JOURNAL OF BACTERIOLOGY, Apr. 1996, p. 2051–2059 Vol. 178, No. 7
0021-9193/96/$04.00+0


# *glc* Locus of *Escherichia coli*: Characterization of Genes Encoding the Subunits of Glycolate Oxidase and the *glc* Regulator Protein

MARIA-TERESA PELLICER, JOSEFA BADÍA, JUAN AGUILAR,* AND LAURA BALDOMÀ

*Department of Biochemistry, Faculty of Pharmacy, University of Barcelona, 08028 Barcelona, Spain*

Received 18 October 1995/Accepted 29 January 1996

**The locus *glc* (min 64.5), associated with the glycolate utilization trait in *Escherichia coli*, is known to contain *glcB*, encoding malate synthase G, and the gene(s) needed for glycolate oxidase activity. Subcloning, sequencing, insertion mutagenesis, and expression studies showed five additional genes: *glcC* and in the other direction *glcD*, *glcE*, *glcF*, and *glcG* followed by *glcB*. The gene *glcC* may encode the *glc* regulator protein. Consistently a chloramphenicol acetyltransferase insertion mutation abolished both glycolate oxidase and malate synthase G activities. The proteins encoded from *glcD* and *glcE* displayed similarity to several flavoenzymes, the one from *glcF* was found to be similar to iron-sulfur proteins, and that from *glcG* had no significant similarity to any group of proteins. The insertional mutation by a chloramphenicol acetyltransferase cassette in either *glcD*, *glcE*, or *glcF* abolished glycolate oxidase activity, indicating that presumably these proteins are subunits of this enzyme. No effect on glycolate metabolism was detected by insertional mutation in *glcG*. Northern (RNA) blot experiments showed constitutive expression of *glcC* but induced expression for the structural genes and provided no evidence for a single polycistronic transcript.**

Glycolate is metabolized in *Escherichia coli* through oxidation to glyoxylate (15, 18) in a reaction catalyzed by the enzyme glycolate oxidase (21, 27). Glyoxylate is a branching point in the metabolic pathway since it is metabolized by two divergent condensation reactions. One reaction condenses glyoxylate with acetyl coenzyme A and is catalyzed by malate synthase G (36), while the other reaction condenses two molecules of glyoxylate in a process catalyzed by glyoxylate carboligase, which simultaneously decarboxylates the condensation product to tartronic semialdehyde (7). This latter compound is reduced to glycerate and subsequently phosphorylated to glycerate-3-phosphate. These three enzyme-catalyzed reactions constitute what is known as the glycerate pathway.

Locus *glc* at min 64.5 of the *E. coli* chromosome has been associated with the glycolate utilization trait (25). This locus contains the *glcB* gene, encoding malate synthase G, and also the gene(s) encoding glycolate oxidase. However, the *gcl* gene, encoding glyoxylate carboligase, is not linked to the *glc* locus; it has been located in a different position, at min 12 (7). The *glcB* gene and enzyme have been characterized previously (25). In this report we describe and characterize the rest of the genes that complete the *glc* locus and identify those encoding subunits of glycolate oxidase and the regulatory gene of the *glc* operon.

## MATERIALS AND METHODS

**Bacterial strains, plasmids, and phages.** All the strains used were *E. coli* K-12 derivatives. The genotypes and sources of the bacterial strains and plasmids are given in Table 1. Phage P1*vir* was used in transduction experiments.

**Growth conditions.** Cells were grown aerobically on Luria broth or minimal medium as described previously (3). For growth on minimal medium, the following compounds were added at the following concentrations unless otherwise specified: D-xylose, 10 mM; glycolate, 30 mM; glyoxylate, 30 mM; and casein acid hydrolysate (CAA), 1%. When necessary, thiamine at 1.65 μg/ml and the following antibiotics at the concentrations indicated were added to the medium: chloramphenicol, 30 μg/ml; ampicillin, 100 μg/ml; and tetracycline, 12.5 μg/ml. 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and isopropyl-β-D-thiogalactoside (IPTG) were used at 30 and 10 μg/ml, respectively.

**Preparation of cell extracts and enzyme assays.** For enzyme assays, cells were harvested at the end of the exponential phase and cell extracts were prepared as described previously (4) in 10 mM Tris-HCl, pH 8.0, containing 1 mM $MgCl_2$. Determinations of total malate synthase activity and the relative concentration of the two forms of malate synthase, A and G, were performed as described by Ornston and Ornston (27).

Glycolate oxidase activity was determined spectrophotometrically according to the method of Lord (21). In this case, to avoid nonspecific reduction of 2,6-dichloroindophenol, the 30% ammonium sulfate fraction resuspended in 10 mM phosphate buffer, as indicated by this author, was routinely used as a source of enzyme.

Protein concentration was determined by the method of Lowry et al. (22) using bovine serum albumin as the standard.

**In vivo expression of plasmid-carried genes in maxicells.** The expression of plasmid-encoded proteins was performed in maxicells as described by Stoker et al. (34) with strain JA120 as a host. Proteins were labeled with L-[$^{35}$S]methionine (50 μCi/ml) and detected by fluorography after separation by sodium dodecyl sulfate–10 or 13.5% polyacrylamide gel electrophoresis (SDS-PAGE).

**DNA manipulation.** Plasmid DNA was routinely prepared by the boiling method (17). For large-scale preparation, a crude DNA sample was purified through a column (Qiagen GmbH, Düsseldorf, Germany). DNA manipulations were performed essentially as described by Sambrook et al. (29). The DNA sequence was determined by the dideoxy-chain termination procedure of Sanger et al. (30) by using the T7 sequencing kit from Pharmacia LKB Biotechnology. Double-stranded plasmid DNA was used as the template. Ordered deletions were obtained with the Erase-a-Base system (Promega Biotec, Madison, Wis.). To avoid sequencing errors, both strands were sequenced by regular dGTP as well as dITP or 7-deaza-dGTP reactions.

DNA sequence analysis and protein alignments were done with the PC/GENE software package (Intelligenetics Inc.) and with the University of Wisconsin Genetics Computer Group package programs.

**Genetic techniques.** Phage P1 transduction experiments were performed as described by Miller (24). The chloramphenicol resistance gene cassette CAT19 was used in the gene inactivation experiments (14, 38) by inserting it into the restriction sites indicated in Fig. 4. This cassette had no terminator designed into the downstream region of the chloramphenicol acetyltransferase (CAT) gene and thus did not always cause polarity when inserted in the same orientation as the interrupted gene. Plasmids carrying inactivated genes (Table 1) were linearized and used to transform strain JC7623 to chloramphenicol resistance ($Cm^r$). This strain efficiently recombines linear DNA into its chromosome (38). P1*vir* lysates obtained from the selected $Cm^r$ recombinants were used to transduce the CAT insertions into the parental strain MC4100. The locations of these mutations in the *glc* locus were verified by cotransductional analysis with the

* Corresponding author. Mailing address: Department of Biochemistry, Faculty of Pharmacy, University of Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain. Phone: 34-3-402 4521. Fax: 34-3-402 1896. Electronic mail address: Jaguilar@far.ub.es.

TABLE 1. Bacterial strains and plasmids used in this study

| Strain or plasmid | Description | Source or reference |
|---|---|---|
| Strains | | |
| XL1Blue | *recA1 lac endA1 gyrA96 thi hsdR17 supE44 relA1* (F′ *proAB lacI*$^q$ *lacZ* Δ*M15* Tn*10*) | Stratagene |
| MC4100 | *araD* Δ*lac rpsL flbB deoC ptsF rbsR* | 6 |
| SØ1023 | *relA1 spoT1 thi-1 nupG511*::Tn*10* | 32 |
| NK6027 | *relA1 spoT1 thi-1* Δ(*gpt-lac*)*5 metC162*::Tn*10* | 32 |
| JC7623 | *arg thi thr leu pro his strA recB21 recC22 sbcB15* | 37 |
| JA120 | *thi-1 thr-1 leuB6 lacY1 tonA21 supE44 recA1 srlA*::Tn*10* | 1 |
| JA151 | Δ(*speC-glc*) *aceB recA*::*cat* | 25 |
| JA154 | MC4100 *glcC*::*cat* | This study |
| JA155 | MC4100 *glcD*::*cat* | This study |
| JA156 | MC4100 *glcE*::*cat* | This study |
| JA157 | MC4100 *glcF*::*cat* | This study |
| JA158 | MC4100 *glcG*::*cat* | This study |
| Plasmids[a] | | |
| Bluescript | Ap$^r$ | Stratagene |
| pBR322 | Ap$^r$ Tc$^r$ | Boehringer Mannheim |
| pCAT19 | Ap$^r$ Cm$^r$; source of *cat* gene | 14 |
| pLB10 | *glcBGFEDC* (11.5-kb *Hin*dIII fragment) in pUC18 | 25 |
| pIM4 | *glcBGF′* (4.5-kb *Hin*dIII-*Eco*RV fragment) in Bluescript | 25 |
| pTP21 | *glcBGFEDC* (9-kb *Hin*dIII-*Eco*RI fragment) | This study |
| pTP22 | *glcBGFED′* (6.5-kb *Hin*dIII-*Sal*I fragment) | This study |
| pTP23 | *glcBGFE′* (5.2-kb *Hin*dIII-*Bam*HI fragment) | This study |
| pTP24 | *glcB′GFEDC* (6.8-kb *Cla*I-*Eco*RI fragment) | This study |
| pTP25 | *glcF′EDC′* (3.8-kb *Eco*RV fragment) | This study |
| pTP26 | *glcE′DC* (3.2-kb *Bam*HI fragment) | This study |
| pTP27 | *glcE′D′* (3.3-kb *Kpn*I-*Sal*I fragment) | This study |
| pTP28 | *glcB′GFE′* (2.8-kb *Cla*I-*Kpn*I fragment) | This study |
| pTP29 | *glcF′E′* (0.8-kb *Eco*RV-*Bam*HI fragment) | This study |
| pTP30 | *glcD′C* (1.3-kb *Mlu*I-*Bam*HI fragment) | This study |
| pTP51 | As pTP26 insert but *glcC*::*cat*, *Sma*I-Cm$^r$ cassette into *Nru*I | This study |
| pTP52 | As pTP26 insert but *glcD*::*cat*, *Sma*I-Cm$^r$ cassette into *Mlu*I | This study |
| pTP53 | As pTP22 insert but *glcE*::*cat*, *Bam*HI-Cm$^r$ cassette into *Bam*HI | This study |
| pTP54 | As pTP22 insert but *glcF*::*cat*, *Sma*I-Cm$^r$ cassette into *Eco*RV | This study |
| pTP55 | As pTP28 insert but *glcG*::*cat*, *Sma*I-Cm$^r$ cassette into *Nru*I | This study |

[a] Plasmids pTP21 to pTP30 are in Bluescript, and plasmids pTP51 to pTP55 are in pBR322. A prime indicates that a particular gene is truncated.

*metC*::Tn*10* (strain NK6027) and *nupG*::Tn*10* (strain SØ1023) markers (data not shown).

**Isolation of RNA and Northern blot hybridization.** For preparation of total RNA, cells of a 25-ml culture grown to an $A_{650}$ of 0.5 were collected by centrifugation at 5,000 × *g* for 10 min and processed as described by Belasco et al. (2). Northern (RNA) blot hybridization was performed with each RNA sample (10 μg) by following the procedure described previously (26).

**Nucleotide sequence accession number.** The sequence reported here is deposited in the EMBL/GenBank database under accession number L43490.

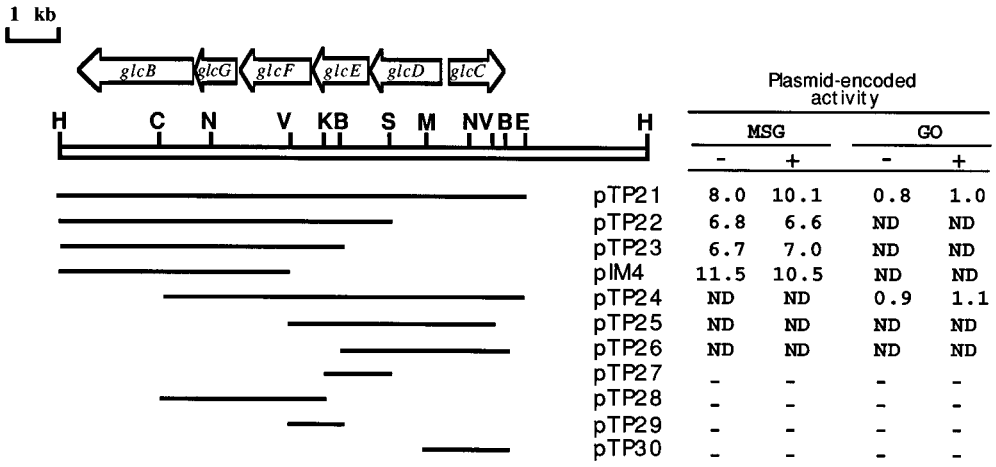
## RESULTS AND DISCUSSION

**Location of glycolate oxidase genes in the *glc* locus.** Previously we have reported that the recombinant plasmid pLB10, containing an 11.5-kb *Hin*dIII fragment, expressed malate synthase G and glycolate oxidase activities and that the pLB10-derived recombinant plasmid pIM4 was able to express only malate synthase G (25). In order to locate the gene encoding glycolate oxidase more precisely (Fig. 1), a set of subclones was derived from pLB10 and used to restore glycolate oxidase activity in strain JA151. Only plasmids containing the 6-kb region between *glcB* and the *Eco*RI restriction site (plasmids pTP21 and pTP24) were able to supply this activity. Enzyme activities were similar in the presence and in the absence of inducer, indicating that expression was from vector promoters.

**Nucleotide sequence of the 5.3-kb fragment upstream of *glcB*.** To identify the genes carried on the genomic fragment upstream of *glcB*, several subclones, covering the region between the *Cla*I (internal to *glcB*) and the second upstream *Bam*HI restriction sites, were constructed (plasmids pTP26, pTP27, pTP28, and pTP29). Serial deletions of these plasmids were obtained and sequenced at least twice on each strand. Figure 2 shows the 5,335 bp of DNA sequenced between the already mentioned *Bam*HI site and the ATG start codon of *glcB* gene previously reported by Molina et al. (25). Five new open reading frames were observed, and they are named *glcD*, *glcE*, *glcF*, and *glcG*, on the same strand, and immediately followed by *glcB*. Diverging from *glcD* on the other strand is *glcC*.

**Expression studies of the sequenced *glc* genes.** The maxicell system (34) was used to identify the gene products of the *glc* locus described above. Strain JA120 was transformed with several plasmids containing different parts of this locus and also with the Bluescript vector. Transformed cells were grown on CAA and CAA plus glycolate. Consistent with the expression from a vector promoter, the same pattern was observed in both conditions. Plasmid pTP21, containing all the genes of this locus, expressed protein bands of 82, 56, and 40 kDa (Fig. 3A, lane 1). Plasmid pTP22, lacking *glcC* and *glcD*, did not express the 56-kDa protein (lane 2), whereas plasmid pTP23, lacking also *glcE*, expressed only the 82-kDa protein (lane 3). Plasmid pIM4, which lacked *glcF*, gave the same expression pattern as pTP23 (lane 4). Consistently, plasmid pTP25 expressed the two proteins of 56 and 40 kDa but not the 82-kDa protein corresponding to malate synthase G (lane 6), while plasmid pTP26 expressed only a protein of 56 kDa (lane 7).

In order to resolve the low-molecular-weight protein corresponding to *glcG*, the expressed products were also separated




| | Plasmid-encoded activity | | | |
|---|---|---|---|---|
| | MSG | | GO | |
| | - | + | - | + |
| pTP21 | 8.0 | 10.1 | 0.8 | 1.0 |
| pTP22 | 6.8 | 6.6 | ND | ND |
| pTP23 | 6.7 | 7.0 | ND | ND |
| pIM4 | 11.5 | 10.5 | ND | ND |
| pTP24 | ND | ND | 0.9 | 1.1 |
| pTP25 | ND | ND | ND | ND |
| pTP26 | ND | ND | ND | ND |
| pTP27 | - | - | - | - |
| pTP28 | - | - | - | - |
| pTP29 | - | - | - | - |
| pTP30 | - | - | - | - |

FIG. 1. Restriction map of the *glc* locus. The open bar represents the genomic fragment present in the recombinant plasmid pLB10 (25). The inserts of the plasmids used in this study are indicated by thin lines below the restriction map. These inserts were cloned in the Bluescript vector with the structural *glc* genes in the same orientation as *lacZ*. For each plasmid, malate synthase G (MSG) and glycolate oxidase (GO) activities are expressed in units per milligram of protein. Enzyme activities were determined in cell extracts of transformed cells of strain JA151 grown on CAA either in the absence (−) or in the presence (+) of glycolate. ND, not detectable. Arrows indicate the extension and direction of transcription of the identified genes. The restriction sites for several cleaving enzymes are indicated as follows: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; M, *Mlu*I; N, *Nru*I; S, *Sal*I; and V, *Eco*RV.

by SDS-13.5% PAGE. A 13-kDa protein was clearly identified in this way. To check that this protein was the product of a *glc* gene and not a product of a truncated gene between *glcB* and the *Hin*dIII restriction site, plasmid pTP28 was also analyzed and proved to express this polypeptide (Fig. 3B).

This expression pattern allowed us to assign the 56-kDa protein to the *glcD* gene, the 40-kDa protein to *glcE*, and the 13-kDa protein to *glcG*. The molecular weights of these expressed proteins were in accordance with those presented below and deduced from the corresponding nucleotide sequence. This system failed to identify *glcF* and *glcC* gene products.

Identification of the *glcF* gene product was achieved when the location of some of these proteins in the cell membrane was determined. To this end, membranes were obtained (28) from *glc*-deleted strain JA151 transformed with plasmid pTP21, using as a control the same strain transformed with the vector or the wild-type strain MC4100 (Fig. 3C). Cultures were grown on CAA plus glycolate. The electrophoretic development of membrane proteins stained by Coomassie blue showed the 56- and 40-kDa proteins encoded by the plasmid and an additional 45-kDa protein, not detected in the maxicell expression experiments, which may correspond to the *glcF* gene product (Fig. 3C). This protein is likely to have a low expression rate and was visualized only after its partial purification through the isolation of membranes.

**Mutational analysis of the sequenced *glc* genes.** The assignment of functions to the sequenced genes was determined by inactivating each of them and studying the effects on glycolate utilization. Inactivation was performed by insertional mutation of the CAT19 cassette in the locus *glc* of the wild-type strain as indicated in Materials and Methods. Insertions in all genes except in *glcG* rendered the cells glycolate negative (Fig. 4).

Determination of enzyme activities in wild-type cells grown under inducing conditions yielded values of 200 mU/mg for malate synthase G and of 60 mU/mg for glycolate oxidase. The inactivation of *glcC* (strain JA154) abolished both malate synthase G and glycolate oxidase, yielding undetectable enzyme activities. Independent inactivation of *glcD*, *glcE*, or *glcF* (strains JA155, JA156, and JA157, respectively) yielded undetectable levels for glycolate oxidase activity but not for malate synthase. These observations strongly suggested the participation of the products of these genes as subunits of the glycolate oxidase structure. Disruption of *glcG* (strain JA158) had no effect on either of the two activities (Fig. 4).

Each inactivation was complemented by a plasmid carrying the fragment in which the insertional mutation was located (Fig. 4), indicating that the inactivation was due to this mutation and that no other cell functions were impaired. The complemented cells displayed glycolate oxidase activities of 40 mU/mg. As expected, in the complementation by *glcC*, both enzyme activities were restored simultaneously (240 mU/mg for malate synthase G and 75 mU/mg for glycolate oxidase).

**Features and predicted properties of *glcC*.** The gene *glcC* starts at position 866 and ends at position 102 (Fig. 2). There is a Shine-Dalgarno sequence (GAGG) 11 bp upstream from the start codon. This gene encodes a 254-amino-acid protein with a calculated molecular mass of 28,901 Da. Its deduced amino acid sequence was compared to protein sequences in the GenBank database and shown to be similar to several prokaryotic regulatory proteins. A helix-turn-helix motif which was homologous to the pyruvate dehydrogenase complex repressor (16) and L-lactate dehydrogenase regulatory protein (13) of *E. coli* and to the *hutC* repressor protein of *Klebsiella aerogenes* (31), among others, was found.

The high similarity of the *glcC* deduced amino acid sequence to other regulatory proteins and the fact that its inactivation simultaneously affected both the glycolate oxidase and the malate synthase G activities indicated that this gene encodes the *glc* regulatory protein. This protein would act as an activator since mutations in *glcC* did not render the expression of the *glc* structural genes constitutive but rather abolished it.

**Features and predicted properties of *glcD* and *glcE*.** The gene *glcD* (positions 1117 to 2616 of Fig. 2) is separated from *glcC* by a 251-bp region. A good ribosome-binding site (GAAG) was identified 11 bp upstream of the ATG codon. This gene encodes a 499-amino-acid polypeptide with a calculated molecular mass of 53,749 Da. The comparison with protein sequences in the GenBank database showed a significant similarity to the sequences of D-lactate dehydrogenases from *Kluyveromyces lactis* (20) and from *Saccharomyces cerevisiae* (19) and also to that from *E. coli*, although the last showed less similarity. Since it has been reported that the glycolate oxidase

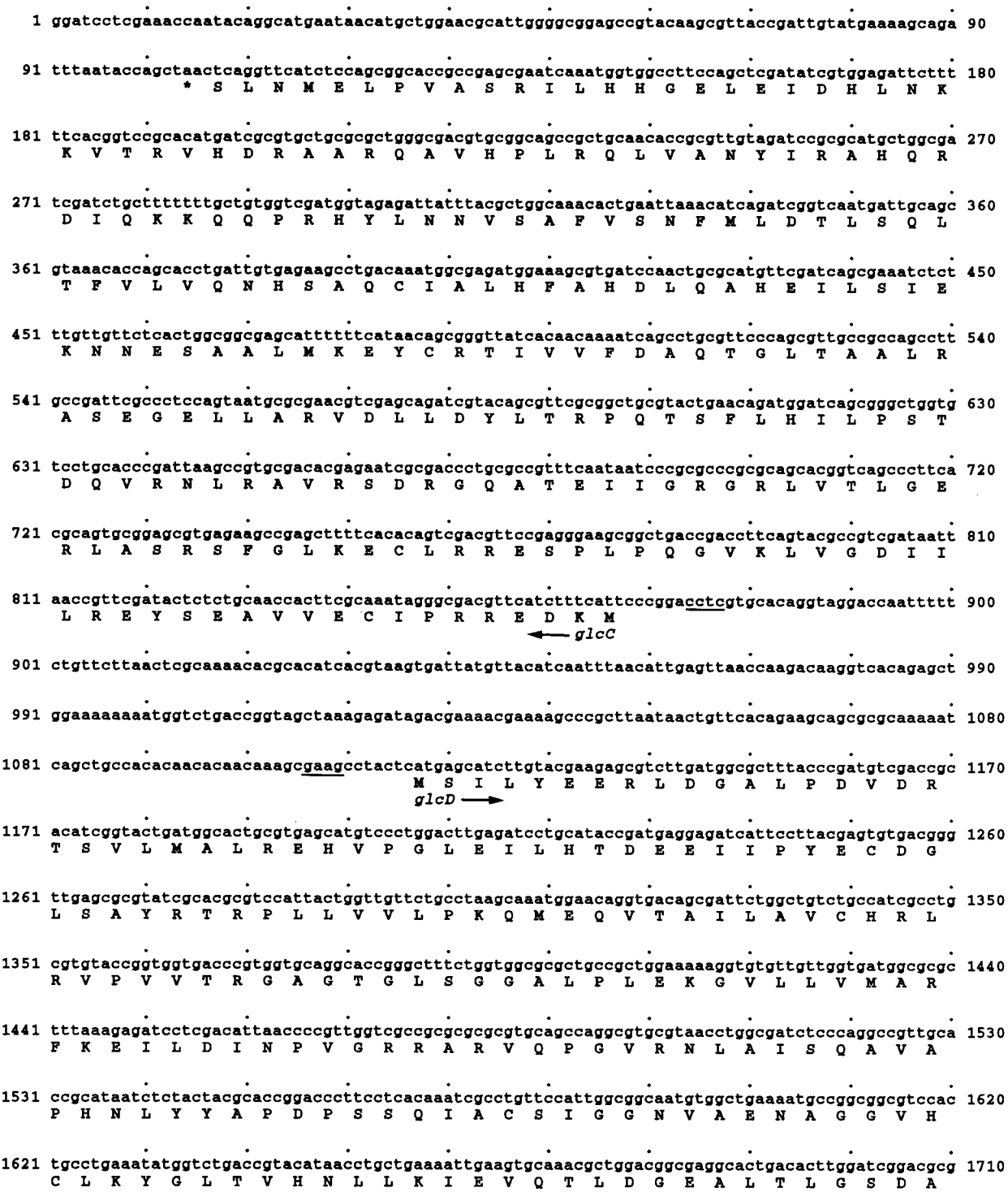



FIG. 2. Nucleotide sequence of the 5.3-kb genomic fragment upstream of *glcB* in the *glc* locus. Only one strand is shown. The genes *glcC* (transcribed clockwise), and *glcD*, *glcE*, *glcF*, and *glcG* (transcribed counterclockwise) have been translated in the one-letter amino acid code; amino acid symbols are written below the first nucleotide of the corresponding codon. Stop codons are indicated by asterisks. The start codon of *glcB* (accession number X74547) is also indicated. Putative ribosome binding sites are underlined. The GenBank accession number of this sequence is L43490.

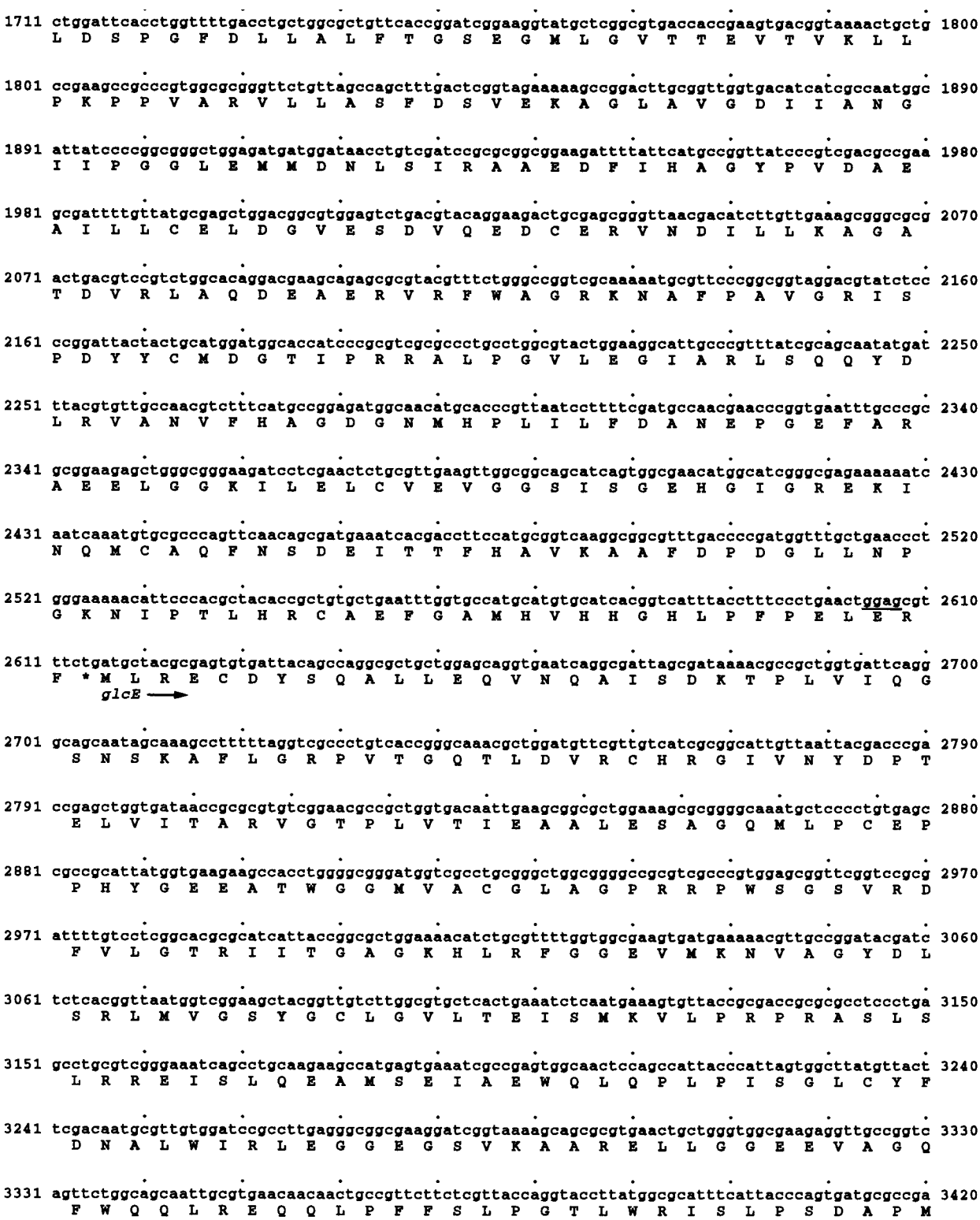



FIG. 2—*Continued.*

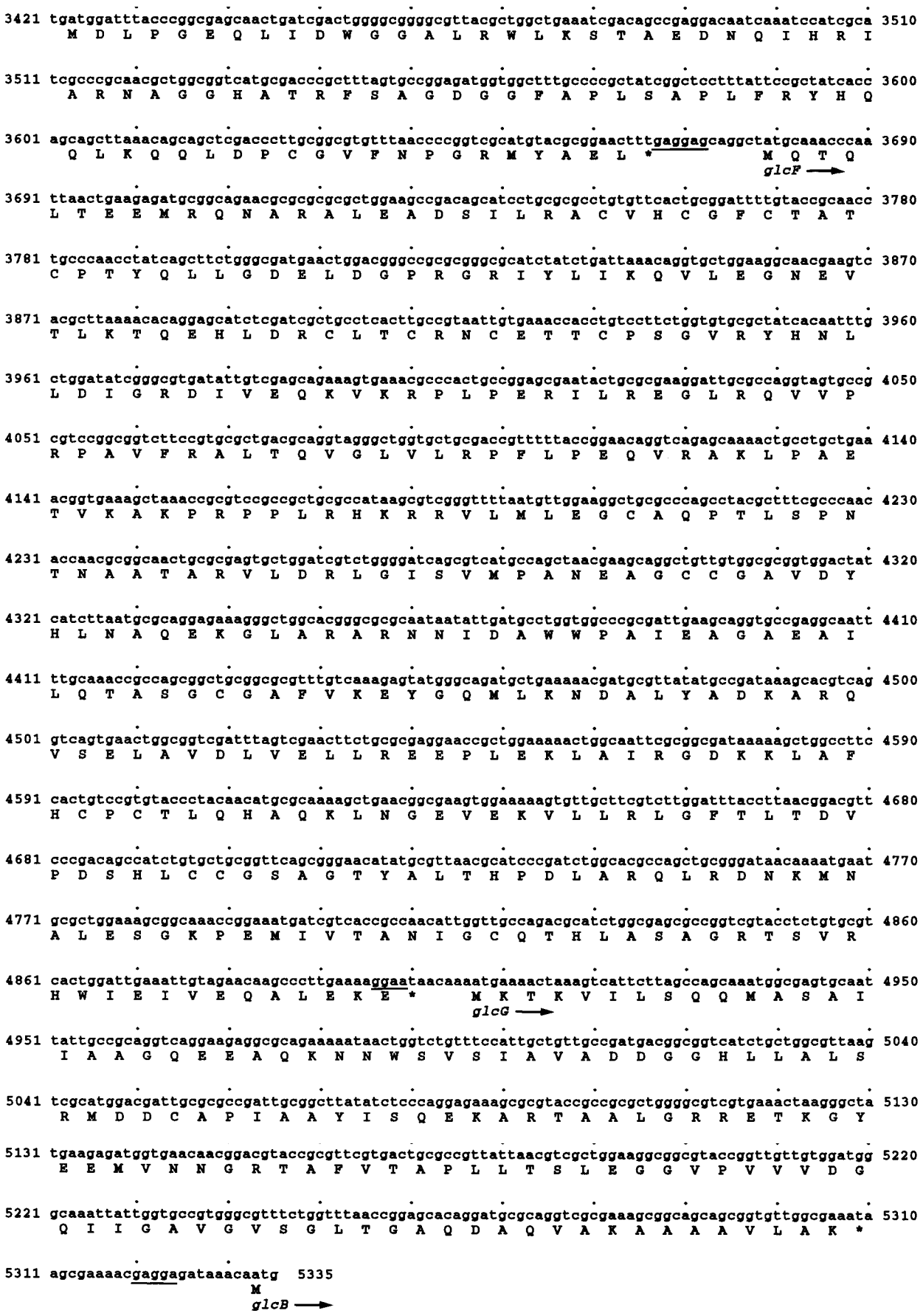



FIG. 2—*Continued.*

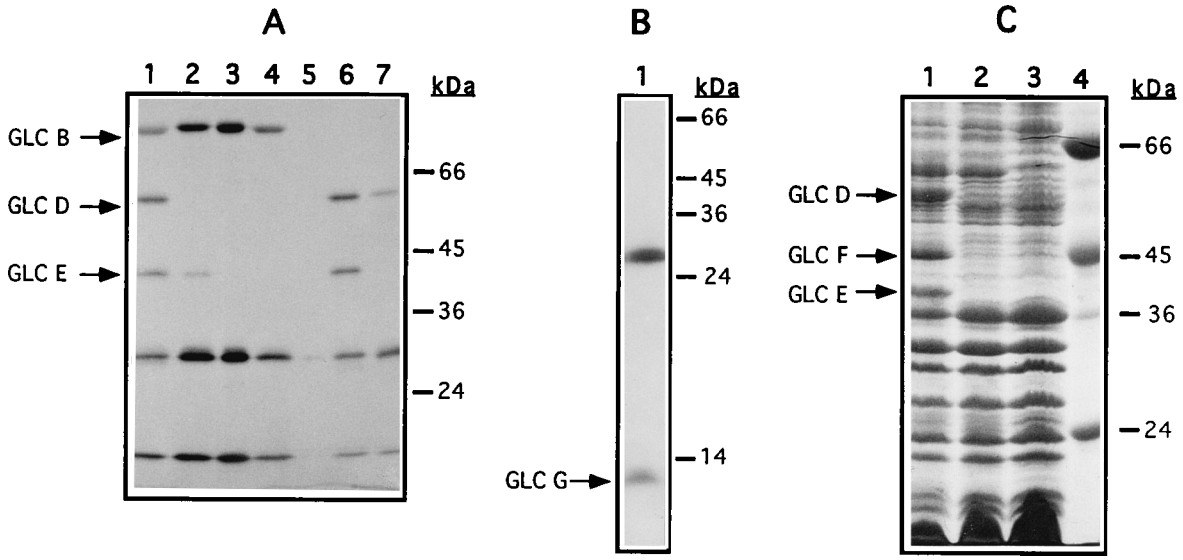



FIG. 3. (A) Autoradiograph of the gene products encoded by recombinant plasmids containing different parts of the *glc* locus and expressed in the maxicell system using strain JA120 as a host. Proteins were separated by SDS–10% PAGE. Lane 1, plasmid pTP21; lane 2, plasmid pTP22; lane 3, plasmid pTP23; lane 4, plasmid pIM4; lane 5, BlueScript vector; lane 6, plasmid pTP25; lane 7, plasmid pTP26. (B) Autoradiograph of the gene products encoded by plasmid pTP28 separated by SDS–13.5% PAGE. (C) SDS-PAGE of crude membrane proteins. Samples of 150 μg of membrane proteins obtained as indicated in the text were applied to a 10% polyacrylamide gel, electrophoresed, and stained with Coomassie blue. Lane 1, strain JA151 transformed with plasmid pTP21; lane 2, strain JA151 transformed with vector Bluescript; lane 3, wild-type strain MC4100. Gene products are indicated on the left of each panel. Molecular mass markers indicated on the right of the three panels were bovine serum albumin (66 kDa), ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), trypsinogen (24 kDa), and lactalbumin (14 kDa).

of *E. coli* can use D-lactate as a substrate (21), this gene could encode a protein involved in the oxidation of glycolate.

The initiation codon of *glcE* overlaps with the termination codon of *glcD*. Its putative Shine-Dalgarno sequence (GGAG) has been located 12 bp upstream of the ATG codon. The *glcE* gene encodes a 350-amino-acid protein with a calculated molecular mass of 38,352 Da. When compared with the protein sequences in the GenBank database, the highest similarity was again found with the D-lactate dehydrogenase from *S. cerevisiae*. On the basis of this similarity, this protein could be assigned to another subunit of glycolate oxidase.

Upon examination of the primary structures of *glcD* and *glcE* gene products, possible flavin-binding domains were easily detected. In the case of *glcD* we have identified a partially conserved consensus sequence like that described by Branden and Tooze (5). This consensus sequence contained the motif Gly-X-Gly-X-X-Gly separated from a conserved glutamate by 17 residues and included several invariant hydrophobic positions (Fig. 5A). In addition, a sequence showing significant homology to the active site of several oxidoreductases was found for both the *glcD* and *glcE* proteins. This sequence contained a histidine residue (His-140 for the *glcD* protein and His-91 for the *glcE* protein) separated from a cysteine residue by 13 and 12 amino acids, respectively, the same distance as in anaerobic glycerol-3-phosphate dehydrogenase (9), fumarate reductase (8), or succinate dehydrogenase (39).

It has been described for some flavin adenine dinucleotide or flavin mononucleotide enzymes like anaerobic glycerol-3-phosphate dehydrogenase that the enzyme activity is higher when measured in the presence of these nucleotides. To test the possible activation of the glycolate oxidase activity of our preparations by these cofactors, we used concentrations of flavin adenine dinucleotide ranging from 2 to 50 μM and flavin mononucleotide from 0.2 to 5 mM in the assay mixture. The

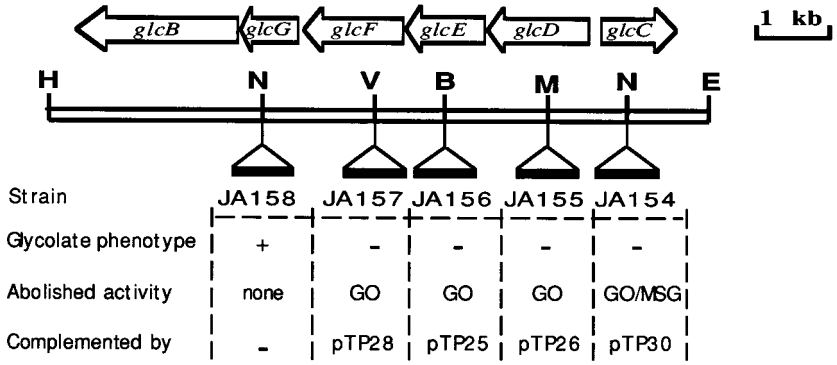



FIG. 4. Physical map of the *glc* locus insertional mutagenesis by CAT19 cassette. The open bar represents the genomic DNA of *E. coli* wild-type strain MC4100, and the arrows represent the extension and direction of transcribed genes. Insertions of the CAT cassette are indicated by black bars below the map. The restriction sites used for each insertion are indicated as follows: B, *Bam*HI; M, *Mlu*I; N, *Nru*I; and V, *Eco*RV. Relevant characterization of each insertion mutant strain is presented. GO, glycolate oxidase; MSG, malate synthase G.

A

| | | | |
|---|---|---|---|
| 84 | TRGA**G**T**G**LS**G**GALPLEKGVLLVMARFK**E** | 111 | **GLC D** |
| 5 | TVIM**G**G**G**LA**G**LLCGLQLQKHGLRCAIVT | 32 | **GLP B** |
| 11 | VIII**G**G**G**AT**G**AGIARDCALRGLRVILV**E** | 38 | **GLP A** |
| 9 | AVVI**G**A**G**GA**G**IARLAQISQSGQTCALLS | 36 | **SDH A** |
| 8 | VVVL**G**A**G**PA**G**YSAAFRCADLGLETVIV**E** | 35 | **LPD** |

B

| | | | |
|---|---|---|---|
| 22 | L**R**AC**V**HC**G**FC**T**AT**CP**TYQLL**G** | 42 | **GLC F** |
| 72 | L**D**RC**L**TC**R**NC**E**TT**CP**SGVRYH | 92 | **GLC F** |
| 6 | F**E**NC**I**KC**T**VC**T**TA**CP**VSRVN**P** | 26 | **GLP C** |
| 53 | L**K**YC**I**NC**K**RC**E**VA**CP**SDVKI**G** | 73 | **GLC C** |
| 145 | L**Y**EC**I**LC**A**CC**S**TS**CP**SFWWN**P** | 165 | **SDH B** |
| 145 | F**S**GC**I**NC**G**LC**Y**AA**CP**QFGLN**P** | 165 | **FRD B** |
| 5 | T**D**EC**I**SC**G**AC**A**AE**CP**VEAIH**E** | 25 | **FDX (CT)** |
| 34 | A**D**TC**I**DC**G**AC**E**AV**CP**TGAVK**A** | 54 | **FDX (CT)** |
| 5 | E**E**TC**I**SC**G**AC**A**AE**CP**VNAIE**Q** | 25 | **FDX (RR)** |
| 34 | A**D**TC**I**DC**G**NC**A**NV**CP**VGAPV**A** | 54 | **FDX (RR)** |

+--C+-C--C---CP-----•

FIG. 5. (A) Consensus sequence of the flavin cofactor binding sites from the following *E. coli* enzymes: GlcD, glycolate oxidase; GlpA and GlpB, anaerobic glycerol-3-phosphate dehydrogenase (9); SdhA, succinate dehydrogenase (39); and Lpd, lipoamide dehydrogenase (33). The invariant positions for glycines and the conserved glutamate are indicated in boldface. The dots mark the rather conserved hydrophobic positions which complete the consensus for the flavin binding site as indicated by Branden and Tooze (5). The numbering corresponds to the amino acid positions in each protein. (B) Amino acid sequences of iron-sulfur centers from the following *E. coli* proteins: GlcF, glycolate oxidase; GlpC, anaerobic glycerol-3-phosphate dehydrogenase (9); SdhB, succinate dehydrogenase (12); and FrdB, fumarate reductase (10). Alignment also includes ferredoxins from *Clostridium thermosaccharolyticum* (35) and *Rhodospirillum rubrum* (23) labeled as Fdx (CT) and Fdx (RR), respectively. The different cysteine clusters are aligned, and the highly conserved residues are shown in boldface. Numbering indicates the position of amino acids in each protein. The consensus is presented at the bottom, where cysteines are indicated by C, proline is indicated by P, conserved hydrophobic amino acids are indicated by plus signs, and an aliphatic amino acid is indicated by a dot.

lack of sensitivity of glycolate oxidase activity in our conditions is noteworthy.

Furthermore, the *glcD*- and *glcE*-encoded subunits of the glycolate oxidase were found to be absolutely dependent on the *glcF* subunit for activity. Preparations obtained from plasmid pTP25 expressing only *glcDE* did not transfer electrons directly to phenazine methosulfate, an acceptor used without the iron-sulfur intermediate by other flavoenzymes such as *glpAB*-encoded subunits of anaerobic glycerol-3-phosphate dehydrogenase (9). This observation reinforced the idea of a different mechanism for glycolate oxidase.

**Features and predicted properties of *glcF*.** Three possible start codons at positions 3532, 3556, and 3679 (Fig. 2) were identified for the gene *glcF*. The first two ATG codons overlapped with the *glcE* gene and had no well-conserved Shine-Dalgarno sequences upstream. The open reading frame, starting at position 3679 and ending at position 4902, displayed a putative ribosome-binding site (GAGGAG) 12 bp upstream of the corresponding ATG codon. The use of this translation initiation signal predicts the expression of a protein of 407 amino acid residues with a molecular mass of 45,083 Da.

The *glcF* subunit, required for the function of the glycolate oxidase enzyme, was a cysteine-rich protein containing 17 residues, most of them organized in two clusters typical of iron-sulfur proteins. Upon alignment of these clusters with those of other respiratory enzymes it was clear that besides the characteristic spatial arrangement of the cysteine residues (Cys-X-X-Cys-X-X-Cys-X-X-X-Cys), additional positions were conserved in these sequences. For instance, a hydrophobic residue located two positions before the first cysteine, which was followed by another hydrophobic residue. The last cysteine residue was followed by a conserved proline and, six residues later, by an aliphatic amino acid (Fig. 5B).

Among the proteins giving high-scoring segment pairs with *glcF* gene product, as indicated by the BLAST program, there were the GlpC subunit of anaerobic glycerol-3-phosphate dehydrogenase from *E. coli* (9) and the iron-sulfur subunit of succinate dehydrogenase from different species. These enzymes are multimeric proteins with similar catalytic mechanisms, oxidizing molecules and transferring electrons through iron-sulfur subunits.

The *glcF* low expression, probably giving a lower cellular concentration, may indicate that several *glcD-glcE* products are associated with one *glcF* product. Alternatively, the subunits encoded by *glcD-glcE* might act catalytically on the *glcF* protein, transferring their electrons to this terminal subunit of the enzyme (9).

**Features and predicted properties of *glcG*.** The initiation codon of *glcG* is located 5 bp downstream of the *glcF* termination codon and is preceded by a Shine-Dalgarno sequence (GGAA) 11 bp upstream. This gene encodes a 134-amino-acid protein with a calculated molecular mass of 13,737 Da. There was no significant similarity to any protein or group of proteins, although some similarity to a 142-amino-acid protein of unknown function was observed (ORFY next to the *dha* regulon from *Citrobacter freundii*, accession number CFU09771) (11).

At present we have been unable to assign any function to the *glcG* gene product, as its inactivation by the CAT cassette insertion did not affect glycolate utilization under the conditions tested. Furthermore, strain JA151 transformed with a plasmid expressing a truncated GlcG protein displayed normal levels of glycolate oxidase activity (not shown). This plasmid was obtained by deletion from the *Cla*I restriction site of plasmid pTP24 (Fig. 1) and the extent of the deletion, which removed up to amino acid 30 of the C-terminal end of GlcG, was ascertained by sequencing.

**Transcription.** Total RNA was prepared as indicated above from cells of strain MC4100 grown either on D-xylose as a noninducing carbon source or on D-xylose plus glyoxylate, D-xylose plus glycolate, or glycolate alone as an inducing carbon source. Northern blot hybridizations of these RNA preparations were performed with a probe of each of the *glc* genes.

For the *glcD* structural gene, a transcript of 2.6 kb was detected only under inducing conditions in the presence of glycolate or glyoxylate, indicating the specificity of the transcription (Fig. 6). Similar results were obtained for the other structural genes, showing transcripts ranging from 2.1 to 2.6 kb (not shown). Thus, none of the bands detected in the Northern experiment was in accordance with the expected size of the possible transcripts. No direct evidence for a polycistronic operon was obtained from Northern blot experiments or polarity effects. Alternatively, the discrete size of the transcripts observed could be indicative of transcript degradation.

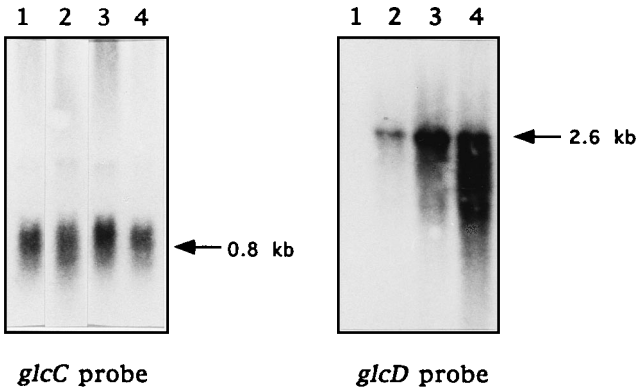



FIG. 6. Northern blots of total RNA from strain MC4100 grown on D-xylose (lanes 1), D-xylose plus glyoxylate (lanes 2), D-xylose plus glycolate (lanes 3), and glycolate (lanes 4). Hybridization was performed with a 500-bp *glcC* specific probe (*Nru*I-*Eco*RV internal fragment) (left) or a 810-bp *glcD* specific probe (*Sal*I internal fragment) (right).

A constitutive transcript of 0.8 kb was detected with the *glcC* probe, well in agreement with the size of this regulatory gene's expected transcript (Fig. 6). Finally, hybridization with probes containing sequences beyond the *glcB* and *glcC* termini displayed no bands in Northern blots of RNAs from either induced or noninduced cells. This seems to indicate that no *glc* genes are found over these limits.

## ACKNOWLEDGMENTS

This work was supported by grant PB-94-0829 from the Dirección General de Investigación Científica y Técnica, Madrid, Spain.

We thank Robin Rycroft for editorial assistance.

## REFERENCES

1. **Badía, J., L. Baldomà, J. Aguilar, and A. Boronat.** 1989. Identification of the *rhaA*, *rhaB* and *rhaD* gene products from *Escherichia coli* K12. FEMS Microbiol Lett. **65:**253–258.
2. **Belasco, J. G., T. Beatty, C. W. Adams, A. von Gabain, and S. N. Cohen.** 1985. Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic *rxcA* transcript. Cell **40:**171–181.
3. **Boronat, A., and J. Aguilar.** 1979. Rhamnose-induced propanediol oxidoreductase in *Escherichia coli*: purification, properties, and comparison with the fucose-induced enzyme. J. Bacteriol. **140:**320–326.
4. **Boronat, A., and J. Aguilar.** 1981. Metabolism of of L-fucose and L-rhamnose in *Escherichia coli*: differences in induction of propanediol oxidoreductase. J. Bacteriol. **147:**181–185.
5. **Branden, C., and J. Tooze.** 1991. Introduction to protein structure, p. 148–150. Garland Publishing, New York.
6. **Casadaban, M. J.** 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J. Mol. Biol. **104:**541–555.
7. **Chang, Y. Y., A. Y. Wang, and J. E. Cronan, Jr.** 1993. Molecular cloning, DNA sequencing, and biochemical analyses of *Escherichia coli* glyoxylate carboligase. J. Biol. Chem. **268:**3911–3919.
8. **Cole, S. T.** 1982. Nucleotide sequence coding for the flavoprotein subunit of the fumarate reductase of *Escherichia coli*. Eur. J. Biochem. **122:**479–484.
9. **Cole, S. T., K. Eiglmeir, S. Ahmed, N. Honore, L. Elmes, W. F. Anderson, and J. H. Weiner.** 1988. Nucleotide sequence and gene-polypeptide relationship of the *glpABC* operon encoding the anaerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. J. Bacteriol. **170:**2448–2456.
10. **Cole, S. T., T. Grundtrom, B. Jaurin, J. J. Robinson, and J. H. Weiner.** 1982. Location and nucleotide sequence of *frdB*, the gene encoding for the iron-sulphur protein subunit of the fumarate reductase of *Escherichia coli*. Eur. J. Biochem. **126:**211–216.
11. **Daniel, R., K. Stuertz, and G. Gottschalk.** 1995. Biochemical and molecular characterization of the oxidative branch of glycerol utilization by *Citrobacter freundii*. J. Bacteriol. **177:**4392–4401.
12. **Darlison, M. G., and J. R. Guest.** 1984. Nucleotide sequence encoding the iron-sulphur protein subunit of the succinate dehydrogenase of *Escherichia coli*. Biochem. J. **223:**507–517.
13. **Dong, J. M., J. S. Taylor, D. J. Latour, S. Iuchi, and E. C. C. Lin.** 1993. Three overlapping *lct* genes involved in L-lactate utilization by *Escherichia coli*. J. Bacteriol. **175:**6671–6678.
14. **Fuqua, W. C.** 1992. An improved chloramphenicol resistance gene cassette for site-directed marker replacement mutagenesis. BioTechniques **12:**223–225.
15. **Hansen, R. W., and J. A. Hayashi.** 1962. Glycolate metabolism in *Escherichia coli*. J. Bacteriol. **83:**679–687.
16. **Haydon, D. J., M. A. Quail, and J. R. Guest.** 1993. A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex in *Escherichia coli* is located within the *pdhR* gene. FEBS Lett. **336:**43–47.
17. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114:**193–197.
18. **Kornberg, H. L., and J. R. Sadler.** 1961. The metabolism of $C_2$ compounds in microorganisms. 8. A dicarboxylic acid cycle as a route for the oxidation of glycollate by *Escherichia coli*. Biochem. J. **81:**503–513.
19. **Lodi, T., and I. Ferrero.** 1993. Isolation of the DLD gene of *Saccharomyces cerevisiae* encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase. Mol. Gen. Genet. **238:**315–324.
20. **Lodi, T., D. O'Connor, P. Goffrini, and I. Ferrero.** 1994. Carbon catabolite repression in *Kluyveromyces lactis*: isolation and characterization of the KIDLD gene encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase. Mol. Gen. Genet. **244:**622–629.
21. **Lord, J. M.** 1972. Glycolate oxidoreductase in *Escherichia coli*. Biochim. Biophys. Acta **267:**227–237.
22. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193:**265–273.
23. **Matsubara, H., K. Inove, T. Hase, H. Hiura, T. Kakuno, and J. Yamashita.** 1983. Structure of the extracellular ferredoxin from *Rhodospirillum rubrum*: close similarity to clostridial ferredoxins. J. Biochem. **93:**1385–1390.
24. **Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Molina, I., M. T. Pellicer, J. Badía, J. Aguilar, and L. Baldomà.** 1994. Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme. Eur. J. Biochem. **224:**541–548.
26. **Moralejo, P., S. M. Egan, E. Hidalgo, and J. Aguilar.** 1993. Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. J. Bacteriol. **175:**5585–5594.
27. **Ornston, L. N., and M. K. Ornston.** 1969. Regulation of glyoxylate metabolism in *Escherichia coli* K-12. J. Bacteriol. **98:**1098–1108.
28. **Sallal, A.-K. J., and N. A. Nimer.** 1989. The intracellular localization of glycolate oxidoreductase in *Escherichia coli*. FEBS Lett. **258:**277–280.
29. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **174:**5463–5467.
31. **Schwacha, A., and R. A. Bender.** 1990. Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Klebsiella aerogenes*. J. Bacteriol. **172:**5477–5481.
32. **Singer, M., T. A. Baker, G. Schnitzler, S. M. Deischel, M. Goel, W. Dove, K. J. Jaacks, A. D. Grossman, J. W. Erickson, and C. A. Gross.** 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol. Rev. **53:**1–24.
33. **Stephens, P. E., H. M. Lewis, M. G. Darlison, and J. R. Guest.** 1983. Nucleotide sequence of the lipoamide dehydrogenase gene of *Escherichia coli* K-12. Eur. J. Biochem. **135:**519–527.
34. **Stoker, N. G., J. M. Pratt, and H. B. Holland.** 1984. "In vivo" gene expression systems in prokaryotes, p. 154–172. *In* B. D. Hames and S. H. Higgins (ed.), Transcription and translation: a practical approach. IRL Press Ltd., Oxford.
35. **Tanaka, M., M. Haniu, K. T. Yasunobu, R. H. Himes, and J. M. Akagi.** 1973. The primary structure of the *Clostridium thermosaccharolyticum* ferredoxin, a heat stable ferredoxin. J. Biol. Chem. **248:**2215–2217.
36. **Vanderwinkel, E., and M. De Vlieghere.** 1968. Physiologie et génétique de l'isocitritase et des malate synthases chez *Escherichia coli*. Eur. J. Biochem. **5:**81–90.
37. **Wackernagel, W.** 1973. Genetic transformation in *E. coli*: the inhibitory role of the *recBC* DNase. Biochem. Biophys. Res. Commun. **51:**306–311.
38. **Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker.** 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. **161:**1219–1221.
39. **Woods, D., M. G. Darlison, R. J. Wilde, and J. R. Guest.** 1984. Nucleotide sequence encoding the flavoprotein and hydrophobic subunits of the succinate dehydrogenase of *Escherichia coli*. Biochem. J. **222:**519–534.