JOURNAL OF BACTERIOLOGY, Feb. 1990, p. 802–807
0021-9193/90/020802-06$02.00/0

Vol. 172, No. 2

Nucleotide Sequence and Transcriptional Analysis of the *Escherichia coli agp* Gene Encoding Periplasmic Acid Glucose-1-Phosphatase

ELIZABETH PRADEL,† CHRISTIAN MARCK, AND PAUL LOUIS BOQUET*

Service de Biochimie, Département de Biologie, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-sur-Yvette Cedex, France

Received 21 April 1989/Accepted 12 September 1989

The nucleotide sequence of the *agp* gene, which encodes a periplasmic glucose-1-phosphatase, was determined. The deduced amino acid sequence corresponds to a 413-amino-acid-residue polypeptide with a typical hydrophobic signal sequence of 22 amino acids. The mature protein lacks the N-terminal signal peptide and has a calculated M_r of 43,514. Its promoter was defined by primer extension of the mRNA made in vivo. Like many genes under positive control, its −35 promoter region does not match the consensus. The *agp* gene is both preceded and followed by transcription termination signals, so it appears to be transcribed as a single unit.

The number of distinct acid phosphatase species present in the periplasmic space of *Escherichia coli* and their biological significance are still uncertain.

The cloning in a wild-type *E. coli* strain of genes which in the multicopy state cause the overexpression of acid phosphatase activities previously led to the identification of *agp*, the structural gene for a periplasmic acid glucose-1-phosphatase (G1Pase). This enzyme, which is optimally active around pH 4, was shown to consist of two subunits of M_r 44,000 each. Deletion analysis and Tn*phoA* transpositions into *agp*, generating Agp-PhoA hybrid proteins of different molecular weight, indicated that *agp* was approximately 1.26 kilobases long (22). By homologous recombination of a plasmid-borne *agp-phoA*-IS*50*$_L$-*neo* (Km^r) construction into the chromosome, *agp* was mapped to minute 22.5 on the *E. coli* linkage map. The presence of a hexose phosphatase with similar relative specificity in the periplasmic space of *Salmonella typhimurium* has been reported (25). The ability of *E. coli* strains lacking the hexose phosphate transport system to utilize glucose-1-phosphate (G-1-P) as the sole carbon source has also been considered as an indication that some periplasmic phosphatase could efficiently cleave G-1-P (13). Inactivating *agp* on the chromosome of a wild-type strain indicated that G1Pase was indeed absolutely required for the growth of *E. coli* in a high-phosphate medium containing G-1-P as the sole carbon source (23). The regulatory characteristics of *agp* (independence from inorganic phosphate availability and apparent submission to catabolite repression) are in good agreement with such a physiological role for G1Pase (23).

We report here the complete nucleotide sequence of *agp*, including 5′ and 3′ flanking regions. The sequenced DNA fragment includes two putative promoters. One of them resembles that of positively controlled operons and is shown here to be the actual promoter for *agp* transcription in vivo. The deduced amino acid sequence of G1Pase indicates the presence of a typical hydrophobic N-terminal signal peptide, and the position of its cleavage by the leader peptidase was confirmed by N-terminal amino acid sequencing of the mature protein.

* Corresponding author.
† Present address: Department of Microbiology, Arizona State University, Tempe, AZ 85287-2701.

MATERIALS AND METHODS

Bacterial strains, bacteriophage, and plasmids. *E. coli* K-12 JM101 [*thi supE* Δ(*lac-proAB*) F′ *traD36 proAB lacI*^{q}ΔM15] was used as a host for the propagation of phage M13mp18 and M13mp19 (33) and was grown in LB medium. Strain SBS1405 (22) [F$^-$ Δ(*ara-leu*)*7679 araD139* Δ*lacX74 galE galK* Δ*phoA20 thi rpsE rpoB argE*(Am) *recA1* Δ*appA*::IS*50*$_L$-*neo* (Km^r)] was used for plasmid propagation and determination of the transcription start site.

Plasmid pEP1376 is a derivative of pBR322 carrying a 1.8-kilobase DNA fragment containing the entire *agp* gene (22).

DNA sequencing. The restriction fragments of the *agp* gene used for DNA sequencing were isolated from pEP1376 and subcloned into phage M13mp18 and M13mp19 (33). All sequences were determined by the dideoxy-chain termination method (27, 29) with [α-^{35}S]dATP (>30 TBq/mmol) and Sequenase (U.S. Biochemical Corporation, Cleveland, Ohio). In some cases, dITP was used in place of dGTP to ensure more reliable reading through areas of high G+C content.

Computer analysis of nucleotide sequence. Sequence data were analyzed with the help of the DNA Strider 1.0 program (21).

RNA transcript mapping. RNA was isolated by the method of Aiba et al. (1). The primer extension experiment was adapted from Ye and Larson (34). The synthetic oligonucleotide (20-mer) primer (10 ng) was added to 12 μg of RNA in 4 μl. The samples were boiled for 2 min and quickly cooled in ice. The hybridized primer was extended by mixing the following constituents: 50 mM Tris hydrochloride (pH 8.0), 50 mM NaCl, 6 mM $MgCl_2$, 5 mM dithiothreitol, 200 μM each dCTP, dGTP, and dTTP, 2 μM [α-^{35}S]dATP (>30 TBq/mmol), 40 U of RNase inhibitor (Promega Corp., Madison, Wis.), and 23 U of avian myeloblastosis virus reverse transcriptase (Appligene Lab., Illkirch, France) in a total volume of 10 μl. The mixture was incubated at 42°C for 30 min, and then 5 ml of a chase solution containing 200 mM each dATP, dCTP, dGTP, and dTTP was added. Incubation

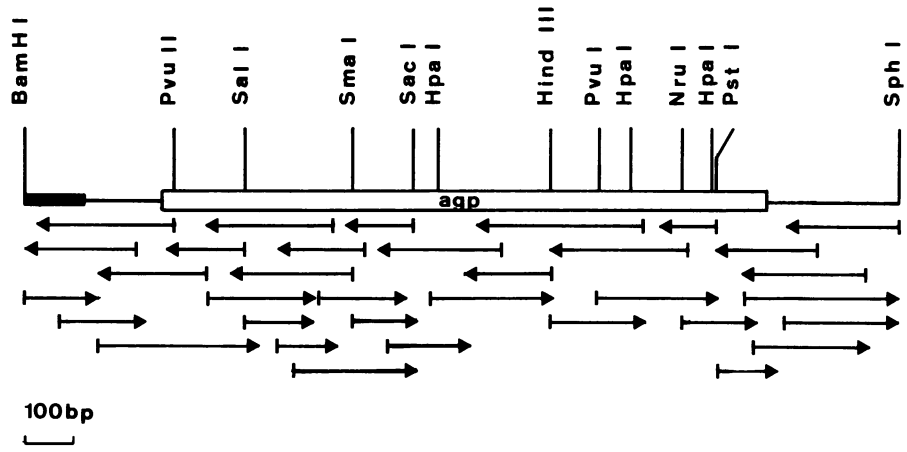

FIG. 1. Strategy for sequencing the insert DNA of pEP1376. Key restriction sites are shown for reference. The arrows indicate the direction of sequencing, and their lengths correspond to accurate sequence readings. The physical location of the *agp* gene is represented by an open box, and the black bar illustrates a 117-bp fragment of the phage Mu "right" extremity (14, 15).

was for 30 min at 30°C. The reaction was stopped by adding 7 μl of a denaturating solution (95% formamide, 0.05% xylene cyanol, 0.05% bromophenol blue, 20 mM EDTA) and heating for 2 min at 90°C before analysis on a sequencing gel. The 450-base-pair (bp) *Bam*HI-*Sal*I fragment from pEP1376 was cloned into M13mp18. The single-stranded DNA template corresponding to the mRNA identical strand was hybridized with the same synthetic oligonucleotide, and the products of the sequencing reactions were loaded on the same gel to give a reference sequencing ladder.

Determination of N-terminal amino acid sequence of G1Pase. G1Pase was purified from the osmotic shock fluid of a 1-liter culture of strain SBS1405(pEP1376) as previously described (22). The concentrated preparation was applied to a reverse-phase high-pressure liquid chromatography C4 column (4.5 by 250 mm, 300-Å porosity) and eluted (0.75 ml · min^{-1}) with a linear gradient of acetonitrile (0 to 60% in 60 min) in 0.1% trifluoroacetic acid. The active fractions were determined by testing for *p*-nitrophenyl phosphate hydrolysis at pH 4 (22). The first nine amino acids of the high-pressure liquid chromatography-purified protein were identified by using an Applied Biosystems 477A gas-phase sequenator and on-line automated high-pressure liquid chromatography.

RESULTS

Nucleotide sequence of *agp*. The nucleotide sequence of the 1.8-kilobase *Bam*HI-*Sph*I fragment present in plasmid pEP1376 (22) was determined by the strategy indicated in Fig. 1. Sequence data were obtained from both strands, with overlaps covering each of the restriction sites used to subclone fragments into M13 derivatives. The first 117 bp correspond to a small fragment of the "right" extremity of the initial cloning phasmid Mu dII4042, which was conserved in the subcloning of *agp* into pBR322 (14, 15, 22). This fragment does not contain any known or potential promoter. The complete nucleotide sequence of *agp* and its flanking regions is shown in Fig. 2. Translation of the nucleotide sequence in the three reading frames and on both strands revealed a single open reading frame of 1,239 nucleotides terminated by a TAA codon (Fig. 2). The size of this open reading frame, its orientation, and its position relative to several restriction sites are in perfect agreement with the data previously obtained from the analysis of some Tn*phoA* insertions into *agp* (22). It has the coding capacity of a protein of 413 amino acid residues, including a presumed signal peptide of 22 residues. The amino acid composition of this polypeptide (Table 1) is in agreement with the periplasmic location of the G1Pase.

Flanking regions. Upstream from this open reading frame, from nucleotides −149 to −121, lies an inverted repeat which should be able to form a hairpin stem-and-loop structure if transcribed into RNA. This potential structure has a calculated free energy of −11.3 kcal · mol^{-1} (35) and could function as a rho-independent transcription termination signal for a similarly oriented gene (genes) next to *agp* on the chromosome (7, 26). From nucleotide −125 to nucleotide −92 is another up-promoter region of dyad symmetry with a stem and loop that overlaps with the last five nucleotides of the above terminator.

In the downstream region 6 bp beyond the ochre termination codon (from positions 1249 to 1273) is a G+C-rich region of dyad symmetry which very likely corresponds to a hairpin structure in the mRNA, with a calculated ΔG value of −28.7 kcal · mol^{-1} (35), and which may function as a rho-independent terminator. Thus, the *agp* gene is enclosed between two potential transcription termination signals.

Putative promoters. At positions −39 to −34 upstream of the ATG start codon lies a possible −10 hexamer, CACACT (putative P1 promoter, 50% homology with TATAAT), but at the appropriate distance (16 bp), the −35 hexamer shows a poor homology to the consensus (CGGAAA compared with TTGACA) (16). Further upstream, however, is another possible −10 hexamer (TCATAT) from nucleotides −51 to −46, with a −35 sequence (TTGCGA) 19 bp upstream from positions −76 to −71 (putative P2 promoter). A presumed Shine-Dalgarno sequence, GAGGTG (28), is located 6 bp upstream of the initiation codon.

To determine which of the putative promoters, P1 or P2, is used in vivo and to map the precise origin(s) of transcription, we used the primer extension method described in Materials and Methods. Cellular RNA was extracted from a strain carrying either the recombinant plasmid pEP1376 or, as a control, plasmid pBR322. The 20-mer oligonucleotide 5′-CCCTGCCACAGCTGCGGCGA-3′ complementary to nucleotides 17 to 36 (Fig. 2) was annealed to the RNA prepa-

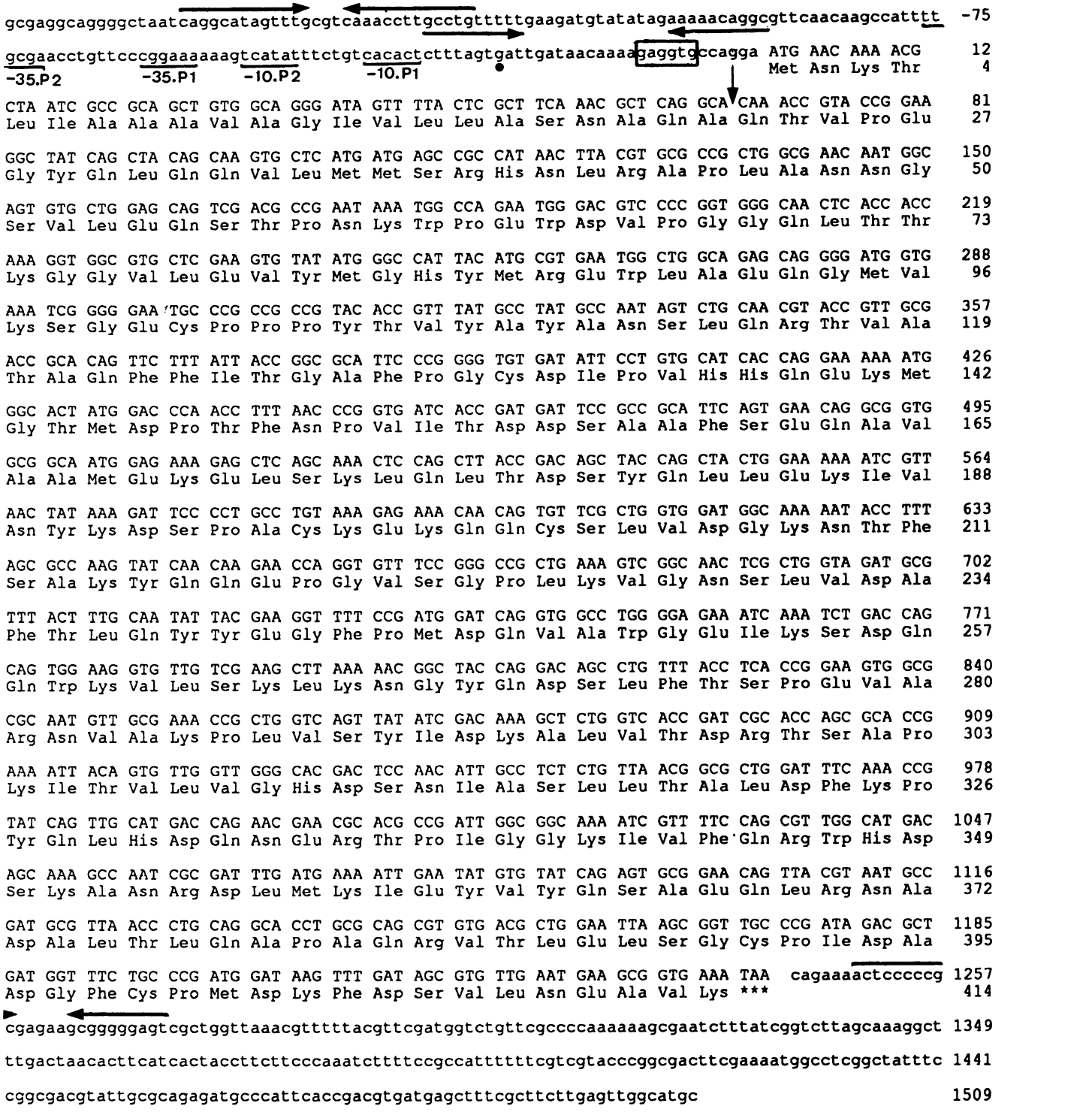

FIG. 2. Nucleotide sequence of the *agp* gene and its flanking regions and deduced amino acid sequence of the G1Pase. The Mu sequence is not shown. Numbering of nucleotides starts at the A of the translation initiation codon. The box indicates the potential ribosome-binding site. Potential promoters P1 and P2 are underlined. Regions of dyad symmetry are indicated by arrows. The dot indicates the transcriptional initiation site. The vertical arrow shows the signal peptide cleavage site.

rations and extended with reverse transcriptase. The primer was also hybridized to the single-stranded DNA of an M13 recombinant clone harboring nucleotides −166 to 165, and a sequencing ladder was prepared by the dideoxy method. An autoradiogram of a portion of the gel corresponding to this experiment is shown in Fig. 3. The cDNA strand extended from the primer annealed to the mRNA of the strain carrying pEP1376 (*agp*) ends at the C at position −26. Consequently, the transcription of *agp* in pEP1376 starts with the G (−26) on the antisense strand. Thus, at least under the conditions

TABLE 1. Amino acid composition of G1Pase protein precursor deduced from the nucleotide sequence

Amino acid	n (%)	Molecular weight (%)
Alanine	38 (9.2)	2,699 (5.9)
Cysteine	6 (1.5)	618 (1.4)
Aspartic acid	24 (5.8)	2,760 (6.0)
Glutamic acid	22 (5.3)	2,838 (6.2)
Phenylalanine	13 (3.1)	1,911 (4.2)
Glycine	25 (6.1)	1,425 (3.1)
Histidine	7 (1.7)	959 (2.1)
Isoleucine	14 (3.4)	1,583 (3.5)
Lysine	27 (6.5)	3,458 (7.6)
Leucine	38 (9.2)	4,297 (9.4)
Methionine	12 (2.9)	1,572 (3.4)
Asparagine	18 (4.4)	2,052 (4.5)
Proline	24 (5.8)	2,329 (5.1)
Glutamine	30 (7.3)	3,841 (8.4)
Arginine	11 (2.7)	1,717 (3.8)
Serine	27 (6.5)	2,349 (5.1)
Threonine	23 (5.6)	2,324 (5.1)
Valine	32 (7.7)	3,170 (6.9)
Tryptophan	6 (1.5)	1,116 (2.4)
Tyrosine	16 (3.9)	2,609 (5.7)

of growth described here, the only promoter used in vivo is P1, which lacks the consensus −35 sequence.

Signal peptide. Most periplasmic proteins are synthesized as precursors, with a hydrophobic amino-terminal signal sequence that is removed when the protein crosses the cytoplasmic membrane. The N-terminal extremity of Agp, deduced from the nucleotide sequence, shows such a typical leader peptide of 22 amino acid residues with a potential site for recognition by the signal peptidase I, Ala-Gln-Ala, and a positively charged amino acid at position 3 (Lys), followed by a stretch of hydrophobic amino acids (Fig. 2) (31). Amino acid sequence analysis of the N-terminal region of the purified mature protein indicated the order NH_2-Gln-Thr-Val-Pro-Glu-Gly-Tyr-Gln-Leu. . ., confirming the position of the cleavage site between residues 22 and 23. Consequently, the signal peptide has an M_r of 2,139 and the mature protein has an M_r of 43,514, which is in good agreement with the value of 44,000 previously estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (22).

DISCUSSION

The genes for four phosphatases found in the periplasmic space of *E. coli* have been identified, mapped, cloned, and sequenced: *phoA* (4, 10, 17), *ushA* (3, 8, 9, 11), *cpdB* (2), and *appA* (5, 12; E. Dassa and P. L. Boquet, manuscript in preparation). The *agp* gene encodes a fifth periplasmic phosphatase with a preferential affinity for the substrate G-1-P which had not been previously described among the activities purified from shock fluids of wild-type *E. coli* strains (22). Unlike *phoA* and *appA*, *agp* expression in vivo is not influenced by inorganic phosphate in the medium, and opposite to *appA*, it is positively controlled by cyclic AMP (cAMP) and the cAMP receptor protein (CRP) (23). Despite differences in the K_m values, the relative efficiency of the *E. coli* G1Pase on several hexose phosphates is similar to that of an *S. typhimurium* periplasmic acid hexose phosphatase (19, 25, 32) also shown to be positively controlled by CRP-cAMP in vivo (20). The product of *agp* was previously shown to be necessary for the utilization of G-1-P as the sole carbon source in a high-phosphate minimal medium (23).

We determined the nucleotide sequence of a DNA fragment containing the whole *agp* gene. The open reading frame that was found is in excellent agreement with our previous data, encoding a protein of 413 amino acid residues, including a hydrophobic signal peptide of 22 amino acid residues. The predicted mature polypeptide contains 83 charged residues among a total of 391 amino acids (Asp + Glu + Arg + Lys = 21 mol%), which is typical of soluble proteins (Table 1) and consistent with its location in the periplasm. The small sequence Asp-174-Ser-Ala-Ala-177 is identical to that found in the active site of alkaline phosphatase and in several other serine hydrolases (6).

Comparison of the G1Pase amino acid sequence with that of the *E. coli* pH 2.5 acid phosphatase (the product of gene *appA*) has shown a significant homology between the two proteins (Dassa and Boquet, in preparation). No homology was found with any of the protein sequences stored in the EMBL data bank with the exception of a published N-terminal fragment of AppA (30).

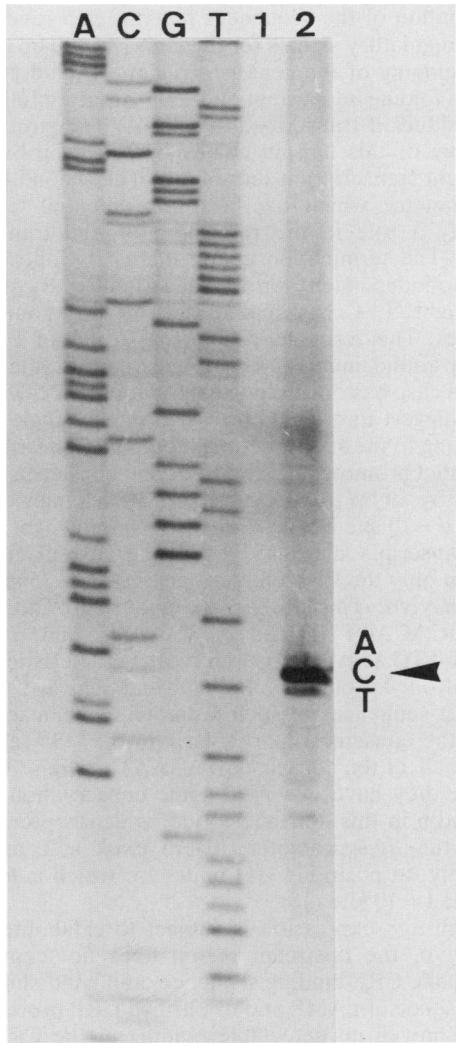

FIG. 3. Primer extension mapping of the *agp* mRNA 5′ termini. Avian myeloblastosis virus reverse transcriptase and the 20-mer oligonucleotide complementary to nucleotides 17 to 36 of the *agp* gene were added to total *E. coli* RNA isolated from SBS1405 (pBR322) (lane 1) or SBS1405(pEP1376) (lane 2). The sequence ladder generated by using the same primer is shown in lanes A, C, G, and T. The arrow indicates the major transcription initiation site.

Examination of the sequences flanking *agp* reveals three potential regulatory signals for the gene. At 120 bp upstream of the beginning of the gene is a region of dyad symmetry capable of folding into a stem-and-loop structure followed by five U residues if transcribed into RNA. The structure and free energy of this hairpin correspond to a probable rho-independent transcription terminator. The second region of dyad symmetry which overlaps this potential terminator could play a role in the regulation of *agp* transcription initiation. The termination codon of *agp* is followed by a typical rho-independent terminator structure. Its stem consists of eight G · C base pairs, and the loop contains five nucleotides. This terminator lacks the stretch of T residues frequently found immediately downstream of the hairpin, but such a case has been reported previously (7). These data strongly suggest that *agp* is transcribed as a single cistron.

According to the sequence, *agp* could be transcribed from two potential promoters. One has a good consensus −35 site separated by 19 bp from a possible −10 hexamer (P2). The other has a −10 site but no −35 site (P1). However, the size of the transcripts measured by primer extension clearly shows that only the P1 promoter, lacking the −35 sequence, is used in vivo. This absence of good −10 and −35 sequences (CACACT instead of TATAAT and CGGAAA instead of TTGACA) is apparently a characteristic of positively controlled promoters. Raibaud and Schwartz (24) have shown that sequences of such promoters significantly deviate from the consensus, particularly in the −35 region and for the first T of the consensus TATAAT in the −10 region. However, they have observed some conservation at positions 2 and 6 in this hexanucleotide, which is precisely the case for the *agp* promoter. There exist a T and a G, respectively, at positions −14 and −15, which is typical of an extended −10 site (18).

Although *agp* expression is subject to catabolite repression in vivo, the upstream region does not contain any consensuslike CRP-binding sequence, and band-shift assays in the presence of cAMP and of purified CRP protein (up to 200 nM) showed no detectable binding of the CRP-cAMP complex to this region (E. Pradel, unpublished data). Since *agp* transcription is very likely monocistronic, this suggests that the positive control exerted by cAMP and its specific receptor is indirect. Preliminary experiments have shown that in addition to the nature of the carbon source used, G1Pase is poorly expressed in minimal media compared with broth-containing media (E. Pradel, unpublished data). Additional information about the in vivo regulation of *agp* is needed to identify possible transcriptional effectors.

ACKNOWLEDGMENTS

We thank Jean-Marie Buhler and Isabelle Treich for their technical advice and the gift of avian myeloblastosis virus reverse transcriptase and Catherine Lepesant for oligonucleotide synthesis. We also thank Jean-Michel Verdier for the gift of RNase inhibitor and Annie Kolb for the purified CRP factor. We are indebted to Gilles Mourier and Eric Gatineau for their kind help in purifying the Agp protein by high-pressure liquid chromatography and to Françoise Bouet for determination of the N-terminal sequence of the protein. We also thank Robert Swanson for kindly reviewing the manuscript.

LITERATURE CITED

1. **Aiba, H., S. Adhya, and B. de Crombrugghe.** 1981. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. J. Biol. Chem. **256**:11905–11910.
2. **Beacham, I. R., and S. Garrett.** 1980. Isolation of *Escherichia coli* mutants (*cpdB*) deficient in periplasmic 2′:3′-cyclic phosphodiesterase and genetic mapping of the *cpdB* locus. J. Gen. Microbiol. **119**:31–34.
3. **Beacham, I. R., and E. Yagil.** 1976. Genetic location of the gene (*ush*) specifying periplasmic uridine 5′-diphosphate glucose hydrolase (5′-nucleotidase) in *Escherichia coli* K-12. J. Bacteriol. **128**:487–489.
4. **Berg, P. E.** 1981. Cloning and characterization of the *Escherichia coli* gene coding for alkaline phosphatase. J. Bacteriol. **146**:660–667.
5. **Boquet, P. L., C. Manoil, and J. Beckwith.** 1987. Use of Tn*phoA* to detect genes for exported proteins in *Escherichia coli*: identification of the plasmid-encoded gene for a periplasmic acid phosphatase. J. Bacteriol. **169**:1663–1669.
6. **Bradshaw, R. A., F. Cancedda, L. H. Ericsson, P. A. Neumann, S. P. Piccoli, M. J. Schlesinger, K. Shrieffer, and K. A. Walsh.** 1981. Amino acid sequence of *Escherichia coli* alkaline phosphatase. Proc. Natl. Acad. Sci. USA **78**:3473–3477.
7. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12**:4411–4427.
8. **Burns, D. M., L. J. Abraham, and I. R. Beacham.** 1983. Characterisation of the *ush* gene of *Escherichia coli* and its protein products. Gene **25**:343–353.
9. **Burns, D. M., and I. R. Beacham.** 1986. Nucleotide sequence and transcriptional analysis of the *E. coli ushA* gene, encoding periplasmic UDP-sugar hydrolase (5′-nucleotidase): regulation of the *ushA* gene, and the signal sequence of its encoded protein product. Nucleic Acids Res. **14**:4325–4342.
10. **Chang, C. N., W.-J. Kuang, and E. Y. Chen.** 1986. Nucleotide sequence of the alkaline phosphatase gene of *Escherichia coli* K-12. Gene **44**:121–125.
11. **Cowman, A., and I. R. Beacham.** 1980. Molecular cloning of the gene (*ush*) from *Escherichia coli* specifying periplasmic UDP-sugar hydrolase (5′-nucleotidase). Gene **12**:281–286.
12. **Dassa, E., and P. L. Boquet.** 1985. Identification of the gene *appA* for the acid phosphatase (pH optimum 2.5) of *Escherichia coli*. Mol. Gen. Genet. **200**:68–73.
13. **Dietz, G. W., and L. A. Heppel.** 1971. Studies on the uptake of hexose phosphates. III. Mechanism of uptake of glucose-1-phosphate in *Escherichia coli*. J. Biol. Chem. **246**:2891–2897.
14. **Groenen, M. A. M., E. Timmers, and P. van de Putte.** 1985. DNA sequences at the ends of the genome of bacteriophage Mu essential for transposition. Proc. Natl. Acad. Sci. USA **82**: 2087–2091.
15. **Groisman, E. A., B. A. Castilho, and M. J. Casadaban.** 1984. *In vivo* DNA cloning and adjacent gene fusing with a mini-Mu-lac bacteriophage containing a plasmid replicon. Proc. Natl. Acad. Sci. USA **81**:1480–1483.
16. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15**:2343–2361.
17. **Inouye, H., S. Michaelis, A. Wright, and J. Beckwith.** 1981. Cloning and restriction mapping of the alkaline phosphatase structural gene (*phoA*) of *Escherichia coli* and generation of deletion mutants in vitro. J. Bacteriol. **146**:668–675.
18. **Keilty, S., and M. Rosenberg.** 1986. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. J. Biol. Chem. **262**:6389–6395.
19. **Kier, L. D., R. Weppelman, and B. N. Ames.** 1977. Resolution and purification of three periplasmic phosphatases of *Salmonella typhimurium*. J. Bacteriol. **130**:399–410.
20. **Kier, L. D., R. Weppelman, and B. N. Ames.** 1977. Regulation of two phosphatases and a cyclic phosphodiesterase of *Salmonella typhimurium*. J. Bacteriol. **130**:420–428.
21. **Marck, C.** 1988. "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
22. **Pradel, E., and P. L. Boquet.** 1988. Acid phosphatases of *Escherichia coli*: molecular cloning and analysis of *agp*, the structural gene for a periplasmic acid glucose phosphatase. J. Bacteriol. **170**:4916–4923.
23. **Pradel, E., and P. L. Boquet.** 1989. Mapping of the *Escherichia coli* acid glucose-1-phosphatase gene *agp* and analysis of its expression in vivo by use of an *agp-phoA* protein fusion. J. Bacteriol. **171**:3511–3517.

24. **Raibaud, O., and M. Schwartz.** 1984. Positive control of transcription initiation in bacteria. Annu. Rev. Genet. **18**:173–206.
25. **Rephaeli, A. W., I. R. Artenstein, and M. H. Saier, Jr.** 1980. Physiological function of periplasmic acid hexose phosphatase in *Salmonella typhimurium*. J. Bacteriol. **141**:1474–1477.
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. **Shine, J., and C. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S-ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
29. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
30. **Touati, E., and A. Danchin.** 1987. The structure of the promoter and amino terminal region of the pH 2.5 acid phosphatase structural gene, *appA*, of *E. coli*: a negative control of transcription mediated by cyclic AMP. Biochimie **69**:215–221.
31. **von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage site. Eur. J. Biochem. **133**:17–21.
32. **Weppelman, R., L. D. Kier, and B. N. Ames.** 1977. Properties of two phosphatases and a cyclic phosphodiesterase of *Salmonella typhimurium*. J. Bacteriol. **130**:411–419.
33. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
34. **Ye, S., and T. J. Larson.** 1988. Structures of the promoter and operator of the *glpD* gene encoding aerobic *sn*-glycerol-3-phosphate deshydrogenase of *Escherichia coli* K-12. J. Bacteriol. **170**:4209–4215.
35. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.