CHARACTERIZATION OF THE *GLPEGR* OPERON OF *ESCHERICHIA COLI* K-12

by

Gang Zeng

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

APPROVED:

Timothy J. Larson, Chairman

Dennis R. Dean

Peter J. Kennelly

Walter G. Niehaus Jr.

Charles L. Rutherford

February, 1996

Blacksburg, Virginia

CHARACTERIZATION OF THE *GLPEGR* OPERON OF *ESCHERICHIA COLI* K-12

by

Gang Zeng

Timothy J. Larson, Chairman

Biochemistry

(ABSTRACT)

In *E. coli*, three open reading frames (ORF) are present in the *glpEGR* operon. The *glpR*-encoded repressor governs the universal negative regulation of the *glp* regulon. The *glpE* and *glpG* genes are located upstream of *glpR*. The initiation codons as well as the translational reading frames used by these two genes were determined. Both proteins were successfully overexpressed and GlpE was purified in one step using immobilized metal affinity chromatography. The translational reading frame of *glpR* in its C-terminal coding region was verified, and confirms the nucleotide sequence reported by other workers. The subcellular locations of GlpE and GlpG were determined, GlpE was in the cytoplasmic fraction and is predicted to be an acidic protein with a molecular weight of 12,082; GlpG was in the cytoplasmic membrane fraction and is predicted to be a basic, hydrophobic protein with a molecular weight of 31,278. *glp* repressor was also overexpressed and purified. The apparent dissociation constant of the *glp* repressor-operator complex was estimated to be 50 nM, which indicates that GlpR has a relatively low affinity for a single *glp* operator site. Isoelectric focusing gel electrophoresis suggested that GlpR exists as three differentially-charged forms *in vitro*. The major form of GlpR had an isoelectric point

of 5.99, and the two minor forms had pI's of 5.92 and 6.05, respectively. The presence and relative amounts of these forms, however, was not dependent on GlpE/GlpG.

The effect of GlpE/GlpG together with GlpR on the regulation of *glp* operons was studied using an *in vivo* system that employed compatible expression vectors for GlpE/GlpG and GlpR. Comparison of the patterns of regulation of *glpK*, *glpD* and *glpT* by the *glp* repressor agreed well with results previously reported on differential repression of these operons. Under the conditions employed, the *glp* repressor conferred more than 90-fold repression on *glpK*. Interestingly, in the presence of GlpE and GlpG, the repression ratio for *glpK* increased to near 170-fold. In contrast, *glpD* was subject to 23-fold repression by GlpR. The presence of GlpE and GlpG, however, decreased this ratio to 12-fold. The *glpTQ* operon was subject to relatively weak repression, about 9- to 12-fold, and repression was independent of the *glpE/glpG* gene products. This is the first time that the effect of the GlpE/GlpG proteins on the regulation of individual *glp* operons has been examined. The results suggest that GlpE/GlpG exert effects that bring about differential regulation of *glpK*, *glpD*, and *glpT*. Since the precise biochemical functions of GlpE and GlpG remain undetermined, the molecular basis for this effect is still unknown.

This work is dedicated to my father, my mother, and my sisters. Without their sacrifices and love, the completion of this research would not have been possible.

ACKNOWLEDGMENTS

My deepest appreciation goes to my mentor, Dr. Timothy J. Larson for his guidance, support, and consideration throughout the course of my graduate study. I especially thank him for introducing me to a scientific career with his unique ways of encouragement and generosity. I admire his extraordinary memory, as well as his talent for explaining scientific data. His numerous ideas and his warm-hearted personality will continuously benefit me in my future life. I thank members of my committee for the scientific philosophies I have learned from them. I want to thank Dr. Dean for his idea of establishing models at all stages of research; I appreciate Dr. Kennelly for the many discussions I had with him, and also his emphasis about the importance of control experiments; I am grateful to Dr. Niehaus for his perceptiveness of linking biochemistry with cell physiology; I also want to express my appreciation to Dr. Rutherford for his suggestion of using antibodies and inhibitors as general biochemical approaches.

Many thanks go to previous and current lab coworkers, Won, Ningyue, Steve, Sherry, Sergey, Tsuneo, Jason, and Andy for their friendship and the time I have spent with them. In particular, I want to express my appreciation to Ali for her patience, warm-heartedness, and many unselfish efforts for making the lab a happy family.

Dr. Jie Leng helped me with protein purification. Dr. Ken Mitton showed me many useful resources on the Internet. Mr. Youlin Tang taught me the basics of IEF. I also want to thank Dr. John Hess for allowing me to use his minigel apparatus. The draft of this dissertation has been proofread by Steve Solow, Tsuneo Hill, and Jason Dietrich.

Most of all, I would like to thank my wife Shubing Zou for her love, understanding, and spiritual support, especially during the difficulties with my research.

LIST OF ABBREVIATIONS

Ap	ampcilin
cAMP	cyclic adenosine 3',5'-monophosphate
Cam	chloramphenicol
CRP	cAMP receptor protein, catabolite gene activator protein
DHAP	dihydroxyacetone phosphate
DTT	dithiothreitol
FAD	flavin adenine dinucleotide
FADH	reduced flavin adenine dinucleotide
FMN	flavin mononucleotide
FNR	anaerobic transcriptional activator
GAP	glyceraldehyde-3-phosphate
Glycerol-P	glycerol 3-phosphate
IMAC	immobilized metal affinity chromatography
IPTG	isopropylthio-β-D-galactopyranoside
K_d	dissociation constant
Km	kanamycin
LB	Luria Broth
MW	molecular weight
OD	optical density
ONPG	*o*-nitrophenyl β-D-galactopyranoside
PAGE	polyacrylamide gel electrophoresis
PNPP	*p*-nitrophenyl phosphate

pfu	plaque forming units
psi	pounds per square inch
rbs	ribosome binding site
SDS	sodium dodecyl sulfate
Tc	tetracycline
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LITERATURE REVIEW

METABOLISM OF GLYCEROL 3-PHOSPHATE IN *ESCHERICHIA COLI*

sn-Glycerol 3-phosphate (glycerol-P) is a ubiquitous compound in nature. Its esterified products, such as phosphatidylglycerol (PG), phosphatidylethanolamine (PE), and cardiolipin (CL) are the major lipid components of all biological membranes (1). Thus, glycerol-P is a direct precursor for phospholipid biosynthesis (2). When glycerol-P or its precursors are present in excess, they can be used as carbon and energy sources for a variety of microorganisms. In order to maintain the cellular concentration of glycerol-P at an optimal level for phospholipid synthesis, organisms must be able to coordinate the levels of catabolic activities with the levels of biosynthesis activities.

In *E. coli*, five operons have been found to encode proteins governing dissimulation of glycerol-P and its precursors, and they are termed the *glp* regulon. Although the five *glp* operons are located at three different locations on the *E. coli* linkage map (Fig. 1), they are all subject to negative regulation by the *glpR*-encoded *glp* repressor. *glpE* and *glpG* are two genes found in the same operon with, but promoter-proximal to, *glpR* (T. J. Larson unpublished). Their functions, however, are unknown.

Fig. 2 summarizes glycerol-P metabolism in *E. coli*, which can be separated into three parts: transport of substrates, catabolism, and glycerophospholipid biosynthesis. These parts are discussed in the following paragraphs.

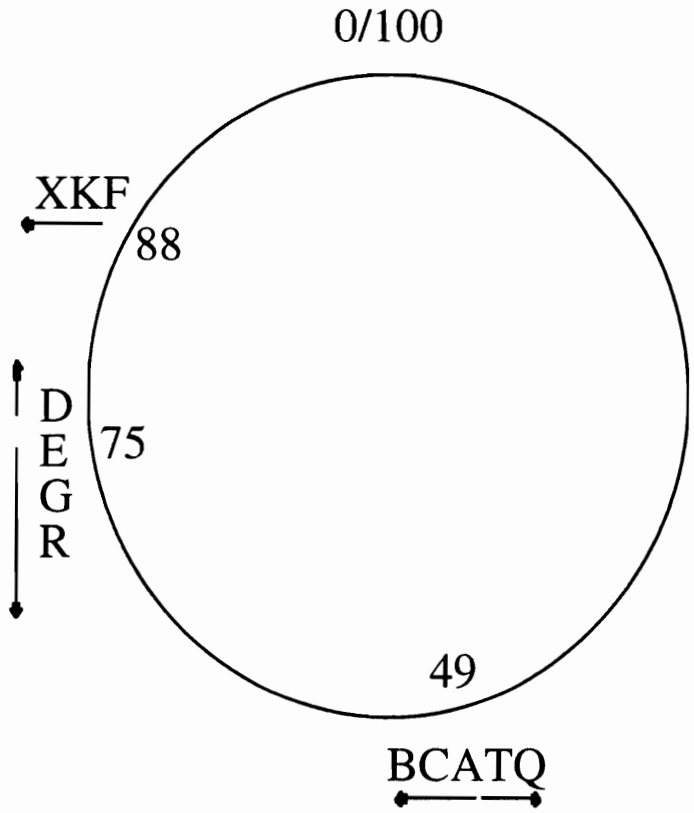


Figure 1. Locations of the *glp* operons on the *E. coli* linkage map. The numbers indicate the location on the chromosome in minutes.

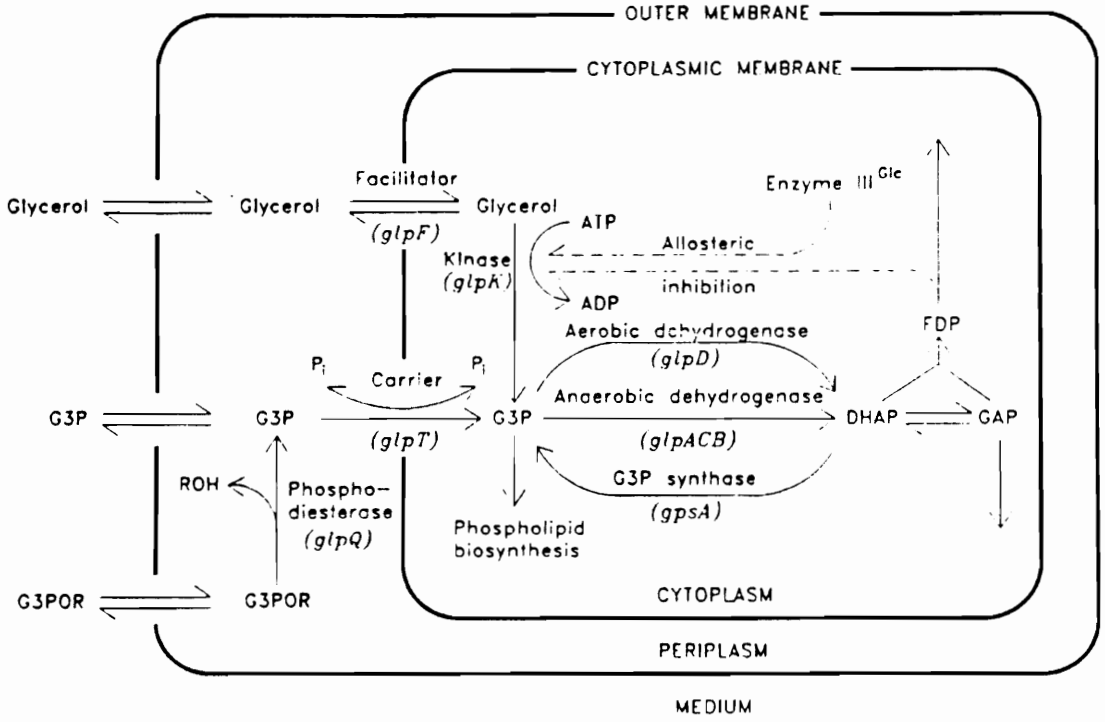


Figure 2. Glycerol-P metabolism in *E. coli* (adapted from Larson *et al.* 3).

Transport of substrates: Among the *glp* genes, *glpF* and *glpT* encode proteins responsible for the transport of the substrates, glycerol and glycerol-P, which are subsequently utilized by *E. coli*. Glycerol enters the cytoplasm of *E. coli* cells via GlpF, a glycerol diffusion facilitator which provides a selective channel with an estimated pore size of 0.4 nm (4). This is the only known example of transport by facilitated diffusion in *E. coli*. (5). It catalyzes the energy-independent equilibration of glycerol between extra- and intracellular space. Once in the cytoplasm, glycerol is phosphorylated by the *glpK*-encoded glycerol kinase forming glycerol-P. This step prevents glycerol from exiting the cell and is the first committed step in its metabolism.

GlpF is a protein of about 30 kDa associated with the cytoplasmic membrane (6, 7). Sequence analysis has revealed six potential membrane spanning helices (7), in contrast to twelve membrane spanning segments typical of other transport proteins, that form a hydrophilic core through which transport occurs. It is therefore proposed that GlpF may form a homodimer with two six-member domains arranged to form the channel. Other compounds that GlpF helps to transport across the cell membrane include small or uncharged molecules such as ribitol, urea, and glycine, but not charged molecules such as glycerol-P and DHAP (5). Because of this broad specificity, as well as temperature insensitivity, the glycerol facilitator was described as a channel rather than a specific carrier protein.

Glycerol kinase (EC 2.7.1.30, K_m glycerol: 10 μM; K_m ATP: 80-100 μM) is a homotetramer of 56 kDa subunits (8). It catalyzes the ATP-dependent phosphorylation of glycerol, and is believed to be the pace-maker for the dissimulation of glycerol.

GlpF is important in glycerol transport and in stimulation of glycerol kinase (9). It was found that upon interacting with GlpF, the K_m of glycerol kinase for glycerol was

decreased from 50 to 5.6 μM and the V_{max} increased from 0.1 to 13.1 nmol/min/10^9 cells (9). However, a Δ*glpF* strain can still grow on glycerol, but the required K_m for normal growth is on the order of 5 mM, compared to 50 μM for the wild-type strain. The fact that the distribution of metabolites is different in the wild-type and the *glpF* mutant support the idea that the metabolic route in these two strains is not identical (9). This might be due to an alternate utilization of glycerol that exhibits a required K_m for normal growth of 5 mM and is not initiated by phosphorylation.

Glycerol-P, on the other hand, is actively accumulated by the cell via several transport systems. The highly specific *ugp* transport system is part of the *pho* regulon and is therefore functional in *pho*-constitutive strains or when phosphate is limiting (10). Glycerol-P has also been reported as a non-specific substrate for the hexose-phosphate transporter UhpT in *uhp*-constitutive or induced strains (11). Nevertheless, the *glpT*-encoded transporter is geared specifically for the import of glycerol-P, with the exchange of two P_i anions (12).

GlpT is found in the cytoplasmic membrane fraction of *E. coli*, and has an apparent size of 50 kDa on SDS gels. Missense mutations in *glpT* confer a negative dominant phenotype, suggesting that the active form of glycerol-P permease may be an oligomer (13). Amino acid sequence homology searching has classified GlpT in a family of P_i-linked transport proteins, including UhpT of *E. coli*. Topological modeling has suggested that they both carry two clusters of six α-helices that span the membrane, which provide an intramolecular dimer structure (13).

GlpT has a K_m for glycerol-P of 12 μM (13), and a K_m for P_i anion of about 1 mM (12). GlpT-mediated glycerol-P:P_i antiport is important in maintaining the carbon/phosphate balance in the cell (12).

A third category of substrates for the *glp* system, phosphodiesters, are hydrolyzed in the periplasmic space by the *glpQ*-encoded phosphodiesterase (14). The resulting glycerol-P is subsequently carried across the cell membrane by the GlpT transporter.

Catabolism of glycerol-P: The catabolic fate of glycerol-P, conversion to DHAP, is catalyzed by two *glp*-gene-encoded glycerol-P dehydrogenases: the aerobic glycerol-P dehydrogenase (GlpD) and the anaerobic glycerol-P dehydrogenase (GlpACB).

The *glpD*-encoded aerobic glycerol-P dehydrogenase (EC 1.1.99.5) is a membrane-bound, flavin-linked, primary dehydrogenase that catalyzes the oxidation of glycerol-P to DHAP. The purified GlpD protein is a homodimer of 58-kDa subunits (15). The K_m for D,L-glycerol-P is 0.8 mM in a purified enzyme preparation and 23 mM in isolated membrane fractions; while the product of the reaction, DHAP is an effective inhibitor with a K_i of 0.5 mM (15). The cofactor for GlpD is FAD, which is reduced to $FADH_2$ with the oxidation of glycerol-P to DHAP. $FADH_2$ then passes electrons on to ubiquinones and ultimately to cytochrome oxidase to generate a proton gradient across the cytoplasmic membrane (16). The proton-pumping ATPase then translocates protons into the cytoplasm, coupled with the reduction of oxygen (aerobic conditions) or nitrate (anaerobic conditions) to produce ATP (17). DHAP, on the other hand, is converted to GAP by triose phosphate isomerase (TPI) and enters the glycolytic pathway.

The *glpACB*-encoded anaerobic glycerol-P dehydrogenase (EC 1.1.99.5) is distinct from the aerobic form of the enzyme in that it consists of two different catalytic subunits,

GlpA and GlpC, and a hydrophobic membrane anchor, GlpB. This is functionally analogous to the *frdABCD* operon, which encodes fumarate reductase. In that case, *frdAB* encodes the catalytic portion of the holoenzyme and the *frdCD* genes encode integral membrane proteins which serve as the membrane anchors for the catalytic subunits (18).

Flavin binding domains within each of the two catalytic subunits, the 62 kDa GlpA and the 43 kDa GlpC, were identified by comparison of their amino acid sequences to that of other *E. coli* flavoproteins (19). It is thought that FAD binds within the Rossman fold found in the GlpA subunit and that GlpC binds FMN. So under anaerobic conditions, glycerol-P is oxidized to form DHAP with FAD and FMN reduced to $FADH_2$ and $FMNH_2$, respectively. After a relatively short transport process, electrons are accepted by exogenous acceptors such as fumarate or nitrate catalyzed by fumarate or nitrate reductase. Thus an anaerobic respiration cycle is completed (20).

A pathway bypassing the glycerol-P step was also discovered. A mutant of a Δ*glpD glpR glpK* strain was isolated as a glycerol-plus pseudorevertant (21). This mutant was found to use an enzyme other than glycerol kinase for the first step of glycerol dissimulation. The enzyme was identified as an NAD^+-linked glycerol dehydrogenase (GLDH), which catalyzed the oxidation of glycerol to DHA, and was purified from this mutant strain (22). The purified enzyme has an apparent molecular weight of 39,000 and was later shown to be identical to the purified D-1-amino-2-propanol oxidoreductase from wild-type *E. coli.* (23). Its structural gene, *gldA* was recently cloned and mapped on the *E. coli* linkage map at 89.2 min. The GLDH from strain MC4100 has a K_m for glycerol of 38 mM, and has a very stringent pH dependence (optimal pH is 9.3) (24). This plus the fact that only hydroxyacetone (but not glycerol) can induce the enzyme may explain the inability of *glpK* derivatives of MC4100 to grow on glycerol at pH 7 even at a 300 mM concentration (21). To enter glycolysis, DHA must first be phosphorylated. But no DHA

kinase has been found in *E. coli* so far. However, a DHA-specific enzyme II of the phosphotransferase system has been proposed (25). Consequently, DHA, produced internally, may diffuse into the medium and be recaptured by the DHA-specific enzyme II delivering DHAP into the cytoplasm, where it could enter glycolysis. Whether this is in fact the alternate route used by *glpD glpR glpK* strains for glycerol utilization, and functions to allow a *glpF* mutant strain to survive in a glycerol medium (9), remains a question.

Glycerophospholipid biosynthesis: In addition to being an energy source, glycerol-P serves another important role for *E. coli*. It is the precursor for glycero-phospholipid biosynthesis.

As the backbone of phospholipids, glycerol-P can be either synthesized from glycerol by *glpK*-encoded kinase or from DHAP by *gpsA*-encoded glycerol-P synthase (26). While both enzymes are located in the cytoplasm, many phospholipid biosynthetic enzymes are found on the inner membrane of *E. coli* (1, 2). Glycerophospholipid biosynthesis begins with the acylation of glycerol-P, which is catalyzed by the *plsB*-encoded acyltransferase (27). The resulting 1-acylglycerol-P is further acylated by the *plsC*-encoded 1-acyl-glycerol-P acyltransferase to form phosphatidic acid (28). The branch point for phospholipid diversification occurs at CDP-diacylglycerol, the activated form of phosphatidic acid and the key intermediate compound in phospholipid synthesis (2). The three major phospholipid species found in *E. coli*, phosphatidylethanolamine (PE) (75%), phosphatidylglycerol (PG) (15-20%), and cardiolipin (CL) (5-10%), are all synthesized from CDP-diacyl glycerol (1).

GENETIC REGULATION OF THE *GLP* REGULON

Expression of the members of the *glp* regulon is subject to three types of regulatory controls: catabolite repression, respiratory repression, and specific transcriptional repression by the *glp* repressor. The *glp* genes show different patterns of responses to each of the three control mechanisms.

Catabolite repression: Catabolite repression is a very common regulatory mechanism controlling carbohydrate metabolism genes in enteric bacteria. When bacteria are exposed to different carbon sources, one is frequently used preferentially. In *E. coli*, the preferred carbon source is usually glucose, but also includes such carbohydrates as fructose, mannose, hexitols, or β-glucosides (29). When the preferred substrate is present, genes encoding metabolic enzymes of other carbohydrates are highly repressed. The abolished expression of enzymes responsible for degradation of other carbon sources is referred as catabolite repression. Repression is initiated by the uptake and subsequent phosphorylation of glucose. The phosphate transfer process takes place via the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) (29), which results in a decrease in intracellular cAMP concentration. CRP (cAMP Receptor Protein) is a central player in catabolite repression, which is required together with cAMP to activate the transcription of the genes encoding the above-mentioned enzymes. Therefore, under circumstances where cAMP is low, these genes undergo the so-called catabolite repression.

Specific CRP-cAMP binding sites within the *glp* regulon have been located by *in vitro* DNase I footprinting and inhibition has been seen during growth on glucose using *in*

vivo assays measuring *glp-lacZ* expression (7). When present in the growth media, glucose brings about a decrease in the intracellular cAMP level. Presumably all of the *glp* operons require CRP-cAMP for maximal expression. Inhibition of *glp* operon transcription by glucose varies from 1.5-fold for *glpD* to 10-fold for *glpTQ*. The degree of inhibition depends on the number of CRP binding sites, the relative positions of CRP sites within the promoters, and their degree of similarity to the consensus CRP binding site (TABLE 1). The sensitivities of the *glp* operons to glucose repression as assessed by using *glp-lacZ* transcriptional fusions (7) (*glpTQ>glpFK>glpACB>glpD*) were similar to those obtained by direct measurement of individual enzyme activities (30) (*glpFK>glpTQ>glpD*). It makes sense that *glpTQ* and *glpFK* are the most sensitive genes, because this would allow a rapid shut down of glycerol and glycerol-P uptake in the presence of the preferred carbon source, glucose.

One of the most common targets of IIA^{Glc} of the PTS is the enzyme or transport system responsible for the uptake of a catabolite carbohydrate or its direct precursors. As a result, the transport of this compound is quickly shut down. This effect is referred to as inducer exclusion (31). Sugars sensitive to catabolite repression and inducer exclusion are classified as the Class I compounds, which include glycerol, lactose, melibiose, and maltose (29). In the *glp* regulon, glycerol kinase is bound and inhibited by nonphosphorylated enzyme IIA^{Glc}, which is the predominant form in the presence of glucose. This specific inactivation of GlpK through binding of IIA^{Glc} also requires the presence of the second substrate, glycerol (29). Upon binding, the glycerol kinase activity of GlpK is inhibited (32). Recently the three-dimensional crystal structure of the glycerol kinase-IIA^{Glc} complex was solved, revealing the exact nature of this allosteric protein-protein interaction (33).

TABLE 1. Summary of catabolite repression of the *glp* operons.

Operon	# of CRP binding sites	Match with consensus	Repression ratio[1]
glpTQ	1	12/14	10.1
glpFK	2	9/14, 7/14	5.5
glpACB	1	9/14	5.5
glpD	1	10/14	1.5

[1] Determined during aerobic growth. This table is adapted from Weissenborn *et al.* (7).

Respiratory control: The only enzyme with an aerobically- and an anaerobically-expressed form in the *glp* regulon is glycerol-P dehydrogenase. The turn-on (off) of GlpD (GlpACB) under aerobic conditions and the opposite action under anaerobic conditions is due to the effect of the transcriptional regulators, FNR and ArcA/B proteins.

As a global regulator, the FNR (fumarate nitrate reduction) protein functions to activate the transcription of genes involved in anaerobic energy-generating pathways, including the *glpACB* operon. It belongs to the same family of transcription regulators as CRP, and possesses a characteristic C-terminal DNA recognition domain (34). Several lines of evidence suggest that a cysteine in the allosteric domain plus several cysteine residues near the N-terminus may form an iron-binding, redox-active site (35). Studies have shown that a mutant FNR (LH28, DA154) containing an Fe-S center, has increased dimerization and DNA binding ability relative to the wild-type enzyme (35). It was

therefore speculated that the effect of such an Fe-S center on FNR dimerization is related to the DNA-binding capacity, and thus may explain the transcriptional activation role of FNR under oxygen limiting circumstances, where the formation of the Fe-S center is favored (36).

In contrast to FNR, the *arcA/arcB* (anaerobic respiratory control) gene products mediate anaerobic repression of genes whose products are needed for growth under aerobic conditions, such as genes encoding the enzymes of the TCA cycle (37). *glpD* is their target in the *glp* regulon. ArcA/ArcB consists of a typical two component regulatory system in *E. coli*. ArcB is the membrane-bound sensor kinase, and ArcA is the cytoplasmic response regulator (20). When the cell is deprived of oxygen, ArcB is autophosphorylated and activates ArcA by donating a phosphoryl group from residue His-292. Upon transfer of the phosphate to an aspartic acid residue of ArcA, ArcA behaves as a repressor (37). The repression of *glpD* is about 2-fold (38), while it varies from 5-fold to about 90-fold on the genes encoding the enzymes of the TCA cycle (39).

Specific repression by the *glp* repressor: All five of the *glp* operons are subject to negative control by the *glpR*-encoded repressor. The *glpR* gene has been cloned downstream of the λ P_L promoter and the repressor was overexpressed. The repressor has been purified by a scheme employing polyethyleneimine (PEI) and ammonium sulfate fractionation, followed by phosphocellulose and DEAE chromatography (40). Purified GlpR was found to be a homotetramer of 29 kDa subunits (40). Each subunit is divided into two domains: the N-terminal domain contributes to DNA binding, and the C-terminal domain is responsible for inducer binding and oligomerization. Interaction of the repressor with glycerol-P, the inducer, was studied using flow dialysis, which revealed 4 binding

sites per repressor tetramer and a dissociation constant of 31 μM for glycerol-P (40). This affinity for for glycerol-P is significantly lower than the affinities of other repressors for their inducers. For instance, LacI has a K_d of 0.6 μM for IPTG (41). One explanation for this is that the intracellular level of glycerol-P must sustain the synthesis of phospholipids, which must always occur in growing cells. On the other hand, the glycerol-P level should be low enough not to cause the induction of *glp* genes, especially the *glpD* gene. Glycerol-P acyltransferase, the enzyme catalyzing the first committed step in phospholipid synthesis, has a K_m of 150 μM for glycerol-P (42). One prediction is that the intracellular glycerol-P pool is about 120 to 240 μM for cells grown on glucose and up to 2.8 mM for cells grown on glycerol (43). However, either of the predicted level of glycerol-P *in vivo* may cause full relief of the GlpR-mediated *glp* gene repression, as it is well above the calculated K_d of GlpR for glycerol-P. An explanation for this paradox is that GlpR may have a much lower affinity for glycerol-P *in vivo* when GlpR is bound to *glp* operators. A large decrease in affinity of *lac* repressor for the inducer (IPTG) occurs when repressor is bound to operator DNA. The affinity for IPTG was found to be approximately 1000-fold lower in this case (41).

Recently, a noninducible version of GlpR ($GlpR^n$) has been identified in mutant strain ECL89. $GlpR^n$ contains an isoleucine substitution at position 101 (T-I) in the C-terminal domain (44). Mutations resulting in noninducible repressor have been observed in other regulatory proteins such as LacI and GalR (45, 46). The noninducible phenotype can be caused by a lower affinity for the inducer and/or a tighter binding to the operator. Determination of *in vitro* inducer binding specific activity showed that $GlpR^n$ has almost no

affinity for glycerol-P (44), which suggests the noninducible nature is due to the lack of binding activity for its inducer.

Recent data obtained from experiments utilizing trypsin cleavage revealed that wild-type GlpR is preferentially cleaved after lysine 71, yielding relatively stable 8 kDa and 22 kDa N-terminal and C-terminal domains. Upon binding glycerol-P, the preferential site of cleavage is moved to arginine 51 and exposes the N-terminal domain to trypsin cleavage (Ali Bhattacharya, personal communication). This provides information on two aspects of GlpR structure. GlpR exists as a protein with two domains linked by an exposed peptide of several amino acids. The fact that trypsin cleavage yields an 8 kDa N-terminal domain and that it does not cleave after arginine 80 provides evidence that the exposed region is not a large peptide. The trypsin experiment also shows that a conformational change takes place as a result of inducer binding. The size of the C-terminal inducer binding domain gets larger, the exposed region moves to around the arginine residue at position 51, while the N-terminal DNA binding domain becomes smaller and is exposed to trypsin. This conformational change may prevent the repressor from binding DNA effectively, and thus abolishes repression.

The N-terminal domain of the repressor has been characterized as a helix-turn-helix DNA binding domain. This was first suggested by the high degree of amino acid sequence similarity between this and other prokaryotic repressors (47). These repressors include those of the glucitol (GutR) (48), fucose (FucR) (49), and deoxyribonucleoside (DeoR) (50) systems of *E. coli*, as well as the *lac* (LacR) (51, 52) repressors of Gram-positive bacteria and the repressor of the agrocinopine (AccR) (53) system of *Agrobacterium*. The overall sequence identity is from 20% to 40%, and the N-terminal domains are even more homologous. The above-mentioned proteins comprise the DeoR family, in contrast to other repressor families in bacteria, such as the LacI-GalR family (54), the AraC-XylS

family (55), and the GntR-PdhR family (56). All of repressors of the DeoR family 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. To investigate whether a change of the recognition helix would yield a change in specificity for operator-binding, the presumptive recognition helix of the *glp* repressor was changed by site-directed mutagenesis to match that of the *deo* repressor. The altered form of *glp* repressor did not recognize either the *glp* or *deo* operator, as determined *in vivo* by monitoring regulation of *glpD-lacZ* and *deoC-lacZ* fusions as well as *in vitro* by gel mobility shift assays (47). The altered form was negatively dominant to wild-type *glp* repressor, indicating the inability to bind DNA with high affinity was likely due to alteration of the DNA-binding domain, and not due to inability to oligomerize or to any instability of the altered repressors (47). The ability of the altered repressor to bind glycerol-P was comparable to that of the wild-type repressor.

Different *glp* operons demonstrate distinct sensitivities to repressor-mediated repression, as summarized in TABLE 2. This differential repression appears to be related to the number and positions of operators present in each operon, and also to the degree of sequence similarity of the individual operator to the consensus. The GlpR-mediated autoregulation of the *glpEGR* operon is still not fully defined. β-galactosidase assays of *glpEGR-lacZ* or *glpE-lacZ* fusions showed a 2- to 6-fold repression in response to the repressor *in vivo* (57), while*in vitro,* DNase I footprinting has failed to identify the proposed operator site as being protected (Bing Yang, unpublished results). Nevertheless, analysis of all the 15 known operators as well as studies involving substitutions at individual positions of the operator have revealed a consensus half site for repressor binding of WATKYTCGWW, where W is A or T, K is G or T, and Y is C or T (58).

TABLE 2. Sensitivities of members of the*glp* regulon to GlpR-mediated repression

Operon	# of operators in promoter region	# of internal operators	Repression ratio during anaerobic growth	Reference
glpFK	4	2	196	7
glpD	2	2	71	59
glpACB	3	1	65	7
glpTQ	1	2 or 3[1]	30	7

[1] Bing Yang, unpublished results.

Even though the *glp* repressor binds to a single *glp* operator, the mechanism by which the repressor binds to operators *in vivo* is believed to be more complicated. Cooperative binding to multiple operator sites clearly provides the basis for repression, and may greatly increase responsiveness to control by the repressor (3). Cooperative binding of the repressor to widely separated operator sites has been demonstrated in the case of *glpTQ-glpACB* operon, where GlpR-mediated DNA looping makes it possible for the involvement of *glpACB* operators in control of *glpTQ* expression (60). It is known that *glp* repressor binds cooperatively to tandemly-arranged operators (58). Binding of repressor to tandem operators is important for control of the *glpD*, *glpACB*, and *glpFKX*

operons (3, 7, 61). Coorperative binding may explain why expression of these operons is more sensitive to repressor when compared with expression of the *glpTQ* operon.

The role of integration host factor (IHF) in facilitating repression of gene expression was also studied. IHF is a sequence specific, asymmetric histone-like protein that plays many important roles in a number of processes including replication, site-specific recombination and transcription (62). It is also known that binding of IHF bends DNA to form a special local nucleo-protein structure and allows IHF to serve as an architectural element in regulating transcription of certain genes (63). Potential IHF binding sites have been found in the *glpTQ-glpACB* promoter-operator region by searching for sequences that match the consensus IHF binding site. DNase I footprinting and gel retardation assays have revealed an IHF-binding site, which is located from +16 to +42 in *glpT* (64). The repression of *glpTQ* by *glp* repressor was found to be 3-fold higher in strains containing IHF. IHF was also found to slightly repress the expression of *glpTQ-lacZ* in the absence of GlpR (64).

The *glp* operons are thus controlled uniquely by the transcriptional modulators as described above. Such an integrative control network allows economical, diverse patterns of gene regulation within a regulon under different environmental and metabolic conditions.

REMAINING QUESTIONS ABOUT THE *GLP* REGULON

Even though our current knowledge of the glycerol-P system seems fairly comprehensive, in fact many questions about glycerol-P metabolism and regulation remain unsolved. Evidence is still needed to resolve whether glycerol dehydrogenase forms part

of the pathway for the utilization of glycerol in Δ*glpFKD* mutant strains. If glycerol dehydrogenase provides an alternate route for utilizing glycerol, the gene(s) responsible for the conversion of DHA to DHAP that permit it to enter glycolysis are still unknown. The hypothesis that GlpR has a lower affinity for glycerol-P in the presence of operator DNA than in the absence of operator DNA has never been proved. Also, there are still three *glp* genes whose functions remain unknown. The *glpX* gene, located downstream of but in the same operon with *glpFK*, was found to encode a protein of 40 kDa, and is probably not membrane-associated (65). To find the biological functions of the other two genes, *glpE* and *glpG*, is the aim of this research.

EXPERIMENTAL PROCEDURES

MATERIALS AND REAGENTS

The Sequenase Version II DNA sequencing system was obtained from US Biochemical Corporation. The Wizard Miniprep DNA purification system, DNA clean up system, and the PCR purification system were obtained from Promega Co. The Fast Stain SDS gel staining reagent was provided by Zoion Company. The gel drying system was from Promega. PCR reagents were purchased from Perkin Elmer, except for the Vent DNA polymerase (Exo^+) which was purchased from New England Biolabs. T4 DNA ligase, polynucleotide kinase, and most restriction enzymes were also obtained from New England Biolabs. From US Biochemical Corporation calf intestine alkaline phosphatase was obtained. Yeast extract, tryptone, casamino acids and bacto agar were from Difco Laboratories. Various salts, sugars, and chloroform, 2-propanol, isoamyl alcohol, phenol, agarose, and acrylamide were from Fisher Scientific Company. Antibiotics, amino acids, ONPG and MOPS were obtained from Sigma Chemical Company. Du-Pont-New England Nuclear supplied [α-^{35}S]dATP (1000-1500 Ci/mmol), [γ-^{32}P]ATP (6000 Ci/mmol), [^{14}C]glycerol-P (144 mCi/mmol) and [^{3}H]glycerol (200 mCi/mmol). The oligonucleotide primers used in this study were synthesized by the phosphoramidate method (66) on an Applied Biosystems 381A Synthesizer and purified by oligonucleotide purification cartridges (Cruachem) as described by the manufacturer. The sequences of all the primers used in this study are listed in TABLE 3.

TABLE 3. Oligonucleotide primers used in this study

serial #	sequence [1]	position [2]
75651	CGTagATCTGCCGCTTCGACGTAAACTGT	303-331
28341	GAAAGAGAGACatATGGATCAGTTCG	360-385
28298	AAAGAGcataTGCTGGTCGATATTCGCG	427-454
34915	TGCTAggatCCACATCGTAGCCCTGTTGC	C637-609
91905	CCGCAGAtcTGGCGTACGGCGCGTAAC	674-701
11360	TAAAagcTTACGCGCCGTACGCCAC	C707-683
181147	GTGGAATTCGCGACAGCAcatATGTTG	723-749
50881	GGCGTTTGTTGATcAtATGGCGACGCAG	785-814
72225	GCACGTACGGGCGcAtaTGGCGCGTTTTC	881-909
38400	GGCGTggaTCCACCGCACCGCCGAGAT	C1248-1222
24551	ATCCCTaagcTTATTTTCGTTTgCGCGCAT	C1585-1555
69882	CCCTGGtcgacTTTTCGTTTTCGCGC	C1583-1557
104350	ATCtCTaGGAATTATTTTCG	C1585-1566
69620	TTTAcAtATGAAACAAACACAACGTCAC	1585-1612
50472	GCAGGAgtcGacCAGCTCCAGTTGAATATGG	C2355-2325
181269	GTCTGACGTGGGtcGaCGTGCA	C2374-2353
82930	TCTCgTCGACCGTATGCAACTG	C2529-2508
253637	CTCACTGtCGACAGCGTGC	C2786-2768
157513	ATTTAGGTGACtCTAgAG	C SP6 primer
168225	TCACTAGTTTGACAGCTTATC[3]	

94425	ACGaaTTCCCgCTAGcAAT[4]	
12390	TAATACGACTCACTATA	T7 promoter
03956	TATGTCA$(CAT)_5$CACCTGGTTCCGCGTGGATCCCA	linker[5]
03921	TATGGGATCCACGCGGAACCAGGT$(GAT)_5$GTGACA	
04145	TCGACCTGGTTCCGCGTGGATCC$(CAC)_5$CATTA	linker[6]
105276	AGCTTAATG$(GTG)_5$GGATCCACGCGGAACCAGG	
138389	GTACGGCGCG$(CAC)_5$CATTA	linker[7]
39309	CCGCGC$(GTG)_5$GTAATTCGA	
86392	CGCGCAAACGAAAA$(CAC)_5$CATTA	linker[8]
50851	GTTTGCTTTT$(GTG)_5$GTAATTCGA	

[1] Mismatched bases with regard to the sequence in Fig. 3 are in lower case; sequences are written from 5' to 3'.

[2] Numbers indicate the position in *glpEGR* sequence (Fig. 3), C represents a complementary sequence.

[3] Complementary to the sequence downstream of the *Cla*I site of pT7-7 (67).

[4] Upstream of the rbs of pT7-7.

[5] Oligonucleotide pair (03956 and 03921) serves as a linker for *Nde*I site and encodes a hexahistidine tag followed by a thrombin cleavage site.

[6] Oligonucleotide pair (04145 and 105276) serves as a linker for *SalI*/*Hind*III site and encodes a thrombin cleavage site followed by a hexahistidine tag.

[7] Oligonucleotide pair (138389 and 39309) serves as a linker for *BsiWI*/*Hind*III site and encodes a hexahistidine tag.

[8] Oligonucleotide pair (86392 and 50851) serves as a linker for *BssH*II/*Hind*III site and encodes a hexahistidine tag.

MICROBIOLOGICAL TECHNIQUES

Bacterial strains and growth media: The strains of *E. coli* K-12 used in this study are described in TABLE 4. Bacteria were routinely grown in LB medium (68) supplemented with the appropriate antibiotics at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 50 μg/ml; kanamycin, 50 μg/ml; spectinomycin, 50 μg/ml; tetracycline, 12.5 μg/ml. When cells were grown for the purpose of IPTG induction, NZCYM medium was usually used (69). AB minimal medium (70) was supplemented with 0.2% casamino acids and 2 μg/ml thiamine. Glucose, glycerol or maltose (0.2%) was provided as carbon source in various experiments, as described in the text. Concentrations of antibiotics were as follows when using minimal media: ampicillin, 40 μg/ml; chloramphenicol, 17 μg/ml; kanamycin, 25 μg/ml; spectinomycin, 20 μg/ml. The indicator plates for testing the Lac phenotype were either MacConkey-lactose agar (71) or various media supplemented with 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal).

Transduction using bacteriophage P1: The method for preparation of bacteriophage P1 lysates was adapted from that described by Silhavy *et al.* (72). A 5-ml culture of a single colony of the donor strain was grown at 37°C in LB for overnight. The overnight culture was diluted 1/100 into 5 ml of LB containing 0.2% glucose and 5 mM

TABLE 4. Strains of *E. coli* K-12

Strain	Genotype	Derivation or reference
DH5αF'	*(F'φ80dlacZΔM15) endA1 recA1 hsdR17 supE44 thi1 gyrA relA1*Δ*(lacZYA-argF)U169*	73
RL22	DH5α (λ*att lacI*^{q}*tetR Sp*^{R}*)*[1]	
MC4100	*F' araD139* Δ*(argF-lac)U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	74
SH305	MC4100 Δ*glpD102 recA1 srl::*Tn*10*	75
TS100	MC4100 *glpR2*	76
TL45	TS100 Δ*(glpT-glpA)593 gyrA*	77
TL73	TS100 *recA1*	13
DA142	MC4100 *glpEG:: Km*^{R} *malT*::Tn*10*	57
GZ533	MC4100 *glpEG:: Km*^{R} *recA1 srl::*Tn*10*[2]	
GD2	MC4100 *glpR2 zih-730*::Tn10 φ(*glpK-lacZ*)hyb λp*lac*Mu	75
GD31	MC4100 φ (*glpK-lacZ*)hyb λp*lac*Mu	P1(GD2) x MC4100, Lac$^+$, Tc^s selection
JC7673	MC4100 *recB21 recC22 sbcB15*	78
NZ411	JC7673 *glpEGR::Km*^{R} *malT::*Tn*10*	Ningyue Zhao, unpublished
NZ45	GD31 *glpEGR::Km*^{R} *(*λ*lacI*^{q} *tetR Sp*^{R}*) recA1 srl::*Tn*10*	Ningyue Zhao, unpublished
GZ555	MC4100 *(*λ*lacI*^{q} *tetR Sp*^{R}*) glpEG:: Km*^{R} *recA1 srl::*Tn*10*[3]	
GZ570	MC4100 *(*λ*lacI*^{q} *tetR Sp*^{R}*)*	P1(RL22) x MC4100, selection for Sp^R

GZ571	GZ570 *glpEGR::Km*^{R}	P1(NZ411) x GZ570
GZ577	GZ571 *recA1 srl::*Tn*10*	P1(SH305) x GZ571
GZ578	GZ577 (λ*glpD-lacZ*)[4]	λ(BY28) lysogen of GZ577
TL658	MC4100 *glpT658* λp1(209)	18
TL661	MC4100 *glpA661* λp1(209)	18
GZ566	TL661 *glpEGR::Km*^{R}	P1(NZ411) x TL661, selection for Km^R, Mal^+
BL21(DE3)	*hsdS gal (λcIts857 ind1 Sam7nin5 lacUV5-T7gene1)*	79
GZ544	BL21(DE3) *glpEG::Km*^{R} *malT::*Tn*10* (?)	P1(DA142) x BL21(DE3), selection for Km^R

[1] DH5α derivative with *lacI*^{q} and *tetR* on a spectinomycin resistant cassette inserted at the λ *att* site, constructed by R. Lutz and T. Larson.

[2] Constructed by P1(TST3) transduction to DA142 cells. Mal^+ was selected, followed by transduction of the resulting cells by P1(SH305). Tetracycline resistance was selected, and UV sensitivity was confirmed.

[3] P1(RL22) x DA142; selection for Sp^R, followed by the same procedure as used for construction of GZ533.

[4] The *glpD-lacZ* transcriptional fusion contains the promoter and operators O_D1 and O_D2.

$CaCl_2$. After 30 minutes, 0.1 ml of P1 lysate (about 10^7-10^8 pfu/ml) was added and the culture was agitated at 37°C for 2-3 hours until the cells lysed. Chloroform (0.1 ml) was added and the culture was vortexed. The cell debris was discarded after centrifugation. The supernatant was then ready to use for P1 transduction. The recipient strain was first grown overnight at 37°C in LB, centrifuged and resuspended in one-half volume of 10 mM $MgSO_4$ containing 5 mM $CaCl_2$. One-tenth milliliter of cell suspension was incubated with 10-100 µl P1 lysate at 37°C without shaking for 30 minutes. One-tenth milliliter of 1 M sodium citrate was added to prevent reinfection after the incubation. The mixture of cell plus lysate was then incubated with 1 ml LB medium at 37 °C for one hour to allow the antibiotic resistant phenotype to develop before the cells were spread on selective media.

Maintenance of cells: Stationary cultures of *E. coli* were routinely stored frozen at -70°C in LB medium supplemented with 15% glycerol.

Construction of strain GZ578: Strain MC4100 (Δ*lacZ*) was chosen as the wild-type parental recipient for a P1 transduction using a lysate from strain RL22 (*lacI*^{q} *tetR* *Sp*^{R}; TABLE 4) as donor. Transductants were selected on medium containing spectinomycin (50 µg/ml). This step introduced the control genes for the Heidelberg plasmids (described below), and all colonies tested expressed the *lac* and *tet* repressors. The resulting strain (GZ570) was transduced using a P1 lysate from NZ411, which has the *glpEGR* region replaced by a Km^R cassette, as well a nearby Tn*10* (Tc^R) insertion in *malT* (min 75.3 with respect to *glpR* at min 75.4). Two antibiotic selective media were

employed to confirm the expected replacement of *glpEGR* and *malT*. In one case, kanamycin resistant transductants were selected, and subsequently tested for tetracycline resistance. Among 50 individual colonies tested, 42 were also tetracycline resistant, indicating an 84% cotransduction frequency. On the other hand, 42 of 50 individual colonies first selected for tetracycline resistance were found to have kanamycin resistance as well. This provided evidence that the Km^R cassette and Tn*10* introduced by transduction are indeed next to each other. The fact that the tetracycline resistant cells were Mal^- provided further evidence that the genetic manipulation was as expected.

The Tc^R and Km^R strain (GZ571) resulting from the above manipulation was chosen as the recipient for the next transduction, which employed P1 from strain SH305 (*malT*$^+$) to select for a Mal^+ derivative (accompanied by sensitivity to tetracycline). In a subsequent transduction, the P1 lysate from SH305 (*recA1 srl*::Tn*10*) was also used to introduce *recA*1 by cotransduction with *srl*::Tn*10*. Replacement of the *recA* gene was confirmed by the characteristic ultraviolet light sensitivity. The resulting strain GZ577 was lysogenized using a λ phage containing *glpD-lacZ* transcriptional fusion on medium with spectinomycin (a lysate of λBY128 was a courtesy of Bing Yang). GZ577 may be employed for construction of other new strains for the study of regulation of *glp* genes by selection of analogous λ lysogens. The new strain GZ578 has all the wild type *glp* genes except *glpEGR*, and contains a *glpD-lacZ* transcriptional fusion controlled by O_D1 and O_D2 (61). GZ578 is also a suitable host for use of the Heidelberg multifunctional plasmids. It should be noted that spectinomycin selection must be maintained to prevent loss of the *lacI*^{q} and *tetR* alleles, which are integrated in tandem with λBY128 at the λ attachment site.

MOLECULAR BIOLOGICAL TECHNIQUES

Transformation: *E. coli* cells were made competent by the following procedure. A fresh overnight culture of the host cell in LB was inoculated into 100 ml of LB at 37°C and allowed to grow to an OD_{600} of 0.6-0.8. Cells were chilled on ice thoroughly and the remaining steps were done in the cold. Cells were first centrifuged at 2,400 g for 15 min, and resuspended in 50 ml of cold 0.1 M $MgCl_2$. The suspension was centrifuged again, resuspended in 25 ml of cold T-salts (75 mM $CaCl_2$, 6 mM $MgCl_2$) and incubated on ice for about 20-30 minutes. After another centrifugation, the cell pellet was resuspended in 3 ml of cold T-salts. Plasmid DNA (0.04 pmol was sufficient to yield about 50 colonies when competent DH5αF' cells were used) was mixed with 100 µl competent cells on ice for about 20 minutes and the mixture was subject to heat shock at 42°C for exactly 2 minutes. The cells were then incubated in 1 ml LB at 37°C for one hour in order for the antibiotic resistant phenotype to develop before spreading on selective media.

Transfection using M13 RF DNA: M13 ligation mixture (0.04-0.1 pmole) was incubated with 0.1 ml DH5αF' competent cells on ice for 30 minutes followed by heat shock at 42°C for 2 minutes. To each sample, 10 µl of 100 mM IPTG, 50 µl of 2% X-gal, and 3 ml of LB top agar were added. After thorough mixing, the samples were poured on pre-warmed LB plates.

Ligation reactions: For ligations involving cohensive ends, a reaction of 10 μl containing about 0.04 pmol of vector DNA and 400 NEB units of T4 DNA ligase was prepared. The ratio of insert to vector was usually kept at 3:1 (molar ratio) for optimal results. The reaction was carried out for more than 6 hours at about 15°C. For blunt-end ligation, the buffer supplied by New England Biolabs was either reduced to half strength or supplemented with PEG at 3% (Ali Bhattacharya, personal communication). For three-fragment ligation, equal amounts of each DNA fragment, about 0.06 pmol, were used. All of the plasmids thus constructed and used in this study are listed in TABLE 5.

The Heidelberg multifunctional plasmid system: The pUH multifunctional plasmid system (R. Lutz and H. Bujard, University of Heidelberg, Germany) was designed for the independently controlled expression of one, two or three different genes in the same cell. The genes for expression were subject to the control by various promoters, for instance, P_{lac}, $P_{lac/ara}$, and P_{N25}. The P_{lac} and P_{N25} promoters are under the control of the gene product of *lacI*^{q} and *tetR*, respectively. Both genes are inserted as a Sp^R cassette in the λ attachment site on the chromosome of the host cell by P1 transduction. $P_{lac/ara}$, in contrast to P_{lac}, is under dual negative and positive regulation, exerted by *lac* repressor and AraC, respectively. Plasmids containing one of these promoters differ in their replication origins and antibiotic resistant markers, so that they are compatible in the same cell and can be independently selected. Another noteworthy feature of this plasmid system is that each element (promoter, origin of replication, antibiotic resistant marker) is cloned as a cassette bracketed by a pair of restriction sites, so they are interchangeable between plasmids. Transcriptional terminators are employed between individual elements.

TABLE 5. Plasmids and M13 phages constructed or used in this study

Plasmid/phage	Source of insert or PCR primers/description	Vector and sites used
pBluescript derivatives		
pGZ80	pNZ80 (*EcoR*I/*Xho*I) filled/*glpE*$^{+}$*G*$^{+}$*R'*	pNZ80 (*EcoR*V/*Xho*I) (44)
pGZ120	91905, 24551 (pNZ80 as template)/*glpG*$^{+}$	pBS (*Bam*HI/*Hind*III)
pSH76 derivatives		
pGZ103	75651, 34915 (pNZ80 as template)/*glpE-lacZ*	pSH76 (*Bam*HI)
pGZ104	91905, 38400 (pNZ80 as template)/*glpG-lacZ*	pSH76 (*Bam*HI)
pGEM3Z derivatives		
pGZ125	91905, 24551 (pNZ80 as template)/*glpG*$^{+}$	pGEM3Z (*Bam*HI/*Hind*III)
pT7-7 derivatives		
pGZ105	28341, 11360 (pNZ80 as template)/*glpE*$^{+}$	pT7-7 (*Nde*I/*Hind*III)
pGZ106	28298, 11360 (pNZ80 as template)/*glpE'*	pT7-7 (*Nde*I/*Hind*III)
pGZ105N	insertion of linker 03956, 03921/*glpE*$^{+}$	pGZ105 (*Nde*I)
pGZ107	insertion of linker 138389, 39309/*glpE*$^{+}$	pGZ105 (*Bsi*WI/*Hind*III)
pGZ108	insertion of linker 138389, 39309/*glpE'*	pGZ106 (*Bsi*WI/*Hind*III)
pGZ110	50881, 24551 (pNZ80 as template)/*glpG'*	pT7-7 (*Nde*I/*Hind*III)
pGZ111	72225, 24551 (pNZ80 as template)/*glpG'*	pT7-7 (*Nde*I/*Hind*III)
pGZ110'	pGZ110 (*Nde*I) filled/*glpG'*	pGZ110 (*Nde*I)
pGZ112	insertion of linker 86392, 50851/*glpG'*	pGZ110 (*Bss*HII/*Hind*III)
pGZ113	insertion of linker 86392, 50851/*glpG'*	pGZ111 (*Bss*HII/*Hind*III)
pGZ114	69620, 253637 (pSY2-6[1] as template)/*glpR*$^{+}$	pT7-7 (*Nde*I/*Hind*III)

pGZ114R	69620, 181269 (pNZ80 as template)/*glpR*$^+$	pT7-7 (*NdeI*/*Hind*III)
pGZ115	69620, 50472 (pNZ80 as template)/*glpR*$^+$	pT7-7 (*NdeI*/*Hind*III)
pGZ116	insertion of linker 04145, 105276/*glpR*$^+$	pGZ114 (*SalI*/*Hind*III)
pGZ117	insertion of linker 04145, 105276/*glpR*$^+$	pGZ115 (*SalI*/*Hind*III)
pGZ118	pGZ117 (*Acc*I) filled/*glpR*$^+$	pGZ117 (*Acc*I)
pGZ119	pGZ117 (*Sal*I) filled/*glpR*$^+$	pGZ117 (*Sal*I)
pGZ126	28341, 82930 (pNZ80 as template)/*glpEGR*$^+$	pT7-7 (*NdeI*/*Hind*III)
pGZ131	pGZ130 (*SalI*/*Hind*III)/*glpRG*$^+$	pGZ114 (*SalI*/*Hind*III)
pGZ133	pGZ132 (*SalI*/*Hind*III)/*glpRE*$^+$	pGZ114 (*SalI*/*Hind*III)
pGZ137	181269, 69620 (pSH79 as template)/*glpRG*$^+$	pGZ131 (*EcoR*V/*Sal*I)
pGZ138	same as pGZ114R but encodes a larger GlpR (SDS-gel)	
pGZ140	pGZ130 (*SalI*/*Hind*III)/*glpRG*$^+$	pGZ138 (*SalI*/*Hind*III)
pGZ141	pGZ132 (*SalI*/*Hind*III)/*glpRE*$^+$	pGZ138 (*SalI*/*Hind*III)
pGZ142	181147, 24551 (pNZ80 as template)/*glpG*	PT7-7 (*NdeI*/*Hind*III)
pUH derivatives		
pGZ127	94425, 168225 (pGZ114 as template)/*glpR*$^+$	pUHE1 P_{lac} (*EcoR*I/*Xba*I)
pGZ128	94425, 104350 (pGZ126 as template)/*glpG*$^+$	pUHE1 P_{lac} (*EcoR*I/*Xba*I)
pGZ129	94425, 168225 (pGZ114 as template)/*glpR*$^+$	pUHE2 P_{N25} (E/X)[(2)]
pGZ130	12390, 157513 (pGZ125 as template)/*glpG*$^+$	pUHE2 P_{N25} (E/X)
pGZ132	94425, 168225 (pGZ105 as template)/*glpE*$^+$	pUHE2 P_{N25} (E/X)
pGZ134	94425, 168225 (pGZ105 as template)/*glpE*$^+$	pUHA4 P_{lac} (*EcoR*I/*Avr*II)
pGZ135	12390, 157513 (pGZ125 as template)/*glpG*$^+$	pUHA4 P_{lac} (*EcoR*I/*Avr*II)
pGZ136	94425, 104350 (pGZ126 as template)/*glpEG*$^+$	pUHA4 P_{lac} (*EcoR*I/*Avr*II)
pGZ148	pGZ135 (E/X)/*glpG*$^+$	pUHA3 P_{N25} (E/X)
pGZ149	pGZ136 (E/X)/*glpEG*$^+$	pUHA3 $P_{lac/ara}$ (E/X)
pGZ150	pGZ136 (E/X)/*glpEG*$^+$	pUHA3 P_{N25} (E/X)

pGZ152	pGZ134 (E/X)/*glpE*$^+$	pUHA4 $P_{lac/ara}$ (E/X)
pGZ153	pGZ135 (E/X)/*glpG*$^+$	pUHA4 $P_{lac/ara}$ (E/X)
pGZ154	pGZ134 (E/X)/*glpE*$^+$	pUHA3 P_{N25} (E/X)
pGZ155	pGZ127 (E/X)/*glpR*$^+$	pUHA3 P_{lac} (E/X)
pGZ156	pGZ158 with N47->R mutation[3]	
pGZ157	pGZ158 with deletion of #2020 T in *glpR*, V119->I[4]	
pGZ158	94425, 168225 (pNZ80 as template)/*glpR*$^+$	pUHE1 P_{lac} (E/X)
pGZ160	94425, 168225 (pGZ142 as template)/*glpG*$^+$	pUHA3 P_{N25} (E/X)
pGZ162	pGZ127 (*EcoRI*/*Avr*II)/*glpR*$^+$	pUHA3 P_{N25} (E/X)
pGZ163	pGZ158 (*EcoRI*/*Avr*II)/*glpR*$^+$	pUHS1 P_{lac} (E/X)
M13mp19 derivatives		
M13mp19-1	pGZ127 (*EcoRI*/*Hind*III)/*glpR*$^+$	M13mp19RF (*EcoRI*/*Hind*III)
M13mp19-2	pGZ156 (*EcoRI*/*Hind*III)/*glpR*$^+$	M13mp19RF (E/H)[2]
M13mp19-3	pGZ157 (*EcoRI*/*Hind*III)/*glpR*'	M13mp19RF (E/H)
M13mp19-4	pGZ158 (*EcoRI*/*Hind*III)/*glpR*$^+$	M13mp19RF (E/H)
M13mp19-5	pGZ148 (*EcoRI*/*Hind*III)/*glpG*$^+$	M13mp19RF (E/H)
M13mp19-6	pGZ154 (*EcoRI*/*Hind*III)/*glpE*$^+$	M13mp19RF (E/H)
M13mp19-7	pGZ150 (*EcoRI*/*Xba*I)/*glpEG*$^+$	M13mp19 (*EcoRI*/*Xba*I)

[1] pSY2-6 (80) was found to encode the P33E,Q34M substitution in GlpR, so pGZ114 and all its derivatives, such as pGZ116, pGZ127, pGZ129, pGZ131, pGZ133, pGZ155, and pGZ162 have this mutation. pGZ114 was reconstructed with the wild-type *glpR* in pT7-7, and was named pGZ114R.

[2] Abbreviations in this table: E/X for *EcoRI*/*Xba*I; E/H for *EcoRI*/*Hind*III.

[3] *glpR* sequence in pGZ156 has a G->A mutation at position 1731, and therefore an N47 to K substitution.

[4] *glpR* sequence in pGZ157 has a V119->I substitution that resulted from a G->A mutation at nucleotide 1945, and a deletion of T at nucleotide 2020. The translational reading frame downstream of position 2020 is thus shifted, which resulted in using the UAA at 2105-2107 as the stop codon.

PCR for screening recombinant plasmid and two-temperature PCR: Regular PCR reactions were carried out by using the protocol provided by Perkin Elmer. PCR was also used to screen recombinant plasmids directly from colonies grown on LB plates. Compared with restriction digestion of recombinant plasmids, this was especially useful when multiple colonies were to be screened. The colony was picked using a pipet tip, and swirled in a microcentrifuge tube containing 50 µl of ddH_2O. The solution was vortexed and boiled for 5 minutes, followed by immediate chilling on ice. Cells were then pelleted at full speed in a microcentrifuge for 15 minutes. Five microliters of the supernatant was used as the template in a 20 µl PCR reaction. PCR products were checked by loading 10 µl of the reaction mixture on an agarose gel. The so-called two temperature PCR was used when primers were no longer than 18 nucleotides. In this procedure, 55°C annealing and 72°C elongation steps were integrated into one step at 45°C.

Preparation of ssDNA for sequencing: An alkaline denaturation method was used for preparation of ssDNA from double stranded plasmid DNA. Fifteen microliters (about 3 µg) of DNA from a standard plasmid DNA preparation (Wizard Miniprep System, Promega Co.) was mixed with water and NaOH (final concentration 0.2 N). The mixture was incubated at 37°C for 5 minutes to allow denaturation, followed by a conventional ethanol precipitation. The dried DNA pellet was dissolved in 7 µl of water and used for one sequencing reaction. The method used for preparation of ssDNA from M13 phage constructs was modified from that described by Messing (81). A single plaque was picked

with a sterile toothpick and the toothpick was dropped into a log-phase culture of DH5αF', followed by agitation at 37°C for 4-5 hours. The cell pellet resulting from centrifugation of 1 ml culture was discarded, and the supernatant was carefully collected into a new microcentrifuge tube and mixed with 300 µl of 20% PEG containing 2.5 M NaCl. The mixture was incubated at room temperature for 15 minutes followed by centrifugation for 5 minutes. The pellet was resuspended in 100 µl TE, and 100 µl TE-saturated phenol was added. The mixture was vortexed and centrifuged for 5 minutes in a microcentrifuge. Next, the aqueous phase was transferred to a fresh tube and mixed with 100 µl chloroform, and centrifuged for 1 minute. Finally, the top phase was removed and a conventional ethanol precipitation was performed. The pelleted DNA was dried, then dissolved in 21 µl water, and was used in 3 sequencing reactions.

BIOCHEMICAL TECHNIQUES

Overproduction of protein and preparation of crude extract for purification: An overnight culture grown in LB or NZCYM medium was diluted 1:100, and allowed to grow to log phase (OD_{600}=0.6-0.8). It was then induced with either 0.1 mM IPTG, or with 2 µg/ml of tetracycline in the case of P_{N25}-controlled expression, for about 2 hours. Cell growth was monitored by measuring the optical density at 600 nm. The total amount of protein was estimated assuming that 100 µl of culture with an OD_{600} of

1 contains 10 µg of total protein. Cells were harvested by centrifugation for 10 minutes at 2,400x g (4°C). The cell pellet was washed with the buffer used for purification, centrifuged and cells were resuspended in the same buffer at about 8 mg/ml total protein concentration based on the above estimation. Cells were broken by passing the mixture through a French press twice at 12,000-20,000 psi. Unbroken cells were pelleted by centrifugation at 2,400 g for 10 minutes, and inclusion bodies were pelleted by centrifugation at 7,500 g for 15 minutes. The supernatant was decanted and used as the crude extract for further purification. If the protein of interest was present in the cytoplasmic fraction, the crude extract was centrifuged one more time at 40,000 rpm for 90 minutes in Beckman TY65 rotor. The supernatant from this centrifugation is the cytoplasmic fraction containing the protein for purification.

Determination of protein concentration: The method of Bradford (82) was used routinely to determine the protein concentration. BSA in the range of 1 µg-6 µg in 100 µl of water was used as the standard.

Subcellular fractionation: This approach was modified from a method described by Larson *et al.* (13). After disrupting the cells by sonication and removal of unbroken cells by centrifugation, a total membrane fraction was obtained by sedimentation through 15% (w/v) sucrose onto a 1.5 ml 70% (w/v) cushion (16 ml Beckman ultracentrifugation tubes for SW41 rotor, 35,000 rpm, 90 min). The membrane fraction was diluted and then layered on top of a second SW41 tube containing 1.5 ml of 70% sucrose and 8.5 ml of 53% sucrose. After centrifugation at 35,000 rpm for 4 hours, the

cytoplasmic and outer membrane fractions were withdrawn from the upper and lower interface regions, respectively. All steps were carried out at 4°C. When purifying a membrane-associated protein, this step was usually included as an initial procedure to enrich the protein of interest.

Overproduction of GlpG and N-terminal amino acid sequence analysis: Strain BL21(DE3)pGZ126 was used for overproduction of GlpG. The crude extract was first fractionated using the procedure described above. The resulting cytoplasmic membrane fraction was loaded on a 15% preparative SDS polyacrylamide gel. After completion of electrophoresis, the gel was washed in 40% methanol containing 1% acetic acid, and stained in 0.1% Coomassie blue R-250 in the above solution. The region containing GlpG was excised and subject to electroelution in 15 mM $(NH_4)_2CO_3$ (pH 8.2), 0.1% SDS buffer by using an Elutrap chamber supplied by Schleicher & Schuell. The protein sample was simultaneously concentrated by electroelution. About 100 pmol of the final product was analyzed by automated Edman degradation on an ABI amino acid sequencer operated in Dr. Charles Rutherford's Laboratory, Department of Biology, Virginia Tech.

Digestion of protein with engineered thrombin site: Thrombin was purchased from Sigma Chemical Company (0.06 NIH unit/μl). The reactions were carried out in 50 mM sodium citrate (pH 6.5), 150 mM NaCl, at 37°C for 45 minutes in a total volume of 20 μl. For 5 μg of protein, a maximum of 50 ng thrombin was used. Samples were analyzed by SDS-PAGE after digestion.

Isoelectric focusing gel electrophoresis: Adams' recipe (83) was used to make slab gels of 10 cm x 7 cm x 0.75 cm. Protein samples were exhaustively dialyzed in ddH_2O containing 1% glycine before loading. The loading buffer used was 6 M urea in 1% glycine (84). The gel was first pre-run at 500 Volts for 30 minutes in a Hoefer Minigel Set with circulating cool water. After samples were loaded, electrophoresis was continued for 90 minutes at a constant voltage of 900 Volts. The gel was either subject to Western blotting or washed in 45% methanol, 10% acetic acid, and stained with Fast Stain from Zoion Co.

Western blot: The protocol was adapted from a method described by Harlow and Lane (85), and modified for use with the GlpR polyclonal antibody (raised from immunized rabbits by Cocalico Company, PA). Approximately 0.5 mg of pure GlpR preparation (40) was injected at three different times (one week between the first two boosts, followed by a 10-day interval until the third boost) into the host rabbit. Antisera used in the experiment was from the second bleed, which was taken 6 weeks after the first injection.

After completion of an SDS or IEF gel electrophoresis, the gel was first soaked in transfer buffer (25 mM Tris, 250 mM glycine, pH 8.3, 20% methanol and 0.05% SDS) for 20 minutes. The nitrocellulose membrane (Schleicher & Schuell Co.) was soaked first in water for 1-5 minutes, then in transfer buffer for 15 minutes. Electrotransfer was performed by using a Bio-Rad semidry transblotter, with a maximum voltage of 20 Volts and a maximum current of 4 mA/cm^2. Transfer time was usually from 45 minutes to 1 hour. The nitrocellulose membrane was blocked with 5% nonfat dry milk in TBST buffer [10 mM Tris, pH 8.0, 150 mM NaCl, 0.1% (v/v) Tween 20] for 1 hour. The milk was

poured off and replaced with TBST buffer containing antibody (2,000- to 10,000-fold dilution), and incubated for 1 hour. The membrane was then washed three times with TBST buffer and soaked with secondary antibody (2,000- to 4,000-fold dilution in TBST) for 1 hour. The washing steps were then repeated, followed by rinsing of the membrane with alkaline phosphatase buffer (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$). The BCIP/NBT substrate was used for detection of alkaline phosphatase. It was made by adding 660 µl NBT (50 mg/ml) and 330 µl BCIP (50 mg/ml) to 100 ml alkaline phosphatase buffer. Upon satisfactory color development, the reaction was stopped by addition of 20 mM EDTA.

Assay of β-galactosidase activity: The activity of β-galactosidase was determined spectrophotometrically by measuring formation of a yellow colored product (*o*-nitrophenol) derived from the hydrolysis of the substrate ONPG. A single colony was inoculated in AB minimal medium containing 2 µg/ml of thiamine, 0.2% casamino acids, and 0.2% of a specified carbon source. The overnight cultures were subcultured and allowed to grow to an OD_{600} of 0.5-0.7. The proper amount of cells were taken, centrifuged, and suspended in Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH 7.0). The OD_{578} was measured to determine the relative amount of cells for assay. A total 0.8 ml of the cells resuspended in Z buffer was assayed at 30°C by the following procedure: SDS (10 µl of 0.1%) and chloroform (20 µl) were added and the mixture was vortexed for 10 seconds. After a 10 minute incubation at 30°C, the addition of 160 µl of pre-warmed ONPG (4 mg/ml in 0.1 M

pH 7.0 phosphate buffer) was added to start the reaction. The reaction was stopped by adding 0.4 ml of 1 M Na_2CO_3 after the appropriate yellow color had developed (OD_{420} in the range of 0.3-0.6). Meanwhile, the reaction time was recorded. The reaction mixture was clarified by microcentrifugation for 2 minutes, and absorbance at 420 nm was measured. The specific activity (Miller units, 71) was calculated using the following formula:

$$\text{Units} = \frac{OD_{420} \times 800}{OD_{578} \times \text{dilution fold} \times T \times V}$$

T: time of the reaction in minutes

V: volume of the culture (ml) used in the assay

Purification of GlpE by immobilized metal affinity chromatography (IMAC): GlpE with an engineered polyhistidine tag at the C-terminus was purified by the so-called IMAC method (86). $NiCl_2$ (0.1 M) was used to charge the chelating sepharose fast flow prepacked column (1 ml, Pharmacia LKB Biotechnology). The column was equilibrated with 50 mM Tris-HCl pH 7.0, 50 mM KCl before loading 10 mg of the cytoplasmic fraction from strain BL21(DE3)pGZ107, which overproduces GlpE$(His)_6$. The column was then washed using the same buffer containing 75 mM imidazole. Next, the column was developed with a 10 ml gradient of imidazole from 75 mM to 250 mM in the same buffer, followed by a constant wash of 250 mM imidazole. Fractions containing 100 to 150 mM imidazole were found to contain the purified GlpE-$(His)_6$ protein.

Operator binding assay for *glp* repressor: PCR primers ACGACTCACTATAGGGCGAATTCG and GAATACTCAAGCTTGCATGCCTGC

were used to amplify the consensus single *glp* operator which was cloned in vector pGEM3Z (58). Operator DNA's 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. The purified operator DNA was then incubated with extract containing GlpR at 30°C for 30 min. The association buffer used was: 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 50 mM KCl, 2 mM $CaCl_2$, 4 mM $MgCl_2$, 5% (v/v) glycerol, 0.025% (v/v) Triton X-100, 100 µg/ml bovine serum albumin, and 5 µg/ml sonicated salmon sperm DNA (7). Samples were analyzed on gels containing 8% polyacrylamide, 0.2% N,N'-methylenobisacrylamide, and one-half strength TBE buffer (pH 8.3). Electrophoresis was performed at 4°C, 160 volts until the dye front migrated about 5 cm into the gel.

Inducer binding assay for *glpR*-encoded repressor: Each assay contained the following: 20 µl of 250 mM MOPS buffer (pH 6.5), 10 µl of casein (4 mg/ml), 50 µl of extract containing the repressor, and water to make a final volume of 80 µl. Twenty microliters of [^{3}H]glycerol-P (0.5 mM, about 110 cpm/pmol) was then added and the mixture was incubated at room temperature for 5 min. Then, 0.3 ml of ice-cold saturated ammonium sulfate was added. The reaction mixture was vortexed and placed on ice for 5 min, followed by centrifugation for 5 min in the cold. The supernatant was aspirated and the pellet was washed with 0.5 ml cold, saturated ammonium sulfate followed by centrifugation for 2 min in the cold. The supernatant was removed as before and the final pellet was resuspended in 100 µl of water. The radioactivity of glycerol-P bound to the

pellet was determined in a scintillation counter. The specific activity for inducer binding (pmol inducer/mg protein) was determined by using the following formula:

$$\text{Specific Activity} = \frac{\text{cpm - background cpm}}{\text{110 cpm/pmol x mg of protein in assay}}$$

Glycerol-P transport assay: The *glpT*-encoded active transport activity was assayed by measuring the uptake of radiolabeled [^{14}C]glycerol-P into cells. Cells were grown in AB minimal medium overnight. They were harvested and washed twice with one-half volume of 0.01 M Tris-HCl (pH 7.5), 0.15 M NaCl. The cells were resuspended in the same buffer to obtain an OD at 578 nm of 0.5. Ten microliters of 5% glycerol was added to 0.5 ml of these cells. The mixture was then incubated for 1 minute before addition of 25 µl of [^{14}C]glycerol-P (final concentration 0.3 µM). Samples (100 µl) from the reaction mixture were filtered at 30, 60, and 90 seconds through a Whatman nitrocellulose membrane (0.45 µm pore size), followed by a wash with 10 ml of the above buffer. Filters were dried in an oven before radioactivity was determined in a scintillation counter. The specific activity for transport (pmol glycerol-P/min/ml of cells at 1 OD_{578}) was determined by using the following formula:

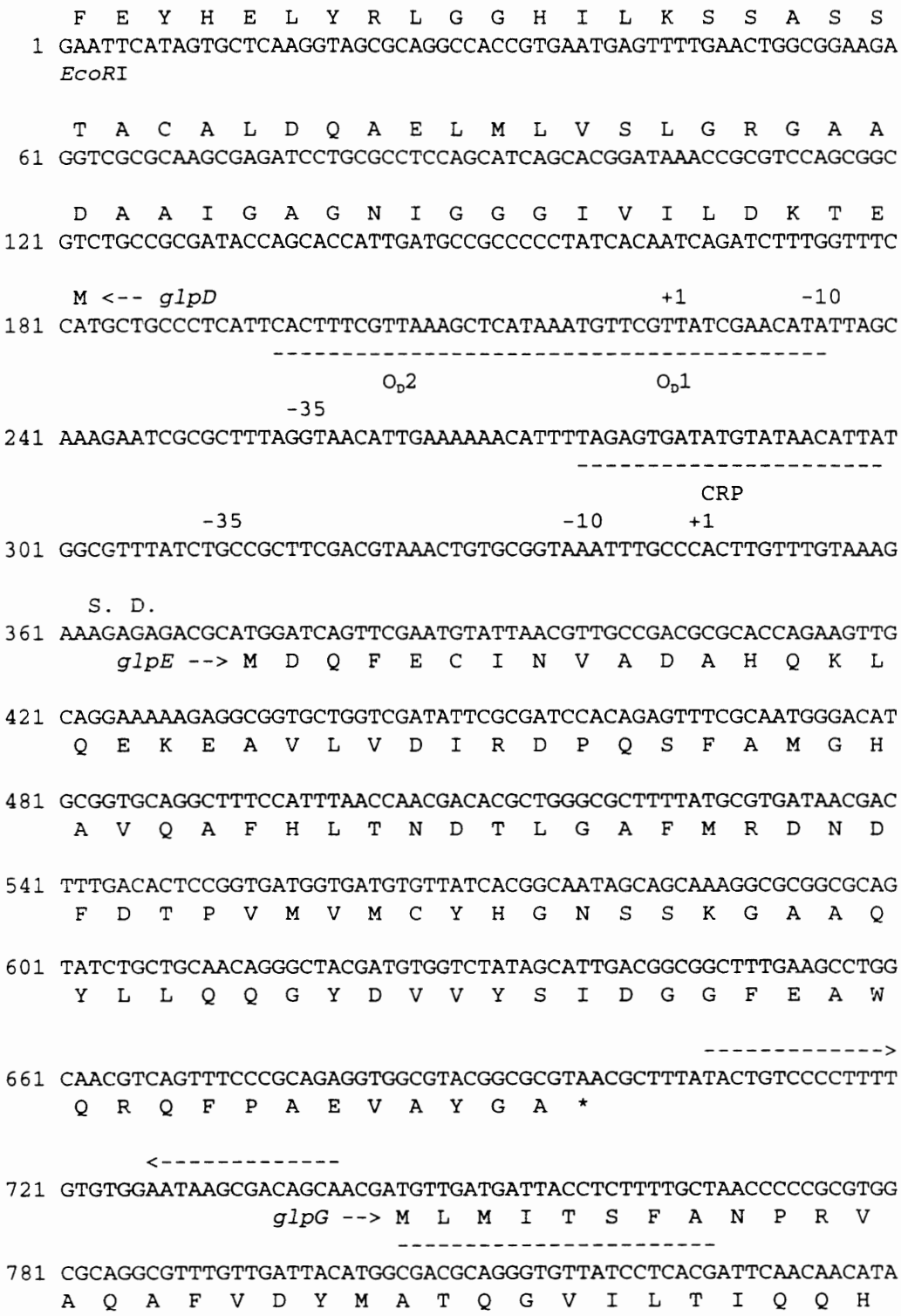
$$\text{Specific Activity (unit)} = \frac{\text{cpm - background cpm}}{\text{280 cpm/pmol x time x ml of cells x } OD_{578}}$$

PART I: CHARACTERIZATION OF THE *GLPE*, *GLPG*, AND *GLPR* OPEN READING FRAMES IN THE *GLPEGR* OPERON

From previous studies, the nucleotide sequence of *glpEGR* was determined and submitted to GenBank with accession number M96795 (80). The region sequenced contains 2895 bp, from an *EcoR*I site within the *glpD* gene to a *Bgl*II site downstream of the *glpE*, *glpG*, and *glpR* genes which are transcribed divergently from *glpD* (87, Fig. 3). Recent studies have shown that the *glpE*, *glpG*, and *glpR* genes are all under the control of the *glpE* promoter, and are therefore in the same operon. My research concentrated on the roles of GlpE/GlpG in *E. coli.*

Determination of the *glpE* start codon and analysis of the deduced amino acid sequence of GlpE: The gene product of *glpE* was first observed by using minicell and maxicell techniques. GlpE is a protein of 13 kDa (75, 87). Computer analysis of the *glpEGR* sequence has revealed two possible start codons located in the region corresponding to the *glpE* gene: the AUG at position 373-375 and the GUG at position 436-438 (Fig. 3). Both of these two possible start codons are predicted to use the UAA at position 697-699 as the stop codon.

To verify the start codon used by *glpE in vivo*, PCR primers (#28341, and #28298, TABLE 3) each bearing an *Nde*I site were made to cover each of the two possible start codons. A downstream primer bearing a *Hind*III site (#11360, TABLE 3) was made to complement the region including the stop codon of *glpE* gene. PCR products from each of the primer pairs were cloned separately between the *Nde*I and *Hind*III site of the expression vector pT7-7 (67). Plasmids containing the first and the second start codon were named pGZ105 and pGZ106, respectively. The plasmids were transformed into strain BL21(DE3). Induction of GlpE expression was as described in the Experimental Procedures.


F E Y H E L Y R L G G H I L K S S A S S
1 GAATTCATAGTGCTCAAGGTAGCGCAGGCCACCGTGAATGAGTTTTGAACTGGCGGAAGA
EcoRI
T A C A L D Q A E L M L V S L G R G A A
61 GGTCGCGCAAGCGAGATCCTGCGCCTCCAGCATCAGCACGGATAAACCGCGTCCAGCGGC
D A A I G A G N I G G G I V I L D K T E
121 GTCTGCCGCGATACCAGCACCATTGATGCCGCCCCCTATCACAATCAGATCTTTGGTTTC
M <-- glpD +1 -10
181 CATGCTGCCCTCATTCACTTTCGTTAAAGCTCATAAATGTTCGTTATCGAACATATTAGC
O_b2 O_b1
-35
241 AAAGAATCGCGCTTTAGGTAACATTGAAAAAACATTTTAGAGTGATATGTATAACATTAT
CRP
-35 -10 +1
301 GGCGTTTATCTGCCGCTTCGACGTAAACTGTGCGGTAAATTTGCCCACTTGTTTGTAAAG
S. D.
361 AAAGAGAGACGCATGGATCAGTTCGAATGTATTAACGTTGCCGACGCGCACCAGAAGTTG
glpE --> M D Q F E C I N V A D A H Q K L
421 CAGGAAAAAAGAGGCGGTGCTGGTCGATATTCGCGATCCACAGAGTTTCGCAATGGGACAT
Q E K E A V L V D I R D P Q S F A M G H
481 GCGGTGCAGGCTTTCCATTTAACCAACGACACGCTGGGCGCTTTTATGCGTGATAACGAC
A V Q A F H L T N D T L G A F M R D N D
541 TTTGACACTCCGGTGATGGTGATGTGTTATCACGGCAATAGCAGCAAAGGCGCGGCGCAG
F D T P V M V M C Y H G N S S K G A A Q
601 TATCTGCTGCAACAGGGCTACGATGTGGTCTATAGCATTGACGGCGGCTTTGAAGCCTGG
Y L L Q Q G Y D V V Y S I D G G F E A W
661 CAACGTCAGTTTCCCGCAGAGGTGGCGTACGGCGCGTAACGCTTTATACTGTCCCCTTTT
Q R Q F P A E V A Y G A *
<----------
721 GTGTGGAATAAGCGACAGCAACGATGTTGATGATTACCTCTTTTGCTAACCCCCGCGTGG
glpG --> M L M I T S F A N P R V
781 CGCAGGCGTTTGTTGATTACATGGCGACGCAGGGTGTTATCCTCACGATTCAACAACATA
A Q A F V D Y M A T Q G V I L T I Q Q H

```
841 ACCAAAGCGATGTCTGGCTGGCGGATGAGTCCCAGGCCGAGCGCGTACGGGCGGACGTGG
    N  Q  S  D  V  W  L  A  D  E  S  Q  A  E  R  V  R  A  D  V

901 CGCGTTTTCTCGAAAACCCGGCAGATCCGCGTTATCTGGCGGCGAGCTGGCAGGCAGGCC
    A  R  F  L  E  N  P  A  D  P  R  Y  L  A  A  S  W  Q  A  G

961 ATACCGGCAGTGGCCTGCATTATCGCCGTTATCCTTTCTTTGCCGCCTTGCGTGAACGCG
    H  T  G  S  G  L  H  Y  R  R  Y  P  F  F  A  A  L  R  E  R

1021 CAGGTCCGGTAACCTGGGTGATGATGATCGCCTGCGTGGTGGTGTTTATTGCCATGCAAA
     A  G  P  V  T  W  V  M  M  I  A  C  V  V  V  F  I  A  M  Q

1081 TTCTCGGCGATCAGGAAGTGATGTTATGGCTGGCCTGGCCATTCGATCCAACACTGAAAT
     I  L  G  D  Q  E  V  M  L  W  L  A  W  P  F  D  P  T  L  K

1141 TTGAGTTCTGGCGTTACTTCACCCACGCGTTAATGCACTTCTCGCTGATGCATATCCTCT
     F  E  F  W  R  Y  F  T  H  A  L  M  H  F  S  L  M  H  I  L

1201 TTAACCTGCTCTGGTGGTGGTATCTCGGCGGTGCGGTGGAAAAACGCCTCGGTAGCGGTA
     F  N  L  L  W  W  W  Y  L  G  G  A  V  E  K  R  L  G  S  G

1261 AGCTAATTGTCATTACGCTTATCAGCGCCCTGTTAAGCGGCTATGTGCAGCAAAAATTCA
     K  L  I  V  I  T  L  I  S  A  L  L  S  G  Y  V  Q  Q  K  F

1321 GCGGGCCGTGGTTTGGCGGGCTTTCTGGCGTGGTGTATGCGCTGATGGGCTACGTCTGGC
     S  G  P  W  F  G  G  L  S  G  V  V  Y  A  L  M  G  Y  V  W

1381 TACGTGGCGAACGCGATCCGCAAAGTGGCATTTACCTGCAACGTGGGTTAATTATCTTTG
     L  R  G  E  R  D  P  Q  S  G  I  Y  L  Q  R  G  L  I  I  F

1441 CGCTGATCTGGATTGTCGCCGGATGGTTTGATTTGTTTGGGATGTCGATGGCGAACGGAG
     A  L  I  W  I  V  A  G  W  F  D  L  F  G  M  S  M  A  N  G

1501 CACACATCGCCGGGTTAGCCGTGGGTTTAGCGATGGCTTTTGTTGATTCGCTCAATGCGC
     A  H  I  A  G  L  A  V  G  L  A  M  A  F  V  D  S  L  N  A

                      S. D.
1561 GAAAACGAAAATAATTCCAGGGATTTATAAATGAAACAAACACAACGTCACAACGGTATT
     R  K  R  K  *          glpR --> M  K  Q  T  Q  R  H  N  G  I
                                      ---------------

1621 ATCGAACTGGTTAAACAGCAGGGTTATGTCAGTACCGAAGAGCTGGTAGAGCATTTCTCC
      I  E  L  V  K  Q  Q  G  Y  V  S  T  E  E  L  V  E  H  F  S
                                                  Helix

1681 GTCAGCCCGCAGACTATTCGCCGCGACCTCAATGAGCTGGCGGAGCAAAACCTGATCCTG
      V  S  P  Q  T  I  R  R  D  L  N  E  L  A  E  Q  N  L  I  L
      Turn           Helix
```

```
1741 CGCCATCATGGCGGTGCGGCGCTGCCTTCCAGTTCGGTTAACACGCCGTGGCACGATCGC
      R  H  H  G  G  A  A  L  P  S  S  S  V  N  T  P  W  H  D  R

1801 AAGGCCACCCAGACCGAAGAAAAAGAGCGCATCGCCCGCAAAGTGGCGGAGCAAATCCCC
      K  A  T  Q  T  E  E  K  E  R  I  A  R  K  V  A  E  Q  I  P

                              EcoRV
1861 AATGGCTCGACGCTGTTTATCGATATCGGCACCACGCCGGAAGCGGTAGCGCACGCACTG
      N  G  S  T  L  F  I  D  I  G  T  T  P  E  A  V  A  H  A  L

1921 CTCAATCACAGCAATTTGCGCATTGTCACCAACAATCTCAACGTTGCTAACACGTTGATG
      L  N  H  S  N  L  R  I  V  T  N  N  L  N  V  A  N  T  L  M

1981 GTAAAAGAAGATTTTCGCATCATTCTCGCCGGTGGCGAATTACGCAGCCGCGATGGCGGG
      V  K  E  D  F  R  I  I  L  A  G  G  E  L  R  S  R  D  G  G

2041 ATCATTGGCGAAGCGACGCTCGATTTTATCTCCCAGTTCCGCCTTGATTTCGGCATTCTG
      I  I  G  E  A  T  L  D  F  I  S  Q  F  R  L  D  F  G  I  L

2101 GGGATAAGCGGCATCGATAGCGACGGCTCGCTGCTGGAGTTCGATTACCACGAAGTTCGC
      G  I  S  G  I  D  S  D  G  S  L  L  E  F  D  Y  H  E  V  R

2161 ACCAAACGCGCCATTATTGAGAACTCGCGCCACGTTATGCTGGTTGTCGATCACTCGAAA
      T  K  R  A  I  I  E  N  S  R  H  V  M  L  V  V  D  H  S  K

2221 TTTGGCCGTAACGCGATGGTCAATATGGGCAGCATCAGCATGGTAGATGCCGTCTACACC
      F  G  R  N  A  M  V  N  M  G  S  I  S  M  V  D  A  V  Y  T

2281 GACGCCCCGCCGCCAGTAAGCGTGATGCAGGTGCTGACGGACCACCATATTCAACTGGAG
      D  A  P  P  P  V  S  V  M  Q  V  L  T  D  H  H  I  Q  L  E

                                                  ----------
2341 CTGTGCTGATCCTGCACGGCTTCCCACGTCAGACCAAAACGCGCCAGGTATTTGCGTACG
      L  C  *
     -->          <----------

2401 CGATCCGCGTCATTGACGCTGGCTTTGCCCTGGCGCGAAACGTCAAAAAGCTGGCGTCCG

2461 GCTGCGGAAAGCGACTTTGCCTGGCGGCAGATAGCGATAACGTGTTCCAGTTGCATACGG

2521 TCGAAGAGATCGATGTTTTCTGCCTCTGCGCCCAGCAACGCCGTAAGCGCGGAGGGGCGG

2581 CTCTCCTGCCAGTTATAGCGCAGACGGTTTATCTCATCTTCAACCACGTCCAGAGTGATG

2641 CGTCCGCTAGTGGCAAAGGTGGCCATCCGCGTGACGCTGGCAGAAAGTTCGCGAAAGTTA

2701 CCGCGCCATGTTGCCTGGGGAGAGGTCGCAAAAGCCAACCAGGCGCGCCGCGCTTCGGTG

2761 TTAAAACGCACGCTGTCGCCAGTGAGTGAGGCGTGGCGCTCCACTTCATAATCCAGGTTC
```

```
2821 GGTTCAATATCTTCCTGGCGCTGGCGTAGACCCGGCAGGGTGAAGGTCCAGAGATTGATC

                BglII
2881 CGCGCGTACAGATCT
```

Figure 3. Nucleotide sequence 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 (S. D.) are indicated by solid underlining. The transcription initiation sites for the *glpD* (61) and *glpE* (88) 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, 61). Inverted repeat sequences are signified by the converging dashed arrows above the sequence. The amino acid sequences at the amino termini of GlpG and GlpR were determined (dashed underlining). The sequence data reported here have been submitted to GenBank and assigned accession number M96795.

The apparent molecular weight of GlpE encoded by pGZ105 was 12,700, which matched the size observed before (75, 87); while that encoded by pGZ106 was only 10,200, as determined by SDS-PAGE (Fig. 4). An open reading frame of 108 codons (residues 373-699) was therefore identified for GlpE (predicted MW=12,082). A weak ribosome binding site (AAGAGA) was found 5 base pairs upstream of the AUG initiation codon (Fig. 3). The predicted GlpE polypeptide is relatively hydrophilic (56% polar and charged residues) and acidic (calculated pI is 4.3). The GlpE protein was found in the soluble fraction of *E. coli* by performing subcellular fractionation as described in Experimental Procedures (Fig. 4). It is most likely that GlpE is a cytoplasmic protein (not periplasmic), since there is no predicted signal sequence that would be required for secretion to the periplasmic space.

Determination of the *glpG* start codon and analysis of the deduced amino acid sequence of GlpG: The open reading frame corresponding to GlpG was found downstream of the *glpE* gene (Fig. 3). Several possible start codons were found in the *glpG* region. These start codons include: the AUG's at position 744-746 and 801-803, as well the GUG at position 897-899.

A similar strategy as that used above for *glpE* was employed to overproduce proteins from the proposed AUG (801-803) and GUG (897-899) codons. No protein overproduction was observed, however, upon induction of strain BL21(DE3) bearing pGZ110 [AUG(801-803)] or pGZ111 [GUG(897-899)]. Another strategy was tried which used primers 91905 and 24551 to amplify the *glpG* region including all the possible start codons. The PCR product was subsequently cloned between the *Bam*HI and *Hind*III sites of pGEM3Z (Promega) such that *glpG* is located downstream of the T7 promoter, resulting in pGZ125. After IPTG induction and subcellular fractionation of BL21(DE3)

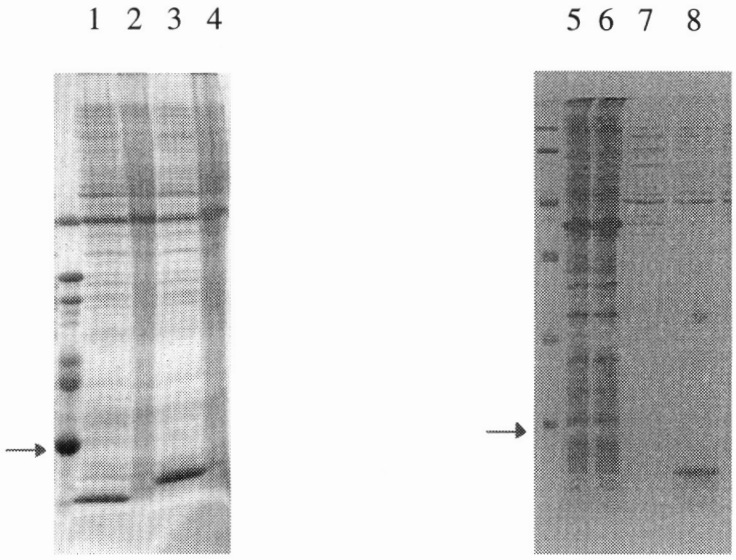


Figure 4. Overexpression and subcellular localization of GlpE. The left panel shows the overexpression of GlpE encoded by pGZ105 (*lane 3*, induced; *lane 4* uninduced) and pGZ106 (*lane 1*, induced; *lane 2*, uninduced). The right panel shows the subcellular localization of GlpE encoded by pGZ105. Both are 15% SDS polyacrylamide gels. *Lanes 5* and *6* contain the membrane fractions of BL21(DE3) cells harboring pT7-7 and pGZ105. *Lane 7* and *8* contain the soluble fractions of BL21(DE3) carrying pT7-7 and pGZ105, respectively. Arrow in both panels mark the 14 kDa molecular weight standard.

cells harboring pGZ125, an overproduced inner membrane protein with an apparent size of 26.2 kDa was found on an SDS polyacrylamide gel (Fig. 5). The apparent size agrees well with the size of the GlpG protein found in minicells and maxicells (26 kDa, 75, 87). Another plasmid, pGZ126 that bears the entire *glpEGR* operon was also found to overproduce the 26.2 kDa protein in its inner membrane fraction (data not shown).

The 26.2 kDa inner membrane protein encoded by pGZ126 was isolated and its N-terminal amino acid sequence was obtained (Experimental Procedures). The N-terminal amino acid residues sequenced matched the predicted GlpG sequence. TABLE 6 shows the results of sequence analysis from which the start codon of *glpG* is identified as the AUG at position 744-746 (Fig. 3).

The GlpG protein, therefore, contains 276 amino acids and has a calculated molecular weight of 31,278. Analysis of its deduced amino acid composition indicated that GlpG is a hydrophobic (47% hydrophobic residues) protein which may possess 6 membrane spanning segments (Fig. 6). The discrepancy between the predicted and apparent MW of GlpG on SDS gels may be due to anomalous binding of SDS to this hydrophobic membrane protein, which was also observed in the case of the GlpF and GlpT proteins (13, 89). Another noteworthy feature of GlpG is its basic nature (calculated pI is 9.3).

Figure 5. Overexpression and subcellular localization of GlpG. Extracts from strain BL21(DE3) harboring pGEM3Z (V) or pGZ125 (G) were fractionated and analyzed on a 15% SDS polyacrylamide gel. An inner membrane protein at 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 weight standards.

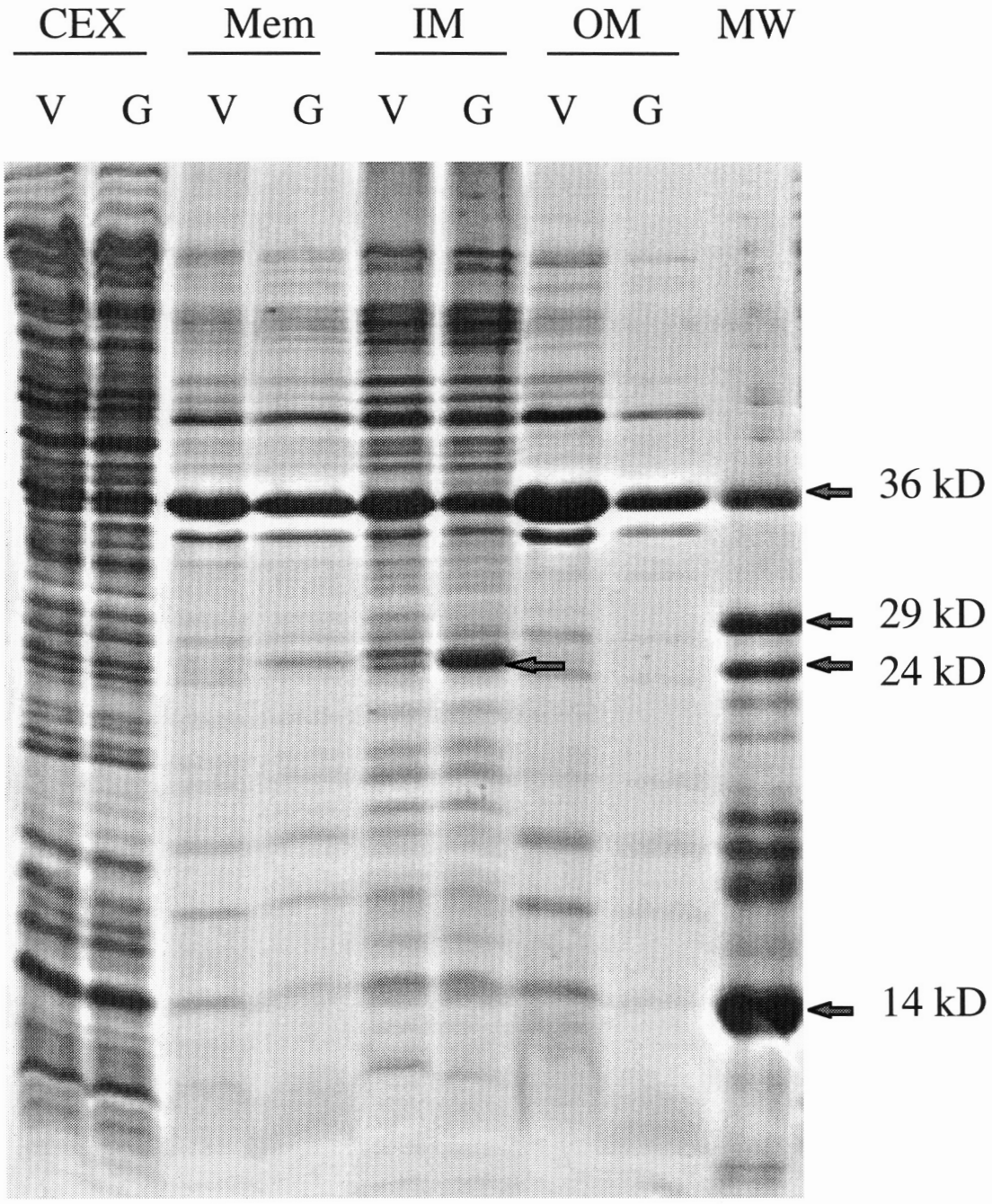

CEX
Mem
IM
OM
MW
V
G
V
G
V
G
V
G
36 kD
29 kD
24 kD
14 kD

TABLE 6. N-terminal amino acid sequence analysis of GlpG

cycle #	predicted amino acid	amino acid identified	pmol detected
1	M	M	31.0
2	L	L	32.9
3	M	M	30.0
4	I	I	27.2
5	T	T	19.6
6	S	S	10.0
7	F	F	23.1
8	A	A	21.2

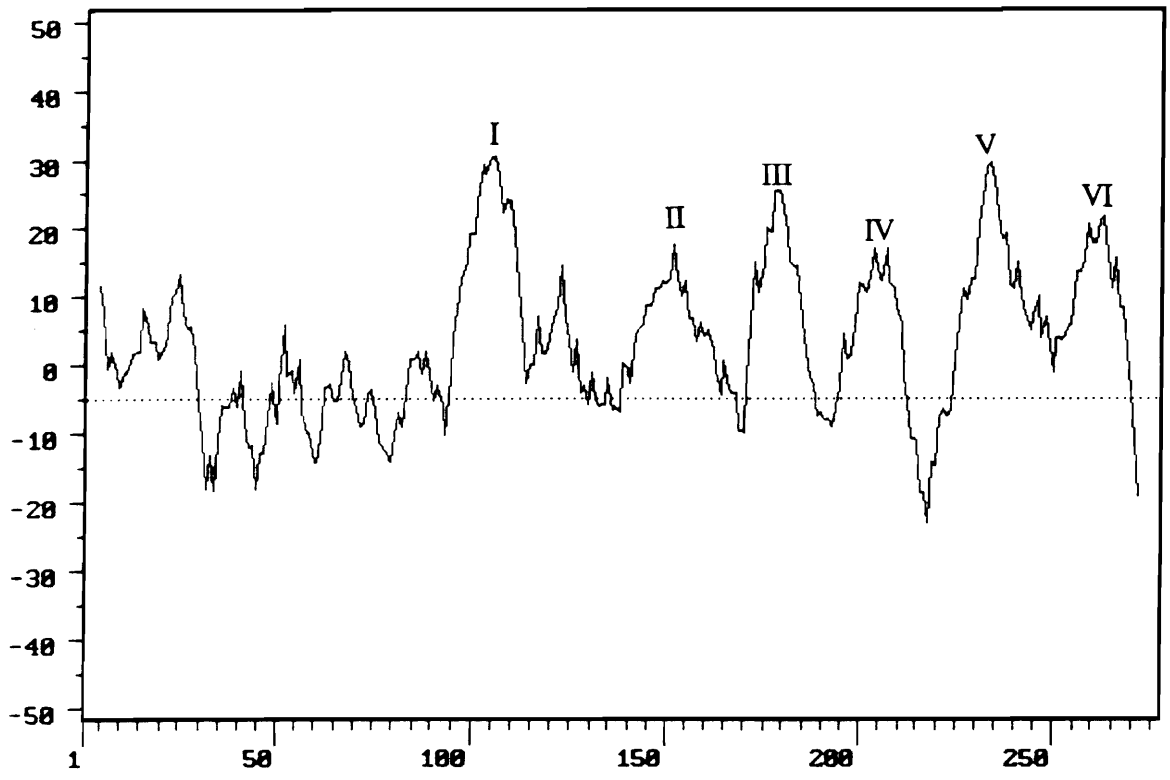


Figure 6. Hydropathy index computation for GlpG. PC/GENE program SOAP was used. The hydropathy scale composed by Kyte and Doolittle (90) was employed to display the hydrophobicity of GlpG using a window of 9 amino acids. Roman numbers indicate the possible membrane-spanning segments in GlpG based on the hydrophobicity prediction.

Determination of the *glpE* and *glpG* internal translational reading frames by construction of *lacZ* translational fusions: To empirically verify the sequence of *glpE* and *glpG*, the translational reading frame near the center of each gene was determined. Primers 28341 and 34915 were used to amplify a segment of *glpE*, which was fused in-frame with the *lacZ* gene of pSH76 according to the deduced translational reading frame of *glpE* (Fig. 7). The resulting pGZ103 had the *glpE* portion from upstream of its start codon to position 690 (Fig. 3), and contains all but the last two codons. Similarly, pGZ104 was constructed by fusing the *glpG* portion from upstream of its start codon to position 1238, and is predicted to contain the first 165 codons fused in-frame with the *lacZ* gene of pSH76. Expression of each fusion was controlled by P_R of λ phage. Both constructs were confirmed by DNA sequence analysis. Strain TL73 (Δ*lacZ recA*) harboring pGZ103 or pGZ104 was grown at 30°C until the OD_{600} reached 0.4. Induction at 42°C, which results in inactivation of λ repressor cI857 controlling P_R, was carried out for 4 hours. The fact that pGZ103 and pGZ104 encoded chimerical proteins (tested by SDS-PAGE, data not shown) with high β-galactosidase activity (TABLE 7) indicates that the predicted reading frames to nucleotide 690 for *glpE* and to nucleotide 1238 for *glpG* were correct.

Figure 7. Construction of pGZ103 (*glpE-lacZ*) and pGZ104 (*glpG-lacZ*) translational fusions. Stop codons (indicated by ▯) upstream of the *glpE* and *glpG* start codons were designed for termination of the translational read-through from the *glpD* gene of pSH76. Both PCR products of *glpE* and *glpG* bear *Bgl*II (Bg) and *Bam*HI (B) sites at the 5' and 3' ends, respectively. Both sites were used for cloning into the *Bam*HI site of vector pSH76. cI857 encodes the heat-labile repressor for the λ promoter P_R. Abbreviations used: E_N, the portion of *glpE* encoding the N-terminal 106 amino acids; G_N, the portion of *glpG* encoding the N-terminal 165 amino acids; D', the truncated *glpD* gene.

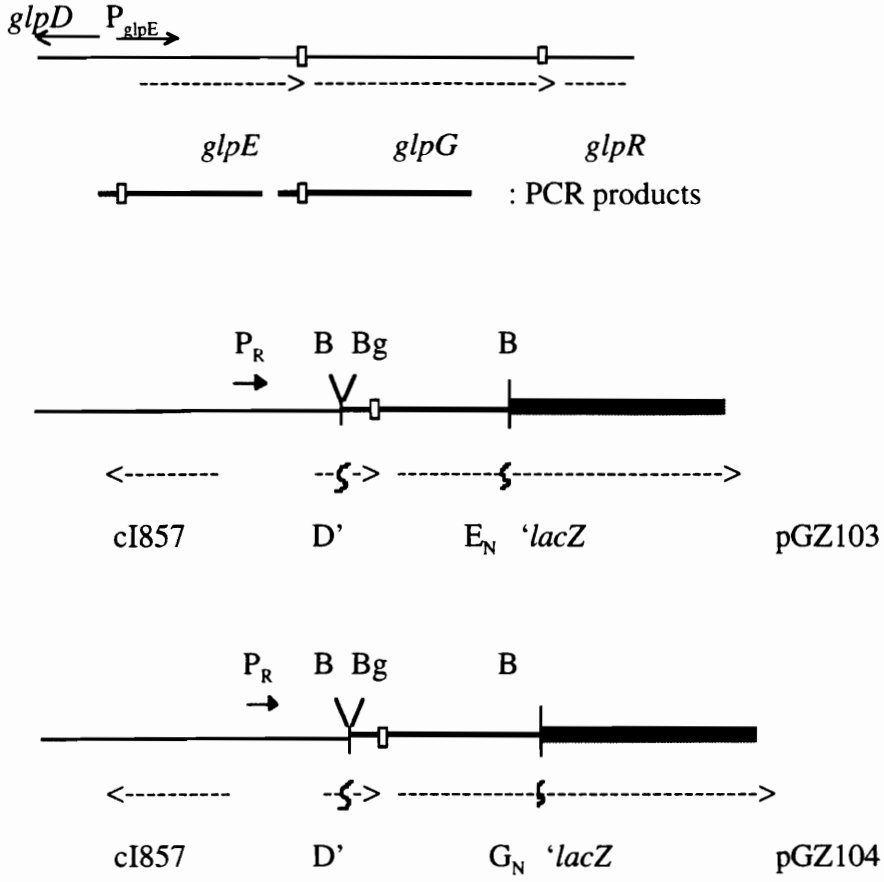

glpD
P glpE
glpE
glpG
glpR
: PCR products
P R
B Bg
B
cI857
D'
E N
'lacZ
pGZ103
P R
B Bg
B
cI857
D'
G N
'lacZ
pGZ104

TABLE 7. β-galactosidase activities of *glpE-lacZ* and *glpG-lacZ* hybrid proteins

Strain [fusion]	Specific Activity at 42°C
TL73 pGZ102 [1]	<10
TL73 pGZ103 [*glpE-lacZ*]	660
TL73 pGZ104 [*glpG-lacZ*]	970

[1] pGZ102 has the *glpG* portion in a reversed orientation, and is used as a negative control.

Determination of the translational reading frame at the C-terminus of GlpR: The initiation codon for the *glpR* gene is found 16 bp downstream of the termination codon for *glpG* (Fig. 3). The sequence that follows contains 252 codons and would be translated into a protein with a calculated molecular weight of 28,048. The beginning of the open reading frame for *glpR* was verified by determination of the N-terminal amino acid sequence of the purified *glp* repressor (47). The amino acid sequence matched that predicted by the nucleotide sequence. 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 (87).

The predicted size of GlpR subunit (28 kDa) is smaller than that found on SDS gels, which is about 30 kDa. Also, the nucleotide sequence predicted a C-terminal cysteine residue, which is uncommon among proteins in *E. coli*. There are several potential stop codons in other reading frames downstream of the putative stop codon of *glpR* (80). Furthermore, there are discrepancies among the sequences submitted by different research groups to the GenBank (47). The sequences reported may contain reading frame shifts in the C-terminal coding region due to the presence of GC-rich compression regions. Therefore, an approach other than DNA sequencing was necessary to resolve the discrepancies.

A PCR product of *glpR* (using primer 69620 and 50472) was cloned between the *Nde*I and *Sal*I sites of pT7-7, resulting in pGZ115. The *Sal*I site was created immediately upstream of the proposed UGA stop codon, and a *Sal*I-*Hind*III linker with an engineered thrombin site was inserted between *Sal*I and *Hind*III of pGZ115 (Fig. 8). The new plasmid, pGZ117, is predicted to encode a protein with a thrombin cleavage site. pGZ117 was then digested with *Sal*I or *Acc*I, which recognize the same site. The resulting 4 nucleotide or 2 nucleotide overhangs were filled in followed by blunt-end ligation

reactions. As a result, the plus one and plus two frame shift variants were created immediately upstream of the oligonucleotide pair, resulting in plasmids pGZ118 and pGZ119 (Fig. 8). The above manipulations were predicted to create new *Pvu*I or *Nru*I restriction sites in pGZ118 and pGZ119, respectively. This was verified by digestion with the appropriate enzymes. The translational reading frame of the *glpR* gene in the C-terminal coding region could then be deduced based on the susceptibility of the three gene products to cleavage by thrombin.

The three plasmids were separately transformed into strain BL21(DE3), and extracts were obtained from cultures induced by IPTG as described in the Experimental Procedures. The crude extracts (5 μg) from each of the three strains were incubated with thrombin before being analyzed by electrophoresis on a 15% polyacrylamide gel (Experimental Procedures). After electrophoresis, a Western blot was performed to detect the sizes and thrombin susceptibilities of each GlpR.

As shown in Fig. 9, GlpR encoded by pGZ117 was cleaved by thrombin to yield a product that comigrated with authentic GlpR. GlpR's encoded by pGZ118 or pGZ119 were slightly larger than authentic GlpR and were not cleaved by thrombin. Since the thrombin site encoded by pGZ117 is fused in the same reading frame predicted by the nucleotide sequence, the results prove that the reading frame predicted for *glpR* is correct as deduced from its DNA sequence, and that cysteine is in fact the C-terminal amino acid residue of the *glp* repressor.

Figure 8. Strategy used to create +1 and +2 frame shifts upstream of the thrombin site. Oligonucleotide pair that encodes a thrombin cleavage site (indicated by the arrow) in the indicated reading frame is shown in Panel A. This oligonucleotide pair also serves as a *Sal*I-*Hind*III linker for cloning purpose. Panel B shows the manipulations used to create the frame-shift derivatives of pGZ117. The generation of *Pvu*I or *Nru*I sites was used to confirm the successful fill-in and ligation reactions. The bold gray line represents the oligonucleotide pair that encodes the thrombin site. The abbreviations used are: P, T7 promoter ϕ10; R, *glpR*.

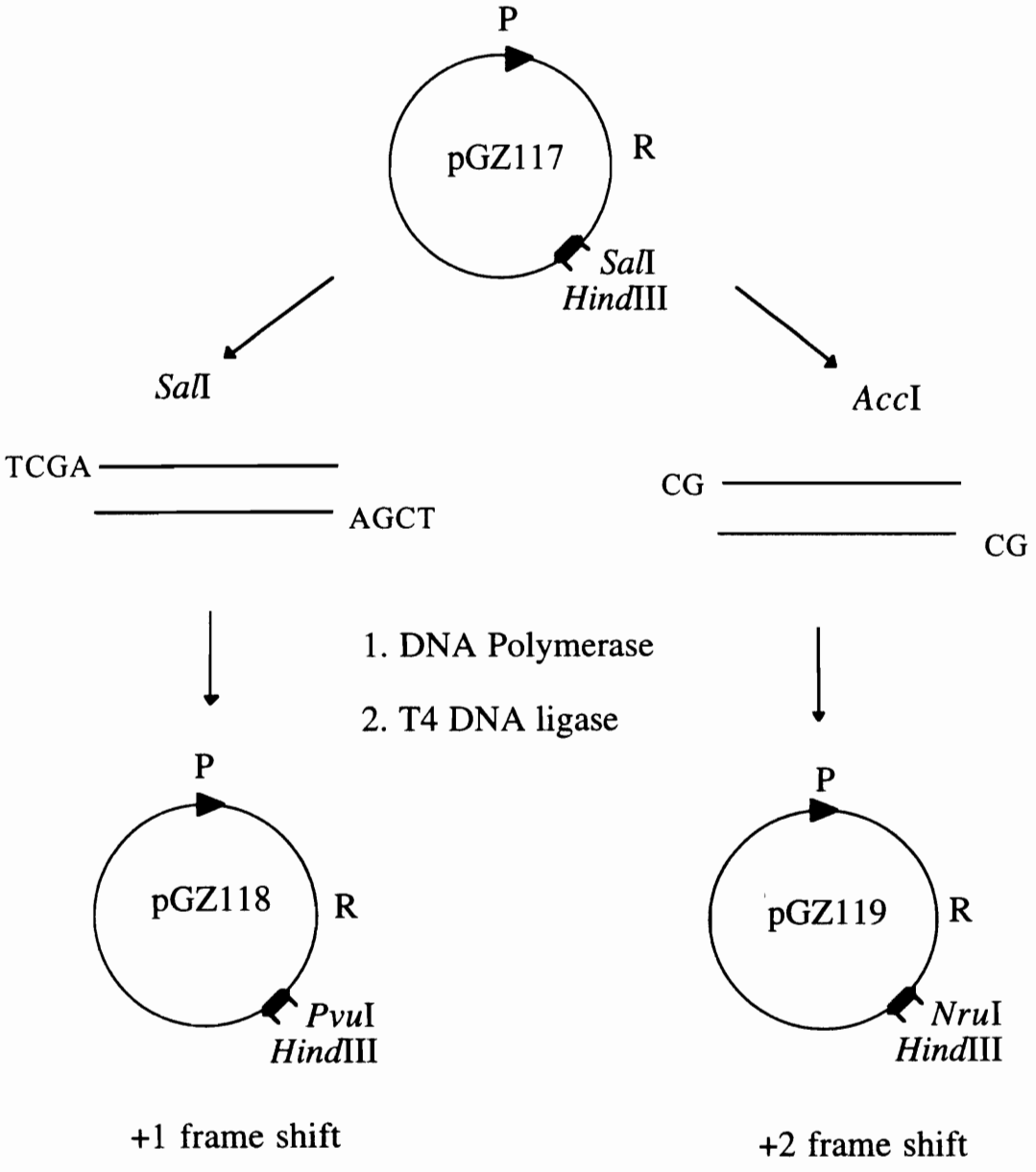
A: *Sal*I *Hind*III

TC GAC CTG GTT CCG CGT GGA TCC $(CAC)_5$ CAT TA

G GAC CAA GGC GCA CCT AGG $(GTG)_5$ GTA ATT CGA

D L V P R ↑ G S H_5 H

B :



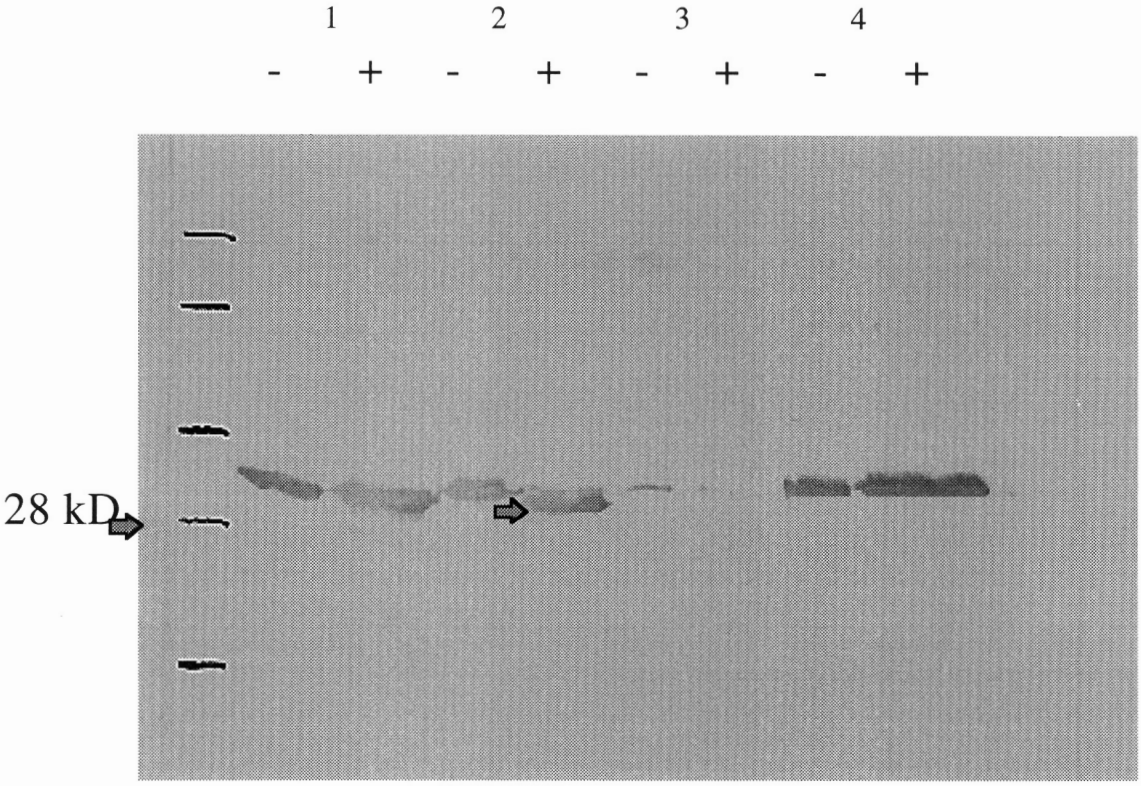


Figure 9. Sensitivities of GlpR derivatives to thrombin cleavage. A Western blot was performed using polyclonal antibodies specific for GlpR, with goat anti-rabbit IgG conjugated with alkaline phosphatase serving as secondary antibody. The pairs of *Lanes* labeled *1, 2, 3,* and *4* contain crude extracts of BL21(DE3) cells containing GlpR encoded by pGZ115, pGZ117, pGZ118, and pGZ119. pGZ115 bears *glpR* without the C-terminal oligo-nucleotide pair. Samples were incubated without or with thrombin, as indicated (-, +). Only GlpR encoded by pGZ117 can be cleaved by thrombin as indicated by the inset arrow.

Purification of GlpE: A pair of oligonucleotides (#04145 and #105276) that also serves as a *Sal*I and *Hind*III linker was ligated between the *Sal*I and *Hind*III sites of pGZ105. The resulting pGZ107 encodes GlpE with a C-terminal hexahistidine tag preceded immediately by an engineered thrombin cleavage site. On the other hand, a similar oligonucleotide pair (#03956 and #03921 that also makes an *Nde*I linker) was ligated in the *Nde*I site of pGZ105. The resulting plasmid, pGZ105N encodes GlpE with an N-terminal hexahistidine tag upstream of an engineered thrombin site. Strain BL21(DE3) harboring pGZ107 or pGZ105N was induced as described in Experimental Procedures. Only pGZ107 encoding GlpE with the C-terminal his tag apparently overproduced the engineered GlpE protein. This result is similar to the case of polyhistidine-tagged PlsX, where it was also found that the N-terminal-tagged PlsX was not overexpressed while the C-terminal-tagged protein was successfully overproduced (Ali Bhattacharya, personal communication). It is possible that artificial sequences at the N-terminus influence the initial ribosome entry, and the consequent translational efficiency. Unlike GlpR, where a large portion of overproduced protein fractionated with the inclusion bodies, the cell tolerated GlpE at levels as high as nearly 30% of the total soluble protein. After a one-step IMAC procedure (Experimental Procedures), GlpE was purified to near homogeneity as shown in Fig. 10. GlpE with the 1.8 kDa thrombin plus polyhistidine tag has an apparent size of approximately 14 kDa on a 16% SDS polyacrylamide gel. The same IMAC strategy was tried for the purification of polyhistidine-tagged GlpR. However, neither the N-terminal nor the C-terminal histidine-tagged GlpR bound to Ni^{2+}, Cu^{2+}, or Zn^{2+} columns under a variety of buffer conditions (data not shown).

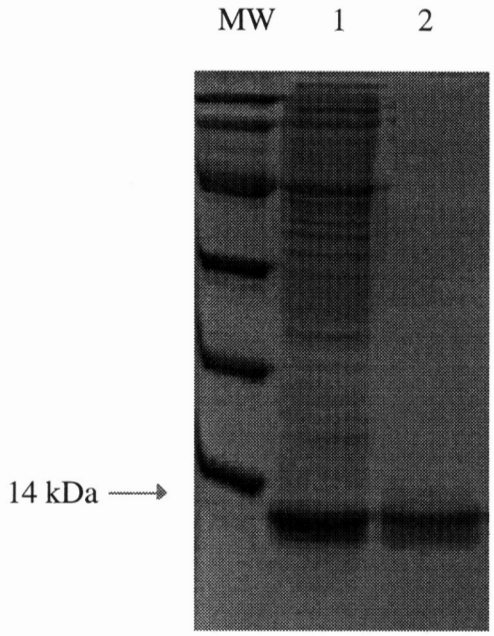


Figure 10. Purification of GlpE protein by a one-step IMAC procedure. *Lane 1* is the soluble fraction of BL21(DE3)(pGZ107). Overexpressed GlpE accounts for more than 30% of the total soluble protein. A total of 5 μg protein was loaded in *Lane 1*. *Lane 2* contains 1 μg of GlpE eluted from the column. The SDS polyacrylamide gel is 16%.

Modification of the GlpR purification scheme: The published procedure used to purify GlpR includes polyethyleneimine (PEI) and ammonium sulfate fractionation, phosphocellulose chromatography, and DEAE chromatography (40). The PEI fractionation step is dependent on protein concentration and ionic strength, and therefore, requires pre-running of the experiment for different crude extracts. This time-consuming step can be eliminated in the new procedure described below. Preparation of the cytoplasmic fraction containing overproduced GlpR is as described in Experimental Procedures. Nucleic acid in the crude extract was sheered by sonication. The entire procedure was carried out at 4 °C and is summarized in TABLE 8.

DE-52 chromatography: Usually, 25 ml of DE-52 resin was used for 150 mg of crude extract protein (about 8 mg/ml). The column (2.0 x 10 cm) was first equilibrated with buffer A, which contained 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1 mM DTT, and 150 mM NaCl. After loading and extensive washing, the column was developed by a one-step elution using 100 ml buffer A with 500 mM NaCl. Fractions containing the glycerol-P binding activity were pooled. This step resulted in approximately 10-fold purification of GlpR (TABLE 8).

Ammonium sulfate fractionation: Saturated and pH balanced (pH set to 7.5 by addition of ammonium hydroxide) ammonium sulfate was slowly added to the DE-52 pool with continuous stirring during 30 minutes. The first cut was made at 20% saturation. After incubation in the cold for another 20 minutes, the suspension was centrifuged at 7,500 g for 15 minutes. The supernatant was saved for the second precipitation, which employed ammonium sulfate at 50% saturation for 20 min of incubation prior to centrifugation as above. The final pellet was dissolved in buffer A (one-half volume of the crude extract).

TABLE 8. Purification of GlpR

Fraction	Protein (mg)	Activity (pmol)	Specific activity (pmol/mg)	Recovery (%)	Enrichment (fold)
Crude Extract	167	13,000	78	100	1
DE-52 Fraction I	18	15,800	880	100	11
Ammonium sulfate Fraction	as a condensation step, total protein was not measured				
Heparin Fraction	1.2	8200	7080	52	91
DE-52 Fraction II	0.67	6560	9790	42	126

Heparin agarose chromatography: Usually 7 ml of heparin agarose resin was used for 15 mg protein from the above step. The heparin column (1.3 x 6 cm) was first equilibrated with buffer A. After loading and washing, the column was developed with a 50 ml linear gradient of 150 mM to 400 mM NaCl in buffer A. Glycerol-P-binding activity was eluted in fractions containing 0.22 to 0.28 M NaCl. GlpR from this step was approximately 90% pure and is suitable for many biochemical assays. To further purify GlpR, ammonium sulfate precipitation at 50% saturation was repeated, and the final solution was applied to a second DE-52 column.

Second DE-52 chromatography: The second DE-52 column was about one-fifth the volume of the first one. Elution was performed using a linear gradient of 150 mM to 400 mM NaCl in buffer A. Fractions from each step were analyzed on a 15% polyacrylamide gel, which are shown in Fig. 11.

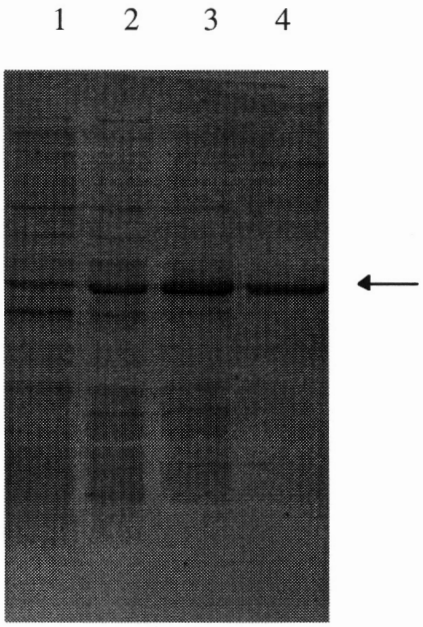


Figure 11. Purification of GlpR from strain GZ544(pGZ114). *Lanes 1* to *3* each contains approximately 4 μg of protein. They are crude extract, DE-52, and Heparin chromatography fractions, respectively. *Lane 4* contains 1 μg of the final product (inset arrow shows GlpR) from elution of the second DE-52 chromatography.

Estimation of the dissociation constant of GlpR for *glp* operator DNA: A Crude extract from strain BL21(DE3)(pSH79) was used as the source of repressor to titrate [^{32}P]-labeled DNA containing a single *glp* operator (Experimental Procedures). As negative controls, a P33E,Q34M altered form of GlpR encoded by pSY2-A (47) and an R38A altered form encoded by pSY2-9 (47) were employed in the same experiment. All repressors were overexpressed (Fig. 12), and binding to operator DNA was assessed by using gel mobility shift assays. As shown in Fig. 13, operator DNA was shifted by the wild type GlpR at about 10^{-8} M GlpR. No obvious DNA retardation was seen for the PQ-EM or the R-A altered forms of repressor even at 10^{-6} M (data not shown). A Shimadzu CS-9000 densitometer was used for quantification by scanning the autoradiogram. Percentages of free DNA versus concentrations of GlpR tetramer were plotted in Fig. 14.

The dissociation constant of the repressor-operator complex was estimated by using the following formula:

$$K_d = \frac{\text{[free DNA] [free repressor tetramer]}}{\text{[DNA-repressor complex]}}$$

Since the concentration of operator DNA is negligible compared with the concentration of GlpR tetramer in these experiments (0.1 vs. 10 nM), [free repressor tetramer] = [total repressor tetramer] - [DNA-repressor complex] can be simplified as [free repressor tetramer] = [total repressor tetramer]. At the point where half of the DNA is shifted, that is [free DNA] = [DNA-repressor complex], the above formula for K_d becomes:

$$K_d = \frac{\text{[free DNA] [free repressor tetramer]}}{\text{[DNA-repressor complex]}}$$

K_d = [free repressor tetramer]

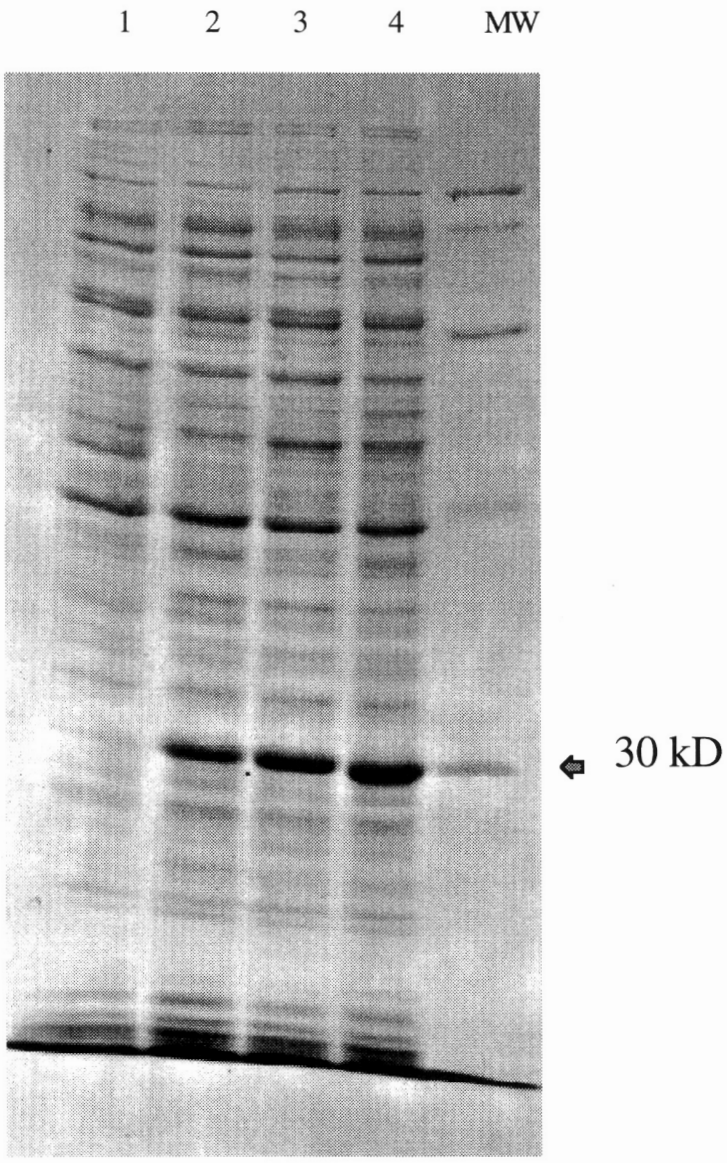


Figure 12. Overexpression of *glp* repressor and its variants. The same amount of protein (25 μg) was loaded in each lane. A 15% polyacrylamide gel was used for analysis of extracts from BL21(DE3) cells with pBluescript, wild-type GlpR encoded by pSH79, $GlpR^{P33E,Q34M}$ encoded by pSY2-A, and $GlpR^{R38A}$ encoded by pSY2-9 (*Lanes 1* to *4*), respectively.

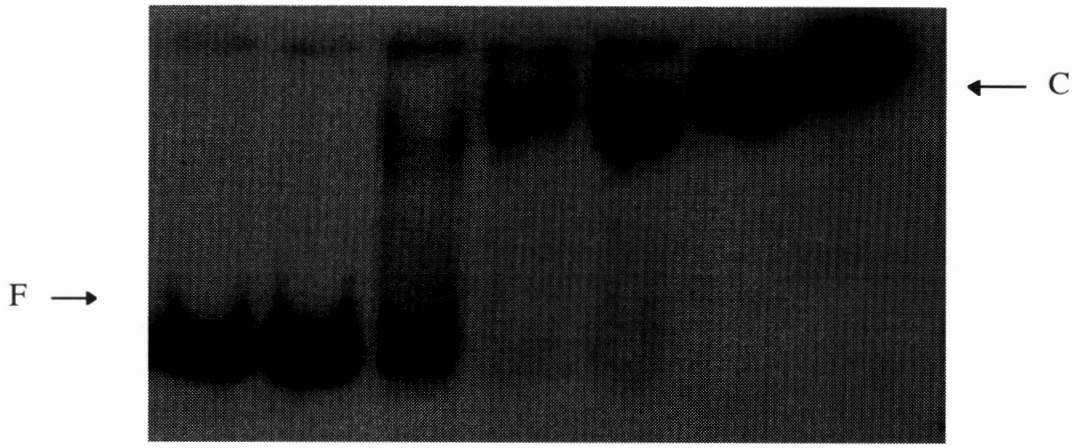


Figure 13. Titration of single *glp* operator consensus DNA by GlpR. Increasing concentrations of GlpR extract was used (from left to right): $1.0x10^{-9}$, $5.0x10^{-9}$, $2.5x10^{-8}$, $1.0x10^{-7}$, $2.0x10^{-7}$, $4.0x10^{-7}$, and $2.0x10^{-6}$ M. Concentrations were determined as GlpR tetramers present in the extract (Fig. 12) under the assumption that GlpR monomer is about 8% of the total protein. Abbreviations: F, free operator DNA; C, repressor-DNA complex.

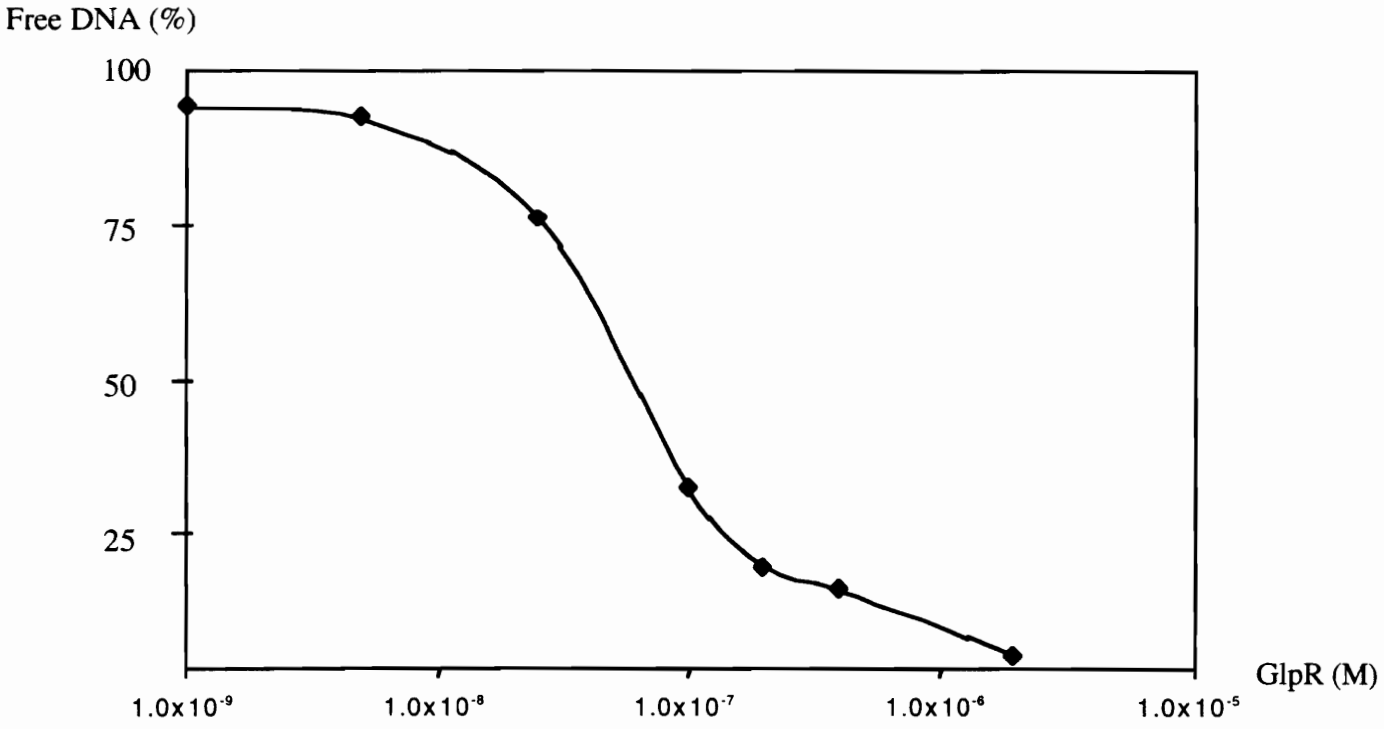


Figure 14. Estimation of the K_d of the *glp* repressor-operator complex. The percentage of free DNA was plotted against the concentration of the GlpR tetramers. The concentration of GlpR tetramer that results in 50% of the operator DNA being shifted is estimated as the K_d of the repressor-operator complex.

K_d = [total repressor tetramer]

Therefore, the apparent K_d for the repressor-operator complex is approximately equal to the concentration of repressor tetramer required for binding half of the operator DNA. By this means, the K_d of GlpR for a single *glp* consensus operator was estimated to be 50 nM (Fig. 14).

PART II: FUNCTIONAL ANALYSIS OF THE GLPE AND GLPG PROTEINS OF *ESCHERICHIA COLI* K-12

The presence of two upstream genes *glpE* and *glpG* in the same operon with *glpR* is not common among *E. coli* genes encoding regulatory proteins. One exception is the *gatYZABCDR* operon (91), in which there are genes encoding EII^{Gat} subunits (*gatABC*), dehydrogenase (*gatD*), and hydrolase (*gatY*) upstream of the repressor gene *gatR*. In this case, the function of the *gatZ* gene product is still unknown. To reveal the possible roles GlpE and GlpG play in *E. coli*, a reasonable starting point is to investigate the impact of GlpE and GlpG on the GlpR-mediated repression and glycerol-P induction of the *glp* regulon.

Amino acid sequence similarities between GlpE and GlpG and other proteins: To gain insight into the functions of the GlpE and GlpG proteins, their amino acid sequences were compared with sequences in GenBank. The best match to both proteins was found in the Gram-negative bacterium *Haemophilus influenzae* (Fig. 15), whose entire sequence was reported recently (92). It was found that *H. influenzae* has all the *glp* genes except *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,* and possibly in the same operon with *glpR*; whereas *glpE* is at another location, adjacent to *tpiA* (92) encoding triosephosphate isomerase.

Similarities between GlpG or GlpE and other polypeptides in GenBank were significantly less compared with the above-mentioned examples (data not shown). Those sequences displaying similarity to GlpG, however, were mainly enzymes of electron transport chains, for instance, NADH dehydrogenase of *Candida parapsilosis*, and NADH-ubiquinone oxidoreductase of *Cyprinus carpio*. Database searches for proteins similar to

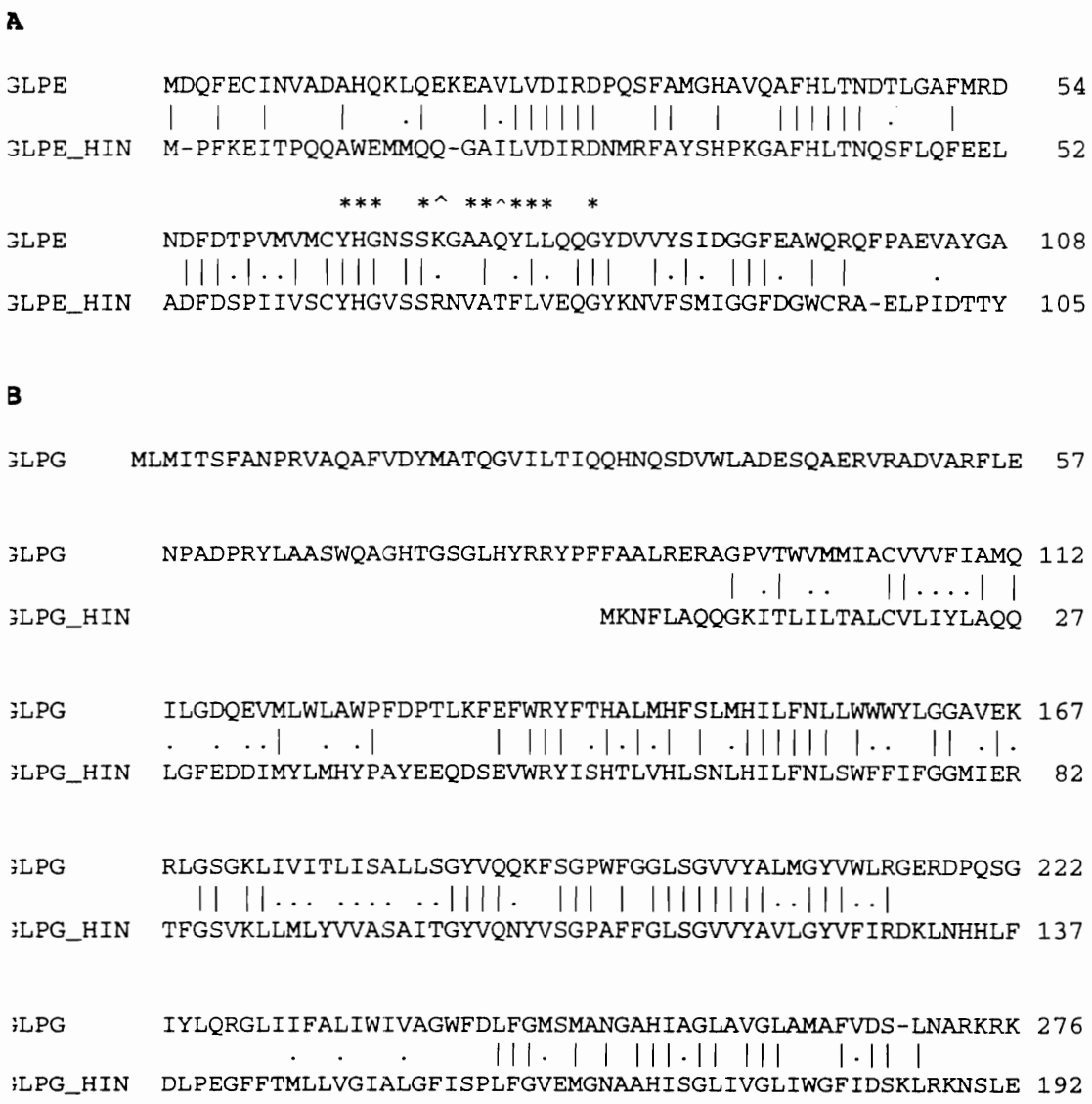


Figure 15. Amino acid sequence similarities between GlpE and GlpG proteins and their counterparts from *H. influenzae*. Also indicated in Panel A is the region (from Y66 to G82) conserved in the SH2 domain of tyrosine phosphatases/kinases. * and ^ indicate residues of GlpE that are identical or similar, respectively, to those found in the SH2 domain.

GlpE revealed several protein tyrosine kinases from eukaryotes. The region with the most similarity (Fig. 15) was found to be a helix in the highly conserved scr2 homologous domain (SH2) of these enzymes. Enzymes having this domain and showing similarity with GlpE include the GAG-ABL-POL polyprotein from feline sarcoma virus (93), and the human proto-oncoprotein tyrosine-protein kinase ABL (94). The SH2 domains are known to bind phosphorylated tyrosine residues. However, the specific role of the first 17 amino acids which are also conserved in GlpE (10/17 identities; 12/17 similarity, Fig. 15) is still unknown. Taken together, no proteins with known functions have extensive similarities to either GlpE or GlpG protein.

Study of the effect of GlpE/GlpG on *glpK-lacZ* gene expression by using the Heidelberg multifunctional plasmid system: The pUH multifunctional plasmid system (R. Lutz and H. Bujard, University of Heidelberg, Germany) was designed for the independently controlled expression of one, two, or three different proteins in the same cell. In my case, I studied the effects of GlpE, GlpG, and GlpEG with respect to GlpR on *glp* gene regulation. To prepare *glpR* for this plasmid system, it was amplified by PCR using expression vector pGZ114R as template. The PCR product contains a strong ribosome binding site upstream of the *glpR* gene. It was then cloned between the *EcoR*I and *Xba*I sites of pUHE1P$_{lac}$, resulting in pGZ158 (Ap^R, ColE1 origin, *glpR* under the control of P$_{lac}$). The entire *glpR* gene in pGZ158 has been sequenced and found to match that reported previously (47). By a similar strategy, *glpE*, *glpG*, and *glpEG* were separately cloned into another vector pUHA4P$_{N25}$, resulting in pGZ154, pGZ148, and pGZ150, respectively (Cm^R, p15A origin, P$_{N25}$ promoter, TABLE 5). Thus, *glpR* was cloned in one plasmid, and *glpE*, *glpG*, or *glpEG* was cloned in a compatible plasmid with a differently controlled promoter and a different antibiotic resistant marker.

Strain NZ45 (Δ*glpEGR recA glpK-lacZ*) was used as the host for testing the repression exerted by combinations of *glpE*, *glpG* or *glpEG* together with *glpR*. The reason *glpK* was selected from the members of the *glp* regulon is that it is most sensitive to negative regulation by the repressor. NZ45 also contains the *lacI*^{q} and *tetR* genes, which encode repressors for control of P_{lac} and P_{N25}, respectively. The genes cloned downstream of P_{lac} or P_{N25} are therefore subject to induction by exogenously-added IPTG or tetracycline, respectively.

Control of the *glpK-lacZ* fusion was determined by measuring β-galactosidase activity in strain NZ45 harboring different pairs of the above-mentioned plasmids. As shown in TABLE 9, even without induction of repressor synthesis by IPTG, the amount of GlpR expressed by pGZ158 was sufficient to achieve more than 200-fold repression of *glpK-lacZ* relative to the constitutive level of expression in NZ45 cell. The presence of GlpE and GlpG increased repression by a factor of two. To test the possibility that plasmids with the high copy number ColE1 origin (such as pGZ158) produced higher than normal levels of GlpR in the cell, the *glpR* cassette was moved into pUHS4P_{lac}. The resulting plasmid, pGZ163, had the pSC101 origin (95) and about one-tenth the copy number compared with pGZ158. Consequently, as shown in TABLE 10, there was an approximate 2.5-fold increase in the repressed level of β-galactosidase activity compared with pGZ158. This suggested that repression of *glpK-lacZ* responded to the amount of GlpR *in vivo*. The relative amount of GlpR encoded by the chromosome, pGZ163, and pGZ158 was assessed by Western analysis (Fig. 16). Strain NZ45 was used as the negative control and as the host for plasmids encoding GlpR; GD31 was the source for chromosomally-encoded GlpR. A Shimadzu CS-9000 densitometer was used for

TABLE 9. Repression of *glpK-lacZ* in strain NZ45 by GlpR (high-copy-number plasmid)[1]

Plasmids [genotype]	β-galactosidase activity	
	- glycerol-P	+ glycerol-P
GD31[2]	50	560
pUHE1P$_{lac}$ [vector]	3200±100	850±90
pGZ158 [*glpR*$^+$]	13±1	1000±30
pGZ158 pGZ150 [*glpR*$^+$] [*glpEG*$^+$]	7.3±0.1	500±50

[1] AB minimal medium supplemented with 0.2% casamino acids and 2 μg/ml thiamine was used in this and the following experiments unless otherwise specified. Data in this table were obtained from three independent experiments.

[2] The top line gives the activities for strain GD31 (chromosomal *glpEGR*$^+$, with the same *glpK-lacZ* fusion as strain NZ45). Results were obtained from Ningyue Zhao (unpublished), who used the same conditions as those employed in this study.

TABLE 10. Effect of GlpE/G on GlpR-mediated repression of *glpK-lacZ* in strain NZ45

Plasmids[1] [proteins encoded by plasmid]	β-galactosidase activity[2]			
	no IPTG		5 μM IPTG	
	no Tc	2 μg/ml Tc	no Tc	2 μg/ml Tc
pGZ163 pUHA4 [R]	29±1	35±0	12±2	14±0
pGZ163 pGZ154 [R/E]	27±6	31±13	18±8	31±0
pGZ163 pGZ148 [R/G]	39±7	51±3	30±3	30±1
pGZ163 pGZ150 [R/EG]	39±1	19±2	21±2	7.6±1.6

[1] Plasmids with low copy-number were used for GlpR expression.

[2] Data shown were obtained from two or more independent experiments. IPTG was used to induce the production of GlpR, whereas tetracycline (Tc) was for induction of GlpE, GlpG, and/or GlpEG.

quantification by scanning of the individual bands, and the relative amount of GlpR was compared to a pure *glp* repressor standard of known amount. By this means, the amount of GlpR encoded by pGZ163 (even without IPTG induction) was estimated to be about 0.14% (650 ng in lane 6 of Fig. 16) of the total cellular protein. GlpR encoded by pGZ158 was about ten-fold the amount of that encoded by pGZ163, which is about 1.4% of the total protein. The relative amount of GlpR encoded by pGZ158 (ColE1 origin) and pGZ163 (pSC101 origin) was proportional to the relative copy-number of each plasmid. Upon induction with 5 μM IPTG, the amount of GlpR encoded by pGZ163 increased about 3-fold to approximately 0.4% of the total protein. GlpE/G did not influence the quantity of GlpR in the cell, thus the 2-fold increase in repression of *glpK-lacZ* observed when GlpE/G was present is not due to a change in the *in vivo* level of GlpR. Surprisely, chromosomally-encoded GlpR in strain GD31 was not detectable under the conditions used. Therefore, the amount of GlpR in lane 2 of Fig. 16 must be below 100 ng, which indicates that chromosomally-encoded GlpR is less than 0.015% of the total cellular protein.

From the data regarding control of *glpK-lacZ* expression, it was concluded that GlpR functions as an effective repressor without the gene products of *glpE* and *glpG* under the conditions employed. However, the gene products of *glpE* and *glpG* had a slight influence on the regulation of *glpK-lacZ* expression. When both were present in the cell, they increased the GlpR-mediated repression of *glpK-lacZ* by about 2-fold in both high- and low-copy-number plasmids. GlpE/G do not seem to have any influence on the *in vivo* level of GlpR encoded by plasmid.

Effects of GlpE/GlpG on the GlpR-mediated repression of *glpD-lacZ*: The *glpK* gene of strain NZ45 is disrupted due to fusion with the *lacZ* gene. In order to

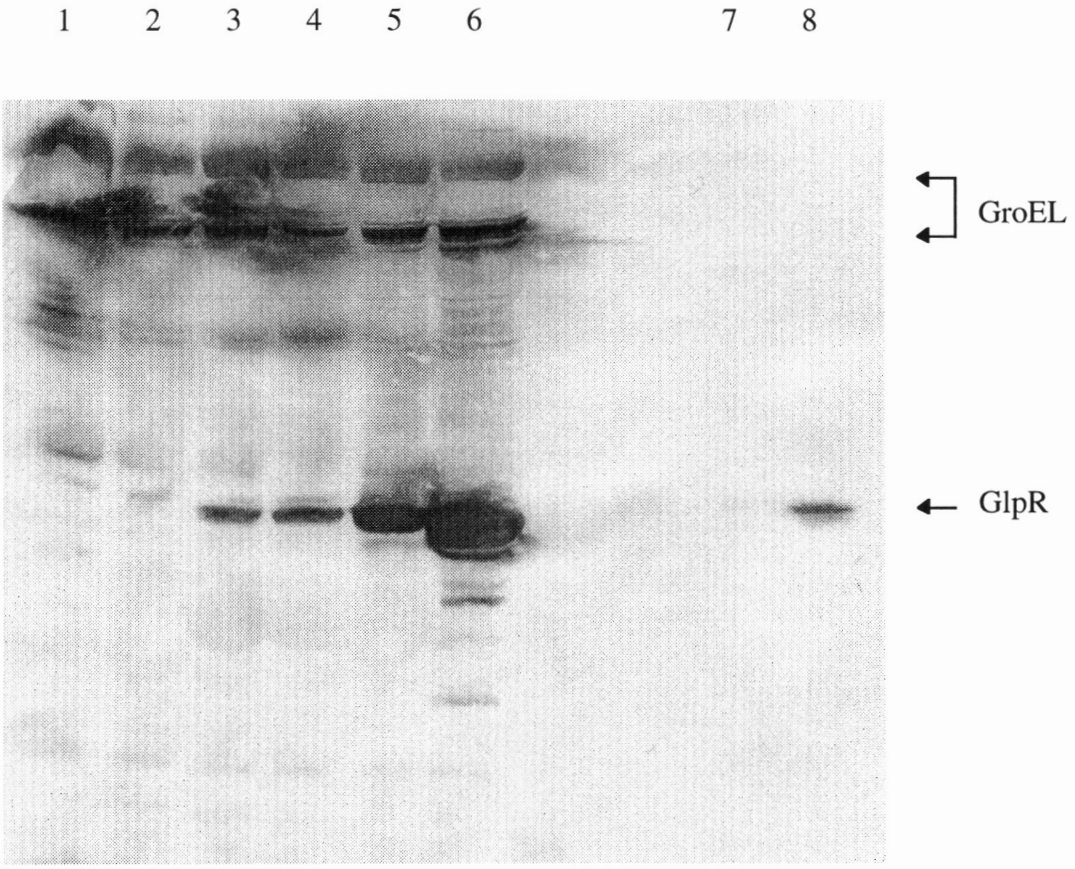


Figure 16. Estimation of the level of GlpR *in vivo* by Western analysis. About 300 μg of the cytoplasmic fraction from strain NZ45, GD31, NZ45(pGZ163), NZ45(pGZ163)(pGZ150) (induced by 2 μg/ml Tc), NZ45(pGZ163) (induced by 5 μM IPTG), and NZ45(pGZ158) were loaded in *lanes 1* to *6*, respectively. Cytoplasmic protein was estimated to be 2/3 of the total cellular protein, thus each lane contains the equivalent of about 450 μg of total cellular protein. *Lanes 7* and *8* contains 100 ng and 480 ng of pure GlpR as the standard. A mixture of polyclonal antibodies against GlpR and GroEL were used as primary antibodies. GroEL was used as the internal control for the quantity of protein present in each lane.

study the regulation of *glp* regulon in a wild-type strain for all *glp* genes except *glpEGR*, strain GZ578 [MC4100 Δ*glpEGR* λBY128 ϕ(*glpD-lacZ*) *recA1*] was constructed. In this strain, *glpD-lacZ* fusion was introduced at the λ attachment site by selecting a lysogen of λBY128, leaving the chromosomal *glpD* gene intact. To reveal whether GlpE/GlpG contribute to differential repression of the members of the *glp* regulon, the effect of GlpE/GlpG on repression of *glpD-lacZ* was determined and compared to that of *glpK-lacZ*.

Strain GZ578 harboring the same sets of plasmids (low copy-number) as used for the study of repression of *glpK-lacZ* were grown in AB minimal medium supplemented with 0.2% casamino acids. Expression of the *glpD-lacZ* fusion was determined by measuring β-galactosidase activity in log-phase cultures. Data are shown in TABLE 11. The fully constitutive level of β-galactosidase expression from this *glpD-lacZ* fusion was about 5500 Miller units. GlpR encoded by pGZ163 alone conferred a 23-fold repression upon *glpD-lacZ*. Repression of *glpD-lacZ* was decreased to 9-12 fold by the copresence of the *glpE/glpG* gene products. As to the expression of *glpK-lacZ* , the constitutive level was about 3200 Miller units (TABLE 9), and GlpR alone conferred an 84-fold repression; in contrast, the copresence of GlpE/GlpG increased repression to almost 170 fold (TABLE 10). Therefore, *glpD-lacZ* was controlled less tightly than *glpK-lacZ*. This agrees with the results obtained by using a plasmid-encoded *glp-lacZ* fusions as the reporter genes (7). Interestingly, the effect of GlpE/GlpG on repression of *glpD* was opposite to that found for repression of *glpK*. The presence of GlpE/GlpG increased repression of *glpK-lacZ* (TABLE 9 and 10) and decreased the repression of *glpD-lacZ* (TABLE 11).

TABLE 11. Effect of GlpE/G on GlpR-mediated repression of *glpD-lacZ* in strain GZ578

Plasmids[1] [proteins encoded by plasmid]	β-galactosidase activity[2]	
	no Tc	2 μg/ml Tc
pUHE1 pUHS4 [vector]	5490±200	7030±290
pGZ163 pUHS4 [R]	240±28	240±25
pGZ163 pGZ154 [R/E]	490±70	710±23
pGZ163 pGZ148 [R/G]	590±80	610±60
pGZ163 pGZ150 [R/EG]	440±90	690±100

[1] GlpR is encoded by low-copy-number plasmid (pSC101 origin).

[2] The activity was measured after growth without IPTG; data shown are from two or more independent determinations. Tetracycline (Tc) was for the induction of GlpE, GlpG, and/or GlpEG.

Effects of GlpE/GlpG on regulation of *glpT*-encoded transport activity: To investigate the regulation of another member of the *glp* regulon, *glpT*-encoded glycerol-P transport activity was measured in the strains described in TABLE 11. The GlpT activity was only subject to 9- to 12-fold repression by pGZ163-encoded GlpR. The presence of the *glpE/glpG* gene products had little or no effect on expression of the *glpT* gene (Fig. 17).

Thus, the degree of repression conferred by GlpR on *glpK*, *glpD*, and *glpT* is different. Each gene also shows a different response to the *glpE* and *glpG* gene products.

Assessment of the repression of *glpD-lacZ* conferred by GlpE/GlpG: To investigate whether GlpE and/or GlpG can directly influence expression of *glp* genes, plasmids encoding GlpE/G were introduced into strain GZ578 and expression of the *glpD-lacZ* gene was measured. Under the conditions employed (in the presence or absence of glycerol), *glpD-lacZ* expression had little or no change with respect to the gene products of *glpE/glpG* (TABLE 12). This result excludes the possibility that GlpE/GlpG are acting independently of GlpR as regulatory proteins for *glp* gene expression.

GlpE/GlpG differentially influence induction of *glpK-lacZ* and *glpD-lacZ*: Strains NZ45 and GZ578 were used to examine the induction ratios of *glp-lacZ* fusions in the presence or absence of GlpE/GlpG. The induction ratio was defined as the activity of *glpD-lacZ* (or in the case of NZ45, *glpK-lacZ*) grown in the presence of an inducer divided by the activity obtained in the absence of an inducer. As shown in TABLE 13, there is a significant difference in induction ratios for *glpK-lacZ* versus *glpD-lacZ*. The more tightly controlled *glpK-lacZ* fusion demonstrated a higher degree of induction than the

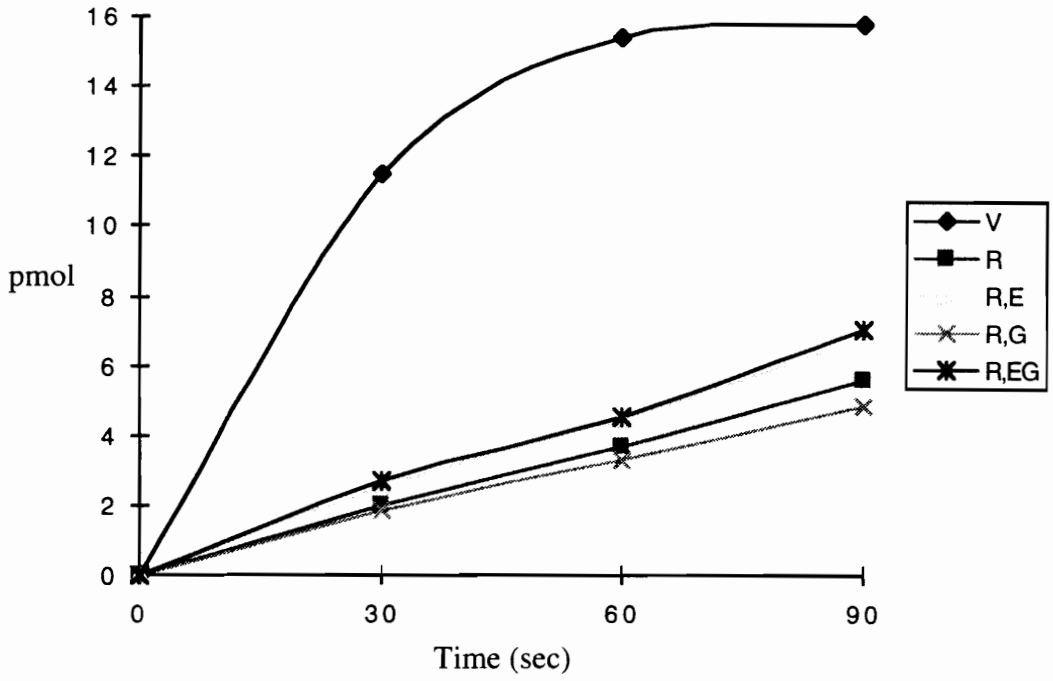


Figure 17. Transport of [^{14}C]glycerol-P into strain GZ578 carrying various plasmids. Growth conditions (in the presence of 2 μg/ml of tetracycline) and strains used in this experiment were the same as those described in TABLE 11. Glycerol-P uptake reached saturation within 60 seconds for strain GZ578 with vector. The activity was estimated according to the initial rate of transport (6.5 pmol in 10 s by 0.1 ml cells at OD_{578} = 0.5). Therefore, the calculated GlpT activity at constitutive level is about 780 pmol/min/ml unit OD_{578} cells. Data are the means of duplicate determinations.

TABLE 12. Effect of GlpE and/or GlpG on *glpD-lacZ* expression in strain GZ578

Plasmids present in GZ578 [proteins encoded by plasmid]	β-galactosidase activity[1]	
	no glycerol	+ glycerol
pUHA4 [vector]	7030	3550
pGZ154 [E]	8970	5090
pGZ148 [G]	8600	4770
pGZ150 [EG]	3830	4850

[1] Where indicated, glycerol (0.2%) was added to AB minimal medium supplemented with 0.2% casamino acids. Tc (2 μg/ml) was included in the medium.

TABLE 13. Effect of GlpE/G on the induction of *glpK* and *glpD*

Plasmids [protein encoded by plasmid]	β-galactosidase activity					
	NZ45(*glpK-lacZ*)			GZ578(*glpD-lacZ*)		
	uninduced	induced	ratio	uninduced	induced	ratio
pUHE1 pUHA4 [vectors]	3200	850	0.3	7030	3550	0.5
pGZ163 pUHA4 [R]	35	1040	30	240	3540	15
pGZ163 pGZ154 [R/E]	31	680	22	650	3880	6.0
pGZ163 pGZ148 [R/G]	51	1230	24	610	5670	9.3
pGZ163 pGZ150 [R/EG]	19	440	23	690	3550	5.1

Definition of induction ratio is described in the text. Cells were grown in the presence of 2 μg/ml tetracycline to induce production of the GlpE and/or GlpG proteins. The inducers employed were 0.4% D,L-glycerol-P for NZ45 derivatives and 0.2% glycerol for GZ578 derivatives. In both cases, GlpR was encoded by a low-copy-number plasmid. Data are the means of duplicate determinations.

less tightly controlled *glpD-lacZ* fusion. One explanation is that GlpK is the pace-maker of the entire *glp* regulon (8), and so a more flexible control with wider range of expression is necessary. This would allow a wider range of control of glycerol-P influx and the subsequent events. In the case of *glpK-lacZ*, neither GlpE nor GlpG had a significant influence on the induction ratio under the conditions employed, whereas the presence of GlpE/GlpG caused 2- to 3-fold decrease in the induction ratio found for the *glpD-lacZ* fusion. In both cases, the induction is mediated by GlpR, as the absence of *glp* repressor abolished any effect of inducers. The observation that the inducer glycerol-P and glycerol decreased the expression of *glpK-lacZ* and *glpD-lacZ* in the Δ*glpEGR* strain suggested the presence of another regulatory factor.

Growth on maltose reduces the induction ratio by glycerol-P: All the results described above were obtained from strains growing in minimal medium supplemented with 0.2% casamino acid as the sole source of energy. When cells harboring high-copy-number plasmids were grown in the presence of maltose, the expression of *glpK-lacZ* gene as well as the induction of *glpK-lacZ* by glycerol-P were greatly reduced (TABLE 14). Even though the mechanism of this effect is undetermined, it is comparable to the so-called catabolite repression and inducer exclusion caused by glucose, as indicated by the parallel data obtained from the same cells grown in the presence of 0.1% glucose (TABLE 14). For instance, the presence of maltose decreased the expression of *glpK-lacZ* in strain NZ45 by a factor of 6 while glucose decreased its expression by about 14-fold. In the presence of glycerol-P, the decrease in the expression of *glpK-lacZ* was 26-fold by maltose and 116-fold by glucose for NZ45 carrying pGZ158.

As stated previously, the presence of GlpE/GlpG increases the GlpR-mediated repression of *glpK-lacZ*. The same effect was also found when maltose or glucose served

TABLE 14. Effect of maltose on the expression of *glpK-lacZ* in strain NZ45

Plasmids [protein encoded by plasmid]	glycerol-P	β-galactosidase activity		
		caa[1]	caa+maltose	caa+glucose
pUHE1 [vector]	-	3200	550	230
pGZ158 [R]	-	13	2.8	3.1
pGZ158 [R]	+	1000	38	8.6
pGZ158 pGZ150 [R EG]	-	7.3	1.2	0.4
pGZ158 pGZ150 [R EG]	+	510	7.2	0.7

[1] Growth media were AB minimal, 2 μg/ml thiamine supplemented with the indicated carbon sources. Concentrations of carbon sources used: casamino acid (caa), 0.2%; D,L-glycerol-P, 0.2%; glucose, 0.1%; and maltose, 0.1%. The presence or absence of glycerol-P is indicated as + and -, respectively. Data shown are the means of duplicate determinations with less than 10% deviation. Cells with pGZ150 were grown in the presence of 2 μg/ml Tc.

as primary carbon source. Interestingly, the effect was more pronounced (8-fold versus 2-fold) when cells were grown in medium with glucose.

GlpE/GlpG do not affect the isoelectric point of GlpR: GlpE/GlpG may achieve its impact on differential regulation of the *glp* regulon by directly interfering with GlpR. To assess the question whether a covalent modification of GlpR is conferred due to the presence of GlpE/GlpG, the net charge of GlpR from a *ΔglpEG* and a *glpEG*$^+$ strain was compared by using isoelectric focusing analysis. Strain GZ578pGZ163 was employed as the source of GlpR in the absence of GlpE/GlpG; and strain GZ578pGZ163 pGZ150 provided GlpR produced in the presence of GlpE/GlpG.

Although it appears on SDS polyacrylamide gel as a single band, GlpR exists as multiple-charged forms as detected by isoelectric focusing (Fig. 18). The major band with a calculated pI of 5.99 is close to the predicted pI of 5.7, calculated from the deduced amino acid sequence of GlpR (47). Two minor bands were found with calculated pI's of 5.92 and 6.05, respectively. Multiple protein bands detected are not likely the results of nonspecific cross reaction to the GlpR antibody, as the pure GlpR which had experienced lengthy purification procedure (40) demonstrated the same pattern with the same relative amount of minor bands. Since GlpR from both Δ*glpEG* and wild-type strains showed the same pattern on IEF gels, no apparent correlation was found between the charges on GlpR and the presence of the gene products of *glpE/G*. The exact nature of this pI difference, and the physiological significance of the differently charged forms of repressor are questions that remain to be answered.

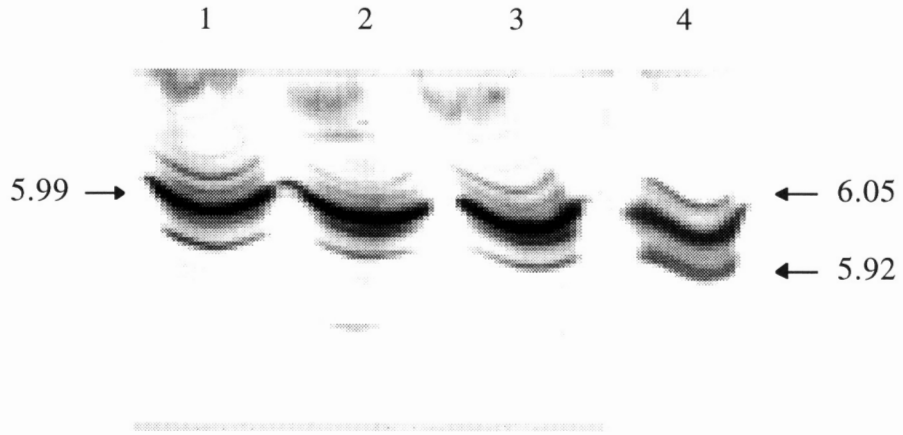


Figure 18. Western analysis of an isoelectric focusing gel of the *glp* repressor. A polyclonal antibody against GlpR was used for detection. *Lane 1* contains crude extract from strain GZ578(pGZ163) (GlpR was induced using 5 μM IPTG). *Lane 2* contains crude extract from strain GZ578(pGZ163)(pGZ150) (GlpR was induced using 5 μM IPTG and GlpEG was induced using 2 μg/ml tetracycline). *Lane 3* contains a mixture of the extracts used in *lane 1* and *2*. *Lane 4* contains a homogeneous preparation of GlpR. pH values at the indicated positions (inset arrows) were also shown.

DISCUSSION

Detailed information regarding the polypeptides encoded by the *glpEGR* operon has been obtained. The presence of two ORF's together with the *glpR*-encoded repressor in a single operon exhibits a unique organization among *E. coli* genes encoding regulatory proteins. Characterization of the *glp* repressor, the central player of the *glp* regulon repression, has revealed distinct features about the regulation of the *glp* regulon. *In vivo* characterization of the effects of GlpE/GlpG on GlpR-mediated repression of *glp-lacZ* has suggested a role for the *glpE* and *glpG* gene products in *glp* gene regulation. The biochemical functions of these two proteins, however, remain unknown.

Characteristic properties of the *glp* repressor and repression of the *glp* operons: The affinity of operator binding by *E. coli glp* repressor was studied by gel mobility shift assays. The apparent dissociation constant for the interaction with the single *glp* consensus operator was about 50 nM. The estimated K_d for operator binding is significantly lower than the affinities of other *E. coli* repressors for their operator DNA's. Examples of K_d's include 0.5 nM for *trp* repressor, which regulates transcription initiation at three operons involved in tryptophan biosynthesis (96); 10 pM or 0.57 nM for *lac* repressor as separately reported by Whitson *et al.* (97) and Fried *et al.* (98), respectively; 5 nM for MarR, which controls two *mar* (multiple antibiotic resistance) operons (99); and 0.2 nM for FadR, which regulates the transcription of many genes involved in fatty acid synthesis and degradation (100). Such a remarkable difference in operator DNA binding affinity for GlpR, from 10-fold compared with MarR to 5,000-fold compared with *lac* repressor seems to indicate that GlpR may require a modulator or a ligand to achieve tighter

binding. The pattern of DNA binding *in vivo* by GlpR may resemble that of PurR (101) or *c1* repressor of P1 bacteriophage (102). PurR represses the genes involved in *de novo* purine nucleotide biosynthesis. Upon binding with corepressor hypoxanthine or guanine, PurR increases operator binding affinity by more than 15-fold (103). In the case of *c1* repressor, *bof* protein enhances the binding to operator Op99a of the *c1* gene, but not to the operator Op72 of the *ban* operon (102). Likely, GlpR may bind a similar corepressor or ligand *in vivo*, even though this type of regulation has not yet been found in any of the repressors of the DeoR family, of which GlpR is a member. Another hypothesis is that the low operator binding affinity may reflect the lack of autoregulation of GlpR synthesis, which would allow a higher level of expression of GlpR, resulting in efficient repression. Furthermore, tandemly repeated *glp* operators occur in the native *glpD*, *glpACB*, *glpFKX* (3, 7, 61), and *glpTQ* operons (Bing Yang, unpublished results). Cooperative binding of the *glp* repressor to these operators results in a much higher degree of repression (58), even though GlpR apparently binds to a single operator with a relatively low affinity.

Another unique feature about GlpR is that although it appears on SDS polyacrylamide gel as a single band, it exists as multiple-charged forms as detected by isoelectric focusing. The major band with a calculated pI of 5.99 is close to the predicted pI of 5.7, calculated from the deduced amino acid sequence of GlpR (47). Two minor bands were found with calculated pI's of 5.92 and 6.05, respectively. Since repressor preparations from freshly made crude extracts and from nearly homogeneous preparations (40) showed the same pattern on IEF gels, the two minor bands could not be caused by the binding of ligand. Covalent modification by an unknown charged group, such as a phosphoryl group, is suggested. However, the pI of GlpR was not changed upon treatment with alkaline phosphatase. Whether the differently charged forms of GlpR are dependent on GlpE/GlpG was also investigated, but no apparent correlation was found.

The exact nature of this pI difference, and the physiological significance of the differently charged forms of repressor are questions that remain to be answered.

The patterns of differential repression of *glpK*, *glpD* and *glpT* by the *glp* repressor agreed well with the results described previously (7). Under the conditions employed, the *glp* repressor conferred more than 90-fold repression on *glpK*. Interestingly, in the presence of GlpE/GlpG, the repression ratio for *glpK* increased to near 170. In contrast, *glpD* was subject to 23-fold repression by GlpR. The presence of GlpE/GlpG, however, decreased this ratio to 120-fold. The *glpTQ* operon was subject to a relatively weak repression (about 9- to 12-fold), which was not affected by the *glpE/glpG* gene products. This is the first time that regulation of individual *glp* operons has been examined with respect to the GlpE/GlpG proteins. The results suggest that GlpE/GlpG exert effects bringing about differential regulation of the *glpK*, *glpD*, and *glpT* genes.

The DNA binding activities of GlpE and/or GlpG to the *glp* operator were examined by measuring the expression of *glpD-lacZ* in the presence of GlpE and/or GlpG encoded by plasmids (TABLE 12). In the presence of GlpE and/or GlpG, the level of expression of *glpD-lacZ* had little or no change. The results excluded the possibility that GlpE/G bind to *glpD* operator. Thus, the effect of GlpE/GlpG on differential regulation was not due to direct interactions between the two proteins and a *cis*-acting target on the operator DNA. Neither do GlpE/GlpG directly influence the level of GlpR *in vivo*, as revealed by Western analysis (Fig. 16). Another possible mechanism whereby GlpE/G confer differential regulation would be through direct interaction with GlpR. Since GlpG is associated with the cytoplasmic membrane, such an interaction would most likely occur between GlpE and GlpR, which are both present in the cytoplasm. A biochemical approach was employed to detect any physical interaction between the two proteins. The polyhistidine-tagged GlpE encoded by pGZ107 was first immobilized on a Ni^{2+} charged affinity column, and then a

crude extract containing overproduced GlpR was loaded onto the column. As a negative control, the extract containing GlpR was also loaded onto the same column but without the preoccupation by GlpE$(His)_6$. GlpE was released from the resin by washing with buffer containing 250 mM imidazole. Analysis of the eluted fractions using polyacrylamide gel electrophoresis revealed no proteins specifically associated with GlpE (data not shown). The conditions used in the experiment (pH 7.0, 150 mM NaCl) are similar to the physiological pH and salt concentration.

A separate approach was used to detect whether GlpE could induce a conformational change of GlpR. Increasing amounts of pure GlpE$(His)_6$ protein were incubated with GlpR and then the mixtures were treated with a low concentration of trypsin. The trypsin cleavage pattern of GlpR remained the same even when GlpE was present at an 8:1 molar ratio over GlpR. In the case where glycerol-P was present, the change in preference for cleavage by trypsin was also found to be independent of GlpE.

We therefore speculate that the effect of GlpE/GlpG on differential regulation is probably indirect, i.e., through influencing the level of another compound relevant to the DNA-binding activity of GlpR. This compound could be a ligand for GlpR. Upon binding this ligand, GlpR must bind more tightly to the *glpK* operators and more weakly to the *glpD* operators; whereas the affinity for the *glpT* operator remains the same. The fact that GlpE/GlpG did not have a remarkable effect on the induction of *glp* repressor by glycerol-P indicates that the binding site of the possible ligand may not be the same as that of the inducer. But the presence of *glpE/glpG* gene products caused a slight decrease in the induction ratio of *glpD-lacZ* and *glpK-lacZ* (TABLE 11), which indicates that binding of the compound may exert a minor effect on glycerol-P binding. The genetic organization of the *glpE* gene in *H. influenzae* shows that *glpE* is adjacent to *tpiA* (93), whose gene product, triose isomerase, is responsible for the isomerization of DHAP and GAP. This

may provide a clue about the function of GlpE. If GlpE interacts with triose phosphate isomerase, it may influence the levels or the ratio of DHAP to GAP *in vivo*. One of them might be the compound that binds to GlpR and is responsible for the differential binding among *glp* operators. It is known, however, that DHAP does not interfere with binding of inducer, glycerol-P, by GlpR (40).

Plasmids of different copy number with *glpR* confer different level of repression of the *glp* operon repressions. For instance, GlpR encoded by the plasmid with a ColE1 origin (pGZ158) yielded a repression ratio of about 250 for *glpK-lacZ* expression. The repression ratio decreased to 100 when the origin was changed to pSC101. The copy numbers of these plasmids are usually 50-100 and 5-10 per cell, respectively. Thus the 2.5-fold decrease in repression may reflect a decrease in the quantity of repressor present. Interestingly, in the presence of an inducer, strains harboring these plasmids achieve the same fold of induction, even though the level of expression is different. Comparison of the induction ratios obtained for *glpK-lacZ* and *glpD-lacZ* also indicated that the more tightly repressed *glpK* gene has a more pronounced responsiveness upon induction. It is believed to be more favorable for cells to have a wide range of control for expression of *glpK*, since glycerol kinase sets the pace for glycerol metabolism *in vivo* (9).

Hypotheses excluded for possible GlpE/GlpG functions: The hypothesis that GlpE/GlpG lead to a covalent modification of GlpR was tested by the IEF experiment described above. The results shown in Fig. 17, however, do not support this idea.

When *glpEG* was to be replaced by a Km^R cassette through homologous recombination in strain TL684 (Δ*glpBglpD*) via P1 transduction, it was observed that all of the transductants selected on LB Km or LB Tc had a glycerol-negative phenotype (T. D.

Clark, unpublished). A possible interpretation is that wild-type *glpD* cotransduced with the nearby Km^R cassette is not able to function when both *glpEG* and *glpB* are defective. This observation led to the hypothesis that GlpD possibly needs another membrane component to function as an effective dehydrogenase *in vivo*, with either GlpB or GlpG providing this role. Genetic experiments were therefore designed to replace *glpEG* or *glpEGR* of strain TL661 (*glpACB-lacZ*) with Km^R cassette by P1 transduction using strains DA142 (Δ*glpEG*) and NZ411 (Δ*glpEGR*) as donors, respectively. Similarly, TL658 (*glpT-lacZ*) was used as a recipient in the same procedure to serve as a control. Transductants were selected on LB Tc with Km, and subsequently tested for their glycerol utilization phenotype. All of the 60 individual TL658 transductants from DA142 donor were able to grow in minimal medium with glycerol as the sole carbon source, and 50 of the 60 TL661 transductants from DA142 donor grew well in the same medium. However, all 30 transductants of TL661 from the NZ411 donor grew well in minimal medium with glycerol. This result did not support the prediction that GlpB or GlpG is required for effective GlpD-encoded dehydrogenase activity.

I also tested the possibility that GlpG binds glycerol-P. Preferential binding of [^{3}H]glycerol-P was tested using protein fractions from strains BL21(DE3)(pGEM3Z) (vector control) and BL21(DE3)(pGZ125), which overproduces GlpG. The experiment was performed according to the inducer binding assay for GlpR. No apparent difference was found using the cytoplasmic membrane preparations from the above strains.

Domain homology search with all known proteins by Eugene Koonin (NIH) has revealed a motif within GlpE that is conserved in a number of enzymes including sulfatase, tyrosine phosphatase, and rhodanese (unpublished). There is also a group of bacterial proteins of unknown functions whose size is similar to GlpE that have the same motif. Based on this information, a phosphatase activity was proposed for GlpE and tested by

using PNPP as the putative substrate. Pure His-tagged GlpE (1 μg) was incubated with the substrate (1 mM) in the presence of 1 mM $MgSO_4$ under various pH conditions (from pH 5.0 to pH 9.0). After incubation at 37°C for 3 hours, the formation of dephosphorylated products was monitored by measuring the absorbance at 420 nm. No obvious difference was found between the blank sample and the samples incubated with GlpE. Another substrate, D,L-glycerol-P (4 mM) was also tried in the presence of 1 μg of GlpE and 1 mM $MgSO_4$ at pH 7.0. The reaction was carried out at 37°C for 4 hours, and subsequent assays were performed to detect the release of inorganic phosphate by using the Malchite Green method (104). No phosphatase activity was found for GlpE.

Working model suggested by the current data: Because limited information was provided by sequence similarity searches, no positive biochemical functions have been suggested for GlpE and GlpG. However, some useful thoughts may shed light on roles for GlpE and GlpG *in vivo* and suggest directions for future work.

The conserved motif of GlpE discovered by Eugene Koonin consists of 3 conserved amino acid residues, Asp, His, and Cys. Interestingly, this is exactly the catalytic triad found extensively among various cysteine proteinases (105). Likely, Asp, His, and Cys of GlpE may form a charge relay system and convert a weakly nucleophilic CH_2SH group to a highly nucleophilic CH_2S^-. Cysteine proteinases are mainly found in the eukaryotic cytosol (106), including vertebrate lysosomal cathepsins B, and S (106), and mammalian cathepsin K(107) and so on. There are also two examples from bacteria falling into this family, the aminopeptidase C from *Lactococcus lactis* (108) and thiol protease Tpr from *Porphyromonas gingivalis*. It should be pointed out that GlpE and the

group of proteins suggested by Eugene Koonin do not match the signature pattern as possessed by all cysteine proteinases. Thus, GlpE may not be a proteinase, but instead may have an activity requiring a nucleophilic attack on compounds with an electron-deficient atom, such as a carbonyl or phosphoryl group. Therefore a kinase/phosphatase or a hydrolase activity may be suggested.

On the other hand, GlpG could be an electron transfer protein as suggested by its sequence similarity to dehydrogenases and cytochromes of many organisms.

A working model has been proposed to explain the data obtained in this research. As discussed above, the model predicts the presence of a compound, maybe DHAP or GAP, that is responsive to *glpE/glpG* gene products. The binding to this compound results in the *glp* repressor with differential affinities to *glp* operons.

Verification of this hypothesis needs further experimental proof. Nevertheless, the effects of GlpE/GlpG on *glpK* and *glpD* expression seem to shed light on the question about how cells respond to inducing levels of glycerol-P. According to previous knowledge, the release of *glp* repressor from binding its operators yields elevated GlpK activity, which results in more production of glycerol-P when glycerol is abundant. It is energetically unfavorable for living cells. However, according to the fact that the presence of *glpE* and *glpG* gene products confer a tighter control by GlpR on *glpK* and a weaker control on *glpD*, the more elevated GlpD activity in the presence of inducing level of glycerol-P may help to quickly consume the inducer and keep the cell healthy. So, GlpE/GlpG may be required to ensure balanced production and utilization of the inducer. Cells may use this mechanism to respond to elevated levels of inducer in the environment.

LITERATURE CITED

1. Cronan, J. E., Jr., and Rock, C. A. (1987) Biosynthesis of membrane lipids. pp.474-479. *In* Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E. (eds) "*Escherichia coli* and *Samonella typhimurium.* Cellular and Molecular Biology". American Society for Microbiology, Washington, D. C.

2. Raetz, C. R. H. (1986) Molecular genetics of membrane phospholipid synthesis. *Annu. Rev. Genet.* **20,** 253-295

3. Larson, T. J., Cantwell, J. S., and van Loo-Bhattacharya, A. (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

4. Lin, E. C. C. (1987) Dissimilatory pathways for sugars, polyols, and carboxylates. pp. 244-84. *In* Neidhardt, F. C. *et al.* (eds.) "*Escherichia coli* and *Samonella typhimurium.* Cellular and Molecular Biology." American Society for Microbiology, Washington, D. C.

5. Heller, K. B., Lin, E. C. C., and Wilson, T. H. (1980) Substrate specificity and transport proteins of the glycerol facilitator of *Escherichia coli. J. Bacteriol.* **144,** 274-278

6. Sweet, G., Gandor, C., Voegele, R., Wittekindt, N., Beuerle, J., Truniger, V., Lin, E. C. C., and Boos, W. (1990) Glycerol facilitator of *Escherichia coli*: cloning of *glpF* and identification of the *glpF* product. *J. Bacteriol.* **172,** 424-430

7. Weissenborn, D., Wittekindt, N., and Larson, T. J. (1992) Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli. J. Biol. Chem.* **267,** 6122-6131

8. Pettigrew, D. W., Ma, D.-P., Conard, C. A., and Johnson, J. R. (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

9. Voegele, R., Sweet, G. D., and Boos, W. (1993) Glycerol kinase of *Escherichia coli* is activated by interaction with the glycerol facilitator. *J. Bacteriol.* **175,** 1087-1094

10. Argast, M., and Boos, M. (1980) Co-regulation in *Escherichia coli* of a novel transport system for *sn*-glycerol 3-phosphate and outer membrane protein Ic (e, E) with alkaline phosphatase and phosphate-binding protein. *J. Bacteriol.* **143,** 142-150

11 . Guth, A., Engel, R., and Troop, B. E. (1980) Uptake of glycerol 3-phosphate and some of its analogs by the hexose phosphate transport system of *Escherichia coli*. *J. Bacteriol.* **143,** 538-539

12 . Maloney, P. C., Ambudkar, S. V., Anantharam, V., Sonna, L. A., and Varadhachary, A. (1990) Anion-exchange mechanisms in bacteria. *Microbiol. Rev.* **54,** 1-17

13 . Larson, T. J., Schumacher, G., and Boos, W. (1982) Identification of the *glpT*-encoded *sn*-glycerol 3-phosphate permease of *Escherichia coli*, an oligomeric integral membrane protein. *J. Bacteriol.* **152,** 1008-1021

14 . Larson, T. J., and van Loo-Bhattacharya, A. T. (1988) Purification and characterization of *glpQ*-encoded glycerophosphodiester phosphodiesterase from *Escherichia coli* K-12. *Archives Biochem. Biophys.* **260,** 577-584

15 . Schryvers, A., Lohmeier, E., and Weiner, J. H. (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

16 . Baltsheffsky, H. and Baltsheffsky, M. (1974) Electron transport phosphorylation. (1974) *Annu. Rev. Biochem.* **43,** 871

17 . Fillingame, R. H. (1980) The proton-translocating pumps of oxidative phosphorylation. *Annu. Rev. Biochem.* **49,** 1079

18 . Ehrmann, M., Boos, W., Ormseth, E., Scheweizer, H., and Larson, T. J. (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

19 . Cole, S. T., Eiglmeier, K., Ahmed, S., Honore, N., Elmes, L., Anderson, W. F., and Weiner J. H. (1988) Nucleotide sequence and gene-polypeptide relationships of the *glpABC* operon encoding the anaerobic *sn*-glycerol 3-phosphate dehydrogenase of *Escherichia coli*. *J. Bateriol.* **170,** 2448-2456

20 . Iuchi, S., and Lin, E. C. C. (1993) Adaptation of *Escherichia coli* to redox environments by gene expression. *Mol. Microbiol.* **9,** 9-15

21 . Martin, E., Freedberg, W., and Lin, E. C. C. (1977) Kinase replacement by a dehydrogenase for *Escherichia coli* glycerol utilization. *J. Bacteriol.* **131,** 1026-1028

22 . Tang, J., Ruch, F., and Lin, E. C. C. (1979) Purification and properties of a nicotinamide adenine dinucleotide-linked dehydrogenase that serves an *Escherichia coli* mutant for glycerol catabolism. *J. Bacteriol.* **140,** 182-187

23 . Kelly, J., and Cekker, E. (1985) Identity of *Escherichia coli* D-1-amino-2-propanol: NAD^+ oxidoctase with *E. coli* glycerol dehydrogenase but not with *Neisseria gonorrhoeae* 1,2-propanediol:NAD^+ oxidoreductase. *J. Bacteriol.* **162,** 170-175

24 . Truniger, V., and Boos, W. (1994) Mapping and cloning of *gldA*, the structural gene of the *Escherichia coli* glycerol dehydrogenase. *J. Bacteriol.* **176,** 1796-1800

25 . Jin, R., and Lin, E. C. C. (1984) An inducible phophoenolpyruvate: dihydroxyacetone phophotransferase system in *Escherichia coli*. *J. Gen. Microbiol.* **130,** 83-88

26 . Cronan, J. E., Jr., and Bell, R. M. (1974) Mutants of *Escherichia coli* defective in membrane phospholipid sythesis: Mapping of the structural gene for L-glycerol 3-phosphate dehydrogenase. *J. Bacteriol.* **118,** 598-605

27 . Rock, C. O., and Cronan, J. E. Jr. (1982) Regulation of bacterial membrane lipid synthesis. *Curr. Top. Membr. Transp.* **17,** 207-233

28 . Coleman, J. (1992) Characterization of the *Escherichia coli* gene for 1-acyl-*sn*-glycerol-phosphate acyltransferase (*plsC*). *Mol. Gen. Genet.* **232,** 295-303

29 . Postma, P. W., Lengeler, J. W., and Jacobson, G. B. (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57,** 543-594

30 . Freedberg, W. B., and Lin, E. C. C. (1973) Three kinds of controls affecting the expression of the *glp* regulon in *Escherichia coli. J. Bacteriol.* **115,** 816-823

31. Nelson, S. O., Schuitema, A. R. J., Benne, R., Van der Ploeg, L. H. T., Plijter, J. S., Aan, F., and Postma, P. W. (1984) Molecular cloning, sequencing and expression of the *crr* gene: the structural gene for III^{Glc} of the bacterial PEP:glucose phosphotransferase system. *EMBO J.* **3,** 1587-1593

32 . De Boer, M., Broekhuizen, C. P., and Postma, P. W. (1986) Regulation of glycerol kinase by enzyme III^{Glc} of the phosphoenolpyruvate:carbohyderate phosphotransferase system. *J. Bacteriol.* **167,** 393-395

33 . Hurley, J. H., Worthylake, D., Faber, H. R., Meadow, N. D., Roseman, S., Pettigrew, D. W., and Remington, S. J. (1993) Structure of the regulatory complex of *Escherichia coli* III^{Glc} with glycerol kinase. *Science* **259,** 673-677

34 . Shaw, D. J., Rice, D. W., and Guest, J. R. (1983) Homology between CAP and FNR, a regulator of anaerobic repression in *Escherichia coli*. *J. Mol. Biol.* **166,** 241-247

35 . Khoroshilova, N., Beinert, H., and Kiley, P. J. (1995) Association of a polynuclear iron-sulfur center with a mutant FNR protein enhances DNA binding. *Proc. Natl. Acad. USA* **92,** 2499-2503

36 . Bates, D. M., Lazazzera, B. A., and Kiley, P. J. (1995) Characterization of FNR mutant proteins indicates two distinct mechanisms for altering oxygen regulation of the *Escherichia coli* transcription factor FNR. *J. Bacteriol.* **177,** 3972-3978

37 . Iuchi, S., and Lin, E. C. C. (1988) *arcA* (*dye*), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proc. Natl. Acad. Sci. USA* **85,** 1888-1892

38 . Iuchi, S., Cameron, D. C., and Lin, E. C. C. (1989) A second global regulator gene (*arcB*) mediating repression of enzymes in aerobic pathways of *Escherichia coli*. *J. Bacteriol.* **171,** 868-873

39 . Iuchi, S., Cole, S. T., and Lin, E. C. C. (1990) Multiple regulatory elements for the *glpA* operon encoding anaerobic glycerol 3-phosphate dehydrogenase and the *glpD* operon encoding aerabic glycerol 3-phosphate dehydrogenase in *Escherichia coli*: further characterization of respiratory control. *J. Bacteriol.* **172,** 179-184

40 . Larson, T. J., Ye, S., Weissenborn, D. L., Hoffmann, H. J. and Schweizer, H. (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

41 . Barkely, M. D., and Bourgeois, S. (1980) Repressor recognition of operator and effectors. *In* "The Operon" (Miller, J. H., and Reznikoff, W. S., eds) pp. 177-220, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

42 . Green, P. R., Merrill, A. H., Jr., and Bell, R. M. (1981) Membrane phospholipid synthesis in *Escherichia coli*:: Purification, reconstitution, and characterization of *sn*-glycerol 3-phosphate acyltransferase. *J. Biol. Chem.* **256,** 11151-11159

43 . Lin, E. C. C. (1976) Glycerol dissimilation and its regulation in bacteria. *Annu. Rev. Microbiol.* **30,** 535-578

44 . Zhao, N. (1993) M.S. thesis. Studies on the operator-repressor-effector interactions in the *glp* regulon of *Escherichia coli* K-12. Virginia Polytechnic Institute and State University

45 . Miller, J. H. (1980) The *lacI* gene: Its role in *lac* operon control and its use .as a genetic system. pp. 31-88. *In* "The Operon" (Miller, J. H., and Reznikoff, W. S., eds) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

47 . Zeng, G., Ye, S., and Larson, T. J. (1996) Repressor for the *sn*-glycerol 3-phosphate regulon of *Escherichia coli* K-12. Primary structure and identification of the DNA-binding domain. manuscript in preparation.

48 . Yamada, M, and Saier, M. H., Jr. (1988) Positive and negative regulators for glucitol (*gut*) operon expression in *Escherichia coli* . *J. Mol. Biol.* **203,** 569-583

49 . Lu, Z., and Lin, E. C. C. (1989) The nucleotide sequence of *Escherichia coli* genes for L-fucose dissimilation. *Nucleic Acids Res.* **17,** 4883-4884

50 . Valentin-Hansen, P., Hojrup, P., and Short, S. (1985) The primary structure of the DeoR repressor from *Escherichia coli* K-12. *Nucleic Acids Res.* **13,** 5927-5936

51 . Oskouian, B., and Stewart, G. C. (1990) Repression and catabolite repression of the lactose operon of *Staphylococcus aureus.* *J. Bacteriol.* **172,** 3804-3812

52 . van Rooijen, R. J., and de Vos, W. M. (1990) Molecular cloning, transcriptional analysis, and nucleotide sequence of *lacR*, a gene encoding the repressor of the lactose phosphotransferase system of *Lanctococcus lactis.* *J. Biol. Chem.* **265,** 18499-18503

53 . Von Bodman, S. B., Hayman, G. T., and Farrand, S. T. (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

54 . Puyet, A., Ibanez, A. M., and Espinosa, M. (1993) Characterization of the *Streptococcus pneumoniae* maltosaccharide regulator MalR, a member of the LacI-GalR family of repressors displaying distinctive genetic features. *J. Biol. Chem.* **266,** 25402-25408

55 . Russel, R. R. B., Aduse-Opoku, J., Sutcliffe, I. C., Tao, L., and Ferretti, J. J. (1992) A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. *J. Biol. Chem.* **267,** 4631-4637

56 . Quail, M. A. and Guest, J. R. (1995) Purification, characterization and mode of action of PdhR, the transcriptional repressor of the *pdhR-aceEF-lpd* operon of *Escherichia coli. Mol. Microbiol.* **15,** 519-529

57 . Austin, D. (1991) Ph. D. dissertation. Characterization of the *glpD* and *glpEGR* operons of *Escherichia coli* K-12. Virginia Polytechnic Institute and State University

58 . Zhao, N., Oh, W., Trybul, D., Thrasher, K. S., Kingsbury, T., and Larson, T. J. (1994) Characterization of the inteaction of the *glp* repressor of *Escherichia coli* with single and tandem *glp* operator variants. *J. Bacteriol.* **176,** 2393-2397

58 . Zhao, N., Oh, W., Trybul, D., Thrasher, K. S., Kingsbury, T., and Larson, T. J. (1994) Characterization of the inteaction of the *glp* repressor of *Escherichia coli* with single and tandem *glp* operator variants. *J. Bacteriol.* **176,** 2393-2397

59 . Yang, B. and Larson, T. J. (1996) Action at a distance for negative control of transcription of the *glpD* gene encoding *sn*-glycerol 3-phosphate dehydrogenase of *Escherichia coli* K-12. In preparation for *J. Bacteriol.*

61 . Ye, S. and Larson, T. J. (1988) Structures of the promoter and operator of the *glpD* gene encoding aerobic *sn*-glycerol-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* **170,** 4209-4215

62 . Friedman, D. I. (1988) Integration host factor, A protein for all reasons. *Cell* **55,** 545-554

63 . Yang, C.-C., and Nash, H. A. (1989) The interaction of *E. coli* IHF protein with its specific binding sites. *Cell* **57,** 869-880

64 . Yang, B., Gerhardt, S. G., and Larson, T. J. (1995) The effect of integration host factor on *glpTQ* transcription in *Escherichia coli* K-12. manuscript submitted.

65 . Truniger, V., Boos, W., and Sweet, G. (1992) Molecular analysis of the *glpFKX* regions of *Escherichia coli* and *Shigella flexneri. J. Bacteriol.* **174,** 6981-6991

66 . Caruthers, M. H. (1985) Gene synthesis machines: DNA chemistry and its uses. *Science* **230,** 281-285

67 . Tabor, S., and Richardson, C. C. (1982) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* .**82,** 1074-1078

68 . Silhavy, T. J., Berman, M. L., and Enquist, L. W. (1984) Media and standard solutions. pp. 217 *In* "Experiments with Gene Fusion". Cold Spring Harbor Laboratory. Cold Spring Harbor, NY

69 . Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Appendix. pp. A.1 *In* Molecular cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

70 . Clark, D. J., and Maaløe, O. (1976) DNA replication and the division cycle in *Escherichia coli. J. Mol. Biol.* **23,** 99-112

71 . Miller, J. H. (1972) *In* "Experiments of Molecular Genetics". Cold Spring Harbor Laboratory. Cold Spring Harbor, NY

72 . Silhavy, T. J., Berman, M. L., and Enquist, L. W. (1984) Preparation of a P1Tn9*clr*100 lysate. *In* "Experiments with Gene Fusion". Cold Spring Harbor Laboratory. Cold Spring Harbor, NY

73 . Liss, L. R. (1987) New M13 host, DH5αF' competent cells. *Focus* **3,** 13

74 . 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.

75 . Schweizer, H., Boos, W., and Larson, T. J. (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

76 . Silhavy, T. J., Hartig-Beecken, I., and Boos, W. (1976) Periplasmic protein related to the *sn*-glycerol 3-phosphate transport system of *Escherichia coli. J. Bacteriol.* **147,** 255-258

77 . Ludtke, D., Larson, T. J., Beck, C., and Boos, W. (1982) Only one gene is required for the *glpT*-dependent transport of *sn*-glycerol 3-phosphate in *Escherichia coli. Mol. Gen. Genet.* **186,** 540-547

78 . Yu, F., Jen, Y., Takeuchi, E., Inouye, M., Nakayama, H., Tagaya, M., and Fukui, T. (1988) α-Glucan phosphorylase from *Escherichia coli. J. Biol. Chem.* **263,** 13706-13711

79 . Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113

80 . Ye, S. (1990) Ph. D. dissertation. Structure and function of the repressor and operators of the *sn*-glycerol 3-phosphate regulon of *Escherichia coli* K-12. Virginia Polytechnic Institute and State University

81 . Messing, J. (1983) New M13 vectors for cloning. *Met. Enzymol.* **101,** 20-78

82 . 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

83 . Adams, L. D. (1987) Two-dimensional gel electrophoresis using the O'Farrell System. pp. 10.4.1. *In* Current Protocols in Molecular Biology (edited by Ausubel, F. M. *et al.*). John Wiley & Sons, New York, NY

84 . Garfin, D. E. (1990) Isoelectric focusing. pp. 459-477. *In* "Guide to Protein Purification" (edited by Deutscher, M. P.). Academic Press, San Diego, CA

85 . Harlow, E., and Lane, D. (1988) Immunoblotting. pp. 471-510. *In* "Antibodies, A Laboratory Manual". Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

86 . Porath, J., and Olin, B. (1983) Immobilized metal ion affinity and immobolized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel-immobolized iron and nickel ions. *Biochemistry* **22,** 1621-1630

87 . Schweizer, H., Sweet, G., and Larson, T. J. (1986) Physical and genetic structure of the *glpD-malT* interval of the *Escherichia coli* K-12 chromosome. *Mol. Gen. Genet.* **202,** 488-492

88 . Choi, Y-L., Kawase, S., Kawamukai, M., Sakai, H., and Komano, T. (1991) Regulation of the *glpD* and *glpE* gene expression by a cyclic AMP-cAMP receptor protein (cAMP-CRP) complex in *Escherichia coli*. *Biochim. Biophys. Acta.* **1008,** 31-35

89 . Weissenborn, D. L. (1990) Ph. D. dissertation. Regulation of the *glpFK* operon of *Escherichia coli* K-12 and characterization of its gene products. Virginia Polytechnic Institute and State University

90 . Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105-132

91 . Nobelmann, B., and Lengeler, J. W. (1995) Sequence of the *gat* operon for galactitol utilization from a wild-type strain EC3132 of *Escherichia coli. Biochim. Biophys. Acta* **1262,** 69-72

92 . Fleischmann, R. D., Adams, M. D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science.* **269,** 496-512

93 . Bergold, P. J., Blumenthal, J. A., D'Andrea, E., Snyder, H. W., Lederman, L., Silverstone, A., Nguyen, H., Besmer, P. (1987) Nucleic acid sequence and oncogenic properties of the HZ2 feline sarcoma virus v-abl insert. *J. Virol.* **61,** 1193-1202

94 . Overduin, M., Mayer B. J., Rios, C. B., Baltimore, D., and Cowburn, D. (1992) *Proc. Natl. Acad. Sci. USA.* **89,** 11673-11677

95 . Cohen, S. N., Chang, A. C. Y., and Hsu, L. (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R factor DNA. *Proc. Natl. Acad. Sci. USA* . **69,** 2110

96 . Carey, J. (1988) Gel retardation at low pH resolves *trp* repressor DNA complexes for quantitative study. *Proc. Natl. Acad. Sci. USA* . **89,**643-647

97 . Whitson, P. A., Olson, J. S., and Matthew, K. S. (1986) Thermodynamic analysis of the lactose repressor-operator DNA interaction. *Biochemistry* **25,** 3852-3858

98 . Fried, M. G., and Crothers, D. M. (1981) Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9,** 6505-6524

99 . Seoane, A. S., and Levy, S. B. (1995) Characterization of MarR, the repressor of the multiple antibiotic resistance (*mar*) operon in *Escherichia coli*. *J. Bacteriol.* **177,** 3414-3419

100 . Raman, N., and DiRusso, C. C. (1995) Analysis of acyl coenzyme A binding to the transcription factor FadR and identification of amino acid residues in the carboxyl terminus required for ligand binding. *J. Biol. Chem.* **270,** 1092-1097

101 . Choi, K. Y., and Zalkin, H. (1992) Structural characterization and corepressor binding of the *Escherichia coli* purine repressor. *J. Bacteriol.* **174,** 6207-6214

102 . Velleman, M., Heirich, M., Gunther, A., and Schuster, H. (1990) A bacteriophage P1-encoded modulator protein affects the P1 c1 repression system. *J. Biol. Chem.* **265,** 18511-18517

103 . Choi, K. Y., Lu, F, and Zalkin, H. (1994) Mutagenesis of amino acid residues required for binding of corepressors to the purine repressor. *J. Biol. Chem.* **269,** 24066-24072

104 . Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., and Candia, O. A. (1979) An improved assay for nanomole amount of inorganic phosphate. *Anal. Biochem.* **100,** 95-97

105 . Bond, J. S., and Butler, P. E. (1987) Intracellular proteases. *Annu. Rev. Biochem.* **56,** 333-364

106 . Ritonja A., Colic, A, Dolenc, I., Ogrinc, T., Podobink, M., and Turk, V. (1991) The complete amino acid sequence of bovine cathepsin S and a partial sequence of bovine cathepsin L. *FEBS Lett.* **283,** 329-331

107 . Shi, G.-P., Chapman, H. A., Bhairi, S. M., Deleeuw, C., Reddy, V. Y., and Weiss, S. J. (1995) Molecular cloning of human cathepsin O, a novel endoproteinase and homologue of rabbit OC2. *FEBS Lett.* **357,** 129-134

108 . Chapot-Chartier, M. P., Nardi, M., Chopin, M. C., Chopin, A., and Gripon, J. C. (1993) Cloning and sequencing of *pepC*, a cysteine aminopeptidase gene from *Lactococcus lactis* subsp. cremoris AM2. *Appl. Environ. Microbiol.* **59,** 330-333

CURRICULUM VITAE

GANG ZENG

PERSONAL:

Date of Birth:	April 28, 1969
Place of Birth:	Shandong Province, China
Business address:	Research Institute for Genetic and Human Therapy Gaithersburg, MD

EDUCATION:

1991-1996 **Ph.D., Biochemistry.** February, 1996
Virginia Polytechnic Institute and State University
Title of Dissertation: "Characterization of the *glpEGR* operon of *E. coli.*" Chairman: Timothy J. Larson, Ph.D.

1987-1991 **B.S., Cell Biology and Genetics.** Beijing University, China

HONORS:

- Eheart/Wu Scholarship, Department of Biochemistry, Virginia Polytechnic Institute and State University, 1995
- Honored undergraduate thesis, Beijing University, China. 1991

RESEARCH INTERESTS:

Mechanisms of gene regulation, especially post-translational modification of DNA-binding proteins and transcription factors; DNA-protein and protein-protein interactions; metabolic signals in regulating gene expression

Gene regulation relating to human diseases; viral vector mediated gene transfer, chimerical protein engineering in gene therapy

EXPERIENCE:

1991-1996 **Graduate Research Assistant**, Department of Biochemistry
Virginia Polytechnic Institute & State University

Aug.- Dec. 1992 **Graduate Teaching Assistant**, Department of Biochemistry
Virginia Polytechnic Institute & State University

Jan. - July 1991 **Undergraduate Researcher**, Department of Human Genetics
Forensic Medicine Institute of the Public Security Ministry, P.R.China

PUBLICATIONS:

Zeng, G., and Larson, T. J. (1995) Identification of two forms of *glp* repressor GlpR in *E. coli.* (Abstract). Protein Engineering. vol.8/suppl., pp.79

Larson, T. J., Clark T. D., Zhao, N., and Zeng, G. (1995) Genetic control of the glycerol 3-phosphate *(glp)* system of *E.coli.* (Abstract). Fall 1995 Cold Spring Harbor Laboratory Meeting: Molecular Genetics of Bacteria and Phages

Zeng, G., Ye, S. and Larson, T. J. (1996) Repressor for the *sn*-glycerol 3-phosphate Regulon of *E. coli.* K-12, Primary Structure and Identification of the DNA-binding Domain. Submitted to *J. Biol. Chem.*

Zeng, G., and Larson, T. J. (1996) Use of an engineered thrombin cleavage site for the determination of translational reading frames. Submitted to *Bio/Techniques*