JOURNAL OF BACTERIOLOGY, Dec. 1996, p. 7080–7089 Vol. 178, No. 24
0021-9193/96/$04.00+0

Repressor for the *sn*-Glycerol 3-Phosphate Regulon of *Escherichia coli* K-12: Primary Structure and Identification of the DNA-Binding Domain

GANG ZENG,† SHANZHANG YE,‡ AND TIMOTHY J. LARSON*

Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0308

Received 5 June 1996/Accepted 8 October 1996

The nucleotide sequence of the *glpEGR* operon of *Escherichia coli* was determined. The translational reading frame at the beginning, middle, and end of each gene was verified. The *glpE* gene encodes an acidic, cytoplasmic protein of 108 amino acids with a molecular weight of 12,082. The *glpG* gene encodes a basic, cytoplasmic membrane-associated protein of 276 amino acids with a molecular weight of 31,278. The functions of GlpE and GlpG are unknown. The *glpR* gene encodes the repressor for the glycerol 3-phosphate regulon, a protein predicted to contain 252 amino acids with a calculated molecular weight of 28,048. The amino acid sequence of the *glp* repressor was similar to several repressors of carbohydrate catabolic systems, including those of the glucitol (GutR), fucose (FucR), and deoxyribonucleoside (DeoR) systems of *E. coli*, as well as those of the lactose (LacR) and inositol (IolR) systems of gram-positive bacteria and agrocinopine (AccR) system of *Agrobacterium tumefaciens*. These repressors constitute a family of related proteins, all of which contain approximately 250 amino acids, possess a helix-turn-helix DNA-binding motif near the amino terminus, and bind a sugar phosphate molecule as the inducing signal. The DNA recognition helix of the *glp* repressor and the nucleotide sequence of the *glp* operator were very similar to those of the *deo* system. The presumptive recognition helix of the *glp* repressor was changed by site-directed mutagenesis to match that of the *deo* repressor or, in a separate construct, to abolish DNA binding. Neither altered form of the *glp* repressor recognized the *glp* or *deo* operator, either in vivo or in vitro. However, both altered forms of the *glp* repressor were negatively dominant to the wild-type *glp* repressor, indicating that the inability to bind DNA with high affinity was due to alteration of the DNA-binding domain, not to an inability to oligomerize or instability of the altered repressors. For the first time, analysis of repressors with altered DNA-binding domains has verified the assignment of the helix-turn-helix motif of the transcriptional regulators in the *deoR* family.

The genes of the *glp* regulon of *Escherichia coli* encode the proteins needed for the dissimilation of *sn*-glycerol 3-phosphate (glycerol-P) and its precursors, glycerol and glycerophosphodiesters (27, 28). The five operons that constitute the *glp* regulon are located at three different positions on the chromosome. Transcription of the *glp* operons is subject to multiple controls, including catabolite repression mediated by cyclic AMP-CRP and respiratory control mediated by the FNR and ArcA/ArcB systems (19, 28). In addition, each of the operons is negatively controlled by a repressor specific for the regulon, the *glp* repressor. The extent of repression is different for each operon. Repression is relieved in the presence of the inducer for the regulon, glycerol-P (28).

The *glpTQ* and *glpACB* operons, located near 51 min of the linkage map, encode the glycerol-P permease/glycerophosphodiesterase and the subunits of the anaerobic glycerol-P dehydrogenase, respectively (10, 13, 14, 47). These operons are transcribed divergently from a common control region that contains five operator sites for binding of the *glp* repressor. Control of these operons by the *glp* repressor appears to be dependent upon repressor-mediated DNA loop formation (23).

The *glpD* operon (2), encoding aerobic glycerol-P dehydrogenase, is located near 77 min, and the *glpFKX* operon (45, 48), encoding glycerol diffusion facilitator and glycerol kinase, is located near 89 min. These operons are relatively tightly controlled by cooperative binding of the *glp* repressor to tandem operator sites which overlap the promoters (52, 56, 60). Comparison of operator sequences and determination of the binding of the *glp* repressor to operator variants in vivo have revealed the operator consensus half site of WATKYTCGWW, where W is A or T, K is G or T, and Y is C or T (60).

The *glpE*, *glpG*, and *glpR* genes (41) are transcribed divergently from the adjacent *glpD* gene. The *glpE* and *glpG* genes encode proteins of unknown function, and the *glpR* gene encodes the *glp* repressor. The *glpR* gene was cloned in an expression vector which facilitated overproduction and purification of the *glp* repressor (25). The purified repressor exhibits a subunit molecular weight of 30,000 and exists as a tetramer under nondenaturing conditions. The purified repressor specifically binds glycerol-P, the inducer (K_d = 31 μM), or DNA fragments harboring operator sites. The binding to operator DNA is diminished in the presence of glycerol-P.

In order to obtain molecular details concerning the structure and function of the *glp* repressor, the nucleotide sequence of the *glpEGR* operon was determined in the present work. On the basis of sequence similarity to other prokaryotic repressors, the helix-turn-helix DNA-binding motif of the *glp* repressor has been identified. Both in vivo and in vitro evidence confirming the identity of the DNA-binding motif was obtained by characterization of *glp* repressor variants with altered DNA recognition helices.

* Corresponding author. Phone: (540) 231-7060. Fax: (540) 231-9070. Electronic mail address: tilarson@vt.edu.

† Present address: Research Institute for Genetic & Human Therapy, Gaithersburg, MD 20879-4117.

‡ Present address: Department of Pharmacology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA 19107.

TABLE 1. Strains of *E. coli* K-12 used in this study

Strain	Genotype	Derivation, source, or reference
MC4100	F$^-$ *araD139* Δ(*argF-lac*)*U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	6
SH305	MC4100 Δ*glpD102 recA1 srl*::Tn*10*	40
GD2	MC4100 *glpR2 zih-730*::Tn*10* Φ(*glpK-lacZ*)(Hyb) λp*lac*Mu	40
GD6	MC4100 *glpR2* Φ(*glpD-lacZ*)(Hyb) λp*lac*Mu	41
GD31	MC4100 Φ(*glpK-lacZ*)(Hyb) λp*lac*Mu	P1(GD2) × MC4100 Lac$^+$ Tet^s selection
SY102	MC4100 *glpR2* Φ(*glpD-lacZ*)(Hyb) λp*lac*Mu *recA1 srl*::Tn*10*	P1(SH305) × GD6 Tet^r UV^s selection
SY105	MC4100 Φ(*glpK-lacZ*)(Hyb) λp*lac*Mu *recA1 srl*::Tn*10*	P1(SH305) × GD31 Tet^r UV^s selection
KH682	*deoR lac* Φ(*deoC-lacZ*)(Hyb) *thi udp upp ton*	K. Hammer
RZ1032	HfrK116 *PO/45* [*lysA*(*61-62*)] *dut-1 ung-1 thi-1 relA1 zbd-279*::Tn*10 supE44*	22
DH5αF′	(F′ ϕ80*dlacZ*Δ*M15*) *endA1 recA1 hsdR17 supE44 thi-1 gyrA relA1* Δ(*lacZYA-argF*)*U169*	29
BL21(DE3)	*hsdS gal* (λc*Its857 ind1 Sam7nin5 lacUV5-T7gene1*)	44

MATERIALS AND METHODS

Materials. Restriction endonucleases, T4 DNA ligase, polynucleotide kinase, and reagents for DNA sequencing and PCR were obtained from Stratagene, Perkin-Elmer Cetus, U.S. Biochemical Corporation, Boehringer Mannheim, Promega, or New England Biolabs. Other chemicals were obtained from Sigma Chemical Co. or Fisher Scientific. Oligonucleotide primers were synthesized on an Applied Biosystems 381A synthesizer and purified by OPC (Applied Biosystems), as recommended by the manufacturer. DuPont/NEN supplied α-^{35}S-dATP and [γ-^{32}P]ATP.

Bacterial strains and growth media. The strains of *E. coli* K-12 used in this study are described in Table 1. Bacteria were routinely grown in Luria-Bertani medium (32) supplemented with the appropriate antibiotics at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 30 μg/ml; kanamycin, 50 μg/ml; and tetracycline, 12.5 μg/ml. Glucose, glycerol, or maltose (0.2%) was provided as a carbon source in various experiments, as described in the text. For minimal medium, the A and B salts of Clark and Maaløe (9) were supplemented with 0.2% maltose, 0.2% Casamino Acids, and 2 μg of thiamine per ml.

Construction of recombinant phage and plasmids. The plasmids used or constructed in this study are listed in Table 2. Plasmid pSH79 contains a 3-kb *Eco*RI-*Sal*I fragment with the entire *glpEGR* operon cloned downstream of the phage T7 promoter of pBluescript KS$^+$ (56) and was used as the source of DNA fragments for sequencing and mutagenesis.

To determine the start codon of *glpE*, two PCR primers, each bearing an *Nde*I site (italicized), were made on the basis of the two potential start codons of *glpE*. They were GAAAGAGAGA*CATATG*GATCAGTTCG and AAAGAG*CATATG*CTGGTCGATATTCGCG (the first underlined ATG is at positions 373 to 375 [see Fig. 1], and the second is changed from the GTG at positions 436 to 438). A downstream primer bearing a *Hin*dIII site (italicized) with its sequence complementary to the *glpE* stop codon (underlined), TAA*AAGCTTA*CGCGCCGT ACGCCAC, was also made. PCR products from each of the primer pairs were cloned separately between the *Nde*I and *Hin*dIII site of the expression vector pT7-7 (46). Plasmids containing the first and second start codons were named pGZ105 and pGZ106, respectively. The plasmids were transformed into strain BL21(DE3). Induction of GlpE expression with IPTG (isopropyl-β-D-thiogalactopyranoside) was performed by the method of Maniatis et al. (31).

An expression vector with a multiple cloning site inserted downstream of the *tac* promoter of pKK223-3 (Pharmacia) (5) and a resident *lacI*^{q} gene was constructed and used for controlled overexpression of the *glpEGR* operon. First, pKK223-3 was digested with *Pst*I and *Hin*dIII, and complementary synthetic oligonucleotides (5′-AGCTTGATATCATCGATCTAGAGGCCTCGAGAGA TCTGCA-3′ and 5′-GATCTCTCGAGGCCTCTAGATCGATGATATCA-3′) were ligated into these sites in order to introduce *Eco*RV, *Cla*I, *Xba*I, *Stu*I, *Xho*I, and *Bgl*II restriction cleavage sites. The resulting plasmid was named pDA223. Next, a 1,300-bp *Eco*RI-*Hin*dIII fragment from pMJR1560 (43) containing the *lacI*^{q} gene was blunt ended and cloned into the *Nru*I site of pDA223. In the resulting plasmid (pSY223), the *lacI*^{q} gene was oriented so that its transcription was opposite that of the *bla* gene.

Sequencing of DNA. Sequence information was obtained using the dideoxy-chain termination technique (39). Templates were generated by direct cloning of appropriate restriction fragments into pBluescript KS$^+$ (Stratagene). The flanking SK, KS, T3 or T7 primers were used to generate sequence data. Additional sequence information was obtained from two series of unidirectional exonuclease III deletions (17), generated by using two derivatives of M13mp18 RF DNA in which a *Bgl*II fragment carrying the *glpEGR* operon was inserted in opposite orientations. Finally, synthetic oligonucleotides were used as primers to obtain the remaining sequence information. Both strands of DNA were sequenced completely by multiple sequencing reactions.

Overexpression of GlpG and N-terminal amino acid sequence analysis. A PCR product containing *glpEGR* (nucleotides 373 to 2524 [see Fig. 1]) was cloned between the *Nde*I and *Sal*I site of expression vector pT7-7. The plasmid thus constructed was named pGZ126. The transformation and induction procedures were the same as those used above for production of GlpE. The cytoplasmic membrane fraction from strain BL21(DE3)(pGZ126) was obtained by sucrose gradient sedimentation as described by Larson et al. (24). GlpG was purified further by using a preparative sodium dodecyl sulfate (SDS)-polyacrylamide gel. The region containing GlpG was excised and subjected to electroelution by using an Elutrap (Schleicher & Schuell). About 100 pmol of the final product was analyzed with an ABI amino acid microsequencer. GlpR was purified as previously described (25) and subjected to N-terminal sequence analysis.

Oligonucleotide-directed mutagenesis. The technique of Kunkel et al. (22) was used for making specific alterations in the nucleotide sequence of the *glpR* gene. For this purpose, the 3-kb *Eco*RI-*Sal*I DNA fragment from pSH79 encoding the *glpEGR* operon was cloned into M13mp19. Templates containing uracil were generated in strain RZ1032 (*dut-1 ung-1*). The mutagenic oligonucleotides used (mismatches are italicized) were TTCTCCGTCAGC*GA*G*AT*GACTATTCG CCG (P33E and Q34M) and GACTATTCGC*GC*CGACCTCAATG (R38A). After mutagenesis, 694-bp *Nsi*I-*Eco*RV fragments harboring the desired alterations were recloned back into pSH79 as cassettes. The nucleotide sequence of the *glpR* gene to the *Eco*RV site was verified in each case. *Bgl*II DNA restriction fragments encoding the wild-type and variant repressors were cloned into the *Bgl*II site of the expression vector pSY223 for analysis of function (pSY2-C [GlpR$^+$], pSY2-IX [GlpR^{P33E,Q34M}], and pSY2-A9 [GlpR^{R38A}] [Table 2]).

Assay of β-galactosidase. Enzyme activity was determined with cells permeabilized with chloroform and SDS (32). Specific activities were expressed in Miller units, and data are the averages of duplicate or triplicate determinations.

Assay of operator binding. Gel mobility shift assays were used to assess the binding of repressor variants to the *glp* and *deo* operators in vitro. PCR primers ACGACTCACTATAGGGCGAATTCG and GAATACTCAAGCTTGCATG CCTGC were used to amplify the consensus single *glp* and *deo* operators cloned in vector pGEM3Z (60). PCR products containing the *glp* or *deo* operator were radiolabeled by using T4 polynucleotide kinase and [γ-^{32}P]ATP. Bio-spin 30 columns (Bio-Rad) were used to separate operator DNA from unincorporated ATP. Purified operator DNA was incubated with extract containing GlpR (or a variant) at 30°C for 30 min. The association buffer used was that previously described (25). Samples were analyzed on 8% polyacrylamide gels containing one-half strength Tris-borate-EDTA buffer (pH 8.3). Electrophoresis was performed at 160 V and 4°C.

Nucleotide sequence accession number. The nucleotide sequence reported in this paper is found in the GenBank database under accession number M96795.

TABLE 2. Plasmids used or constructed in this study

Plasmid	Description	Reference
pKH31	pBR322 derivative carrying *deoR*	12
pSH79	pBluescript KS$^+$ derivative carrying *glpEGR*	56
pSY2-A	pSH79 derivative encoding the P33E and Q34M GlpR variant	This work
pSY2-9	pSH79 derivative encoding the R38A GlpR variant	This work
pDA223	pKK223-3 with expanded multiple cloning site	This work
pSY223	pDA223 with *lacI*^{q}	This work
pSY2-C	pSY223 with *glpEGR* (167–2895)	This work
pSY2-IX	pSY2-C derivative encoding the P33E and Q34M GlpR variant	This work
pSY2-A9	pSY2-C derivative encoding the R38A GlpR variant	This work
pGZ105	pT7-7 derivative carrying *glpE* (373–703)	This work
pGZ106	pT7-7 derivative carrying *glpE* (436–703)	This work
pGZ125	pGEM3Z (Promega) derivative carrying *glpG* (679–1578) on a *Bgl*II-*Hin*dIII fragment cloned downstream of the T7 promoter	This work
pGZ126	pT7-7 derivative carrying *glpEGR* (373–2524)	This work

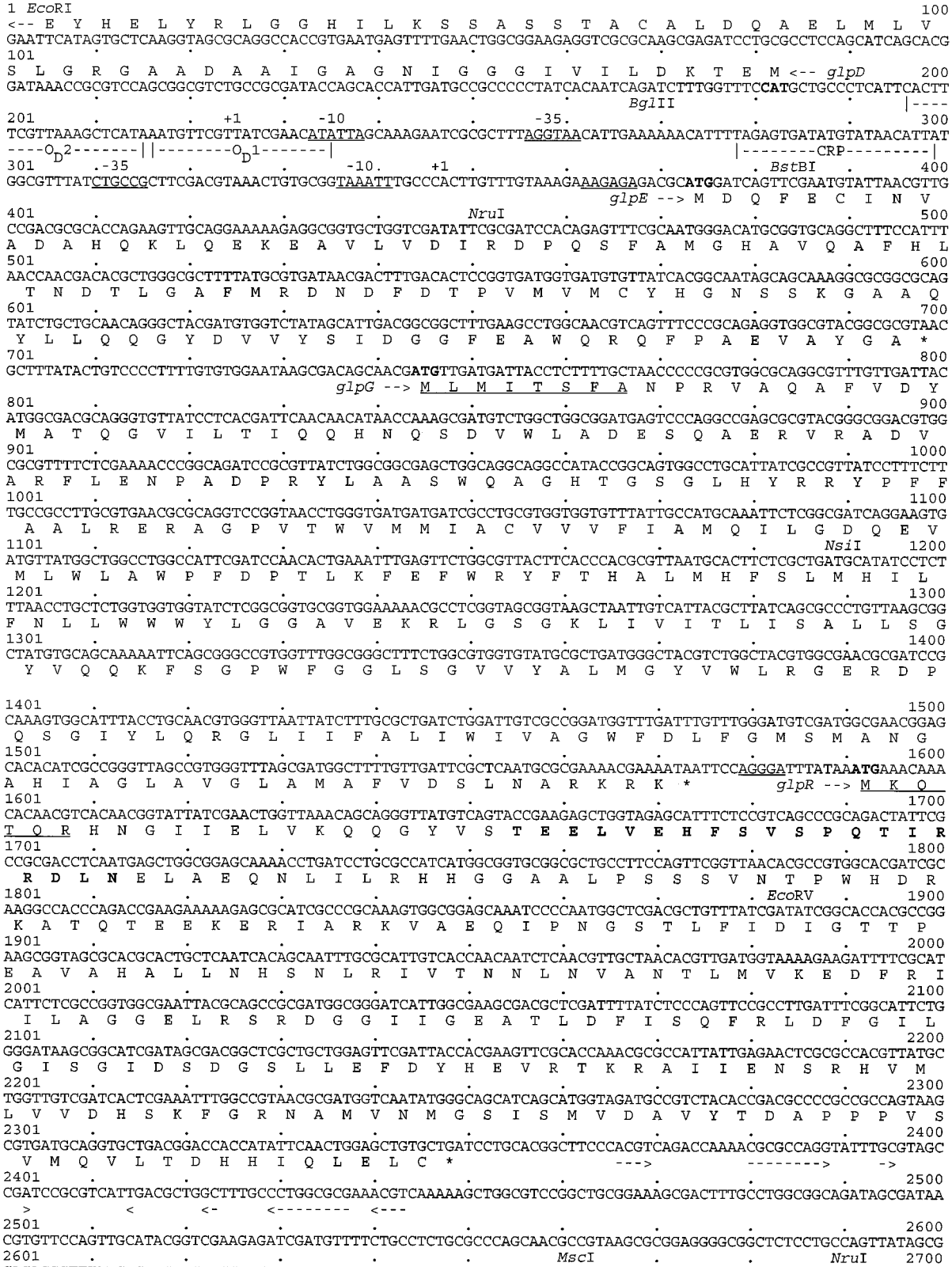
1 EcoRI
100
<-- E Y H E L Y R L G G H I L K S S A S S T A C A L D Q A E L M L V
GAATTCATAGTGCTCAAGGTAGCGCAGGCCACCGTGAATGAGTTTTGAACTGGCGGAAGAGGTCGCGCAAGCGAGATCCTGCGCCTCCAGCATCAGCACG
101
200
S L G R G A A D A A I G A G N I G G G I V I L D K T E M <-- glpD
GATAAACCGCGTCCAGCGGCGTCTGCCGCGATACCAGCACCATTGATGCCGCCCCCTATCACAATCAGATCTTTGGTTTCCATGCTGCCCTCATTCACTT
BglII
201
300
TCGTTAAAGCTCATAAATGTTCGTTATCGAACATATTAGCAAAGAATCGCGCTTTAGGTAACATTGAAAAAACATTTTAGAGTGATATGTATAACATTAT
----OD2-------||--------OD1-------|
|--------CRP---------|
301
400
GGCGTTTATCTGCCGCTTCGACGTAAACTGTGCGGTAAATTTGCCCACTTGTTTGTAAAGAAAGAGAGACGCATGGATCAGTTCGAATGTATTAACGTTG
BstBI
glpE --> M D Q F E C I N V
401
500
CCGACGCGCACCAGAAGTTGCAGGAAAAAGAGGCGGTGCTGGTCGATATTCGCGATCCACAGAGTTTCGCAATGGGACATGCGGTGCAGGCTTTCCATTT
NruI
A D A H Q K L Q E K E A V L V D I R D P Q S F A M G H A V Q A F H L
501
600
AACCAACGACACGCTGGGCGCTTTTATGCGTGATAACGACTTTGACACTCCGGTGATGGTGATGTGTTATCACGGCAATAGCAGCAAAGGCGCGGCGCAG
T N D T L G A F M R D N D F D T P V M V M C Y H G N S S K G A A Q
601
700
TATCTGCTGCAACAGGGCTACGATGTGGTCTATAGCATTGACGGCGGCTTTGAAGCCTGGCAACGTCAGTTTCCCGCAGAGGTGGCGTACGGCGCGTAAC
Y L L Q Q G Y D V V Y S I D G G F E A W Q R Q F P A E V A Y G A *
701
800
GCTTTATACTGTCCCCTTTTGTGTGGAATAAGCGACAGCAACGATGTTGATGATTACCTCTTTTGCTAACCCCCGCGTGGCGCAGGCGTTTGTTGATTAC
glpG --> M L M I T S F A N P R V A Q A F V D Y
801
900
ATGGCGACGCAGGGTGTTATCCTCACGATTCAACAACATAACCAAAGCGATGTCTGGCTGGCGGATGAGTCCCAGGCCGAGCGCGTACGGGCGGACGTGG
M A T Q G V I L T I Q Q H N Q S D V W L A D E S Q A E R V R A D V
901
1000
CGCGTTTTCTCGAAAACCCGGCAGATCCGCGTTATCTGGCGGCGAGCTGGCAGGCAGGCCATACCGGCAGTGGCCTGCATTATCGCCGTTATCCTTTCTT
A R F L E N P A D P R Y L A A S W Q A G H T G S G L H Y R R Y P F F
1001
1100
TGCCGCCTTGCGTGAACGCGCAGGTCCGGTAACCTGGGTGATGATGATCGCCTGCGTGGTGGTGTTTATTGCCATGCAAATTCTCGGCGATCAGGAAGTG
A A L R E R A G P V T W V M M I A C V V V F I A M Q I L G D Q E V
1101
1200
ATGTTATGGCTGGCCTGGCCATTCGATCCAACACTGAAATTTGAGTTCTGGCGTTACTTCACCCACGCGTTAATGCACTTCTCGCTGATGCATATCCTCT
NsiI
M L W L A W P F D P T L K F E F W R Y F T H A L M H F S L M H I L
1201
1300
TTAACCTGCTCTGGTGGTGGTATCTCGGCGGTGCGGTGGAAAAACGCCTCGGTAGCGGTAAGCTAATTGTCATTACGCTTATCAGCGCCCTGTTAAGCGG
F N L L W W W Y L G G A V E K R L G S G K L I V I T L I S A L L S G
1301
1400
CTATGTGCAGCAAAAATTCAGCGGGCCGTGGTTTGGCGGGCTTTCTGGCGTGGTGTATGCGCTGATGGGCTACGTCTGGCTACGTGGCGAACGCGATCCG
Y V Q Q K F S G P W F G G L S G V V Y A L M G Y V W L R G E R D P
1401
1500
CAAAGTGGCATTTACCTGCAACGTGGGTTAATTATCTTTGCGCTGATCTGGATTGTCGCCGGATGGTTTGATTTGTTTGGGATGTCGATGGCGAACGGAG
Q S G I Y L Q R G L I I F A L I W I V A G W F D L F G M S M A N G
1501
1600
CACACATCGCCGGGTTAGCCGTGGGTTTAGCGATGGCTTTTGTTGATTCGCTCAATGCGCGAAAACGAAAATAATTCCAGGGATTTATAAATGAAACAAA
A H I A G L A V G L A M A F V D S L N A R K R K * glpR --> M K Q
1601
1700
CACAACGTCACAACGGTATTATCGAACTGGTTAAACAGCAGGGTTATGTCAGTACCGAAGAGCTGGTAGAGCATTTCTCCGTCAGCCCGCAGACTATTCG
T Q R H N G I I E L V K Q Q G Y V S T E E L V E H F S V S P Q T I R
1701
1800
CCGCGACCTCAATGAGCTGGCGGAGCAAAACCTGATCCTGCGCCATCATGGCGGTGCGGCGCTGCCTTCCAGTTCGGTTAACACGCCGTGGCACGATCGC
R D L N E L A E Q N L I L R H H G G A A L P S S S V N T P W H D R
1801
1900
AAGGCCACCCAGACCGAAGAAAAAGAGCGCATCGCCCGCAAAGTGGCGGAGCAAATCCCCAATGGCTCGACGCTGTTTATCGATATCGGCACCACGCCGG
EcoRV
K A T Q T E E K E R I A R K V A E Q I P N G S T L F I D I G T T P
1901
2000
AAGCGGTAGCGCACGCACTGCTCAATCACAGCAATTTGCGCATTGTCACCAACAATCTCAACGTTGCTAACACGTTGATGGTAAAAGAAGATTTTCGCAT
E A V A H A L L N H S N L R I V T N N L N V A N T L M V K E D F R I
2001
2100
CATTCTCGCCGGTGGCGAATTACGCAGCCGCGATGGCGGGATCATTGGCGAAGCGACGCTCGATTTTATCTCCCAGTTCCGCCTTGATTTCGGCATTCTG
I L A G G E L R S R D G G I I G E A T L D F I S Q F R L D F G I L
2101
2200
GGGATAAGCGGCATCGATAGCGACGGCTCGCTGCTGGAGTTCGATTACCACGAAGTTCGCACCAAACGCGCCATTATTGAGAACTCGCGCCACGTTATGC
G I S G I D S D G S L L E F D Y H E V R T K R A I I E N S R H V M
2201
2300
TGGTTGTCGATCACTCGAAATTTGGCCGTAACGCGATGGTCAATATGGGCAGCATCAGCATGGTAGATGCCGTCTACACCGACGCCCCGCCGCCAGTAAG
L V V D H S K F G R N A M V N M G S I S M V D A V Y T D A P P P V S
2301
2400
CGTGATGCAGGTGCTGACGGACCACCATATTCAACTGGAGCTGTGCTGATCCTGCACGGCTTCCCACGTCAGACCAAAACGCGCCAGGTATTTGCGTAGC
V M Q V L T D H H I Q L E L C *
2401
2500
CGATCCGCGTCATTGACGCTGGCTTTGCCCTGGCGCGAAACGTCAAAAAGCTGGCGTCCGGCTGCGGAAAGCGACTTTGCCTGGCGGCAGATAGCGATAA
2501
2600
CGTGTTCCAGTTGCATACGGTCGAAGAGATCGATGTTTTCTGCCTCTGCGCCCAGCAACGCCGTAAGCGCGGAGGGGCGGCTCTCCTGCCAGTTATAGCG
MscI
NruI
2601
2700
CAGACGGTTTATCTCATCTTCAACCACGTCCAGAGTGATGCGTCCGCTAGTGGCAAAGGTGGCCATCCGCGTGACGCTGGCAGAAAGTTCGCGAAAGTTA
2701
2800
CCGCGCCATGTTGCCTGGGGAGAGGTCGCAAAAGCCAACCAGGCGCGCCGCGCTTCGGTGTTAAAACGCACGCTGTCGCCAGTGAGTGAGGCGTGGCGCT
BglII
2801
2895
CCACTTCATAATCCAGGTTCGGTTCAATATCTTCCTGGCGCTGGCGTAGACCCGGCAGGGTGAAGGTCCAGAGATTGATCCGCGCGTACAGATCT

FIG. 1. Nucleotide sequences of the *glpE*, *glpG*, and *glpR* genes and the deduced amino acid sequences. The sequence beginning at the *Eco*RI site within the 5′ end of the *glpD* gene is shown. The restriction sites referred to in the text are indicated. Sequences resembling the consensus sequences for the −10 and −35 regions and for ribosome binding are indicated by solid underlining. The transcription initiation sites for the *glpD* (56) and *glpE* (8) genes are indicated by +1. O_D1 and O_D2 are the operator sites for the *glpD* gene, and CRP is the interaction site for CRP (dashed underlining) (56). Inverted repeat sequences are signified by the converging dashed arrows below the sequence downstream of *glpR*. The amino acid sequences at the amino termini of GlpG and GlpR were determined (underlining) (Table 3). The helix-turn-helix motif of GlpR is in boldface type.

RESULTS

Nucleotide sequence of the *glpEGR* operon. To obtain detailed information about the structures of the *glpE*, *glpG*, and *glpR* genes, including their promoters and operators, the nucleotide sequence of the DNA between an upstream *Eco*RI site within the *glpD* gene (2) and a *Bgl*II site downstream of the *glpR* gene was determined (Fig. 1). This region contained 2,895 bp, including *glpE*, *glpG*, and *glpR*, which are transcribed divergently from *glpD* (41). Potential signals for transcription initiation and termination of the *glpEGR* operon are indicated in Fig. 1. The point of initiation of transcription has previously been determined by Choi et al. (8). There is a region of hyphenated dyad symmetry downstream of *glpR* that may serve as a transcription termination signal (Fig. 1). Recent results have revealed that *glpE*, *glpG*, and *glpR* are all expressed from the *glpE* promoter and are therefore in the same operon (55).

The positions of the initiation codons and verification of the reading frames for *glpE*, *glpG*, and *glpR* were determined as follows. Previous work had established that the *glpE* gene is promoter proximal and that its product is a protein of 13 kDa (40, 41). The following two possible start codons were located in the region corresponding to *glpE*, the AUG at positions 373 to 375 and the GUG at positions 436 to 438 (Fig. 1). To determine which of these initiation codons is used, the sizes of the proteins encoded by pGZ105 and pGZ106, which produce GlpE proteins by using the two potential start codons (see Materials and Methods) were determined. The apparent molecular weight of GlpE encoded by pGZ105 was 12,700 (AUG codon at 373), while that encoded by pGZ106 was only 10,200 (GUG at 436), as determined by SDS-polyacrylamide gel electrophoresis (data not shown). The first initiation codon is therefore likely to be used for translation of *glpE*. Translation initiated at position 373 would continue for 108 codons and yield a GlpE protein with a predicted molecular weight of 12,082. The reading frame in the C-terminal coding region of *glpE* was verified by construction of a translational fusion with the *lacZ* gene. A hybrid protein with β-galactosidase activity contained the GlpE portion encoded by nucleotides 373 to 690, with in-frame fusion to the *lacZ* gene. The reading frame agrees with that predicted by its nucleotide sequence. The predicted GlpE polypeptide is relatively hydrophilic (56% polar and charged residues) and acidic (calculated pI = 4.3).

An open reading frame corresponding to *glpG* was found downstream of *glpE* (Fig. 1). There are several potential initiation codons in this region, none of which is preceded by a good ribosome binding site. To determine which initiation codon is utilized, the N-terminal amino acid sequence of GlpG encoded by pGZ126 was determined (see Materials and Methods). The results (Table 3) indicate that the start codon for *glpG* is the AUG at positions 744 to 746 (Fig. 1). The reading frame in the middle of the gene (to position 1238) was confirmed by construction of a *glpG-lacZ* translational fusion, as described above for GlpE. The GlpG protein contains 276 amino acids and has a calculated molecular weight of 31,278. The apparent molecular weight, as determined by SDS-polyacrylamide gel electrophoresis of GlpG encoded by pGZ125 or pGZ126 or synthesized in minicells or maxicells, is 26,000 (40, 41). The discrepancy between the predicted and apparent mo-

TABLE 3. N-terminal amino acid sequence analysis of GlpG and GlpR[a]

Cycle no.	GlpG		GlpR	
	Amino acid identified	Amt detected (pmol)	Amino acid identified	Amt detected (pmol)
1	M	31	M	206
2	L	33	K	137
3	M	30	Q	142
4	I	27	T	129
5	T	20	Q	127
6	S	10	R	50
7	F	23	ND[b]	ND
8	A	21	ND	ND

[a] About 100 pmol of GlpG and 800 pmol of GlpR were subjected to Edman degradation to determine the N-terminal amino acid sequences.
[b] ND, not determined.

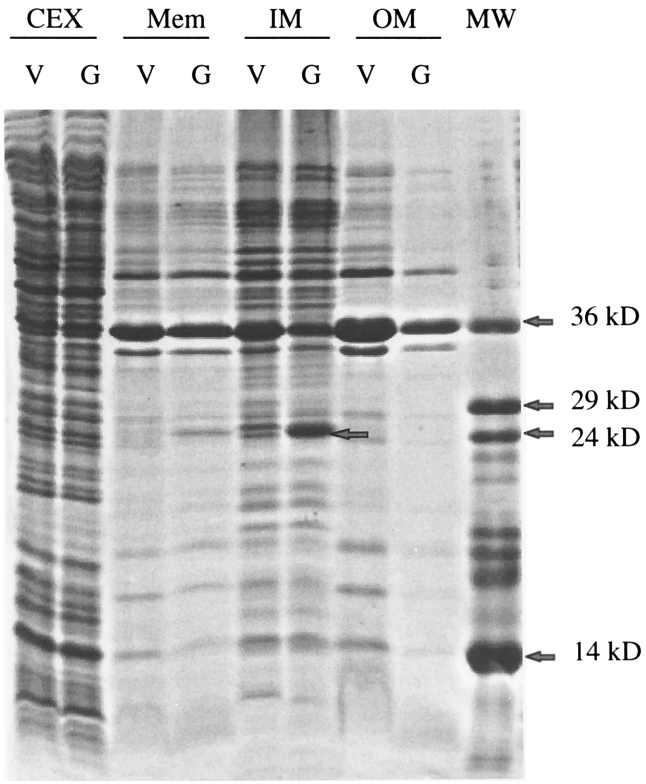

FIG. 2. Overexpression and subcellular localization of GlpG. Extracts from strain BL21(DE3) harboring vector pGEM3Z (V) or pGZ125 (G) were fractionated and analyzed on an SDS–15% polyacrylamide gel. A cytoplasmic membrane protein of about 26 kDa was overexpressed in BL21(DE3) cells carrying pGZ125 (inset arrow). The N terminus of the protein was sequenced, and the protein was identified as GlpG. Abbreviations: CEX, crude extract; Mem, total membrane fraction; IM, inner membrane fraction; OM, outer membrane fraction; MW, molecular mass standards (in kilodaltons [kD]).

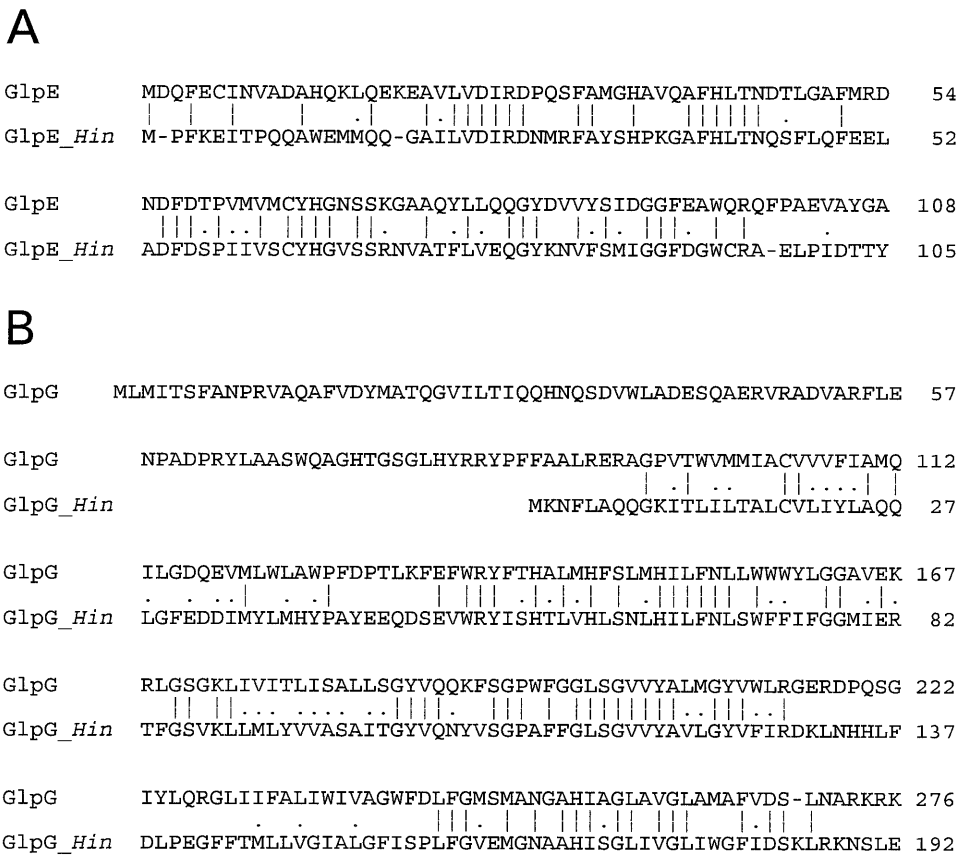

FIG. 3. Alignment of the amino acid sequences of GlpE, GlpG, and GlpR with those of related proteins. Alignment of the GlpE (A) and GlpG (B) proteins of *E. coli* and *H. influenzae* (*Hin*) and selected repressor sequences (C). The amino acid sequences were aligned by using the CLUSTAL program of PC/GENE. The position of the helix-turn-helix motif (Helix-t-Helix) is indicated at the top of panel C. Positions conserved in all sequences are indicated by asterisks, positions highly conserved are indicated by plus signs, and well conserved positions are indicated by dots. GlpR, GutR, FucR, and DeoR are the repressors of the *glp*, *gut* (54), *fuc* (30), and *deo* (49) systems of *E. coli*, respectively. R_YIHW is a hypothetical repressor protein from *E. coli* that is closely related to GlpR (36). GlpR_*Hin* is the presumptive repressor of the *glp* system of *H. influenzae* (15). LacR_*Lla* is the *lac* repressor of *L. lactis* (50). AccR is the *acc* repressor of *A. tumefaciens* (51). IolR_*Bsu* is the presumptive repressor for the inositol utilization operon of *B. subtilis* (57). Dashes indicate gaps introduced in order to maximize sequence alignment.

lecular weights of GlpG may be due to anomalous binding of SDS to this relatively hydrophobic (47% hydrophobic residues) membrane protein. GlpG protein contains at least six potential transmembrane segments, and the protein fractionates with the cytoplasmic membrane (Fig. 2). Another noteworthy feature of GlpG is its basic nature (calculated pI = 9.3).

The predicted initiation codon for *glpR* was confirmed by N-terminal sequence analysis of the purified *glp* repressor (Table 3). Translation of *glpR* is initiated at the AUG codon located 16 bp downstream from the termination codon for *glpG* (Fig. 1). The sequence that follows contains 252 codons and would be translated into a protein with a calculated molecular weight of 28,048. The predicted molecular weight agrees reasonably well with the apparent molecular weight of the purified repressor (M_r = 30,000) (25). The pI was predicted to be 5.7 and agreed well with that determined by isoelectric focusing (pI = 6.0). The reading frame in the middle of the *glpR* gene was verified by determination of the nucleotide sequence across the fusion joint of the *glpR-lacZ* translational fusion in pSH53 (41). This fusion was constructed by ligation of the 5′ end of the *glpEGR* operon (the 1,883-bp *Eco*RI-*Eco*RV fragment) to pMC1403 (7) cleaved with *Eco*RI and *Sma*I. The nucleotide sequence at the *Eco*RV-*Sma*I fusion joint was TTT ATC GAT | GGG GAT CCC, with the indicated reading frames for *glpR* and *lacZ*. The C-terminal residue of GlpR is predicted to be Cys, the sole Cys residue in the protein. Purified, reduced, and dialyzed repressors were titrated with 5,5′-dithiobis-2-(nitrobenzoic acid). The results were consistent with the prediction of one Cys per repressor monomer (data not shown). In addition, a translational fusion at this point in the predicted reading frame to a DNA segment encoding a thrombin cleavage site yielded a fusion protein that was cleavable by thrombin (58).

Comparison of GlpE, GlpG, and GlpR sequences with other protein sequences. The amino acid sequences of GlpE, GlpG, and GlpR were compared with the translated sequences in the GenBank database (3). The best matches to GlpE and GlpG were the corresponding proteins from *Haemophilus influenzae* (Fig. 3), whose entire sequence was reported recently (15). *H. influenzae* has all of the *glp* genes except for *glpD*, including *glpE*, *glpG*, and *glpR*. However, the arrangement of *glpE*, *glpG*, and *glpR* is different from that of *E. coli*. A 5′ truncated version of *glpG* is immediately upstream of *glpR*, whereas *glpE* is at another location, adjacent to *tpiA*, encoding triose phosphate isomerase (15). Similarities between GlpG or GlpE and other polypeptides in the GenBank database were significantly less compared with the examples mentioned above.

Eighteen proteins displayed significant sequence similarity with the *glp* repressor (Fig. 3C and Table 4). The *glp* repressor was most closely related to what is likely to be the *glp* repressor

C

```
                                          Helix-t-Helix
GlpR            MKQTQRHNGIIELVKQQGYVSTEELVEHFSVSPQTIRRDLNELAEQNLIL  50
GlpR_Hin        MKQSLRHQKIIKLVEQSGYLSTEELVAALDVSPQTIRRDLNILAELDLIR  50
R_YIHW     MSLTELTGNPRHDQLLMLIAERGYMNIDELANLLDVSTQTVRRDIRKLSEQGLIT  55
GutR            MKPRQRQAAILEYLQKQGKCSVEELAQYFDTTGTTIRKDLVILEHAGTVI  50
IolR_Bsu         MKLMRIQEMEEYILSHGTVSLDELCQVFNVSKNTVRRDINKLTEKGAIE  49
FucR             MKAARQQAIVDLLLNHTSLTTEALSEQLKVSKETIRRDLNELQTQGKIL  49
AccR_Atu    MVFNSTQDRQAKIVELLRDEQFLAIGRLTEHFQISVATARRDLSELHEAGLLR  53
LacR_Lla        MNKKRRLEKILDMLKIDGTITIKEIIDELDISDMTARRDLDALEADGLLT  50
DeoR          METRREERIGQLLQELKRSDKLHLKDAAALLGVSEMTIRRDLNNHSAPVVLL  52
                     *   +   .   . .   +. . . .+  *.*+*+  +   . .

GlpR       RHHGGA---ALPSSSVNTPWHDRKATQTEEKERIARKVAEQIPNGSTLFIDIGTT 102
GlpR_Hin   RHHGGA---ASPSSAENSDYVDRKQFFSLQKNNIAQEVAKLIPNGASLFIDIGTT 102
R_YIHW     RHHGGA---GRASSVVNTAFEQREVSQTEEKKAIAEAVADYIPDGSTIFITIGTT 107
GutR       RTYGGV---VLNKEESDPPIDHKTLINTHKKELIAEAAVSFIHDGDSIILDAGST 102
IolR_Bsu   KVYGGV---TSIEKTALVPFENRTIQHQDEKTKIAHYASRFIEDHDLVFIDSGTT 101
FucR       RNHGRAKYIHRQNQDSGDPFHIRLKSHYAHKADIAREALAWIEEGMVIALDASST 104
AccR_Atu   RTHGGA--VSVTQVTQDKPNAARAVWNRAEKAAIAGVVAGMIVEGDTVLLDAGTT 106
LacR_Lla   RTHGGAQ-LLSSKKPLEKTHIEKKSLNTKEKIDIAKKACSLIKDGDTIFIGPGTT 104
DeoR       ---GGYI-VLEPRSASHYLLSDQKSRLVEEKRRAAKLAATLVEPDQTLFFDCGTT 103
           +  *+                    .      .*  +*  .   +  + ..... ++*

GlpR       PEAVAHALLNHS--NLRIVTNNLNVANTLMVKEDFR--IILAGGELRSRDGGIIG 153
GlpR_Hin   PEAVANALLGHE--KLRIVTNNLNAAHLLRQNESFD--IVMAGGSLR-MDGGIIG 152
R_YIHW     VEHVARALLNHN--HLRIITNSLRVAHILYHNPRFE--VMVPGGTLRSHNSGIIG 158
GutR       VLQMVPLLSRFN--NITVMTNSLHIVNALSELDNEQT-ILMPGGTFRKKSASFHG 154
IolR_Bsu   TKSILDTLDPAK--NVTILTNSLDIINAASALKNINL--IIIGNNYKRKTRSFVG 152
FucR       CWYLAR-QLPDI--NIQVFTNSHPICHELGKRERIQL--ISSGGTLERKYGCYVN 154
AccR_Atu   ALEVAKKLADRR--NLTFISNGLDIVEELTRGEG-KS-IYSVGGEYTETNRSFRG 157
LacR_Lla   LVQLALELKGRKGYKIRVITNSLPVFLIL--NDSETIDLLLLGGEYREITGAFVG 157
DeoR       TPWIIEAIDNEIPF--TAVCYSLNTFLAL--KEKPHCRAFLCGGEFHASNAIFKP 154
              .+  .       . ..++.+ ..  +             *+...     . +

GlpR       EATLDFISQFRLDFGILGISGIDSDGSLL-EFDYHEVRTKRAIIENSRHVMLVVD 207
GlpR_Hin   EATVNFISQFRLDFGILGISAIDADGSLL-DYDYHEVQVKRAIIESSRQTLLVAD 206
R_YIHW     PSAASFVADFRADYLVTSVGAIESDGALM-EFDVNEANVVKTMMAHARNILLVAD 212
GutR       QLAENAFEHFTFDKLFMGTDGIDLNAGVT-TFNEV-YTVSKAMCNAAREVILMAD 207
IolR_Bsu   MDDPAMLDKYNINKAFMSATGTTLTHGLT-NSDLLEYEIKKRISEKAKEVYLLAD 206
FucR       PSLISQLKSLEIDLFIFSCEGIDSSGALW-DSNAINADYKSMLLKRAAQSLLLID 208
AccR_Atu   PLAEQFIRQFNVDKLILNAASIDVDRGLICTSSPVNASVARAMIEVSSRVIVVAD 212
LacR_Lla   SMASTNLKAMRFAKAFVRANAVTHNSIA--TYSDKEGVIQQLALNNAVEKFLLVD 210
DeoR       IDFQQTLNNFCPDIAFYSAAGVHVSKGATC-FNLEELPVKHWAMSMAQKHVLVVD 208
                 .     .       ...  .        .              .     . *

GlpR       HSKFGRNAMVNMGSISMVDAVYTDAPPPVSVMQVLTDHHIQLELC           252
GlpR_Hin   HSKFTRQAIVRLGELSDVEYLFTGDVPE-GIVNYLKEQKTKLVLCNGKVR      255
R_YIHW     HTKYHASAAVEIGNVAQVTALFTDELPPAALKSRLQDSQIEIILPQEDA       261
GutR       SSKFGRKSPNVVCSLESVDKLITDAGIDPAFRQALEEKGIDVIITGESNE      257
IolR_Bsu   HSKFGKSTLLTYAPFDRLHCIVTSQPLDDEYTQYCNEHQIGIHLA           251
FucR       KSKFNRSGEARIGHLDEVTHIISDERQVATSLVTA                     243
AccR_Atu   HSKFTKSSLSVTARIEDVGVIVTDSGTRTIIETIPEKLRKKFVVAN         258
LacR_Lla   STKFDRYDFFNFYNLDQLDTIITDNQISPQHLEEFSQYTTILKAD           255
DeoR       HSKFGKVRPARMGDLKRFDIVVSDCCPEDEYVKYAQTQRIKLMY            252
            +*+ .         .  .  ...+
```

FIG. 3—*Continued.*

of *H. influenzae* (15) and to the hypothetical transcriptional regulatory protein YIHW of *E. coli* (36). The *glp* repressor of *E. coli* was also closely related to the *gut* (54), *fuc* (30), *gat* (21), *aga* (37), and *deo* (49) repressors of *E. coli* as well as the *lac* repressors of the gram-positive bacteria *Lactococcus lactis* (50), *Staphylococcus aureus* (34), and *Streptococcus mutans* (38). The amino acid sequence of the *glp* repressor was also similar to that of the *iol* repressor of *Bacillus subtilis* (57) and the *acc* repressor of *Agrobacterium tumefaciens* (51). The sizes of these proteins are similar, ranging from 242 to 257 amino acids. One of the most highly conserved regions in this family of repressors is found in the amino-terminal region (Fig. 3C). This region is likely to be the helix-turn-helix DNA-binding motif, based on the sequence similarities found in this region of the *gut*, *gal*, *lac*, and other repressors. Thus, the helix-turn-helix motif of the *glp* repressor was proposed to lie between residues 22 and 41 (Fig. 1 and 3C).

Identification of the DNA recognition helix of the *glp* repressor. The second helix of the helix-turn-helix motif, the so-called recognition helix, is strikingly well conserved among the repressors listed in Fig. 3C. The first two residues of the recognition helix differ among the repressors and are likely to be critical for DNA sequence binding specificity (26). Of the repressors listed in Fig. 3C, the corresponding operator sequences are known for the *glp* and *deo* repressors (16, 60). The two operator sequences are very similar, with the only major difference being the base pair at position 4 (with respect to the center of operator symmetry), where a C-G base pair is present

TABLE 4. Pairwise comparison of GlpR with related transcriptional regulators[a]

Protein	Identity (%)										
	GlpR_*Hin*	R_YIHW	Accr	GutR	FucR	FucR_*Hin*	DeoR	LacR_*Lla*	LacR_*Smu*	LacR_*Sau*	IolR_*Bsu*
GlpR	61	44	32	29	28	25	26	27	29	30	26
GlpR_*Hin*		39	31	24	28	26	23	24	24	25	21
R_YIHW			26	26	28	22	21	28	30	31	25
AccR				31	28	29	19	26	24	24	26
GutR					27	31	22	29	26	23	30
FucR						34	15	26	24	21	26
FucR_*Hin*							21	26	28	28	29
DeoR								25	25	24	24
LacR_*Lla*									42	41	26
LacR_*Smu*										63	26
LacR_*Sau*											26

[a] Sequences were aligned by using the PALIGN program of PC/GENE. Abbreviations are as defined in the legend to Fig. 3. In addition, the FucR protein from *H. influenzae* (FucR_*Hin*) and the LacR proteins from *Streptococcus mutans* (LacR_*Smu*) and *Staphylococcus aureus* (LacR_*Sau*) are compared.

in the *glp* consensus operator and an A-T base pair is present in the *deo* consensus operator. It is likely that the amino acyl residues present at positions 1 and 2 of the recognition helices play a role in the recognition of the operator base pair at position 4.

To provide evidence for a role of amino acyl residues 1 and 2 in discriminating between the *glp* and *deo* operators and furthermore to prove that this region of the *glp* repressor is in fact the recognition helix, altered forms of the *glp* repressor were generated by directed mutagenesis. In the first construct, residues 1 and 2 of the presumptive recognition helix of the *glp* repressor were changed so that the entire recognition helix would be identical to that of the *deo* repressor (P33E and Q34M). It was anticipated that this altered form of *glp* repressor might change its recognition specificity to that of the *deo* repressor. Changes in specificity have been generated in the case of several DNA-binding proteins, including the *lac* repressor (20), bacteriophage lambda and 434 repressors (18, 53), and CRP and FNR proteins (42). In another construct, which served as a negative control, residue 6 of the recognition helix of the *glp* repressor was changed from arginine to alanine (R38A). It was anticipated that this change would greatly decrease the affinity of the *glp* repressor for its operator. Arginine is highly conserved at this position in the DeoR family of transcriptional regulators. Arginine is also present at the corresponding position of the helix-turn-helix motif of the *lac* repressor, is critical for high-affinity binding to the *lac* operator, and is thought to interact specifically with the G-C base pair of the TGT operator motif (20).

The functions of the wild-type and variant repressors were tested both in vivo and in vitro. For in vivo experiments, the corresponding genes were introduced into pSY223, in which the *glpEGR* operon was placed downstream of the *tac* promoter. The resulting plasmids were introduced into strains SY102 (*glpR glpD-lacZ*) and KH682 (*deoR deoC-lacZ*). The efficacies of the repressor variants were assessed by measuring β-galactosidase activities in cells growing in minimal medium with or without IPTG, the inducer of the *tac* promoter (Table 5). Induction of the synthesis of the wild-type *glp* repressor resulted in a 10-fold increase in repression of the *glpD-lacZ* fusion (from 46 to 550; Table 5). Neither of the *glp* repressor variants exerted significant repression on either fusion, with or without induction by IPTG (Table 5). These results suggest that neither the P33E and Q34M nor R38A repressor variant was able to bind either the *glp* or *deo* operator with high affinity in vivo.

To rule out the possibility that lack of repression by the repressor variants is due to instability or defects in oligomerization, the ability of the repressor variants to interfere with normal repressor function was tested. Both the P33E and Q34M repressor variant and R38A repressor variant were found to be negatively dominant to the wild-type repressor (Table 6), which indicates that these forms of the repressor are stable and are able to oligomerize with the normal repressor, thereby interfering with the ability of the normal repressor to bind the operator. Further evidence in support of the conclusion that the above amino acid substitutions affected only the DNA-binding function and not the overall structure of the repressor was obtained by measuring the specific activity of inducer ([^{3}H]glycerol-P) binding in crude extracts of IPTG-induced strains. The specific activities for the variant repressors were similar to those obtained for the wild-type repressor (values ranged from 80 to 120 pmol of [^{3}H]glycerol-P bound per mg of protein).

For in vitro experiments, the repressor variants were overexpressed in strain BL21(DE3). The relative amount of repressor produced by each strain was determined by SDS-polyacrylamide gel electrophoresis (Fig. 4). Although expressed at high levels, each repressor variant was soluble and in a native conformation, as assessed by inducer-binding assays (data not shown). To compare the affinities of wild-type and mutant repressors for both *glp* and *deo* operators, gel mobility shift assays were performed. As shown in Fig. 5, only the wild-type *glp* repressor bound the *glp* operator when approximately 100 nM repressor tetramer was used. Neither P33E and Q34M nor R38A altered repressor bound either the *glp* or *deo* operator. These findings agree with those obtained in vivo for repression of the *glpD-lacZ* or *deoC-lacZ* fusion. These results, along with the observation that the repressor variants exerted a negative-dominant phenotype, provide the first published experimental data indicating that the proposed region of the *glp-deo* repressor family is indeed the helix-turn-helix DNA-binding motif.

DISCUSSION

Here we have reported the nucleotide sequence of the *glpEGR* operon of *E. coli*. Three open reading frames corresponding to *glpE*, *glpG*, and *glpR* have been identified and verified. The *glpR* gene encodes the repressor for the glycerol-P regulon. The *glpE* and *glpG* genes are promoter proximal to *glpR* in the same operon. The biochemical functions of GlpE and GlpG are unknown.

The *glp* repressor is a member of the DeoR family of transcriptional regulators, which at present contains 19 members

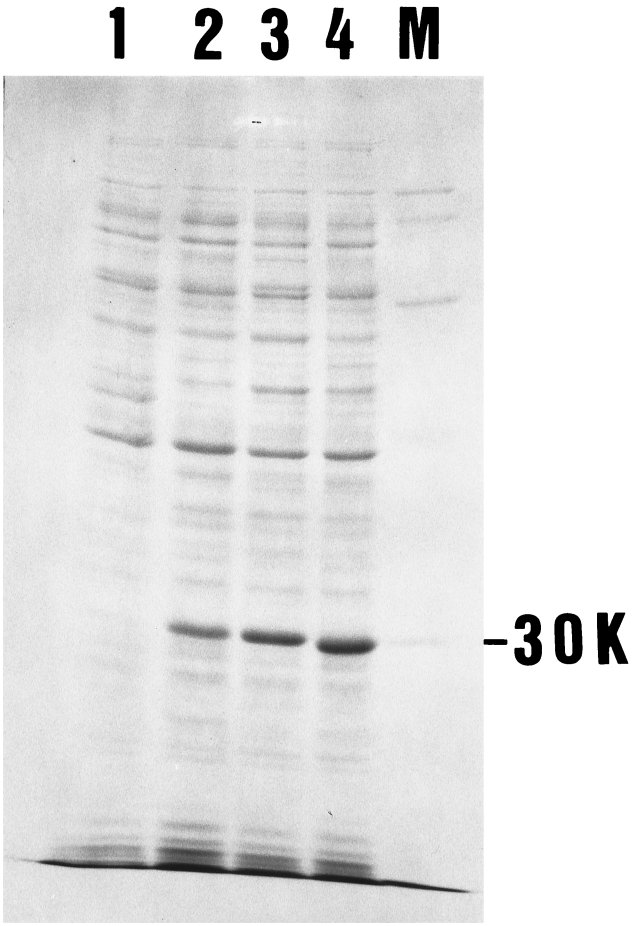

FIG. 4. Overexpression of the *glp* repressor and its variants. Induction of expression was carried out as described for GlpE in Materials and Methods. Cells were lysed by using a French pressure cell, and crude extracts were prepared (25). The same amount of protein (25 μg), as determined by the method of Bradford with bovine serum albumin as the standard (4), was loaded in each lane. Lanes 1 to 4 contained extracts from strains BL21(DE3) with pBluescript, wild-type GlpR encoded by pSH79, $GlpR^{P33E,Q34M}$ encoded by pSY2-A, and $GlpR^{R38A}$ encoded by pSY2-9, respectively. Lane M, molecular mass standards (in kilodaltons [K]).

on the basis of similarity searches of the databases. The members of this protein family are present in a variety of bacterial organisms. *E. coli* contains 10 members, 6 of which have defined functions (Table 4). The genome of *H. influenzae* encodes three family members, including those corresponding to GlpR and FucR of *E. coli* (15). Family members are also found in gram-positive organisms and include the LacR repressors from lactose-utilizing bacteria and the inositol repressor IolR from *B. subtilis*. Finally, one member (PIR locus S48604) is from *Mycoplasma capricolum*, an organism with a relatively small, AT-rich genome.

Features common to the members of the DeoR family have been pointed out by van Rooijen and de Vos (50) and by Reizer et al. (37). All of these proteins contain about 250 amino acid residues. In many cases, the effector molecule that interacts with the transcriptional regulator in order to achieve induction is a phosphorylated sugar (DeoR, GlpR, LacR, and AccR). The proteins contain several highly conserved regions, one of which is the second helix of the helix-turn-helix motif. Other conserved regions distal to the DNA-binding motif may be involved in oligomerization or in binding of the inducer.

TABLE 5. Repression in vivo of Φ(*glpD-lacZ*) and Φ(*deoC-lacZ*) by the wild-type *glp* repressor and *glp* repressor variants

Strain and repressor encoded by plasmid	β-Galactosidase sp act		Repression[a]	
	−IPTG	+IPTG[b]	−IPTG	+IPTG
SY102 [*glpR2* Φ(*glpD-lacZ*)]				
None[c]	5,500	5,500		
Wild type	120	10	46	550
P33E and Q34M	3,500	1,600	2	3
R38A	5,700	2,900	1	2
KH682 [*deoR* Φ(*deoC-lacZ*)]				
None[d]	240	290		
Wild type	250	220	1	1.3
P33E and Q34M	240	130	1	2.3
R38A	260	140	1	2.1
DeoR[e]	8	ND[f]	30	ND

[a] Repression was defined as the specific activity of β-galactosidase in the absence of repressor divided by the specific activity of β-galactosidase in the presence of the *glp* repressor or a variant.
[b] IPTG (0.2 mM) was added (+) to the maltose minimal medium as indicated.
[c] β-Galactosidase activity was determined for strain SY102(pSY223).
[d] β-Galactosidase activity was determined for strain KH682(pSY223).
[e] β-Galactosidase activity was determined for strain KH682(pKH31).
[f] ND, not determined.

The DeoR and GlpR proteins have been purified and characterized in some detail (1, 25, 33). Both DeoR and GlpR form higher-order oligomers under native conditions (octamers and tetramers, respectively) and achieve repression by simultaneous binding of repressor to widely separated operator sites, with the intervening DNA forming a loop (1, 11, 12, 16, 23).

Among the repressors, GlpR and DeoR were found to have striking similarities in the recognition helices and corresponding operator sites (16, 60). The first two residues of the recognition helix are not conserved among the DeoR family members and are therefore thought to confer DNA sequence specificity for repressor binding to the operator. Therefore, it was surprising that the operator binding specificity of GlpR was not changed to that of DeoR when these two residues of GlpR were changed to match those of DeoR. The P33E and Q34M form of the *glp* repressor did not bind either the *glp* or *deo* operator in vivo or in vitro. The same results were obtained with a Q34M singly substituted variant and with a E23K, P33E, and Q34M triply substituted variant of the *glp* repressor (59). These results suggest that not only is the recognition helix important for determining binding specificity for the DeoR family members but that the supporting structure also must play a role in DNA sequence binding specificity. There is very little sequence conservation within the first helix of the helix-

TABLE 6. Negative dominance of the P33E and Q34M *glp* repressor and the R38A *glp* repressor

Plasmid (repressor variant)	β-Galactosidase sp act[a]	
	−Glycerol-P	+Glycerol-P
pSY223 (None)	7.4	75
pSY2-C (wild type)	0.5	15
pSY2-IX (P33E and Q34M)	1,020	1,120
pSY2-A9 (R38A)	380	1,000

[a] Strains constructed by transformation of GD31 [$glpR^+$ Φ(*glpK-lacZ*)] with the indicated plasmids were grown on maltose minimal medium containing 0.25 mM IPTG and 1 mM glycerol-P (+) as indicated. Cells were harvested during log phase, and the β-galactosidase specific activity was determined.

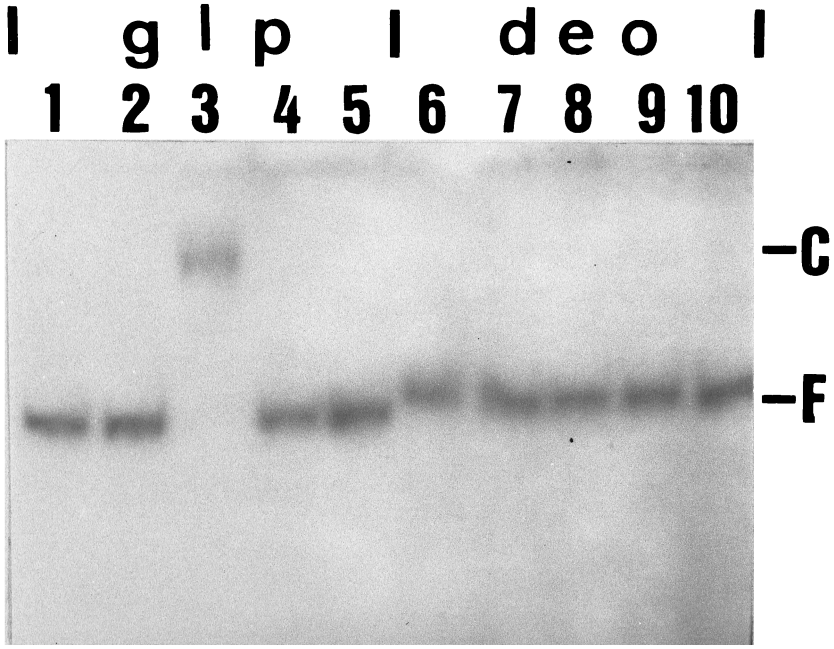

FIG. 5. Binding of the *glp* repressor and its variants to operator DNA. Gel mobility shift assays were performed using radiolabeled DNA fragments with the *glp* (lanes 1 through 5) or *deo* (lanes 6 through 10) operator. The reaction mixtures loaded in lanes 1 and 6 contained only DNA. The reaction mixtures loaded in lanes 2 and 7 contained crude extract from strain BL21(DE3)(pBluescript). The reaction mixtures loaded in lanes 3 and 8 were the wild-type *glp* repressor from strain BL21(DE3)(pSH79); lanes 4 and 9 contained $GlpR^{P33E,Q34M}$ from strain BL21(DE3)(pSY2-A); and lanes 5 and 10 contained $GlpR^{R38A}$ from strain BL21(DE3)(pSY2-9). The concentration of each repressor tetramer used was about 100 nM. Abbreviations: F, free DNA; C, repressor-DNA complex.

turn-helix DNA-binding motif; therefore, the different amino acid sequences could confer some sequence specificity to DNA binding and/or could specifically support different recognition helices. The aliphatic residues in this region that are conserved among family members are likely to be buried within the repressor structure and thus would not be expected to contribute directly to sequence-specific binding of DNA. The inability of the *glp* repressor variants to bind the *deo* operator may be due to loss of a stabilizing interaction that occurs between the two helices of the normal *glp* repressor. Loss of such an interaction could result in an alteration of the angle between the two helices and an inability to bind the *deo* operator. For example, a critical hydrogen bond may exist between Q-34 and one of the glutamic acid residues of the first helix of the *glp* repressor. The loss of such a bond may destabilize the helix-turn-helix DNA-binding motif. Such a hydrogen bond occurs between a glutamine at the start of the recognition helix and a glutamine in the first helix of lambda and 434 repressors (35). An alternate explanation for the apparent lack of binding of GlpR variants to the *deo* operator is that the operator arrangements for the *glp* and *deo* systems are fundamentally different. In the *deo* system, the repressor binds to multiple operators that are widely separated (12). In contrast, the *glp* repressor binds with highest affinity to adjacent, tandemly arranged operators (60). However, we found that the P33E and Q34M *glp* repressor variant was unable to bind tandemly arranged *deo* operators in vitro (data not shown).

Generation of a GlpR variant with altered DNA sequence binding specificity would provide direct information concerning amino acid residues that contact the operator. Although we were unable to generate a *glp* repressor variant with *deo* operator binding specificity, the fact that all amino acid substitutions within the second helix of the proposed helix-turn-helix motif conferred a negative-dominant phenotype, along with the observed high degree of sequence similarity of this region to the helix-turn-helix motif of other transcriptional regulators, provides the first experimental evidence for the assignment of this region of GlpR and other family members as the helix-turn-helix DNA-binding motif. Details of the interactions occurring between the DNA-binding domain and the operator await further structural characterization.

ACKNOWLEDGMENTS

G. Zeng and S. Ye contributed equally to this work, and both should be considered first authors.

This research was supported by National Science Foundation grants DMB-8819128 and MCB-9118757.

We thank Ali T. van Loo-Bhattacharya for skillful technical assistance and T. Denise Clark for construction of pDA223. Karin Hammer, Gaye Sweet, and Herbert Schweizer provided strains for this work. N-terminal amino acid sequence analysis was done in the laboratory of Charles L. Rutherford, Department of Biology, Virginia Polytechnic Institute and State University.

REFERENCES

1. **Amouyal, M., L. Mortensen, H. Buc, and K. Hammer.** 1989. Single and double loop formation when *deoR* repressor binds to its natural operator. Cell **58:**545–551.
2. **Austin, D., and T. J. Larson.** 1991. Nucleotide sequence of the *glpD* gene encoding aerobic *sn*-glycerol 3-phosphate dehydrogenase of *Escherichia coli* K-12. J. Bacteriol. **173:**101–107.
3. **Benson, D. A., M. Boguski, D. J. Lipman, and J. Ostell.** 1996. GenBank. Nucleic Acids Res. **24:**1–5.
4. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
5. **Brosius, J., and A. Holy.** 1984. Regulation of ribosomal RNA promoters with a synthetic *lac* operator. Proc. Natl. Acad. Sci. USA **81:**6929–6933.
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. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detec-

tion and cloning of translational initiation signals. J. Bacteriol. **143:**971–980.
8. **Choi, Y. L., S. Kawase, M. Kawamukai, H. Sakai, and T. Komano.** 1991. Regulation of *glpD* and *glpE* gene expression by a cyclic AMP-cAMP receptor protein (cAMP-CRP) complex in *Escherichia coli*. Biochim. Biophys. Acta **1088:**31–35.
9. **Clark, D. J., and O. Maaløe.** 1967. DNA replication and the division cycle in *Escherichia coli*. J. Mol. Biol. **23:**99–112.
10. **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* K-12. J. Bacteriol. **170:**2448–2456.
11. **Dandanell, G.** 1992. DeoR repression at-a-distance only weakly responds to changes in interoperator separation and DNA topology. Nucleic Acids Res. **20:**5407–5412.
12. **Dandanell, G., and K. Hammer.** 1985. Two operator sites separated by 599 base pairs are required for *deoR* repression of the *deo* operon of *Escherichia coli*. EMBO J. **4:**3333–3338.
13. **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.
14. **Eiglmeier, K., W. Boos, and S. T. Cole.** 1987. Nucleotide sequence and transcriptional startpoint of the *glpT* gene of *Escherichia coli*: extensive sequence homology of the glycerol-3-phosphate transport protein with components of the hexose-6-phosphate transport system. Mol. Microbiol. **1:**251–258.
15. **Fleischmann, R. D., M. D. Adams, et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
16. **Hammer, K., L. Bech, P. Hobolth, and G. Dandanell.** 1993. DNA specificity of *Escherichia coli deoP1* operator-DeoR repressor recognition. Mol. Gen. Genet. **237:**129–133.
17. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
18. **Hochschild, A., J. Douhan, and M. Ptashne.** 1986. How lambda repressor and lambda Cro distinguish between O_R1 and O_R3. Cell **47:**807–816.
19. **Iuchi, S., and E. C. C. Lin.** 1993. Adaptation of *Escherichia coli* to redox environments by gene expression. Mol. Microbiol. **9:**9–15.
20. **Kisters-Woike, B., N. Lehming, J. Sartorius, B. von Wilcken-Bergmann, and B. Müller-Hill.** 1991. A model of the *lac* repressor-operator complex based on physical and genetic data. Eur. J. Biochem. **198:**411–419.
21. **Komoda, Y., M. Enomoto, and A. Tominaga.** 1991. Large inversion in *Escherichia coli* K-12 1485IN between inversely oriented IS3 elements near *lac* and *cdd*. Genetics **129:**639–645.
22. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154:**367–382.
23. **Larson, T. J., J. S. Cantwell, and A. T. van Loo-Bhattacharya.** 1992. Interaction at a distance between multiple operators controls the adjacent, divergently transcribed *glpTQ-glpACB* operons of *Escherichia coli* K-12. J. Biol. Chem. **267:**6114–6121.
24. **Larson, T. J., G. Schumacher, and W. Boos.** 1982. Identification of the *glpT*-encoded *sn*-glycerol-3-phosphate permease of *Escherichia coli*, an oligomeric integral membrane protein. J. Bacteriol. **152:**1008–1021.
25. **Larson, T. J., S. Ye, D. L. Weissenborn, H. J. Hoffmann, and H. Schweizer.** 1987. Purification and characterization of the repressor for the *sn*-glycerol 3-phosphate regulon of *Escherichia coli* K-12. J. Biol. Chem. **262:**15869–15874.
26. **Lehming, N., J. Sartorius, B. Kisters-Woike, B. von Wilcken-Bergmann, and B. Müller-Hill.** 1990. Mutant *lac* repressors with new specificities hint at rules for protein-DNA recognition. EMBO J. **9:**615–621.
27. **Lin, E. C. C.** 1976. Glycerol dissimilation and its regulation in bacteria. Annu. Rev. Microbiol. **30:**535–578.
28. **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. American Society for Microbiology, Washington, D.C.
29. **Liss, L. R.** 1987. New M13 host. Focus **9:**13.
30. **Lu, Z., and E. C. C. Lin.** 1989. The nucleotide sequence of *Escherichia coli* genes for L-fucose dissimilation. Nucleic Acids Res. **17:**4883–4884.
31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Mortensen, L., G. Dandanell, and K. Hammer.** 1989. Purification and characterization of the *deoR* repressor of *Escherichia coli*. EMBO J. **8:**325–331.
34. **Oskouian, B., and G. C. Stewart.** 1990. Repression and catabolite repression of the lactose operon of *Staphylococcus aureus*. J. Bacteriol. **172:**3804–3812.
35. **Pabo, C. O., A. K. Aggarwal, S. R. Jordan, L. J. Beamer, U. R. Obeysekare, and S. C. Harrison.** 1990. Conserved residues make similar contacts in two repressor-operator complexes. Science **247:**1210–1213.
36. **Plunkett, G. I., V. Burland, D. L. Daniels, and F. R. Blattner.** 1993. Analysis of the *Escherichia coli* genome. III. DNA sequence of the region from 87.2 to 89.2 minutes. Nucleic Acids Res. **21:**3391–3398.
37. **Reizer, J., T. M. Ramseier, A. Reizer, A. Charbit, and M. H. Saier.** 1996. Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative *N*-acetylgalactosamine metabolic pathway in *Escherichia coli*. Microbiology **142:**231–250.
38. **Rosey, E. L., and G. C. Stewart.** 1992. Nucleotide and deduced amino acid sequences of the *lacR*, *lacABCD*, and *lacFE* genes encoding the repressor, tagatose 6-phosphate gene cluster, and sugar-specific phosphotransferase system components of the lactose operon of *Streptococcus mutans*. J. Bacteriol. **174:**6159–6170.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
40. **Schweizer, H., W. Boos, and T. J. Larson.** 1985. Repressor for the *sn*-glycerol-3-phosphate regulon of *Escherichia coli* K-12: cloning of the *glpR* gene and identification of its product. J. Bacteriol. **161:**563–566.
41. **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 structural genes of the *glp*-regulon. Mol. Gen. Genet. **202:**488–492.
42. **Spiro, S., K. L. Gaston, A. I. Bell, R. E. Roberts, S. J. W. Busby, and J. R. Guest.** 1990. Interconversion of the DNA-binding specificities of two related transcription regulators, CRP and FNR. Mol. Microbiol. **4:**1831–1838.
43. **Stark, M. J. R.** 1987. Multicopy expression vectors carrying the *lac* repressor gene for regulated high level expression of genes in *Escherichia coli*. Gene **51:**255–267.
44. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
45. **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.
46. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
47. **Tommassen, J., K. Eiglmeier, S. T. Cole, P. Overduin, T. J. Larson, and W. Boos.** 1991. Characterization of two genes, *glpQ* and *ugpQ*, encoding glycerophosphoryl diester phosphodiesterases of *Escherichia coli*. Mol. Gen. Genet. **226:**321–327.
48. **Truniger, V., W. Boos, and G. Sweet.** 1992. Molecular analysis of the *glpFKX* regions of *Escherichia coli* and *Shigella flexneri*. J. Bacteriol. **174:**6981–6991.
49. **Valentin-Hansen, P., P. Hojrup, and S. Short.** 1985. The primary structure of the DeoR repressor from *Escherichia coli* K-12. Nucleic Acids Res. **13:**5927–5936.
50. **van Rooijen, R. J., and W. M. de Vos.** 1990. Molecular cloning, transcriptional analysis, and nucleotide sequence of *lacR*, a gene encoding the repressor of the lactose phosphotransferase system of *Lactococcus lactis*. J. Biol. Chem. **265:**18499–18503.
51. **von Bodman, S. B., G. T. Hayman, and S. T. Farrand.** 1992. Opine catabolism and conjugal transfer of the nopaline Ti plasmid pTiC58 are coordinated by a single repressor. Proc. Natl. Acad. Sci. USA **89:**643–647.
52. **Weissenborn, D. L., N. Wittekindt, and T. J. Larson.** 1992. Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. J. Biol. Chem. **267:**6122–6131.
53. **Wharton, R. P., and M. Ptashne.** 1987. A new-specificity mutant of 434 repressor that defines an amino acid-base pair contact. Nature (London) **326:**888–891.
54. **Yamada, M., and M. H. J. Saier.** 1988. Positive and negative regulators for glucitol (*gut*) operon expression in *Escherichia coli*. J. Mol. Biol. **203:**569–583.
55. **Yang, B., T. D. Clark, N. Zhao, and T. J. Larson.** Unpublished data.
56. **Ye, S. Z., 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.
57. **Yoshida, K., S. Seki, M. Fujimura, Y. Miwa, and Y. Fujita.** 1995. Cloning and sequencing of a 36-kb region of the *Bacillus subtilis* genome between the *gnt* and *iol* operons. DNA Res. **2:**61–69.
58. **Zeng, G., and T. J. Larson.** 1996. Use of engineered thrombin cleavage site for determination of translational reading frames. BioTechniques **21:**822–824.
59. **Zhao, N., and T. J. Larson.** Unpublished data.
60. **Zhao, N., W. Oh, D. Trybul, K. S. Thrasher, T. J. Kingsbury, and T. J. Larson.** 1994. Characterization of the interaction of the *glp* repressor of *Escherichia coli* with single and tandem *glp* operator variants. J. Bacteriol. **176:**2393–2397.