STRUCTURE AND FUNCTION OF THE REPRESSOR AND OPERATORS OF THE *sn*-GLYCEROL-3-PHOSPHATE REGULON OF *ESCHERICHIA COLI* K-12

by

Shanzhang Ye

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:

Timothy J. Larson, Chairman

Malcolm Potts

Thomas O. Sitz

John L. Johnson

Karl Pedersen

November, 1990

Blacksburg, Virginia

STRUCTURE AND FUNCTION OF THE REPRESSOR AND OPERATORS OF THE *sn*-GLYCEROL-3-PHOSPHATE REGULON OF *ESCHERICHIA COLI* K-12

by

Shanzhang Ye

Timothy J. Larson, Chairman

Biochemistry and Nutrition

(ABSTRACT)

The *glpD* gene, which encodes aerobic *sn*-glycerol 3-phosphate dehydrogenase, and the *glpR* gene, which encodes a repressor that negatively regulates the expression of the *glp* regulon, map near minute 75 on the linkage map of *Escherichia coli* K-12. In the present study, the nucleotide sequence of the 2895 base pair of DNA containing the *glpD* control region and the *glpE, glpG, glpR* genes was determined. The translation initiation codons with adjacent ribosome-binding sites were found for these four genes. The transcription start site of the *glpD* gene was identified 42 base pairs upstream from the proposed methionine start codon, preceded by a region containing typical -10 and -35 sequences found in bacterial promoters. A binding site for the cyclic AMP-cAMP receptor protein complex was located just upstream from the -35 sequence, centered at position -63. The interaction site for the *glp* repressor was identified by using DNase I footprinting. This region contained two tandemly repeated sequences which started at the -10 sequence and continued to position +38. The *glp* repressor contained 252 amino acid residues and had a molecular weight of

28,046 which was deduced from the nucleotide sequence. The position of the initiation codon was verified by determination of the amino acid sequence of the N-terminus of the purified *glp* repressor. The presumptive helix-turn-helix region of the repressor was located near the N-terminus (amino acids 22 to 41) at a position analogous to that found for the operator binding domain of other repressors such as the *deo* and *lac* repressors. The recognition helix of the *glp* repressor and the nucleotide sequence of the *glp* operator were very similar to those of the *deo* system. The presumptive *glpR* recognition helix was changed to the *deoR* recognition helix and the sixth amino acid arginine of the recognition helix was changed to alanine by site-directed mutagenesis. The mutant forms of the repressor had a greatly reduced affinity for the *glpD* operators *in vivo*, determined by measuring β-galactosidase activity in a strain carrying a *glpD-lacZ* fusion.

Acknowledgements

I wish to especially thank my advisor, Dr. Timothy J. Larson for his guidance, encouragement, support and patience throughout the course of this research. Thanks are also due to all the members of my committee, Drs. M. Potts, T. O. Sitz, J. L. Johnson and K. Pedersen for their valuable advice and help during my PhD study.

I am so grateful to my friends and colleagues in our laboratory, Deborah L. Weissenborn, Ali T. van Loo-Bhattacharya, Denise Austin and Won Oh, who made the lab a pleasant place in which to work. Thanks go to all other faculty members and friends in the Biochemistry Department who have helped in numerous ways to make this research a success.

I would like to thank the Department of Biochemistry and Nutrition for financial support during the period of my graduate studies.

I would also like to thank all my friends who have supported and made difficult times bearable, and helped make my tenure at Virginia Tech a memorable experience.

Finally, I would like to thank Dan Zhou for sharing her expertise during preparation of this dissertation.

Dedication

This dissertation is dedicated to my beloved mother, Weide Huang

and devoted father, Weiqin Ye for their

love, care, guidance and support.

List of Abbrreviations

A	absorbance
AA	amino acid
ATP	adenosine 5′-triphosphate
Ap	ampicillin
bp	base pairs
BSA	bovine serum albumin
cAMP	cyclic adenosine 3′, 5′-monophosphate
CIP	calf intestinal alkaline phosphatase
Cm	chloramphenicol
CRP	cAMP receptor protein
DAP	DL-2,6-diaminoheptanedioic acid
DHAP	dihydroxyacetone phosphate
dNTP	deoxynucleoside triphosphates
ddNTP	dideoxynucleoside triphosphates
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetate

EtBr	ethidium bromide
FAD	flavin adenine dinucleotide
FDP	fructose 1,6-bisphosphate
FNR	anaerobic transcriptional activator
HTH	helix-turn-helix
GAP	D-glyceraldehyde 3-phosphate
Glycerol-P	*sn*-glycerol 3-phosphate
IPTG	isopropylthio-β-D-galactopyranoside
kb	kilobase
Kd	dissociation constant
kDa	kilodalton
Km	kanamycin
Km	Michaelis-Menten constant
LB	Luria broth
MOPS	3-[N-morpholino]propanesulfonic acid
mRNA	messenger RNA
MW	molecular weight
NAD	nicotinamide adenine dinucleotide
ONPG	orthonitrophenyl-β-D-galactopyranoside
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
pfu	plaque-forming units

PMSF	phenylmethylsulfonyl fluoride
RF	replicative form
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
St	streptomycin
TB	Terrific broth
Tc	tetracycline
TE	Tris-EDTA
Tris	tris(hydroxymethyl)aminomethane
Tris-HCl	Tris-hydrochloride
U	units
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Table of Contents

List of Illustrations

List of Tables

INTRODUCTION

sn-Glycerol-3-phosphate has two metabolic fates in *Escherichia coli.* It is not only the precursor for phospholipid biosynthesis, but may also serve as the sole source of carbon and energy due to the presence of glycerol-P dehydrogenases. The aerobic glycerol-P dehydrogenase catalyzes the reaction which converts cytoplasmic glycerol-P to dihydroxyacetone phosphate for further catabolism. It is encoded by the *glpD* gene, one of the members of the *glp* regulon. This regulon encodes the proteins that carry out the catabolic steps for dissimilation of glycerol-P and its precursors. The *glp* regulon consists of at least five operons which are all regulated by a specific repressor encoded by the *glpR* gene. Repression mediated by the *glp* repressor can be relieved to differing extents in the presence of glycerol-P, the inducer for the system.

Many regulatory proteins bind their targets by using a bihelical domain for DNA recognition. In order to obtain molecular details of how *glp* repressor negatively regulates the expression of various *glp* operons, it was necessary to determine the nucleotide sequence of the genes, their structure and organization on the chromosome, and the function of the repressor as well. I report here the nucleotide sequence of the *glpEGR* genes, the organization of the control region for the *glpD* gene, and the structure, function and the interaction of the *glp* repressor with the operator DNA. The amino acid sequence of the GlpR protein

was deduced from the nucleotide sequence and compared with the sequences of other procaryotic regulatory proteins. Based upon the similarities of the proposed recognition helix of the *glp* repressor and the nucleotide sequence of the *glp* operator to those of the *deo* system, a putative DNA binding domain located in the amino-terminal region of the *glp* repressor was identified. The amino acid sequence of the proposed recognition helix of the *glp* repressor was altered by site-directed mutagenesis. The activities of both the wild-type and mutant forms of the repressor were assayed *in vivo*. The results strongly suggest that the putative DNA binding domain of the repressor is in fact involved in DNA recognition.

LITERATURE REVIEW

GLYCEROL PHOSPHATE METABOLISM IN *ESCHERICHIA COLI*

sn-Glycerol-3-phosphate (glycerol-P) is a ubiquitous compound in nature. Esterified derivatives of glycerol-P, the glycerophospholipids, are present in virtually all biological membranes. Thus, glycerol-P is a direct precursor for phospholipid biosynthesis (1). When glycerol-P or its precursors are present in excess, they can be used as sources of carbon and energy for a variety of microorganisms. The catabolism of glycerol-P and its precursors in *Escherichia coli* is mediated by the components of the *glp* regulon. This regulon is comprised of at least five operons under the control of a common repressor encoded by the *glpR* gene (Fig. 1) (2, 3). The affinity of the *glp* repressor for its various operators is reduced upon binding of glycerol-P, the inducer for the regulon.

In order to maintain the cellular concentration of glycerol-P at a level so that phospholipid synthesis can proceed at an optimal rate, organisms must be able to coordinate the levels of the catabolic activities with the levels of the biosynthetic activities. When the exogenous supply of glycerol-P is depleted, it would be desirable for the cell to prevent continued degradation of glycerol-P. If the degradation were not halted, a futile cycle involving the catabolic enzyme and the biosynthetic enzyme (glycerol-P synthase) catalyzing the reverse reaction

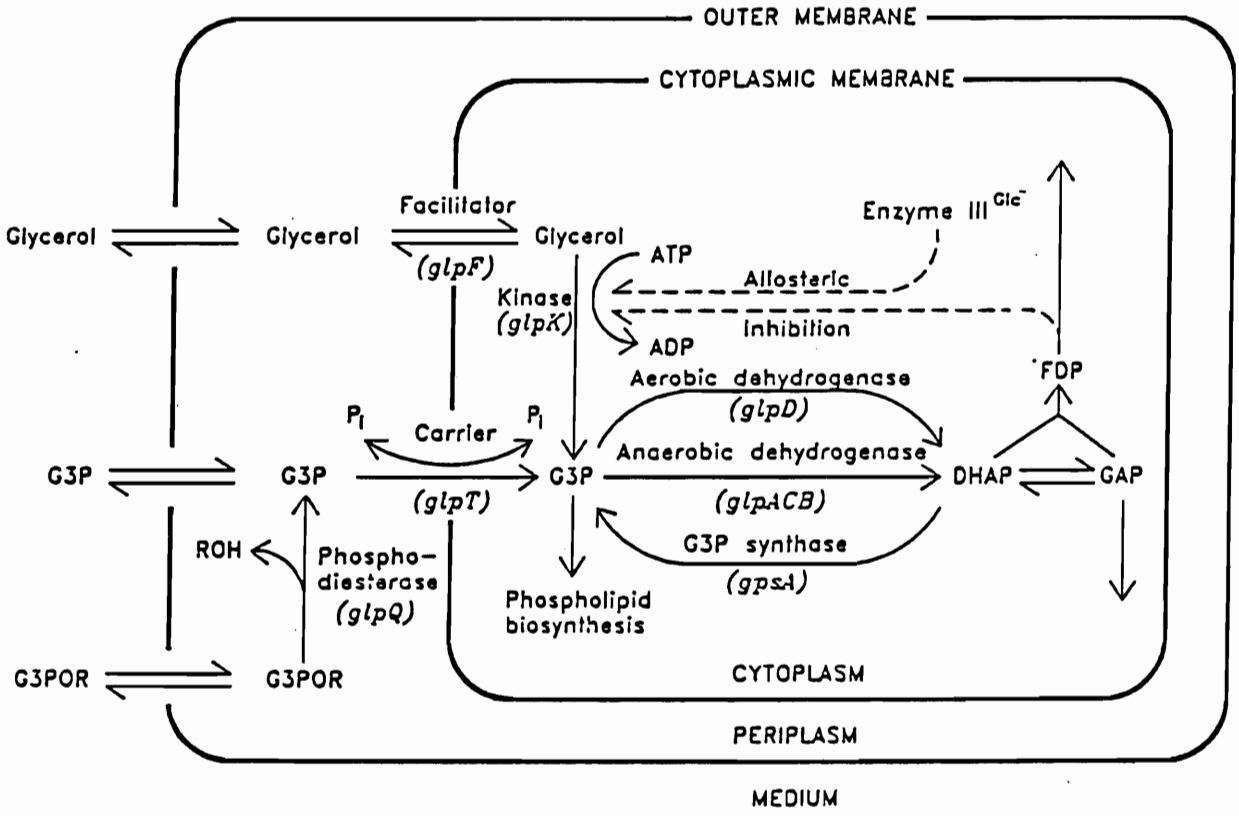


Figure 1. Glycerol-P metabolism in *E. coli*: The gene designations are *glpF*, glycerol diffusion facilitator; *glpK*, glycerol kinase; *glpD*, aerobic glycerol-P dehydrogenase; *glpACB*, anaerobic glycerol-P dehydrogenase; *glpT*, glycerol-P permease; *glpQ*, glycerophosphodiester phosphodiesterase; *gpsA*, glycerol-P synthase. Abbreviations used are: G3P, glycerol-P, G3POR, glycerophosphodiester; DHAP, dihydroxyacetone phosphate; FDP, fructose 1,6-bisphosphate; GAP, D-glyceraldehyde 3-phosphate; Enzyme III^{Glc}, enzyme III of the phophotransferase system specific for glucose.

would occur. On the other hand, when exogenous sources of glycerol-P become plentiful, synthesis of the catabolic enzymes is induced.

Glycerol-P catabolism: The utilization of glycerol-P as an energy source is totally dependent on the glycerol-P dehydrogenases. In *E. coli*, glycerol-P is oxidized to DHAP by a membrane-associated respiratory dehydrogenase that in turn transfers electrons on to ubiquinone (or menaquinone) and utimately to oxygen, nitrate or fumarate (4). Genetic studies have shown that, under aerobic growth conditions in the presence of glycerol, a glycerol-P dehydrogenase is induced which passes reducing equivalents from glycerol-P to an electron transport chain terminating with oxygen as the electron acceptor. Under anaerobic growth conditions in the presence of glycerol and nitrate or fumarate, a distinct glycerol-P dehydrogenase is induced. This enzyme passes the reducing equivalents from glycerol-P to a short electron transport chain terminating with nitrate or fumarate as the ultimate electron acceptor (4). Both respiratory pathways supply the cell with metabolic energy by generating a proton-motive force across the cytoplasmic membrane, which can be coupled to ATP synthesis.

Aerobic glycerol-P dehydrogenase has an apparent *Km* for glycerol-P in the range of 0.8-2 mM (2). The enzyme consists of two identical 58 kDa subunits as estimated by SDS-PAGE (5). The purified protein contains FAD noncovalently bound at the N-terminus. Examination of the nucleotide sequence of the *glpD* gene reveals a protein of 501 amino acid residues with a calculated MW of 56,747 (6). Cells missing aerobic glycerol-P dehydrogenase do not grow on glycerol or glycerol-P in the presence of molecular oxygen (2). This enzyme is

competitively inhibited by DHAP, PEP, phosphoglycolic acid, GAP and D-2- and D-3-phosphoglyceric acid (5). The activity of dehydrogenase is regulated in a complex manner by purine nucleotides. Both ATP and GTP at 10 mM inhibit enzyme activity about 60%. At low concentration (0.5 mM), ATP appears to stimulate activity slightly (5). The gene *glpD*, encoding this catalytically active protein, is transcribed divergently from the repressor gene *glpR* (Fig. 2). Between these two genes are the *glpG* and *glpE* genes, the functions of which are not clear (7). Recent studies have revealed that *glpE, glpG* and *glpR* are organized in one operon (personal communication, D. Austin).

Anaerobic glycerol-P dehydrogenase has an apparent *Km* for glycerol-P in the range of 0.1-0.3 mM (in the presence of flavin mononucleotide as an activator) (8). The enzyme consisits of three subunits with molecular weights of 62,000, 43,000 and 44,000 (9). Mutants missing anaerobic glycerol-P dehydrogenase do not grow anaerobically on either glycerol or glycerol-P, with fumarate as the hydrogen acceptor (10). The promoter proximal *glpAC* genes encode the catalytic subunits of dehydrogenase, whereas *glpB* codes for the membrane-bound subunit (10, 11). GlpB is tightly associated with the envelope fraction and appears to function as the membrane anchor to accept the reducing equivalents from the GlpAC catalytic dimer in the electron transport chain. GlpB contains two clusters of cysteine residues which are typical iron-sulfur centers. Analysis of the GlpB primary structure indicates that the protein lacks extended hydrophobic sequences with the potential to form α-helices but contains several long segments capable of forming transmembranal amphipathic helices (9). The product of the

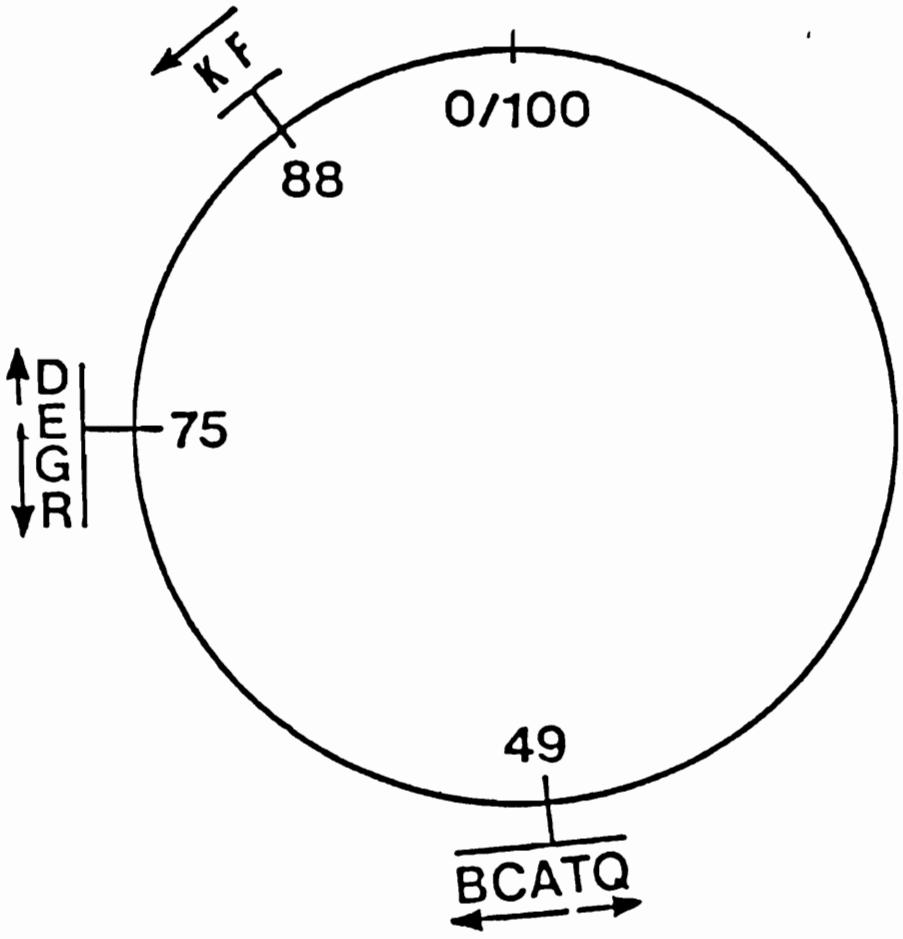


Figure 2. Genetic organization of the *glp* regulon of *E. coli*: The positions on the linkage map and the orientations of transcription are indicated. Gene designations not defined in Figure 1 are *glpR*, repressor for the *glp* regulon; *glpE* and *glpG*, two genes of unknown function.

dehydrogenase reaction, DHAP, is then isomerized to GAP which enters the glycolytic metabolic pathway.

There are several sources of glycerol-P that may be utilized by *E. coli.* Glycerol enters the cytoplasm of the cell via a glycerol diffusion facilitator protein which provides a selective channel with an estimated pore size of 0.4 nm (12). The facilitator catalyzes the energy-independent equilibration of glycerol between extra- and intracellular space (13). This is the only example of transport by facilitated diffusion in *E. coli.* Some nonmetabolizable polyhydric alcohols, such as ribitol, and other small molecules like urea and glycine are also substrates of the glycerol facilitator. Charged molecules such as glycerol-P and DHAP are not substrates (14). The glycerol diffusion facilitator is encoded by the *glpF* gene (15). The predicted MW of GlpF based on the nucleotide sequence is 29,780 (16). Analysis of hydrophobicity distribution of GlpF indicates that it has extremely low polarity and is an integral membrane protein consisting of six potential membrane spanning helices. It may function as a homodimer with the two six-member domains arranged in the membrane to form a channel through which glycerol passes (16).

Cytoplasmic glycerol is phosphorylated to glycerol-P by the action of an ATP-dependent *glpK*-encoded glycerol kinase that can also phosphorylate dihydroxyacetone and L-glyceraldehyde. Glycerol kinase (EC 2.7.1.30; *Km* for glycerol: 10 μM; *Km* for Mg-ATP: 80-100 μM and 400-500 μM) (2) consists of four identical 55 kDa subunits (17). For mutants lacking glycerol kinase, glycerol can neither serve as carbon and energy source nor act as an inducer for the re-

maining proteins of the *glp* system (2). Glycerol kinase is subject to noncompetitive allosteric inhibition by FDP, the key metabolite in glycolysis, and by the nonphosphorylated form of enzyme III^{Glc}, a feature that explains the extreme effectiveness of glucose utilization in preventing overconsumption of glycerol (18, 19). Wild-type glycerol kinase consists of four binding sites for glycerol and four for FDP. The maximum inhibition by FDP is about 80%, with a *Ki* of 0.5 mM (20). The *glpF* and *glpK* genes are present in the same operon (15).

Extracellular glycerophosphodiesters (the deacylation products of all glycerophospholipids) entering the periplasmic compartment are hydrolyzed to glycerol-P plus an alcohol by the action of a periplasmic glycerophosphodiester phosphodiesterase (EC 3.1.4.2) encoded by *glpQ* (21). The native enzyme is a dimer composed of two identical 38 kDa subunits (22). The activity of the phosphodiesterase is calcium-dependent. The secreted enzyme is synthesized as a precursor with a 25 amino acid signal sequence at the N-terminus (23).

Glycerol-P liberated by hydrolysis in the periplasm or derived from the external environment is actively accumulated via the cytoplasmic membrane-associated permease encoded by *glpT* (24). The ability of this transport system to accumulate external glycerol-P without prior hydrolysis is unique. It conserves the energy which otherwise would be required for the ATP-dependent phosphorylation of glycerol (25). The *glpT*-encoded permease mediates transport of glycerol-P ($Km = 12\ \mu M$) (2) through a glycerol-P/Pi antiport mechanism, which mediates the efflux of cytoplasmic Pi in response to externally added Pi or glycerol-P (26, 27). The apparent MW of permease is dependent upon the con-

ditions used for solubilization of the proteins prior to electrophoresis. When samples are heated at 50°C, it displays a MW of 44,000. Following treatment at 95°C, the apparent MW is 33,000 (28). The discrepancy between the observed MW and that deduced from the nucleotide sequence, 50,251 (29), is probably due to the high content of hydrophobic amino acid residues present in permease (25). Mutants defective in this transport system cannot grow on glycerol-P, but can still grow on glycerol or glucose-6-phosphate (30).

The *glp* genes are located at three different positions on the linkage map of *E. coli* (Fig. 2). The *glpTQ* operon (31) is at min 48.6. The *glpACB* operon is located directly adjacent to and transcribed divergently from the *glpTQ* operon. The *glpD* and *glpEGR* operons are transcribed divergently at min 75.3. The *glpFK* operon is transcribed in the counterclockwise direction at min 88.4 (4, 32).

Biosynthesis of glycerol-P: Glycerol-P forms the backbone of all phospholipid molecules and can be synthesized by two metabolic pathways. In the growth medium with glycerol as sole carbon source, glycerol kinase phosphorylates glycerol to glycerol-P. In the absence of exogenous glycerol or glycerol-P, the glycerol-P required for phospholipid synthesis is derived from direct reduction of DHAP with NADH by glycerol-P synthase (EC 1.1.1.94) encoded by *gpsA* (33) (Fig. 1). Mutants missing glycerol-P synthase are auxotrophic for glycerol or glycerol-P. The cellular concentration of glycerol-P is stringently regulated by feedback inhibition of this enzyme by the product, glycerol-P. Thus, the central reactions of glycerol-P dissimilation are very well regulated by the system itself.

Phospholipid biosynthesis: The key phospholipid synthetic intermediate, phosphatidic acid, is formed by the acylation of glycerol-P and then converted to the three major glycerophospholipid classes: phosphatidylethanolamine (75%), phosphatidylglycerol (20%), and cardiolipin (1~5%) (1). The *plsB*-encoded cytoplasmic membrane-associated glycerol-P acyltransferase (*Km* for glycerol-P = 0.15 mM) catalyzes the condensation of glycerol-P and fatty acid thioesters to yield 1-acyl-glycerol-P (34). The second fatty acid is added by 1-acyl-glycerol-P acyltransferase (35). Phosphatidic acid turns over rapidly to CDP-diacylglycerol. The phosphatidyl group of CDP-diacylglycerol is transfered to either serine or glycerol-P by the enzyme phophatidylserine synthetase or phosphatidyl-glycerophosphate synthetase (*Km* for glycerol-P = 0.32 mM). Thus, a glycerol-P concentration of 0.5 mM is needed to allow optimal phospholipid synthesis. This level of glycerol-P should not result in induction of the *glp* regulon, because phospholipid synthesis must continue even when exogenous glycerol-P is absent.

GENETIC REGULATION OF GLYCEROL-P METABOLISM

Expression of the members of the *glp* regulon is subject to three types of regulatory control (36): specific repression by the product of the *glpR* gene; catabolite repression; and respiratory repression. The *glp* operons show different patterns of response to the three control mechanisms (37, 38).

Specific repression: The five operons of the *glp* regulon are coordinately and, at the same time, differentially controlled by the *glp* repressor. Early studies

indicated that expression of *glpD* was the most sensitive to the *glp* repressor; *glpT* less sensitive, and *glpK* the least sensitive to control by *glp* repressor. When the cell is deprived of glycerol-P, synthesis of the dehydrogenase would be repressed before the glycerol-P permease or glycerol kinase. This would prevent unnecessary degradation of glycerol-P needed for phospholipid synthesis. If glycerol or glycerol-P is provided again, higher basal levels of glycerol-P permease and glycerol kinase would facilitate accumulation of the inducer, glycerol-P (39). Repression is relieved to different extents in the presence of glycerol-P, which binds to the *glp* repressor, thereby decreasing its affinity for the operator sites. The affinity of the repressor for glycerol-P is relatively low ($Kd = 31$ μM) (40), compared to the affinities of other repressors, such as *lac*, for inducer ($Kd = 0.6$ μM) (41). Because glycerol-P has dual functions in metabolism, intracellular glycerol-P must be maintained at a level allowing optimal synthesis of phospholipids, without necessarily causing induction of the members of the *glp* regulon.

Catabolite repression: Catabolite repression occurs when glucose is available as an energy source. Glucose is used in preference to other sugars. The presence of glucose in the growth medium results in a decrease in the intracellular levels of cAMP which controls the synthesis of other inducible enzymes subject to catabolite repression (42). The transcription of the genes encoding these enzymes (catabolite sensitive genes) is activated by the cAMP-dependent binding of CRP at specific promoters (43). Expression of all of the *glp* operons is sensitive to catabolite repression by glucose (37). All of the *glp* operons presumably re-

quire CRP-cAMP for maximal expression. The degree of sensitivity to catabolite repression by glucose is reversed when compared to specific repression (36). Expression of *glpK* is the most sensitive, *glpT* less sensitive, and *glpD* the least sensitive. The observed order of sensitivity allows rapid shut down of glycerol kinase and glycerol-P permease when glucose is provided. Exogenous cAMP can overcome catabolite repression completely (44).

Respiratory control: The expression of the *glpD* and the *glpACB* operons is subject to respiratory control (36). Expression of *glpD* is maximal under well-oxygenated growth conditions, and expression of the *glpACB* operon is maximal under anaerobic growth conditions with nitrate or fumarate as electron acceptor (12). The FNR protein (encoded by the *fnr* gene) is structurally homologous to CRP (45). FNR serves as a transcriptional activator for a number of genes, such as the *glpACB* operon, participating in anaerobic respiration (10). Recently, FNR has been shown to exist in two different forms *in vivo,* an aerobic and an anaerobic one (46). Spiro et al. (47) have proposed that the cysteine cluster of FNR at the N-terminus might be a metal binding site. It was suggested that Fe^{2+} is the redox-sensitive group, which binds to FNR during anaerobiosis. The FNR protein then adopts an active conformation responsible for the activation of anaerobic respiratory genes (46, 47). The interconversion of two forms might be a redox process carried out by reversible binding of this divalent metal ion.

The *arcA* and *arcB* gene products mediate anaerobic repression of genes whose products are needed for growth under aerobic conditions, such as the genes encoding the enzymes of the TCA cycle (48). The anaerobic repression of the

glpD operon is 2-fold (49), while the repression for the genes of the TCA cycle varies from 5-fold to 90-fold (50). The *arcA* and *arcB* genes appear to belong to the family of two-component regulatory systems. It seems that ArcA receives the respiratory signal from ArcB, a sensor protein for the redox or energy state, and then represses the expression of *glpD* and other genes. Since the *arc* system includes several enzymes that are induced by different effectors, it appears that the term regulon no longer applies to the *arc* system. The term modulon has been proposed to describe a group of operons and/or regulons that are under the modulation of a common pleiotropic regulatory protein (48). Thus, the *glp* regulon is a member of the Arc, FNR and CRP modulons.

STRUCTURE AND FUNCTION OF TRANSCRIPTIONAL REGULATORY PROTEINS

Regulatory proteins which bind to specific target sites on DNA may activate or inhibit the transcription of the structural genes. These regulatory proteins share common features necessary for carrying out their functions. They are usually oligomers with three domains: the DNA-binding domain, the ligand-binding domain and the oligomerization domain. These domains are localized in separate regions of the protein.

lac repressor: Gilbert and Müller-Hill first purified the *lac* repressor and demonstrated that it was a protein consisting of four identical 38 kDa subunits (51). When the repressor is treated with trypsin, it is cleaved preferentially at

amino acid 59 (long headpiece) and amino acid 51 (short headpiece). Both the short and long headpieces retain the ability to bind to DNA (52); they make the same pattern of contacts achieved by intact repressor but interact more weakly with operator DNA. This suggests that the structure of the headpiece is independent of the rest of the protein and that the DNA binding domain is localized at the N-terminus. The C-terminal domain retains the ability to aggregate into a tetramer and to bind inducer; but it cannot bind operator (53). Fine-structure genetic mapping also proves separate functions for the N- and C-terminal domains of the repressor. The mapping shows that *lacI* mutations occur throughout the gene. Negative dominant ($lacI^{-d}$) mutations are clustered at the extreme 5′-end of the gene (54, 55). These negatively dominant repressor subunits are able to oligomerize with one another and with wild-type repressor subunits but cannot bind to operator and therefore identify the DNA-binding domain of the repressor subunit. Uninducible mutations ($lacI^{s}$) are located beyond residue 62, and tend to occur in two clusters, roughly between positions 80-120 and 190-280. These are the inducer binding sites (55). The $lacI^{-}$ mutations (the monomers cannot aggregate into tetramers) are found between residues 220 and 290 (56), which are responsible for oligomerization. The nine amino acids of the C-terminus (residues 352-360) are involved in the dimer-tetramer equilibrium (57).

gal and deo repressor: The negative regulatory element of the *gal* operon is the Gal repressor. It contains two identical 38 kDa subunits (58). The inducer is D-galactose or D-fucose (59). The GalR monomer contains two domains. The N-terminal domain binds DNA and the C-terminal domain is responsible for

dimerization and inducer contact (59). Binding of the sugar to the C-terminal domain inactivates specific DNA-binding ability of the other domain by allosteric alteration of the protein.

Expression of the majority of genes (*deoCABD*) involved in uptake and catabolism of nucleosides and deoxyribonucleosides in *E. coli* is regulated negatively by DeoR. The inducer is deoxyribose-5-phosphate (60). Similar to LacR and GalR, the DNA binding domain of DeoR is located in the N-terminus of the protein. Distinctively, native *deo* repressor exists as an octamer in solution (61).

Lambda repressor: Since the early 1980's, crystallographic studies have revealed the three-dimensional structures of site-specific DNA-binding proteins such as the Cro and cI repressors of bacteriophage λ (62, 63). The λ repressor monomer is a polypeptide of 27 kDa with two distinct domains (64). The N-terminal domain (residues 1-92) binds specifically to the λ operator sites and mediates both positive and negative control of transcription (65). The C-terminal domain (residues 132-236) is responsible for dimerization, which permits the binding energies of two N-terminal domains to be coupled; thus, the dimer binds several order of magnitude more tightly to the operator than does the monomer (65). The two domains are joined by a connector of 40 residues (64). The N-terminal domain of λ repressor, consists of an N-terminal arm of the first eight residues and five α-helices. This class of repressor, unlike a typical enzyme, binds to operator with a protruding α-helix structure rather than a cleft (54). Two of the helical regions, helices 2 and 3, form the so called 'helix-turn-helix motif', and are responsible for binding DNA. The third helix, the 'recognition helix', fits into

the major groove. Amino acid side chains along the helix are poised to contact the edges of base pairs and make sequence-specific interactions. Helix two of the λ repressor lies across the major groove and makes sequence-independent interactions with the sugar-phosphate backbone (66). Two-fold rotational symmetry of the monomers present in the dimeric λ repressor allows each monomer to contact one half of the palindromic operator site. Since a dimeric repressor is needed for binding to one symmetric operator site, a tetrameric repressor molecule would be needed for simultaneous binding to two operator sites.

CRP protein: The CRP protein is a dimer composed of two identical subunits with 209 AA each (67). Strikingly different from the repressor proteins, the C-terminal domain of CRP binds DNA, while the N-terminal domain binds one cAMP molecule and provides most of the dimer contacts. Crystallographic studies have revealed the structure of CRP (68). The C-terminal domain contains three α-helices D, E, F and two pairs of short antiparallel β-sheet (68). Models for the CRP-DNA complex have been proposed where CRP binds to right-handed B-DNA and uses its F-helix which clearly protrudes from the surface of the CRP dimer to contact the specific major groove (69). The E helix lies across the major groove (helix-turn-helix). Binding of CRP to its recognition site enhances the ability of RNA polymerase holoenzyme to initiate transcription at the promoter. Recent experiments strongly suggest that binding of CRP to DNA causes the DNA to bend (70). It has been suggested the active CRP is only obtained after the cAMP-CRP complex interacts with DNA and this active form has a different configuration from that which is crystallized (43).

glp repressor: The *glpR* gene was cloned downstream from the strong pL promoter of bacteriophage λ. This allowed overproduction of the repressor upon thermal induction of a cryptic λ lysogen harboring the cI857 gene (40). The purified repressor exhibits a subunit MW of 30,000. Under nondenaturing conditions, the *glp* repressor is a tetramer (40). The *glp* repressor negatively regulates the expression of various *glp* operons by interacting with DNA containing *glp* operators.

Despite differences in size, domain organization and tertiary structure, a number of DNA-binding proteins share sequence homologies with λ repressor and CRP in the helix-turn-helix motif. Conserved amino acid residues appear at points where the two helices interact with each other and in the turn which determines the angle between the two helices. Several research groups have predicted that proteins such as *lacR, deoR, galR*, etc., also use the helix-turn-helix motif for DNA binding (71, 72). It is likely that GlpR does too.

STRUCTURES OF PROMOTER AND OPERATOR DNA

One of the landmark events in the history of Molecular Biology was the proposal of the operon model of genetic regulation by Jacob and Monod in 1961, based on their studies of the *lac* genes of *E. coli* and of the genes of bacteriophage lambda (73). The operon is a unit of gene expression, including structural genes and elements that control their expression. The activity of the *lac* operon is controlled by the repressor gene *lacI*, whose product interacts with the control ele-

ments — *lac* operators, and denies RNA polymerase access to *lac*P, and therefore prevents transcription of three genes *lacZYA* from being initiated.

Transcription is one of the principal stages at which gene expression is controlled. Transcription initiation has been proposed as a rather simple model involving the three overall steps shown below (74):

$$R + P \underset{K_B}{\rightleftharpoons} RP_c \xrightarrow[k_f]{} RP_o \xrightarrow{NTP} \rightarrow \rightarrow \ldots \rightarrow RNA$$

1) binding 2) isomerization 3) promoter clearance

This scheme involves the initial binding of RNA polymerase to the promoter with a binding constant K_B, to form an inactive closed complex (RP_c), which subsequently isomerizes with rate constant k_f to form the transcriptionally active open complex (RP_o). The promoter has three components consisting of consensus sequences. The -35 hexamer is separated by 17 bp from the -10 hexamer, which in turn lies 7 bp upstream of the start point. The function of the -35 sequence is to provide the signal for recognition by RNA polymerase (K_B), while the -10 sequence allows the complex to convert from closed to open form (k_f) which facilitates the initiation of transcription.

The relative positions of DNA binding sites within promoter sequences distinguishes activators from repressors. In particular, all of the known activator binding sites are located near or upstream from the -35 region (75), while the repressor binding sites usually overlap the promoter or lie between the promoter and the cluster of structural genes. The operator has a feature common to many

recognition sites: it contains 14 to 20 nucleotides and has an axis of dyad symmetry, which reflects a symmetry in the protein. Most of the regulatory proteins are multimers of identical subunits, each of which must have the same DNA binding site. Each half of the operator is contacted in the same way by an individual repressor subunit.

During early studies, it was commonly believed that repressor binding to an overlapping operator inhibits transcription initiation by hindering the binding of RNA polymerase to the promoter. The following section will discuss the evidence of the multipartite nature of the operators, their locations at sites far away from the promoters, and in some systems where repressor does not inhibit RNA polymerase binding. Based on the evidence, Adhya pointed out the various biochemical stages at which a repressor could act to freeze transcription (76), such as the stage of isomerization or initiation (77), as well as elongation (78). Repressors from different systems may function at different stages (76).

INTERACTION OF REGULATORY PROTEINS WITH DNA

The rate of transcription initiation is determined by interaction of regulatory proteins with specific DNA sites, which affects the activity of RNA polymerase. Regulatory proteins are of two types, a repressor or an activator (76). Some regulators can perform both functions. The λ repressor-operator system is an elegant example of how a single protein uses multipartite, but adjacent operators to achieve negative and positive controls simultaneously.

λ repressor-operator interactions: The λ repressor regulates transcription by differentially occupying two operator regions of the phage chromosome. Each operator contains three contiguous repressor-binding sites. Genetic and biochemical studies have established that λ repressor simutaneously activates P_{RM}, thus stimulates transcription of its own *cI* gene and represses P_R (a promoter needed for lytic development of λ, divergent from P_{RM}) in one physiological condition and allows transcription from P_R but not from P_{RM} in the other (79, 80). The λ repressor binds to the operator in a pairwise cooperative manner. When repressor concentration is low, a dimer binds first to the rightmost and strongest site O_R1. That bound repressor then assists the binding of an additional dimer to the adjacent site O_R2 (65). Occupancy of O_R1, and thus O_R2, physically blocks access of RNA polymerase to P_R, and thus turns off transcription of the lytic genes. At high concentrations, repressor binds to O_R3 which overlaps with the RNA polymerase binding site P_{RM} and prevents transcription of *cI* gene (80). Therefore, the negative control of P_{RM} and P_R by λ repressor has been ascribed to steric interference of RNA polymerase binding to the two promoters. Cooperativity makes the binding of proteins extremely sensitive to small changes in their concentration, thereby enabling genes to switch on and off very efficiently. On the other hand, a different relationship is shown between O_R and the P_{RM} for transcription of *cI*. The RNA polymerase binding site is just about adjacent to O_R2. Occupancy of repressor dimer at O_R2 positions repressor so direct contacts can be made between an N-terminal domain of repressor and RNA

polymerase bound at P_{RM}. This protein-protein interaction activates P_{RM}. A mutant defective both in O_R1 and O_R3 does not affect the activation (81).

lac repressor-operator interactions: The affinity of a repressor for a binding site on DNA is dependent upon the sequence of the operator, the presence of a secondary operator nearby and the topology of the DNA. A fully symmetric 'ideal' *lac* operator binds *lac* repressor *in vitro* 8~10 times tighter than wild-type *lac* operator; however, in *vivo*, the difference in relative affinities is only 1.5 to 3 fold (82). This indicates that optimizing protein binding is not simply a matter of providing symmetric operator sites. Other factors, including the spacing, and hence the relative rotational orientation of half operator sites within the context of the DNA helix, play a significant role (83). The two operator sites need not be directly adjacent to one another to influence repressor binding. The *lac* operon contains a primary operator (O1), and two other regions which bind repressor with lower affinity. One is located 92 bp upstream in the *lacI* gene (O3), the other 401 bp downstream in the *lacZ* gene (O2) (84). In the presence of three operators, expression of β-galactosidase is repressed by a factor of 1,300. The destruction of O1 leads to almost total loss of repression. Inactivation of either O2 or O3 results in a slight decrease of repression. However, the combined loss of both O2 and O3 leads to a 70-fold decrease in repression. Although O1 exhibits its crucial role for cooperative repression, single occupation of just O1, yields an 18-fold repression (85). Full repression is observed only upon cooperative binding between O1 and a remote operator sequence via DNA loop formation.

Loop formation and regulation of the deo operon: Studies of the *deo* system provide evidence of DNA loop formation when DeoR binds to its operator sites. The *deo* operon contains three operators (Oe, O_1 and O_2), in the *deo*P1-P2 regulatory region. DeoR repression of P1 or P2 transcription is weak (2-fold) on a promoter fragment which only contains one operator O_1 or O_2. Full repression by DeoR is 80~100-fold higher on promoter fragments containing three operator sites (86). Upon binding of DeoR to any two operators, the intervening DNA forms a single loop. The repressor octamer may bind to all three operators simultaneously and thus form a double loop structure of the DNA observed by electron microscopy (87). This DNA loop structure is expected to possess an enhanced stability because of the low probability of simultaneous dissociation from both or all three sites. Such structures are also thought to play a role in obtaining full repression of the *galETK* and *araBAD* operons (88, 89).

Effect of the DNA topology and spacing of operators: The topology of the DNA was found to influence formation of looped structures in the gene of the *lac* operon (90). Supercoiling strongly favors interaction between two sites as close together as 99 bp. The distance between two sites could be as far as 5 kb, and still influence repression mediated by a site located in the promoter (91).

The formation of a loop structure is favored and efficient repression is maintained only when an integral number of helical turns of DNA is added or deleted between two operator sites 5~10 helical turns apart, because they are still located on the same face of the helix (92). Introduction of half-integral helical

turns would place the protein on opposite sides of the helix and the energy required both to bend and twist DNA would be prohibitive (93).

Function and specificity of the 'recognition helix': In order to function as specific regulator of gene expression in transcription, a regulatory protein must be able to recognize and to bind preferentially to certain DNA sequences against the background of a large number of competing sites of similar sequence that are also present in the genome. The DNA-recognition helix plays an extremely important role for recognition of primary sequence. Hydrogen bond donors and acceptors are the major recognition elements in DNA-protein interactions. Secondary sequence recognition, including hydrophobic interactions (especially with the methyl group of thymine), the steric fit of DNA groove geometries, and the conformation and flexibility of the DNA helix, also seem to be important parts of the recognition process (94). However, secondary sequence recognitions are not sufficiently limiting by themselves to define a specific recognition sequence for a protein, but serve mostly as binding modulators within the context of a hydrogen bond matrix. α-Helices on the surface of a regulatory protein often have a characteristic pattern of nonpolar residues at positions 4, 5 and 8 of helix-2 and 15, 18 and 19 of helix-3. Because the helical repeat is 3.6 residues/turn, the homologous proteins could form similar bihelical units and have predominantly nonpolar side chains facing the hydrophobic core of the repressor. The amino acid at positions 1-3, 6-7, 11-14 and 16-17 are solvent exposed and hydrophilic, and may be used for DNA recognition (69). Lehming et al. (95) suggest that the recognition helix of *lac* repressor is oriented such that the N-terminus of the re-

cognition helix enters the major groove close to the center of symmetry of the *lac* operator.

A 'helix-swap' experiment was performed to find out if the recognition helix is the major determinant of binding specificity for each of two proteins — bacteriophage 434 repressor (434 R) and P22 repressor (P22 R) (96). A hybrid recognition-helix in 434 repressor was constructed by replacing five amino acids derived from P22 repressor, which presumably face the 'outside' surface of the helix. Its 'inside' surface facing the core was derived from that of 434 repressor:

S N V A I S Q W E R	P22 R
T Q Q S I E Q L E N	434 R
S N V S I S Q L E R	434 $R^{P22\ R}$

The binding specificity of the resulting hybrid repressor (434 $R^{P22\ R}$), as measured *in vivo* and *in vitro*, is that of P22 repressor which donates its exposed surface of the recognition helix (97). It is concluded that the 'outside' surface of the recognition helix is sufficient to distinguish between two different DNA sequences. Therefore, in this case it is the major determinant of its binding specificity.

A similar experiment was done using the CRP and FNR proteins. The amino acid sequences of these two transcriptional activators are homologous, particularly in the DNA-binding domain. The residues of CRP that contact cAMP are not conserved in FNR (45). Based on studies of CRP mutants and model building, three amino acids presumably on the 'outside' surface of the putative DNA-recognition helix of FNR were replaced by those found at the equivalent position in CRP (98):

R E T V G R I L K	CRP
V E T I S R L L G	FNR
R E T I G R L L K	FNR^{CRP}

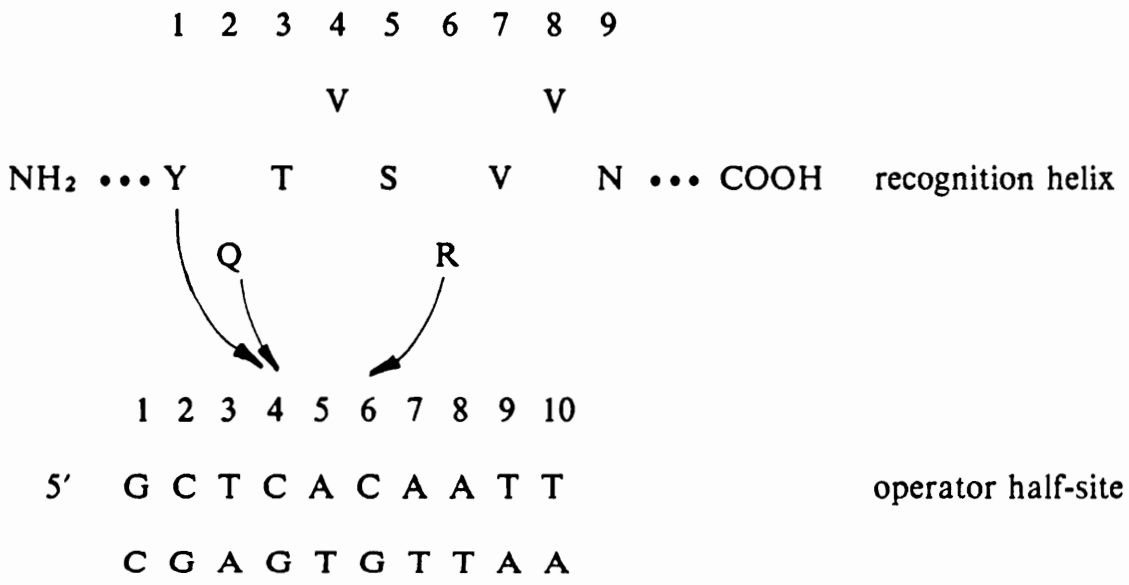
In the presence of altered FNR, expression of CRP-dependent genes is activated during anaerobiosis and is unaffected by glucose. The results indicate that these changes induce a CRP-like activity of FNR interaction with DNA. As a matter of fact, the consensus DNA site for CRP resembles that for FNR (99):

5 5	
AAATTTGATATATATCAAATTT	FNR consensus
AAATGTGATCTAGATCACATTT	CRP consensus

Substitution of G:C by T:A at bp 5 of each of the *lac* DNA half-sites for CRP binding results in FNR-dependent, anaerobiosis-induced transcription of the *lac* promoter (99). It is likely that the overall three-dimensional structure of the specific FNR•DNA complex is extremely similar to that of the specific CRP•DNA complex. Therefore, the specificity of protein-DNA interaction can be exchanged between two homologous and functionally similar proteins by manipulation of the amino acid sequence of the recognition helix or by changing the nucleotide sequence of the site recognized on the DNA.

Müller-Hill et al have devised a systematic method for testing interactions between any repressor and operator. The system consists of two plasmids mutually compatible. One contains a *lac* operon in which the *lac* operator has been replaced by a unique restriction site into which synthetic operators can be cloned (100). The other plasmid carries the *lac* repressor gene which allows the exchange

of specific mutations in a cassette-like manner (101). Repressor-operator interaction is tested by measuring the repression of β-galactosidase in a strain of *E. coli* carrying a deletion of the natural *lac* system. Results from studies where interactions between numerous altered repressor and operator variants were characterized suggest that residues 1 and 2 of the recognition helix interact with bp 4 and to a lesser extent with bp 5 of *lac* operator. Residue 6 of the recognition helix recognizes operator bp 6 (102):



When Y1Q2 of the *lac* repressor recognition helix are changed to V1A2 (GalR-like) or Q1M2 (DeoR-like), the effect of the change of two amino acids is cooperative in recognizing operator variants having sequences closer to the *gal* or *deo* operator (101). The recognition of bp 4 and 5 is totally independent from the recognition of bp 6.

In addition to specific nucleotide sequence recognition, different orientations, different spacing from the center of symmetry and different angles of

helices make the difference in specificity of protein-DNA recognition possible (95).

The objective of this research was to determine the molecular basis for the interaction between *glp* repressor and operator DNA. To achieve this, the structures of the operator and repressor were determined. To determine the structure of the operator, the promoter/operator region of the *glpD* operon was characterized, because it was known that this operon is controlled tightly by repressor. The results of this work are described in part one (below). Part two focuses on determination of the structure of the repressor, and describes results of experiments aimed at definition of the DNA binding domain of the *glp* repressor. The *glp* system is ideal for the study of gene regulation and the interaction of regulatory proteins with DNA. The *glp* regulon contains at least five operons which are under the control of a common inducer glycerol-P and a common regulatory protein, the *glp* repressor. Moreover, the *glp* regulon is regulated in *trans* by multiple regulatory elements. In addition to the specific repression by the *glpR* product, expression of all the structural genes is transcriptionally regulated to varying degrees by the *crp, fnr* and *arc* modulons.

EXPERIMENTAL PROCEDURES

MATERIALS AND METHODS

The DNA sequencing kit was from Stratagene or US Biochemical Corporation. The Klenow fragment of DNA polymerase I, T4 DNA ligase, glycerol-P dehydrogenase, uridine, X-gal and IPTG were purchased from US Biochemical Corporation. M13mp18, M13mp19, DNase I, RNase I and RNase inhibitor were obtained from Boehringer Mannheim Biochemicals. Restriction endonucleases were purchased from New England Biolabs, US Biochemical Corporation and Boehringer Mannheim Biochemicals. Polynucleotide kinase was obtained from New England Biolabs. The Erase-a-Base™ System kit, CIP and reverse transcriptase were supplied by Promega. Tryptone, yeast extract, bacto agar and MacConkey agar were from Difco Laboratories. Deoxynucleoside triphosphates were from Pharmacia. Antibiotics, amino acids, protein molecular weight standards, lysozyme, cAMP, glycerol-P, DAP, ONPG, MOPS and Coomassie brilliant blue were from Sigma Chemical Company. DTNB was from Aldrich. Cesium chloride was from International Biotechnologies, Inc. Agarose, chemicals and carbohydrates were obtained from Fisher Scientific Company. Reagents for polyacrylamide gel electrophoresis were obtained from Bio-Rad.

The oligonucleotide primers specific for sequencing of *glpEGR* and for mutagenesis of the *glp* repressor were synthesized by the phosphoramidite method (103) on an Applied Biosystems 381A Synthesizer and purified by Oligonucleotide Purification Cartriges as described in the manual for the DNA synthesizer (104). The sequences of the synthetic primers are listed in Table 1 and Table 2. New England Nuclear Corp. supplied [α-^{35}S]dATP and [α-^{35}S]dGTP. cAMP receptor protein (CRP) was a gift from Dr. James Harman (105). The *glp* repressor was purified to homogeneity as described previously (40). Pustell Sequence Analysis Programs (Version 4) were purchased from International Biotechnologies, Inc., and were used for DNA sequence analysis.

MICROBIOLOGICAL AND RECOMBINANT DNA TECHNIQUES

Bacterial strains: The genotypes of the *E. coli* K-12 strains used in this study are listed in Table 3.

Media: Various kinds of media were used in this study to grow bacteria or propagate phage, and are listed in Table 4. Proper amounts of antibiotics were supplemented when needed as follows: ampicillin (Ap), 100μg/ml; chloramphenicol (Cm), 30 μg/ml; kanamycin (Km), 50 μg/ml; streptomycin (St), 100 μg/ml; tetracycline (Tc), 12.5 μg/ml. Bacto agar (1.5%) was added to the media when plates were prepared. Glucose, glycerol, maltose or lactose (0.2%) were provided as carbon source depending on the requirement. The lactose

Table 1. Synthetic oligonucleotide primers used for sequencing of the *glpEGR* operon

Number	Nucleotide sequence	Position in sequence
12	TGGCAACGTCAGTTTCC	658 to 674
13	AACCTGGGTGATGATGA	1013 to 1047
14	ATTTTTGCTGCACATAG	1317 to 1300
15	GGCGGCAAAGAAAGGAT	1007 to 991
24	GCCCGCAAAGTGGCGG	1834 to 1849
7	GGCATTCTGGGGATAAG	2092 to 2108
8	GCTGTGCTGATCCTGCA	2330 to 2356
9	TACGGTCGAAGAGATCG	2516 to 2532
10	TGGCAGGAGAGCCGCCC	2591 to 2575
11	CCGTGCAGGATCAGCAC	2359 to 2343
28	GCCGTCGCTATCGATG	2127 to 2112
29	GCGTCGAGCCATTGGG	1873 to 1858
123	GATTCGCTCAATGCGCG	1545 to 1561

Table 2. Synthetic oligonucleotide primers used for site-directed mutagenesis

Number	Nucleotide sequence	Position in sequence
WT	TTCCAGGGATTTATAAATGAAAC	1574 to 1597
23	TTCCAGGAATTCATAAATGAAAC	
	*Eco*RI	
WT	TTCTCCGTCAGCCCGCAGACTATTCGCCG	1675 to 1703
AA	Pro Gln	
16	TTCTCCGTCAGCGAGATGACTATTCGCCG	
AA	Glu Met	
WT	ACTATTCGCCGCGACCTCAATG	1693 to 1714
AA	Arg	
21	ACTATTCGCCAGGACCTCAATG	
AA	Gln	
WT	ACTATTCGCCGCGACCTCAATG	1693 to 1714
AA	Arg	
122	ACTATTCGCGCCGACCTCAATG	
AA	Ala	

Table 3. Bacterial strains used in this study

Strain	Genotype	Source
DH5αF′	*ϕ80dlacZ△M15 endA1 recA1 hsdR17 supE44 thi-1 gyrA relA1 △(lacZYA-argF)U169*	106
JM107	*endA1 gyrA96 thi-1 hsd17 supE44 relA1 △(lac-proAB)* (F′ *traD36 proAB*$^+$ *lacI*^{q} *lacZ△M15*	107
JM109	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA △(lac-proAB)* (F′ *traD36 proAB*$^+$ *lacI*^{q} *lacZ△M15*	107
KH682	*deoR lac ϕ(deoC-lacZ)hyb thi udp upp ton*	Karin Hammer
XL1 Blue	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA* (F′ *proAB*$^+$ *lacI*^{q} Tn*10 lacZ△M15*)	Stratagene
RZ1032	HfrK116 PO/45 [*lysA*(61-62)] *dut1 ung1 thi1 relA1 zbd-279*::Tn*10 supE44*	108
MC4100	F$^-$ *araD139 △(argF-lac)U169 rpsL150 rbsR deoC1 relA1 thiA1 ptsF25 flbB5301*	109
GD1	MC4100 *glpR2 zih-730*::Tn*10*	3
GD2	MC4100 *glpR2 ϕ(glpK-lacZ)hyb* λ*plac*Mu *zih-730*::Tn*10*	3
GD6	MC4100 *glpR2 ϕ(glpD-lacZ)hyb* λ*plac*Mu	7
GD31	MC4100 *ϕ(glpK-lacZ)hyb* λ*plac*Mu	P1(GD2)→MC4100 Lac$^+$ Tet^s selection
SH305	MC4100 *△glpD102 recA1 srl*::Tn*10*	3
SH309	MC4100 *△(malA-asd)3 ugpA704*::Tn*10*	7
SY101	MC4100 *glpR2 ϕ(glpD-lacZ)hyb* (F′ *proAB*$^+$ *lacI*^{q} Tn*10 lacZ△M15*)	GD6 × XL1 Blue
SY102	MC4100 *glpR2 ϕ(glpD-lacZ)hyb* λ*plac*Mu *recA1 srl*::Tn*10*	P1(SH305)→GD6
TL73	MC4100 *glpR2 recA1*	28

Table 4. Media used in this study

Type	Components/liter	Usage/Comment	Reference
LB	10 g Tryptone 5 g yeast extract 10 g NaCl	grow cells for storage and enzyme assays; adjust pH to 7.5	110
TB	12 g Tryptone 24 g yeast extract 0.17 M KH_2PO_4 0.72 M K_2HPO_4 4 ml glycerol	amplify plasmids	111
2 × YT	16 g Tryptone 10 g yeast extract 5 g NaCl	grow phages for isolation of DNA; adjust pH to 7.4	113
M9	6 g Na_2HPO_4 3 g KH_2PO_4 1 g NH_4Cl 0.5 g NaCl 2 mM $MgSO_4$ 0.1 mM $CaCl_2$ 0.2% glucose	minimal medium for selection and for growth of plasmids; adjust pH to 7.4	110
A + B	A: 3 g KH_2PO_4 6 g Na_2HPO_4 2 g $(NH_4)_2SO_4$ 3 g NaCl B: 2 mM $MgCl_2$ 75 μM $CaCl_2$ 2 μM $FeCl_3$	minimal medium for β-galactosidase activity assay; adjust pH to 7.2	112
H-top	10 g tryptone 8 g NaCl 8 g agar	preparation of M13 phage	Amersham
R-top	10 g tryptone 8 g NaCl 1 g yeast extract 8 g agar 2 ml 1 M $CaCl_2$ 5 ml 20% glucose	preparation of P1 lysate	113

phenotype of strains carrying *glp-lac* fusions was determined with MacConkey-lactose medium (113) or on various minimal media or LB medium containing 40 μg/ml of X-gal.

Growth and maintenance of cells: *E. coli* strains were initially streaked on appropriate media and their phenotypes were tested. A single colony was used to inoculate 2 ml of LB or corresponding medium, and grown overnight at 37°C on a roller drum. For propagation of M13, a single plaque was inoculated into LB with DH5αF′ and grown for six hours. Aliquots of overnight cultures were stored at -70°C in the presence of 15% glycerol (110). Phage supernates and *E. coli* cells on solid media were stored at 4°C.

Isolation of DNA: Plasmid DNA was isolated from 1.5 ml overnight cultures by a rapid alkaline detergent lysis method (114). Closed circular DNA for progressive unidirectional deletions was purified by conventional equilibrium sedimentation in CsCl-EtBr gradients (110). For sequencing reactions and for preparation of radiolabeled DNA, purification followed the procedure described by (115). Single-stranded DNA was isolated by using a protocol outlined by BRL (116). DNA yield could be quantified either by reading the absorbance at 260 nm or by analysis on an agarose gel in the presence of EtBr.

Agarose gel eletrophoresis: For DNA restriction enzyme digestion analysis, plasmid DNA was analyzed on a 0.6~2% agarose gel at constant voltage

(10-12V/cm) at room temperature. Low melting agarose gels at 4°C were used for eluting DNA fragments (117). TBE buffer (0.05 M Tris-HCl, 0.05 M boric acid, 0.002 M EDTA, pH 8.0) was used in both cases with 0.5 μg/ml of EtBr.

Dephosphorylation of DNA with calf intestinal alkaline phosphatase (CIP): The terminal 5′ phosphates were removed from DNA by treatment with CIP (110). The DNA was digested to completion with restriction enzyme, extracted once with phenol/chloroform and precipitated with ethanol. Then the DNA pellet was resuspended in a minimum volume of 10 mM Tris-HCl (pH 8.0) and incubated with CIP (0.01 unit/μg DNA) in CIP buffer (50 mM Tris-HCl, pH 9.0, 1.0 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 1.0 mM spermidine) in a volume of 50 μl at 37°C for 30 minutes. A second aliquot of CIP was then added and incubated for 30 minutes for 5′ protruding termini. With blunt ends or recessed 5′ termini of DNA, incubation was at 37°C and 56°C repectively. The reaction was stopped by adding 40 μl of H_2O, 10 μl of 10 × STE buffer (100 mM Tris-HCl, pH 8.0, 1 M NaCl, 10 mM EDTA), and 5μl of 10% SDS. The sample was heated to 68°C for 15 minutes, followed by extraction with phenol/chloroform (1:1), chloroform/isoamyl alcohol (24:1), and precipitation with ethanol. The DNA pellet was resuspended in a minimum volume of water and used for ligation.

Construction of recombinant plasmids: Plasmid DNA was digested with restriction endonucleases using conditions recommended by their suppliers. The DNA fragments were then eluted from a low melting agarose gel. The vector

DNA fragments were usually dephosphorylated with CIP before ligation. In some cases, the 5′ or 3′ overhang DNA generated by restriction enzymes was converted into blunt ended DNA, which could be used for ligation. After complete digestion of DNA, the Klenow fragment of DNA polymerase I (1 U/μg DNA) and dNTP mixture (0.25 mM) were added to the digestion mixture. The filling-in reaction was performed at 37°C for 15 minutes and stopped by heating at 75°C for 10 minutes. After electrophoresis on an agarose gel to remove extra DNA fragments, the fragment for ligation was eluted from the gel. Ligation with T4 DNA ligase was carried out by manufacturer's guidelines.

Construction of progressive unidirectional deletions for DNA sequencing: A 2,723 bp *Bgl*II-*Bgl*II DNA fragment of pSH79 containing part of the *glpD* gene and the *glpEGR* operon was cloned into M13mp18 DNA which was cleaved by *Bam*HI. After transfection to DH5αF′, two different oriented template DNA's were selected by the C-test (118). Progressive unidirectional deletions were generated based on the procedure developed by Henikoff (119). Each of the DNA's was digested with *Pst*I and *Xba*I to produce a 4-base 3′ and a 5′ overhang respectively. ExoIII digests of the double-cut DNA proceeded synchronously from the 5′-protruding end into the insert. Aliquots were removed at intervals and digested with S1 nuclease to remove single stranded DNA after ExoIII digestion. Then Klenow fragment of DNA polymerase I was added to fill in the ends, which were ligated to circularize the deleted molecules. Transfection of DH5αF′ fol-

lowed and subclones from each time point were screened to select appropriate intervals between deletions for DNA sequencing.

Transformation: Transformation was done by using a modified Maniatis et al. method (110). *E. coli* cells were made competent as follows. A fresh overnight culture (1 ml) was inoculated into 100 ml of LB at 37°C and allowed to grow to an $OD_{600} \simeq 0.6 \sim 0.8$. Cells were chilled on ice thoroughly for 15 minutes, centrifuged, and resuspended in 50 ml of cold 0.1M $MgCl_2$. The suspension was incubated on ice for 15 minutes, and centrifuged again. The cells were resuspended in 25 ml of pre-chilled T-salts (75 mM $CaCl_2$, 6 mM $MgCl_2$) and incubated on ice for 20~30 minutes. After centrifugation, the cell pellet was resuspended in 3 ml of cold T-salts. Plasmid was mixed with *E. coli* competent cells on ice for 45 minutes and the mixture was subjected to a heat shock at 42°C for 2 minutes. 1 ml LB was added to the cells and incubated at 37°C for 1 hour before plating.

Sequencing of DNA: Double-stranded plasmid and single-stranded phage DNA were used as templates for sequencing reactions carried out by the dideoxy-chain termination technique (120). Sequencing reactions were analyzed on 6% wedge-shaped polyacrylamide gels. Gels were pre-electrophoresed for 15 to 30 minutes before loading the samples. Electrophoresis was carried out at constant temperature (~50°C) by application of constant power (~60 W). To maximize the length of sequence determined from each clone, two loadings were necessary.

One loading was electrophoresed for 4½ hours and the other for 1¾ hours. In this way, up to 400 nucleotides from the primer were determined.

Band-shift electrophoresis: Interaction of the *glp* repressor with radiolabeled DNA restriction fragments was assessed using the methods developed by Garner and Revzin (121), and Fried and Crothers (122). For these experiments appropriate DNA fragments containing the *glpD* and *deo* operator regions were cloned into pGEM 3Z. The fragments were end-labeled by filling recessed 3′ ends using the Klenow fragment of DNA polymerase I or Sequenase and [α-^{35}S]dATP or [α-^{35}S]dGTP. A radiolabeled vector DNA fragment was used as control. To assess binding of the *glp* repressor to the DNA fragments, the proposed operator DNA (0.7~1.5 nM) and carrier salmon sperm DNA (6 ng) were incubated together at room temperature in buffer A (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 50 mM KCl, 5% glycerol, 0.025% Triton X-100) for 10 minutes. Repressor protein was then added and incubation was continued at 30°C in a final volume of 7.5 μl. After 30 minutes, 18.7 μl of buffer A, 100 μg/ml BSA, and 0.025% bromphenol blue were added, and the sample were applied immediately to a 6% polyacrylamide gel containing 0.07% bisacrylamide. Electrophoresis was carried out at room temperature for 2-3 hours at 1.5 mA/cm of gel using TBE buffer.

DNase I footprinting: DNase I footprinting was done by a combination of previously described methods (123, 124). A DNA fragment containing the *glpD* operator was end labeled with [α-^{35}S]dGTP and the Klenow fragment of DNA

polymerase I at a *Not*I site present in the vector (located near the *Bgl*II site). A second digestion with *Asu*II and subsequent purification by agarose gel electrophoresis resulted in a 233 bp fragment labeled 20 and 21 bp downstream of the *Bgl*II cleavage site. The labeled DNA fragment was diluted to 1.0 to 1.5 nM in 0.1 ml of buffer B (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 50 mM KCl, 4 mM $MgCl_2$, 2 mM $CaCl_2$, 5% glycerol, 0.025% Triton X-100), 100 μg/ml of BSA, and 5 μg/ml of sonicated salmon sperm DNA. Various concentrations of purified *glp* repressor (with or without inducer, glycerol-P) or CRP (with or without cAMP) were added. The reactions were incubated for 15 minutes at 37 °C. DNase I (0.1 ng) was then added, and the incubation continued for 3.5 minutes. DNA was precipitated by the addition of 10 μg of sonicated salmon sperm DNA, 25 μl of 7.5 M ammonium acetate, and 2.5 volumes etnanol to each reaction. The precipitated DNA was suspended in 4 μl TE/formamide, heated for 3 minutes, and then loaded onto a sequencing gel. The position of the protected regions was located by co-electrophoresis of sequencing reactions as size markers.

Isolation of RNA: To determine the transcription start site for the *glpD* gene, total mRNA was isolated from two strains isogenic except for *glpD* [SH309(pSH56, $glpD^+$) and [SH309(pACYC177, *glpD*)] by the method of Bialkowska-Hobrzanska et al. (125). Cells were grown in 20 ml LB to log phase with 100 μg/ml Ap, 0.4% glucose and 100 μg/ml DAP. To stop growth, 5 ml of 2 M NaN_3 was added. The culture was quickly poured over frozen, crushed

buffer (0.15 M NaCl, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA). The mixture was centrifuged at 7,000 rpm (Beckman JA-20 rotor) for 10 minutes at 4°C and cells were suspended in 2 ml of TE-NaCl buffer (10 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 mM NaCl). An equal volume of TE-NaCl buffer (~100°C) containing 1% SDS was added to the suspension and the incubation continued for 5 minutes. The mixture was cooled to 60°C and an equal volume of phenol equilibrated with 50 mM sodium acetate (pH 5.5) at 60°C was used for extraction. The aqueous phase was extracted twice with $CHCl_3$-isoamyl alcohol (24:1). The RNA was then precipitated with two volumes ethanol at -20°C. The precipitate was collected by centrifugation at 12,500 rpm (Beckman JA-20 rotor) for 10 minutes at 4°C and resuspended in 150 μl TE buffer. To check the quality of the RNA preparation, an aliquot (3 μg) was treated with DNase I (15 ng) or RNase I (27 ng) for 15 minutes at 37°C. Buffer B was used for DNase I digestion and buffer C (50 mM Tris-HCl, pH 7.4, 1 mM EDTA) was used for RNase I digestion. The samples were then run on an agarose gel. The DNase I digestion was scaled up to 40-50 μg. The digestion was stopped by adding an equal volume of STES buffer (100 mM Tris-HCl, pH 7.4, 10 mM EDTA, pH 8.0, 1 M sodium acetate, pH 5.2) followed by extraction with phenol/$CHCl_3$ and ethanol precipitation. The precipitated RNA was collected by centrifugation at 14,000 rpm (Beckman JA-20 rotor) for 30 minutes, and resuspended in 3.5 μl TE buffer. RNase-free DNase was prepared by preincubation of DNase I (1 mg/ml) in 20 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, and 3.5 mM $CaCl_2$ for 5 minutes at 37°C (126). To this solution was added ⅛ volume of proteinase K (10 mg/ml), and the

enzymes were incubated at 37°C for 30 minutes. The freshly treated DNase I was used immediately to digest DNA present in RNA preparations.

Primer extension analysis: Primer extension assays were adapted from the procedure of Alam et al. (127). A 10 ng amount of synthetic oligonucleotide primer was added to 20 μg of RNA in 4 μl. The tubes were boiled for 2 minutes, followed by quick cooling in ice water. The hybridized primer was extended under the following conditions: 34 mM Tris-HCl (pH 8.3), 50 mM NaCl, 6 mM $MgCl_2$, 5 mM DTT, 200 μM each dCTP, dGTP, and dTTP, 3.5 μM [α-^{35}S]dATP (181 Ci/mmol), 2,500 U/ml of RNasin, and 1,500 U/ml of reverse transcriptase in a total volume of 10 μl. The reaction was incubated at 42°C for 30 minutes, after which 5 μl of a chase solution containing 800 μM dNTP was added. The mixture was incubated for 30 minutes at 30°C. The reactions were extracted with phenol and chloroform and then precipitated with ethanol. The pellets were suspended in 80% formamide and analyzed on a sequencing gel. The position of the transcription start point was determined by running sequencing reactions that used the same oligonucleotide as the sequencing primer.

Site-directed mutagenesis: Site-directed mutagenesis was carried out by the method described by Kunkel et al (108). Uracil-containing M13 DNA templates were isolated from intact M13 phage grown on *E. coli* strain RZ1032. An overnight subculture of RZ1032 was diluted 50-fold into 10 ml LB and grown with vigorous shaking at 37°C. After 1.5 hours, 0.25 μg/ml uridine and template

phage stock containing the *glpR* gene or the *glpEGR* operon were added, and shaken continuously for six hours. The culture was centrifuged and the clear supernate contained the phage at about 10^{10}~10^{11} pfu/ml. Before preparing viral template DNA, the phage titers were compared on *ung*$^-$ (RZ1032) and *ung*$^+$ (DH5αF′) hosts. When the survival ratio in these two strains reached 1 × 10^{-5} or less, the template DNA was isolated as described above for single-stranded DNA. The 5′-hydroxyl group of the synthetic oligonucleotide (100 to 150 ng) was phosphorylated (for subsequent ligation) in a 10~20 μl reaction volume containing 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 1 mM ATP, 2 U of T4 polynucleotide kinase. The reaction was incubated at 37°C for 60 minutes and stopped by adding 3 μl 100 mM EDTA and heating at 70°C for 10 minutes. Annealing was carried out with 500 to 600 ng template DNA and 2~4 ng phosphorylated oligonucleotide in a 10 μl reaction volume (40 mM Tris-HCl, pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl). The mixture was heated at 65°C for 5 minutes and cooled slowly to 35°C, left at room temperature for 15 minutes and then on ice for 10 minutes. Polymerization was carried out by adding 2.5 μl dNTP's (0.5 mM), 1.5 μl 10 mM ATP, 1 U Sequenase, 2 U T4 DNA ligase. Incubation was at 37°C for 15 minutes. The reaction mixture was heated at 65°C for 10 minutes, cooled, and then used for transformation of strain DH5αF′. The sequence of the mutagenized regions and the entire *glpR* gene were verified by sequencing.

Site-directed mutagenesis was carried out along with some unexpected spontaneous mutations in the *glpEGR* operon, possibly caused by sequence

homology between the synthetic oligonucleotides used and the template DNA. The DNA fragment with mutant forms of the *glp* repressor containing two unique sites *Nsi*I (400 bp upstream from the *glpR* gene) and *Eco*RV (293 bp downstream from the repressor translation start codon) was cloned into plasmid pSH79 containing the wild-type *glpR* gene. The *Nsi*I and *Eco*RV DNA fragment allowed introduction of specific mutations into wild-type repressor in a cassette-like manner without incurring unexpected mutations on the rest of the DNA. The newly constructed mutations were verified by sequencing the 700 bp *Nsi*I-*Eco*RV DNA fragment and used for β-galactosidase activity assays.

Conjugation: Conjugation was carried out by the method of Miller (113). The donor XL1 Blue (F′ *lacI*^{q} Tn*10* St^s Tc^r) was subcultured from a fresh overnight culture until the density was 2-3 × 10^8 cells/ml (OD ≃ 0.25). An overnight culture of recipient GD6 [*glpR φ(glpD-lacZ)hyb*, St^r Tc^s] was diluted 1:20. A mating mixture was prepared with 0.5 ml of each culture in a test tube and placed on a 30 rpm roller at 37°C for 60 minutes. A loopful of the mating mixture was directly streaked onto glucose minimal selection plate with St (100 μg/ml) and Tc (5 μg/ml). A loopful of either XL1 Blue or GD6 alone was plated on the same selection plates as control. Since all of the donor cells were unable to grow in the presence of St while the recipient cells were unable to grow in the presence of Tc, only those GD6 cells which have received the F′ *lacI*^{q} Tn::*10* Tc^r from XL1 Blue formed colonies on the selection plate. The exconjugate strain SY101 [F′ *lacI*^{q} Tn::*10 glpR2 φ(glpD-lacZ)hyb* St^r Tc^r] was constructed in this way (Table 3).

Transduction: Transduction was adapted from the procedures of Silhavy et al. and Miller (128, 113). One drop (30 μl) of overnight SH305 (*recA1 srl*::Tn*10*) was subcultured in 3 ml of LB containing 5×10^{-3} M $CaCl_2$ and 10 μg/ml of Tc, and allowed to grow to a density of 2×10^8 cells/ml ($OD_{600} \simeq 0.25$). $P1_{vir}$ (10^7 pfu) was added to this culture and incubated for 20 minutes at 37°C. 2.5 ml R-top agar was added immediately and poured onto a freshly made R plate. After 8 hours incubation at 37 °C, the soft agar was scraped into a corex tube and 120 μl of chloroform was added and mixed vigorously. After 10 minutes incubation at room temperature, the cell debris and agar were removed by centrifugation. The supernate (P1 lysate) was stored in the refrigerator. A single colony of recipient strain GD6 (*recA*$^+$) was inoculated into 5 ml of LB and grown at 37°C overnight. The cells were centrifuged and resuspended in 2.5 ml 10 mM $MgSO_4$, 5 mM $CaCl_2$. Various amounts of P1 lysate were added to GD6 cells in a series of test tubes. The mixtures were incubated at 30°C for 30 minute. Then 100 μl of 1 M sodium citrate and 3 ml of H-top agar were added to each tube. The mixture was plated on LB-Tc plates containing 20 mM sodium citrate. The next day, different sizes of transductants were streaked on duplicate LB-Tc-sodium citrate plates. One of the plates was exposed to UV light, using GD6 and SH305 as controls. The plates were wrapped with aluminum foil and incubated at 37°C for 10 hours. The colonies sensitive to UV light were tested with UV light again. These strains were the newly constructed strain SY102 [*recA1 ϕ(glpD-lacZ)*] (Table 3).

Construction of expression vector pSY223 and derivatives with the* glpR *gene: A synthetic oligonucleotide containing a multiple cloning site was inserted between the *Pst*I and *Hin*dIII sites of the expression vector pKK223-3 (129). The resulting plasmid was named pDA223 (personal communication, D. Austin). A 1,300 bp *Eco*RI-*Hin*dIII fragment from pMJR1560 (130) containing the *lacI*^{q} gene was blunt-ended with dNTP's and Klenow, and then cloned into the *Nru*I site of pDA223 (Fig. 3). After transformation of strain DH5αF′ to Ap resisitance, the counterclockwise orientation of the *lacI*^{q} gene in the expression vector pSY223 was verified by *Eco*RV, *Pvu*II and *Bst*E II restriction digests.

The repressor genes were cloned into pSY223 vector in two ways. First, pSY223 was partially digested with *Eco*RI and heated at 70°C for 10 minutes followed by a second digestion with *Stu*I and treatment with CIP. A 1,312 bp *Eco*RI-*Sma*I DNA fragment containing wild-type or mutant forms of the *glpR* gene from mp18-44 was cloned into this 5.9 kb partially digested vector. Alternatively, a 2,700 bp *Bgl*II-*Bgl*II DNA fragment containing the *glpEGR* operon with wild-type or mutant forms of the *glpR* gene was cloned into the *Bgl*II site of pSY223. In both of the cases, the ligation mixture was used for transformation of strain SY102 or DH5αF′ by selection for Ap resistance. In the case of SY102, transformants were tested on MacConkey-Ap plates. The pink-colored transformants were the clones containing the wild type *glpR* repressor; cells harboring the vector or mutant forms of the repressor were dark red.

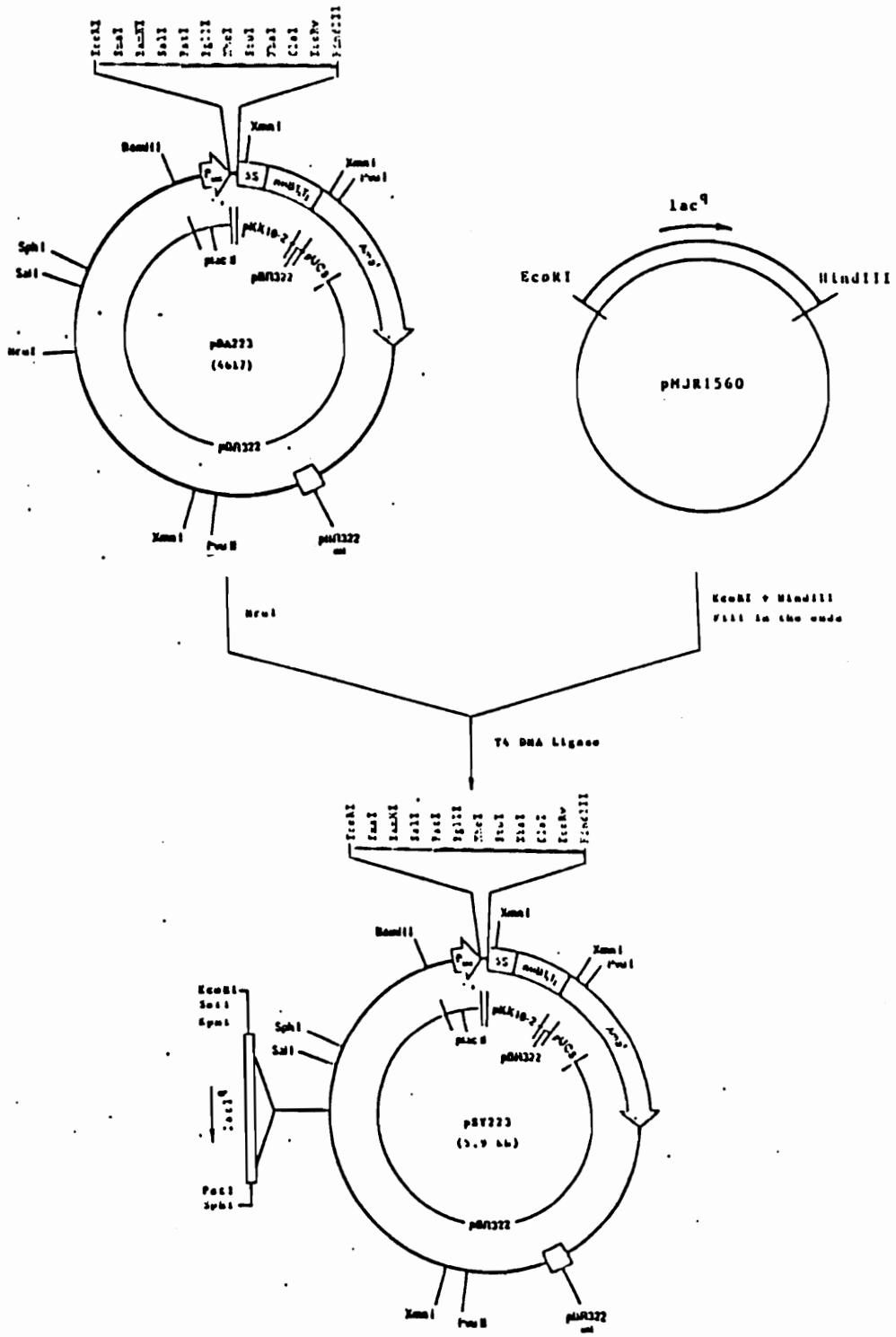


Figure 3. Construction of pSY223: A 1,300 bp *Eco*RI-*Hin*dIII fragment containing the *lacI*^{q} gene from plasmid pMJR1560 was blunt-ended and cloned into the *Nru*I site of plasmid pDA223. The counterclockwise orientation of the *lacI*^{q} gene in the expression vector pSY223 was verified by endonuclease restriction digestion.

BIOCHEMICAL TECHNIQUES

Overproduction of the* glp *repressor: SY102 harboring pSY102-6 (wt), pSY102-11 (R38Q) or pSY102-3 (P33E,Q34M), where the various *glpR* genes are positioned downstream from the *tac* promoter in pSY223, was grown in LB-Ap medium without IPTG or with 500 μM IPTG for 2 hours at 37°C. SY102 (pSY223) was grown under the same condition and served as control. An aliquote (25 μg protein) of each sample was applied to a 10% SDS-PAGE. DNA from the overexpressed clones was then isolated and transformed into translational fusion strains SY102 and KH682.

Protein gel electrophoresis: Protein samples were analyzed on 10% SDS-PAGE using the buffer system of Laemmli (131) at a constant voltage (~200V) for 3-4 hours. Protein bands were stained with Coomassie Brilliant Blue G-250 and gels were destained in a solution of methanol (30%) and acetic acid (10%).

Determination of protein concentration: Protein concentrations were determined spectrophotometrically by the method of Bradford (132) using BSA as the standard.

Preparation of [2-^{3}H] glycerol-P: A 1.65 ml phosphorylation reaction mixture containing 4 mM [2-^{3}H] glycerol (200 mCi/mmol), 50 mM Tris-HCl (pH 8.0), 10 mM ATP, 20 mM $MgCl_2$, 20 mM β-mercaptoethanol, 1 mg/ml serum albumin

and 10 μg/ml *E. coli* glycerol kinase was prepared. The mixture was incubated at 37°C for 3 hours and then heated in a boiling water bath for 5 minutes and centrifuged for 10 minutes. The supernate was applied to a Dowex-1 (formate) column (1 × 5 cm) and eluted with 15 ml distilled water. Unreacted [2-^{3}H] glycerol was not absorbed on the column and appeared in the water wash. The column was eluted with 2 N formic acid and the eluent was collected in 2-ml fractions. Tubes containing radioactivity were pooled, evaporated to dryness, and the residue dissolved in 1.625 ml H_2O. Radioactivity was determined by liquid scintillation counting.

The concentration of glycerol-P was determined by measuring the formation of NADH from NAD during the oxidation of glycerol-P to DHAP by rabbit muscle glycerol-P dehydrogenase. The assay followed the description of Beisenherz et al. (133) as follows: In the reaction mixture, 0.97 ml 1 M hydrazine-glycine buffer (2.43 ml hydrazine hydrate, 0.75 g glycine, adjust pH to 9.0 with 1 ml 6 N HCl in a final volume of 50 ml), 0.01 ml 25 mM NAD, 0.01 ml glycerol-P dehydrogenase (4 U/μl) were mixed as blank, and equilibrated at 30°C for 10 minutes. Then various amounts of [^{3}H] glycerol-P were added and the absorbance at 340 nm was measured ($\simeq$ 0.5-1.0). After the reaction was complete, the molar concentration of glycerol-P was calculated based upon the molar extinction coefficient of NAD (6.3×10^3 M^{-1}). The specific radioactivity of glycerol-P was 82 cpm/pmol.

Assay of* glp *repressor: The *glp* repressor was monitored using a modification of the ligand-binding assay of the catabolite gene activator protein of *E. coli* (40). Assays (in 1.5 ml microcentrifuge tubes) contained 0.05 M MOPS (pH 6.5), 0.4 mg/ml casein, 50 μM [2-^{3}H] glycerol-P (82 cpm/pmol), and repressor in 0.1 ml. After a 5 min incubation at RT, 0.35 ml of ice-cold, saturated ammonium sulfate was added; the tubes were vortexed and centrifuged for 5 min at 4°C. The supernate was aspirated, and the pellet was washed with 0.5 ml of cold ammonium sulfate. The final pellet was resuspended in 0.1 ml of water, and the radioactivity was determined by adding 1 ml of scintillation fluid. The assay was linear from 0~2 μg of repressor. One unit of repressor activity is defined as the amount of repressor required to bind 1 pmol of glycerol-P.

Partial purification of* glp *repressor: Based on the purification scheme of Larson et al (40), the following purification procedure for the wild-type and mutant forms of the repressor expressed in pSY223 was developed. All steps were carried out at 0-4°C. *E. coli* cells (5 g) were suspended in 40 ml of Buffer A (0.02 M Tris-HCl, pH 7.5, 0.4 M NaCl, 0.1 mM EDTA, 0.2 mM DTT, 0.2 mM PMSF) and then disrupted by a single passage through a French pressure cell. Cell debris and inclusion bodies were removed by ultracentrifugation at 20,000 rpm (Beckman TY65 rotor) for 60 minutes. The supernate was diluted with Buffer A to 7 mg/ml (crude extract).

Assay of β-galactosidase activity: The activity of β-D-galactosidase was determined spectrophotometrically by measuring the absorption at 420 nm of the yellow colored product (o-nitrophenol) from the hydrolysis of the substrate ONPG (113). A single colony was inoculated in A and B salts minimal medium supplemented with 0.2~0.4% of the specified carbon source and 2 μg of thiamine per ml. The overnight cultures were subcultured for 2 to 3 hours to an OD_{600} of 0.28~0.70. Then 100 μg/ml of Cm was added and the cells were cooled on ice. After 20 minutes, the cultures were assayed in Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH 7.0) in a final volume of 1 ml. Two drops chloroform and one drop 0.1% SDS were added and vortexed for 10 minutes to disrupt the cell membrane partially, allowing ONPG to diffuse into the cells. The mixtures were incubated at 28°C for 5 minutes. Reactions were started by adding 0.2 ml of ONPG (4 mg/ml in 0.1 M phosphate buffer, pH 7.0) and stopped by adding 0.5 ml of 1 M Na_2CO_3 after the yellow color developed. Reaction time was recorded. The cell debris was removed by centrifugation. The absorbance at 420 nm was measured versus a blank assay containing no cells. Miller units (113) were calculated from the formula:

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

t = time of the reaction in minutes

v = volume of culture used in the assay in milliliters

These units were proportional to the increase in o-nitrophenol per minute per bacterium.

PART I: CHARACTERIZATION OF THE PROMOTER AND OPERATOR OF THE GLPD GENE

RESULTS

In order to characterize the structure of the promoter and operator for the *glpD* gene, the nucleotide sequence of the *glpD* control region and transcription start site for the *glpD* have been determined. The *glpD* operator sequences and CRP binding site have been identified.

Nucleotide sequence of the* glpD *control region: Previous studies (134) have indicated that the promoter for the *glpD* gene is located just to the left of the *Bgl*II restriction site and is transcribed toward the *Eco*RI site (Fig. 4). To obtain detailed information about the structure of the promoter and operator for the *glpD* gene, the nucleotide sequence of a 690 bp DNA fragment between an upstream *Rsa*I site and the *Eco*RI site was determined by using the strategy outlined in the legend to Fig. 4. Plasmids constructed for subcloning are listed in Table 5. Located just upstream from the *Bgl*II site is a translational start codon which precedes an open reading frame that continues for 61 codons to the end of the DNA that was sequenced. This presumably encodes the amino-terminal portion of the aerobic glycerol-P dehydrogenase. Recently, experimental evidence has demonstrated that the proposed start codon is correct (6). It is the only translation initiation codon in this region that is preceded by a good ribosome-binding site. Also, it is located downstream from sequences analogous to those found in

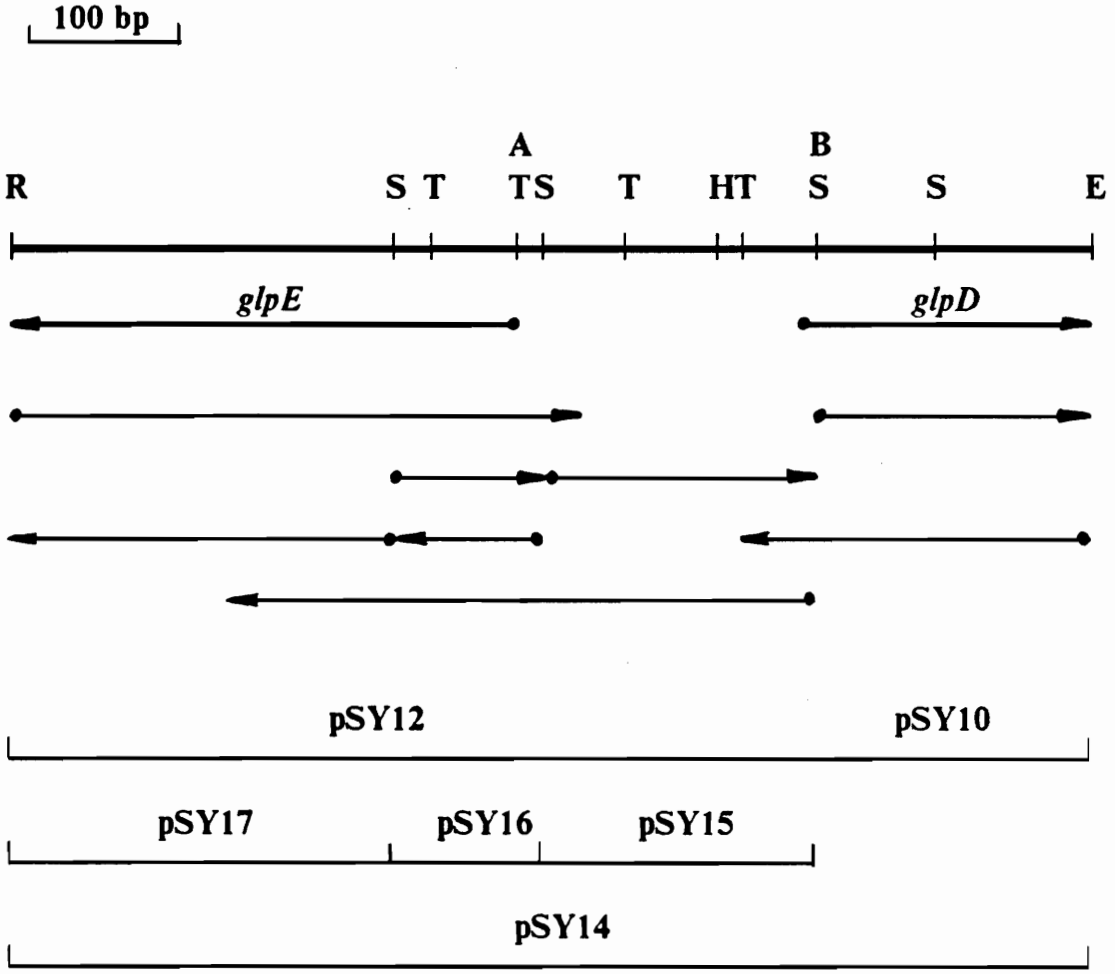


Figure 4. Strategy for sequencing the control region of the *glpD* gene: The extent of the sequence information obtained from the plasmids (indicated on the lower part of the diagram) is indicated by the arrows. Restriction endonuclease cleavage sites: A, *Asu*II; B, *Bgl*II; E, *Eco*RI; H, *Hin*fI; R, *Rsa*I; S, *Sau*3A; T, *Taq*I.

Table 5. Construction of plasmids used in this study

Plasmid	Insert (bp)	Source of insert	Vector[1] sites used
pSH79	*Eco*RI-*Sal*I (3,000)	pSH21	*Eco*RI-*Sal*I
pSY10	*Eco*RI-*Bgl*II (166)	pSH79	*Eco*RI-*Bam*HI
pSY12	*Bgl*II-*Rsa*I (517)	pSH79	*Bam*HI-*Eco*RV
pSY14	*Eco*RI-*Rsa*I (683)	pSH79	*Eco*RI-*Eco*RV
pSY15	*Xba*I-*Sau*3A (224)	pSY12	*Xba*I-*Bam*HI
pSY16	*Sau*3A-*Sau*3A (78)	pSY12	*Bam*HI-*Bam*HI
pSY17	*Sau*3A-*Hin*dIII (235)	pSY12	*Bam*HI-*Hin*dIII

[1]The vector in each case was Bluescript KS M13+ (Stratagene).

catabolic operons in *E. coli*. These include a typical -10 sequence, a -35 sequence, and a sequence similar to the consensus sequence for binding of the CRP (134). These sites are labeled in Fig. 5.

Transcription start site for the* glpD *gene: To prove that the promoter elements discussed above are in fact those employed in vivo, the 5′-end of the *glpD* mRNA was mapped by primer extension analysis. Total mRNA was isolated from two strains isogenic except for *glpD*. A 20-mer oligonucleotide probe complementary to bases 517 to 536 (Fig. 5) was hybridized to the RNA preparations and extended in the presence of reverse transcriptase. The same oligonucleotide was used as a primer in standard sequencing reactions with pSY14 as the template; the latter reactions served as standards. The major product of the primer extension reaction with RNA from the *glpD*$^+$ strain (Fig. 6, lane 1) was a polynucleotide terminating at position 466 (Fig. 5). Thus, the end of the mRNA begins 5′-AAC--. Some minor products of shorter length were also observed. These may be the result of premature termination of transcription by reverse transcriptase caused by secodary structure of the mRNA in this region, which is expected because of the regions of dyad symmetry present in the *glpD* operator (see below). No products were observed when RNA from the *glpD* deletion strain was employed as the template (Fig. 6, lane 2). The observed transcription start site is located at the expected position relative to the proposed -10 and -35 sequences shown in Fig. 5. Therefore, it is likely that these sequences constitute the *glpD* promoter.

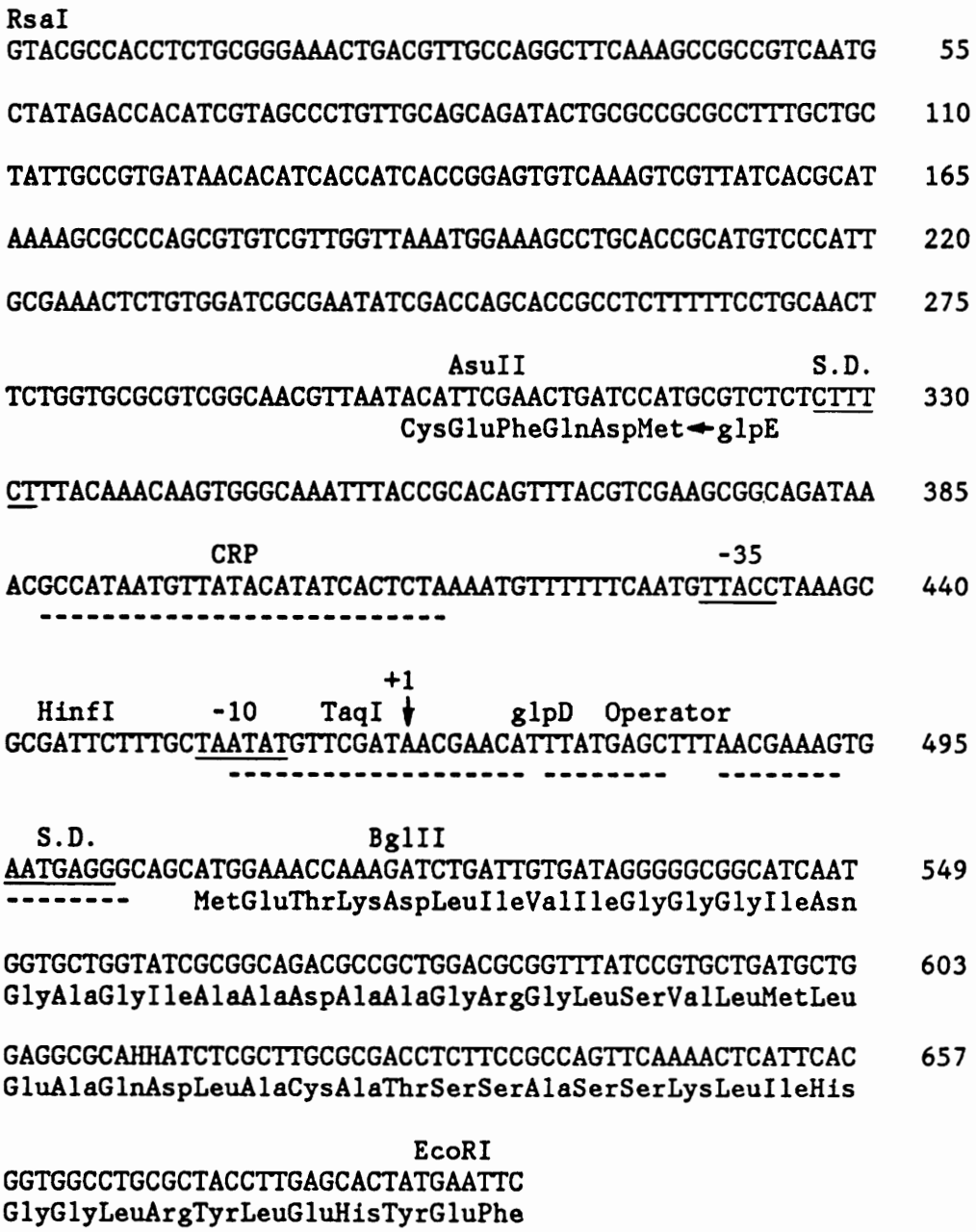


Figure 5. Nucleotide sequence of the control region of the *glpD* gene: The restriction sites referred to in the text are indicated. The 5′-end of the *glpD* transcript is indicated by the arrow labeled +1. Sequences resembling the consensus sequences for ribosome binding (S.D.) and sequences -10 and -35 are underlined. Dashed underlining indicates the regions protected from DNase I digestion by binding of CRP or the *glp* repressor. The proposed amino acid sequence of the amino-terminal part of the glycerol-P dehydrogenase is shown. The proposed translation start codon of *glpE* is indicated.

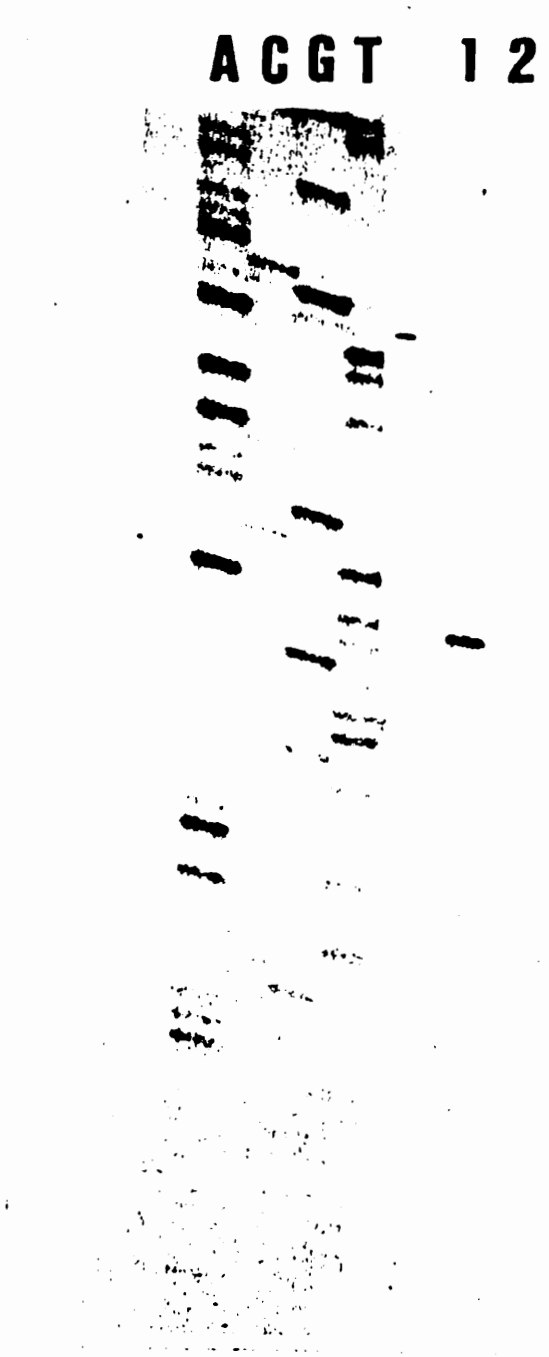


Figure 6. Identification of the start site for transcription of the *glpD* gene: Primer extension assays were performed with a 20-nucleotide primer specific for the 5'-end of the *glpD* gene with total RNA from strain SI1309(pSI156) (lane 1) or from strain SI1309(pACYC177) (lane 2) serving as the template. The sequence ladder (lanes A, C, G, and T; generated by using the same primer) was coelectrophoresed with the products of the primer extension reactions.

Identification of the* glpD *operator: Prokaryotic operators are in general palindromic sequences contained in or near elements of the promoter which they control. Several approaches were used to localize the *glpD* operator. First, band-shift electrophoresis assays (40) employing purified *glp* repressor and restriction fragments end labeled at the *Eco*RI site (Fig. 5) were used to assess the distance between the *Eco*RI site and the repressor binding site. Binding of the repressor caused retardation of the mobility of the *Rsa*I- *Eco*RI (Fig. 7, lane 2), *Asu*II-*Eco*RI (lane 4), and *Hin*fI-*Eco*RI (lane 6) restriction fragments. The presence of the repressor had no significant effect on the mobility of the *Taq*I-*Eco*RI (lane 8) or *Bgl*II-*Eco*RI (lane 10) restriction fagments. The results indicate that the operator must be located between the *Hin*fI and *Eco*RI restriction sites and is not located between the *Taq*I site shown in Fig. 5 and the *Eco*RI site. A palindromic sequence was found which began between the *Hin*fI and *Taq*I sites (nucleotides 456 to 475). With the exception of two positions, this 20-bp sequence is a perfect palindrome (see Fig. 8). If this is the binding site for the *glp* repressor, the presence of repressor bound at this site should prevent digestion by *Taq*I, which has a cleavage site in this palindrome (Fig. 5). To test this possibility, DNA end labeled at the *Eco*RI site was treated with *Taq*I in the presence or absence of *glp* repressor. The presence of the *glp* repressor decreased cleavage at this site but did not affect cleavage at other sites (assessed by PAGE followed by autoradiography; data not shown). The results suggest that this palindromic sequence is the operator for the *glpD* gene.

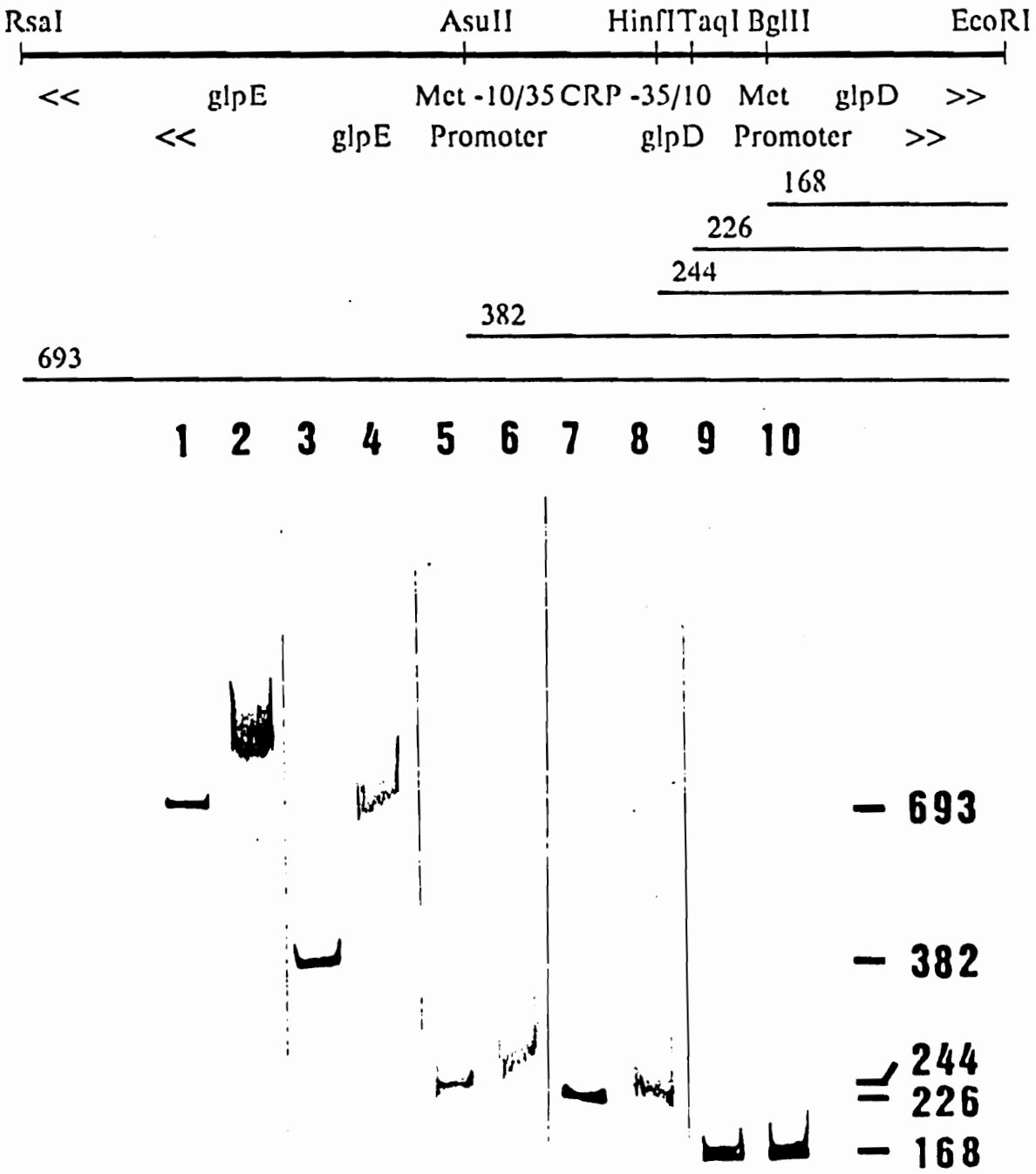


Figure 7. Localization of the binding site for the *glp* repressor by band shift electrophoresis: The 690-bp *Rsa*I-*Eco*RI fragment shown in Fig. 5 was labeled at the *Eco*RI end with [α-^{35}S]dATP and the Klenow fragment, subsequently digested with *Asu*II (lanes 3 and 4), *Hin*fI (lanes 5 and 6), *Taq*I (lanes 7 and 8), or *Bgl*II (lanes 9 and 10). Samples run in the even-numbered lanes had 50 nM repressor tetramers added before electrophoresis. The size of the fragments are indicated to the right.

Operator 1:	TATGTTCGAT	AACGAACATT	(456-475)
	ATACAAGCTA	TTGCTTGTAA	
Operator 2:	TATGAGCTTT	AACGAAAGTG	(476-495)
	ATACTCGAAA	TTGCTTTCAC	

O1	(top strand):	TATGTTCGAT
O1	(bottom strand):	AATGTTCGTT
O2	(top strand):	TATGAGCTTT
O2	(bottom strand):	CACTTTCGTT
Consensus sequence:		TATGTTCGTT

Figure 8. ***glpD* operators and the consensus sequence for repressor binding:** Both strands of each operator are shown in the upper part of the figure, with the centers of symmetry indicated by the asterisks.

To identify more precisely the repressor-binding site, DNase I footprinting analysis was performed. The results (Fig. 9A) revealed a region of approximately 40 bp that was protected from DNase I digestion in the presence of the *glp* repressor (labeled O1 and O2). Protection did not occur when the inducer, glycerol-P, was added to the reactions (Fig. 9A, lanes 9 through 11). The protected region included not only the palindrome mentioned above (O1) but also extended downstream for an additional 20 bp (O2). The regions protected are indicated by dashed underlining in Fig. 5. Closer scrutiny of the downstream sequence revealed that it is similar to the first palindrome:

O1 TATGTTCGAT*AACGAACATT (nucleotides 456 through 475)

O2 TATGAGCTTT*AACGAAAGTG (nucleotides 476 through 495)

Thus, the operator may be tandemly repeated. The centers of symmetry are indicated by asterisks (*).

To determine whether the repressor binds independently to each of the two putative operators, the footprinting reactions were carried out starting at very low repressor concentrations. Protection of the operator region did not occur until the repressor concentration was increased from 0.08 nM to 0.5 nM (Fig. 9A, lanes 3 and 4), and then there was no apparent preferential binding to either of the two proposed operators (O1 and O2; Fig. 9A). Therefore, the repressor binds to a relatively large region emcompassing 40 bp, or individual repressor subunits may bind cooperatively to the two proposed operators. At 400 nM repressor, protection of the entire fragment from DNase I digestion was observed (Fig. 9A, lane 8).

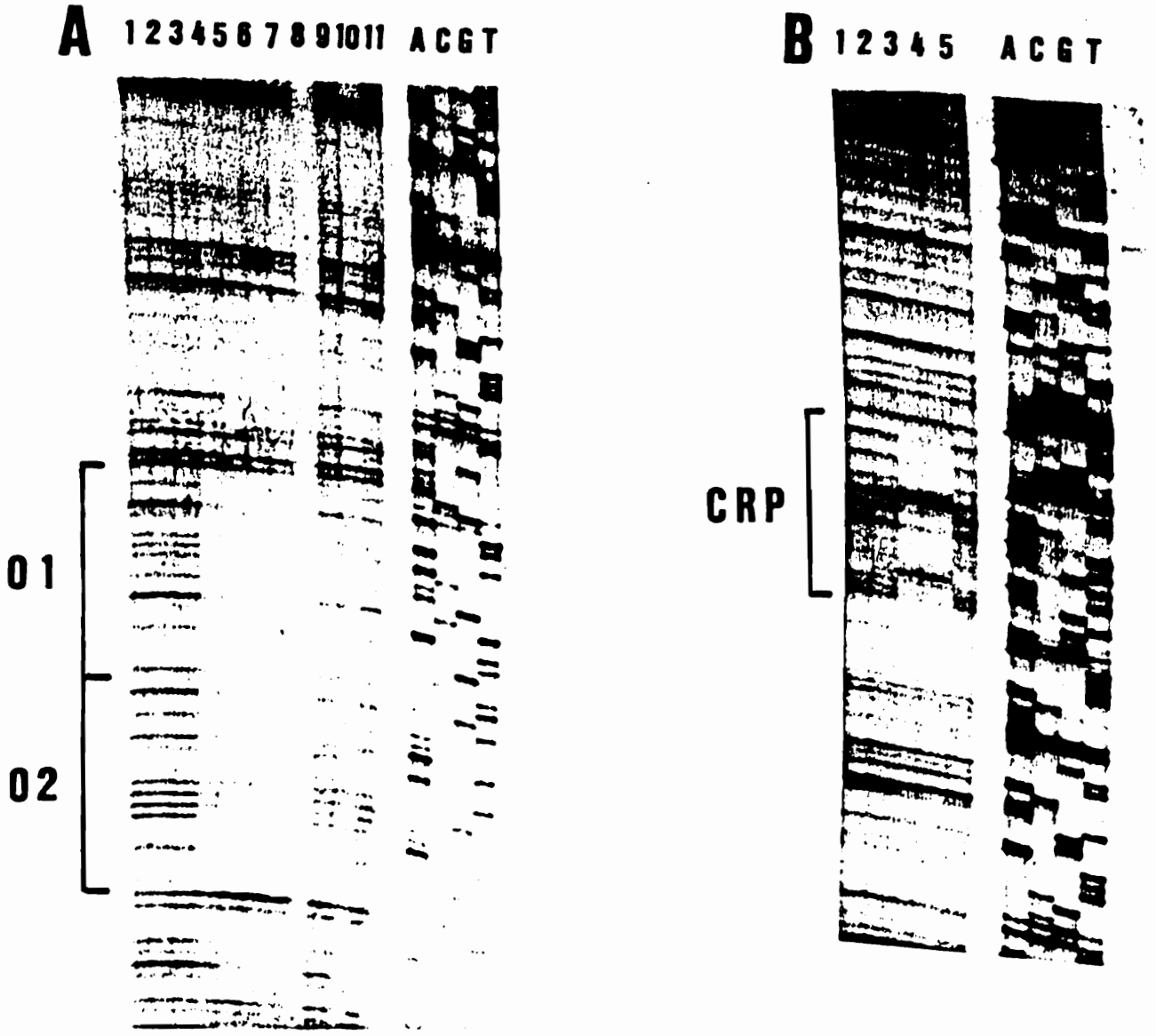


Figure 9. Identification of the binding sites for the *glp* repressor and CRP in the control region of the *glpD* gene by using DNase I footprinting: Positions of the regions protected by *glp* repressor (A) or by CRP (B) were determined by coelectrophoresis of DNase I footprinting reaction with the sequence standard (Lanes A, C, G, and T). (A) Reactions were carried out using the following concentrations of *glp* repressor tetramers from lanes 1 to 8 (nM): 0, 0.02, 0.08, 4, 5, 20, 100, 400; 9 through 11, 5 nM repressor and 0.5, 2 and 10 mM glycerol-P, respectively. (B) Reactions were carried using the following concentrations of CRP dimers from lanes 1 to 5 (nM): 0, 14, 70, 140, 70. Reactions run in lanes 1 through 4 also contained 0.25 mM cAMP, and the reaction run in lane 5 contained no cAMP.

There is some indication that the protected region extends further downstream to include the Shine-Dalgarno sequence (nucleotides 496 through 503; Fig. 9A and Fig. 5). This sequence displays some similarity to the first half of the O1 palindrome. At present, it is not clear whether protection of this region represents binding of additional repressor subunits to an operator half-site by virtue of its proximity to the other two proposed operators or whether repressor bound at the other site(s) results in protection of this region due to the size of the repressor.

Identification of the binding site for the CRP-cAMP complex: Expression of *glpD* is partially sensitive to catabolic repression. The level of expression of the *glpD* gene decreases by a factor of 2 when glucose is added to the growth medium (2). Therefore, it is possible that the CRP-cAMP complex is a positive activator of transcription of the *glpD* gene. Scrutiny of the sequence upstream from the -35 sequence revealed the presence of a sequence (nucleotides 392 through 413, Fig. 5) that is similar to the consensus sequence (135) for binding of CRP:

AA- TGTGA - - - - - - TCACA -TT (consensus)

TAATGTTATACATATCACTCTA (nucleotides 392 through 413)

To determine whether the CRP-cAMP complex binds to this region, a DNase I footprinting experiment with purified CRP and end-labeled DNA was performed (Fig. 9B). Indeed, the region containing nucleotides 388 through 413 was protected in the presence of 70 or 140 nM CRP plus cAMP (Fig. 9B, lanes 3 and 4; indicated by dashed underlining in Fig. 5). No protection was observed

when cAMP was omitted from the reactions (Fig. 9B, lane 5). Thus, it is likely that binding of the CRP-cAMP complex at this site activates transcription of the *glpD* gene.

DISCUSSION

The *glpD* operator is the first among the members of the *glp* regulon to be characterized. Binding of the *glp* repressor to DNA from the control region of the *glpD* gene protected a relatively large region from digestion by DNase I. The protected region contained a tandemly repeated 20 bp element possessing hyphenated dyad symmetry. The repressor-binding site overlapped the -10 region of the *glpD* promoter. Therefore, binding of repressor to the *glpD* operator may effectively prevent binding of RNA polymerase to the promoter and inhibit initiation of transcription of the *glpD* gene.

The apparent presence of two operators in the *glpD* promoter allowed formulation of a consensus sequence for binding of the *glp* repressor from the four half sites, all read starting at the 5′ end (Fig. 8). The consensus sequence for the half site is TATGTTCGTT. Bases at positions 2, 7, and 10 of the half site are invariant among the four half sites, and those at all other positions except the first are the same in three of four half sites. Each of the symmetrical 20 bp operator sites presumably binds two symmetrically arranged repressor subunits, which may be part of a dimeric or tetrameric repressor. The *glp* repressor is a tetramer in solution at high protein concentrations under nondenaturing conditions (40). Although it is not clear whether the repressor is tetrameric when diluted to the concentrations used for the present study, it is likely that a minimum of a dimeric

repressor binds to each 20 bp operator sequence, if binding of *glp* repressor to DNA is analogous to binding of other repressors to operators which possess dyad symmetry (65).

Tandemly repeated operators are found in the control regions of several well-characterized bacterial and phage transcription units. In the case of phage lambda, three tandemly repeated operators control transcription initiation at p_L and p_R (65). Cooperative binding of lambda repressor dimers at adjacent sites helps to ensure tight repression of the two promoters (65). The LexA repressor also exhibits cooperative binding to the tandemly repeated operators present in the control regions for the colicin E1 gene (136) and the *lexA* gene (137). Tandemly repeated operators are also found in the control regions for several members of the *arg* (138, 139) and *tyr* (140) regulons.

Results of experiments employing a strain with a thermolabile *glp* repressor indicated that the *glpD* gene was the most sensitive to control by the *glp* repressor (2). It is possible that the tandemly repeated operator plays a role in the relatively tight control of expression of the *glpD* gene. Results obtained in the DNase I footprinting experiments are consistent with the idea that cooperative binding of the *glp* repressor to the two operators occurs. Simutaneous binding to both O1 and O2 occured as the concentration of repressor was increased (Fig. 9A). Results obtained with band shift eletrophoresis indicated that the *glp* repressor bound with higher affinity to DNA fragments containing both O1 and O2 (Fig. 7, lanes 1 through 6) when compared with a DNA fragment containing only O2 and part of O1 (Fig. 7, lanes 7 and 8). In the latter case, there

was only a slight indication that the repressor interacted with the *Taq*I-*Eco*RI restriction fragment. It should be pointed out, however, that the half sites present in O2 each deviate from the consensus sequence for repressor binding at three positions, and therefore O2 may be expected to bind repressor less avidly than does O1 (each half site of O1 deviates from the consensus sequence at only one position) or the two sites combined. It may be of interest to note that the first six bases present in the bottom strand of O1 (AATGTT; Fig. 8) are found three times just upstream from the *glpD* operator (nucleotides 393 through 398, 415 through 420, and 426 through 431; Fig. 5). No evidence for binding of repressor to these regions was obtained in the footprinting experiments; if, however, the repressor has a higher affinity for these partial half sites than for general sites on DNA, these partial sites could help localize *glp* repressor on DNA very near the *glpD* operator. Further experiments are required to find out whether the presence of tandem operators or the partial operator sequences upstream from O1 and O2 is the reason for the highest apparent affinity of the *glp* repressor for the *glpD* control region relative to the other operons controlled by the *glp* repressor. Relative tight control of expression of the gene encoding glycerol-P dehydrogenase is desirable so that when exogenous supplies of glycerol-P become depleted, endogenously synthesized glycerol-P may be channeled into synthesis of the phopholipids (1, 2).

DNase I footprinting experiments were also performed in our laboratory to locate the positions of the operator in the *glpTQ-glpACB,* and *glpFK* control regions. One binding site for *glp* repressor was found to coincide with the CRP

site in the *glpTQ* promoter (141) and four other sites were found in the *glpACB* (141) and *glpFK* (16) promoters. Comparison of the sequences protected by *glp* repressor with the operators found in *glpD* allowed formulation of a consensus operator (Table 6). Like other binding sites for regulatory proteins in prokaryotes, it is a palindromic sequence possessing hyphenated dyad symmetry. With the exception of *glpTQ*, control by tandemly repeated operators is a common feature of negative regulation in the *glp* regulon. Cooperative binding of repressor to tandemly repeated operator sites (with the exception of *glpTQ*) may facilitate tight control of transcription of the *glpD*, *glpACB* and *glpFK* operons.

Lin reported that expression of *glpTQ* is more responsive to repressor than *glpFK* (2, 12). If this is the case, even in the absence of a close match of the *glpT* operator with the consensus and the lack of tandem operator (Table 6), then one or more of the *glpA* operators might interact with the *glpT* operator. Recent studies revealed possible loop formation between the *glpT* and *glpA3* or *glpA4* operators (141). Therefore, the *glpT* operator does not function alone, but requires cooperation from a suitably oriented *glpA* operator, which is mediated by tetrameric *glp* repressor to form a stable repression loop (141).

Repressor-mediated DNA loop formation has also been implicated in the control of the *glpFK* operon (36). Four and two repressor binding sites were found in the *glpFK* promoter and *glpK* gene, respectively. In this case, the two internal operators in the *glpK* coding region had a synergistic effect on affinity of repressor for the *glpFK* promoter region, and thus are likely to play a role in the control of expression of the *glpFK* operon (16). Possibly, they function as

Table 6. Comparison of the *glp* operators

Operator	Position[1]	Sequence		% Match with Consensus
glpD1	-10	TATGTTCGAT	AACGAACATT	100
glpD2	+11	TATGAGCTTT	AACGAAAGTG	65
glpA1	-60	AATGTTCAAA	ATGACGCATG	55
glpA2	-28	ACTTTCGAAT	TATGAGCGAA	35
glpA3	-8	TATGCGCGAA	ATCAAACAAT	70
glpA4	+33	AATGGTAAAA	AACGAACTTC	65
glpT	-51	TGTGTGCGGC	AATTCACATT	65
glpF1	-89	ATGGCGCGAT	AACGCTCATT	65
glpF2	-68	TATGACGAGG	CACACACATT	55
glpF3	-47	TAAGTTCGAT	ATTTCTCGTT	65
glpF4	-27	TTTGCTCGTT	AACGATAAGT	70
glpK1	+993	CGCGGTCGTA	ATGGATCACG	45
glpK2	+1035	GCAGCGCGAA	TTTGAGCAAA	40
Consensus		TATGTTCGAT	AACGAACATT	(100)

[1] The number refers to the position of the first base listed relative to the start point of transcription. All operators are written in the same direction as the direction of transcription of the indicated gene.

enhancer-like cis-acting transcriptional regulatory elements in eukaryotes, or tetrameric repressor interacts simutaneously with the operator sites of *glpK* and *glpF* with the intervening DNA forming a loop.

The binding site for the CRP-cAMP complex in the *glpD* control region was identified during the course of this work. It is located two bp further upstream relative to the -35 region when compared with the CRP-binding site in the *lac* operon (135). The CRP-binding site in the *glpD* promoter does not correspond as closely to the consensus sequence (135) for CRP binding when compared with the analogous site in other catabolic operons such as *lac*:

AA- TGTGA - - - - - - TCACA- TT (consensus)

TAATGTTATACATATCACTCTA (*glpD*; 10 of 14 matches)

TAATGTGAGTTAGCTCACTCAT (*lac*; 11 of 14 matches)

The underlined positions are those which match the consensus sequence.

One of the differences between the site in *glpD* and the consensus (and *lac*) is the T at position 7 of the site in *glpD*. In the case of the CRP site in the *lac* promoter, the G at position 7 and the C located symmetrically at position 16 are very important for binding of CRP. Glutamic acid residues (Glu-181) of the CRP dimer are known to make direct contacts with these bases (135). A G-to-T substitution at position 7 of the site in *lac* decreases the affinity of CRP by a factor of 29 (135). Therefore, it is likely that the *glpD* promoter binds CRP with relatively low affinity and thus may be less dependent on or responsive to CRP-cAMP than is the *lac* promoter. This may help to explain why the expression of the *glpD* gene is not particularly sensitive to the addition of glucose

to the growth medium (2). Expression of *glpD* may be more dependent upon other elements in the promoter relative to its dependence upon CRP-cAMP.

During the course of this work, the sequences and positions of the CRP sites for the other *glp* operons were determined by other members of the lab. Also, the relative sensitivity of the *glp* operons to catabolite repression was found to be *glpTQ* > *glpFK* = *glpACB* > *glpD*. This order is similar but not identical to that found by Freedberg and Lin (36). Comparing the CRP-binding sites of the *glp* operons (Table 7), it seems that there are two factors which affect the sensitivity of the operons to catabolite repression. One is the degree of similarity to the consensus sequence; the other is the position in the promoter sequence. It would appear that a CRP binding site overlapping the -35 sequence (as in the *glpTQ, glpACB* and *glpFK2* promoters) renders these operons more sensitive to catabolite repression than *glpD* (Table 7). Moreover, the CRP binding site in the *glpTQ* promoter corresponds closer to the consensus sequence than *glpFK*, and so *glpTQ* is more sensitive to catabolite repression. Cooperative binding of CRP to tandemly repeated DNA sites in the *glpFK* promoter may be the reason why the *glpFK* operon is more sensitive to catabolite repression than is the *glpD* operon.

Table 7. Comparison of the CRP binding sites of *glp* operons

Operator	Position[1]	Sequence	% Match with Consensus
glpTQ	-51 to -30	ATGTGTGCGGCAATTCACATTT	86
glpD	-74 to -53	TAATGTTATACATATCACTCTA	71
glpACB	-51 to -30	AATGACGCATGAAATCACGTTT	64
glpFK1	-71 to -50	TTTTATGACGAGGCACACACAT	64
glpFK2	-48 to -27	TTAAGTTCGATATTTCTCGTTT	50
		or	
	-46 to -25	AAGTTCGATATTTCTCGTTTTT	64
Consensus sequence		AA-TGTGA--------TCACA-TT	

[1] The number refers to the position of the first base listed relative to the start point of transcription. All binding sites are written in the same direction as the direction of transcription of the indicated gene.

The CRP consensus sequence is that based on comparison of many CRP sites (99).

PART II: STRUCTURE AND FUNCTION OF THE GLP REPRESSOR

RESULTS

The control region for the *glpD* gene has been characterized in greatest detail (142). But the organization and structures of the adjacent *glpE, glpG* and *glpR* genes have not been determined. The *glpR* gene product is the negative regulator of the *glp* regulon. The functions of *glpE* and *glpG* are unknown. The *glp* repressor exists as a tetramer of 30 kDa subunits under nondenaturing conditions. The interaction of the inducer, glycerol-P, with the repressor was characterized. An apparent *Kd* of 31 μM was demonstrated (40). The *glp* repressor bound DNA fragments containing the control regions for the *glpD, glpF* and *glpT-A* genes. Binding of DNA by the repressor was diminished in the presence of glycerol-P. In order to study molecular details of how *glp* repressor negatively regulates the expression of the various *glp* operons, the nucleotide sequence of the *glpEGR* operon has been determined. Based on sequence similarities to other bacterial repressors, the putative DNA binding domain of the *glp* repressor has been proposed. The amino acid sequence of the proposed recognition helix of the *glp* repressor has been altered to verify the function of the DNA-binding domain.

Nucleotide sequence of the* glpEGR *operon: To obtain detailed information about the structures of the *glpE, glpG* and *glpR* genes including their promoters and operators, the nucleotide sequence of the DNA between an upstream *Bgl*II

site within the *glpD* gene and the *Bgl*II site downstream of the *glpR* gene was determined by using the strategy outlined in Fig. 10. Plasmids and M13 clones constructed for this study are listed in Table 8. This region contained 2.72 kb including the *glpE, glpG* and *glpR* genes which are transcribed divergently from the *glpD* gene (7). Recent studies have revealed that the *glpE, glpG* and the *glpR* genes are all under the control of *glpE* promoter, so that they are in the same operon (personal communication, D. Austin). The potential open reading frames capable of encoding the GlpE, GlpG and GlpR proteins were sought.

The complete sequence of the *glpEGR* operon is shown in Fig. 11. An open reading frame of 108 codons beginning with an ATG start codon preceded by a Shine-Dalgarno sequence A-G-A-A-A-G was identified for GlpE (residues 373-699, Fig.11). This is the only reading frame encoding a protein of 12 kDa, which is consistent with the apparent size (13 kDa) observed in minicells (7). An open reading frame was identified for GlpG also. It contains 258 amino acids beginning with an ATG start codon preceded by a Shine-Dalgarno sequence A-G-G-C-G (residues 801-1574, Fig. 11). The calculated MW is 29,130.

Search for a reading frame which could encode the *glp* repressor revealed one of 252 codons which corresponds to the repressor gene. The calculated MW based on the nucleotide sequence of the *glp* repressor is 28,046, which matched the size of the gene product (30,000) determined by SDS-PAGE (40). This potential translated sequence begins at the ATG codon which is preceded by a ribosome binding site A-G-G-G-A (residues 1579-1583, Fig. 11). The beginning of this open reading frame was verified by determination of the sequence of the

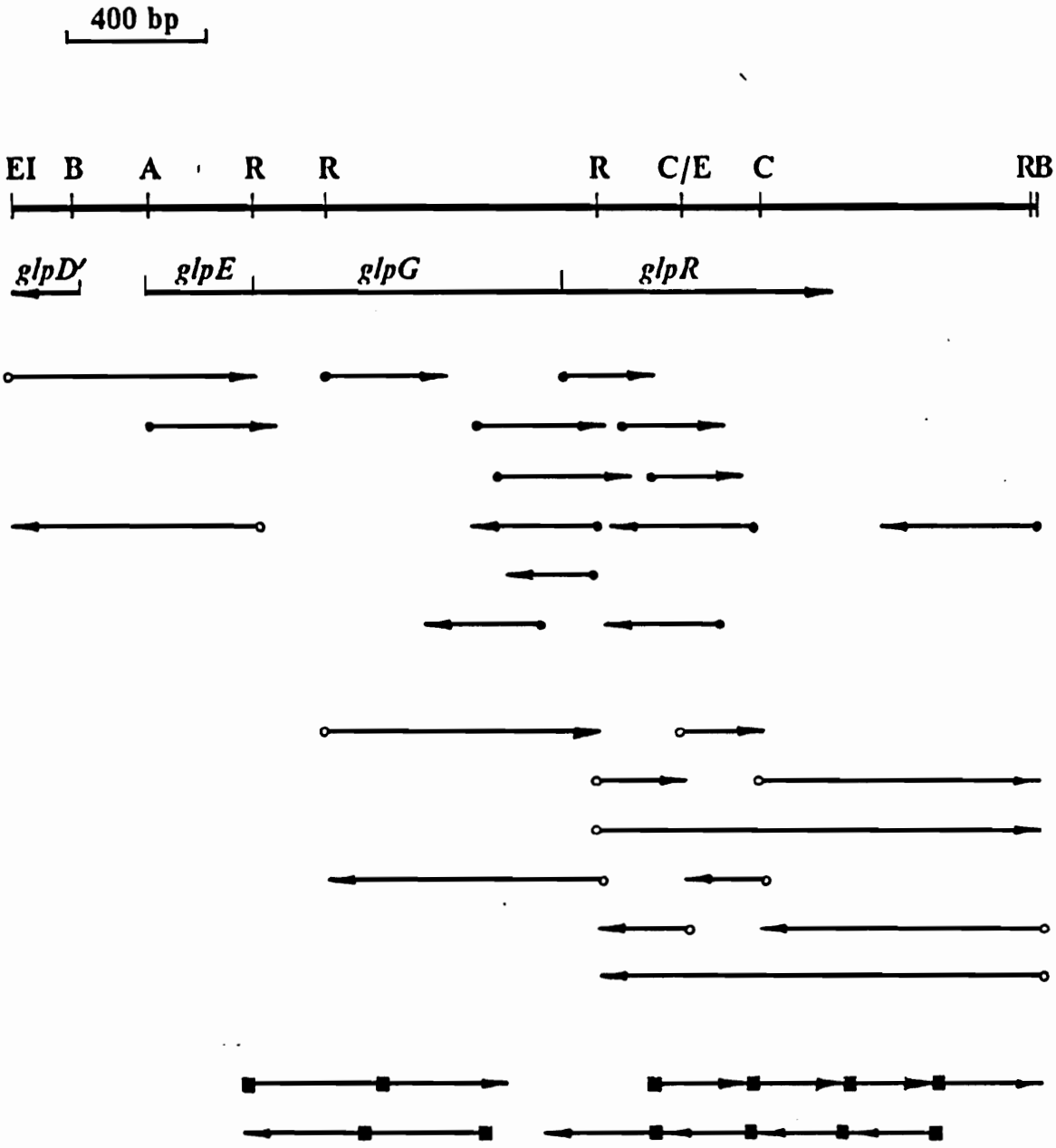


Figure 10. Strategy for sequencing the *glpE*, *glpG* and *glpR* genes: The sequence information obtained from the M13 universal primer is indicated by •, the one from SK, KS, T3, T7 primers is indicated by ○, and that from the synthetic oligonucleotide primers is indicated by ■. Restriction endonuclease cleavage sites: EI, *Eco*RI; B, *Bgl*II; A, *Asu*II; R, *Rsa*I; C, *Cla*I; E, *Eco*RV.

Table 8. Plasmids and M13 clones used in this study

Plasmid or M13 phage	Derivation	Position in sequence[2]
Plasmid[1]		
pSH79	*Eco*RI-*Sal*I from pSH21	1 to 2890
pSY20	*Rsa*I-*Rsa*I from pSH79	886 to 1653
pSY21	*Rsa*I-*Cla*I from pSH79	1654 to 1879
pSY22	*Cla*I-*Cla*I from pSH79	1880 to 2113
pSY23	*Cla*I-*Bgl*II from pSH79	2114 to 2890
pSY24	*Rsa*I-*Rsa*I from pSH79	1654 to 2887
pSY2-A (P33E,Q34M)	*Nsi*I-*Eco*RV from mp19-S	1191 to 1885
pSY2-9 (R38A)	*Nsi*I-*Eco*RV from mp19-S	1191 to 1885
M13 phage[3]		
mp18-E	*Bgl*II-*Bgl*II from pSH79	167 to 2890
mp18-8	deletion from mp18-E	880 to 2890
mp18-44	deletion from mp18-E	1558 to 2890
mp18-44-E	creation *Eco*RI of mp18-44	1558 to 2890
mp18-44-ED	P33E,Q34M mutation from mp18-44-E	1558 to 2890
mp18-44-EQ	R38Q mutation from mp18-44-E	1558 to 2890
mp18-3	deletion from mp18-E	392 to 2890
mp18-181	deletion from mp18-E	1324 to 2890
mp18-71	deletion from mp18-E	1724 to 2890

mp18-52	deletion from mp18-E	1366 to 2890
mp18-73	deletion from mp18-E	1802 to 2890
mp18-A	*Bgl*II-*Bgl*II from pSH79	2890 to 167
mp18-116	deletion from mp18-A	2890 to 167
mp18-117	deletion from mp18-A	2102 to 167
mp18-185	deletion from mp18-A	1658 to 167
mp18-7C	deletion from mp18-A	1644 to 167
mp18-7H	deletion from mp18-A	1998 to 167
mp18-55	deletion from mp18-A	1492 to 167
mp19-S	*Eco*RI-*Sal*I from pSH79	
mp19-S-D	P33E,Q34M mutation from mp19-S	
mp19-S-A	R38A mutation from mp19-S	

Expression vector[4] with:

glpR gene		
pSY2-6 (wt)	*Eco*RI-*Sma*I from mp18-44	1551 to 2890
pSY2-3 (P33E,Q34M)	*Eco*RI-*Sma*I from mp18-44-E	1551 to 2890
pSY2-11 (R38Q)	*Eco*RI-*Sma*I from mp18-44-E	1551 to 2890
glpEGR operon		
pSY2-C (wt)	*Bgl*II-*Bgl*II from pSH79	168 to 2890
pSY2-IX (P33E,Q34M)	*Bgl*II-*Bgl*II from pSY2-A	168 to 2890
pSY2-A9 (R38A)	*Bgl*II-*Bgl*II from pSY2-9	168 to 2890

[1]The vector in each case was Bluescript KS M13+ (Stratagene).

[2]The orientation of the deletion clones is indicated by the numbers given for the position in the sequence.

[3]The vector in each case was M13mp18 or M13mp19 as indicated.

[4]The vector in each case was pSY223.

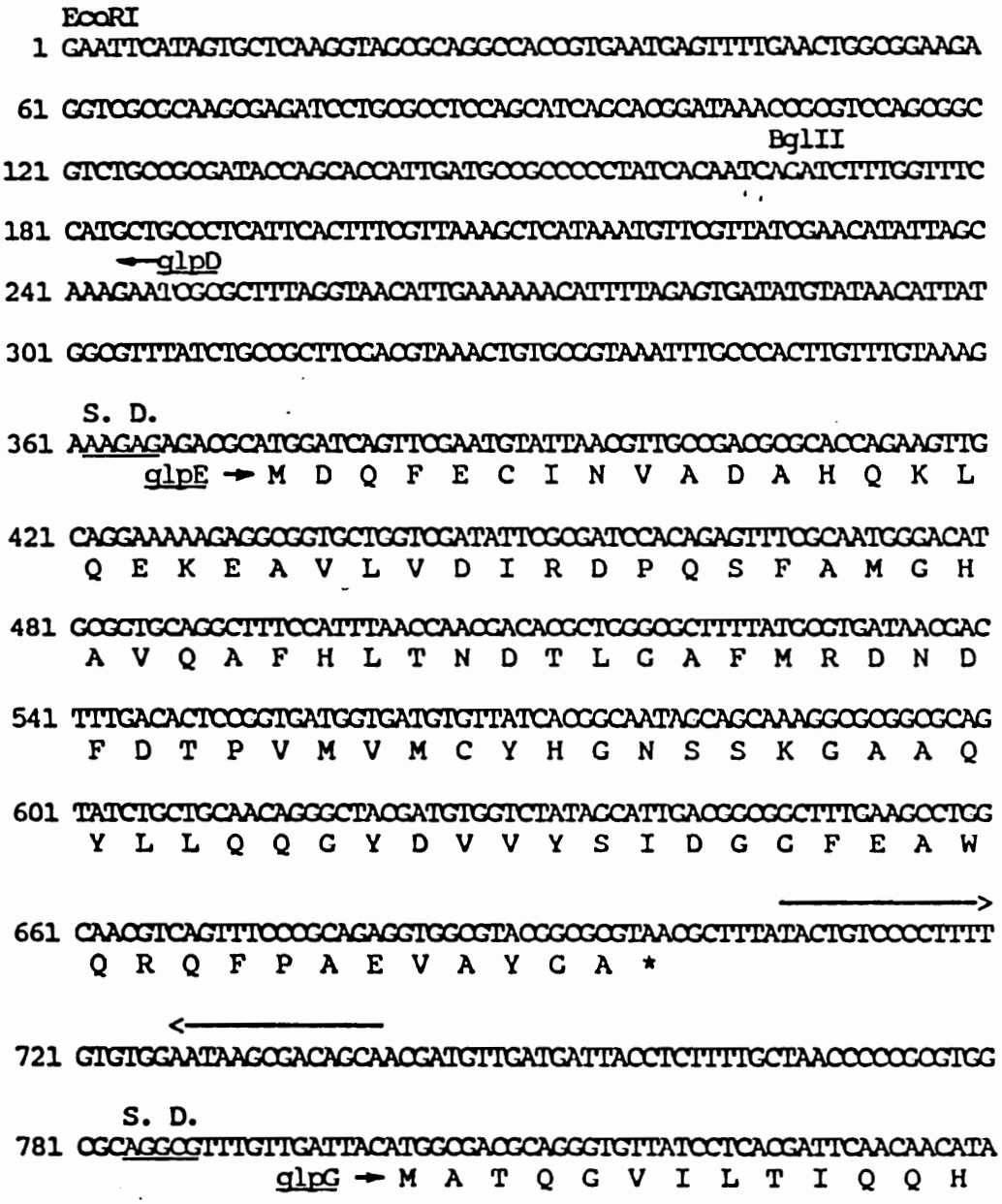
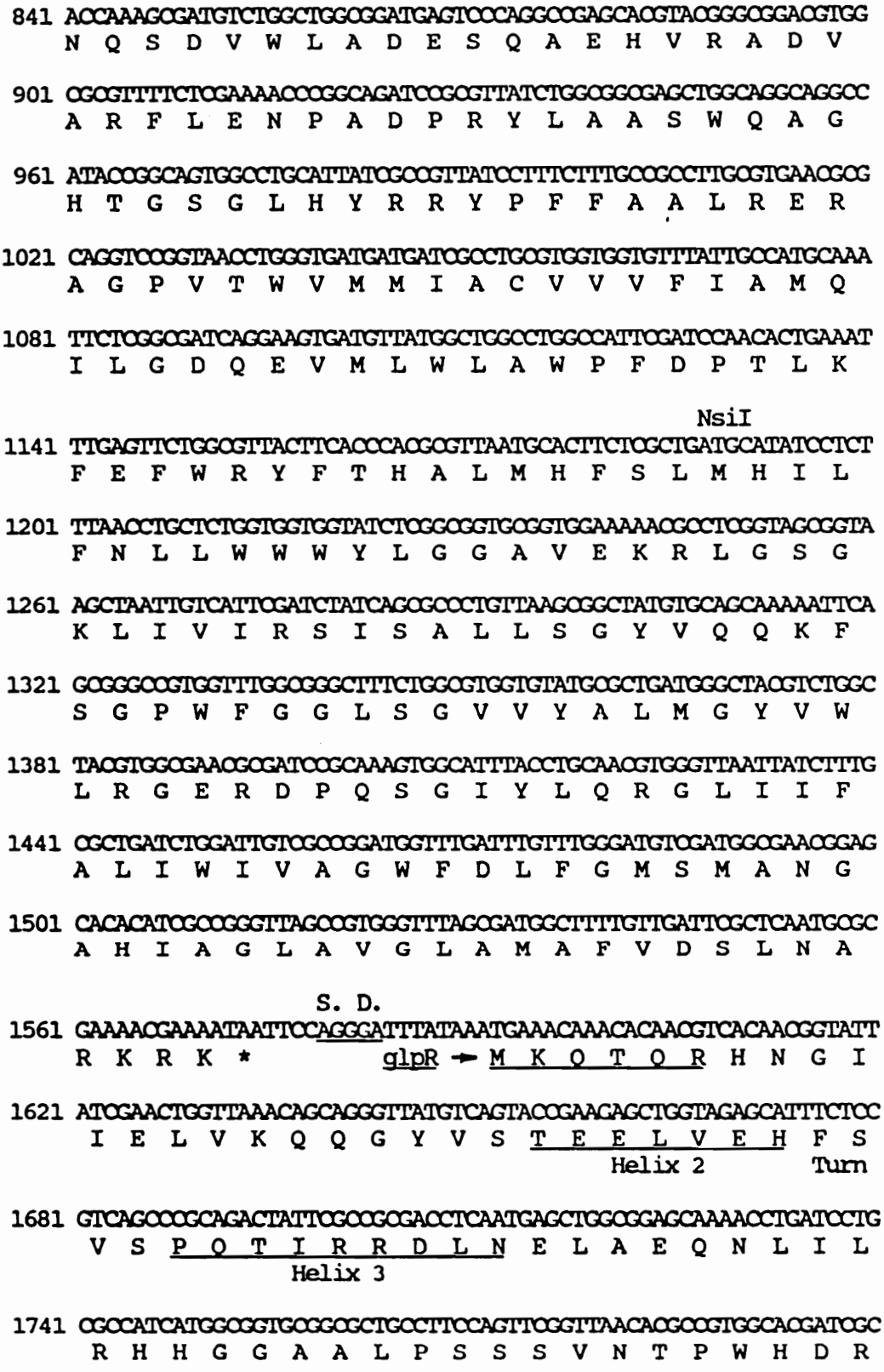
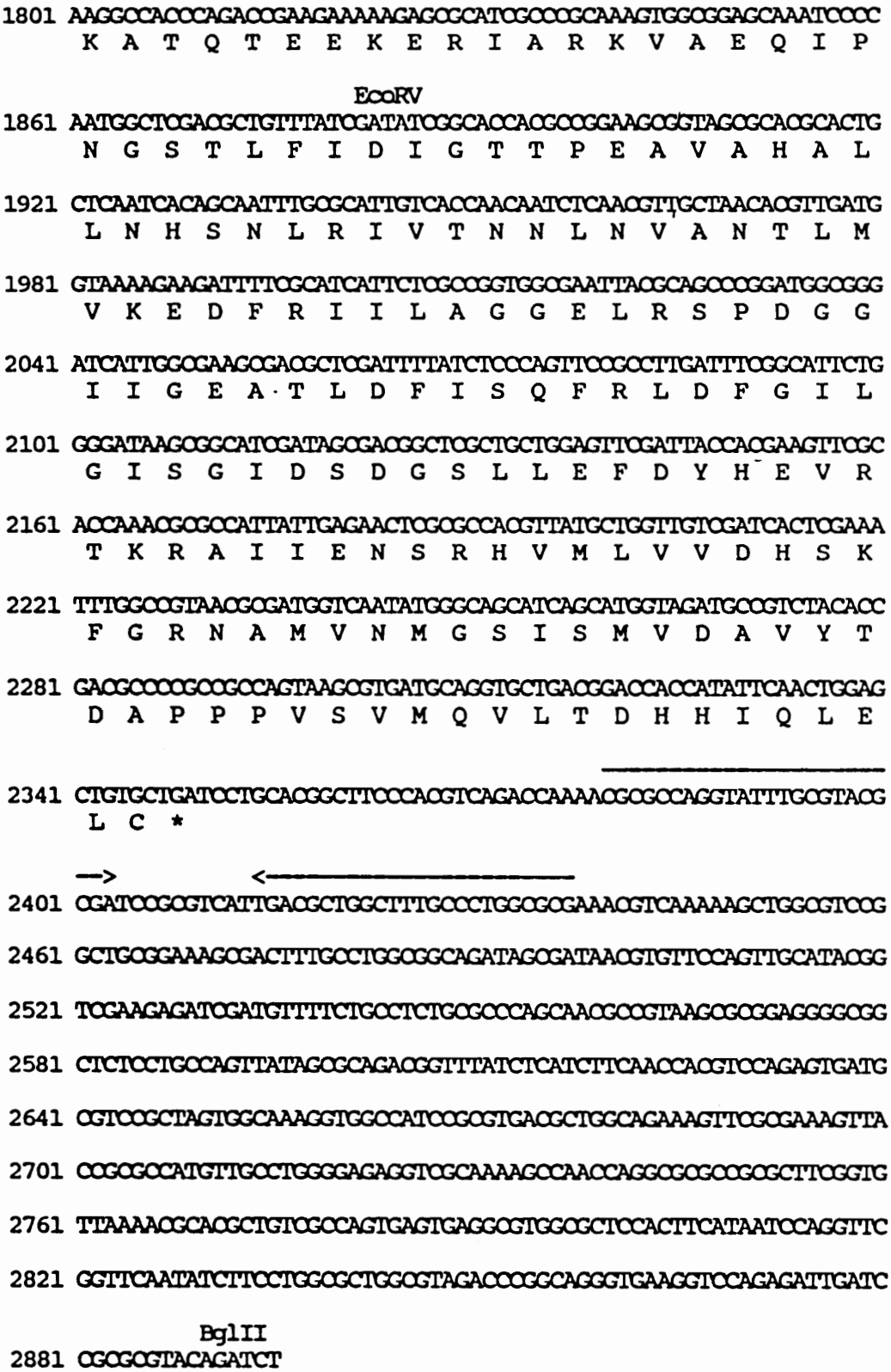


Figure 11. Nucleotide sequence of the *glpEGR* operon and deduced amino acid sequence of GlpE, GlpG and *glp* repressor: GlpE, GlpG and GlpR are polypeptides containing 108, 276 and 252 amino acid residues (12,082, 29,130 and 28,046 daltons, respectively). Sequences resembling the consensus sequence for ribosome binding (S.D.), repressor helix-2 and helix-3 of the DNA binding domain and the six N-terminal amino acids determined by sequence analysis of purified *glp* repressor are underlined. Potential rho-dependent transcription terminators are indicated by long dashed convergent arrows.


841 ACCAAAGCGATGTCTGGCTGGCGGATGAGTCCCAGGCCGAGCACGTACGGGCGGACGTGG
N Q S D V W L A D E S Q A E H V R A D V
901 CGCGTTTTCTCGAAAACCCGGCAGATCCGCGTTATCTGGCGGCGAGCTGGCAGGCAGGCC
A R F L E N P A D P R Y L A A S W Q A G
961 ATACCGGCAGTGGCCTGCATTATCGCCGTTATCCTTTCTTTGCCGCCTTGCGTGAACGCG
H T G S G L H Y R R Y P F F A A L R E R
1021 CAGGTCCGGTAACCTGGGTGATGATGATCGCCTGCGTGGTGGTGTTTATTGCCATGCAAA
A G P V T W V M M I A C V V V F I A M Q
1081 TTCTCGGCGATCAGGAAGTGATGTTATGGCTGGCCTGGCCATTCGATCCAACACTGAAAT
I L G D Q E V M L W L A W P F D P T L K
NsiI
1141 TTGAGTTCTGGCGTTACTTCACCCAGCGTTAATGCACTTCTCGCTGATGCATATCCTCT
F E F W R Y F T H A L M H F S L M H I L
1201 TTAACCTGCTCTGGTGGTGGTATCTCGGCGGTGCGGTGGAAAAACGCCTCGGTAGCGGTA
F N L L W W W Y L G G A V E K R L G S G
1261 AGCTAATTGTCATTCGATCTATCAGCGCCCTGTTAAGCGGCTATGTGCAGCAAAAATTCA
K L I V I R S I S A L L S G Y V Q Q K F
1321 GCGGGCCGTGGTTTGGCGGGCTTTCTGGCGTGGTGTATGCGCTGATGGGCTACGTCTGGC
S G P W F G G L S G V V Y A L M G Y V W
1381 TACGTGGCGAACGCGATCCGCAAAGTGGCATTTACCTGCAACGTGGGTTAATTATCTTTG
L R G E R D P Q S G I Y L Q R G L I I F
1441 CGCTGATCTGGATTGTCGCCGGATGGTTTGATTTGTTTGGGATGTCGATGGCGAACGGAG
A L I W I V A G W F D L F G M S M A N G
1501 CACACATCGCCGGGTTAGCCGTGGGTTTAGCGATGGCTTTTGTTGATTCGCTCAATGCGC
A H I A G L A V G L A M A F V D S L N A
S. D.
1561 GAAAACGAAAATAATTCCAGGGATTTATAAATGAAACAAACACAACGTCACAACGGTATT
R K R K * glpR → M K Q T Q R H N G I
1621 ATCGAACTGGTTAAACAGCAGGGTTATGTCAGTACCGAAGAGCTGGTAGAGCATTTCTCC
I E L V K Q Q G Y V S T E E L V E H F S
Helix 2 Turn
1681 GTCAGCCCGCAGACTATTCGCCGCGACCTCAATGAGCTGGCGGAGCAAAACCTGATCCTG
V S P Q T I R R D L N E L A E Q N L I L
Helix 3
1741 CGCCATCATGGCGGTGCGGCGCTGCCTTCCAGTTCGGTTAACACGCCGTGGCACGATCGC
R H H G G A A L P S S S V N T P W H D R


1801 AAGGCCACCCAGACCGAAGAAAAAGAGCGCATCGCCCGCAAAGTGGCGGAGCAAATCCCC
K A T Q T E E K E R I A R K V A E Q I P
EcoRV
1861 AATGGCTCGACGCCTGTTTATCGATATCGGCACCACGCCGGAAGCGGTAGCGCACGCACTG
N G S T L F I D I G T T P E A V A H A L
1921 CTCAATCACAGCAATTTGCGCATTGTCACCAACAATCTCAACGTTGCTAACACGTTGATG
L N H S N L R I V T N N L N V A N T L M
1981 GTAAAAGAAGATTTTCGCATCATTCTCGCCGGTGGCGAATTACGCAGCCCGGATGGCGGG
V K E D F R I I L A G G E L R S P D G G
2041 ATCATTGGCGAAGCGACGCTCGATTTTATCTCCCAGTTCCGCCTTGATTTCGGCATTCTG
I I G E A T L D F I S Q F R L D F G I L
2101 GGGATAAGCGGCATCGATAGCGACGGCTCGCTGCTGGAGTTCGATTACCACGAAGTTCGC
G I S G I D S D G S L L E F D Y H E V R
2161 ACCAAACGCGCCATTATTGAGAACTCGCGCCACGTTATGCTGGTTGTCGATCACTCGAAA
T K R A I I E N S R H V M L V V D H S K
2221 TTTGGCCGTAACGCGATGGTCAATATGGGCAGCATCAGCATGGTAGATGCCGTCTACACC
F G R N A M V N M G S I S M V D A V Y T
2281 GACGCCCCGCCGCCAGTAAGCGTGATGCAGGTGCTGACGGACCACCATATTCAACTGGAG
D A P P P V S V M Q V L T D H H I Q L E
2341 CTGTGCTGATCCTGCACGGCTTCCCACGTCAGACCAAAACGCGCCAGGTATTTGCGTACG
L C *
2401 CGATCCGCGTCATTGACGCTGGCTTTGCCCTGGCGCGAAACGTCAAAAAGCTGGCGTCCG
2461 GCTGCGGAAAGCGACTTTGCCTGGCGGCAGATAGCGATAACGTGTTCCAGTTGCATACGG
2521 TCGAAGAGATCGATGTTTTCTGCCTCTGCGCCCAGCAACGCCGTAAGCGCGGAGGGGCGG
2581 CTCTCCTGCCAGTTATAGCGCAGACGGTTTATCTCATCTTCAACCACGTCCAGAGTGATG
2641 CGTCCGCTAGTGGCAAAGGTGGCCATCCGCGTGACGCTGGCAGAAAGTTCGCGAAAGTTA
2701 CCGCGCCATGTTGCCTGGGGAGAGGTCGCAAAAGCCAACCAGGCGCGCCGCGCTTCGGTG
2761 TTAAAACGCACGCTGTCGCCAGTGAGTGAGGCGTGGCGCTCCACTTCATAATCCAGGTTC
2821 GGTTCAATATCTTCCTGGCGCTGGCGTAGACCCGCAGGGTGAAGGTCCAGAGATTGATC
BglII
2881 CGCGCGTACAGATCT

first six amino acids at the N-terminus of the *glp* repressor (Table 9). The reading frame in the middle of the *glpR* gene was verified by sequencing the fusion joint of a *glpR-lacZ* translational fusion (Fig. 12). A commercially available 17-mer primer was used which was homologous to *lacZ* sequences starting 22 nucleotides from the *Bam*HI site of pMC1403 and reading toward the *glpR* gene.

The deduced C-terminal codon from nucleotide sequence of the *glpR* gene was cysteine. Carboxypeptidase hydrolysis was performed to cleave several amino acid residues at the C-terminus, followed by amino acid sequence analysis. The results only showed one peak of leucine which was predicted to be next to cysteine (data not shown). The open reading frame that we predicted was the only possible reading frame that contained a leucine residue at C-terminus. Cysteine is easily oxidized to cystine or the sulfhydryl group oxidized to $-SO_3H$ during the process of sequence analysis, so that the peak for cysteine would not be seen.

Inspection of the DNA sequences in the region downstream from the open reading frame of the *glpE* and *glpR* genes revealed the presence of a long dyad symmetry which, once transcribed, could form a stable stem-and-loop structure containing 5 or 9 contiguous bp in the respective stem (Fig. 11). The stem-loop structures could be rho-dependent transcription terminators.

There are discrepancies between the previously reported nucleotide sequences of the *glpEGR* operon (143) and the one reported in this study. They are listed in Table 10. The size of the *glp* repressor we identified is smaller than the one previously reported. Since N-terminal and C-terminal amino acid composition analysis have been made, and open reading frame in the middle of the

Table 9. N-terminal amino acid sequence analysis of *glp* repressor

Cycle	Amino Acid[1]	pmole
1	Met	142
2	X[2]	
3	Glu	105
4	Thr	68
5	Gln	79
6	Arg	24

[1]N-terminal amino acid sequence analysis of *glp* repressor was performed by UM Protein Sequencing Facility, University of Michigan Medical School.

[2]No residue was detected.

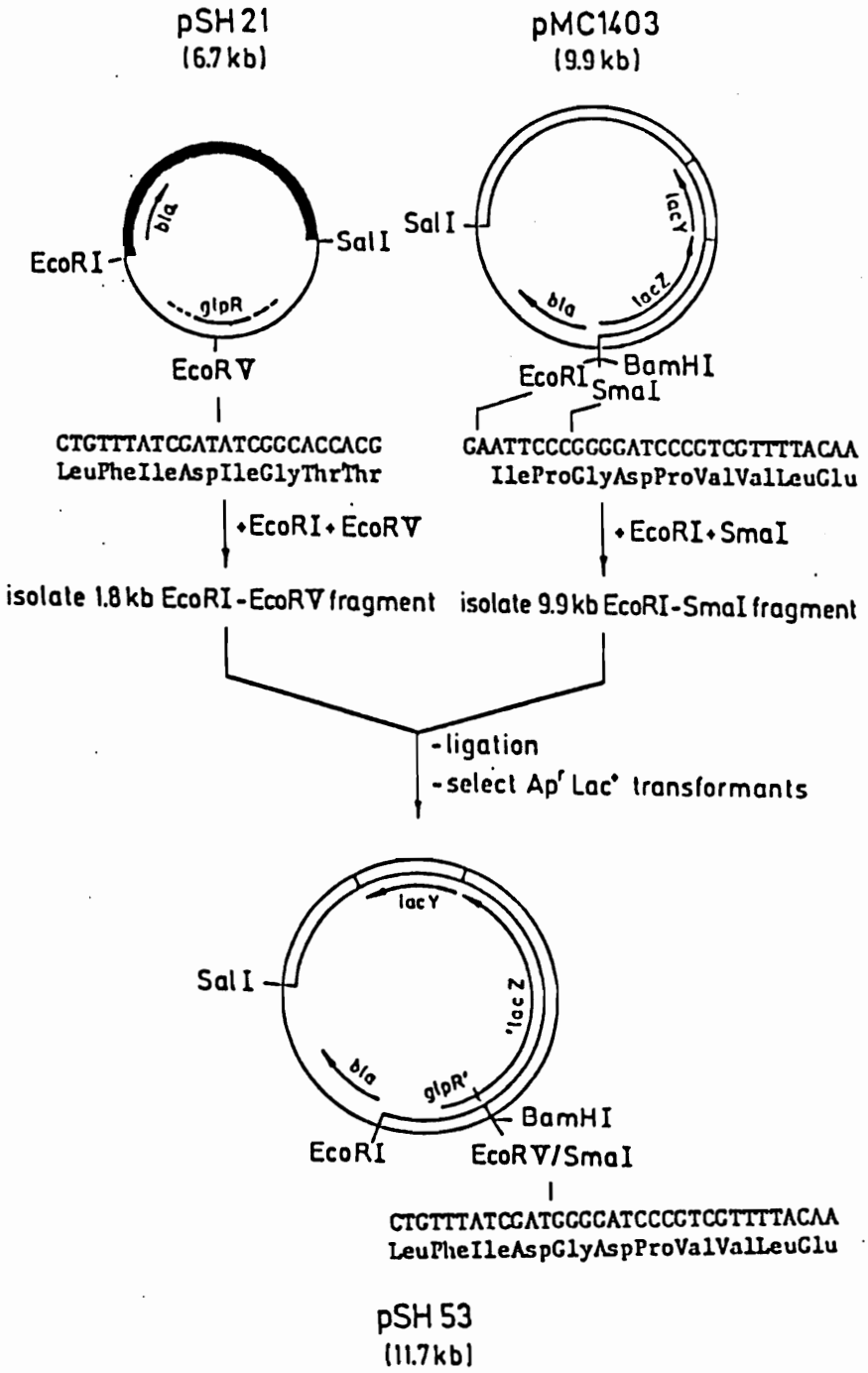


Figure 12. Establishment of the translational reading frame at the center of *glpR*: A 1.8 kb *Eco*RI-*Eco*RV DNA fragment containing the promoter and 5'-end of the *glpR* gene was cloned into *Eco*RI + *Sma*I digested pMC1403. This created an in-frame translational fusion (pSH53) of *glpR* to *lacZ*. The 17-mer primer which is complementary to the 5'-end of the *lacZ* gene was then used to sequence into the *glpR* gene and established the reading frame at the fusion joint.

Table 10. Discrepancies between previously reported nucleotide sequence and the sequence of *glpEGR* reported in this study

Sequence (this study)[1]	Previously reported sequence	Results of error
AGCG	ACG[2]	Frameshift
(116)	(deletion)	
GCGC	GCC	Frameshift
(694)	(deletion)	
GCGG	CCGG	Substitution
(1321)	(base substitution)	
AGCGCACG	ACGACG	Frameshift
(1909-1912)	(deletions)	
CCGGA	CGGCA	Substitution
(2030-2032)	(base substitution)	
TGGCGGGA	TGGGCGA	Frameshift
(2037-2040)	(inversion and deletion)	
GCG	CGG	Substitution
(2053-2054)	(inversion)	
CCTT	CCGTT	Frameshift
(2084)	(insertion)	

CCCC	CCGCC	Frameshift
(2287)	(insertion)	
AGCC	CAGC	Substitution
(2398-2401)	(insertion and deletion)	
CCG	GCG	Substitution
(2405)	(substitution)	
AGCA	ACGA	Substitution
(2555-2556)	(inversion)	
TCCG	TCG	Frameshift
(2668)	(deletion)	
GCGC	GCC	Frameshift
(2745)	(deletion)	

[1]The numbers in parentheses indicate the position of the discrepancy within the nucleotide sequence of *glpEGR* (Fig. 11).

[2]Sequence reported by Choi et al (142).

glpR gene has been verified, we believe that the nucleotide sequence shown here is correct.

Proposed DNA-binding domain for* glp *repressor: Like other prokaryotic repressors, *glp* repressor presumably has three domains responsible for its action: DNA binding, inducer binding and oligomerization. The interactions between repressor and operator DNA have been well studied in several bacterial systems, such as *lac*, *gal* and *deo* operons. The helix-turn-helix sequences have been identified in these operons and they are highly conserved between *lac* and *gal*, and *gal* and *deo* repressors. The essential differences between the putative recognition helices occur in residues 1 and 2 of the *lac*, *gal*, and *deo* repressors. The essential differences between the operators of the *gal*, *deo*, and *lac* systems seemed to be limited to 1 bp in each operator half site (75). It was pointed out by B. Müller-Hill that the *glp* operator is very similar to that of the *deo* system, which suggested that the recognition helices would be similar as well. Therefore, the amino acid sequence of the *glp* repressor was compared to that of the *deo* repressor. It was found that there are striking sequence similarities throughout the entire primary sequences of the *deo* and *glp* repressors (Fig. 13). The length of these two polypeptides is identical — 252 amino acid residues. Among them, 64 residues are identical and 44 additional residues are conserved. The similarities were 43%. Four regions of homology were found: between residues 30-41, 69-84, 94-103 and 203-213 (Fig. 13). Based on the sequence homologies of the *deo* repressor with *gal*, and of *gal* with *lac* repressors, the HTH sequence of the

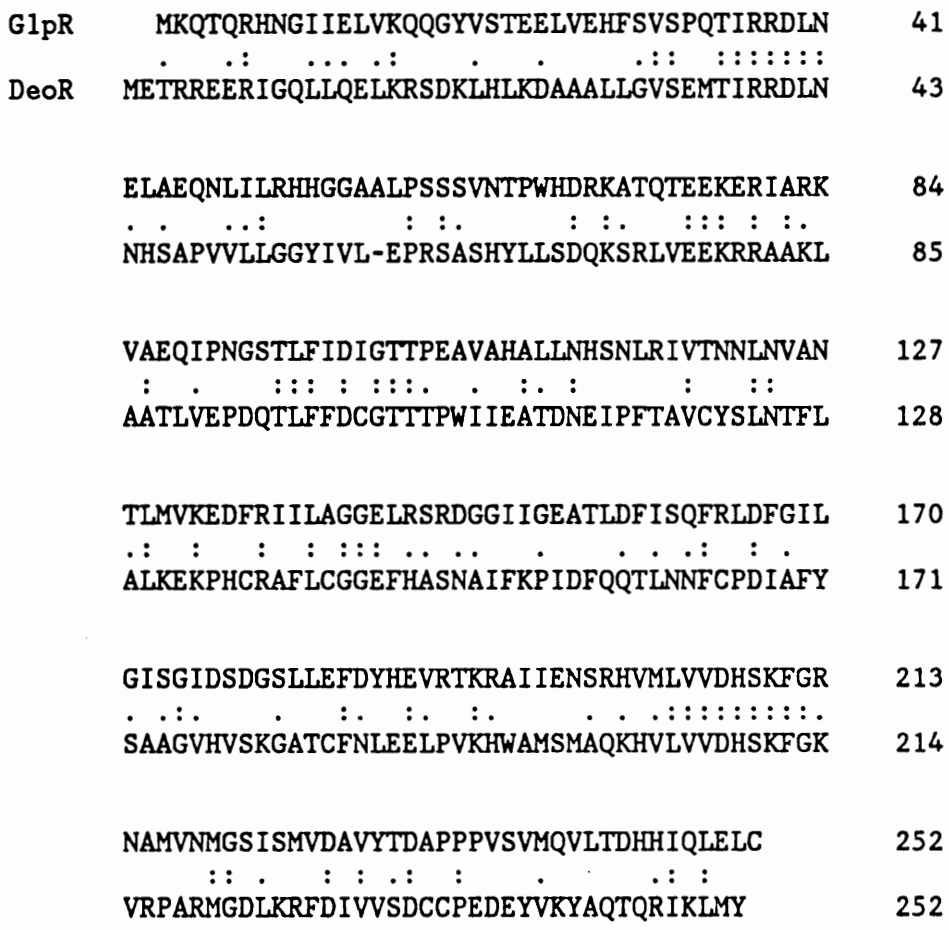


Figure 13. Alignment of the GlpR and DeoR amino acid sequences: Amino acid identities are indicated by a colon. Conservative amino acid substitutions are indicated by a dot. Conservative substitutions are those occuring within the groups ILVM, GASTP, DENQ, WFY, RKH.

glp repressor was proposed to lie between residues 22 and 41 (Figs. 13 and 14). The putative recognition helices of GlpR and DeoR are identical except at the first two residues. The structural relatedness is reflected in the sequence similarities of the respective operators (Fig. 14). Eight out of ten nucleotides are identical. The critical exception is position four, which probably interacts with residues 1 and 2 of the recognition helix (102). The inducer molecules for the two repressors, deoxyribose-5-phosphate and glycerol-P, are structurally similar also. Thus, one of the other conserved sequences could be required for binding of inducer.

Site-directed mutagenesis of the proposed recognition helix of the* glp *repressor encoded by the* glpR *gene:

1. ***Strategy, method and verification:*** If the first two residues of the recognition helix of the *glp* and *deo* repressors differentiate binding to the respective operators, it should be possible to change the specificity of the *glp* repressor to that of *deo* by changing these two AA's. Generation of a mutant repressor with altered DNA binding specificity would provide conclusive evidence that the proposed recognition helix is in fact the recognition helix. If the sixth AA of the recognition helix is also crucial for DNA recognition, it should be the case that change of the residue at position six of the recognition helix from arginine to glutamine would greatly decrease the affinity of the repressor for the operator.

To facilitate studies of function of the *glp* repressor, a controllable expression vector pSY223 was constructed for production of large amounts of

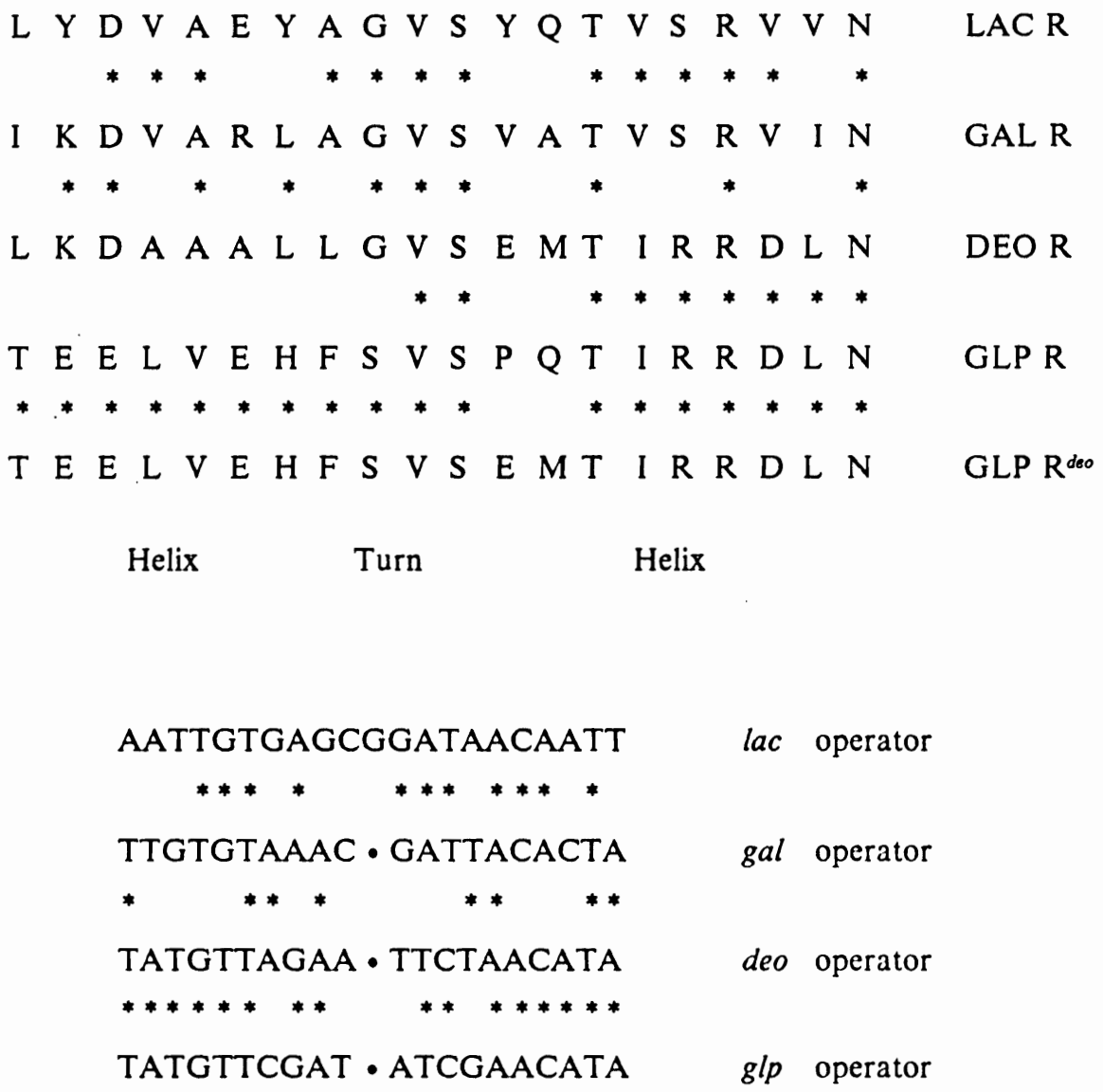


Figure 14. Amino acid sequence similarities used to identify the helix-turn-helix motif of the *glp* repressor: Amino acid or nucleotide identities between two repressors or operators are indicated by asterisks (*). The operator sequences are indicated at the bottom.

repressor and for characterization of wild-type and mutant forms of the *glp* repressor both *in vivo* and *in vitro*. Plasmid pSY223 contains a strong hybrid *tac* promoter (the -35 region of the *trp* promoter fused to the *lac*UV5 -10 and *lac* operator sequences) (144) and adjacent multiple cloning site for the insertion of genes to be expressed at high level between the regulatable *tac* promoter and the *rrnB* transcription terminators (145). The *tac* promoter is three to ten times stronger then the *trp* and *lac*UV5 promoters (144), respectively. A *lacI*^{q} gene of pMJR1560 was deliberately cloned into the vector (Fig. 3), such that it could produce a sufficient amount of *lac* repressor to provide tight control of the *tac* promoter until the promoter is induced to its full strength by addition of IPTG to the medium. Thus, the amount of *glp* repressor produced would be proportional to the amount of IPTG added to the medium.

Initially, an *Eco*RI site (GAATTC) was introduced directly upstream of the start codon of the *glpR* gene of plasmid M13mp18-44 by site-directed mutagenesis (Fig. 15). This was done so that the *glpR* gene could be expressed using the Shine-Dalgarno sequence (A-G-G-A-A) of the vector. The resulting phage (M13mp18-44E) was used as the template for mutagenesis of the recognition helix. Two mutant forms of the repressor were created. In the first case, the first two amino acids of the putative recognition helix were changed from P1Q2 to E1M2 of DeoR (P33E,Q34M). This double mutation is expected to change the specificity of DNA recongition from *glp* operators to *deo* operators. The second mutation changed R6 of GlpR into Q6 (R38Q) (Fig. 15). Both of the mutations were predicted to drastically reduce the binding affinity of *glp* repressor for *glp*

Creation of *glpR* mutations:

```
23-mer for R38A:                              GACTATTCGCGCCGACCTCAATG

22-mer for R38Q:                               ACTATTCGCCAGGACCTCAATG

29-mer for P33E,Q34M:      TTCTCCGTCAGCGAGATGACTATTCGCCG

   DNA  (wild type):       TTCTCCGTCAGCCCGCAGACTATTCGCCGCGACCTCAATG

   AA Seq.  (WT):            F  S  V  S  P  Q  T  I  R  R  D  L  N

   P33E,Q34M                           E  M

   R38Q                                                 Q

   R38A                                                 A
```

*Eco*RI

```
24-mer to create EcoRI:        ATTCCAGGGAATTCTAAATGAAAC

    DNA  (wild type):          ATTCCAGGGATTTATAAATGAAAC

                                   S. D.        M  K  •  •  •
```

Figure 15. Oligonucleotide-directed mutagenesis of the *glpR* gene: The Kunkel technique and M13 ssDNA templates were used. The mutagenized nucleotides are underlined.

operators. The altered sequences at the mutagenized sites were verified for the presence of the desired mutations. Moreover, the entire *glpR* gene was sequenced to rule out the possibility of spontaneous mutations which may cause a mutant phenotype. The 1.3 kb *Eco*RI-*Sma*I fragments containing the repressor genes were then cloned into the expression vector as described in Fig. 16. The effects of mutations on the repressor function were assessed both *in vivo* and *in vitro*.

2. ***Characterization of repressor variants encoded by the wild-type or mutant* glpR *gene:*** The wild-type and mutant forms of the *glp* repressor were overproduced in the presence of 500 μM IPTG as assessed by SDS-PAGE analysis of cells of strain SY102 harboring the appropriate plasmids (Fig. 17). Strain SY102 [*glpR2* ϕ(*glpD-lacZ*)] was used for testing the effects of the mutations on repressor function *in vivo*. The activity of the β-galactosidase was measured in cells grown with or without IPTG. The *glpD-lacZ* fusion was not repressed nearly as much as expected even when repressor was highly overexpressed in the presence of IPTG (Table 11). The repression calculated from the observed specific activities of β-galactosidase was only 1 to 4, which suggested that the repressor did not function properly. Many different conditions were tried to see if repression could be improved. These included different growth media, varying the concentration of IPTG, and the optimal time for addition of IPTG. None of the variants tested improved repression significantly.

Cell free extracts containing wild-type and the P33E,Q34M forms of repressor were prepared. Both extracts contained large amounts of repressor, as determined by binding of inducer and by SDS-PAGE (Fig. 17). The specific ac-

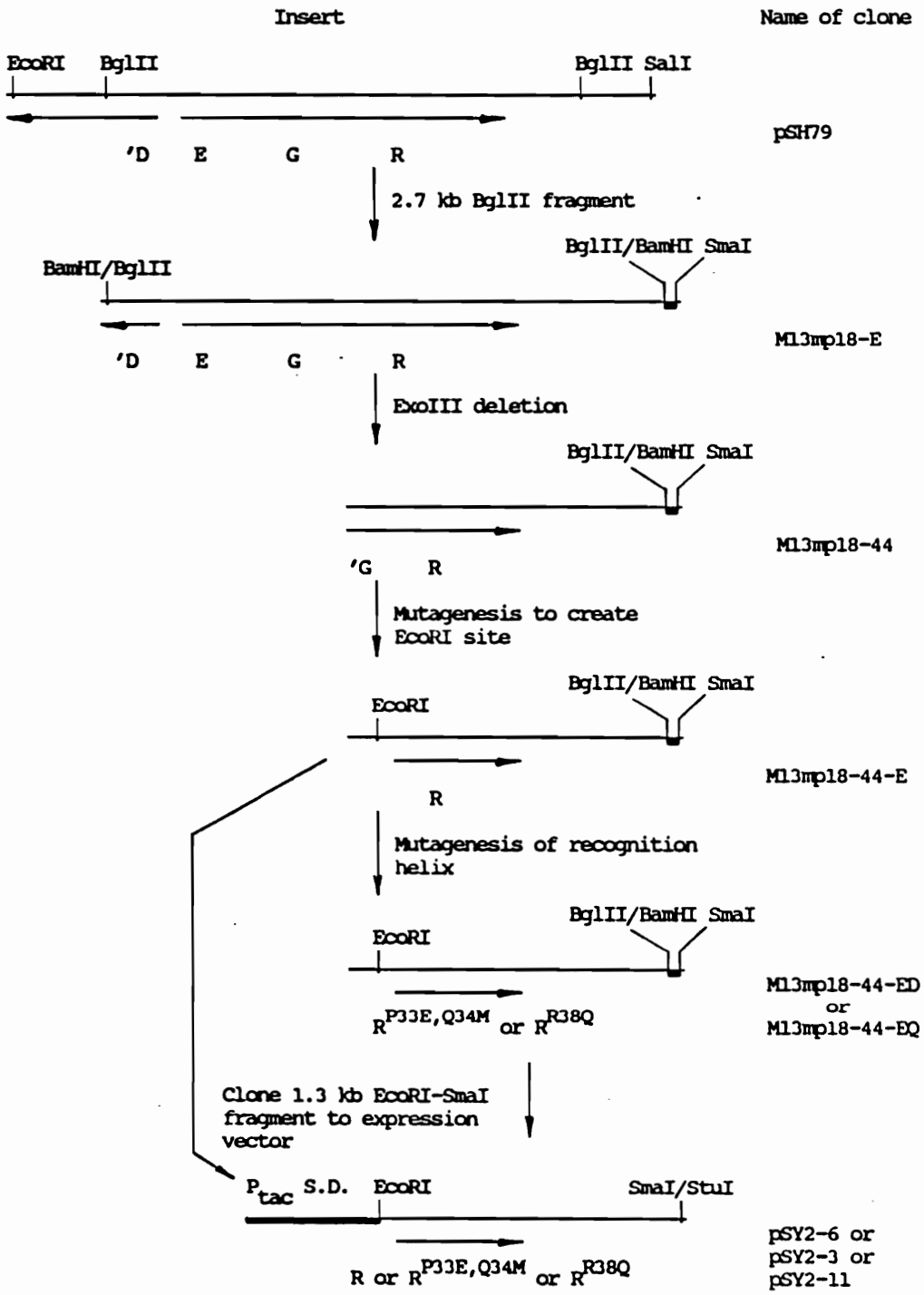


Figure 16. Construction of plasmids containing wild-type and mutant forms of *glp* repressor in expression vector pSY223: Plasmid pSH79 and pSY223 have been described in the text. The heavy lines designate vector DNA.

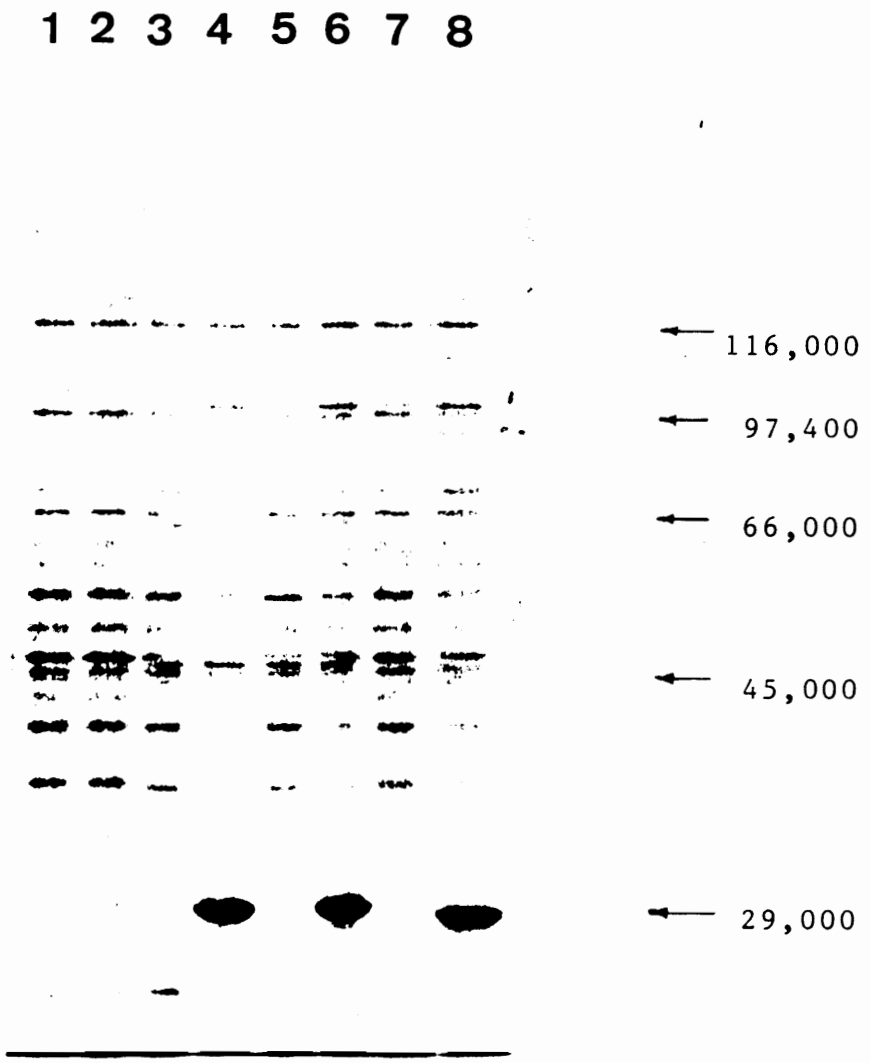


Figure 17. Overexpression of the wild-type and mutant *glpR* genes: *Eco*RI-*Sma*I DNA fragments containing the wild-type or mutant *glpR* gene in M13mp18 were cloned into the expression vector pSY223 in SY102. The *glp* repressor was overproduced in the expression vector. Lane 1 & 2, pSY223 (vector); lane 3 & 4, pSY2-6 (wt); lane 5 & 6, pSY2-3 (P33E,Q34M); lane 7 & 8, pSY2-11 (R38Q). IPTG (0.5) mM was added to the medium for the cells analyzed in the even-numbered lanes. Protein MW standards are at right as indicated.

Table 11. Repression of ϕ(*glpD-lacZ*) or ϕ(*glpK-lacZ*) by *glp* repressor variants

Repressor	Specific Activity of β-Galactosidase		Repression in vivo	
	-IPTG	+IPTG[1]	-IPTG	+IPTG
in SY102 [*glpR* ϕ(*glpD-lacZ*)][2]				
pSY2-6 (*glpR*$^+$)	3,200	1,100	1.1	3.4
pSY2-3 (P33E,Q34M)	2,900	900	1.3	4.2
pSY2-11 (R38Q)	3,400	1,200	1.1	3.1
in GD31 [*glpR*$^+$ ϕ(*glpK-lacZ*)][3]				
pSY223 (vector)	5	5	340	340
pSY2-6 (*glpR*$^+$)	290	32	5.9	53.1
pSY2-3 (P33E,Q34M)	394	44	4.3	38.6
pSY-11 (R38Q)	82	10	20.7	170

[1]IPTG (0.5 mM) was added to the maltose minimal medium where indicated.

[2]Repression is defined as specific activity of β-galactosidase in the absence of *glp* repressor [strain SY102(pSY223); specific activity was 3,600 and 3,800 in the absence and presence of IPTG, respectively] divided by the specific activity of β-galactosidase in strain SY102 containing the indicated plasmids.

[3]Repression is defined as specific activity of β-galactosidase in strain GD31 (pSY223) grown on glycerol (specific activity was 1,700 both in the absence and presence of IPTG) divided by the specific activity of β-galactosidase in strain SY102 containing the indicated plasmids.

tivities determined by the inducer binding assay were 1,500 for wild-type and 3,400 for P33E,Q34M repressor, respectively. These specific activities are 3 to 7 fold higher than that (470 pmol/mg) obtained when the repressor was expressed from the λ pL promoter (40). These results indicate that the repressor was expressed in large amounts in both of the cases in a form that was capable of inducer binding. An *in vitro* gel retardation assay to assess DNA binding was per formed. Both wild-type and P33E,Q34M mutant repressor bound to both *glp* and *deo* tandem operators, but not to vector DNA. Thus, no binding specificity was observed.

An explanation for the above results could be that the *glp* repressor was highly overproduced in a largely inactive form. Alternatively, there could be some other regulatory elements involved in the regulation of the *glpD* gene. An indication that the overproduced repressor was inactive with regard to operator binding but still capable of oligomerization was obtained when overproduced subunits were mixed with subunits encoded by a wild-type chromosomal *glpR* gene. The wild-type and mutant forms of the *glp* repressor encoded by the plasmid had a negative-dominant effect on the wild-type *glp* repressor encoded by the chromosome of strain GD31 [$glpR^+$ $\phi(glpK\text{-}lacZ)$] (Table 11). This indicated that the plasmid-encoded forms of repressor did not function properly. If the defective subunits acted in a negatively dominant fashion, they are sequestering the wild-type subunits from the chromosome into defective dimers, thus lowering their effective concentration. Since chromosomally-encoded repressor functioned properly, the results indicate that there may be other ele-

ments involved in regulation which are not present for activation of repressor overproduced from the plasmids.

Site-directed mutagenesis of the proposed recognition helix of the* glp *repressor encoded by the* glpEGR *operon:

1. ***Rationale and method:*** During the course of the above studies, it was found that the *glpE, G* and *R* genes are present in the same operon (6). In addition, preliminary results indicated that the *glpE* and the *glpG* genes are essential for efficient repression (D. Austin, personel communication). The constructs described above did not contain the *glpE* and *glpG* genes. To find out if these two genes would affect the efficiency of the repressor expressed from pSY223, a 2.7 kb *Bgl*II DNA fragment containing the *glpEGR* operon from plasmid pSH79 was cloned into the expression vector and named pSY2-C. A much higher repression value, 46, was obtained even without induction of repressor synthesis (Table 12). This indicated that GlpE and GlpG might be involved in the activation of the *glp* repressor. Therefore, the entire experiment involving the mutant forms of the repressor was redesigned by using the DNA fragment containing the intact *glpEGR* operon instead of the *glpR* gene alone. A 3 kb *Eco*RI-*Sal*I fragment containing the *glpEGR* operon of pSH79 was cloned into M13mp19 (Fig. 18). The resulting phage, M13mp19-S, was used as the template for site-directed mutagenesis of the recognition helix. As before, the first two amino acids of the recognition helix of GlpR were changed to those of DeoR (P33Q,E34M). In a separate case, R6 was replaced by A6 (R38A) (Figs. 15 and 18). The wild-type and mutant forms of the repressor gene were verified by sequencing. During the

Table 12. Repression of ϕ(*glpD-lacZ*) or ϕ(*deo*C-*lacZ*) by *glp* repressor variants

Repressor	Specific Activity of β-Galactosidase		Repression[1] in vivo	
	-IPTG	+IPTG[2]	-IPTG	+IPTG[2]
in SY102 [*glpR2 φ(glpD-lacZ)*]				
pSY223 (vector)	5,500	5,500		
pSY2-C (*glpR*$^+$)	120	10	46	550
pSY2-IX (P33E,Q34M)	3,500	1,600	2	3
pSY2-A9 (R38A)	5,700	2,900	1	2
in KH682 [*deoR φ(deoC-lacZ)*]				
pSY223 (vector)	240	290		
pSY2-C (*glpR*$^+$)	250	220	1	1.3
pSY2-IX (P33E,Q34M)	240	130	1	2.3
pSYA9 (R38A)	260	140	1	2.1

[1]Repression is defined as in Table 11.

[2]IPTG (0.2 mM) was added to the maltose minimal medium where indicated.

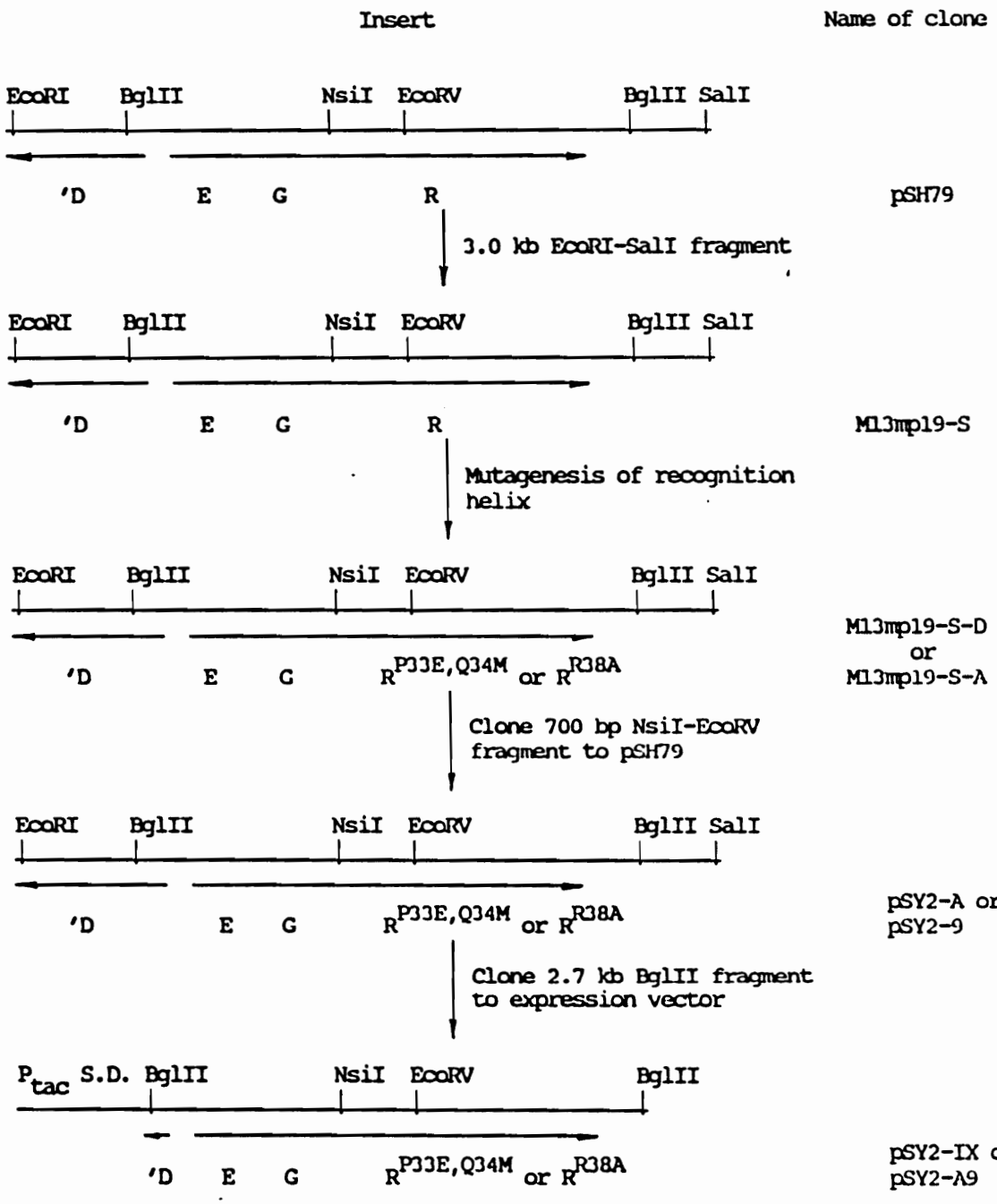


Figure 18. Construction of plasmids containing *glpEGR* operon with mutant forms of *glp* repressor in expression vector: In addition to the above constructs, the 2.7 kb *Bgl*II fragment containing *glpEGR* wild-type operon was cloned directly from pSH79 into the expression vector, named pSY2-C.

process of site-directed mutagenesis, spontaneous mutations occured downstream of the *Eco*RV site, so the 700 bp *Nsi*I-*Eco*RV fragment containing the mutations was subcloned back into pSH79 in a cassette-like manner (Fig. 18). The insertion was verified by restriction digestion. The 2.7 kb *Bgl*II fragments of pSY2-A (P33E,Q34M) and pSY2-9 (R38A) were then cloned into the expression vector pSY223 (Fig. 18). The orientation was determined by restriction enzyme *Eco*RV digestion.

2. ***Characterization of repressor variants encoded by the wild-type or mutant* glpEGR *operon:*** Plasmids where the *glpEGR* operon is positioned properly downstream of the *tac* promoter were transformed into SY102 ϕ*(glpD-lacZ)* and KH682 ϕ(*deoC-lacZ*). The efficacies of the wild-type and mutant forms of the repressor were assessed by measuring β-galactosidase activities. The specific activity of β-galactosidase produced by the *glpD-lacZ* fusion was greatly reduced from 5,500 to 10 for wild-type repressor and slightly reduced from 5,500 to 1,600 or 2,900 for the mutations, respectively (Table 12). The repression value was 550 for wild-type repressor and 3 or 2 for the mutant forms. This indicated that the repressor encoded by the wild-type *glpEGR* operon functioned extremely well. In the P33E,Q34M mutant, the repression was greatly reduced from 550 to 3 (155-fold), which suggested that alteration of the first two amino acids of the putative recognition helix disrupted the operator binding activity of the *glp* repressor. Thus, the first two amino acids are critical for the DNA-recognition. In the case of the R38A mutation, the sixth amino acid arginine was changed to alanine. Repression was almost lost completely in this

case. These results indicate that arginine contributes even more to high affinity operator binding than the first two amino acids. When the polarity, size and charge of arginine have been changed by introduction of a totally different amino acid, alanine, the activity was practically abolished and the affinity of the repressor for the operator was drastically reduced.

Since the recognition helix of GlpR was mutagenized to mimic that of DeoR, we expected that the P33E,Q34M repressor would recognize ϕ(*deoC-lacZ*) instead of ϕ(*glpD-lacZ*). The repression in KH682 was a little lower for the wild-type *glp* repressor than for P33E,Q34M (1.3 versus 2.3, Table 12), which meant that the P33E,Q34M mutant repressor may recognize the *deoC* operator a little better than the wild-type repressor. Comparing the repression of P33E,Q34M and R38A, the repressions were similar. There is no reason for R38A to recognize the *deo* operator, however. It seems that the differences in repression are insignificant. The results suggest that none of the repressors caused significant repression of ϕ(*deoC-lacZ*).

The results of *in vivo* assay of the repressor variants encoded by the wild-type and mutant *glpEGR* operon demonstrated that the wild-type repressor functions properly. Site-directed mutagenesis introduced changes at one or two residues on the recognition helix, which drastically reduced operator binding affinity. This suggests that the proposed recognition helix is in fact the recognition helix, if the mutant forms of the repressor are produced and are stable *in vivo*. To further prove that the repressor has been produced in a stable form both in wild-type and mutant cases, SDS-PAGE and inducer binding assay were per-

formed. Strain TL73 (*glpR2*) was transformed with the expression vector encoding wild-type and mutant forms of the *glp* repressor. These strains were cultured in glycerol minimal medium with 0.1% casamino acids. Glycerol medium was used to overcome potential autoregulation of repressor synthesis. The cells were subcultured in the absence or presence of 200 μM IPTG for 6~8 hours and then cell-free extracts were prepared. In this case, overexpression was not apparent as assessed by SDS-PAGE (data not shown). The lack of overexpression might be due to the large distance between the *tac* promoter and the *glpR* gene, or due to a less efficient ribosome binding site relative to the constructs described earlier, or due to autoregulation, or a combination of these factors. Autoregulation might occur because the repressor would bind to the *glpD* and *glpE* operators, thus preventing the efficient transcription of the *glpEGR* operon from the upstream *tac* promoter.

Because SDS-PAGE did not provide evidence for the presence of the mutant repressors, the specific activities were determined by inducer binding assays (Table 13). Extracts containing wild-type repressor had a specific activity of 29 units/mg. This value is much lower than that in extracts containing repressor produced from the λ pL promoter, 470 units/mg (40), which indicates that the repressor was not strongly overproduced. However, it is much higher than the specific activity found for the strain with the vector (*glpR*$^-$), which proved that the repressor was made. Significantly, the specific activities of the mutant forms of repressor were similar or higher than that of the wild-type. This indicates that the various forms of the repressor were produced and that they are

functional with respect to inducer binding. The results suggest that only the function of DNA binding domain was disrupted by the mutagenesis. Therefore, proposed recognition helix of the *glp* repressor is in fact the recognition helix.

Taken together, all of the results strongly suggest that the *glp* repressor functions as a negative regulatory protein in the *glp* regulon. The repressor activity requires the presence of the gene products of *glpE* and *glpG*. Alteration of the amino acid residues on the presumptive recognition helix of the *glp* repressor greatly reduced the affinity for the operators. Therefore, the recognition helix was identified.

Table 13. Specific activity of the *glp* repressor in cell-free extracts

Repressor	Specific activity[1] (units/mg)
pSY223 (vector *glpR*$^-$)	9
glpR$^+$	29
P33E,Q34M	60
R38A	38

[1]The protein concentration was determined by measuring the A_{280} and A_{260}.

DISCUSSION

Determination of the nucleotide sequence of the *glpEGR* operon demonstrated the potential open reading frames and allowed calculation of the MW of GlpE, GlpG and GlpR, which are in good agreement with the protein sizes derived from SDS-PAGE. The nucleotide sequence of the N-terminus and middle of the open reading frame were verified for *glpR*, but not for *glpE* and *glpG*. Study of the structure and function of the *glpEGR* operon revealed that the *glpR* gene is special among bacterial repressors. It is probably the only repressor gene which is in the same operon with other regulatory genes in *E. coli*. This makes the repressor very unusual. The presence of *glpR* within the polycistronic *glpEGR* operon is presumably related to the unusual function of the *glp* repressor. GlpE or/and GlpG might interact with and cause a conformational change of the repressor, or they might catalyze a covalent modification, such as phosphorylation, acetylation, or adenylylation (or the reverse reaction), which is essential for activation of the repressor. As a matter of fact, the data derived from *in vivo* assays apparently showed that the repressor functions very well in the presence of GlpE and GlpG, but does not function normally in the absence of these two proteins. The repression was more than 500 for wild-type repressor in the presence of GlpE and GlpG, but only 3.4 for wild-type repressor in the absence of GlpE and GlpG. A hypothesis was proposed that GlpE might be an

activator for the *glp* repressor, while GlpG might be a sensor protein which senses the metabolic or physiological state. When GlpE receives the repression signal from GlpG, it activates GlpR through covalent modification to repress the expression of the *glp* regulon. In the absence of GlpE and GlpG, GlpR does not respond to the environmental state properly. An observation suggestive of covalent modification was the inability to identify lysine at position 2 during N-terminal sequence analysis. If this residue was modified, the peak expected for lysine would not be present in the chromatogram during AA sequence analysis.

To test whether or not *glpE* and *glpG* can activate the repressor in *trans*, a plasmid pACYC184 containing the *glpE* and *glpG* genes compatible with pSY223 was transformed into strain SY102, and then pSY223 containing the *glp* repressor variants (without *glpE* and *glpG*) was cotransformed into the same strain. The activity of the β-galactosidase was then measured. There was no apparent change in repression compared to the case without GlpE and GlpG. Since the wild-type repressor with GlpE and GlpG functions well (the repression higher than 500), it seems that the *glpE* and *glpG* genes must work *in cis* with *glpR*.

Another possible reason why *glpR* did not function well without *glpE* and *glpG* could be that spontaneous mutations occurred during site-directed mutagenesis. This explanation is not very likely, however, because after the mutations were generated, the entire sequence of *glpR* was verified by sequencing of the M13 derivative. But when the *glpR* gene was subcloned into the expression vector, the sequence was not verified again.

A third possible reason for lack of repressor function in the absence of *glpE* and *glpG* is that the expression vector worked so efficiently that overproduction of the repressor under *tac* promoter control led to the formation of inclusion bodies. Formation of inclusion bodies could be a highly cooperative, irreversible process which would lead to sequestering of repressor in an insoluble, inactive aggregate. This explanation is not likely, either, because the repressor variants produced in the absence of GlpE and GlpG were still capable of binding glycerol-P. Thus, the most likely explanation for the abnormal function of repressor in these variants is the absence of GlpE and/or GlpG.

The sequence and structural similarities of the repressors and operators between *glp* and *deo* system allowed prediction of the DNA-binding domain of the *glp* repressor. There might be some evolutionary factors involved in the relatedness of these two metabolic systems, which play an important role of breaking down macromolecules. The *glp* regulon is involved in the uptake of the catabolites of phospholipids, while the *deo* regulon is involved in degradation of nucleic acids. Since the sequence and structure of the repressor, operator and inducer are all similar, it appears that study of the relatedness of the fuction of these two repressors is significant.

Our rationale in changing the specificity of *glp* repressor to *deo* repressor is two fold. First, the alignment of the amino acid sequences reveals significant structural similarities between GlpR and DeoR. Without this high degree of structural homology, we imagined that replacing two amino acids from GlpR with those from DeoR might produce an unstable mutant protein. Second, the

similarities between the *glp* and *deo* genes allowed the exchange of two residues from P33Q34 to E33M34 in order to convert the proposed recognition helix of GlpR to that of DeoR. We expected that the E33M34 mutant repressor would not bind the *glp* operators. This mutant repressor should instead bind *deo* operators.

The first of these predictions was correct. Change of the first two amino acids of the putative recognition helix of GlpR to those of DeoR abolished the DNA binding activity with respect to *glp* operators. However, the P33E,Q34M mutant repressor was not able to repress ϕ(*deoC-lacZ*) and thus had no apparent affinity for *deo* operators. Harrison et al. have recently pointed out that besides the second helix of HTH, several mechanisms contribute to recognition, such as direct contacts between amino acid side chains and the edges of bp in the major groove, free energetics of DNA conformation, and hydrogen bonds between amino acids and sugar phosphate backbones (146). The unit for DNA "recognition" is really an entire complex which contributes to recognition specificity. Different mechanism contributes to different extents in different cases. In the *glp* and *deo* systems, the amino acid sequences of helix-3 are identical except the first two residues. The sequences of helix-2, on the other hand, are quite different. Among seven amino acids of helix-2 and four amino acids of the turn, only two residues are identical. It is likely that the amino acid sequence of helix-2 and turn might contribute to position the repressor on the operator DNA also. Moreover, in these two systems, repressors bind to different types of operators. The *glp* repressor binds cooperatively to tandem operators in

the various *glp* operons (except *glpTQ*), while the *deo* repressor binds cooperatively to widely separated operators. Thus, the *deo* repressor prefers to recognize distant operators to form a loop. It would be interesting to find out if the P33E,Q34M mutant form of the *glp* repressor would bind to tandem *deo* operators.

The *glp* repressor presumably binds to its operators by inserting a 'recognition' α-helix into the major groove of the DNA. We have proposed AA-bp contacts that determine at least part of the DNA-binding specificity of the *glp* repressor. As shown in Fig. 14, the glutamine 34 side chain of wild-type repressor may contact the G:C bp 4 near the symmetric center of the operator site by making hydrogen bonds to the O6 and/or N7 positions of guanine 4 within the major groove. Glutamine 34 thus makes a major contribution to the stability of the *glp* reperssor-operator complex. In λ and 434 repressors, glutamine residues at the beginning of helix-3 also donate a hydrogen bond to a residue in helix-2 (147). Thus, glutamine 34 of the *glp* repressor may also hydrogen bond to one of the glutamic acids at the start of the HTH unit, such as E23, E24 or E27. The last residue in the HTH unit is asparagine, a highly conserved residue in many HTH units. In λ and 434 repressors, this asparagine donates two hydrogen bonds. One of them is to a phosphate oxygen on the DNA backbone, and the other is to a carbonyl oxygen on the preceding turn of the α helix. It is likely that similar contacts are made in the case of the *glp* repressor. This hydrogen-bonding network would be important not only for site-specific DNA recognition, but also for maintaining the overall structure of the HTH unit.

From the foregoing discussion, it is apparent that substitution of glutamine by methionine in the P33E,Q34M mutant would disrupt the recognition framework. The methionine 34 methyl side chain can be placed so that it specifically recognizes a T:A bp by making a van der Waals contact with the 5-methyl group of the thymine at position 4 of the *deo* operator. But methionine would be unable to hydrogen bond with a glutamic acid residue at the start of the HTH in order to establish a recognition framework. This could be an explanation for the inability of the P33E,Q34M repressor to bind either *glp* or *deo* operators. Specific amino acid-bp contacts were abolished, and the overall structure of the HTH may have been lost. It is possible that the structure of the HTH unit of the *deo* repressor is maintained by a hydrogen bond between E33 and the lysine at position two of helix-2. It would be interesting to construct the E23K, P33E, Q34M mutant to test this possibility. Likewise, arginine 6 can donate hydrogen bonds to hydrogen bond acceptors in the major groove. Arginine is polar, positively charged and bulky. When it was changed to alanine, the side chain capable of forming hydrogen bonds in the major groove or with the DNA backbone in the wild-type was substituted with a small, nonpolar, uncharged residue, unable to make hydrogen bonds. Thus, the R38A mutant repressor was unable to bind operator DNA with high affinity.

Much has been learned about the structure and function of the *glp* repressor during the course of this study. Further study will be needed to differentiate the function between the wild-type and P33E,Q34M repressors. It will be interesting to elucidate the relationship between the structure of the recogni-

tion helix and the rest of the structure of the protein for DNA recognition specificity. If all the residues of HTH were altered from GlpR to DeoR, it would be predicted that this mutant repressor would repress ϕ(*deoC-lacZ*). If tandem *deo* operators were supplied as the target for the P33E,Q34M repressor, higher affinity binding may be observed relative to the distant operators employed in the present work.

LITERATURE CITED

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

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

3. 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

4. Lin, E. C. C., and Kuritzkes, D. R. (1987) Pathways for anaerobic electron transport. pp.201-221. in Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E. (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* American Society for Microbiology, Washington, DC.

5. Schryvers, A., Lohmeier, E., and Weiner, J. H. (1978) Chemical and functional properties of the native and reconstituted forms of the membrane-bound, aerobic glycerol-3-phosphate dehydrogenase of *Escherichia coli. J. Biol. Chem.* **253,** 783-788

6. Austin, D., and Larson, T. J. (1991) Nucleotide sequence of the *glpD* gene encoding aerobic *sn*-glycerol 3-phosphate dehydrogenase from *Escherichia coli* K-12. *J. Bacteriol.* **173** (in press)

7. 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. Identification of two new structural genes of the *glp*-regulon. *Mol. Gen. Genet.* **202,** 488-492

8. Kistler, W. S., and Lin, E. C. C. (1972) Purification and properties of the flavin-stimulated anaerobic L-α-glycerophosphate dehydrogenase of *Escherichia coli. J. Bacteriol.* **112,** 539-547

9. 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

10. Kuritzkes, D. R., Zhang, L.-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.* **159,** 591-598

11. 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

12. Lin, E. C. C. (1987) Dissimilatory pathways for sugars, polyols, and carboxylates. pp.244-284. in Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E. (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. American Society for Microbiology, Washington, DC.

13. 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

14. 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

15. 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

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

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

18. Riel, J. K., and Paulus, H. (1978) Subunit dissociation in the allosteric regulation of glycerol kinase from *Escherichia coli.* 2. Physical evidence. *Biochemistry* **17,** 5141-5145

19. 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

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

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

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

23. Tommassen, J., Eiglmeier, K., Cole, S. T., Overduin, P., Larson, T. J., and Boos, W. (1990) Characterization of two genes, *glpQ* and *ugpQ*, encoding glycerophoryl diester phosphodiesterases of *Escherichia coli*. *Mol. Gen. Genet.* (in press)

24. 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

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

26. Elvin, C. M., Hardy, C. M., and Rosenberg, H. (1985) Pi exchange mediated by the GlpT-dependent *sn*-glycerol-3-phosphate transport system in *Escherichia coli*. *J. Bacteriol.* **161,** 1054-1058

27. Larson, T. J. (1987) *glpT*-dependent transport of *sn*-glycerol-3-phosphate in *Escherichia coli*. pp.164-169. in Torriani-Gorini, A., Rothman, F. G., Silver, S., Wright, A., and Yagil, E., eds. *Phosphate Metabolism and Cellular Regulation in Microorganisms*. American Society for Microbiology, Washington, DC.

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. 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

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

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

32. 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

33. Edgar, J. R., and Bell, R. M. (1978) Biosynthesis in *Escherichia coli* of *sn*-glycerol-3-phosphate, a precursor of phospholipid. Kinetic characterization of wild-type and feedback-resistant forms of the biosynthetic *sn*-glycerol-3-phosphate dehydrogenase. *J. Biol. Chem.* **253,** 6354-6363

34. Cronan, J. E., Jr., and Rock, C. O. (1987) Biosynthesis of membrane lipids. pp.474-497. in Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E. (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* American Society for Microbiology, Washington, DC.

35. Rock, C. O., and Cronan, J. E., Jr. (1982) Regulation of bacterial membrane lipid synthesis. *Curr. Top. Membr. Transp.* **17,** 209-233

36. 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

37. 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

38. Koch, J. P., Hayashi, S., 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

39. Hayashi, S.-I., and Lin, E. C. C. (1965) Product induction of glycerol kinase in *Escherichia coli. J. Mol. Biol.* **14,** 515-521

40. Larson, T. J., Ye, S., Weissenborn, D. L., Hoffman, H. J., and Schweizer, H. (1987) Purification and characterization of the repressor for the *sn*-glycerol 3-phosphate regulon of *Escherichia coli* K 12. *J. Biol. Chem.* **262,** 15869-15874

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

42. de Crombrugghe, B., Perlman, R. L., Varmus, H. E., and Pastan, I. (1969) Regulation of inducible enzyme synthesis in *Escherichia coli* by cyclic adenosine 3′,5′-monophosphate. *J. Biol. Chem.* **244,** 5828-5835

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

44. Ullmann, A., and Monod, J. (1968) Cyclic AMP as an antagonist of catabolite repression in *Escherichia coli. FEBS Lett.* **2,** 57-61

45. 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

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

47. Spiro, S., Roberts, R. E., and Guest, J. R. (1989) FNR-dependent repression of the *ndh* gene of *Escherichia coli* and metal ion requirement for FNR-regulated gene expression. *Mol. Microbiol.* **3,** 601-608

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

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

50. 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

51. Gilbert, W., and Müller-Hill, B. (1970) The lactose repressor. pp.93-109. in *The lactose operon* (ed.) Beckwith, J. R., and Zipser, D. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY

52. Kania, J., and Brown, D. T. (1976) The functional repressor parts of a tetrameric *lac* repressor-β-galactosidase chimaera are organized as dimers. *Proc. Natl. Acad. Sci. USA* **73,** 3529-3533

53. Beyreuther, K. (1980) Chemical structure and functional organization of *lac* repressor from *Escherichia coli.* pp.123-154. in *The Operon.* Miller, J. H., and Reznikoff, W. S. ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY

54. Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Müller-Hill, B., Pfahl, M., and Schmitz, A. (1972) How *lac* repressor binds to DNA. *Nature* **237,** 322-327

55. Müller-Hill, B. (1975) *Lac* repressor and *lac* operator. *Prog. Biophys. Mol. Biol.* **30,** 227-252

56. Schmitz, A., Schmeissner, U., Miller, J. H., and Lu, P. (1976) Mutations affecting the quaternary structure of the *lac* repressor. *J. Biol. Chem.* **251,** 3359-3366

57. Miller, J. H., Reznikoff, W. S., Silverstone, H. E., Ippen, K., Signer, E. R., and Beckwith, J. R. (1970) Fusions of the *lac* and *trp* regions of the *Escherichia coli* chromosome. *J. Bacteriol.* **104,** 1273-1279

58. Majumder, A., and Adhya, S. (1984) Demonstration of two operator elements in *gal: in vitro* repressor binding studies. *Proc. Natl. Acad. Sci. USA* **81,** 6100-6104

59. von Wilcken-Bergmann, B., Koenen, M., Griesser, H.-W., and Müller-Hill, B. (1983) 72 residues of *gal* repressor fused to β-galactosidase repress the *gal* operon of *Escherichia coli. EMBO J.* **2,** 1271-1274

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

61. Mortensen, L., Dandanell, G., and Hammer, K. (1989) Purification and characterization of the *deoR* repressor of *Escherichia coli. EMBO J.* **8,** 325-331

62. Anderson, W. F., Ohlendorf, D. H., Takeda, Y., and Matthews, B. W. (1981) Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290,** 754-758

63. Pabo, C. O., and Lewis, M. (1982) The operator-binding domain of λ repressor structure and DNA recognition. *Nature* **298,** 443-447

64. Sauer, R. T., Anderegg, R. (1978) Primary structure of the λ repressor. *Biochemistry* **17,** 1092-1100

65. Gussin, G. N., Johnson, A. D., Pabo, C. O., and Sauer, R. T. (1983) Repressor and cro protein: Structure, function, and role in lysogenization. pp.93-121. in Hendrix, R. W., Roberts, J. W., Stahl, F. W., and Weisberg, R. A. (ed.), *Lambda II.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY

66. Johnson, A. D., Meyer, B. J., and Ptashne, M. (1979) Interactions between DNA-bound repressors govern regulation by the λ phage repressor. *Proc. Natl. Acad. Sci. USA* **76,** 5061-5065

67. Takeda, Y., Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1983) DNA-binding proteins. *Science* **221,** 1020-1026

68. McKay, D. B., Weber, I. T., and Steitz, T. A. (1982) Structure of catabolite gene activator protein at 2.9-Å resolution. *J. Biol. Chem.* **257,** 9518-9524

69. Pabo, C. O., and Sauer, R. T. (1984) Protein-DNA recognition. *Annu. Rev. Biochem.* **53,** 293-321

70. Kim, J., Zweib, C., Wu, C., and Adhya, S. (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene* **285,** 15-23

71. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., and Pabo, C. O. (1982) Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. *Nature* **298,** 447-451

72. Weber, I. T., McKay, D. B., and Steitz, T. A. (1982) Two helix DNA binding motif of CAP found in *lac* repressor and *gal* repressor. *Nucleic Acids Res.* **10,** 5085-5102

73. Jacob, F., and Monod, J. (1961) On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* **26,** 193-209

74. Walter, G., Zillig, W., Palm, P., and Fuchs, E. (1967) Initiation of DNA-dependent RNA synthesis and the effect of heparin on RNA polymerase. *Eur. J. Biochem.* **3,** 194-201

75. McClure, N. R. (1985) Mechanism and control of transcription in prokaryotes. *Annu. Rev. Biochem.* **54,** 171-204

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

77. Majumdar, A., and Adhya, S. unpublished results.

78. Sellitti, M. A., Parco, P. A., and Steege, D. A. (1987) *Lac* repressor blocks in vivo transcription of *lac* control region DNA.. *Proc. Natl. Acad. Sci. USA* **84,** 3199-3203

79. Ptashne, M. (1986) *A Genetic Switch: Gene Control and Phage λ.* Cell Press and Blackwell Scientific Publications, Cambridge, MA and Palo Alto, CA

80. Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., and Meyer, B. J. (1980) How the λ repressor and Cro work. *Cell* **19,** 1-11

81. Meyer, B. J., Maurer, R., and Ptashne, M. (1980) Gene regulation at the right operator (O_R) of bacteriophage λ. II. O_R1, O_R2, and O_R3: Their roles in mediating the effects of repressor and cro. *J. Mol. Biol.* **139,** 163-194

82. Sadler, J. R., Sasmor, H. M., and Betz, J. L. (1983) A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. *Proc. Natl. Acad. Sci. USA* **80,** 6785-6789

83. Sasmor, H. M., and Betz, J. L. (1990) Symmetric *lac* operator derivatives: effects of half-operator sequence and spacing on repressor affinity. *Gene* **89,** 1-6

84. Oehler, S., Eismann, E. R., Krämer, H., and Müller-Hill, B. (1990) The three operators of the *lac* operon cooperate in repression. *EMBO J.* **9,** 973-979

85. Hsieh, W. T., Whitson, P. A., Matthews, K. S., and Wells, R. D. (1987) Influence of sequence and distance between two operators on interaction with the *lac* repressor. *J. Biol. Chem.* **262,** 14583-14591

86. Valentin-Hansen, P., Albrechtsen, B., and Larsen, J. E. L. (1986) DNA-protein recognition: demonstration of three genetically separated

operator elements that are required for repression of the *Escherichia coli deoCABD* promoters by the DeoR repressor. *EMBO J.* **5,** 2015-2021

87. Amouyal, M., Mortensen, L., Buc, H., and Hammer, K. (1989) Single and double loop formation when *deoR* repressor binds to its natural operator sites. *Cell* **58,** 545-551

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

89. Martin, K., Huo, L., and Schleif, R. F. (1986) The DNA loop model for *ara* repression. AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc. Natl. Acad. Sci. USA* **83,** 3654-3658

90. Whitson, P. A., Hsieh, W.-T., Wells, R. D., and Matthews, K. S. (1987) Influence of supercoiling and sequence context on operator DNA binding with *lac* repressor. *J. Biol. Chem.* **262,** 14592-14599

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

92. Hochschild, A., and Ptashne, M. (1986) Cooperative binding of λ repressor to sites separated by integral turns of the DNA helix. *Cell* **44,** 681-687

93. Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M., and Chambon, P. (1986) Requirement of stereo specific alignments for initiation from the simian virus 40 early promoter. *Nature* **319,** 121-126

94. Berg, O. G., and von Hippel, P. H. (1988) Selection of DNA binding sites by regulatory proteins. *Trends Biochem. Sci.* **13,** 207-211

95. Lehming, N., Sartorius, J., Oehler, S., von Wilcken-Bergmann, B., and Müller-Hill, B. (1988) Recognition helices of *lac* and λ repressor are oriented in opposite directions and recognize similar DNA sequences. *Proc. Natl. Acad. Sci. USA* **85,** 7947-7951

96. Wharton, R. P., and Ptashne, M. (1986) An α-helix determines the DNA-binding specificity of a repressor. *Trends Biochem. Sci.* **11,** 71-73

97. Wharton, R. P., and Ptashne, M. (1985) Changing the binding specificity of a repressor by redesigning an α-helix. *Nature* **316,** 601-605

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

99. Zhang, X., and Ebright, R. H. (1990) Substitution of 2 base pairs (1 base pair per DNA half-site) within the *Escherichia coli lac* promoter DNA site for catabolite gene activator protein place the *lac* promoter in the FNR regulon. *J. Biol. Chem.* **265,** 12400-12403

100. Besse, M., von Wilcken-Bergmann, B., and Müller-Hill, B. (1986) Synthetic *lac* operator mediates repression through *lac* repressor when introduced upstream and downstream from *lac* promoter. *EMBO J.* **5,** 1377-1381

101. Lehming, N., Sartorius, J., Niemöller, M., Genenger, G., von Wilcken-Bergmann, B., and Müller-Hill, B. (1987) The interaction of the recognition helix of *lac* repressor with *lac* oprator. *EMBO J.* **6,** 3145-3153

102. Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1990) Mutant *lac* repressors with new specificities hint at rules for protein-DNA recognition. *EMBO J.* **9,** 615-621

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

104. User Bulletin 13 (1987) The evaluation and purification of synthetic oligonucleotides. DNA synthesizer Model 380/381 from Applied Biosystems Inc.

105. Harman, J. G., Mckenney, K., and Peterkofsky, A. (1986) Structure-function analysis of three cAMP-independent forms of the cAMP receptor protein. *J. Biol. Chem.* **261,** 16322-16339

106. Liss, L. R. (1987) New M13 Host: DH5αF′ competent cells. *Focus* **9:3,** 13

107. 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,** 103-119

108. Kunkel, T. A., Robert, J. D., and Zakour, R. A. (1989) Rapid and efficient site-specific mutagenesis without phenotypic selection. pp.587-601. in Wu, R., Grossman, L., and Moldave, K. (eds.) *Recombinant DNA Methodology.* Academic Press, Inc., San Diego, CA

109. Schultz, J., Silhavy, T. J., Berman, M. L., Fiil, N., and Emr, S. D. (1982) A previously unidentified gene in the *spc* operon of *Escherichia coli* K-12 specifies a component of the protein export machinery. *Cell* **31,** 227-235

110. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

111. Tartof, K. D., and Hobbs, C. A. (1987) Improved media for growing plasmids and cosmid clones. *BRL Focus* **9:2,** 12

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

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

114. Morelle, G. (1989) A plasmid extraction procedure on a miniprep scale. *Focus* **11:1,** 7

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

116. BRL in *M13 Cloning/Dideoxy Sequencing,* pp.44-49, BRL

117. 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

118. Hackett, P. B., Fuchs, J. A., and Messing, J. W. (1988) in *An Introduction to Recombinant DNA Techniques.* pp.53-55, Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA

119. Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. *Gene* **28,** 351-359

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

121. Garner, M. M., and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *E. coli* lactose operon regulatory system. *Nucleic Acids Res.* **13,** 3047-3060

122. Fried, M., and Crothers, D. M. (1981) Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9,** 6505-6525

123. de Lorenzo, V., Wee, S., Herrero, M., and Neilands, J. B. (1987) Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. *J. Bacteriol.* **169,** 2624-2630

124. Jones, K. A., Yamamoto, K. R., and Tjian, R. (1985) Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. *Cell* **42,** 559-572

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

126. Orozco, E. M., Jr., Mullet, J. E., and Chua, N. (1985) An *in vitro* system for accurate transcription initiation of chloroplast protein genes. *Nucleic Acids Res.* **13,** 1283-1302

127. 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

128. Silhavy, T. J., Berman, M. L., and Enguist, L. W. (1984) in *Experiments with Gene Fusions,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

129. Brosius, J., and Holy, A. (1984) Regulation of ribosomal RNA promoters with a synthetic *lac* operator. *Proc. Natl. Acad. Sci. USA.* **81,** 6929-6933

130. Stark, M. J. R. (1987) Multicopy expression vectors carrying the *lac* repressor gene for regulated high-level expression of genes in *Escherichia coli. Gene* **51,** 255-267

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

132. 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

133. Beisenherz, G., Bucher, T., and Garbade, K. (1955) α-Glycerolphosphate dehydrogenase from rabbit muscle. *Methods Enzymol.* **1,** 391-397

134. 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

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

136. Ebina, Y., Takahara, Y., Kishi, F., Nakazawa, A., and Brent, R. (1983) LexA protein is a repressor of the colicin E1 gene. *J. Biol. Chem.* **258,** 13258-13261

137. Brent, R., and Ptashne, M. (1981) Mechanism of action of the *lexA* gene product. *Proc. Natl. Acad. Sci. USA* **78,** 4204-4208

138. Cunin, R., Glansdorff, N., Pierard, A., and Stalon, V. (1986) Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* **50,** 314-352

139. Lim, D., Oppenheim, D., Eckhardt, T., and Maas, W. K. (1987) Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor. *Proc. Natl. Acad. Sci. USA* **84,** 6697-6701

140. Pittard, A. J. (1987) Biosynthesis of the aromatic amino acids. pp.368-394 in Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E. (eds.) *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* American Society for Microbiology, Washington, DC.

141. Larson, T. J., and van Loo-Bhattacharya, A. T., unpublished data, 1990

142. 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

143. Choi, Y., Kawase, S., Nishida, T., Sakai, H., Komano, T., Kawamukai, M., Utsumi, R., Kohara, Y., and Akiyama, K. (1988) Nucleotide sequence of the *glpR* gene encoding the repressor for the

glycerol-3-phosphate regulon of *Escherichia coli* K 12. *Nucleic Acids Res.* **16,** 7732

144. de Boer, H. A., Comstock, L. J., and Vasser, M. (1983) The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. *Proc. Natl. Acad. Sci. USA.* **80,** 21-25

145. Brosius, J., Dull,T. J., Sleeter, D. D., and Noller, H. F. (1981) Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli. J. Mol. Biol.* **148,** 107-127

146. Harrison, S. C., and Aggarwal, A. K. (1990) DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* **59,** 933-969

147. Pabo, C. O., Aggarwal, A. K., Jordan, S. R., Beamer, L. J., Obeysekare, U. R., and Harrison, S. C. (1990) Conserved residues make similar contacts in two repressor-operator complexes. *Science* **247,** 1210-1213

Vita

Shanzhang Ye was born in Shanghai, China on December 8, 1948. She received the Bachelor of Science degree in Biochemistry with honor from the East China University of Chemical Technology in Shanghai, China in 1982. From 1982 to 1985, she worked in the Shanghai Institute of Biochemistry, Chinese Academy of Science and in the Shanghai Institute of Biological Products for one and one-half years each.

In July, 1985 she moved to the United States and worked as a research technician in the Department of Cell Biology, Baylor College of Medicine. In 1986, she moved to Blacksburg to pursue the PhD degree in the Department of Biochemistry and Nutrition at Virginia Tech. During her graduate studies, she joined Dr. Larson's laboratory and studied regulation of gene expression in *Escherichia coli.*

Shanzhang Ye received her degree of Doctor in Philosophy in November 1990.