Regulation of the *glpFK* Operon of *Escherichia coli* K-12 and Characterization of its Gene Products

by

Deborah L. Weissenborn

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 and Nutrition

APPROVED:

T.J. Larson, Chairman

D.R. Dean

J.L. Hess

T.O. Sitz

M. Potts

April, 1990

Blacksburg, Virginia

Regulation of the *glpFK* Operon of *Escherichia coli* K-12 and Characterization of its Gene Products

by

Deborah L. Weissenborn

Committee Chairman: Timothy J. Larson
Biochemistry and Nutrition

(ABSTRACT)

The *glpF* gene, which encodes a cytoplasmic membrane protein that facilitates the diffusion of glycerol into the cell, and the *glpK* gene, which encodes glycerol kinase, map near minute 88 on the linkage map of *Escherichia coli* K-12. In the present work, the nucleotide sequence of the 843 base pair *glpF* gene, 430 base pairs of the *glpK* gene, and the intervening sequence between the two genes were determined. The control region for the *glpFK* operon was identified and sequenced. The *glpK* gene product was purified to near homogeneity by streptomycin sulfate and ammonium sulfate fractionation with subsequent DEAE Sephadex chromatography. N-terminal amino acid analysis identified the startpoint of translation for the *glpK* gene. The transcription start site was identified 71 base pairs upstream from the proposed translation start codon for *glpF*. Preceding the transcription start site were -10 and -35 sequences similar to the consensus sequences for gram$^-$ bacterial promoter elements. DNase I

footprinting was used to identify two binding sites for the cAMP-cAMP receptor protein (CRP) complex upstream from and overlapping the putative -35 sequence. Four binding sites for the *glp* repressor were located sequentially along the DNA extending from -89 (relative to the start point of transcription) to within the -10 region. Two additional repressor binding sites were identified within the *glpK* coding region. Interaction of these operator sites with those in the control region was identified. The affinity of the *glp* repressor for the control regions of the *glpD*, *glpACB-glpTQ*, and *glpFK* operons was compared by titration studies using a strain harboring a *glpT:lacZ* fusion and a $glpR^n$ mutation.

Acknowledgements

I would like to express my deep appreciation to my advisor, Dr. Timothy Larson, for all of his patience, guidance, and support during my time under his tutelage. He has provided an example of integrity and ability to which I will always strive.

I would also like to thank the members of my committee, Drs. D.R. Dean, J.L. Hess, T.O. Sitz, and M. Potts for their help and advice. Thank you to other members of this faculty for their counsel and friendship.

Thanks, too, to the members of our lab, Shanzhang Ye, Ali Bhattacharya, Denise Austin, and Won Oh, for making the lab an efficient, pleasant place to work and learn. Special thanks to Ali Bhattacharya and Denise Austin for providing unpublished results used in this dissertation.

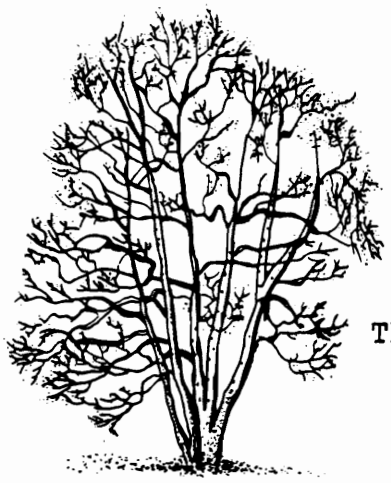
I would like to thank the proficient and helpful Biochemistry and Nutrition office staff for all their help during the last four years.

I would also like to thank Dave Ross for all of his technical assistance in the preparation of this dissertation.

Thank you to my good friends who have supported and helped me through these years, especially Karen and Dave Ross, Deb Peterson, Dan Paris, Pat Trihey, Dr. Gary Evans, and the members of OSLC.

To my family, especially my wonderful mother, Gert "Masie" Weissenborn: your faith in me and the support you have given me have made this possible.

And to my husband, Neil Ames: no one could have helped me more, believed in me more, or stood by me as well as you. I thank you with all my heart.



This work is dedicated to the memory of
my father, Otto H. Weissenborn.
He never took away the hammer.

List of Abbreviations

Ap	ampicillin
ATP	adenosine 5'-triphosphate
bp	base pairs
cAMP	cyclic adenosine 3',5'-monophosphate
Cm	chloramphenicol
dNTP	deoxynucleoside triphosphate
ddNTP	dideoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetate
FBP	fructose 1,6-bisphosphate
Glycerol-P	glycerol 3-phosphate
IPTG	isopropylthio-ß-D-galactopyranoside
kb	kilobase pairs
Km	kanamycin
K_m	Michaelis constant
NAD	nicotinamide adenine dinucleotide
ONPG	orthonitrophenyl-ß-D-galactopyranoside
PAGE	polyacrylamide gel electrophoresis
SDS	sodium dodecyl sulfate
Tc	tetracycline
Tris	tris(hydroxymethyl)aminomethane
X-gal	5-brome-4-chloro-3-indolyl-ß-D-galactopyranoside

Table of Contents

List of Illustrations

List of Tables

INTRODUCTION

Escherichia coli is able to utilize glycerol as a carbon source through the action of a cytoplasmic membrane protein (glycerol facilitator), glycerol kinase, and glycerol 3-phosphate dehydrogenase. The *glpF* and *glpK* genes encode the facilitator and the kinase, respectively, and are part of the *glp* regulon, which is responsible for the catabolism of *sn*-glycerol 3-phosphate (glycerol-P) and its derivatives. The regulon is comprised of at least five operons and is under control of the *glpR*-encoded repressor. Glycerol-P, the product of the reaction catalyzed by glycerol kinase, is the inducer for the regulon. It is a precursor for phospholipid biosynthesis and is found in an esterfied form in biological membranes. When glycerol-P is present in excess, it is degraded as a carbon source. The fact that glycerol-P has a dual role necessitates regulation of the expression of the catabolic *glp* genes. In addition to the negative regulation provided by the repressor, members of the *glp* regulon are subject to catabolite repression by glucose; thus these genes are maximally expressed only in the presence of the cAMP-CRP complex. Both the repressor and the cAMP-CRP complex exert their influence by binding to specific sequences of DNA. In order to study the regulation of the *glpF* and *glpK* genes, it was necessary to determine the nucleotide sequence of the

genes and their organization on the chromosome. I report here the sequencing of the genes, the organization of the control region of the operon, and the interaction of the regulatory elements with the operon. The amino acid sequence for the GlpF protein was deduced from the nucleotide sequence and its structure proposed. The substrate-induced alteration of the native state of glycerol kinase was examined.

LITERATURE REVIEW

GLYCEROL METABOLISM IN *ESCHERICHIA COLI*

Escherichia coli is a facultative anaerobic bacterium, able to metabolize a number of different carbon sources in the presence or absence of oxygen. Under aerobic conditions, oxygen serves as the terminal electron acceptor as part of the substrate is oxidized to CO_2. Nitrate or fumarate serves as the electron acceptor under anaerobic conditions. Although glucose is the substrate of choice for *E. coli*, a variety of other substrates, including *sn*-glycerol 3-phosphate (glycerol-P), glycerophosphodiesters, and glycerol can serve as sole carbon and energy sources (1, 2). These substrates are products of phospholipid and triglyceride degradation and are dissimilated by the components of the *glp* regulon (Fig. 1). The entry of glycerol into the cell is mediated by a cytoplasmic membrane protein, the glycerol facilitator (3, 4). The facilitator is the only known cytoplasmic membrane porin in *E. coli* and catalyzes the energy-independant equilibration of intracellular and extracellular glycerol concentrations (2, 3, 5). In addition to glycerol, the facilitator also allows the transport of glyceraldehyde, glycine, urea, and several polyhydric alcohols (6). The glycerol facilitator has been

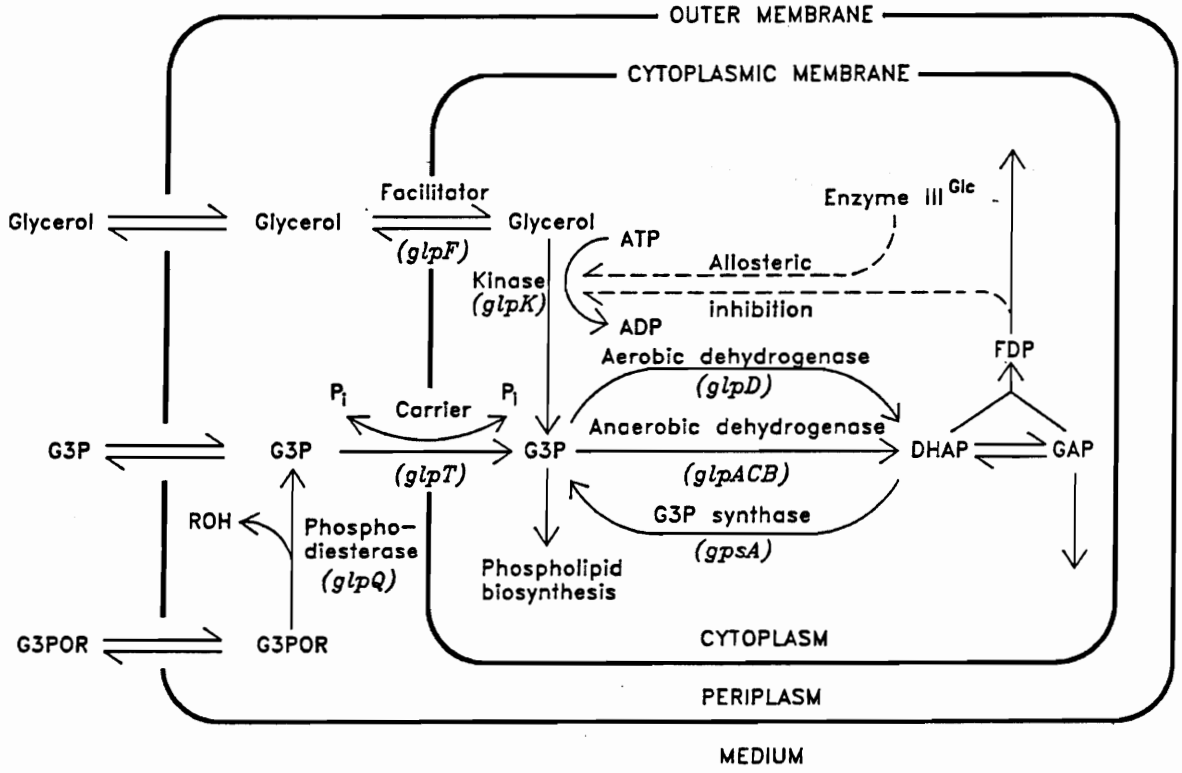


Fig. 1. **Metabolism of glycerol-P in *E. coli*.** Abbreviations: G3P, glycerol 3-phosphate; G3POR, glycerophosphodiester; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; FDP, fructose 1,6-bisphosphate.

expressed in maxicells and identified as a membrane protein with an apparent molecular weight of 25,000 by using SDS-PAGE (7). In the process of cloning and sequencing random pieces of bent DNA, Muramatsu and Mizuno observed an open reading frame coding for 281 amino acids downstream from the BENT-6 segment, which had been identified by two-dimensional polyacryamide gel electrophoresis (8, 9). The position of the reading frame (upstream from *glpK*) led them to suggest that this sequence corresponds to the *glpF* gene. Similarities between the deduced amino acid sequence of the open reading frame and that of two other membrane proteins, the bovine lens major intrinsic protein and the soybean nodulin-26 protein, were observed (10). Each of these proteins is proposed to function in the transport of small molecules across membranes (11, 12).

Once glycerol enters the cell, it is trapped by the MgATP-dependent phosphorylation catalyzed by glycerol kinase (EC 2.7.1.30, ATP:glycerol 3-phosphotransferase; K_m glycerol: 10 μM; K_m ATP: 80-100 μM and 400-500 μM). The addition of the charged phosphate group prevents the passage of glycerol-P back through the cytoplasmic membrane. *E. coli* mutants which lack glycerol kinase are not able to accumulate radioactive material when incubated with [^{14}C]glycerol (13). Glycerol kinase has been termed the pacemaker for glycerol metabolism and is subject to two forms of noncompetitive inhibition (Fig.

1), an unusual means of control for a catabolic enzyme. Fructose 1,6-bisphosphate (FBP), an intermediate in the metabolism of glucose and in gluconeogenesis, is responsible for feedback inhibition of the kinase (K_i=0.5 mM) (14-16). Strains which constitutively express a mutated *glpK* gene that encodes glycerol kinase refractory to this inhibition quickly die when exposed to glycerol as the carbon source (14, 15). This lethal response is not due to the buildup of glycerol-P because mutations in the enzyme responsible for the aerobic dehydrogenation of glycerol-P result only in growth stasis, not death, in the presence of glycerol (1). Methylglyoxal, which is derived from dihydroxyacetone phosphate (the product of the dehydrogenase reaction), is the toxic factor (16). Its lethal effect is thought to be at the level of protein synthesis.

Growth on glucose results in a second form of allosteric inhibition of glycerol kinase. This inhibition is exerted by the unphosphorylated form of enzyme III^{glc} of the bacterial phosphotransferase system (K_i=10 μM at pH 7.0) (17, 18). When glucose is available to the cell, the unphosphorylated form of enzyme III is the predominant species. The resulting inhibition of glycerol kinase prevents the use of glycerol as a carbon source when the preferred substrate, glucose, is present.

Purified glycerol kinase subjected to equilibrium

centrifugation displays a molecular weight of 217,000; SDS-PAGE gives one band of molecular weight 55,000, indicating a tetrameric protein (19). Sequencing of the *glpK* gene revealed an open reading frame which codes for a polypeptide of 502 amino acids with a calculated molecular weight of 56,106 (20). The tetrameric form of the protein undergoes reversible dissociation into a dimeric form at low enzyme concentrations (K_D=2.8 x 10^{-8} M) (21). The dimer exhibits reduced binding of one of the noncompetitive allosteric inhibitors of the enzyme, FDP. The catalytic activities of the dimer and tetramer are approximately equal, but inhibition of the tetramer occurs upon binding of FDP (21). Crystallization and X-ray studies in the presence of glycerol show that the kinase consists of pairs of nonidentical subunits with at least one dimer in an asymmetric unit (22). The kinase is able to catalyze phosphorylation of not only glycerol, but also of 28 nitrogen, sulfur, and alkyl-substituted analogues of glycerol, and thus is useful in the stereospecific synthesis of phosphorylated organic substances (23).

A calcium-dependent periplasmic phosphodiesterase (EC 3.1.4.2) specific for glycerophosphodiesters provides the cell with another source of glycerol-P (Fig. 1) (24, 25). The phosphodiesterase is able to catalyze hydrolysis of substrates with hydrophilic alcohols; K_m values range from 0.24 mM for glycerophosphoethanolamine to 1.0 mM for glycerophospho-

inositol (26). The enzyme exhibits a subunit molecular weight of 40,000 on SDS-PAGE; gel filtration chromatography and nondenaturing gradient polyacrylamide gel electrophoresis indicate that the enzyme is dimeric in its native state (26). Analysis of the protein sequence deduced from the nucleotide sequence indicates that the enzyme is made in a precursor form of molecular weight 40,362; cleavage of the 25 amino acid signal sequence results in the mature form of the protein of molecular weight 37,722 (Eiglmeier, Boos, and Larson, unpublished results).

Glycerol-P obtained extracellularly or from the phosphodiesterase reaction is transported across the cytoplasmic membrane via active transport catalyzed by the glycerol-P permease (27). After solubilization of the cytoplasmic membrane at 50°C, the permease displays a molecular weight of 44,000 by SDS-PAGE (28). Following sample treatment at 95°C, a molecular weight of 33,000 is seen by SDS-PAGE. This observation and the fact that missense mutations mapping in the carboxy terminal portion of the permease result in a negative dominant phenotype suggests an oligomeric structure for the permease (28, 29). The nucleotide sequence of the gene encoding the permease has been used to deduce the amino acid sequence of the protein (30). The permease consists of 402 amino acids with a calculated molecular weight of 50,251. The proposed structure of this

hydrophobic protein is similar to that of other transport proteins. Six hydrophobic domains which have the potential to traverse the cytoplasmic membrane in a α-helical conformation are located on each side of a central hydrophilic region (31). The permease-mediated transport of glycerol-P (K_T=20 μM in intact cells) occurs by an antiport mechanism with P_i exported out of the cytoplasm as glycerol-P enters through the cytoplasmic membrane (Fig. 1) (25, 32).

Under aerobic conditions, cytoplasmic glycerol-P is converted to dihydroxyacetone phosphate (DHAP) by the aerobic dehydrogenase (2). Examination of the nucleotide sequence of the gene which encodes the dehydrogenase revealed a protein of 504 amino acids with a molecular weight of 56,000 (33). Revision of the nucleotide sequence shows a protein of 501 amino acids with a calculated molecular weight of 56,747 (D. Austin and T.J. Larson, unpublished results). The enzyme, which exists as a dimer, has an apparent K_m for glycerol-P of 2 mM (2). The aerobic dehydrogenase transfers electrons to cytochrome oxidases or the nitrate reductase complex (2).

Under conditions of limited oxygen availability, an anaerobic dehydrogenase converts glycerol-P to DHAP, conducting electrons to the fumarate or nitrate reductase complexes (2, 27). The anaerobic enzyme has an apparent K_m for glycerol of 0.3 mM. It consists of three subunits with molecular weights of 62,000, 43,000, and 44,000 (29,34). The

product of the dehydrogenase reaction, DHAP, is then isomerized to glyceraldehyde 3-phosphate, an intermediate in both the glycolytic and gluconeogenic pathways.

Cytoplasmic glycerol-P serves both as a carbon and energy source and as a precursor for phospholipid biosynthesis and is found in an esterified form in biological membranes (35). Thus, the intracellular concentration glycerol-P must be maintained at a level which supports phospholipid biosynthesis. Glycerol-P acyltransferase, which catalyzes the first committed step in phospholipid biosynthesis, has a K_m for glycerol-P of 150 µM (36). A second enzyme involved in phospholipid synthesis, phosphatidylglycerophosphate synthetase, has a K_m of 320 µM for glycerol-P (37). The intracellular concentration of glycerol-P is estimated at 120 to 240 µM for cells grown on glucose and 2.8 mM for cells grown on glycerol (1). When the concentration of glycerol-P falls below that needed to support phospholipid biosynthesis, DHAP is converted to glycerol-P by the action of glycerol-P synthase (Fig. 1). The enzyme is subject to regulation by feedback inhibition by glycerol-P (K_i=4.4 µM) (38).

GENETIC REGULATION OF GLYCEROL METABOLISM

The ability of *Escherichia coli* cells to grow and divide

is largely determined by the ability of the bacteria to respond to environmental conditions and nutrient supply. To provide the organism with the proteins necessary for maximum growth and yet avoid wasteful synthesis of unnecessary proteins, the bacteria must be able to regulate the expression of genes in response to the existing physiological conditions. In multistep degradative systems, the availability of the molecule to be degraded often determines the regulation of expression of the genes involved in the degradative pathway. This type of regulation, in the form of both induction and repression, is typified by that found in the lactose system of *E. coli*.

The concept of the inducibility of expression was first presented by Jacob, Monod, and collaborators (39). Chromosomal mutations which resulted in constitutive synthesis of ß-galactosidase were identified at a locus termed the *i locus* for "inducibility". The product of this locus (now termed *lacI*) affects the synthesis of three *lac* genes, *lacZ* (which encodes ß-galactosidase), *lacY* (encoding galactoside permease), and *lacA* (encoding thiogalactoside transacetylase). In strains which are $lacI^+$, expression of the three genes is inducible in the presence of lactose. Expression of the genes is constitutive in *lacI* strains. Mapping experiments revealed that the three genes are arranged in a linear fashion on the *E. coli* chromosome; mutations in one locus, the operator,

affect the synthesis of all three genes (40). The three genes are thus organized in an operon, a unit of coordinately expressed genes which are preceded by operator and promoter sites.

Expression of the *lac* genes is dependent on the presence of an inducer and a second regulatory element, the cAMP-CRP complex. After lactose enters the cell by the action of the permease, it is converted to the true inducer of the system, allolactose, through the action of ß-galactosidase (41, 42). In the absence of the inducer, the *lac* repressor binds to the *lac* operator, preventing transcription. Allolactose binds to the repressor, changing its conformation such that it is no longer able to bind to the operator (43). This allows RNA polymerase access to the *lac* promoter. However, RNA polymerase exhibits a low affinity for the *lac* promoter and binds to the promoter only in the presence of the cAMP-CRP complex (44-46). Following the binding of cAMP-CRP and RNA polymerase, the genes of the *lac* operon are expressed.

Subsequent to the proposal of the operon model of positive and negative regulation in the lactose system, similar situations were identified in the arabinose and galactose catabolic systems (47-50). In each situation, expression of the genes of the operon is under both negative and positive control. Stringent control over the expression of these and other systems is provided by the presence of

multiple repressor binding sites (51-55). These operator regions can be adjacent to each other, as in bacteriophage lambda. Three tandem operators, O_{R1}, O_{R2}, and O_{R3} (separated by 3- and 7-bp spaces), lie between two divergent promoters (56). The repressor is a dimer; the amino domain of each subunit binds one-half of each operator. Thus, one repressor dimer binds one operator. O_{R1} has the highest intrinsic affinity for the repressor, binding the repressor approximately 10 times more tightly than do the other operators (57). However, two repressor dimers bind to O_{R1} and O_{R2} in a cooperative manner due to interaction of the carboxy domains of the dimeric repressor molecules and the two operators are filled virtually simultaneously. This cooperativity results in very efficient repression of the system. A very small increase in the concentration of active repressor greatly enhances repression. When O_{R1} and O_{R2} are separated by an integral number of helical turns of DNA, protein-protein interaction between repressor molecules still causes cooperative binding to the operators (58). This binding causes the intervening segment of DNA to form a loop both *in vivo* and *in vitro*. Repression loops formed between a regulatory protein and two widely separated operators are also seen in the *lac*, *ara*, *gal*, and *deo* systems (52-55).

In the wild-type *lac* operon three operator sequences exist. The first operator (O1) overlaps the startpoint of

transcription, while the second functional operator (O2) lies 401 bp downstream within the *lacZ* coding region, and a third operator (O3) lies 93 bp upstream of O1 (52, 59). The formation of a repression loop by the interaction of O1, O2 and the *lac* repressor has been verified by electron microscopy (60). *In vivo* binding studies show that repressor binding at O2 strengthens binding at O1 approximately 3-fold and blocks elongation in the expression of *lacZ* (61). Mutations in O2 result in a significant decrease in repression of the *lac* operon *in vivo* (62). Simultaneous binding of the *lac* repressor to O1 and O3 was observed in both *in vitro* and *in vivo* experiments (59, 63, 64). Deletion of O1 greatly reduced the binding of the repressor to O3 (64).

Binding of the AraC protein to two operator sites (O_2 and I) is necessary for repression of the P_C and P_{BAD} promoters in the arabinose system (65). The *araI* site is located between the two promoters while the O_2 site lies 200 bp downstream from the P_C promoter, within the transcribed region of the *araC* gene. The formation of a loop by the AraC/operator interaction prevents AraC from acting as a transcription activator. When operator I is deleted, O_2 binds AraC extremely weakly *in vivo*, demonstrating cooperative binding (65).

In the galactose system, a looping mechanism is proposed involving O_E, located upstream of the promoters for the *gal*

operon, and O_I, located 114 bp downstream within the *galE* structural gene (66). The two operators exhibit unequal affinities for the *gal* repressor, with the internal site functioning primarily under conditions of severe repression.

The expression of the *deo* operon coding for the enzymes involved in the catabolism of deoxyribonucleosides and ribonucleosides is controlled by the binding of the *deo* repressor to three operator sites (67, 68). Binding of the repressor to O_1 and O_2, which overlap the two promoters for the system, is necessary for efficient repression of the operon. Looping of the intervening DNA when the repressor is bound to the two sites has been identified by electron microscopy (68). In addition, looping is seen when the repressor is bound to either of these two operators and to a third operator, O_E, located upstream from the promoters. If all three operator sites are present, not only single loops form, but double loops result from the simultaneous binding of the repressor to the three sites (68).

These complex systems are necessary for efficient regulation of these operons. Regulation of the expresson of the genes involved in glycerol metabolism in *E. coli* can be expected for two reasons. First, it is economical for the organism to synthesize the *glp* proteins only when glycerol or glycerol-P is present in the medium. Second, in order to maintain phospholipid biosynthesis, *E. coli* must regulate the

expression of the catabolic *glp* genes and thus maintain the intracellular concentration of glycerol-P. Without regulation, degradation of glycerol-P would continue even at decreased concentrations of glycerol-P, resulting in a reduction of phospholipid biosynthesis. Conversely, if excess glycerol-P is present, the organism must be able to induce synthesis of the catabolic genes.

Inducibility of the *glp* genes was first observed by Koch and collaborators in 1964 (69). Levels of glycerol kinase, the aerobic glycerol-P dehydrogenase, and the glycerol-P transport system in cells grown on casein hydrolysate were very low when compared to those seen when cells were grown on media containing glycerol or glycerol-P. Evidence that one repressor was responsible for this inducibility was obtained by the use of mutants which constitutively utilized glycerol-P. When all three proteins were produced constitutively, it was determined that a single mutation was responsible for the phenotypic change. It was further determined that the three proteins were encoded by more than one operon. The addition of glucose to the growth media resulted in repression of the activity of the three proteins. However, this catabolite repression was unequal; kinase activity was repressed 20-fold, transport repressed 8-fold, and the dehydrogenase repressed only 2-fold. This differential sensitivity indicated that each protein was encoded by a separate operon

(69). A network of operons which responds to the same regulatory signals but are located at different regions of the chromosome is termed a regulon. Hence, these operons are part of the *glp* regulon. Since the early work by Koch and his collaborators, it has been determined that the *glp* genes which encode the glycerol-P catabolic proteins are organized into five known operons located at three different positions on the linkage map of *E. coli* (Fig. 2) (2). Near minute 49, the *glpACB* operon encoding the anaerobic dehydrogenase is located adjacent to, and is transcribed divergently from, the *glpTQ* operon encoding the glycerol-P permease (28) and the glycerophosphodiesterase (29). The *glpD* gene (70), which encodes the aerobic dehydrogenase, maps near minute 75 and is transcribed divergently from the *glpEGR* genes. The *glpR* gene encodes the repressor (71), while *glpE* and *glpG* encode genes whose functions are not known. Preliminary results indicate that these genes are also organized in an operon (D. Austin and T.J. Larson unpublished results). Finally, the *glpFK* operon is located near minute 88 on the linkage map (7, 72, 73).

Negative regulation of the *glp* genes is provided by the *glpR*-encoded repressor. Early *in vivo* experiments indicated that the repressor exhibits apparent differential binding affinity for various *glp* operons (75). Assays of glycerol-P transport, glycerol kinase activity, and glycerol-P

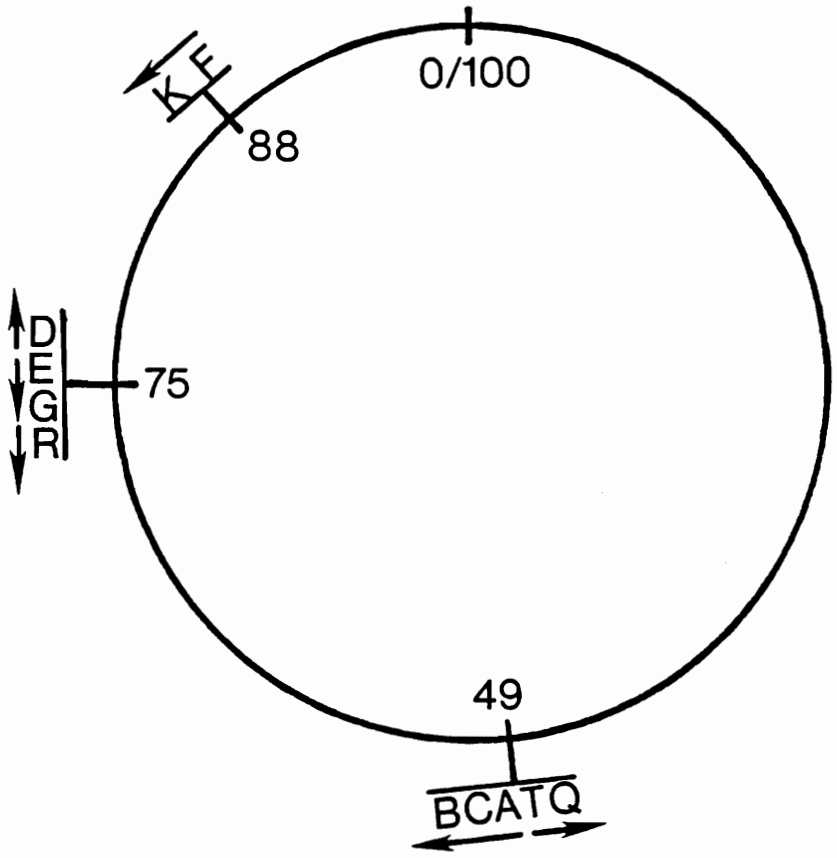


Fig. 2. Genetic map of *E. coli* showing the locations of the *glp* genes. Arrows indicate the direction of transcription of the genes. *Q* denotes the phosphodiesterase; *T*, glycerol-P permease; *ACB*, anaerobic dehydrogenase; *D*, aerobic dehydrogenase; *R*, repressor; *E* and *G*, proteins of unknown function; *K*, glycerol kinase; and *F*, glycerol facilitator.

dehydrogenase activity were performed on cultures from two *E. coli* strains. One strain harbored a temperature sensitive *glpR* allele while the second carried a *glpR* deletion. Assays of cultures grown over a temperature range from 26.5°C to 42°C were used to develop specific activity repression ratios, where a fully derepressed state would have a repression ratio of 0, and higher represson ratios would indicate greater repression of gene expression. Results were interpreted to indicate that *glpD* is most responsive to the repressor, *glpT* is less responsive, and *glpK* is least responsive. The theory was proposed that *glpD* needs to be most tightly regulated to prevent degradation of glycerol-P required for the biosynthesis of phospholipids in the absence of exogenous glycerol and glycerol-P (75). The *glpE* and *glpG* genes also appear to be under the regulation of the repressor. The level of expression of the two proteins was found to be higher in strains haboring plasmids with a mutated *glpR* gene compared to those with a wild-type allele (76).

The *glpR* gene was cloned behind the strong P_L promoter of lambda, overexpressed, and purified to near homogeneity (74). Purified repressor exhibits a subunit molecular weight of 30,000 by SDS-PAGE (74). Examination of the protein under nondenaturing conditions indicates that the *glp* repressor exists as a tetramer. Recently, the interaction of the *glpR*-encoded repressor with two tandemly arranged operator

sequences of the *glpD* gene was demonstrated (77). A consensus binding sequence was proposed: TATGTTCGTT AACGAACATA. Binding of the repressor to the two operators appears to be cooperative; both sites are protected from digestion by DNase I at the same concentration of repressor (0.08 nM). It is thought that one subunit of the tetrameric repressor molecule binds to one half of the palindromic operator sequence.

The inducer for the *glp* system is glycerol-P. Although the addition of either glycerol or glycerol-P to a growth medium of casein hydrolysate will induce the system in a wild-type strain, glycerol is unable to act as the inducer in a *glpK* strain (69). Flow dialysis experiments indicate that four molecules of glycerol-P bind to each repressor tetramer; the apparent K_d is 31 µM (74). The *lac* repressor, which is the best characterized repressor for a catabolic system, exhibits a much lower K_d (0.6 µM) for allolactose (78). The high K_d for the *glp* repressor makes physiological sense in light of the fact that glycerol-P is used both for phospholipid biosynthesis and as a carbon and energy source. Intracellular glycerol-P must be maintained at a concentration which allows phospholipid biosynthesis without necessarily inducing the *glp* regulon. Although the affinity of glycerol-P for the *glp* repressor-operator complexes has not been quantified, the affinity of the inducer for the *lac* repressor-operator complex is almost 1000 times lower than it is for the

free repressor (78). If the situation in the *glp* regulon is similar, the concentration of glycerol-P needed to effect full induction of the regulon would be higher than that needed for synthesis of phospholipids. The binding sites for glycerol-P on the *glp* repressor have not been identified, but it is assumed that binding of the molecule changes the conformation of the repressor and thus decreases its affinity for the *glp* operators.

As in the case of the *lac* system, the genes of the *glp* regulon are also subject to catabolite repression by glucose; thus these genes are maximally expressed only in the presence of the cAMP-CRP complex (2). Evidence suggests that CRP stimulates transcription by first binding to a specific site on DNA near the promoter, then enhancing the binding of RNA polymerase (79, 80). A consensus binding sequence on the target DNA has been proposed (81):

AA-TGTGA------TCACA-TT

Each subunit of the dimeric protein is composed of two domains. The carboxy portion is involved with the binding of DNA, while cAMP binds to the amino portion of the subunit (79). The response of the *glp* operons to cAMP-CRP was first observed by Koch and coworkers in 1964 when they identified multiple operons in the *glp* system (69). The sensitivity of the *glp* genes to this regulation was the inverse of the sensitivity to the repressor. That is, expression of *glpK* was

found to be most sensitive and that of *glpD* least sensitive to catabolite repression (75). Expression of the anaerobic dehydrogenase (encoded by *glpA*) also responded to the presence of the cAMP-CRP complex (75).

The *glpD* and *glpACB* operons are also subject to respiratory control. The *glpD* gene is maximally expressed when molecular oxygen is available to serve as the electron acceptor while the *glpACB* operon is maximally expressed under anaerobic conditions with fumarate as the electron acceptor (1, 2, 75). The FNR protein, which is necessary for full expression of genes needed for growth under anaerobic conditions, is required for maximal expression of the *glpACB* operon (82, 83). Although it is not known how FNR regulates transcription in response to the availability of oxygen, the protein is known to assume a different conformation in the absence of oxygen (84). The amino terminus of FNR contains a cluster of four cysteine residues which is speculated to be part of a metal-binding site (85). It is thought that a metal ion, possibly Fe^{2+}, acts as the redox-sensitive group, binding to FNR under anaerobic conditions and converting the protein to a form in which it acts as a transcriptional activator of anaerobic genes (86). The amino acid sequence of FNR (deduced from the nucleotide sequence of the *fnr* gene) has three regions which display sequence homology with CRP, one of which corresponds to the site which bind DNA (87). Like CRP, the

carboxy portion of FNR binds to DNA. A consensus FNR-binding sequence which resembles the CRP-binding sequence has been proposed (85):

FNR A---TTGA----TATCAAT-A

CRP AA-TGTGA------TCACA-TT

Repression of genes required for growth under aerobic conditions is mediated by the presence of the *arcA* and *arcB* gene products when anaerobic conditions exist (89-91). ArcA and ArcB appear to belong to the family of two-component regulatory systems. The ArcB protein is thought to communicate the respiratory state to ArcA (90) which in turn exerts repression on the expression of *glpD* (91). Mutations in the *arcA* or *arcB* genes eliminate the anaerobic repression of the *glpD* gene (91).

While it is known that the regulation of glycerol-P and phospholipid biosynthetic enzymes is largely determined by product feedback inhibition, the regulation of the expression of the glycerol-P catabolic genes is not fully defined. It appears that the members of the *glp* regulon are subject to three systems of transcriptional regulation by the *crp*, *fnr*, and *arc* modulons in addition to the regulation exerted by the *glp* repressor.

The goal of this research was to examine the organization and regulation of the *glpFK* operon, to examine the gene products of the operon, and to compare the regulation by the

glp repressor of the *glpFK* operon with that of the other *glp* genes. This work will help clarify the complex regulation of this catabolic system.

MATERIALS AND METHODS

MICROBIOLOGICAL AND RECOMBINANT DNA TECHNIQUES

Materials: Yeast extract, tryptone, casamino acids, and agar were purchased from Difco Laboratories. Agarose, Tris, EDTA, 2-propanol, isoamyl alcohol, calcium chloride, magnesium chloride, glucose, sucrose, and glycerol were obtained from Fisher Scientific. Antibiotics, lysozyme, DNase I, cAMP, and glycerol-P were supplied by Sigma Chemical Company. New England Nuclear supplied [α-^{35}S]dATP and [α-^{35}S]dGTP. DNA sequencing kits were purchased from Pharmacia and U.S. Biochemical Corporation. X-Gal, IPTG, and Sequenase were obtained from U.S. Biochemical Corporation. RNasin and reverse transcriptase were from Promega Biotec Company. Restriction endonucleases, Klenow fragment of DNA polymerase I, and T4 DNA ligase were obtained from Boehringer, New England Biolabs, or U.S. Biochemical Corporation. The *glp* repressor and CRP were purified to homogeniety in this laboratory as previously described (74; D. L. Weissenborn and T.J. Larson, in preparation).

Bacterial strains and plasmids: The *E. coli* K-12 strains used were from the laboratory collection and are listed in Table

I. Plasmids Bluescript KS M13$^+$ (Stratagene; see Appendix), pBR322 (92), pPLa2311 (93), pK3 (7), pCB267 (102), pACYC184 (103), pSH50, pSH58, pSH79, pS2, pHH11, pSY12, pAB10, and pDA901 were from the collection of this laboratory. Plasmid pGC2 was obtained from G. Sweet, University of Konstanz, Federal Republic of Germany.

Media and growth conditions: Strains were maintained on LB plates (104) supplemented with 0.4% glucose and 10 µg/ml Tc, 50 µg/ml Km, 25 µg/ml Cm, and 100 µg/ml Ap where appropriate. Growth of liquid cultures was at 37°C in LB and appropriate antibiotics unless otherwise noted. When β-galactosidase activity was to be measured, cells were grown in LB with 0.4% glucose with antibiotics until stationary phase was reached. Cultures were centrifuged and the pellet resuspended in 10 mM $MgSO_4$. These suspensions were used to inoculate minimal medium A (104) containing 0.4% casamino acids and antibiotics at 50% of the concentrations listed above. The cultures were grown at 37°C until the assays were performed (OD_{600} 0.28-0.80). The induction medium used in the overproduction of glycerol kinase contained 3.2% tryptone, 2% yeast extract, M9 salts (104), 2 mM $MgCl_2$, 0.075 mM $CaCl_2$, 0.002 mM $FeCl_2$, and 50 µg/ml Km.

Table I. *E. coli* K-12 strains used in this study

Strain	Genotype	Source or reference
MC4100	F⁻ *araD139* Δ*(argF-lac)U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	94
TS100	MC4100 *glpR*	95
ECL1	HfrC *phoA8 relA1* spoT1 pit-10 *fhuA22 ompF627 fadL1701*	27
ECL4	ECL1 *glpK*	13
ECL89	ECL1 *glpR*^{n}	96
TL95	ECL4 *zih-730*::Tn*10 glpK*$^+$	P1·Tn*10* pool⟶ECL4
TL100	ECL4 *zih-730*::Tn*10 glpK*	P1·TL95⟶ECL4
TL298	TS100 *zih-730*::Tn*10 glpK*	P1·TL100⟶TS100
TL300	TL298 Tc^s	Fusaric acid selection
TST3	MC4100 *malT*::Tn*10*	97
TL658	MC4100 (*glpT-lacZ*)658	29
TJS1	ECL89 *malT*::Tn*10*	P1(TST3)⟶ECL89
TJS2	MC4100 *glpR*^{n} *glpT-lacZ malT*::Tn*10*	P1(TJS1)⟶TL658
CB806	Δ*lacZ lacY*$^+$ *galK phoA8 rpsL thi recA*56	98

Table I. *con't*

Strain	Genotype	Source or reference
JM109	(F' *traD36 proAB*$^{+}$ *lacI*^{q} *lacZ*ΔM15) *recA1 endA1 gyrA96 thi hsdR17 supE44 relA1 Δ(lac-proAB)*	99
DH5αF'	(F' φ80d*lacZ* M15) *endA1 recA1 hsdR17 supE44 thi1* λ^{-} *gyrA relA1 Δ(lacZYA-argF)U169*	100
N4830	*gal*$^{+}$ *Δ8 cI857 ΔBAM ΔH1*	101

Construction of recombinant plasmids: Recombinant plasmids were constructed using standard protocols. Plasmid DNA was digested with restriction endonucleases according to the recommendations of suppliers. Following electrophoresis on low gelling temperature agarose, fragments were eluted by the method of Wieslander (105) then ligated with T4 ligase following manufactures' guidelines.

Transformation: *E. coli* cells were made competant by the method of Silhavy (97). Plasmid DNA was mixed with *E. coli* cells and placed on ice for 45 min. The mixture was subjected to a heat shock at 42°C for 2 min, 1 ml LB added, and the cells allowed to grow at 37°C for 1 hour. Transformants were selected by growth on plates of LB medium supplemented with 10 µg/ml Tc, 50 µg/ml Km, 25 µg/ml Cm, 100 µg/ml ampicillin, 40 µg/ml X-Gal, and 0.5 mM IPTG where appropriate.

Isolation of DNA: Plasmid DNA utilized for transformation and verification of structure was isolated from 1.5-ml cultures (106). Single-stranded DNA for sequencing was prepared by the method outlined by Hackett, *et al.* (107). Double-stranded DNA for sequencing and DNase I footprinting was isolated from 100-ml cultures by a detergent/alkaline lysis method. Cells were pelleted at 5000 x *g*, 4°C, for 15

min. The pellet was resuspended in 2.5 ml SET buffer (25 mM Tris-HCl pH 7.5, 10 mM EDTA, 15% sucrose) with 2 mg/ml lysozyme and incubated 20 min in ice water. Incubation was continued for 10 min after the addition of lysis buffer (0.2 N NaOH, 1% SDS). After the addition of 3 ml 3 M sodium acetate (pH 5.2), the mixture was incubated on ice for 20 min. Cellular debris was removed by centrifugation (15,000 x *g*, 4°C, 20 min). The supernatant was subjected to sequential extractions with equal volumes of 1) phenol; 2) phenol:$CHCl_3$:isoamyl alcohol (25:24:1 v/v); and 3) $CHCl_3$:isoamyl alcohol (24:1). Two volumes 95% EtOH were added to the supernatant; the mixture was kept at -70°C for 20 min, then centrifuged for 15 min at 4°C, 12,500 x *g*. Following a wash with 70% EtOH, the pellet was dried, then resuspended in 300 µl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) with 10 µl of RNase (10 mg/ml). The mixture was incubated at 37°C for 30 min. After the addition of 350 µl sterile water, 160 µl 4 M NaCl, and 800 µl 13% polyethylene glycol, the solution was incubated for 60 min in ice water. Centrifugation at 10,000 x g for 10 min at 4°C was followed by a subsequent wash with 70% EtOH and a second centrifugation. The pellet was dried and resuspended in 150 µl TE buffer.

Agarose gel electrophoresis: Plasmid DNA was subjected to

analysis on 0.6 to 2.0% agarose gels at constant voltage (10-12 V/cm) at room temperature. TBE buffer (97) was used in both the gel solution and as the running buffer. Electrophoresis on low gelling temperature agarose gels was performed at 4°C for the elution of DNA fragments (105).

Sequencing of DNA: Double- and single-stranded plasmids were used as templates for the dideoxynucleotide chain termination method of Sanger, *et al.* (108). Bluescript KS M13+ (Stratagene; Appendix A) was used as the vector for double-stranded DNA fragments, while the cloning vectors M13mp18 and M13mp19 (107) were used for single-stranded sequencing. Sequencing reactions were analyzed on 6% wedge-shaped (0.4 - 0.8 mm) polyacrylamide gels. Electrophoresis was performed at 50-55°C for 1.5 hours for analyzing sequences up to 250 bp in length, and for up to 5 hours for analysis beyond that point.

DNase I footprinting: The 149 base pair (bp) *Sau*3A-*Bam*HI fragment of pDW51 (which contains the control region of the *glpFK* operon) and the 242 bp *Hin*dIII-*Eco*RV fragment of pDW30 (which contains a portion of the coding region of *glpK*) were subjected to DNase I footprinting analysis using modifications of previously described methods (77). The *Eco*RI-*Not*I fragment of pDW51 was isolated by elution from

low gelling temperature agarose (105). The fragment was then end-labelled at the *Eco*RI site with [α-^{35}S]dATP using the Klenow fragment of DNA polymerase I. The *Hin*dIII-*Eco*RV fragment of pDW30 was isolated in the same manner and end-labelled at the *Hin*dIII site with [α-^{35}S]dATP and [α-^{35}S]dGTP. Purified repressor (in the presence or absence of glycerol-P) or CRP (in the presence or absence of cAMP) was added to the labelled DNA in a reaction mixture containing 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 50 mM KCl, 4 mM $MgCl_2$, 5% glycerol, 0.025% Triton X-100, 100 µg/ml BSA, and 5 µg/ml sonicated salmon sperm DNA in a total volume of 0.1 ml. The reactions were incubated for 30 min at 37°C, followed by the addition of 0.5 ng DNase I. Incubation was continued for 3.5 min. The DNA was precipitated by the addition of 5 µg sonicated salmon sperm DNA, 25 µl 7.5 M ammonium acetate, and 400 µl ethanol. The DNA was resuspended in 4 µl Tris-HCl/EDTA:formamide (1:4), heated to 70°C for 2.5 min and analyzed on a sequencing gel. Co-electrophoresis of sequencing reactions was used to define the positions of the protected regions.

Isolation of RNA: RNA was isolated by a previously described method (109) from 10 ml cultures of DH5αF'(pGC2) which were grown to log phase in LB medium which contained 100 µg/ml Ap and 0.4% glucose or glycerol. After the

addition of 2.5 ml 2 M NaN_3, the cultures were quickly poured into centrifuge tubes containing 5.0 ml frozen Buffer A (50 mM Tris-HCl pH 8.0, 50 mM EDTA, 150 mM NaCl). Upon thawing of the buffer, the mixture was centrifuged at 7,000 x *g* for 10 min at 4°C. The pellet was resuspended in 1.0 ml Buffer B (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 mM NaCl). Buffer B containing 1.0% SDS was heated to 100°C and 1 ml added to the suspension, which was then heated at 100°C for 5 min. The mixture was allowed to cool slowly to 60°C, then extracted twice with 2 ml 60°C phenol (equilibrated with 50 mM sodium acetate, pH 5.5). A final extraction with 2 ml chloroform was followed by precipitation with 2 ml 2-propanol and centrifugation at 12,500 x *g* for 10 min at 4°C. The pellet was washed with 80% EtOH, recentrifuged, and dried. The pellet was resuspended in 75 µl TE buffer. A 40 µg aliquot was treated with RNase-free DNase for 40 min at 37°C. The reaction was stopped by the addition of an equal volume of 100 mM Tris-HCl pH 7.4, 10 mM EDTA, 1 M sodium acetate (pH 5.2), then sequentially extracted with equal volumes of phenol:$CHCl_3$ and $CHCl_3$. The RNA was precipitated by adding an equal volume of 2-propanol and centrifuged at 14,000 x *g* for 30 min. The pellet was washed with 80% EtOH, dried, and resuspended in 3.5 µl TE buffer.

Primer extension analysis: The primer extension analysis

protocol was adapted from the procedure of Alam, *et al.* (110). A mixture of 20 µg RNA and 30 ng synthetic oligonucletide primer was incubated for 2 min at 100°C, then cooled on ice. The synthetic primers are complementary to an area of the transcriptional unit of *glpF* (ATCCTGAAGAGTTAATGTTT; nucleotides 175-194, Fig. 5) or *glpK* (TCGAGCGCAACGATATATTT nucleotides 1081-1100, Fig. 5). The RNA-primer mix was added to a reaction mixture containing: 34 mM Tris-HCl pH 8.3, 50 mM NaCl, 6 mM $MgCl_2$, 5 mM DTT, 15 units reverse transcriptase, 40 units RNasin, 12.5 µCi [α-^{35}S]dATP, and 200 µM each dCTP, dGTP, and dTTP in a final volume of 40 µl. After incubation at 42°C for 30 min, an 800 µM dNTP chase (equal concentrations of dATP, dCTP, dGTP, and dTTP) was added and incubation was continued for 30 min at 30°C. Extraction with an equal volume of phenol:$CHCl_3$ preceded precipitation by the addition of an equal volume of isopropanol, centrifugation at 14,000 x g for 12 min, an 80% EtOH wash, recentrifugation, and drying. The pellet was resuspended in 6 µl TE:stop solution (0.3% xylene cyanol, 0.3% bromophenol blue, 10 mM Tris-HCl, 10 mM EDTA made up in 95% deionized formamide). Aliquots of this mixture were applied to a sequencing gel; co-electrophoresis of a sequencing reaction using the corresponding primer and area of DNA was used to determine the position of the transcription startpoint.

***Repressor titration by operators in* trans:** The response of the *glp* operons to the *glp* repressor was examined by using two methods employing titration of the repressor.

a. The promoter/operator regions from the *glpD*, *glpACB-glpTQ*, and *glpFK* operons were isolated by restriction endonuclease digestion and ligated into the promoter-probe vector pCB267 (102), creating transcriptional fusions of the *glp* promoters to the *lacZ* gene. The recombinant plasmids were cotransformed into CB806 and TL73 (Table I) with either a plasmid carrying the *glpR* gene (pSH58) or the vector plasmid without the *glpR* gene (pACYC184). The β-galactosidase activities were determined (as described in Biochemical Techniques).

b. The operator sequences from the *glpD*, *glpACB-glpTQ*, and the *glpFK* promoter/operator regions, the first operator of the *glpD* operon, and the operator sequences in *glpK* were isolated by restriction digestion, eluted from low gelling temperature agarose (105), and ligated into pBS+ (see Appendix). The resulting plasmids were used to transform strain TJS2 (Table I) which encodes a noninducible repressor and harbors a *glpT-lacZ* transcriptional fusion. The β-galactosidase activities of the transformed strain were determined (see Biochemical Techniques).

BIOCHEMICAL TECHNIQUES

Materials: Streptomycin sulfate, DEAE Sephadex, Sephadex G-200-120, Sephadex G-100-120, glycerol-P dehydrogenase from rabbit muscle, β-amylase (sweet potato), carbonic anhydrase (bovine erythrocytes), albumin (bovine serum and chicken egg), alcohol dehydrogenase (yeast), NAD, ATP, FDP, and ONPG were purchased from Sigma Chemical Company. Fisher Scientific supplied ammonium sulfate. Reagents for polyacrylamide gel electrophoresis were obtained from Bio-Rad.

Assay of glycerol kinase: The assay used to monitor glycerol kinase activity is based on the coupling of kinase reaction with the glycerol-P dehydrogenase reaction at pH 9.5 (111):

$$\text{glycerol-P} + \text{NAD}^+ \rightleftharpoons \text{dihydroxyacetone-P} + \text{NADH} + \text{H}^+$$

The reaction mixture contained 50 µmoles of ATP, 40 µmoles of $MgCl_2$, 5 µmoles of NAD, 1 mmole of hydrazine, 360 µmoles of glycine, 50 µmoles of glycerol, 0.14 mg of glycerol-P dehydrogenase from rabbit muscle, and 30-60 units glycerol kinase in a final volume of 2.0 ml. Reactions were initiated with the addition of glycerol and were performed at 30°C. NADH formation was followed spectrophotometrically at 340 nm. One unit of glycerol kinase activity is defined

the amount of enzyme which catalyzes the formation of 1 µmole of glycerol-P per min.

Determination of protein concentration: Protein concentration was estimated spectrophotometrically by the Bradford method (112) using bovine serum albumin as the standard.

Overproduction of glycerol kinase: A 5.0 kb *Eco*RI - *Pst*I fragment containing the *glpK* gene was ligated into the expression vector pPLa2311 (93), creating pDW23, which was then introduced into strain N4830 (Table I). The transformed strain was grown in 300 ml of induction medium for 4 hours at 30°C to OD_{600} = 0.65, then induced by the addition of an equal volume of medium preheated to 65°C. Growth was continued at 42°C for 3.25 hours, after which the cells were harvested by centrifugation.

Purification of glycerol kinase: Purification of the kinase was performed at 0-4°C as outlined below:

1. *Preparation of crude extract:* Cells (3.5 grams) from 600 ml of induced culture (see above) were suspended in 25 ml Buffer A (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 mM DTT) with 50 mM NaCl, then broken by passage through a French pressure cell. Centifugation at 40,000 x *g*

for 40 minutes removed cell debris. Buffer A was used to dilute the supernatant to 6.7 mg protein/ml.

2. *Streptomycin sulfate fractionation:* Keeping the crude extract (70 ml) in an ice bath, 7 ml of 20% streptomycin sulfate was added with stirring. After 1 hour, the solution was centrifuged at 12,000 x *g* for 20 min. The supernatant comprised the streptomycin sulfate fraction.
3. *Ammonium sulfate fractionation:* Ammonium sulfate (0.3 g/ml) was added to the streptomycin sulfate fraction and stirred on ice 30 min. Centrifugation at 17,000 x *g* for 15 minutes removed the insoluble material. Ammonium sulfate (0.45 g/ml) was added with stirring to the supernatant. After 30 min, the insoluble matter was pelleted by centrifugation at 17,000 x *g* for 15 min. The pellet was resuspended in 8 ml Buffer A.
4. *DEAE Sephadex chromatography:* Buffer A was used to dilute the ammonium sulfate fraction to 100 ml. This was applied to a 2.5 x 40 cm column of DEAE Sephadex at 38 ml/hr. The column was washed with 18 ml Buffer A with 50 mM NaCl, followed by elution with a 500 ml gradient of 50 mM to 500 mM NaCl in Buffer A. Kinase activity was eluted between 0.25 and 0.35 M NaCl. Fractions containing activity were pooled and concentrated by ammonium sulfate precipitation (0.55

g/ml). Aliquots of this DEAE Sephadex fraction were stored at -80°C, with or without 10% glycerol.

Protein gel electrophoresis: The buffer system of Laemmli (113) was utilized for SDS-polyacrylamide gel electrophoresis on 12% slab gels. Non-denaturing gradient gel electrophoresis was performed using the same buffer system without SDS (114). A 0-25% linear gradient of sucrose was used to stabilize slab gels containing a 5-25% linear gradient of polyacrylamide.

Gel filtration chromatography: Buffer C (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT) was used to equilibrate a 1.5 x 50 cm column of Sephadex G-200 at 7 ml/hr (4°C). Standard proteins and purified glycerol kinase were applied individually to the column and eluted wilh Buffer C. Subsequent applications of glycerol kinase were eluted with Buffer C supplemented with one of the following: 5 mM FDP; 10 mM glycerol; 20 mM phosphate; or 20 mM phosphate and 10 mM glycerol. In each case, the column was equilibrated with the supplemented buffer, and standard proteins were applied to the column. Elution volumes (V_e) were determined by counting the number of drops collected from application of the sample until peak absorbance at 280 nm was observed by a flow cell. The void (V_o) and total

(V_t) volumes were determined by application of plasmid DNA and ATP, respectively. The molecular weight of the glycerol kinase was estimated on a standard curve of log molecular weight *versus* K_{av}, where $K_{av} = (V_e - V_o)/(V_t - V_o)$.

N-terminal sequence analysis of glycerol kinase: Purified glycerol kinase was subjected to additional gel filtration chromatography on Sephadex G-100 to eliminate low molecular weight contaminants. The column was equilibrated with buffer D (10 mM $NaPO_4$, 100 mM NaCl, 0.5 mM EDTA, 0.1 mM DTT, 5 mM glycerol) and the protein applied from the bottom at a rate of 8 ml/hr. Protein was monitored at A_{280} with a flow cell; aliquots of the peak fractions were subjected analysis on SDS-PAGE. The peak fractions were combined and filtered through 0.2 µm filters. The 2.0 ml glycerol kinase sample was then dialyzed against 800 ml H_2O for 24 hours with four changes of water. The dialyzed glycerol kinase was then lyophilized in eight tubes. Approximately 625 pmol of the dried glycerol kinase was subjected 10 cycles of Edman degradation. The procedure was performed at the UM Protein Sequencing Facility, University of Michigan Medical School, Ann Arbor, MI.

Assay of ß-galactosidase activity: ß-galactosidase activity was measured spectrophotometrically by the absorption at 420

nm resulting from the hydrolysis of the substrate ONPG (104). Overnight cultures were grown in minimal media A (104) to OD_{600} 0.25-0.7. The assay mixture contained 0.05-0.5 ml of culture in Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol) in a total volume of 1 ml. The cell membranes were partially disrupted by the addition of two drops chloroform and one drop 0.1% SDS, and vortexing for 10 seconds. The mixtures were incubated for 5 minutes in a 28° waterbath. Reactions were started by the addition of 0.2 ml of ONPG (4 mg/ml in 0.1 M phosphate buffer pH 7.0), reaction times noted, and the reactions stopped by the addition of 0.5 ml of 1 M Na_2CO_3 after the development of yellow color. The reactions were centrifuged for 2 minutes at 14,000 x *g*. The absorbance of the reaction supernatants was measured at 420 nm. Miller units were determined by the formula:

$$\text{Units} = \frac{1000 \times OD_{420}}{t \times v \times OD_{600}}$$

where *t* is the time of the reaction and *v* is the volume of culture used in the assay, in ml. The OD_{600} value is the density of the culture. The OD_{420} value was determined by subtracting the OD_{420} value of the blank (a reaction mixture which does not contain any culture) from the observed OD_{420} of a reaction and is a function of the amount of *o*-nitrophenol produced. The units are proportional to the increase in *o*-nitrophenol per minute per bacterium.

RESULTS

STRUCTURE, ORGANIZATION, AND REGULATION OF THE *glpFK* OPERON

The *glpFK* operon of *Escherichia coli* is the second operon of the glycerol-P regulon examined as to the organization of its promoter/operator region and its regulation. Previous work by Ye and Larson (77) identified the control elements of the *glpD* operon and examined the regulation of the operon by the *glp* repressor. This work was designed to expand our understanding of the regulation of the glycerol-P regulon. To this end, it was necessary to obtain the DNA sequence of the operon, identify the startpoint of transcription and the promoter and operator elements, and characterize the interaction of regulatory molecules with the DNA. The results of this work follow.

Subcloning of* glpK *and* glpF*: To examine the organization of the *glpFK* operon, the genes were cloned and subclones generated for sequencing. Plasmid pK3 (7) served as a source of DNA for isolating much of the *glpFK* operon. Digestion of pK3 with *Eco*RI and *Pst*I yielded a 5.0 kb fragment which was ligated to similarly digested pBR322 (92), creating pSH50. This same 5.0 kb fragment was cloned

into *Eco*RI-*Pst*I digested pPLa2311 (93), yielding pDW23, where the P_L promoter of lambda is adjacent to the *Eco*RI site (Fig. 3). High levels of glycerol kinase activity obtained when strain TL300 (*glpK*) was transformed with pDW23 verified the presence of the *glpK* gene. Thermal inducibility of glycerol kinase in strain of N4830 (pDW23), a cryptic lambda cI857 lysogen (Table I), demonstrated that the direction of transcription of *glpK* is from the *Eco*RI restriction site towards the *Pst*I site. Subclones generated from pDW23 are shown in Table II. Plasmid pGC2 (7) served as the source of DNA upstream of the *Eco*RI-*Pst*I fragment indicated above. Subclones from pGC2 were generated and are shown in Table II.

Sequencing of the* glpFK *operon: Derivatives of pDW23 and pGC2 were used in sequencing the *glpF* gene and the 5' end of the *glpK* gene (Table II; Fig. 4). This sequence revealed an open reading frame of 281 codons upstream of *glpK* extending from base 201 to base 1043 (Fig. 5). The position of the open reading frame and its direction of transcription correspond to the promoter-proximal location of *glpF*, which encodes the glycerol facilitator (7). The nucleotide sequence of *glpF* obtained differs from that published (9) by one base (nucleotide 703, marked x on Fig. 5) and has been confirmed by multiple sequencing reactions on both strands

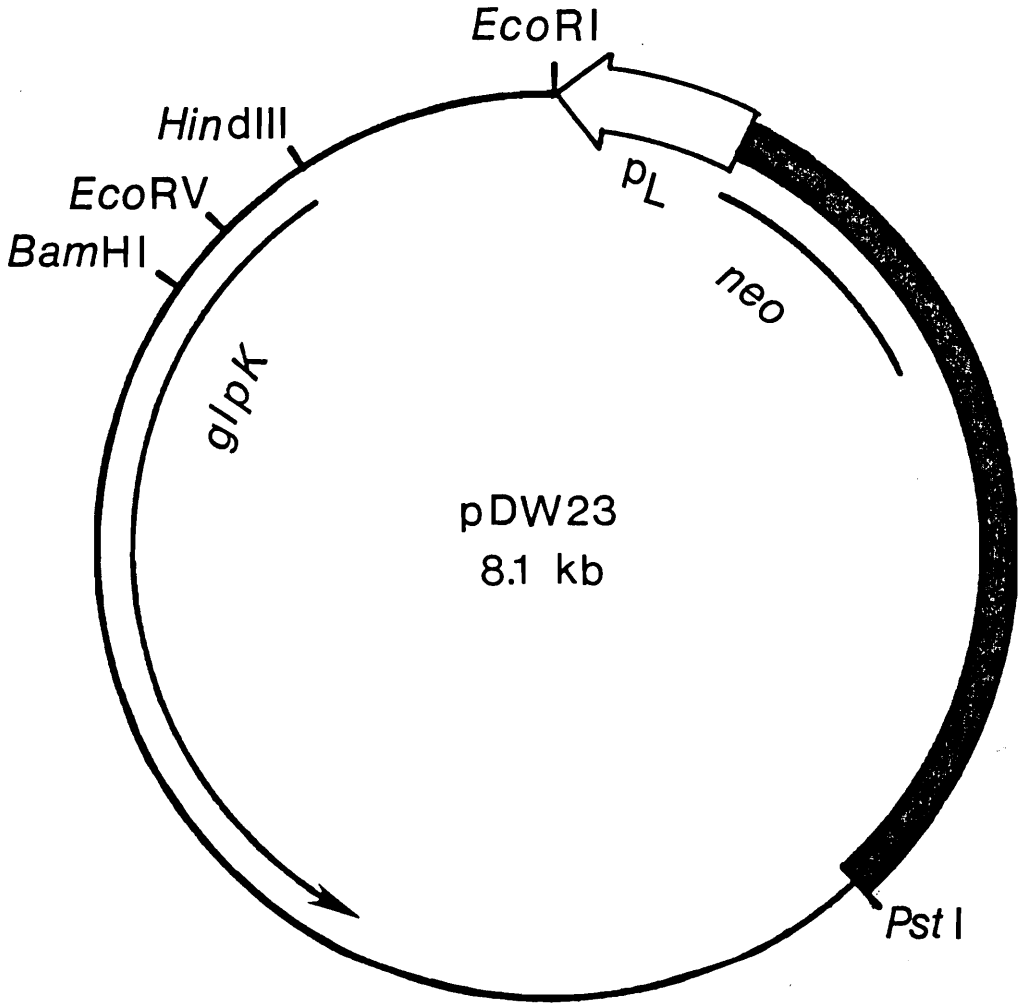


Fig. 3. Structure of pDW23. The arrow indicates the direction of transcription of the *glpK* gene.

Table II. Construction of plasmids and phages

Plasmid or phage[a]	Insert (bp)	Source of insert	Vector sites used
Bluescript			
pDW25	*Eco*RI-*Hin*dIII (793)	pDW23	*Eco*RI-*Hin*dIII
pDW28	*Hin*dIII-*Bam*HI (466)	pDW23	*Bam*HI-*Hin*dIII
pDW30	*Hin*dIII-*Eco*RV (242)	pDW28	*Eco*RV-*Hin*dIII
pDW31	*Eco*RV-*Bam*HI (224)	pDW28	*Bam*HI-*Eco*RV
pDW39	*Sau*3A-*Sau*3A (98)	pDW25	*Bam*HI
pDW50	*Eco*RV-*Eco*RI (430)[b]	pGC2	*Eco*RI-*Eco*RV
pDW51	*Sau*3A-*Bam*HI (191)	pGC2	*Bam*HI
M13			
42-19	*Eco*RI-*Sau*3A (445)	pDW25	*Eco*RI-*Bam*HI
43-18	*Eco*RI-*Sau*3A (347)	pDW25	*Eco*RI-*Bam*HI
44-18,-19	*Sau*3A-*Hin*dIII (446)	pDW25	*Bam*HI-*Hin*dIII
45-18,-19	*Sau*3A-*Hin*dIII (348)	pDW25	*Bam*HI-*Hin*dII

a: The vectors used were pBluescript KS(+) and M13mp18 or M13mp19.

b: Nucleotides 1-240 (Fig. 2) plus 190 bp from the vector.

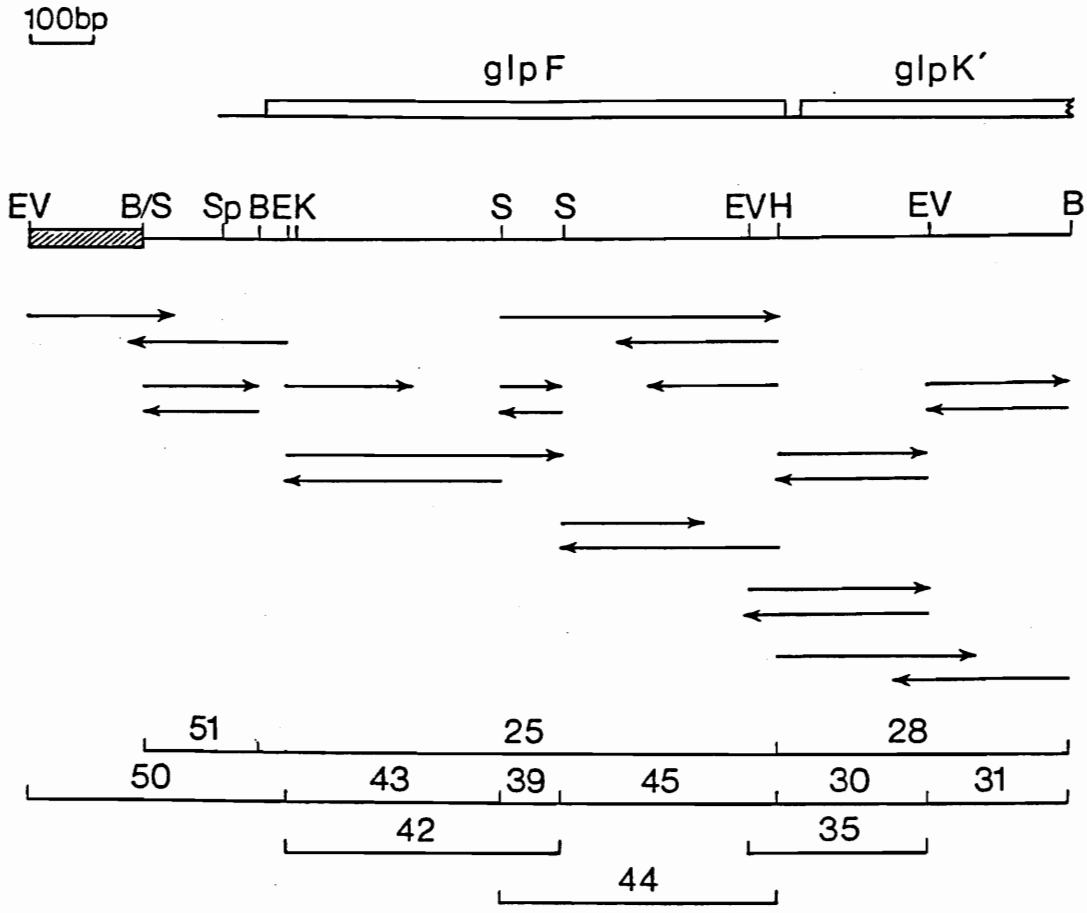


Fig. 4. Strategy for sequencing. The glycerol facilitator and kinase coding regions are indicated on the top portion of the figure. Plasmids or phages corresponding to regions sequenced are indicated on the lower portion of the figure; numbers refer to the recombinant plasmids shown in Table II. The arrows indicate the direction and extent of sequencing. The hatched box represents sequence from the vector pBR322 (92). Abbreviations used: B, *Bam*HI; E, *Eco*RI; EV, *Eco*RV; H, *Hin*dIII; K, *Kpn*I; S, *Sau*3A; Sp, *Sph*I.

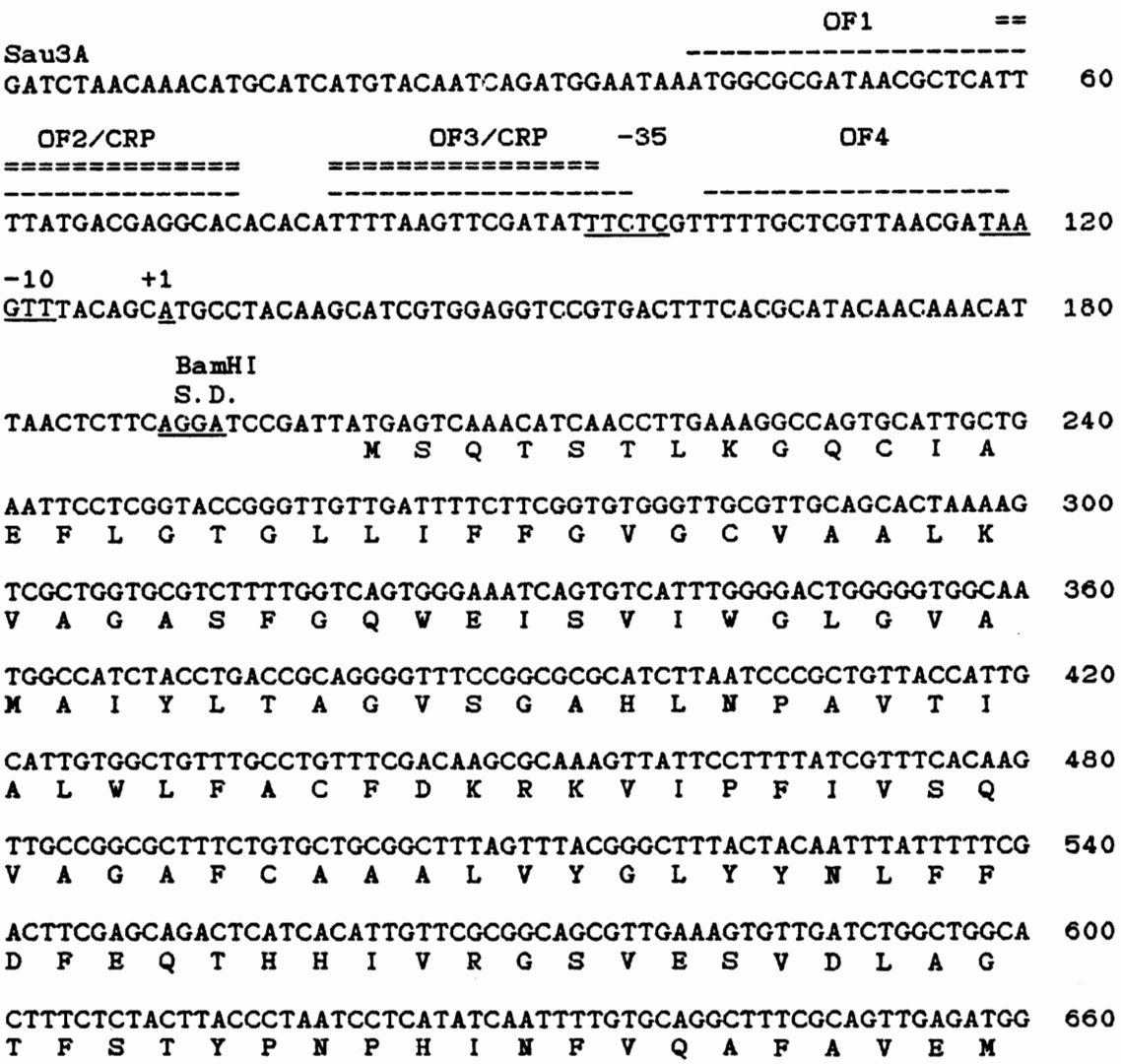
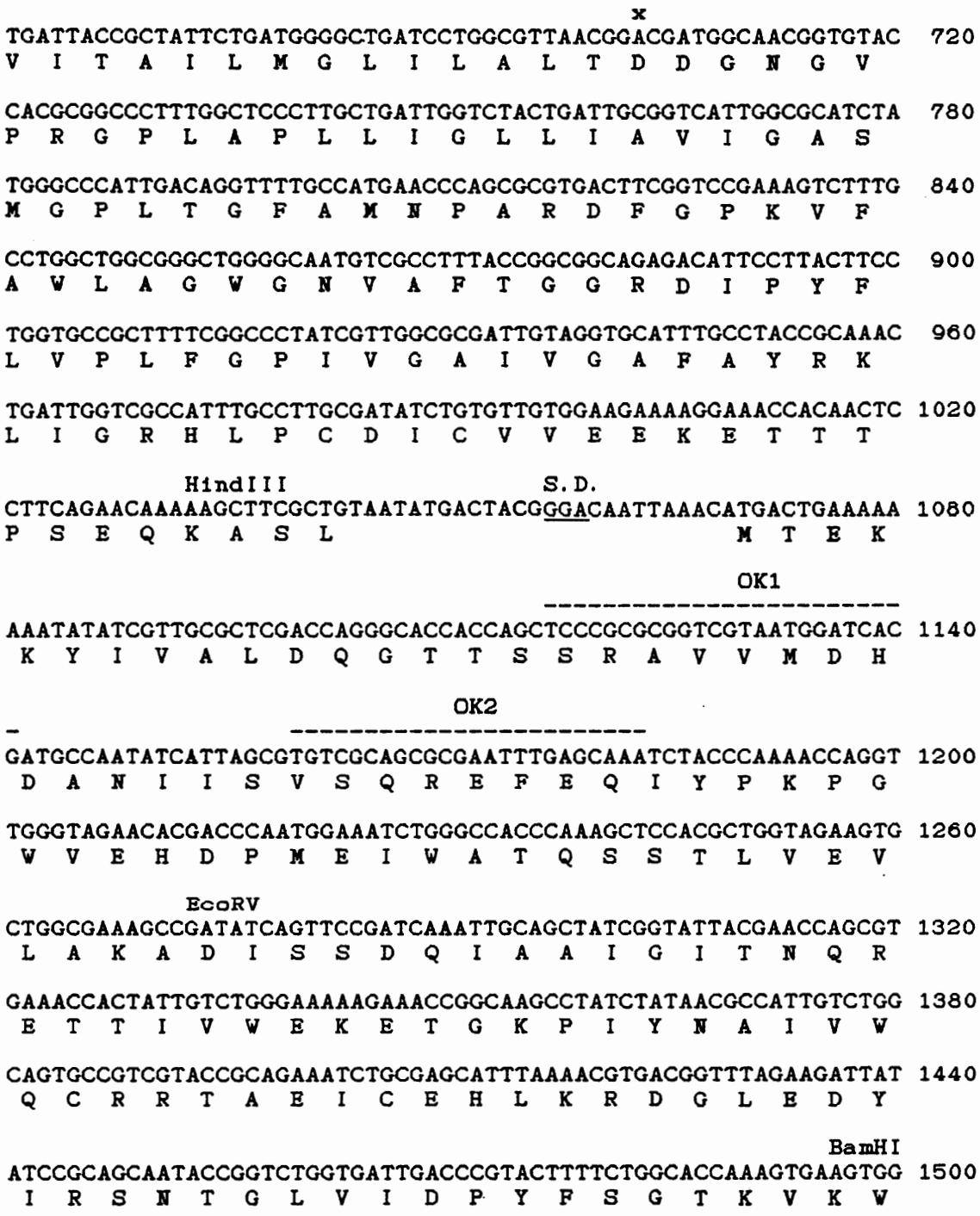


Fig. 5. Nucleotide sequence of *glpF*, the 5' end of *glpK*, and the control region of the *glpFK* operon. The 5' end of the *glpFK* transcript is indicated by +1. Sequences resembling the consensus sequences for the -10 and -35 regions and for ribosome binding (S.D.) are indicated by underlining. The restriction sites shown are those utilized in DNase I footprinting experiments. Single-dashed lines above the sequence indicate areas protected from DNase I digestion by the *glp* repressor (OF1-OF4). Double-dashed lines indicate protection by cAMP-CRP. The deduced amino acid sequence of the glycerol facilitator and the amino terminal portion of glycerol kinase are shown.


x
TGATTACCGCTATTCTGATGGGGCTGATCCTGGCGTTAACGGACGATGGCAACGGTGTAC 720
V I T A I L M G L I L A L T D D G N G V
CACGCGGCCCTTTGGCTCCCTTGCTGATTGGTCTACTGATTGCGGTCATTGGCGCATCTA 780
P R G P L A P L L I G L L I A V I G A S
TGGGCCCATTGACAGGTTTTGCCATGAACCCAGCGCGTGACTTCGGTCCGAAAGTCTTTG 840
M G P L T G F A M N P A R D F G P K V F
CCTGGCTGGCGGGCTGGGGCAATGTCGCCTTTACCGGCGGCAGAGACATTCCTTACTTCC 900
A W L A G W G N V A F T G G R D I P Y F
TGGTGCCGCTTTTCGGCCCTATCGTTGGCGCGATTGTAGGTGCATTTGCCTACCGCAAAC 960
L V P L F G P I V G A I V G A F A Y R K
TGATTGGTCGCCATTTGCCTTGCGATATCTGTGTTGTGGAAGAAAAGGAAACCACAACTC 1020
L I G R H L P C D I C V V E E K E T T T
HindIII
S.D.
CTTCAGAACAAAAAGCTTCGCTGTAATATGACTACGGGACAATTAAACATGACTGAAAAA 1080
P S E Q K A S L M T E K
OK1
AAATATATCGTTGCGCTCGACCAGGGCACCACCAGCTCCCGCGCGGTCGTAATGGATCAC 1140
K Y I V A L D Q G T T S S R A V V M D H
OK2
GATGCCAATATCATTAGCGTGTCGCAGCGCGAATTTGAGCAAATCTACCCAAAACCAGGT 1200
D A N I I S V S Q R E F E Q I Y P K P G
TGGGTAGAACACGACCCAATGGAAATCTGGGCCACCCAAAGCTCCACGCTGGTAGAAGTG 1260
W V E H D P M E I W A T Q S S T L V E V
EcoRV
CTGGCGAAAGCCGATATCAGTTCCGATCAAATTGCAGCTATCGGTATTACGAACCAGCGT 1320
L A K A D I S S D Q I A A I G I T N Q R
GAAACCACTATTGTCTGGGAAAAAGAAACCGGCAAGCCTATCTATAACGCCATTGTCTGG 1380
E T T I V W E K E T G K P I Y N A I V W
CAGTGCCGTCGTACCGCAGAAATCTGCGAGCATTTAAAACGTGACGGTTTAGAAGATTAT 1440
Q C R R T A E I C E H L K R D G L E D Y
BamHI
ATCCGCAGCAATACCGGTCTGGTGATTGACCCGTACTTTTCTGGCACCAAAGTGAAGTGG 1500
I R S N T G L V I D P Y F S G T K V K W

(Fig. 5 con't)

of the DNA. The sequence of the 5' end of *glpK* agrees with the published sequence of the gene from the HindIII restriction site 35 bp upstream of the translation start point to the *Bam*HI site 431 bp within the gene (Fig. 5) (20). A potential ribosome binding sequence (AGGA) for *glpF* is located 8 bp upstream from the ATG of the open reading frame. A potential ribosome binding site (-AGG-) for *glpK* is also seen 10 bp upstream of the coding region for the gene.

***Codon usage in* glpF *and* glpK:** Codon usage patterns for the *glpF* and *glpK* genes were examined and compared to trends in codon usage bias for other *E. coli* genes. (Table III). Relative synonymous codon usage (RSCU) values have been derived which reflect the frequency of use of a particular codon with respect to the expected frequency if all codons for an amino acid were used equally (115). Codon usage bias is influenced by the abundance of cognate tRNA species. There is a very high bias against 10 "rare" codons which are used very infrequently in very highly expressed genes (114). These rare codons have RSCU values of <0.05 in very highly expressed genes, while bias against these codons decreases (and the RSCU value increases) in nonhighly expressed genes. The *glpF* gene contains 12 of the rare codons with RSCU values of 0.13-0.86. Eight of the rare codons appear in the

Table III. Codon usage in the *glpF* and *glpK* genes.

Amino acids	Codons	*glpF*	*glpK*	Amino acids	Codons	*glpF*	*glpK*
Arg	*CGA	0	1	Gly	*GGA	1	0
	CGC	5	14		GGC	15	22
	*CGG	0	0		*GGG	5	5
	CGU	1	18		GGU	11	16
	*AGA	1	0	Val	GUA	2	5
	*AGG	0	0		GUC	5	5
Leu	*CUA	2	0		GUG	6	20
	CUC	1	5		GUU	13	6
	CUG	13	22	Pro	CCA	3	6
	CUU	3	1		*CCC	2	0
	UUA	3	2		CCG	2	7
	UUG	8	6		CCU	8	1
Ser	UCA	3	1	Trp	UGG	5	13
	UCC	1	8	Asn	AAC	2	15
	*UCG	1	1		AAU	5	5
	UCU	3	6	Gln	CAA	3	5
	AGC	1	4		CAG	4	15
	AGU	3	1	Ile	*AUA	0	1
His	CAC	1	5		AUC	7	17
	CAU	4	4		AUU	14	16

TABLE III. *con't*

Amino acids	Codons	*glpF*	*glpK*	Amino acids	Codons	*glpF*	*glpK*
Thr	ACA	3	0	Asp	GAC	5	15
	ACC	7	21		GAU	3	12
	ACG	1	7	Tyr	UAC	7	8
	ACU	4	9		UAU	0	10
Met	AUG	6	15	Cys	UGC	3	5
Ala	GCA	8	10		UGU	3	0
	GCC	7	13	Phe	UUC	12	7
	GCG	8	19		UUU	9	7
	GCU	11	3	Lys	AAA	6	18
Glu	GAA	7	28		AAG	2	4
	GAG	2	12				

*Rare codons

glpK gene with RSCU values of 0.09-0.47, indicating that codon usage for these genes is typical of genes that are moderately to lowly expressed.

The promoter region of the* glpFK *operon: Preceding the open reading frame of *glpF* is a potential promoter region between nucleotides 94 and 130. This area contains a -10 sequence (TATGTT) and a -35 sequence (TTCTCG) which has low similarity to the -35 consensus sequence (TTGACA). Primer extension analysis was used to map the 5' end of the *glpFK* transcript in order to determine if these promoter elements are those used *in vivo*. Total RNA was isolated from strain DH5αF'(pGC2) (Table I). A synthetic oligonucleotide primer complementary to nucleotides 175 to 194 (Fig. 5) was hybridized to the RNA and extended by the use of reverse transcriptase (Fig. 6); the location of the product of this reaction is indicated as +1 on Figure 5. When the RNA used in the primer extension reactions was isolated from cells grown on glycerol, the major product marking the startpoint of transcription was much darker than one obtained when the cells were grown on glucose (Fig. 6). This indicates that expression of the operon is sensitive to catabolite repression. A second band seen with this analysis is interpreted as the result of premature termination of transcription by reverse transcriptase, possibly caused by

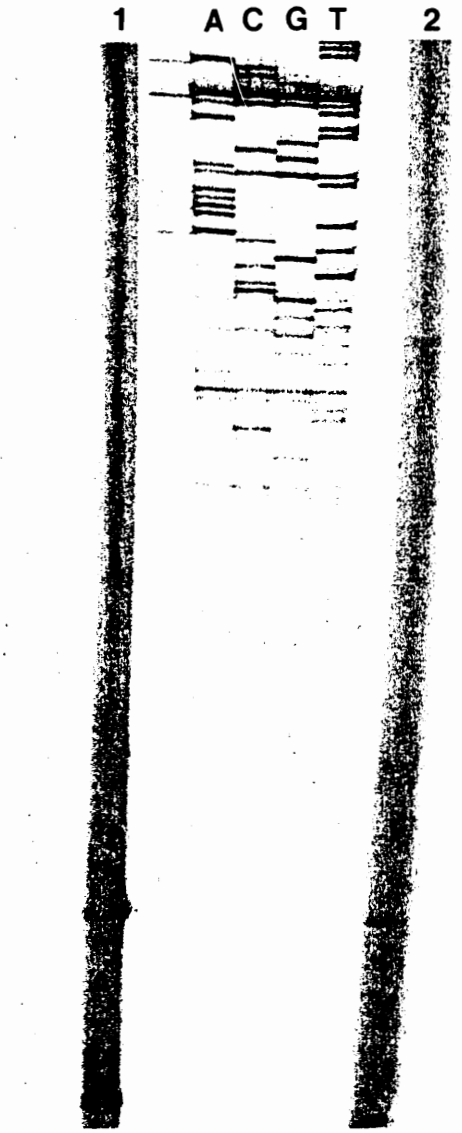


Fig. 6. Start site of transcription of the *glpFK* operon. Primer extension reactions were performed using a 20-nucleotide primer specific for the 5' end of the *glpF* gene with total RNA from strain DH5αF'(pGC2), grown on either glycerol (lane 1) or glucose (lane 2). The sequence ladder (lanes A,C,G, and T; generated by the use of the same primer) was co-electrophoresed as a position marker.

secondary structure of the mRNA. The position of the transcription start site indicates that the promoter elements proposed above most likely function *in vivo*.

In order to determine if *glpK* has its own promoter, primer extension reactions were performed using a synthetic primer complementalry to nucleotides 1081-1100 (Fig. 5) and RNA from DH5α-F'(pGC2) (Table I). There was no indication of a second transcription start site; no bands resulted from reactions with RNA from either glycerol- or glucose-grown cells.

Characterization of the* glpFK *control region: The genes of the glycerol-P regulon are subject to negative regulation by the *glpR* gene product (2). The *glp* repressor, a tetrameric protein of MW 130,000 (74), binds to a tandemly repeated palindromic sequence (TATGTTCGAT ATCGAACATT) in the regulatory region of the *glpD* gene (77). Several sequences similar to this sequence were identified in the region upstream of the start of transcription, between nucleotides 40 and 130 (Fig. 5). To determine which of these may serve as operator sites controlling expression of the *glpFK* operon, DNase I footprinting experiments were performed (Fig. 7). Four sequences within the protected region similar to the repressor binding sites found in the *glpD* gene were identified:

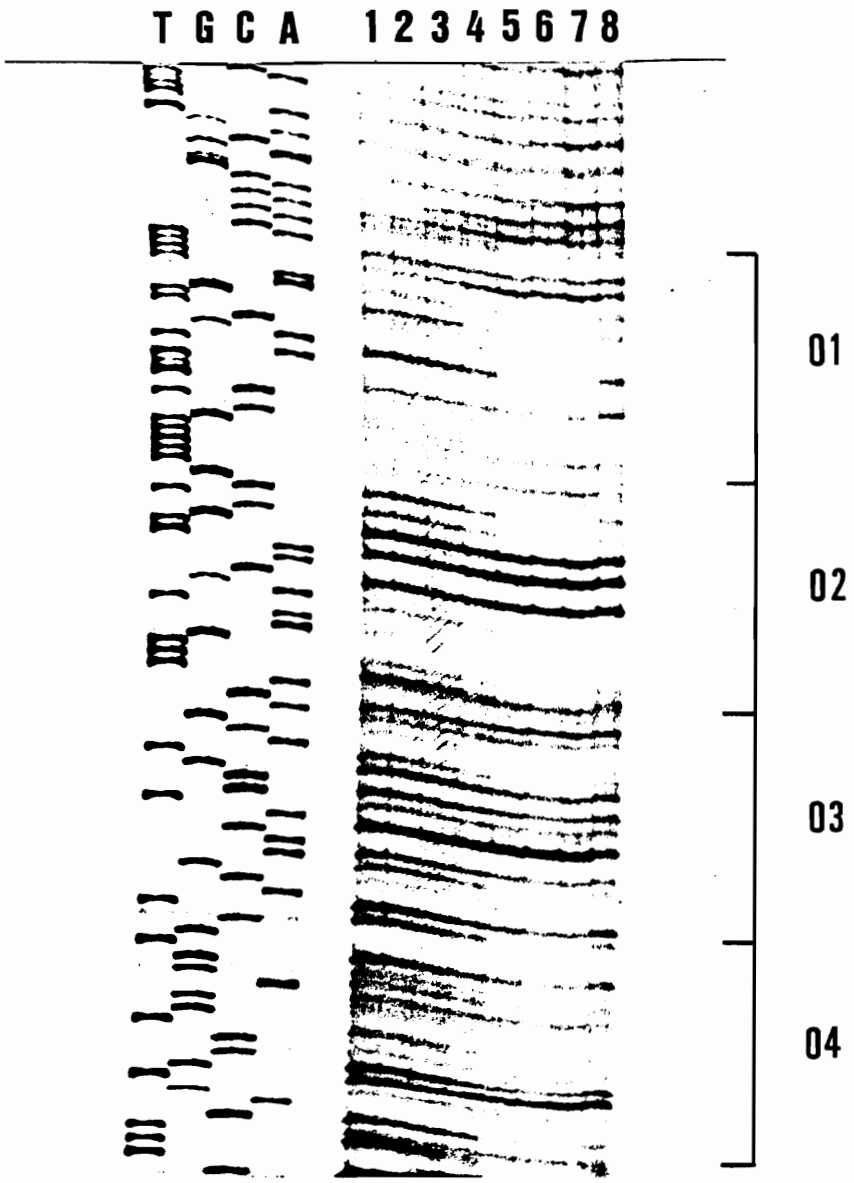


Fig. 7. **Identification of binding sites for the *glp* repressor in the *glpFK* control region.** DNase I footprinting reactions were carried out using the following concentrations of *glp* repressor tetramers: Lanes 1, 0 nM; 2, 0.1 nM; 3, 0.25 nM; 4, 0.5 nM; 5, 1.0 nM; 6, 2.5 nM; 7, 5 nM; 8, 5 nM repressor and 5 mM glycerol-P. The sequence of pDW51 (A, C, G, T; T3 primer) serves as a position ladder.

O1: ATGGCGCGATAACGCTCATT (nucleotides 41-60)

O2: TATGACGAGGCACACACATT (nucleotides 62-81)

O3: TAAGTTCGATATTTCTCGTT (nucleotides 83-102)

O4: TTTGCTCGTTAACGATAAGT (nucleotides 103-122)

The protected areas extended fron nucleotide -89 (relative to the startpoint of transcription) to the -10 region of the promoter for the operon. These areas are indicated by single-dashed lines on Figure 5 and are shown in Figure 7. The addition of glycerol-P to the footprinting reactions eliminated the protection by the repressor. The use of increasing concentrations of repressor (0.1 nM to 5 nM) revealed no preferential binding to any one of these sites. Very little protection was seen at 0.5 nM while protection of all four sites was complete at a concentration of 1.0 nM, suggesting cooperativity in the binding of the *glp* repressor to the tandemly repeated operator sites.

Binding of the repressor to the *Hin*dIII-*Eco*RV fragment of pDW30 (within the coding region of *glpK*) was demonstrated previously by gel retardation assays (74). To further characterize the binding of the repressor to this region, DNase I footprinting was performed. These studies revealed two areas which were protected by the *glp* repressor from digestion (nucleotides 1117-1141 and 1160-1184, Fig. 5; Fig. 8). The affinity of the repressor for these sites is much lower than that for the sites in the regulatory region, as

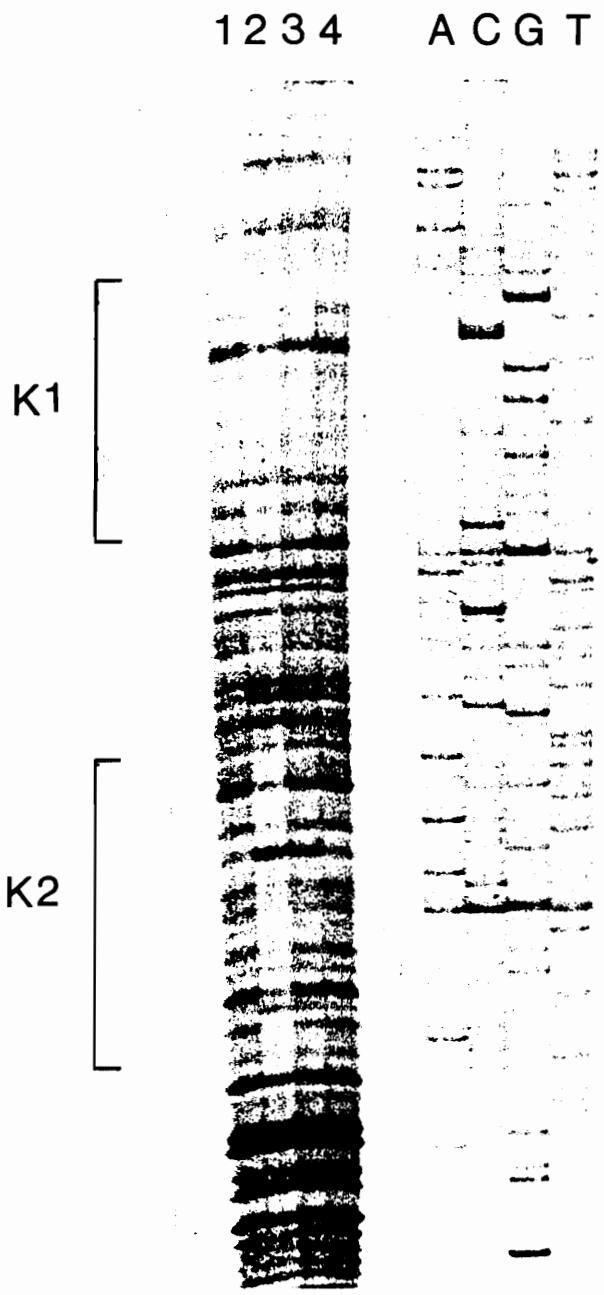


Fig. 8. Identification of binding sites for the *glp* repressor in the *glpK* coding region. DNase I footprinting reactions contained the following concentrations of repressor tetramers: Lanes 1, 0 nM; 2, 50 nM; 3, 10 nM; 4, 50 nM repressor and 5 mM glycerol-P. Sequencing reactions of pDW30 (SK primer) was used as the position marker.

50 nM repressor was needed for protection. Addition of glycerol-P eliminated this protection, indicating that these binding sites are specific for the *glp* repressor.

In addition to regulation by the *glp* repressor, transcription of the genes of the glycerol-P regulon is dependent on the cAMP-CRP complex (2). The regulatory region of pDW51 was examined for areas of similarity to the consensus sequence for the binding of the cAMP-CRP complex (81). Two possible binding sites were observed, both having 64% identity to the consensus sequence. DNase I footprinting performed in the presence and absence of CRP revealed an extensive region (Fig. 9), large enough for the binding of two CRP dimers, protected by 50 nM CRP (nucleotides 60 to 100, Fig. 5; Fig. 9). Protection was not apparent when cAMP was omitted from the reaction mixture. As expected, no evidence of binding of the cAMP-CRP complex was found in the promoter-proximal end of the *glpK* gene, either by gel retardation assays or DNase I footprinting experiments (data not shown).

The role of the* glpK *repressor binding sites: To determine if the two operator sequences located within the *glpK* coding region have any effect on the apparent affinity of the *glpFK* operon for the repressor, *in vivo* competition experiments were performed. The recombinant plasmids pDW25, 30, 51, 58,

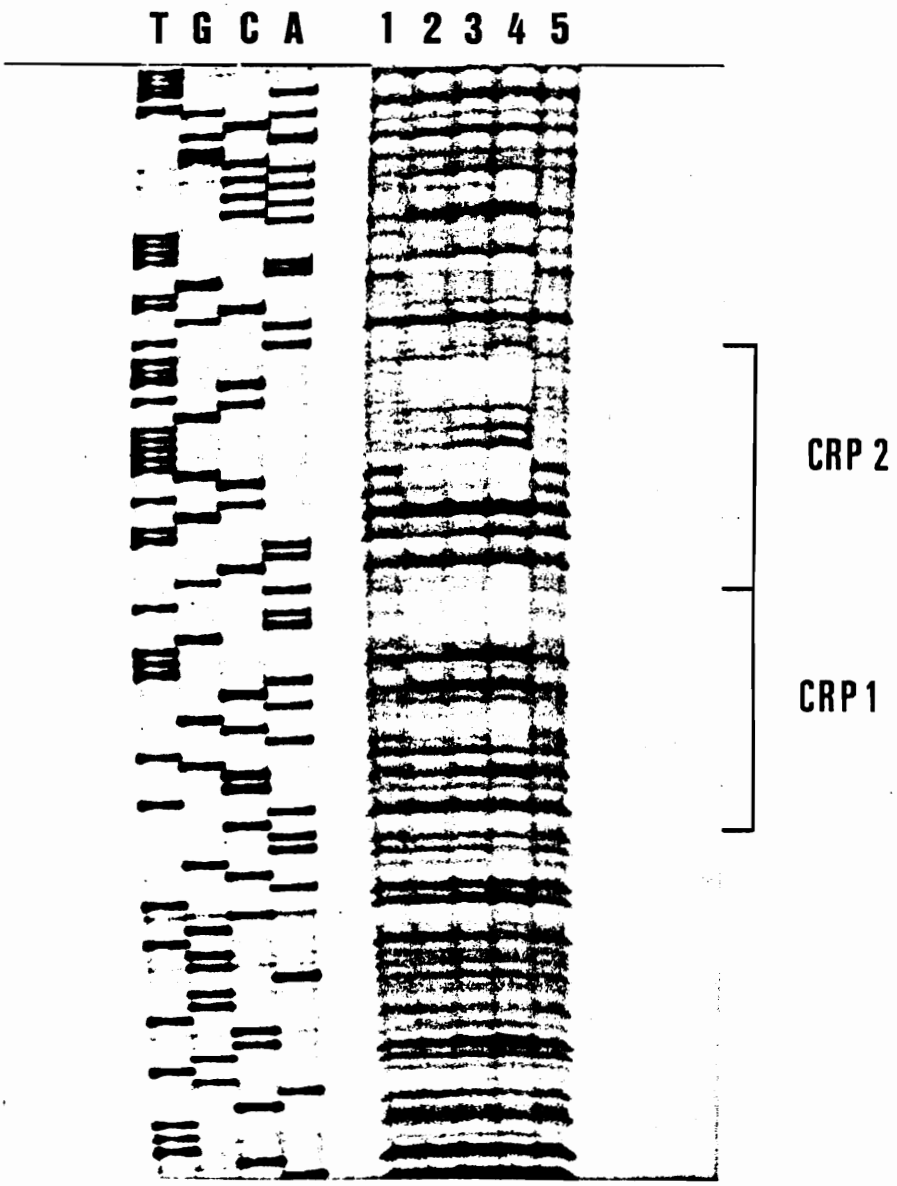


Fig. 9. Identification of binding sites for CRP in the *glpFK* control region. DNase I footprinting reactions were carried out using the following concentrations of CRP dimers: Lanes 1, 0 nM; 2, 50 nM; 3, 100 nM; 4, 150 nM; 5, 150 nM. Reactions run in lanes 1 through 4 also contained 0.25 mM cAMP; the reaction run in lane 5 contained no cAMP. Sequencing reaction of pDW51 (T3 primer) served as the position reference.

and 59 (Fig. 10; Table IV) and the vector pBS+ were used to transform TJS2 (*glpR*^{n} *glpT-lacZ*; Table I). The *glpR*^{n} allele confers a Lac^{-} phenotype on the host. Repressor binding sites present on multicopy plasmids transformed into this strain compete with repressor binding to the single-copy *glpT-lacZ* fusion, allowing *lacZ* to be expressed. Thus, ß-galactosidase activity is proportional to the affinity of the repressor for the operator(s) present on the multicopy plasmid. The ß-galactosidase activities of the transformed strain were determined (Table V). When TJS2 was transformed with pDW25 (which contains most of the *glpF* coding region but no operator sites), the activity level was the same as that obtained with the vector pBS+. The presence of the operator sites within the *glpK* coding region resulted in a very slight increase in the amount of ß-galactosidase activity detected (pDW30 and pDW58, Table V). Transformation with pDW51 (which contains the promoter/operator region of the *glpFK* operon) resulted in a large increase in the level of activity. By far the highest level of ß-galactosidase activity seen was when both the promoter/operator region and the intragenic operator sites were present on the same plasmid (pDW59). The presence of this construct resulted in levels of activity that were twice that found when only the promoter/operator sequences were involved. This synergistic effect indicates that the

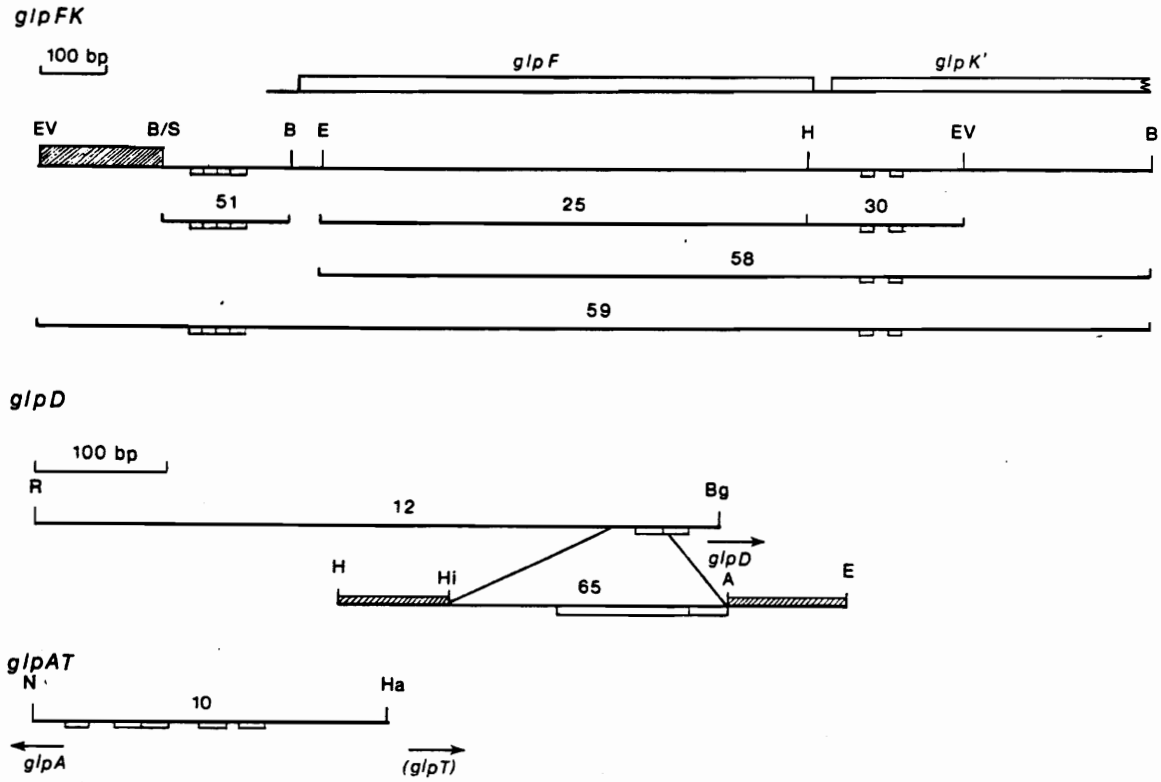


Fig. 10. Inserts used in the analysis of the affinity of the *glp* operators for the *glp* repressor. The open boxes at the top of the figure indicate the coding regions of the *glpF* and *glpK* genes. Small open boxes within the figure indicate operator sites. Restriction sites are abbreviated as in Fig. 4. In addition: N, *Nru*I; Bg, *Bgl*II; Ha, *Hae*III; R, *Rsa*I; Hi, *Hin*PI; A, *Alu*I. Numbers indicate recombinant plasmid designations. The insert of pDW65 contains a 42 bp *Alu*I-*Hin*PI fragment of pSY12 which was first cloned into M13mp18, creating pS2. A 78 bp *Hin*dIII-*Eco*RI fragment of pS2 was then cloned into pBS+, creating pDW65.

Table IV. Plasmids used in the examination of the affinity of the *glp* operators for the repressor.

Plasmid	Insert	Source of insert	Vector sites used
Bluescript			
pDW25	*Eco*RI-*Hin*dIII	pDW23	*Eco*RI-*Hin*dIII
pDW30	*Hin*dIII-*Eco*RV	pDW28	*Eco*RV-*Hin*dIII
pDW51	*Sau*3A-*Bam*HI	pGC2	*Bam*HI
pDW58	*Eco*RI-*Bam*HI	pDW23	*Eco*RI-*Bam*HI
pDW59	*Eco*RI-*Xho*I	pDW50	*Eco*RI-*Xho*I
	*Eco*RI-*Xho*I	pDW58	
pSY12	*Rsa*I-*Bgl*II	pSH79	*Bam*HI-*Eco*RV
pDW65	*Hin*dIII-*Eco*	pS2	*Hin*dIII-*Eco*RI
pAB10	*Hae*III-*Nru*I	pHH11	*Sma*I
pCB267			
pDW53	*Sau*3A-*Xba*I	pDW51	*Bam*HI-*Xba*I
pDW54	*Xba*I-*Sal*I	pAB10	*Xba*I-*Sal*I
pDA901	*Sau*3A-*Bgl*I	pSY12	*Bam*HI

Table V. Affinity of the *glp* operators for the *glp* repressor as reflected by ß-galactosidase activities.

Plasmids and Operator regions	Units ß-galactosidase activity[a]
Bluescript M13+[b]	5 ± 0.5
pDW25: No operators	5 ± 0.4
pDW30: K_1, K_2	8 ± 0.6
pDW58: K_1, K_2	8 ± 0.7
pDW51: F_1, F_2, F_3, F_4	74 ± 2.9
pDW59: F_1, F_2, F_3, F_4, K_1, K_2	154 ± 4.2
pSY12: D_1, D_2	63 ± 2.7
pDW65: D_1	28 ± 1.2
pAB10: A_1, A_2, A_3, A_4, T	78 ± 3.1

a: Values represent the average of a minimum 12 assays for the *glpFK* operators and a minimum of 6 assays for all others.

b: Control

intragenic operator sequences of *glpK* are likely to play a physiological role in the expression of the *glpFK* operon.

Differential regulation of the* glp *operons by the* glp *repressor: Early *in vivo* experiments indicated that the *glp* operons exhibit differential response to the *glp* repressor. Because those conclusions were based on the activities of glycerol kinase, glycerol permease, and the aerobic dehydrogenase rather than the level of transcription of the genes, experiments were designed to measure the response of the *glp* operators to the *glp* repressor. The first methodology used to examine this response was an extension of the repressor titration system used to examine the role of the *glpK* operators. In addition to the plasmids containing the promoter/operator sequences of the *glpFK* operon, plasmids containing the *glpD* (pSY12) and the *glpACB-glpTQ* (pAB10) promoter/operator regions were used (Fig. 10; Table IV). The primary operator of the *glpD* gene (O_1) plus 6 bp of the second operator was also isolated and ligated into the vector pBS+, creating pDW65. Strain TJS2 was transformed with each of these plasmids and ß-galactosidase activity was measured. As is seen in Table V, the activity levels found with the promoter/operator regions of the *glpACB-glpTQ* and *glpFK* operons were very similar, while that obtained with the *glpD* control region was slightly lower.

The enzyme activity level resulting from transformation with the plasmid containing the single *glpD* operator was less than half of that found when both operators were present. Again, the activity level determined for the plasmid with all six operator sites of the *glpFK* operon was much higher than that found with any of the other constructs. This system was very stable. Numerous assays (a minimum of 12 for the *glpFK* operators, 6 for all others) performed over 7 months resulted in consistent values for all strains.

The differential response of the *glp* operons to the *glp* repressor was also examined using transcriptional fusions of the *glp* promoters to the *lacZ* gene in the promoter-probe vector pCB267 (102) (Fig. 11; Table VI). The recombinant plasmids were cotransformed into CB806 (Table I) with either a plasmid carrying the *glpR* gene (pSH58) or the vector without *glpR*, pACYC184 (103). A repression ratio was determined by the comparison of the ß-galactosidase activities obtained under induced and repressed conditions (Table VI). Induction was achieved by growing the strains with pACYC184 and the recombinant plasmids in minimal media A (104) with casamino acids and glycerol. Repression of the system occured when the strains harboring pSH58 and the recombinant plasmids were grown on minimal media A with only casamino acids. Because the presence of a chromosomal copy of *glpR* on CB806 could result in some repression even under

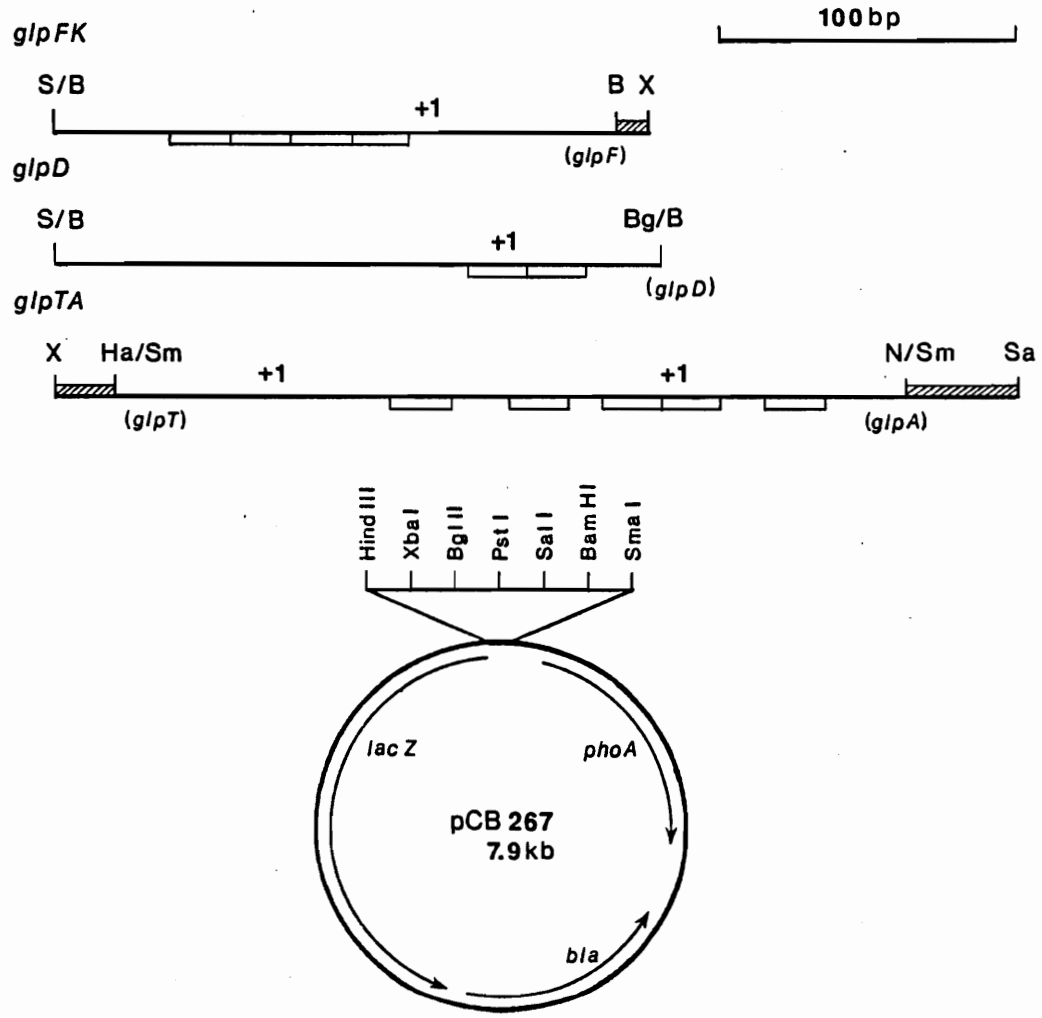


Fig. 11. Construction of plasmids used to examine the differential response of the *glp* operators for the *glp* repressor. Abbreviations used: S, *Sau*3A; B, *Bam*HI; X, *Xba*I; Bg, *Bgl*II; Ha, *Hae*III; N, *Nru*I; Sa, *Sal*I; Sm, *Sma*I. The first of each pair of restriction sites separated by "/" indicates the insert restriction site, while the second indicates the restriction site utilized on the vector to construct the recombinant plasmid.

Table VI. Affinity of the *glp* operons for the *glp* repressor under induced vs repressed conditions.

Host	Operon	Repression ratio[a]
CB806		
	glpD	20
	glpFK	36
	glpTQ	20
TL73		
	glpD	39
	glpFK	68
	glpTQ	36[b]

a: The repression ratio is defined as:

$$\frac{\text{ß-galactosidase activity with pACYC184}}{\text{ß-galactosidase activity with pSH58}}$$

b: T.J. Larson, unpublished results.

inducing conditions, repression ratios were also determined in strain TL73 (*glpR*) (Table I; Table VI). The higher repression ratios obtained with TL73 indicates that the chromosomal *glpR* did indeed exert repression on the CB806 strains. This system proved to be problematic in that the strains did not appear to be stable. Growth rates and enzyme activity levels varied widely, although in each case the represson ratios were consistent. In each case, the *glpFK* operon was subject to a greater level of repression than the *glpD* or *glpTQ* operons.

CHARACTERIZATION OF GLPF AND GLPK

Analysis of the deduced amino acid sequence of GlpF: Sequencing of the DNA upstream from *glpK* revealed an open reading frame from nucleotide 201 to nucleotide 1043 which was preceded by a good ribosome binding site. Other possible start codons in the same reading frame lacked potential ribosome binding sites, and several termination codons exist in the other reading frames. The amino acid sequence of GlpF deduced from the uninterrupted reading frame was comprised of 281 amino acids with a calculated molecular weight of 29,780 (Fig. 5). Because the second amino acid is serine, the N-formylmethionine residue is likely to be cleaved post-translationally (116). The *glpF* gene product exhibits an apparent molecular weight of 25,000 on SDS-polyacrylamide gel electrophoresis (7). This discrepancy can be accounted for by the aberrant mobility of hydrophobic proteins on sodium dodecyl sulfate-polyacrylamide gels. The GlpF protein has extremely low polarity (72% nonpolar, 28% polar residues; Table VII), as revealed by analysis of the deduced amino acid composition.

Hydrophobicity analysis of GlpF: Analysis of hydrophobicity distribution can be used to predict secondary structure and transmembrane helices. The hydropathy scale composed by

Table VII. Amino acid compositon of the glycerol facilitator protein.

Polar		(%)	Nonpolar		(%)
Arg	7	(2.5)	Phe	21	(7.5)
Lys	8	(2.8)	Leu	30	(10.7)
Asp	8	(2.8)	Ile	21	(7.5)
Asn	7	(2.5)	Val	26	(9.3)
Glu	9	(3.2)	Trp	5	(1.8)
Gln	7	(2.5)	Ala	34	(12.1)
His	5	(1.8)	Tyr	7	(2.5)
Thr	15	(5.3)	Gly	32	(11.4)
Ser	12	(4.3)	Met	6	(2.1)
			Cys	6	(2.1)
			Pro	15	(5.3)
Subtotal	78	(27.7)		203	(72.3)
Total	281	(100.0)			

Kyte and Doolittle is based on the hydrophobic/hydrophilic properties of the side chains of 20 amino acids (117). Using a window of 7 amino acid residues,the hydropathy profile of GlpF was computed (Fig. 12). The average hydropathy value for the glycerol facilitator protein was 0.73, which can be compared to a hydropathy value of -0.4 for soluble proteins (117). Hydropathy averages greater than +1.6 over at least a 19-residue span indicate that the sequence is likely to span the membrane (117). Using a window of 21 residues, six potential membrane-spanning regions were identified in the GlpF protein with average hydropathy values of +2.1, 1.7, 1.6, 2.2, 2.0, and 1.9 (labeled in Figs. 12 and 13). These potential transmembrane alpha helices were also predicted by hydrophobic moment analysis (118) and by the averaging of secondary structure propensities, charge profile, and hydrophobicity (119). A model of the GlpF protein based on these analyses has been derived (Fig. 13). Assuming that the amino terminus of GlpF lies in the cytoplasm and that the protein contains six transmembrane regions, the model indicates that the carboxyl terminus of the protein (31 amino acid residues) also lies in the cytoplasm.

Homology between GlpF and other membrane proteins: A search for homology between the amino acid sequence of the glycerol facilitator and other proteins revealed similarities between

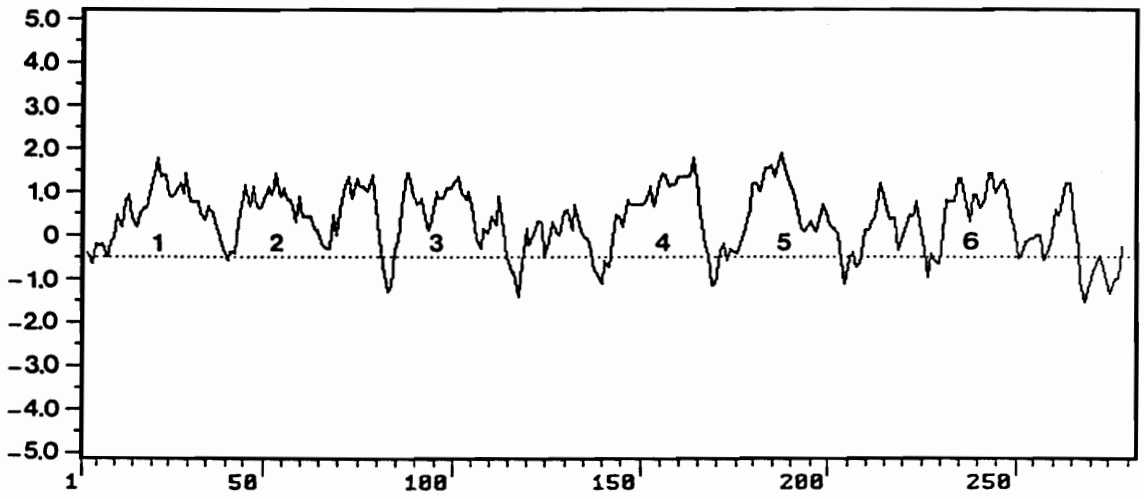


Fig. 12. Hydropathy profile of the predicted GlpF protein. Amino acid residues are numbered on the *X* axis. Six 21-residue, putative membrane-spanning segments with hydropathy values of greater than +1.6 are numbered within the figure.

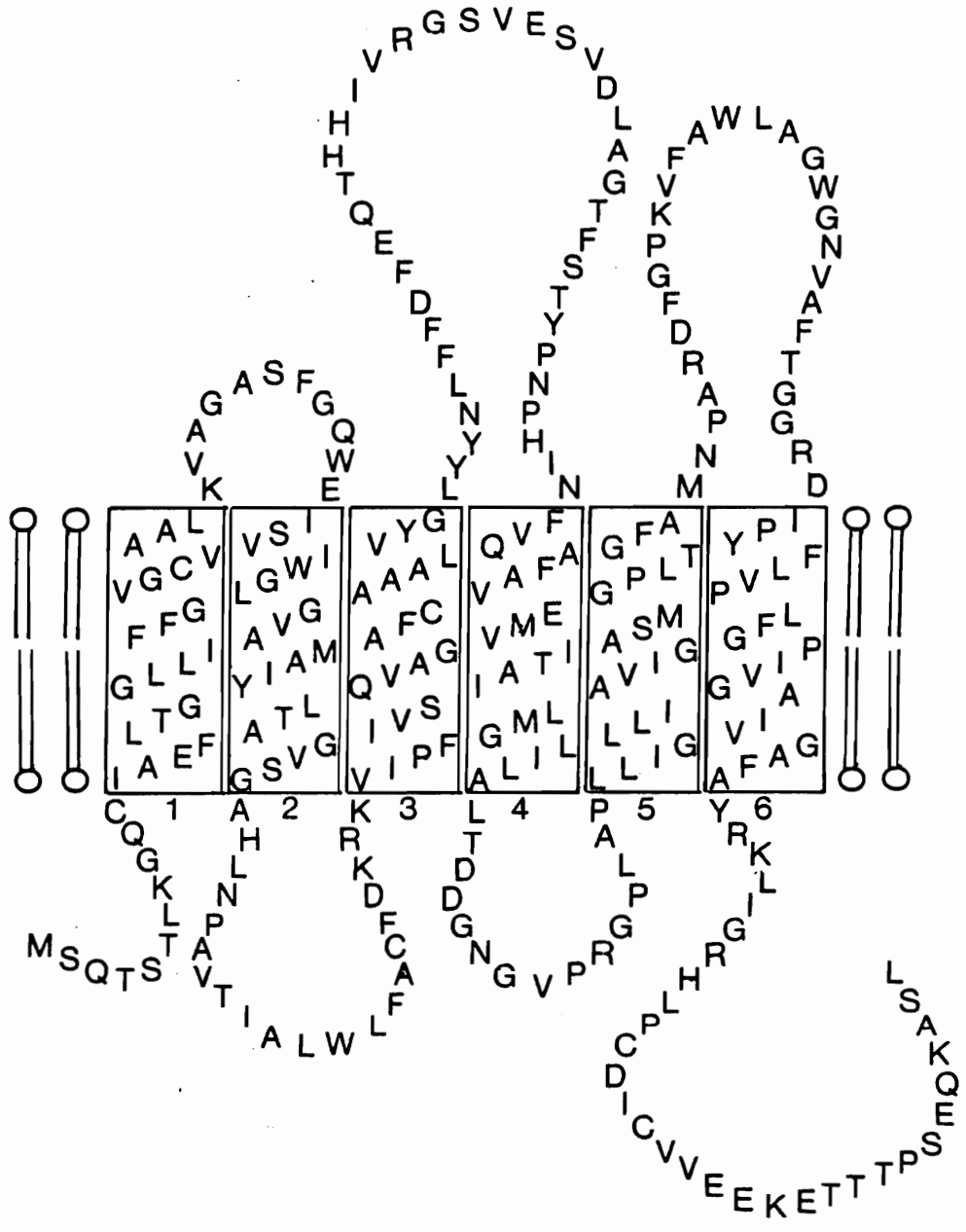


Fig. 13. **A model for the arrangement of GlpF in the membrane based on the deduced amino acid sequence.** Numbered rectangles indicate putative membrane-spanning α-helices.

GlpF and the major intrinsic protein (MIP) of the bovine lens fiber membrane (Fig. 14). MIP is thought to function as a gap junction protein and is proposed to have six membrane spanning α-helices (11). Using an overlap of 107 amino acids at the amino terminal end of each protein, a 30.8% identity is seen between GlpF and MIP, both in hydrophobic regions proposed to be transmembrane spanning helices and in hydrophilic regions (11). Another area of identity is seen between the fifth and sixth putative transmembrane helices of the two proteins. In each case, this region is proposed to extend extracellularly and could provide interaction between these membrane proteins and molecules on the outside of the cell.

Sequence similarities were also found between GlpF and Nod-26, a protein located in the peribacteroid membrane of soybean root nodules. Analysis of Nod-26 suggested that it is a transmembrane protein with four regions hydrophobic enough to span the membrane (12). However, only one of these regions can be folded into stable α-helix; the other regions are proposed to be ß-sheets which traverse the membrane. GlpF and Nod-26 exhibited a 33.8% identity in a 65 amino acid overlap at the amino terminal end of the proteins (Fig. 14). Again, a second area of considerable identity was found near the carboxyl end of the proteins. The two areas of identity align well between these three membrane proteins.

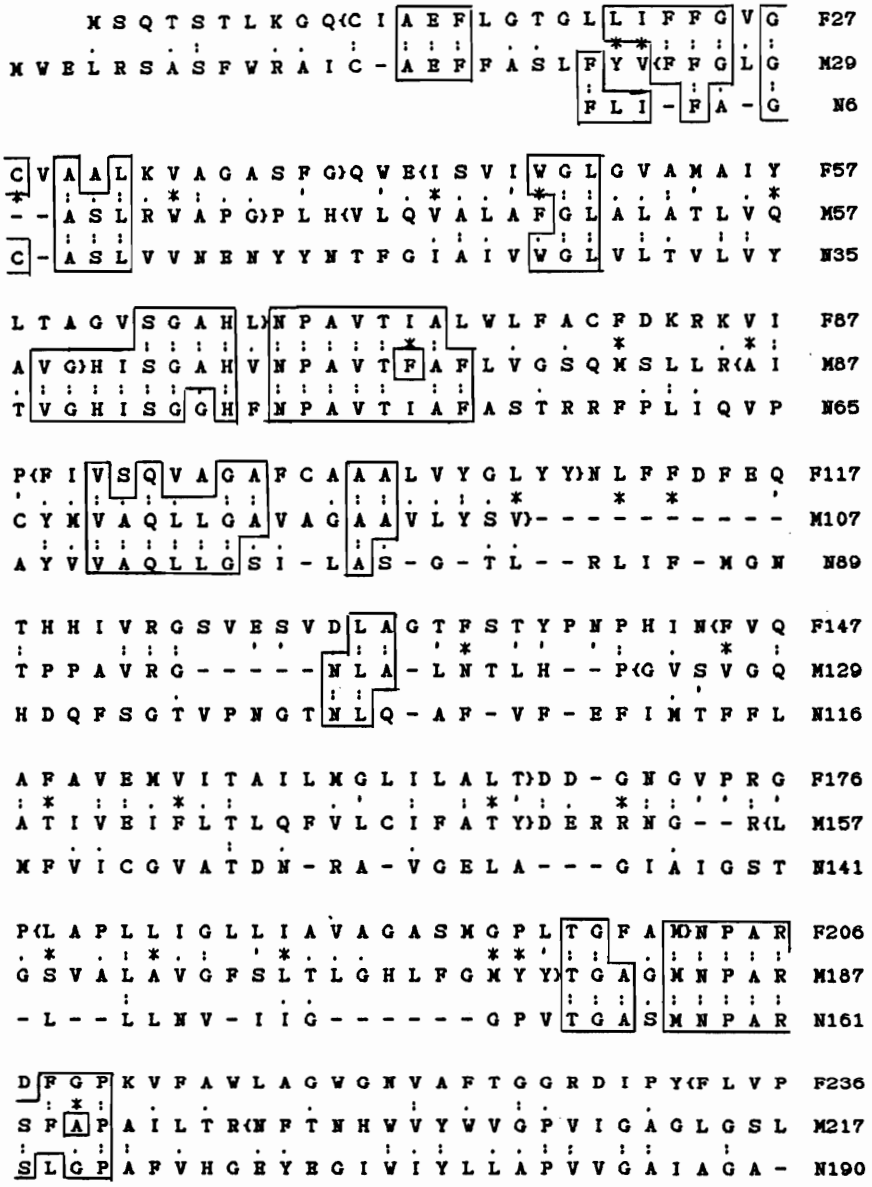


Fig. 14. **Alignment of the GlpF, MIP, and Nod-26 amino acid sequences.** Colons show identities and periods show conservative replacements between adjacent sequences. * denotes identities and ' denotes conservative replacements between GlpF and Nod-26. Boxed areas indicate areas which are highly conserved between the three proteins. Amino acids are grouped as follows (GASTP) (ILVM) (QNED) (HKR) (C).

Overproduction of glycerol kinase: In order to determine the amino acid sequence of the N-terminus of glycerol kinase, and thereby align the sequence with that predicted by the nucleotide sequence, purified kinase was needed. Overproduction of glycerol kinase was accomplished by the construction of pDW23 (Fig. 3), in which the *glpK* coding region is positioned behind the strong pL promoter of phage lambda. Thermal induction of the cryptic lambda lysogen N4830 harboring pDW23 yielded extracts in which glycerol kinase constituted approximately 25% of the total protein, as identified on SDS-PAGE (Fig. 15, Lane 1). Cells harvested from this thermally induced culture served as starting material for the purification of the kinase.

Purification of glycerol kinase: Glycerol kinase was purified to near homogeneity in a three-step purification process modified from the procedure of Hayashi and Lin (120) which utilized streptomycin sulfate and ammonium sulfate precipitation and DEAE Sephadex chromatography. A summary of the purification protocol is shown in Table VIII. The final product required a 4-fold purification with a 47% yield, resulting in 50 mg of pure protein (starting with 3.5 grams of cells from a 600 ml culture). Aliquots from each step of the purification were subjected to SDS-PAGE (Fig. 15). The DEAE Sephadex chromatography step removed the bulk of the

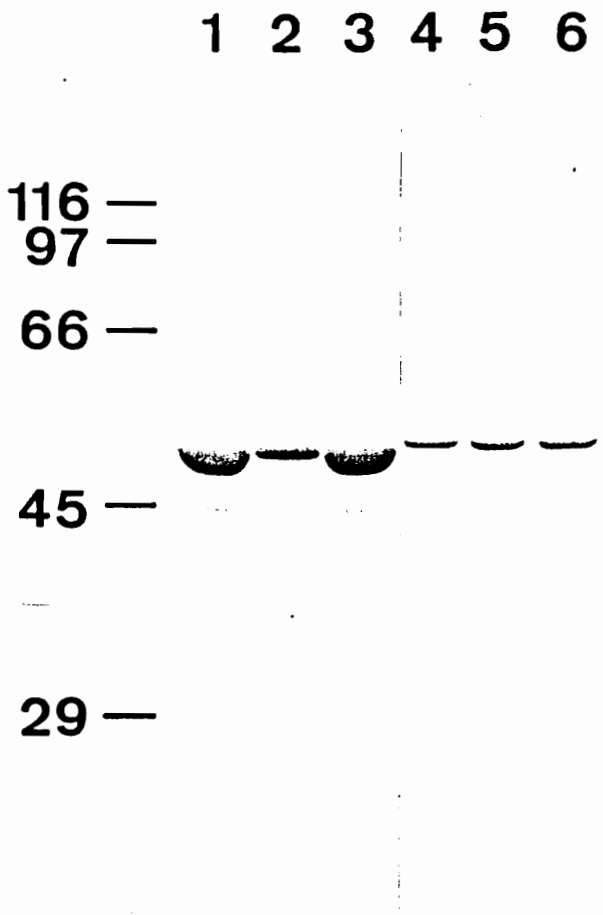


Fig. 15. SDS-polyacrylamide gel electrophoretic analysis of the purification of glycerol kinase. *Lanes: 1,* crude extract (60 µg protein); *2,* streptomycin sulfate fraction (40 µg); *3,* ammonium sulfate fraction (60 µg); *4* (5 µg), *5* (8 µg), *6* (8 µg), and *7* (5 µg), peak fractions of DEAE Sephadex chromatography. The molecular weights ($x10^3$) of protein standards (ß-galactosidase, phosphorylase *b,* bovine serum albumin, ovalbumin, and carbonic anhydrase) are indicated on the left.

Table VIII. Purification of glycerol kinase from *E. coli*.

Purification step	Volume ml	Protein mg/ml	Specific activity units/mg	Yield %
Crude extract	70	6.7	4.9	(100)
Streptomycin sulfate	74	4.5	3.7	52
Ammonium sulfate	120	1.4	13.2	97
DEAE Sephadex	10	5.0	21.8	47

contaminants; no Coomassie Blue-stainable contaminants were seen in the DEAE Sephadex peak fractions. The subunit molecular weight was 55,000 on SDS-PAGE and this value is in good agreement with the molecular weight of 56,106 predicted by the nucleotide sequence of *glpK* (20). The specific activity of the purified glycerol kinase was 21.8 µmol/min per mg of protein. Glycerol kinase purified by Thorner and Paulus displayed a specific activity of 21 µmol/min per mg of protein (19), indicating that overproduction did not affect the catalytic properties of the enzyme.

N-terminal sequence of glycerol kinase: Sequence analysis by Edman degradation of approximately 625 pmol of the N-terminal portion of glycerol kinase revealed the sequence:

T E K K Y I V A L D (Table IX)

This concurred with the amino acid sequence deduced from the DNA sequence of the *glpK* gene (Fig. 5) and revealed post-translational cleavage of the initial methionine residue. This finding indicates that the processing enzyme is present in enough quantity to process the formyl-methionine even in the case of massive overproduction. Similar loss of the terminal methionine was observed with overproduced D-lactate dehydrogenase (121).

Molecular weight of glycerol kinase: In the process of

Table IX. Analysis of the N-terminal sequence of glycerol kinase.

Cycle	Amino acid residue		Yield (pmol)
1	T	Threonine	362
2	E	Glutamate	342
3	K	Lysine	281
4	K	Lysine	239
5	Y	Tyrosine	280
6	I	Isoleucine	272
7	V	Valine	300
8	A	Alanine	277
9	L	Leucine	180
10	D	Aspartate	175

preparing glycerol kinase for N-terminal amino acid analysis, the kinase was subjected to chromatography on Sephadex G-200 to remove low molecular weight contaminants. The purified enzyme exhibited a molecular weight of 110,000 (Fig. 16). This value contrasts with the value (210,000) determined by Thorner and Paulus (19), who found the enzyme was tetrameric with a subunit molecular weight of 56,000. Dissociation of glycerol kinase into a dimeric form is favored when the enzyme concentration is low. Tetramer formation is promoted in the presence of FDP, an allosteric inhibitor of the kinase (21). Addition of 5 mM FDP to the gel filtration system resulted in an increase in apparent molecular weight to 180,000 (Fig. 16, Table X). When 10 mM glycerol was added to the buffer system, the glycerol kinase exhibited an apparent molecular weight of 195,000. While the addition of 20 mM phosphate had no significant effect of the apparent molecular weight of glycerol kinase, the addition of 20 mM phosphate with 10 mM glycerol to the system resulted in an apparent molecular weight of 220,000 (Fig. 16, Table X). Results indicated that the presence of glycerol alone caused an increase in apparent molecular weight of the kinase in this system.

The native size of glycerol kinase was also examined using nondenaturing gradient polyacrylamide gels. Glycerol, FDP, and combinations of glycerol plus FDP or glycerol plus phosphate were added to the electrophoresis gel and buffer.

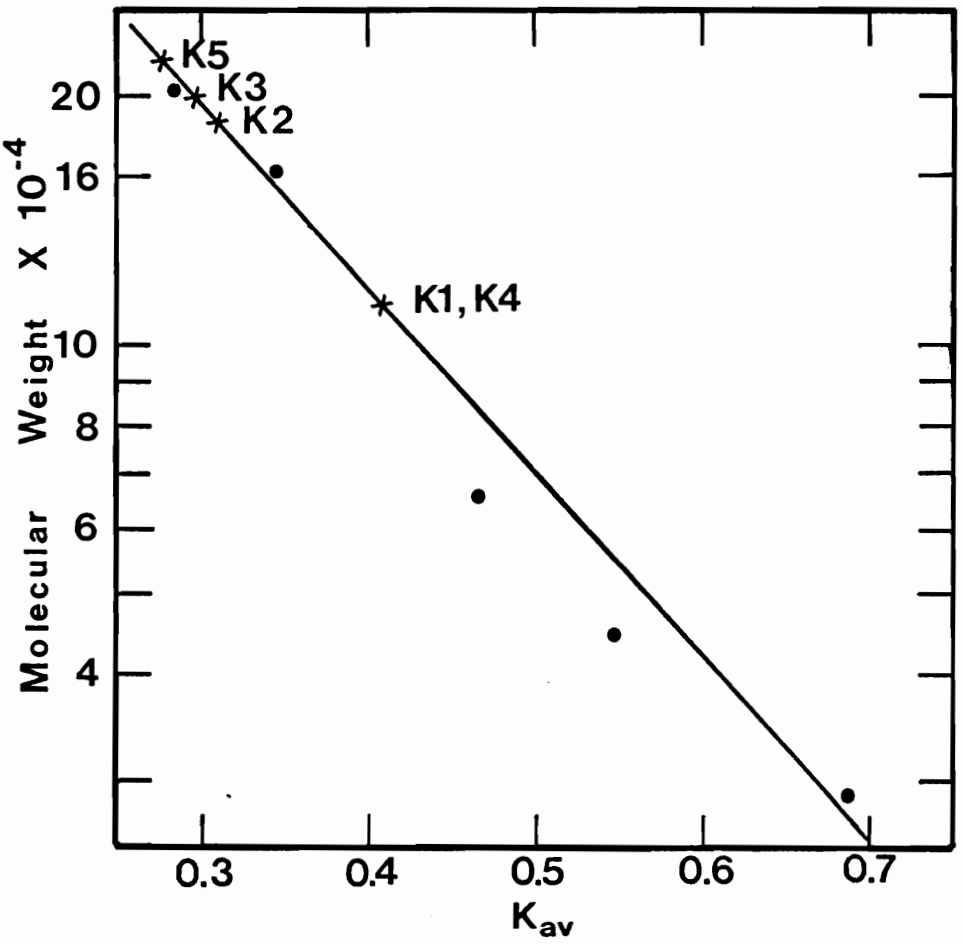


Fig. 16. **Chromatography on Sephadex G-200 of glycerol kinase.** The elution positions of glycerol kinase are indicated by K1-K5. Additions to the elution buffer were : K1, none; K2, 5 mM FDP; K3, 10 mM glycerol; K4, 20 mM phosphate; K5, 20 mM phosphate, 10 mM glycerol. The molecular weights ($x10^3$) of the standard proteins were: β-amylase (202); alcohol dehydrogenase (164); bovine serum albumin (66); ovalbumin (45); carbonic anhydrase (29). The elution buffer was 20 mM Tri-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT. Data are summarized in Table X.

Table X. Apparent molecular weight of glycerol kinase determined by chromatography on Sephadex G-200.

Sample	Additions to elution buffer	Molecular weight
K1	None	110,000
K2	5 mM FDP	180,000
K3	10 mM glycerol	195,000
K4	20 mM phosphate	112,500
K5	20 mM phosphate, 10 mM glycerol	220,000

The kinase exhibited an apparent molecular weight of 98,000 with no additions to the system (Fig. 17, Table XI). Addition of only 0.5 mM glycerol resulted in an apparent increase to 126,000. Addition of increased concentrations of glycerol or glycerol in combination with FDP or phosphate also caused aggregation or a conformational change such that the apparent molecular weight increased to 140,000. These data indicate that the presence of even low concentrations of glycerol have a profound effect on the native structure or conformation of glycerol kinase. The influence of glycerol may be due to interaction at the active site, or to another site, or both, of glycerol kinase. Binding of the substrate to the active site of glycerol kinase is possible in the absence of ATP because the reaction mechanism is ordered, with glycerol binding to the enzyme first (122).

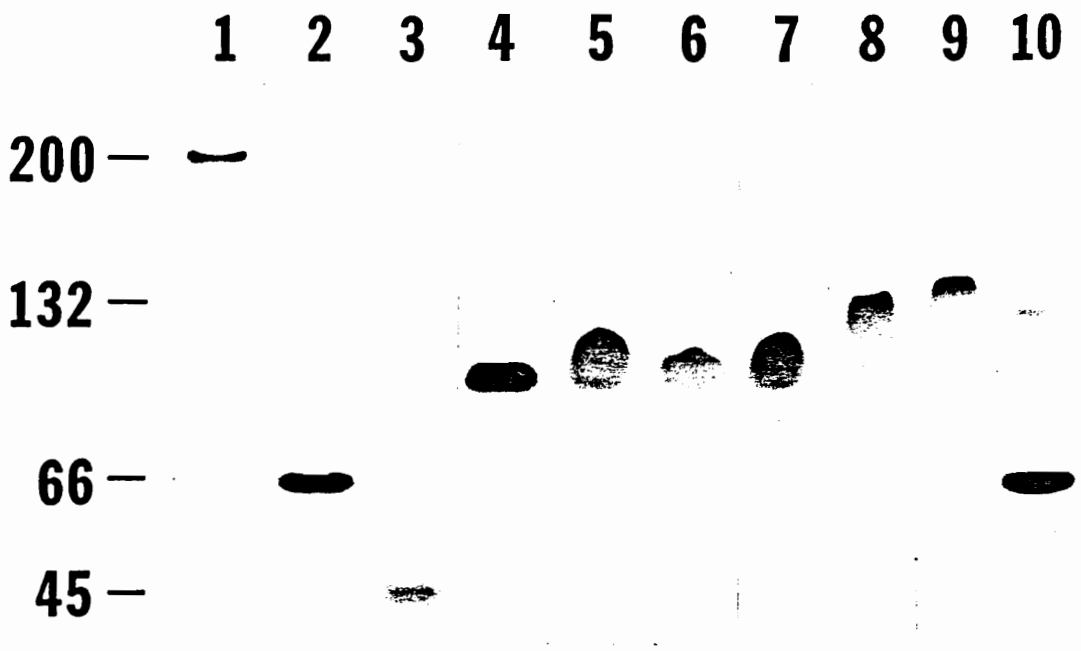


Fig. 17. **Electrophoretic mobility of glycerol kinase on gradient polyacrylamide gels.** Additions to gels and buffers are indicated by lanes. *Lanes: 4*, no glycerol; *5*, 0.5 mM glycerol; *6*, 0.5 mM glycerol*; *7*, 1 mM glycerol; *8*, 5 mM glycerol; *9*, 100 mM glycerol. *Lanes 1*, *2*, *3*, and *10* contain the molecular weight standards β-amylase, bovine serum albumin, ovalbumin, and bovine serum albumin, respectively. Data are sumarized in Table XI.

* *Lane 6* contains glycerol kinase which had not been exposed to glycerol during purification or storage. All other kinase samples had been stored at -70°C in buffer containing 10% glycerol.

Table XI. Apparent molecular weight of glycerol kinase determined by electrophoresis on gradient polyacrylamide gels.

Additions*	Molecular weight
None	98,000
0.5 mM glycerol	126,000
1 mM glycerol	126,000
5 mM glycerol	132,000
100 mM glycerol	138,000
5 mM FDP	124,000
5 mM FDP, 5 mM glycerol	138,000
10 mM phosphate, 10 mM glycerol	140,000

*Additions to gel and electrophoresis buffers.

DISCUSSION

ORGANIZATION AND REGULATION OF THE *GLPFK* OPERON

Previous studies indicate that the *glpF* and *glpK* genes are organized in an operon with *glpF* promoter proximal (7). Results reported above, including the sequence of the *glpF* gene, the identification of operator and promoter sequences upstream from the *glpF* coding region, the absence of a typical transcription termination signal at the 3'-flanking region of the *glpF* gene, and the absence of promoter elements at the 5'-flanking region of the *glpK* gene confirm this conclusion. The organization of the operon is similar to other prokaryotic operons, consisting of an operator/promoter region, an untranslated leader sequence, the coding region for the *glpF* gene followed by a 25 bp untranslated region, and the coding region for the *glpK* gene. During the time the *glpK* gene was being sequenced, the same sequence was published by Pettigrew, *et al* (20). The sequence of the *glpF* gene and the control region for the operon had been completed when it was published by Muramatsu and Mizuno as part of a study of segments of bent DNA (9); a discrepancy found between the published sequence and that determined in this study is indicated by "x" on Figure 5.

The region that Muramatsu and Mizuno (9) identified as a bent DNA segment (BENT-6) by two-dimensonal polyacrylamide gel electrophoresis is positioned upstream from the coding sequence for *glpF* and contains the regulatory elements for the operon (8,9). This region contains several tracts of T_n (n>3) with approximately 10 bp phasing, a feature which causes bending of DNA (123, 124). Bending of DNA is also caused by the binding of the cAMP-CRP complex in the *lac* and *ara* promoters (125-127). Both forms of bending may allow RNA polymerase to interact with the CRP protein and the -35 and -10 sites on the DNA, resulting in an increased rate of transcription of the *glpFK* operon, and partially offsetting the effects of the low degree of similarity of the -35 sequence of the *glpFK* operon to the -35 consensus sequence.

The role of the cAMP-CRP complex in transcription activation:

The cAMP-CRP complex stimulates transcription from other prokaryotic promoters of genes required for sugar metabolism. The mechanism of action of the complex is not fully defined. It is proposed that RNA polymerase interacts with CRP, resulting in stronger binding of the polymerase to the DNA. A second proposed mechanism involves the alteration of the structure of DNA by the binding of CRP, allowing the formation of an open complex between RNA polymerase and DNA. Wu and Crothers (126) argue against each of these models. Variations

exist in the spacing of the binding sites for RNA polymerase and CRP, arguing against the first model. In opposition to the second model, binding of CRP only unwinds one-half of a helical turn in the *lac* promoter, not enough to promote the formation of the open complex (126). Instead, Wu and Crothers propose that the CRP-induced bending may alter the conformation of the DNA in a way that allows the polymerase to interact with the DNA.

Variations in the position of the cAMP-CRP binding site are seen in the *gal*, *lac*, and *glp* systems. Two tandem binding sites for the complex are present in the *gal* operon from -50 to -30 and from -72 to -51 (80, 128). Tandem binding sites for the complex are also found in the *glpFK* operon, located from -46 to -25 and from -71 to -50 relative to the startpoint of transcription. The first binding site in both the *gal* and *glpFK* operons covers the -35 site on the DNA. A CRP binding site in the *glpTQ* promoter in the -35 region (-52 to -31) was identified by sequence analysis by Eiglmeier and collaborators (30) and has been verified by footprinting experiments (A. Bhattacharya and T.J. Larson, unpublished results). The cAMP-CRP complex also binds to the regulatory region of *glpD* (-74 to -53) (76). This position is similar to that in the *lac* operon (-71 to -50) (129).

The two binding sites in the *glpFK* operon are both identical with the consensus sequence at 9 of 14 bases, the

glpT site matches the consensus sequence at 11 of the 14 bases, and the binding site of *glpD* is identical with the consensus binding sequence for CRP in 10 of 14 bases:

Consensus:	AA-TGTGA------TCACA-TT	Position
glpFK:	AAGTTCGATATTTCTCGTTTTT	-46 to -25
glpFK:	TTTTATGACGCGGCACACACAT	-71 to -50
glpT:	ATGTGTGCGGCAATTCACATTT	-52 to -31
glpD:	TAATGTTATACATATCACTCTA	-74 to -53

Nucleotides at positions 7 and 16 of the consensus sequence (underlined above) make direct contact with residue Glu^{181} of CRP (81). Position 16 of the CRP binding site is invariant in *glpFK*, *glpTQ*, and *glpD* and matches the consensus sequence. The G:C base pair found at position 7 of both *glpFK* binding sites and the *glpTQ* site also matches the consensus sequence. However, a T:A substitution at position 7 is found in the *glpD* promoter. CRP is able to discriminate between this substitution and the consensus sequence by a factor of 29 (81). Earlier studies indicated that *glpK* is most responsive to catabolite repression, *glpT* is less responsive, and *glpD* is least responsive (74). This conclusion is supported by the identification of the tandem CRP binding sites in the *glpFK* operon, both with essential bases present at positions 7 and 16, and the high degree of match of the *glpT* site with consensus (with both essential bases present). In contrast, *glpD* has contains only a single binding site for

CRP, a site which has a substitution at position 7.

The response of the* glp *operators to the repressor:

The expression of the various members of the *glp* regulon also differs in response to the presence of the *glp* repressor. In *in vivo* experiments which measured the activity of the gene products, the *glpD* gene exhibited a greater sensitivity to repression than did the *glpT* gene or the *glpK* gene (75). The results of the *in vivo* titration experiments in this study contradict this conclusion. When the affinities of the *glpD*, *glpACB-TQ*, and *glpFK* operators for the *glp* repressor were measured using the *in vivo* titration assay, the construct in which all six *glpFK* operators were present displayed the greatest affinity for the repressor (Table V). While these results do not agree with the early proposal of regulation of Freedberg and Lin (75), the methodology utilized in this study can be seen as a more accurate measure of the expression of these operons in response to the repressor. The early studies concentrated on the activities of the gene products rather than the expression of the genes but did not take into account the two forms of allosteric inhibition to which glycerol kinase is subject. The following discussion focuses on several factors which may contribute to the differential response of the *glp* operons to the repressor: the number of repressor binding sites and their positions relative to the

promoter elements of the operon, the intrinsic affinity of the individual binding sites for the repressor, and cooperativity between the binding sites. Comparison of the *glpD*, *glpACB-TQ*, and *glpFK* operons shows that all of these factors except the last differ.

The number and positions of the* glp *operators:

Although each of these operons has tandemly repeated operators which bind the *glp* repressor cooperatively, the positions and number of binding sites for the *glp* repressor vary (Fig. 18, Table XII). In the case of the *glpFK* operon, six of these elements are seen. Two operators are located within the *glpK* coding region; their function will be examined separately. Four operators are located in the promoter/operator region of the *glpFK* operon, arranged tandemly beginning at -89 and extending to and overlapping the -10 sequence. The second and third operators coincide with the cAMP-CRP binding sites and binding of the repressor could prevent the binding of CRP. The third operator also covers the proposed -35 region of the promoter. The initial reversible contact between RNA polymerase and DNA is believed to occur at the -35 region, followed by nearly irreversible contact at the -10 site. Binding of the repressor to the *glpFK* operators could, therefore, block the initial contact and the binding of RNA polymerase to the promoter.

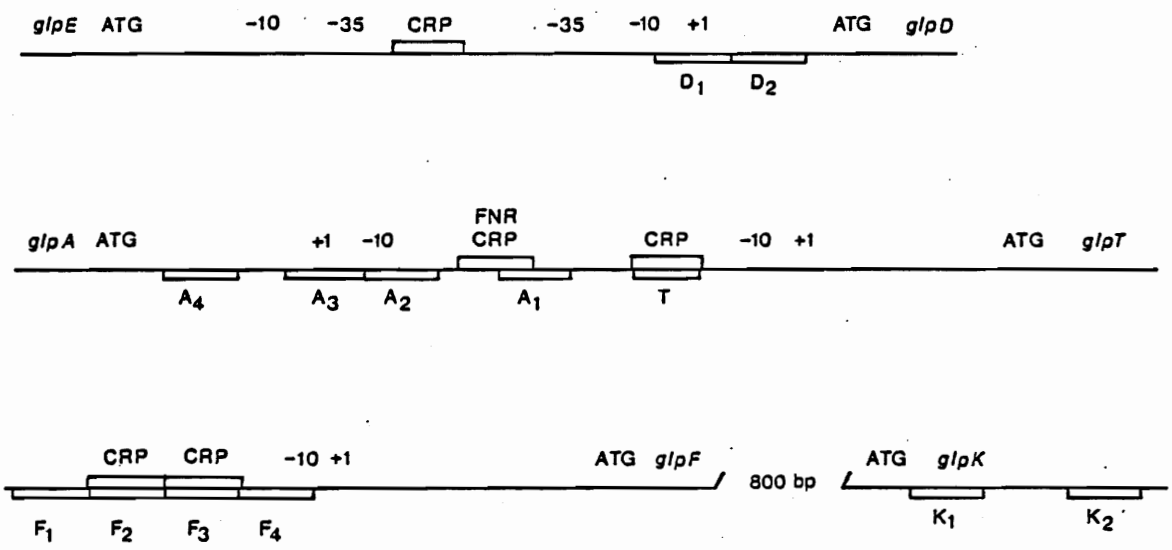


Fig. 18. **Arrangement of operator sites of the *glp* operons.** The promoter/operator regions for each operon are depicted, with promoter sequences indicated above and operator sites shown below each operon.

Table XII. Comparison of operators for the glycerol-P genes.

Operator	Position[a]	Sequence	% Match with Consensus
*glpD*1	-10	TATGTTCGAT AACGAACATT	100
*glpD*2	+11	TATGAGCTTT AACGAAAGTG	65
*glpA*1	-60	AATGTTCAAA ATGACGCATG	55
*glpA*2	-28	ACTTTCGAAT TATGAGCGAA	40
*glpA*3	-8	TATGCGCGAA ATCAAACAAT	70
*glpA*4	+33	AATGGTAAAA AACGAACTTC	65
glpTQ	-51	TGTGTGCGGC AATTCACATT	65
*glpF*1	-89	ATGGCGCGAT AACGCTCATT	65
*glpF*2	-68	TATGACGAGG CACACACATT	55
*glpF*3	-47	TAAGTTCGAT ATTTCTCGTT	65
*glpF*4	-27	TTTGCTCGTT AACGATAAGT	70
*glpK*1	+993	CGCGGTCGTA ATGGATCACG	45
*glpK*2	+1035	GCAGCGCGAA TTTGAGCAAA	40
Consensus		TATGTTCGAT AACGAACATT	(100)

a: The number refers to the position of the first base listed relative to the start point of transcription. All operators are written in the direction of transcription of the indicated gene.

The *glpACB-TQ* operons have a complex organization of five operator sites. Two tandem operators are located from +12 to -28 relative to the startpoint of transcription of *glpACB*. Single operator sites are positioned at -60 to -40 and +33 to +53 relative to the startpoint of transcription for *glpACB*, and -51 to -31 relative to the startpoint of transcription for *glpTQ* (A. Bhattacharya and T.J. Larson, unpublished results). The -60 operator of *glpACB* overlaps the binding site for the CRP and FNR proteins. Although no operator site is located directly at the -35 sequence of the *glpACB* operon, binding of the repressor to the two flanking operators could sterically hinder RNA polymerase from making contact with the -35 region. In addition, the *glpACB* tandem operators fully cover the -10 sequence for the operon and could further hinder initiation of transcription by the polymerase. Initial contact by the polymerase may be blocked at the -35 region of the *glpTQ* operon by binding of the repressor at the single operator in this region. The *glpTQ* operator overlaps the cAMP-CRP binding site and thus binding of the repressor could also block the binding of CRP.

In contrast, the operators of the *glpD* gene are positioned from -10 to +30 (77), a situation which could allow the polymerase to make initial contact with the DNA and perhaps invade the -10 site, resulting in some transcription of the *glpD* gene even in the presence of the repressor.

The significance of the position of an operator:

The role of the position of an operator has been studied in detail in the *lac* and *trp* systems. Varying the position of the operator in the *lac* operon results in differing degrees of repression (130). Repression of *lacZ* obtained by the use of a centrally positioned operator (-12 to -33) was 50- to 70-fold greater than that found with an operator at +1 to +21 or -39 to -60. Both the *glpFK* and *glpACB* operons have operators which occupy this central position; the *glpD* gene does not. Studies with the *trp* repressor of *E. coli* show a similar situation exists with the operators of the *trp* and *aroH* operons and the autoregulation of the *trpR* gene (131). The *trp* repressor binds to the *trp* operator at -22 to -2 (similar to the situations in the *glpFK* and *glpACB* operon), to the *aroH* operon from -49 to -29 (like *glpFK*), and to the *trpR* operator from -11 to +10 (similar to the situation of *glpD*). Although the *trp* repressor binds to each operator with similar affinity *in vitro*, repression of transcription *in vivo* was 300-fold for the *trp* operon, 6-fold for the *aroH* operon, and only 3-fold for the *trpR* gene. This result suggests that the relative positions of the operators can cause differential regulation of expression and, in the case of these three *glp* operons, that one might expect the *glpFK* and *glpACB-TQ* operons to be more responsive to the *glp* repressor than is the *glpD* gene.

The intrinsic affinity of the glp *operators for the repressor:*

Differential regulation could also result from the variable inherent affinity of the *glp* repressor for the individual binding sites of the operons. Comparison of the repressor binding sites of the *glpFK* operon to those identified for the *glpD* (77), *glpACB*, and *glpTQ* (A. Bhattacharya and T.J. Larson, unpublished results) operators has allowed the formulation of a repressor-binding consensus sequence: TATGTTCGAT AACGAACATT. As seen in Table XI, the binding sites in the *glpD* operator region have identities of 100% and 65% to the consensus. The *glpACB-TQ* operators display identities of 55%, 40%, 70%, 65%, and 65%, while the binding sites in the *glpFK* control region have identities of 65%, 55%, 65%, and 70% with the consensus. The operator sites within the *glpK* coding region display identities of 45% and 40%.

Previous *in vivo* studies have shown the expression of the *glpD* gene to be most sensitive to the *glp* repressor (75). The high percentage of identity of $glpD_1$ with the consensus sequence could play a role in this observed sensitivity. Evidence of binding is seen at a repressor concentration of 0.5 nM for the *glpD* and *glpACB-TQ* operators (77; A. Bhattacharya and T.J. Larson, unpublished results), while a concentration of 1.0 nM repressor is necessary for total protection of the *glpFK* operators from DNase I digestion. In

the *glpFK* and *glpD* operons, binding appears to be cooperative; either all or none of the sites in the regulatory region are protected from DNase I digestion by a given concentration of the *glp* repressor.

The role of the* glpK *intragenic operators:

Additional control over the expression of the *glpFK* operon is provided by the presence of two repressor binding sites within the coding region of the *glpK* gene. When the *glpFK* operators are present on multicopy plasmids, the *glp* repressor is titrated off of the *glpT-lacZ* fusion, resulting in expression of ß-galactosidase in a $glpR^{n}$ strain. The level of expression is a function of the number of operators and the affinity of the operators for the repressor. The affinity of the four tandem operators in the promoter/operator region was approximately 10 times higher than that of the two *glpK* operators. When all six operators are present, the affinity for the repressor is higher than the sum of the affinities for the two sets of operators, indicating cooperativity. This cooperativity implies a multipartite interaction between the repressor and the operators with the formation of a loop of DNA.

Although additional binding sites within the coding regions of other *glp* genes have not been demonstrated, similar arrangements are observed in the *lac* and *gal* systems (52, 53).

In these systems, the existence of a repression loop formed by protein-protein interaction of regulatory molecules bound to distal operators is proposed. *In vivo* studies show that binding of the *lac* repressor at the intragenic O2 strengthens binding at O1 approximately 3-fold (61). A significant decrease in repression is seen upon mutation of O2 (62). Looping has been verified *in vitro* by electron microscopy (60). Derepression is also seen when either O_I or O_E of the *gal* operon undergoes mutation (66). In the case of the *gal* operon, experimental results indicate looping, but looping has not been verified. Although the role that a DNA loop plays in repression has not been defined, the "chelating" action of two separated operators binding to repressor is thought to result in a high degree of repression even if the affinity of the repressor for each site is low (67). It has been proposed that the formation of the loop may either increase the stability of the primary repressor-operator complex or change the DNA structure in a way that would make isomerization by RNA polymerase difficult (55). Although the *glpK* operator sites are approximately 1 Kb from the operator sites in the promoter/operator region of the operon, a distance greater than that usually seen in regulatory looping, looping of 880 bp of DNA *in vivo* has been verified in the *deo* system (67). Insertion of an operator up to 4 kb downstream from the start of transcription causes looping which results in greater

repression of expression of the *deo* operon (132).

The existence and positions of the 4 binding sites within the *glpFK* control region and the presence of the two intragenic in *glpK* suggests that the regulation of the *glpFK* operon should be very responsive to the repressor. The *in vivo* repressor titration experiments corroborate this proposal. Freedberg and Lin supported their assertion that *glpD* was more responsive to the repressor by the necessity of the cell to maintain the intracellular concentration of glycerol-P for phospholipid biosynthesis (75). Although a more stringent regulation of the *glpFK* appears to be contrary to this need, the formation of the repression loop in the operon provides the cell with a sensitive, reversible control of the system. The loop allows binding of the repressor at low concentrations to sites with low intrinsic affinity for the protein. The dissociation rates from these sites is high, permitting small alterations in the concentration of the repressor to result in effective regulation. Small decreases in active repressor concentration can cause a large decrease in the occupancy of the binding sites. Thus, the appropriate intracellular concentration of glycerol-P is maintained. While the regulation of the *glpACB-TQ* operons will be addressed in future communications, it should be mentioned here that evidence for a looping mechanism for the regulation of expression of these operons has been obtained (A.

Bhattacharya and T.J. Larson, unpublished results). This supports the conclusion presented here that the expression of the *glpACB-TQ* operons exhibits a greater response to the repressor than does *glpD*. In addition, the aerobic dehydrogenase faces regulation by the high K_m (1-2 mM) for its substrate (2). Although this value has been determined only *in vitro* and is dependent on the electron acceptor, the K_m is sufficiently high to maintain enough glycerol-P in the cell to support phospholipid biosynthesis, reducing the need for stringent control at the level of gene expression.

CHARACTERIZATION OF THE GLPF AND GLPK PROTEINS

Analysis of the deduced amino acid sequence of GlpF indicates that it has extremely low polarity and is an integral membrane protein consisting of 6 potential membrane spanning helices. The proposed structure is unique among prokaryotic transport proteins. The amino acid sequences of the arabinose-H^+, xylose-H^+ (133), lactose-H^+ (134), and melibiose-Na^+ (135) transport proteins of *E. coli* have been used to analyze the structures of these membrane transporters. Each is proposed to have 12 membrane spanning α-helices. Indeed, with the exception of several eukaryotic mitochondrial and chloroplast ion transport proteins and the major intrinsic

protein (MIP) of the bovine lens fiber membrane (11) which display six transmembrane helices, all membrane carriers identified thus far appear to consist of 12 membrane spanning helices (136). These helices are believed to be arranged in two bundles of six helices each. If this is the general structure of membrane transporters, the GlpF protein may function as a homodimer with the two six-member domains arranged in the membrane to form a channel through which substrates pass. It is interesting to note that the structure of MIP, which displays a high degree of identity with GlpF in a protein similarilty search, appears to be similar to that of the facilitator. The regions of the amino acid sequences of the two genes which have the highest levels of similarities occur at similar locations in the proposed structures of the two genes. The first of these areas is located in the first proposed transmembrane α-helix of each protein. The second region with a high degree of similarity begins in the second transmembrane sequence and extends into the cytoplasm. The last region of similarity is situated in the periplasm between the fifth and sixth transmembrane sequences. Although the Nod-26 protein shares these regions of similarity, its proposed structure is not similar to that of the facilitator and MIP. This peribacteroid membrane Nod-26 has four regions which display sufficient hydrophobicity to traverse the membrane, only one of these domains can be folded into a

stable α-helix. The other segments have ß-sheet structures (12). Baker and Saier used the ALIGN program to assess the homology of these three proteins and determined that the proteins probably do arise from a common ancestor (10). It is interesting to note that this analysis was performed using an amino acid sequence of the facilitator which differs from ours (at residue 34) and that the last region of similarity between the three proteins was not identified.

The GlpK protein has been characterized in detail by a number of researchers (13, 19-23). The method of overproduction and purification of the kinase utilized in this study provided a substantial amount of the protein (50 mg kinase from 600 ml culture). The kinase is subject to post-translational modification; the N-formylmethionine residue is cleaved, as is seen by the N-terminal amino acid analysis of the protein.

Preparation of the purified glycerol kinase for the amino acid analysis resulted in the serendipitous discovery of the effect of the presence of glycerol on the native structure of the kinase. All preparations to date have been maintained in concentrations of 10% glycerol or higher. Analysis by G-200 gel filtration showed an increase in the apparent molecular weight of the enzyme from 110,000 (in the absence of glycerol) to a value of 195,000 in the presence of 10 mM glycerol. An increase in apparent molecular weight was also seen with the

addition of glycerol in nondenaturing gel electrophoresis experiments. After the initial findings in this study which indicated a dimeric native structure, preparations of glycerol kinase were maintained both in 10% glycerol and in glycerol-free conditions. The catalytic properties of the samples did not vary with the methods of storage. These results, taken with those obtained by nondenaturing gel electrophoresis and gel filtration studies, indicate that glycerol exerts an effect on the native structure of glycerol kinase, causing a change in conformation or aggregation of subunits.

LITERATURE CITED

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

2. Lin, E.C.C. (1987) Sugars, polyols, and carboxylates. pp. 263-264. In F.C. Neidhardt (ed. in chief), J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, American Society for Microbiology, Washington, D.C.

3. Sanno, Y., Wilson, T.H., and Lin, E.C.C. (1968) Control of permeation to glycerol in cells of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **32**, 344-349.

4. Alemohammad, M.M., and Knowles, C.J. (1974) Osmotically induced volume and turbidity changes of *Escherichia coli* due to salts, sucrose and glycerol, with particular reference to the rapid permeation of glycerol into the cell. *J. Gen. Microbiol.* **82**, 125-142.

5. Saier, M. H,, Jr. (1985) *In* Mechanisms and regulation of carbohydrate transport in bacteria. Academic Press, New York, NY.

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

7. 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.

8. Mizuno, T. (1987) Random cloning of bent DNA segments from *Escherichia coli* chromosome and primary characterization of their structures. *Nucleic Acids Res.* **15**, 6827-6841.

9. Muramatsu, S., and Mizuno, T. (1989) Nucleotide sequence of the region encompassing the *glpFK* operon and its upstream region containing a bent DNA sequence of *Escherichia coli*. *Nucleic Acids Res.* 17, 4378.

10. Baker, M.E., and Saier, M.H. Jr. (1990) A common ancestor for bovine lens major intrinsic protein, soybean nodulin-26, and *E. coli* glycerol facilitator. *Cell* **60**, 185-196.

11. Gorin, M.B., Yancey, S.B., Cline, J., Revel, J.-P., and Horwitz, J. (1984) The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning. *Cell* **39**, 49-59.

12. Fortin, M.G., Morrison, N.A., and Verma, D.P.S. (1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. *Nucleic Acids Res.* **15**, 813-824.

13. Hayashi, S.-I., and Lin, E.C.C. (1965) Capture of glycerol by cells of *Escerichia coli. Biochim. Biophys. Acta.* **94**, 479-487.

14. Zwaig, N., Kistler, W.S., and Lin, E.C.C. (1970) Glycerol kinase, the pacemaker for the dissimilation of glycerol in *Escherichia coli. J. Bacteriol.* **102**, 753-759.

15. Berman, M., and Lin, E.C.C. (1971) Glycerol-specific revertants of a phosphoenolpyruvate phosphotransferase mutant: suppression by the desensitization of glycerol kinase to feedback inhibition. *J. Bacteriol.* **105**, 113-120.

16. Freedberg, W.B., Kistler, W.S., and Lin, E.C.C. (1971) Lethal synthesis of methylglyoxal by *Escherichia coli* during unregulated glycerol metabolism. *J. Bacteriol.* **108**, 137-144.

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

18. Novotny, M.J., Frederickson, W.L., Waygood, E.B., and Saier, M.H., Jr. (1985) Allosteric regulation of glycerol kinase by enzyme III^{glc} of the phosphotransferase system in *Escherichia coli* and *Salmonella typhimurium. J. Bacteriol.* **162**, 810-816.

19. Thorner, J.W., and Paulus, H. (1971) Composition and subunit structure of glycerol kinase from *Escherichia coli. J. Biol. Chem.* **246**, 3885-3894.

20. Pettigrew, D.W., Ma, D.-P., Conrad, 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.

21. de Riel, J.K., and Paulus, H. (1978) Subunit dissociation in the allosteric regulation of glycerol kinase from *Escherichia coli*. 1. Kinetic evidence. *Biochemistry* **17**, 5134-5140.

22. Faber, H.R., Pettigrew, D.W., and Remington,S.J. (1989) Crystallization and preliminary X-ray studies of *Escherichia coli* glycerol kinase. *J. Mol. Biol.* **207**, 637-639.

23. Crans, D.C., and Whitesides, G.M. (1985) Glycerol kinase: substrate specificity. *J. Am. Chem. Soc.* **107**, 7008-7018.

24. Larson, T.J., Ehrmann, M., and Boos, W. (1983) Periplasmic glycero-phosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the *glp* regulon. *J. Biol. Chem.* **258**, 5428-5432.

25. Larson, T.J. (1990) Glycerol 3-phosphate/phosphate antiport. (in press) *In* M.G.P. Page and P.J.F. Henderson (eds.) Ion-coupled Sugar Transport in Microorganisms. CRC Uniscience Series, Boca Raton, FL.

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

27. Hayashi, S.-I., Koch, J.P., and Lin, E.C.C. (1964) Active transport of L-α-glycerophosphate in *Escherichia coli*. *J. Biol. Chem.* **239**, 3098-3105.

28. 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.

29. Ehrmann, M., Boos, W., Ormseth, E., Schweizer, 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.

30. Eiglmeier, K., Boos, W., and Cole, S.T. (1987) Nucleotide sequence and transcriptional startpoint of the *glpT* gene of *Escherichia coli*: extensive sequence homology of the glycerol-3-phosphate transport protein with components of the hexose-6-phosphate transport system. *Mol. Microbiol.* **1**, 251-258.

31. Gött, P., and Boos, W. (1988) The transmembrane topology of the *sn*-glycerol-3-phosphate permease of *Escherichia coli* analyzed by *phoA* and *lacZ* protein fusions. *Mol. Microbiol.* **2**, 655-663.

32. Ambudkar, S.V., Larson, T.J., and Maloney, P.C. (1986) Reconstitution of sugar phosphate transport systems of *Escherichia coli*. *J. Biol. Chem.* **261**, 9083-9086.

33. Choi, Y.-L., Kawase, S., Kawamukai, M., Utsumi, R., Sakai, H., and Komano, T. (1989) Nucleotide sequence of the glycerol-3-phosphate dehydrogenase gene of *Escherichia coli* and regulation by the cAMP-CRP complex. *Agric. Biol. Chem.* **53**, 1135-1143.

34. 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* K-12. *J. Bacteriol.* **170**, 2448-2456.

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

36. Green, P.R., Merrill, A.H., Jr., and Bell, R.M. (1981) Membrane phospholipid synthesis in *Escherichia coli*. *J. Biol. Chem.* **256**, 11151-11159.

37. Hirabayashi, T., Larson, T.J., and Dowhan, W. (1976) Membrane-associated phosphatidylglycerophosphate synthetase from *Escherichia coli*: purification by substrate affinity chromatography on cytidine 5'-diphospho-1,2-diacyl-*sn*-glycerol sepharose. *Biochemistry* **15**, 5205-5211.

38. Edgar, J.R., and Bell, R.M. (1978) Biosynthesis in *Escherichia coli* of *sn*-glycerol 3-phosphate, a precursor of phospholipid. *J. Biol. Chem.* **253**, 6354-6363.

39. Pardee, A.B., Jacob, F., and Monod, J. (1959) The genetic control and cytoplasmic expression of "inducibility" in the synthesis of ß-galactosidase

by *E. coli*. *J. Mol. Biol.* **1**, 165.

40. Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318-356.

41. Jobe, A., and Bourgeois, S. (1972) *lac* repressor-operator interaction. VI. The natural inducer of the *lac* operon. *J. Mol. Biol.* **69**, 397-408.

42. Müller-Hill, B., Rickenberg, H.V., and Wallenfels, K. (1964) Specificity of the induction of the enzymes of the *lac* operon in *Escherichia coli*. *J. Mol. Biol.* **10**, 303-318.

43. Beckwith, J. (1987) The lactose operon. In F.C. Neidhardt (ed. in chief), J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, American Society for Microbiology, Washington, D.C.

44. Beckwith, J., Grodzicker, T., and Arditti, R. (1972) Evidence for two sites in the *lac* promoter region. *J. Mol. Biol.* **69**, 155-160.

45. Hopkins, J. (1974) A new class of promoter mutation in the lactose operon of *Escherichia coli*. *J. Mol. Biol.* **87**, 715-724.

46. Majors, J. (1975) Specific binding of CAP factor to *lac* promoter DNA. *Nature* **256**, 672-674.

47. Brown, C.E., and Hogg, R.W. (1972) A second transport system for L-arabinose in *Escherichia coli* B/r controlled by the *araC* gene. *J. Bacteriol.* **111**, 606-613.

48. Casadaban, M. (1972) Regulation of the regulatory gene for the arabinose pathway, *araC*. *J. Mol. Biol.* **104**, 557-566.

49. Nakanishi, S., Adhya, S., Gottesman, M., and Pastan, I. (1973) Studies on the mechanism of action of the *gal* repressor. *J. Biol. Chem.* **248**, 5937-5942.

50. Musso, R.E., DiLauro, R., Adhya, S. and de Crombrugghe, B. (1977) Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. *Cell* **12**, 847-854.

51. Ptashne, M. (1986) Genetic regulation by proteins acting at a distance. *Nature* 322, 697-701.

52. Reznikoff, W.S., Winter, R.B., and Hurley, C.K. (1974) The location of the repressor binding sites in the *lac* operon. *Proc. Natl. Acad. Sci. U.S.A.* 71, 2314-2318.

53. Irani, M.H., and Orosz, L. (1983) A control element within a structural gene: the *gal* operon of *Escherichia coli*. *Cell* 32, 783-788.

54. Lee, N.L., Gielow, W.O., and Wallace, R.G. (1981) Mechanism of *araC* autoregulation and the domains of two overlapping promoters, P_C and P_{BAD}, in the L-arabinose regulatory region of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 78, 752-756.

55. Dandanell, G., and Hammer, K. (1985) Two operator sites separated by 559 base pairs are required for *deoR* repression of the *deo* operon of *Escherichia coli*. *EMBO J.* 4, 3333-3338.

56. Adhya, S. (1989) Multipartite genetic control elements: communication by DNA loop. *Annu. Rev. Genet.* 23, 227-250.

57. Ptashne, M. (1986) A genetic switch: gene control and phage. Cell Press and Blackwell Scientific Publications, Cambridge, MA and Palo Alto, CA.

58. Hochschild, A., and Ptashne, M. (1988) Interaction at a distance between repressors disrupts gene activation. *Nature* 336, 353-357.

59. Borowiec, J.A., Zhang, L., Sasse-Dwight, S., and Gralla, J.D. (1987) DNA supercoiling promotes formation of a bent repression loop in *lac* DNA. *J. Mol. Biol.* 196, 101-111.

60. Krämer, H., Niemöller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1987) *lac* repressor forms loops with linear DNA carrying two suitably spaced *lac* operators. *EMBO J.* 6, 1481-1491.

61. Flashner, Y., and Gralla, J.D. (1988) Dual mechanism of repression at a distance in the *lac* operon. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8968-8972.

62. Eismann, E., von Wilcken-Bergman, B., and Müller-Hill, B. (1987) Specific destruction of the second *lac* operator

decreases repression of the *lac* operon in *Escherichia coli* fivefold. *J. Mol. Biol.* **195**, 949-952.

63. Winter, R.B., and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor-operator interaction: equilibrium measurements. *Biochemistry* **20**, 6948-6960.

64. Sasse-Dwight, S., and Gralla, J.D. (1988) Probing co-operative DNA-binding *in vivo*. The *lac* O_1:O_3 interaction. *J. Mol. Biol.* **202**, 107-119.

65. Hou, L., Martin, K.J., and Schleif, R. (1988) Alternative DNA loops regulate the arabinose operon in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5444-5448.

66. Haber, R., and Adhya, S. (1988) Interaction of spatially separated protein-DNA complexes for control of gene expression: operator conversions. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9683-9687.

67. Valentin-Hansen, P., Albrechtsen, B., and Love-Larson, J.E. (1986) DNA-protein recognition: demonstration of three genetically separated operator elemants that are required for repression of the *Escherichia coli deoCABD* promoters by the DeoR repressor. *EMBO J.* **5**, 2015-2021.

68. Amouyal, M., Mortensen, L., Buc, H., and Hammer, K. Single and double loop formation when *deoR* repressor binds to its natural operator sites. *Cell* **58**, 545-551.

69. Koch, J.P., Hayashi, S.-I., and Lin, E.C.C. (1964) The control of dissimilation of glycerol and L-α-glycerophosphate in *Escherichia coli*. *J. Biol. Chem.* **239**, 3106-3108.

70. Schweizer, H., and Larson, T.J. (1987) Cloning and characterization of the aerobic *sn*-glycerol-3-phosphate dehydrogenase structural gene *glpD* of *Escherichia coli* K-12. *J. Bacteriol.* **169**, 507-513.

71. 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.

72. Cozzarelli, N.R., and Lin, E.C.C. (1966) Chromosomal location of the structural gene for glycerol kinase in *Escherichia coli*. *J. Bacteriol.* **91**, 1763-1766.

73. Berman-Kurtz, M., Lin, E.C.C., and Richey, D.P. (1971) Promoter-like mutant with increased expression of the glycerol kinase operon of *Escherichia coli*. *J. Bacteriol.* **106**, 724-731.

74. Larson, T.J., Ye, S., Weissenborn, D.L., Hoffmann, H.J., and Schweizer, H.S. (1987) Purification and characterization of the repressor for the *sn*-glycerol 3-phosphate regulon of *Escherichia coli* K12. *J. Biol. Chem.* **262**, 15869-15874.

75. 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.

76. 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.

77. Ye, S., and Larson, T.J. (1988) Structure of the promoter and operator of the *glpD* gene encoding aerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* **170**, 4209-4215.

78. Barkely, M.D., and Bourgeois, S. (1980) *In* The Operon (Miller, J.H., and Reznikoff, W.S., eds) pp 177-220 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

79. de Crombrugghe, B., Busby, S., and Buc, H. (1984) Cyclic AMP receptor protein: role in transcription activation. *Science* **224**, 831-838.

80. Shanblatt, S.H., and Revzin, A. (1987) Interactions of the catabolite activator protein (CAP) at the galactose and lactose promoters of *Escherichia coli* probed by hydroxyl radical footprinting. *J. Biol. Chem.* **262**, 1141-11427.

81. Ebright, R.H., Kolb, A., Buc, H., Kunkel T.A., Krakow, J.S., and Beckwith, J. (1987) Role of glutamic-181 in DNA-sequence recognition by the catabolic gene activator protein (CAP) of *Escherichia coli*. Altered DNA-sequence-recognition properties of [Val^{181}]CAP and [Leu^{181}]CAP. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6083-6087.

82. Kuritzkes, D.R., Zhang, X.-Y., and Lin, E.C.C. (1984) Use of Φ(*glp-lac*) in studies of respiratory regulation of the *Escherichia coli* anaerobic *sn*-glycerol-3-

phosphate dehydrogenase genes (*glpAB*). *J. Bacteriol.* **157**, 591-598.

83. Stewart, V. (1988) Nitrate respiration in relation to facultative metabolism in enterobacteria. *Microbiol. Rev.* **52**, 190-232.

84. Trageser, M., and Unden, G. (1989) Role of cysteine residues and of metal ions for the regulatory function of FNR, the transcriptional regulator of anaerobic respiration in *Escherichia coli*. *Mol. Microbiol.* 3, 593-599.

85. Trageser, M., Spiro, S., Duchêne, A., Kojro, E., Fahrenholz, F., Guest, J.R., and Unden, G. (1990) Isolation of intact FNR protein (M_r 30000) of *Escherichia coli*. *Mol. Microbiol.* **4**, 21-27.

86. Unden, G., Trageser, M., and Duchêne, A. (1990) Effect of positive redox potentials (>+400 mV) on the expression of anaerobic respiratory enzymes in *Escherichia coli*. *Mol. Microbiol.* **4**, 315-319.

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

88. Spiro, S., and Guest, J.R. (1987) Activation of the *lac* operon of *Escherichia coli* by a mutant FNA protein. *Mol. Microbiol.* **1**, 53-58.

89. 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. U.S.A.* **85**, 1888-1892.

90. Iuchi, S., Cameron, D.C., 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.

91. 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 aerobic glycerol-3-phosphate dehydrogenase in *Escherichia coli*: further characterization of respiratory control. *J. Bacteriol.* **172**, 179-184.

92. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., and Boyer, H.W. (1977) Construction and characterization of new cloning vehicles, II. A multipurpose cloning system. *Gene* **2**, 95-113.

93. Remaut, E., Stanssens, P., and Fiers, W. (1981) Plasmid vectors for high-efficiency expression controlled by the p_L promoter of coliphage lambda. *Gene*, **15**, 81-93.

94. Casadaban, M. (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.

95. Silhavy, T.J., Hartig-Beecken, I., and Boos, W. (1976) Perisplasmic protein related to the *sn*-glycerol-3-phosphate transport system of *Escherichia coli*. *J. Bacteriol.* **126**, 951-958.

96. Cozzarelli, N.R., Freedberg, W.B., and Lin, E.C.C. (1968) Genetic control of the L-α-glycerophosphate system in *Escherichia coli*. *J. Mol. Biol.* **31**, 371-387.

97. Silhavy, T.J., Berman, M.L., and Enquist, L.W. (1984) *In* Experiments with Gene Fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

98. Schneider, K., and Beck, C.F. (1987) New expression vectors for identifying and testing signal structures for initiation and termination of transcription. *Methods in Enzymology* **153**, 452-461.

99. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-109.

100. Focus (1987) **9**, pp. 3, 13. Bethesda Research Laboratories. Gaithersburg, MD.

101. Adhya, S., and Gottesman, M. (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell* **29**, 939-944.

102. Schneider, K., and Beck, C.F. (1986) Promoter-probe vectors for the analysis for divergently arranged promoters. *Gene* **42**, 37-48.

103. Chang, A.C.Y., and Cohen, S.N. (1978) Construction and Characterization of amplifiable multicopy DNA cloning

vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141-1156.

104. Miller, J.H. (1972) *In* Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

105. Wieslander, L, (1979) A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature gels. *Anal. Biochem.* **98**, 305-309.

106. Rodriguez, R.L., and Tait, R.C. (1983) *In* Recombinant DNA techniques: an introduction. pp. 164-165. Addison-Wesley Publishing Co., Reading, MA.

107. Hackett, P.B., Fuchs, J.A., and Messing, J.W. (1988) *In* An Introduction to Recombinant DNA Techniques. Benjamin/Cummings Publishing Co, Inc., Menlo Park, CA.

108. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.

109. Bialkowska-Hobrzanska, H., Gilchrist, C.A., Denhardt, D.T. (1985) *Escherichia coli rep* gene: identification of the promoter and terminus of the Rep protein. *J. Bacteriol.* **164**, 1004-1010.

110. Alam, J., Whitaker, R.A., Krogmann, D.W., and Curtis, S.E. (1986) Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **168**, 1265-1271.

111. Bublitz, C., and Wieland, O. (1962) Glycerokinase. *Methods Enzymol.* **5**, 354-361.

112. 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.

113. Laemmli, U. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

114. Rock, C.O., and Cronan, J.E., Jr. (1981) Acyl carrier protien from *Escherichia coli*. *Methods Enzymol.* 71, 341-364.

115. Sharp, P.M., and Li, W.-H. (1986) Codon usage in

regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. *Nucleic Acids Res.* **14**, 7737-7749.

116. Miller, C.G. (1987) Protein degradation and proteolytic modification. pp. 680-691. In F.C. Neidhardt (ed. in chief), J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, American Society for Microbiology, Washington, D.C.

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

118. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125-142.

119. Novotny, J., and Auffray, C. (1984) A program for prediction of protein secondary structure from nucleotide sequence data: application to histocompatibility. *Nucleic Acids Res.* **12**, 243-255.

120. Hayashi, S.-I., and Lin, E.C.C. (1967) Purification and properties of glycerol kinase from *Escherichia coli*. *J. Biol. Chem.* **242**, 1030-1035.

121. Rule, G.S., Pratt, E.A., Chin, C.C.Q., Wold, F., and Ho, C. (1985) Overproduction and nucletide sequence of the respiratory D-lactate dehydrogenase of *Escherichia coli*. *J. Bacteriol.* **161**, 1059-1068.

122. Thorner, J.W., and Paulus, H. (1973) Catalytic and allosteric properties of glycerol kinase from *Escherichia coli*. *J. Biol. Chem.* **248**, 3922-3932.

123. Hagerman, P.J. (1985) Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochemistry* **24**, 7033-7036.

124. Koo, H.-S., Wu, H.-M., and Crothers, D.M. (1986) DNA bending at adenine-thymine tracts. *Nature* **320**, 501-506.

125. Liu-Johnson, H.-N., Gartenberg, M.R., and Crothers, D.M. (1986) The DNA binding domain and bending angle of *E. coli* CAP protein. *Cell* **47**, 995-1005.

126. Wu, H.-M., and Crothers, D.M. (1984) The locus of

sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-513.

127. Lichenstein, H.S., Hamilton, E.P., and Lee, N. (1987) Repression and catabolite gene activation in the *araBAD* operon. *J. Bacteriol.* **169**, 811-822.

128. Shanblatt, S.H., and Revzin, A. (1986) Role of a second catabolite activator protein molecule in controlling initiation of transcription at the galactose operon of *Escherichia coli*. *Biochemistry* **25**, 5539-5546.

129. Nick, H., and Gilbert, W. (1985) Detection *in vivo* of protein-DNA interactions within the *lac* operon of *Escherichia coli*. *Nature* **313**, 795-798.

130. Lanzer, M., and Bujard, H. (1988) Promoters largely determine the efficiency of repressor action. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8973-8977.

131. Klig, L.S., Carey, J., and Yanofsky, C. (1988) *trp* repressor interactions with the *trp aroH* and *trpR* operators. *J. Mol. Biol.* **202**, 769-777.

132. Dandanell, G., Valentin-Hansen, P., Love Larsen, J.E., and Hammer, K. (1987) Long-range cooperativity between gene regulatory sequences in a prokaryote. *Nature* **325**, 823-826.

133. Maiden, M.C.J., Davis, E.O., Baldwin, S.A., Moore, D.C.M., and Henderson, P.J.F. (1987) mammalian and bacterial sugar transport proteins are homologous. *Nature* **325**, 641-643.

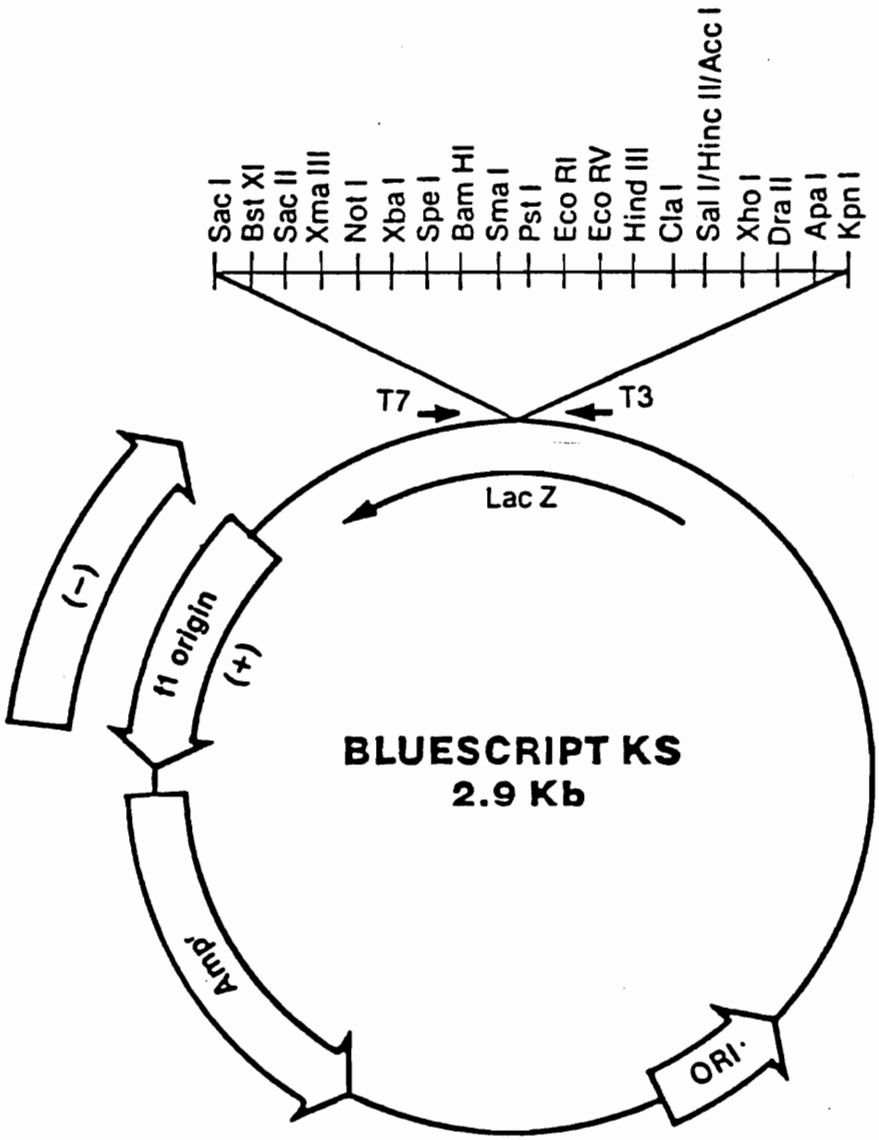
134. Foster, D.L., Boublik, M., and Kaback, H.R. (1983) Structure of the *lac* carrier protein of *Escherichia coli*. *J. Biol. Chem.* **258**, 31-34.

135. Yazyu, H., Shiota-Niiya, S., Shimamoto, T., Kanazawa, H., Futai, M., and Tsuchiya, T. (1984) Nucleotide sequence of the *melB* gene and characteristics of deduced amino acid sequence of the melibiose carrier in *Escherichia coli*. *J. Biol. Chem.* **259**, 4320-4326.

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

APPENDIX

STRUCTURE OF THE pBLUESCRIPT VECTOR KS+



Curriculum Vitae

Name: Deborah L. Weissenborn

Business Address: Department of Plant Pathology, Physiology and Weed Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 231-6757

Home Address: Rt. 3, Box 163
Christiansburg, Virginia 24073

Date of Birth: September 27, 1951

Education:

Ph.D.	1986-1990	Virginia Polytechnic Institute and State University Blacksburg, Virginia 24061 Major: Biochemistry
	1984-1986	University of North Dakota Grand Forks, North Dakota 58202 Major: Biochemistry and Molecular Biology
B.S.	1981-1984	Bemidji State University Bemidji, Minnesota 56601 Major: Environmental Science Minor: Biology
B.A.	1981-1984	Bemidji State University Bemidji, Minnesota 56601 Major: Chemistry

Honors:

Virginia Commonwealth Fellow, 1989-1990
Graduated *Summa cum laude*, Bemidji State University
Recipient of Outstanding Student in Environmental Studies, Bemidji State University, 1983
Recipient of Lakehead Science Scholarship, Bemidji State University, 1983-1984
Listed in Who's Who among Students in American Universities and Colleges, 1984

Professional Affiliation:

American Society for Microbiology, Student Member
Gamma Sigma Delta
Phi Lambda Upsilon

Publications:

1. Larson, T.J., Ye, S., Weissenborn, D.L., Hoffmann, H., and Schweizer, H.S. (1987) Purification and characterization of the repressor for the *glp* regulon of *Escherichia coli* K-12. *J. Biol. Chem.* **262**, 15869-15874.

2. Weissenborn, D.L., and Larson, T.J. Substrate-induced alteration of the native state of glycerol kinase of *Escherichia coli* K-12. In preparation.

3. Weissenborn, D.L., and Larson, T.J. Characterization of the control region of the *glpFK* operon of *Escherichia coli* and the nucleotide sequence of the *glpF* gene encoding the glycerol diffusion facilitator. In preparation.

Abstracts:

1. Weissenborn, D.L., and Larson, T.J. (1985) The effect of the *plsB26* allele on glycerol 3-phosphate acyltransferase activity in *Escherichia coli*. West Central States Biochemistry Conference. Vermillion, South Dakota.

2. Weissenborn, D.L., and Larson, T.J. (1987) Substrate-induced alteration of the native state of glycerol kinase of *Escherichia coli* K-12. Fed. Proc. **46**, 2009. Philadelphia, Pennsylvania.

3. Larson, T.J., Ye, S., Weissenborn, D.L., and Hoffmann, H.J. (1988) Regulation of glycerol 3-phosphate metabolism in *Escherichia coli* by the *glpR*-encoded repressor. International Congress of Biochemistry, Prague.

4. Weissenborn, D.L., Larson, T.J., Sweet, G.D., Wittekindt, N., and Gandor, C. (1989) The *glpFK* operon of *Escherichia coli* K-12. *J. Cell Biol.* **107**, 738a. San Francisco, California.

5. Larson, T.J., Weissenborn, D.L., Ye, S., Austin, D., and Leffler, C.T. (1989) Tandem operators control *sn*-glycerol 3-phosphate (*glp*) gene expression in

Escherichia coli. Gordon Res. Conf., Meriden, New Hampshire.

6. Weissenborn, D.L., and Larson, T.J. (1989) Regulation of the *glpFK* operon of *Escherichia coli* K-12. Virginia Biochemists' Conference, Mountain Lake, Virginia.

David L. Weissenborn