb* and STB2-*arb* motifs lie upstream and downstream of the *E. chrysanthemi arbG* gene, respectively; the T1-*bgl* and the T2-*bgl* motifs lie upstream and downstream of the *E. coli bglG* gene, respectively (48); the BS*bgl*-IR, BS*sacR*/*sacB*, and BS*sacR*/*sacP* motifs lie upstream of the *B. subtilis bgl* endoglucanase gene (32), the *sacB* levansucrase gene (9), and the *sacPA* sucrose metabolic operon (13), respectively.

3). None exhibited perfect matching to the −35 and −10 consensus sequences.

A search for potential stem-loop structures that could be involved in transcription termination was conducted. Two were located flanking the *arbG* gene; each was followed by a run of T residues, as expected for rho-independent terminators (Fig. 3). The free energy of formation, if the stem-loop structures were folded in RNA, would be −21.7 kcal/mol (ca. 90.8 kJ/mol) and −30 kcal/mol (ca. 125.5 kJ/mol) for the upstream (STB1) and downstream (STB2) structures, respectively. Furthermore, the STB1 and STB2 structures, along with each immediate upstream region, were found to share a high level of identity. In addition, both exhibited motifs, called BoxA and BoxB, which were earlier identified in the regions involved in the antitermination regulatory circuit of the *E. coli bgl* and *B. subtilis sacT* and *sacB* genes (9, 13, 20, 27, 46, 48). Alignment of the *E. chrysanthemi arb* STB1-containing region was straightforward: this region exhibited a sequence highly related to the BoxA and BoxB motifs, with the exception of the 3′ end of BoxB, which was demonstrated not to be essential for antiterminator recognition in *E. coli* (20). In contrast, good alignment of the *E. chrysanthemi arb* STB2-containing region seemed more difficult to achieve, since part of BoxB seemed to have been duplicated. Hence, the distance between BoxA and BoxB could be either 8 or 17 nucleotides (Fig. 8). It might be interesting to note that the 17-bp spacing region could be folded in a small hairpin, bringing together BoxA and BoxB such that they would now be at a distance similar to that described in the *E. coli* and *B. subtilis* systems (data not shown). In addition, a third potential rho-independent terminator structure was found downstream of the *arbB* gene (Fig. 3).

Conclusion. The determination of the *E. chrysanthemi arb* gene nucleotide sequence allowed the identification of three genes, *arbG*, *arbF*, and *arbB*. Phenotypic analysis of deletion derivatives and sequence comparison studies showed that the ArbF and ArbB proteins constitute a PTS-dependent EII permease and a phospho-β-glucosidase, respectively. The function of ArbG was solely inferred from a sequence comparison study, and ArbG was proposed to be an antiterminator. Both the genetic order and the sequences of the encoded products demonstrated that the expressed *E. chrysanthemi* Arb system is structurally equivalent to the *E. coli* cryptic *bgl* operon.

Two differences between the Arb and Bgl systems were noted; both involved the nontranslated regions. First, with the exception of the terminator-containing regions, no similarity was found between the regions lying upstream of the *arbG* gene and the *bgl* promoter. This finding precluded any attempt to identify discrete sites or determinants which could be responsible for the cryptic nature of the *bgl* promoter. This finding was surprising, since work by Diaz-Torres and Wright showed that a sequence highly similar to the *bgl* promoter occurs in the chromosomes of other members of the family *Enterobacteriaceae* (14). The second difference concerned the intercistronic region lying between the antiterminator and the permease genes which, in the Arb system, is unusually long, i.e., 300 bp versus 130 bp in *bgl* (48). The existence of this 300-bp intercistronic region points to a question yet to be solved in the *arb* system: are all three genes organized as an operon? Actually, preliminary evidence suggests that both the *arbF* and the *arbB* genes can be expressed independently from *arbG*. Furthermore, in our previous work, we showed that the expression of *arbF* was 3-fold inducible while the expression of *arbB* was increased 15-fold in the presence of β-glucosides (16).

The comparison of the *arb* and *bgl* gene nucleotide sequences failed to provide a clear-cut explanation of why the *E. coli* system is cryptic and the *E. chrysanthemi* is expressed. A possible answer is that the Arb system contains additional (or more efficient internal) promoters which are not controlled by the antitermination mechanism. Knowledge of the structural features of the *arb* genes has now provided the groundwork for investigating these questions.

ACKNOWLEDGMENTS

We are thankful to A. Wright, O. Amster-Choder, and M. Diaz-Torres for discussions and communications of unpublished data. Thanks are also due to J. Haiech for interest in part of this work. We are indebted to one of the reviewers for help in editing the manuscript.

This work was supported by grants from the European Communities, from the CNRS, and from La Fondation pour la Recherche Médicale.

REFERENCES

1. **Amster-Choder, O., F. Houman, and A. Wright.** 1989. Protein phosphorylation regulates transcription of the β-glucoside utilization operon in *E. coli*. Cell **58:**847–855.
2. **Amster-Choder, O., and A. Wright.** 1990. Regulation of activity of a transcriptional antiterminator in *E. coli* by phosphorylation *in vivo*. Science **249:**540–542.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
4. **Barras, F., M. El Hassouni, J. P. Chambost, and M. Chippaux.** 1989. The β-glucoside metabolism in *Erwinia chrysanthemi*: preliminary analysis and comparison to *Escherichia coli* systems. FEMS Microbiol. Rev. **63:**143–148.
5. **Boizet, B., D. Villeval, P. Slos, M. Novel, G. Novel, and A. Mercenier.** 1988. Isolation and structural analysis of the phospho-β-galactosidase gene from *Streptococcus lactis* Z268. Gene **62:**249–261.
6. **Breidt, F., and G. C. Stewart.** 1987. Nucleotide and deduced amino acid sequences of the *Staphylococcus aureus* phospho-β-galactosidase gene. Appl. Environ. Microbiol. **53:**969–973.
7. **Chatterjee, A. K., and A. K. Vidaver (ed.).** 1986. Advances in plant pathology, vol. 4, p. 1–218. Academic Press, Inc., New York.
8. **Chung, C. T., and R. H. Miller.** 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucleic Acids Res. **16:**3580.
9. **Crutz, A. M., M. Steinmetz, S. Aymeric, R. Richter, and D. Le Coq.** 1990. Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. J. Bacteriol. **172:**1043–1050.
10. **Cubellis, M. V., C. Rozzo, P. Montecucchi, and M. Rossi.** 1990. Isolation and sequencing of a new β-galactosidase-encoding archaebacterial gene. Gene **94:**89–94.
11. **Davis, T., M. Yamada, M. Elgort, and M. H. Saier.** 1988. Nucleotide sequence of the mannitol (*mtl*) operon in *Escherichia coli*. Mol. Microbiol. **2:**405–412.
12. **Dayhoff, M. O.** 1978. Survey of new data and computer methods of analysis, p. 1–8. *In* M. O. Dayhoff (ed.), Atlas of protein sequences and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
13. **Débarbouillé, M., M. Arnaud, A. Fouet, A. Klier, and G. Rapoport.** 1990. The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. J. Bacteriol. **172:**3966–3973.
14. **Diaz-Torres, M. R., and A. Wright (Tufts University, Boston, Mass.).** 1991. Personal communication.
15. **DiNardo, S., K. A. Voelkel, R. Sternglanz, A. E. Reynolds, and A. Wright.** 1982. *E. coli* DNA topoisomerase I mutants have

compensatory mutations in DNA gyrase genes. Cell **13**:43–51.

16. **El Hassouni, M., M. Chippaux, and F. Barras.** 1990. Analysis of the *Erwinia chrysanthemi arb* genes, which mediate metabolism of aromatic β-glucosides. J. Bacteriol. **172**:6261–6267.
17. **Gonzàlez-Candelas, L., D. Ramon, and J. Polaina.** 1990. Sequences and homology analyses of two genes encoding β-glucosidases from *Bacillus polymyxa*. Gene **95**:31–38.
18. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280**:309–316.
19. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.
20. **Houman, F., M. R. Diaz-Torres, and A. Wright.** 1990. Transcriptional antitermination in the *bgl* operon of *E. coli* is modulated by a specific RNA binding protein. Cell **62**:1153–1163.
21. **Lemesle-Varloot, L., B. Henrissat, C. Gaboriaud, V. Bissery, A. Morgat, and J. P. Mornon.** 1990. Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D representation of protein sequences. Biochimie **72**:555–574.
22. **Lipman, D. J., S. F. Altschul, and J. D. Kececioglu.** 1989. A tool for multiple sequence alignment. Proc. Natl. Acad. Sci. USA **86**:4412–4415.
23. **Little, S., P. Cartwright, C. Campbell, A. Prenneta, J. McChesney, A. Mountain, and M. Robinson.** 1989. Nucleotide sequence of a thermostable β-galactosidase from *Sulfolobus solfataricus*. Nucleic Acids Res. **17**:7980.
24. **Lopilato, J., and A. Wright.** 1990. Mechanisms of activation of the cryptic *bgl* operon in *Escherichia coli* K-12, p. 435–444. *In* K. Drlica and M. Riley (ed.), The bacterial chromosome. American Society for Microbiology, Washington, D.C.
25. **Love, D. R., R. Fischer, 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.
26. **Mahadevan, S., A. E. Reynolds, and A. Wright.** 1987. Positive and negative regulation of the *bgl* operon of *Escherichia coli*. J. Bacteriol. **169**:2570–2578.
27. **Mahadevan, S., and A. Wright.** 1987. A bacterial gene involved in transcription antitermination: regulation at a *rho*-independent terminator in the bgl operon of *E. coli*. Cell **50**:485–494.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Mantei, N., M. Villa, T. Enzler, H. Wacker, W. Boll, P. James, W. Hunziker, and G. Semenza.** 1988. Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. EMBO J. **7**:2705–2713.
30. **Meagher, R. B.** 1977. Protein expression in *E. coli* minicells by recombinant plasmids. Cell **10**:521–536.
31. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Murphy, N., D. J. McConnell, and B. A. Cantwell.** 1984. The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme β-glucanase. Nucleic Acids Res. **12**:5355–5367.
33. **Parker, L. L., and B. G. Hall.** 1988. A fourth *Escherichia coli* gene system with the potential to evolve β-glucoside utilization. Genetics **119**:485–490.
34. **Parker, L. L., and B. G. Hall.** 1990. Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. Genetics **124**:455–471.
35. **Porter, E. V., and B. M. Chassy.** 1988. Nucleotide sequence of the β-D-phosphogalactoside galactohydrolase gene of *Lactobacillus casei*: comparison to analogous *pbg* genes of other gram-positive organisms. Gene **62**:263–276.
36. **Reynolds, A. E., J. Felton, and A. Wright.** 1981. Insertion of DNA activates the cryptic bgl operon in Escherichia coli. Nature (London) **293**:625–629.
37. **Reynolds, A. E., S. Mahadevan, S. Legrice, and A. Wright.** 1986. Enhancement of bacterial gene expression by insertion elements or mutation in a CAP-cAMP binding site. J. Mol. Biol. **191**:85–95.
38. **Rissler, J. L., M. O. Delorme, H. Delacroix, and A. Henaut.** 1988. Amino acid substitutions in structurally related proteins. Determination of a new and efficient scoring matrix. J. Mol. Biol. **204**:1019–1029.
39. **Saier, M. H., P. K. Werner, and M. Müller.** 1989. Insertion of proteins into bacterial membranes: mechanism, characteristics and comparison with the eucaryotic process. Microbiol. Rev. **53**:333–366.
40. **Saier, M. H., Jr., M. Yamada, B. Erni, K. Suda, J. Lengeler, R. Ebner, P. Argos, B. Rak, K. Schnetz, C. A. Lee, G. G. Stewart, F. Breidt, E. B. Waygood, K. G. Peri, and R. F. Doolittle.** 1988. Sugar permeases of the bacterial phosphoenolpyruvate-dependent phosphotransferase system: sequence comparisons. FASEB J. **2**:199–208.
41. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
42. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
43. **Schaefler, S.** 1967. Inducible system for the utilization of β-glucosides in *Escherichia coli*. I. Active transport and utilization of β-glucosides. J. Bacteriol. **93**:254–263.
44. **Schaefler, S., and W. K. Maas.** 1967. Inducible systems for the utilization of β-glucosides in *Escherichia coli*. J. Bacteriol. **93**:264–272.
45. **Schaefler, S., and A. Malamy.** 1969. Taxonomic investigations on expressed and cryptic phospho-β-glucosidases in *Enterobacteriaceae*. J. Bacteriol. **99**:422–433.
46. **Schnetz, K., and B. Rak.** 1988. Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. EMBO J. **7**:3271–3277.
47. **Schnetz, K., S. L. Sutrina, M. H. Saier, and B. Rak.** 1990. Identification of catalytic residues in the β-glucoside permease of *Escherichia coli* by site-specific mutagenesis and demonstration of interdomain cross-reactivity between the β-glucoside and glucose systems. J. Biol. Chem. **265**:13464–13471.
48. **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**:2579–2790.
49. **Semenza, G., H. C. Curtins, O. Raunhardt, P. Hore, and M. Müller.** 1969. The configuration at the anomeric carbon of the reaction products of some digestive carbohydrases. Carbohydr. Res. **10**:417–428.
50. **Sinnott, M. L.** 1990. Catalytic mechanisms of glycosyl transfer. Chem. Rev. **90**:1171–1202.
51. **Warkarchuk, 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.
52. **Withers, S. G., A. J. R. Warren, I. P. Street, K. Rupitz, J. B. Kempton, and R. Aebersold.** 1990. Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. J. Am. Chem. Soc. **112**:5887–5889.