JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 101–107
0021-9193/91/010101-07$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 173, No. 1

# Nucleotide Sequence of the *glpD* Gene Encoding Aerobic *sn*-Glycerol 3-Phosphate Dehydrogenase of *Escherichia coli* K-12

DENISE AUSTIN AND TIMOTHY J. LARSON*

*Department of Biochemistry and Nutrition, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*

Received 4 June 1990/Accepted 11 October 1990

**Aerobic *sn*-glycerol 3-phosphate dehydrogenase, encoded by the *glpD* gene of *Escherichia coli*, is a cytoplasmic membrane-associated respiratory enzyme. The nucleotide sequence of *glpD* was determined. An open reading frame of 501 codons was preceded by a consensus Shine-Dalgarno sequence. The proposed translational start and reading frame of *glpD* were confirmed by determining the nucleotide sequence across the fusion joint of a *glpD-lacZ* translational fusion. The predicted molecular weight, 56,750, corresponds well with the reported value of 58,000 for purified *sn*-glycerol 3-phosphate dehydrogenase. The flavin-binding domain, located at the amino terminus, was identified by comparison with the amino acid sequences of other flavoproteins from *E. coli*. Repetitive extragenic palindromic sequences were identified downstream of the *glpD* coding region. The site for transcription termination was located between 87 and 216 bp downstream of the translation stop codon.**

*Escherichia coli* is capable of utilizing glycerol, *sn*-glycerol 3-phosphate (glycerol-P), and glycerophosphodiesters as carbon sources and as precursors for phospholipid biosynthesis. The dissimilation of glycerol-P and its precursors is catalyzed by proteins encoded by the genes of the *glp* regulon (17, 18). This metabolic system acts essentially as a salvage pathway for utilization of glycerol and glycerol-P derived from the breakdown of phospholipids and triacylglycerol. The regulon is induced by glycerol-P. The *glp* genes, under negative genetic regulation by the *glpR*-encoded repressor protein (16, 17), have been mapped to three separate regions on the *E. coli* chromosome and are arranged in at least five different operons (17).

The *glpQ* gene encodes a periplasmic glycerophosphodiesterase that catalyzes hydrolysis of glycerophosphodiesters to glycerol-P and an alcohol (17). Glycerol-P is actively transported into the cytoplasm by the *glpT*-encoded permease (17). The *glpTQ* operon is located at min 49 on the linkage map of *E. coli*. It is adjacent to, and is transcribed divergently from, the *glpACB* (also called *glpABC* [7]) operon encoding the three subunits of anaerobic glycerol-P dehydrogenase (8, 17, 18). The *glpFK* operon (19, 35) is located near min 88. The *glpF* gene encodes a cytoplasmic membrane protein responsible for facilitating the diffusion of glycerol across the membrane (17). The *glpK* gene codes for glycerol kinase (21). Phosphorylation of glycerol traps glycerol-P in the cytoplasm.

The *glpD* gene encodes aerobic glycerol-P dehydrogenase and maps near min 75 (17). The *glpD* gene is transcribed divergently from the adjacent *glpE*, *glpG*, and *glpR* genes (6, 31–33). Aerobic glycerol-P dehydrogenase is expressed maximally under aerobic growth conditions (17). Anaerobic repression of *glpD* expression is mediated by the two-component regulatory system encoded by *arcA* and *arcB* (13, 14). Glycerol-P dehydrogenase is associated with the cytoplasmic membrane and is a primary dehydrogenase in the respiratory chain. It catalyzes the oxidation of glycerol-P to dihydroxyacetone phosphate, with concurrent reduction of flavin adenine dinucleotide (FAD) to FADH. FADH passes electrons on to ubiquinone and ultimately to oxygen or nitrate (17, 24). The purified enzyme exists in the native state as a dimer of identical 58-kDa subunits (30).

Previous work from this laboratory has established the physical location and direction of transcription of the *glpD* gene on cloned DNA fragments (31–33). Also, the promoter-operator region has been characterized and a portion of the *glpD* gene encoding the amino-terminal end of aerobic glycerol-P dehydrogenase has been sequenced (38). The objective of this study was to obtain the remainder of the nucleotide sequence of the *glpD* gene. While this work was in progress, a nucleotide sequence for the *glpD* gene was published (5). We found several discrepancies between our sequence and the published sequence (5) which significantly change the predicted amino acid sequence of the dehydrogenase. Also, we present data which establish the translational start site and reading frame, and the approximate location for transcription termination, for the *glpD* gene. Finally, amino acid sequence similarity between GlpD, GlpA, and triosephosphate isomerase (TPI) was discovered which allowed prediction of the glycerol-P-binding domain of the dehydrogenases.

## MATERIALS AND METHODS

**Bacterial strains, phages, and plasmids.** *E. coli* DH5αF′ [*recA1 endA1 hsdR17* Δ(*lacZYA-argF*)*U169* φ80*d lacZ*Δ*M15 supE44 thi-1 gyrA96 relA1*] was used as the host strain for preparation of plasmid and phage M13 DNA. DNA fragments of sizes amenable for sequencing were cloned into pBluescript KS+ (Stratagene), M13mp18, or M13mp19 (37). The plasmids used or constructed are described in Table 1. The site of transcription termination was estimated by cloning of fragments from pSH60 into pCB267 (29).

* Corresponding author.

TABLE 1. Plasmids constructed or used in this study

| Plasmid[a] | Insert | Source | Position in sequence |
|---|---|---|---|
| pSY10[b] | *Rsa*I-*Bgl*II | pSH79 | −120–54 |
| pSY12[b] | *Bgl*II-*Eco*RI | pSH79 | 55–220 |
| pDA1101 | *Eco*RI-*Hin*dIII | pSH60 | 221–2645 |
| pDA1104 | *Eco*RI-*Eco*RV | pDA1101 | 221–606 |
| pDA1103 | *Eco*RV-*Eco*RV | pDA1101 | 607–966 |
| pDA1102 | *Eco*RV-*Hin*dIII | pDA1101 | 967–2645 |
| pDA701 | *Rsa*I-*Rsa*I | pDA1101 | 498–865 |
| pDA20 | *Bam*HI-*Stu*I | pSH60 | −120⁺–1634[c] |
| pDA21 | *Bam*HI-*Ase*I | pSH60 | −120⁺–1840 |

[a] The vector in each case, with the exceptions of pDA20 and pDA21 (Fig. 2), was pBluescript KS+ (Stratagene).

[b] Sequence determined by Ye and Larson (38).

[c] The *Bam*HI site of pSH60 is that present in a Tn*1000* insertion in the adjacent, divergently transcribed *glpE* gene (32). −120⁺ indicates that the exact point of insertion of Tn*1000* in pSH60 has not been determined.

**Media, enzymes, and reagents.** LB culture medium (1% tryptone, 0.5% yeast extract [Difco Laboratories], 1% NaCl) was used for growth of *E. coli*. Strains transformed with plasmid DNA were grown in LB fortified with glucose (0.4%) plus 200 μg of ampicillin (Sigma) per ml. Screening for the expression of the alpha fragment of β-galactosidase was carried out on LB agar plates containing 40 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; United States Biochemical Corp.) per ml. Screening for the expression of alkaline phosphatase was carried out on LB agar plates containing 0.4% glycerol and 40 μg of 5-bromo-4-chloro-3-indolylphosphate per ml. Restriction endonucleases and T4 DNA ligase were purchased from New England BioLabs, United States Biochemical Corp., and Boehringer Mannheim.

**Isolation of DNA.** Plasmid DNA was isolated by an alkaline detergent lysis method (27). Cells transfected with M13 were grown in H broth (1% tryptone, 0.8% NaCl), and single-stranded DNA was isolated as described previously (10).

**Sequencing of DNA.** A DNA sequencing kit containing Sequenase, a modified T4 DNA polymerase, was purchased from United States Biochemical Corp. [$\alpha$-$^{35}$S]dATP was purchased from DuPont/New England Nuclear Corp. Both double-stranded plasmid DNA and single-stranded M13 DNA were sequenced by the dideoxy-chain termination method (28). Synthetic oligonucleotides were used as primers to sequence DNA downstream of *glpD*. Their sequences and relative positions were GAGAAAAATAGCATTAA (1826 to 1842), AATTTTTATGCTGCGCT (1867 to 1850), TGCCCCTTCTGGACACT (2085 to 2101), TCACCAATAAATAGAAC (2160 to 2144), TGAAGAACTGCGTCGTC (2511 to 2495), and ACCTGATGCAGCTCCTC (2443 to 2459). The primers were purified, after addition of sodium acetate (pH 4.6) to 0.3 M, by phenol extraction followed by a phenol-chloroform (50:50) extraction and ethanol precipitation.

**Sequence analysis programs.** PC/GENE microcomputer software from IntelliGenetics and the Pustell sequence analysis programs from IBI were used for analysis of protein and nucleic acid sequences.

**Nucleotide sequence accession number.** The sequence data reported here have been submitted to GenBank and assigned accession number M55989.

## RESULTS AND DISCUSSION

**Nucleotide sequence of the *glpD* gene.** Previous studies (32, 38) indicated that the *glpD* operator consists of tandemly repeated repressor-binding sites located just upstream of the *Bgl*II site at nucleotide 55 (Fig. 1 and 2). It has been shown that the *glpD* gene is transcribed from the *Bgl*II restriction site (32) toward the *Eco*RI site (nucleotide 220; Fig. 1 and 2), and the transcription start (+1) site has been identified (38). To obtain the remaining nucleotide sequence of *glpD*, the 2,430-bp *Eco*RI-*Hin*dIII fragment from pSH60 (32) was ligated to pBluescript KS+, creating pDA1101. Clones appropriate for sequencing of *glpD* were generated (Table 1 and Fig. 1). Sequence data for one strand were confirmed by generating sequence data for the entire complementary strand. The sequence across all restriction sites, used for creation of subclones, was obtained by sequencing overlapping clones (Fig. 1).

An open reading frame of 501 codons beginning with an ATG start codon was identified (Fig. 2). The translated sequence of *glpD* encodes a protein of molecular weight 56,750. The previously published nucleotide sequence of *glpD*

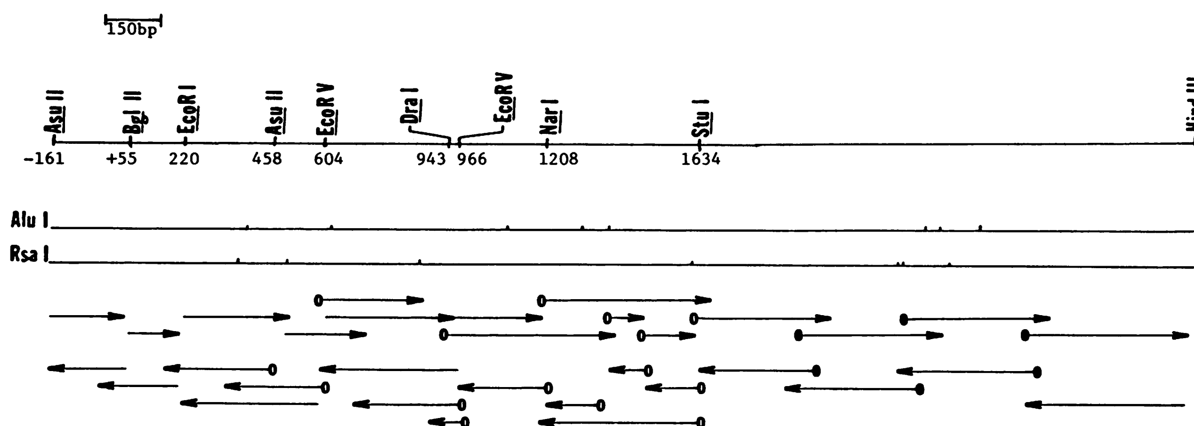


FIG. 1. Strategy for sequencing of the *glpD* gene. The extent and direction of sequencing are indicated by arrows. Arrows with open circles represent sequence data obtained from M13 clones by using primers that hybridize to the vector. Arrows with closed circles represent sequence data obtained from M13 clones by using chemically synthesized oligonucleotide primers. The remaining arrows represent sequence data obtained from plasmid DNA. The nucleotide sequence at the 5′ end, from *Asu*II to *Eco*RI, was determined previously (38).

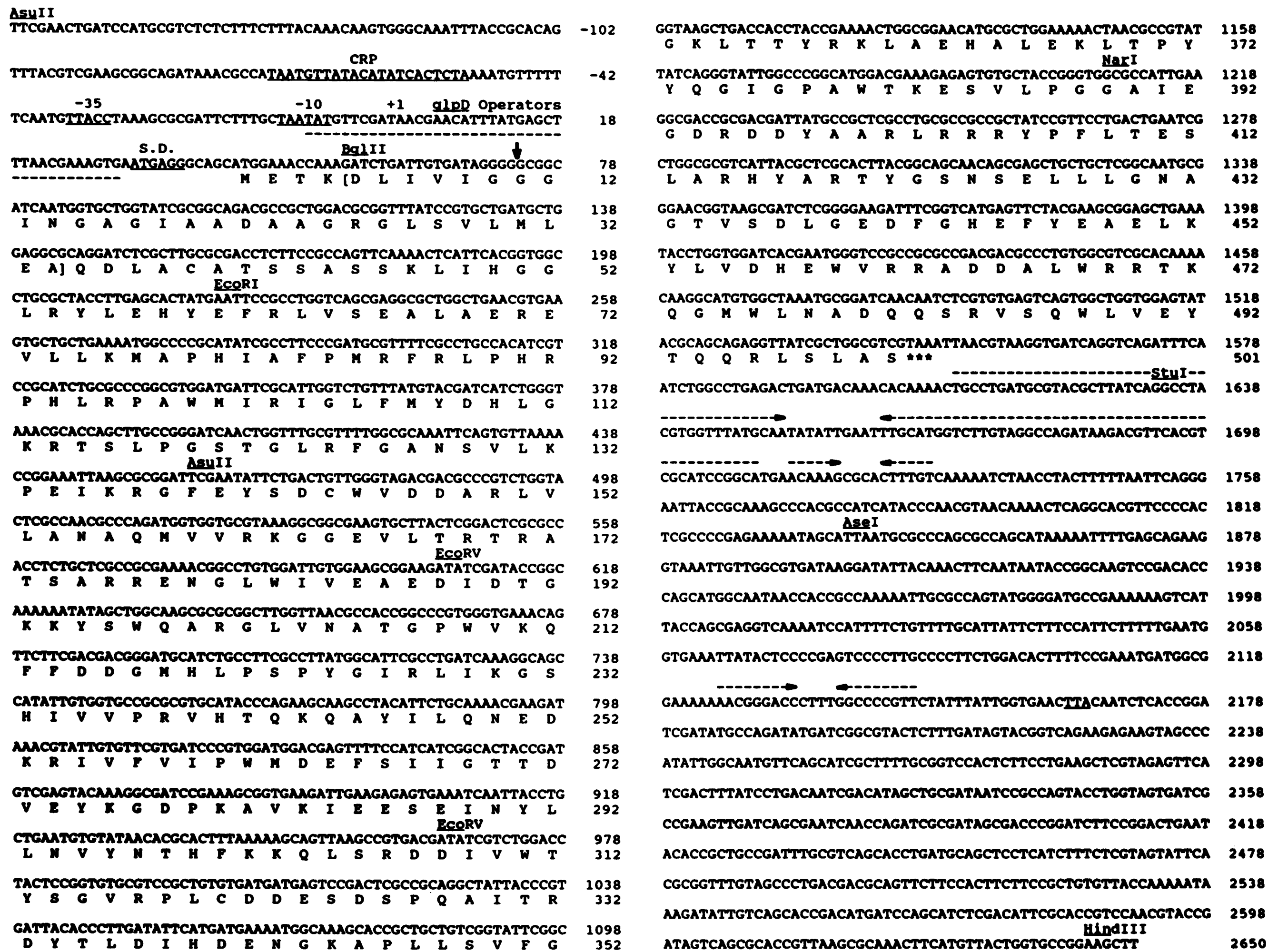

FIG. 2. Nucleotide sequence and corresponding amino acid sequence of the aerobic glycerol-P dehydrogenase gene (*glpD*). The control region for *glpD* (38) contains promoter and operator sequences (solid and dashed underlining, respectively) comparable to those of other *glp* genes of *E. coli*. The transcription start site is designated +1. The flavin-binding domain is bracketed. The position of the fusion joint in the *glpD-lacZ* translational fusion present in pSH85 (32, 33) is indicated by the vertical arrow between nucleotides 73 and 74. The nucleotide sequence (from λ p*lac*Mu) in pSH85 following the fusion joint was TG AAG CGG CGC ACG AAA with the indicated reading frame (3). REP sequences identified in the 3'-flanking DNA (nucleotides 1611 through 1709) are indicated by the long dashed arrows over the sequence. A potential transcription terminator for *glpD* is indicated by the short dashed arrows (nucleotides 1713 to 1728). The reported rho-independent transcription terminator (dashed arrows) and translation termination codon (underlined) for the convergent *glgP* gene (4, 39) are centered at bases 2136 and 2164, respectively.

(5) reports an open reading frame of 504 codons encoding a protein of molecular weight 56,131. The published sequences of the *glpD* (5) and *glgP* (4, 39) genes were compared with the nucleotide sequence presented in this report. The *glgP* gene, encoding α-glucan phosphorylase, is located downstream of *glpD* and is transcribed convergently toward *glpD*. The termination codon for *glgP* is centered at 2164 (Fig. 2). Four areas of sequence discrepancy in the *glpD* coding region, one in the intergenic region, and one in the coding region for the *glgP* gene were located (Table 2). After careful examination of the autoradiograms, we have determined that for both strands of each region in question, the sequence presented here is in complete agreement with our data.

The proposed reading frame for *glpD* was confirmed by determination of the nucleotide sequence of a *glpD-lacZ* translational fusion. A universal primer complementary to the 5' end of *lacZ* was hybridized to the gene fusion in pSH85 (32, 33), and nucleotide sequence data were generated according to the Sequenase protocol. The results indicated that the proposed reading frame is correct. Furthermore, the location of the fusion joint very near the 5' end of the gene (Fig. 2, vertical arrow) proves that the proposed translational start site is correct. There were only three possible start codons between the 5' end of the *glpD* mRNA and the fusion joint (GTG, ATG, and GTG, located at nucleotides 28, 43, and 64, respectively; Fig. 2). There is no likely ribosome-binding site preceding the first GTG. The second GTG (at position 64), closest to the fusion joint, is located in the flavin-binding domain and therefore could not be the start codon for *glpD*. The ATG codon (position 43) is preceded by a good Shine-Dalgarno sequence and is in the correct reading frame. It is therefore likely to be the translational start site for the *glpD* mRNA.

The deduced amino acid composition was compared with

TABLE 2. Discrepancies between previously reported nucleotide sequences and the nucleotide sequence of *glpD* reported in this study

| Sequence | | Result of error |
|---|---|---|
| This study | Previously reported | |
| CCGCTGG (110) | CCGGTGG[b] (substitution) | Substitution |
| GCAGGCT (1026–1028) | GCACGT[b] (deletion and inversion) | Frameshift |
| GGCGGTA (1100) | GGCGGCTA[b] (insertion) | Frameshift |
| CGACGAC (1433) | CGCACGAC[b] (insertion) | Frameshift |
| AGTCCCC (2080) | AGTGCCC[c] (substitution) | None |
| ACCGCTG (2424–2425) | ACCCGTG[c] (inversion) | Substitution |

[a] Numbers in parentheses give the position of the discrepancy within the nucleotide sequence of *glpD* (Fig. 2).
[b] Sequence reported by Choi et al. (5).
[c] Sequence reported by Yu et al. (39).

the amino acid composition reported for the purified protein (30). There is very good agreement, with minor differences, between the predicted and chemically determined compositions (Table 3).

A search for possible secondary structures in the nucleotide sequence flanking the coding region for *glpD* revealed the presence of two palindromes that begin approximately 60 bp downstream of the termination codon (Fig. 2). These palindromes were identified as repetitive extragenic palindromic (REP) sequences and are tandemly repeated, inverted with respect to each other, a typical arrangement found for REP sequences (12). They are nearly identical (85 and 94%, respectively) to the following consensus for REP sequences (12): GCCKGATGGCGRCGY. . . . .RCGYCT TATCMGGCCTAC, where K is G or T, R is A or G, Y is C or T, and M is C or A.

Experimental data generated in vitro as well as in vivo indicate that REP sequences, present in mRNA, will orient themselves to form the most stable secondary structure possible, with $\Delta G$ values near $-51$ kcal (ca. $-213$ kJ)/mol (12). A potential secondary structure for the *glpD* mRNA in this region having a $\Delta G$ value of $-57$ kcal (ca. 238 kJ)/mol was predicted by using the RNA fold program.

There are at least 1,000 copies of REP sequences on the *E. coli* chromosome (12). In many cases, REP sequences are located between genes within polycistronic operons and mediate differential expression of the genes by stabilizing upstream mRNA, possibly by protecting against 3′-5′ exonuclease activity (12). The possibility of additional open reading frames downstream of *glpD* was suggested by the presence of REP sequences. However, analysis of this DNA revealed no open reading frames of any significance between the end of REP and the end of the convergent *glgP* gene (4, 39). Furthermore, the 3′ end of *glpD* was localized to within 100 bp of the end of the REP sequences, as discussed below.

It has been proposed that REP sequences may be the physiological sites of DNA gyrase action (36). Action of DNA gyrase at the REP sequences between the convergent *glpD* and *glgP* genes might be important for relief of positive supercoils generated when both genes are transcribed. Another example of REP sequences occurring between convergently transcribed genes, *rcsB* and *rcsC*, has been reported recently (34).

**Estimation of the transcription termination point.** To rule out the possibility of genes located downstream of *glpD* in the same operon, the approximate position of the 3′ end of the *glpD* mRNA was determined. This was done by cloning of fragments containing the 5′ region of *glpD* truncated either at *Stu*I (1635) or *Ase*I (1840; Fig. 1 and 2; Table 1) into the promoter probe vector pCB267 (29). This vector contains promoterless, divergently arranged *lacZ* and *phoA* genes separated by a multiple cloning site. If transcription of *glpD* proceeds across the *Stu*I or *Ase*I restriction site, an active *glpD-phoA* transcriptional fusion would be created. An active *glpE-lacZ* transcriptional fusion was created in either case, because each fragment also contained the divergent *glpE* promoter (6, 33, 38). When the *Stu*I site (pDA20) was ligated to the promoterless *phoA* gene in pCB267, PhoA$^+$ LacZ$^+$ transformants of DH5αF′ were obtained. When the *Ase*I site was used (pDA21), the LacZ$^+$ transformants were PhoA$^-$. These results indicate that the transcript generated from pDA20 continued through *Stu*I into *phoA* whereas the transcript generated from pDA21 was terminated before the *Ase*I site. Therefore, the *glpD* transcript must terminate between these two restriction sites.

Evaluation of the nucleotide sequence between 1635 (*Stu*I) and 1840 (*Ase*I) revealed one possible rho-independent terminator (23) consisting of a string of thymine residues at 1745 preceded by a weak putative stem loop. A possible rho-dependent terminator (23) also exists in this region, however, with a 50% G+C-rich stem loop centered on 1720 followed by five adenine residues. This proposed terminator is situated such that the first base in the stem-loop is one base removed from the end of the REP sequences.

**Characteristics of the amino acid sequence of GlpD.** Glycerol-P dehydrogenase is a primary dehydrogenase in the electron transport chain. It functions in vitro only when associated with the cytoplasmic membrane. The enzyme has been purified from the membrane fraction after solubilization with deoxycholate (30). To elucidate how this enzyme may be associated with the membrane, the program designed by Rao and Argos (25) was used to search for potential transmembrane helices in the deduced amino acid sequence. Only one possible transmembrane helix, spanning residues 6 to

TABLE 3. Comparison of deduced and chemically determined (30) amino acid compositions of glycerol 3-P dehydrogenase

| Amino acid | No. of codons | | |
|---|---|---|---|
| | Protein[a] | DNA[b] | DNA[c] |
| Gly | 42.6 | 41 | 42 |
| Ala | 43.6 | 41 | 45 |
| Val | 32.0 | 30 | 36 |
| Leu | 53.3 | 53 | 47 |
| ILe | 25.2 | 26 | 28 |
| Pro | 17.4 | 20 | 22 |
| Phe | 15.5 | 15 | 15 |
| Tyr | 20.3 | 22 | 22 |
| Trp | 7.7 | 12 | 9 |
| Met | 12.6 | 10 | 13 |
| Ser | 31.0 | 30 | 30 |
| Thr | 28.1 | 25 | 24 |
| Cys | 3.9 | 3 | 7 |
| Asp/Asn | 43.6 | 45 | 38 |
| Glu/Gln | 46.5 | 49 | 42 |
| Lys | 26.2 | 25 | 27 |
| Arg | 36.8 | 39 | 39 |
| His | 14.5 | 15 | 18 |
| Total | 500.8 | 501 | 504 |

[a] Values have been corrected to 501 codons.
[b] Predicted by the nucleotide sequence shown in Fig. 2.
[c] Predicted by the nucleotide sequence reported by Choi et al. (5).

32, was located. Because this is the flavin-binding domain (see below), it is unlikely to be involved in localization of the enzyme to the membrane. The hydropathy plot, generated by the method of Kyte and Doolittle (15), indicated that the protein has a hydrophobic amino terminus but is overall sufficiently hydrophilic to exist as a cytosolic protein. It is possible that GlpD becomes associated with the membrane by interaction with a membrane-anchoring subunit(s), as is the case for other respiratory enzymes such as succinate dehydrogenase, anaerobic glycerol-P dehydrogenase, and fumarate reductase (17, 18, 24). However, in each of these cases, the catalytic and membrane-anchoring subunits are encoded by genes present in a polycistronic operon. There are no genes in the immediate vicinity of *glpD*, either upstream (32) or downstream, that are needed for aerobic glycerol-P dehydrogenase function. Thus, it is likely that GlpD is similar to NADH dehydrogenase and D-lactate dehydrogenase of *E. coli*, which are each encoded by single genes. Both enzymes are overall hydrophilic but are tightly membrane associated (24). Recent studies regarding the topography of membrane proteins have indicated that hydrophobicity analysis is insufficient as a sole means of predicting membrane association and that beta structures, containing as few as six residues, are capable of spanning the membrane in proteins virtually devoid of alpha-helical structure (20).

The hybrid protein encoded by the *glpD-lacZ* fusion of pSH85 has an apparent molecular weight only slightly larger than that of β-galactosidase, sediments with the particulate fraction, and can be solubilized by Triton X-100 (32). These results were interpreted to mean that the extreme amino-terminal region of GlpD is sufficient for membrane localization. In fact, the hybrid protein contains only 10 amino acids from GlpD (Fig. 2) and 38 amino acids encoded by Mu DNA (From λ p*lac*Mu). Thus, the interpretation presented above may be incorrect.

A flavin-binding domain, exhibiting 77% similarity with the binding site present in the *glpA*-encoded dehydrogenase, was found between residues 5 and 30 of GlpD. A comparison of the flavin-binding sites within selected flavoproteins with those of GlpD, GlpA, and GlpC is presented in Fig. 3. Although the overall similarities vary from 37 to 77%, the characteristic residues and spacing of the consensus sequence are conserved.

Unlike anaerobic glycerol-P dehydrogenase, consisting of three distinct subunits, one of which is an iron-sulfur protein possibly involved in electron transport (7), a scan of the amino acid sequence of GlpD revealed no clustering of cysteines typical of iron-sulfur proteins.

Comparison of the predicted GlpD polypeptide sequence with bacterial and plant sequences in GenBank (2) revealed one sequence, GlpA (7; GenBank accession number M20938), that displayed significant sequence similarity to GlpD. Alignment of the amino acid sequences of GlpD and GlpA (Fig. 4) revealed 80% identity in the amino-terminal regions with periodicities of substantial homology elsewhere, particularly in the carboxy-terminal regions. Overall, the two polypeptides were 25% identical and 46% similar.

The region displaying the most sequence similarity between GlpD and GlpA is the FAD-binding site. The other region exhibiting a high degree of sequence similarity (residues 264 to 370) could be the glycerol-P-binding site of the dehydrogenase enzymes. To investigate this possibility, this region of GlpD was used as the query sequence to search GenBank for similar amino acid sequences. The results revealed significant amino acid sequence similarity between

| Amino acid sequence | Origin | % Similarity |
|---|---|---|
| **GXGXXGXXXAXXXXXXXGXXXX** | | |
| DLIVIGGGINGAGIAADAAGRGLSVLMLEA | GlpD | (100) |
| DVIIIGGGATGAGIARDCALRGLRVILVER | GlpA | 77 |
| DTVIMGGGLAGLLCGLQLQKHGLRCAIVTR | GlpB | 53 |
| DYIAIGGGSGGIASINRAAMYGQKCALIEA | GR-*E.* | 50 |
| DYLVIGGGSGGLASARRAAELGARAAVVES | GR-H | 57 |
| DLAIVGAGGAGLRAAIAAAQANAKIALISK | FrdA | 47 |
| KIVIVGGGAGGLEMATQLGHKKAKITLVDR | NDH | 57 |
| KLLILGSGPAGYTAAVYAARANLQPVKITG | TrxB | 53 |
| QVVVLGAGPAGYSAAFRCADLGLETVIVER | LPD-*E.* | 60 |
| DVTVIGSGPGGYVAAIKAAQLGFKTVCIEK | LPD-H | 50 |
| DVTVIGSGPGGYVAAIKAAQLGFKRVCIEK | LPD-P | 50 |
| DAVVIGAGGAGIARLAQISQSGQTCALLSK | SdhA | 53 |
| QICVVGSGPAGFYTAQHLLKHHSRAHVDIY | AdrR-B | 37 |
| QICVVGSGPAGFYTAQHLLKHP-QAHVDIY | AdrR-H | 37 |
| QVAVIGSGGAAMAAALKAVEQGAQVTLIER | MerR-*P.* | 53 |
| DNVIVGTGLAGVEVAFGLRASGWEGNIRLV | PutR | 47 |

FIG. 3. Comparison of the amino acid sequences of proposed flavin-binding domains in GlpD (aerobic glycerol-P dehydrogenase), GlpA and GlpB (anaerobic glycerol-P dehydrogenase from *E. coli*) (7), GR-*E.* (glutathione reductase from *E. coli*) (11), GR-H (human glutathione reductase) (11), FrdA (fumarate reductase from *E. coli*) (7), NDH (NADH dehydrogenase from *E. coli*) (7), TrxB (thioredoxin reductase from *E. coli*) (11), LPD-*E.* (lipoamide dehydrogenase from *E. coli*) (7), LPD-H (human lipoamide dehydrogenase) (11), LPD-P (porcine lipoamide dehydrogenase) (11), SdhA (succinate dehydrogenase from *E. coli*) (7), AdrR-B (bovine adrenodoxin reductase) (11), AdrR-H (human adrenodoxin reductase) (11), MerR-P. (mercuric reductase from *Pseudomonas* sp.) (11), and PutR (putidaredoxin reductase from *Pseudomonas putida* (11). Percent similarity reflects the degree of similarity that each protein exhibits with respect to the GlpD sequence. For comparison, amino acids were grouped according to chemical similarity (GASTP, QNED, FYW, ILVM, HRK, C).

this region of GlpD and TPIs (from yeast cells [1] and *E. coli* [22]; GenBank accession numbers J01366 and X00617). The alignment of the GlpD and GlpA sequences with that of *E. coli* TPI is shown in Fig. 4. It is reasonable to propose that the glycerol-P-binding site of glycerol-P dehydrogenase resembles the substrate-binding site of TPI in that the product of the glycerol-P dehydrogenase reaction (dihydroxyacetone phosphate) is the substrate for the reaction catalyzed by TPI. It is known that residues 168 to 177, 209 to 212, and 232 to 234 are highly conserved among known TPI sequences (1, 22) and that these residues are involved in phosphate binding. Residues 168 to 172 form a flexible loop that folds down on the phosphate end of the substrate (1); the phosphate is held in place by hydrogen bonds from the main chain of two glycine-containing loops (residues 209 to 212 and 232 to 234 [1]; overlined in Fig. 4). It is interesting that the corresponding regions of GlpD and GlpA contain highly conserved sequences that could function in an analogous manner. We also noticed that a portion of GlpC, the other catalytic subunit of the anaerobic glycerol-P dehydrogenase, exhibited sequence similarity to GlpA and GlpD in this region (Fig. 4). It has been pointed out that residues 74 to 90 of GlpC may be involved in catalysis (7). All of these observations suggest that the glycerol-P-binding site for glycerol-P dehydrogenase encompasses, at least in part, residues 300 to 390. The similarity between GlpD and TPI provides a basis for experimental identification of residues critical for substrate binding and enzyme activity.

```
       GlpD
    METKDLIVIGGGINGAGIAADAAGRGLSVLMLEAQDLACATSSASSKLIHGGLRYLEHYEFRLVSEALAEREVLLKMAPHIAFPMRFRLPHRPHL   95
GlpA  ** :*:*::::  ::::: : : ::: :***:  :*: ****    :*:*: ::     *       : **: **: **: :    :            :
MKTRDSQSSDVIIIGGGATGAGIARDCALRGLRVILVERHDIATGATGRNHGLLHSGARY-AVTDAESARECISENQILKRIARHCVEPTNGLFITLPED   99

RPAWMIRIGLFMYDHLGKRTSLPGSTGLRFGANSVLKPEIKRGFEYSDCWVDDARLVLANAQMVVRKGGEVLTRTRATSARRENG-LWIVEAEDIDTGKK  194
 **     *    *   :   *     *    :    ** * : *   *    *:   ::   ::   ::        *:* *::     :*  :: * *   :    *  ::
DLSFQAT-FIRACEEAGISAEAIDPQQAR-IIEPAVNPALIGAVKVPDGTVDPFRLTAANMLDAKEHGAVILTAHEVTGLIREGATVCGVRVRNHLTGET  197

YSWQARGLVNATGPWVKQFFDDGMHLPSPYGIRLIKGSHIVVPRVHTQKQAYILQNEDKRIVFVIPWMDEFSIIGTT--DVEYKGDPKAVKIEESEINYL  292
 *  :   *:::*: :  :    *    :            ::: *** * * *                   **:  :  :*::::   **:   *     **    :** :
QALHAPVVVNAAGIWG-QHIAEYADLRIR-MF-PAKGSLLIMDH-RINQHVINRCRKPSDADILVP-GDTISLIGTTSLRIDY-NEIDDNRVTAEEVDIL  291

                |
TPI (160) AVIAYEPVWAIGTGKSATPAQAQAVHKFI-RDHIA-KVDANIAEQVIIQ-YGGSVNASNAAELFAQPDIDGALVGGASLKADAFAVIVKA  242
              *    * ::: ** **  ****** *  : :: *    : : * * **  *:: * *        *:    * * : :***  *    :*    :
LNVYNTHFKKQLSRDDIVWTYSGVRPLCDDESDSP-QAITRDYTLDIHDEN-GKAPLLSVFGGKLTTYRKLAEHALEKLTPYYQGIGPAWTKESVLPGGA  389
:      *     ***  :* *:::::::     * :**     **:     :    : : *:   * *** :::: ::: *:: : * *          :   :  *   :::*
LR-EGEKLAPVMAKTRILRAYSGVRPLVASDDDPSGRNVSRGIVLLDHAERDGLDGFITITGGKLMTYRLMAEWATDAVCRKLGNTRPCTTADLALPGS-  389
              :   *    **:: * :     :      *::*   * * ** *   *       *:    :*    *:*     *:  ** *:**:      * * **
GlpC (66) SLRQQAPAHPYSLLEPQRVLDLACQAQALIAESGAQLQGSVELAHQRVTPLGTLRSTWLSSPEVPVWPLPAKKICVVGISGL . . . (149)

AEAAKQA   249   TPI
 :* **
IEGDRDDYAARLRRRYPFLTESLARHYARTYGSNSELLLGNAGTVSDLGEDFGHEFYEAELKYLVDHEWVRRADDALWRRTKQGM-WLNADQQSRVSQWL  487
 :*    *   * ::*   * :* *:              :   *     :***       :   *        *:* : :*    :      : : :::* ::     ***  *   *   :
QEP--AE-VT-LRKVIS-LPAPLRGSAVYRHGDRTPAWLSEGRLHRSLVCE-CEAVTAGEVQYAVENLNVNSLLD-LRRRTRVGMGTCQGELCACRAAGL  480

VEYTQQRLSLAS   501   GlpD
**   *    :   :
LQRFNVTTSAQSIEQLSTFLNERWKGVQPIAWGDALRESEFTRWVYQGLCGLEKEQKDAL   542   GlpA
```

FIG. 4. Alignment of the GlpD, GlpA, GlpC, and TPI amino acid sequences. Amino acid identities are indicated by colons. Conservative amino acid substitutions (asterisks) are defined in the legend to Fig. 3. Where TPI and GlpC are compared, the identities or similarities are indicated with respect to either GlpD or GlpA. Dashed overlining above the TPI sequence indicates residues 168 to 177, 209 to 212, and 232 to 234, which interact with phosphate of the substrate (1). Active site E165 of TPI is indicated by the vertical line. The numbering system for TPI is that used for the yeast sequence (1).

Sequence and structural similarities between enzymes catalyzing sequential steps of the glycolytic pathway have been sought (9). Phosphoglycerate mutase and diphosphoglycerate mutase are clearly related and probably have evolved following a gene duplication event. Glyceraldehyde 3-phosphate dehydrogenase and phosphoglycerate kinase also appear to be distantly related. The TPI/glycerol-P dehydrogenase pair represents another example for which sequence similarity is observed; the similarity is likely to be the result of convergent evolution, however, because glycerol-P dehydrogenase is very different from TPI in size, subcellular location, and function.

There are many reports describing enzymes with active-site arginyl residues that act as positively charged receptors for anionic substrates (26). Most of the glycolytic enzymes, all of which have phosphorylated substrates or cofactors, contain active-site arginyl residues (26). The region of substantial sequence identity in the glycerol-P dehydrogenase (Fig. 4) contains conserved arginyl residues that may aid in the binding of glycerol-P or dihydroxyacetone phosphate. It should be pointed out, however, that cationic side chains are generally used for phosphate binding when the phosphate is hydrolyzed or transferred and not when phosphate is simply bound by an enzyme during catalysis (1).

The observed similarity between the sequences of GlpD, GlpA, and GlpC suggests that the genes encoding these proteins evolved from a common ancestral gene. Thus, GlpD acquired the ability to associate with the membrane and the aerobic electron transport chain. It functions as a homodimer. The anaerobic glycerol-P dehydrogenase functions as a heterodimer of related subunits (GlpA and GlpC) and requires a third subunit (GlpB) for interaction with the membrane and with the anaerobic electron transport chain. The control regions for the operons diverged such that each enzyme is expressed under the appropriate environmental condition.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant DMB-8696072.

Joel Weiner pointed out the sequence similarity between the amino termini of GlpA and GlpD.

## REFERENCES

1. **Alber, T. C., R. C. Davenport, Jr., D. A. Giammona, E. Lolis, G. A. Petsko, and D. Ringe.** 1987. Crystallography and site-directed mutagenesis of yeast triosephosphate isomerase. What can we learn from a "simple" enzyme? Cold Spring Harbor Symp. Quant. Biol. **52:**603–613.
2. **Bilofsky, H. S., and C. Burks.** 1988. The GenBank® genetic sequence data bank. Nucleic Acids Res. **16:**1861–1864.
3. **Bremer, E., T. J. Silhavy, and J. M. Weissemann, and G. M. Weinstock.** 1984. λ p*lac*Mu: a transposable derivative of bacteriophage lambda for creating *lacZ* protein fusions in a single step. J. Bacteriol. **158:**1084–1093.
4. **Choi, Y.-L., M. Kawamukai, R. Utsumi, H. Sakai, and T. Komano.** 1989. Molecular cloning and sequencing of the glycogen phosphorylase gene from *Escherichia coli*. FEBS Lett. **243:**193–198.
5. **Choi, Y.-L., S. Kawase, M. Kawamukai, R. Utsumi, H. Sakai, and T. Komano.** 1989. Nucleotide sequence of the glycerol-3-phosphate dehydrogenase gene of *Escherichia coli* and regulation by the cAMP-CRP complex. Agric. Biol. Chem. **53:**1135–1143.
6. **Choi, Y.-L., S. Kawase, T. Nishida, H. Sakai, T. Komano, M. Kawamukai, R. Utsumi, Y. Kohara, and K. Akiyama.** 1988.

Nucleotide sequence of the *glpR* gene encoding the repressor for the glycerol-3-phosphate regulon of *Escherichia coli* K12. Nucleic Acids Res. **16**:7732.

7. **Cole, S. T., K. Eiglmeier, S. Ahmed, N. Honore, L. Elmes, W. F. Anderson, and J. H. Weiner.** 1988. Nucleotide sequence and gene-polypeptide relationships of the *glpABC* operon encoding the anaerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli*. J. Bacteriol. **170**:2448–2456.
8. **Ehrmann, M., W. Boos, E. Ormseth, H. Schweizer, and T. J. Larson.** 1987. Divergent transcription of the *sn*-glycerol-3-phosphate active transport (*glpT*) and anaerobic *sn*-glycerol-3-phosphate dehydrogenase (*glpA glpC glpB*) genes of *Escherichia coli* K-12. J. Bacteriol. **169**:526–532.
9. **Fothergill-Gilmore, L. A.** 1986. Domains of glycolytic enzymes, p. 85–174. *In* D. G. Hardie and J. R. Coggins (ed.), Multidomain proteins—structure and evolution. Elsevier, Amsterdam.
10. **Hackett, P. B., J. A. Fuchs, and J. W. Messing.** 1988. An introduction to recombinant DNA techniques: basic experiments in gene manipulation, p. 153–157. Benjamin/Cummings Publishing Co., Reading, Mass.
11. **Hanukoglu, I., and T. Gutfinger.** 1989. cDNA sequence of adrenodoxin reductase: identification of NADP-binding sites in oxidoreductases. Eur. J. Biochem. **180**:479–484.
12. **Higgins, C. F., R. S. McLaren, and S. F. Newbury.** 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion?—a review. Gene **72**:3–14.
13. **Iuchi, S., D. C. Cameron, and E. C. C. Lin.** 1989. A second global regulator gene (*arcB*) mediating repression of enzymes in aerobic pathways of *Escherichia coli*. J. Bacteriol. **171**:868–873.
14. **Iuchi, S., S. T. Cole, and E. C. C. Lin.** 1990. Multiple regulatory elements for the *glpA* operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the *glpD* operon encoding aerobic glycerol-3-phosphate dehydrogenase in *Escherichia coli*: further characterization of respiratory control. J. Bacteriol. **172**:179–184.
15. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
16. **Larson, T. J., S. Ye, D. L. Weissenborn, H. J. Hoffmann, and H. Schweizer.** 1987. Purification and characterization of the repressor for the *glp* regulon of *Escherichia coli* K-12. J. Biol. Chem. **262**:15869–15874.
17. **Lin, E. C. C.** 1987. Dissimilatory pathways for sugars, polyols, and carboxylates, p. 244–284. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
18. **Lin, E. C. C., and D. R. Kuritzkes.** 1987. Pathways for anaerobic electron transport, p. 201–221. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
19. **Muramatsu, S., and T. Mizuno.** 1989. Nucleotide sequence of the region encompassing the *glpKF* operon and its upstream region containing a bent DNA sequence of *Escherichia coli*. Nucleic Acids Res. **17**:4378.
20. **Paul, C., and J. P. Rosenbusch.** 1985. Folding patterns of porin and bacteriorhodopsin. EMBO J. **4**:1593–1597.
21. **Pettigrew, D. W., D.-P. Ma, C. A. Conrad, and J. R. Johnson.** 1988. *Escherichia coli* glycerol kinase. Cloning and sequencing of the *glpK* gene and the primary structure of the enzyme. J. Biol. Chem. **263**:135–139.
22. **Pichersky, E., L. D. Gottlieb, and J. F. Hess.** 1984. Nucleotide sequence of the triose phosphate isomerase gene of *Escherichia coli*. Mol. Gen. Genet. **194**:314–320.
23. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
24. **Poole, R. K., and W. J. Ingledew.** 1987. Pathways of electrons to oxygen, p. 170–200. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
25. **Rao, M. J. K., and P. Argos.** 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta **869**:197–214.
26. **Riordan, J. F., K. D. McElvany, and C. L. Borders, Jr.** 1977. Arginyl residues: anion recognition sites in enzymes. Science **195**:884–886.
27. **Rodriguez, R. L., and R. C. Tait.** 1983. Recombinant DNA techniques: an introduction, p. 164–165. Addison-Wesley Publishing Co., Reading, Mass.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Schneider, K., and C. F. Beck.** 1986. Promoter-probe vectors for the analysis of divergently arranged promoters. Gene **42**:37–48.
30. **Schryvers, A., E. Lohmeier, and J. H. Weiner.** 1978. Chemical and functional properties of the native and reconstituted forms of the membrane-bound, aerobic glycerol-3-phosphate dehydrogenase of *Escherichia coli*. J. Biol. Chem. **253**:783–788.
31. **Schweizer, H., W. Boos, and T. J. Larson.** 1985. Repressor for the *sn*-glycerol 3-phosphate regulon of *Escherichia coli*: cloning of the *glpR* gene and identification of its product. J. Bacteriol. **161**:563–566.
32. **Schweizer, H., and T. J. Larson.** 1987. Cloning and characterization of the aerobic *sn*-glycerol-3-phosphate dehydrogenase structural gene *glpD* of *Escherichia coli* K-12. J. Bacteriol. **169**:507–513.
33. **Schweizer, H., G. Sweet, and T. J. Larson.** 1986. Physical and genetic structure of the *glpD-malT* interval of the *Escherichia coli* K-12 chromosome. Identification of two new genes of the *glp* regulon. Mol. Gen. Genet. **202**:488–492.
34. **Stout, V., and S. Gottesman.** 1990. RcsB and RcsC: a two-component regulator of capsule synthesis in *Escherichia coli*. J. Bacteriol. **172**:659–669.
35. **Sweet, G., C. Gandor, R. Voegele, N. Wittekindt, J. Beuerle, V. Truniger, E. C. C. Lin, and W. Boos.** 1990. Glycerol facilitator of *Escherichia coli*. Cloning of *glpF* and identification of the *glpF* product. J. Bacteriol. **172**:424–430.
36. **Yang, Y., and G. F.-L. Ames.** 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. Proc. Natl. Acad. Sci. USA **85**:8850–8854.
37. **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.
38. **Ye, S., and T. J. Larson.** 1988. Structures of the promoter and operator of the *glpD* gene encoding aerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. J. Bacteriol. **170**:4209–4215.
39. **Yu, F., Y. Jen, E. Takeuchi, M. Inouye, H. Nakayama, M. Tagaya, and T. Fukui.** 1988. Alpha-glucan phosphorylase from *Escherichia coli*. J. Biol. Chem. **263**:13706–13711.