STUDIES ON THE OPERATOR-REPRESSOR-EFFECTOR INTERACTIONS IN THE *glp* REGULON OF *ESCHERICHIA COLI* K-12

by

Ningyue Zhao

Thesis submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirments for the degree of
MASTER OF SCIENCE

in

Biochemistry and Anaerobic Microbiology

APPROVED:

Timothy J. Larson, Chairman

Patrick K. Bender

Brenda W. Shirley

September, 1993

Blacksburg, Virginia

C.2

STUDIES ON THE OPERATOR-REPRESSOR-EFFECTOR INTERACTIONS IN THE *glp* REGULON OF *ESCHERICHIA COLI* K-12

by

Ningyue Zhao

Timothy J. Larson, Chairman

Biochemistry and Anaerobic Microbiology

(ABSTRACT)

All of the *glp* genes of *Escherichia coli* are subject to negative regulation by the *glpR*-encoded repressor (GlpR). Comparison of the repressor binding affinity for consensus and altered consensus operator regions showed that positions 3, 4, 5, 7 and 8 bp removed from the center of operator symmetry are most important for repressor binding. Cooperative binding of repressors to tandem operators was demonstrated. Cooperativity was maximal when two 20-base pair operators were directly repeated and decreased with the deletion of 2 base pairs or the addition of 4 base pairs between the operators. Cooperative binding was eliminated by a 6 base pair insertion between tandem operators.

The inducer for the *glp* regulon is glycerol 3-phosphate. Strain ECL89 has a noninducible phenotype for the members of the *glp* regulon. In order to determine the molecular basis for the noninducible phenotype, *glpR12* DNA from ECL89 was amplified using the polymerase chain reaction, cloned and sequenced. The DNA-binding domain of the noninducible repressor was identical to the wild-type repressor. The alteration which resulted in amino acid substitution at position 101 (T $\rightarrow$ I) was responsible for the noninducibility. Results from inducer binding assays suggested that the noninducible phenotype was the result of a lowered binding affinity of *glp* repressor for the inducer.

TO YUNZHEN YANG and XIAFANG ZHAO

Acknowledgements

I wish to express my deepest thanks to my advisor, Dr. Timothy J. Larson for his support, guidance and patience during the course of this study. Thanks are also due to the members of my committee, Dr. B. W. Shirley and Dr. P. K. Bender for their help and suggestions.

I am grateful to Ali Bhattacharya for her assistance and making the lab a pleasant place to work. I also would like to thank Won Oh for her technical help during the beginning of my research. I would like to thank all my friends for their help and friendship.

Finally, I wish to especially thank my husband, Jiuyi Cheng for his love, support and understanding.

List of Abbreviations

A	absorbance
Amp	ampicillin
bp	base pair
BSA	bovine serum albumin
cAMP	cyclic adenosine 3', 5'-monophosphate
CRP	cAMP receptor protein
dH_2O	distilled water
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetate
EtBr	ethidium bromide
FAD	flavin adenine dinucleotide
FDP	fructose 1,6-bisphosphate
FNR	anaerobic transcriptional activator
Glycerol-P	*sn*-glycerol 3-phosphate
HTH	helix-turn-helix
IPTG	isopropylthio-β-D-galactopyranoside
kb	kilobase
Kd	dissociation constant

kDa	kilodalton
LB	Luria broth
MOPS	3-[N-morpholino]propanesulfonic acid
MW	molecular weight
PMSF	phenylmethylsulfonyl fluoride
S.D.	Shine-Dalgarno
SDS	sodium dodecyl sulfate
Tris-HCl	tris(hydroxymethyl)aminomethane-hydrochloride
U	units
X-gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside

Table of Contents

List of Illustrations

List of Tables

INTRODUCTION

The *glp* regulon of *Escherichia coli* is composed of five operons encoding the enzymes responsible for dissimilation of glycerol 3-phosphate and its precursors. The *glp* operons are negatively controlled by the *glpR*-encoded repressor (GlpR). A total of 13 operators have been identified in the *glp* operons. A consensus operator was deduced by sequence comparison: 5'-WATGTTCGWT AWCGAACATW-3' (W = A or T). Tandemly repeated operator sites are present in the *glpACB*, *glpD* and *glpFKX* operons and cooperative binding of repressors has been suggested in these operons. In order to learn more about the molecular details of repressor-operator interaction, experiments aimed at identification of the nucleotides in the operator that are critical for repressor binding and the demonstration of cooperative binding between two adjacent operators were carried out. The results of the experiments are reported here.

GlpR is a tetramer with four identical 30 *kDa* subunits. Each subunit contains a binding site for glycerol 3-phosphate and a putative helix-turn-

helix DNA-binding motif. The affinity of the *glp* repressor for its operator sites is decreased upon binding of glycerol 3-phosphate, the inducer for the regulon.

Strain ECL89 has a noninducible phenotype for the *glp* regulon. In order to determine the molecular basis for the noninducible phenotype, the nucleotide sequence of the *glpR12* gene from this strain was determined. The inducer binding activity was measured for both the wild-type and noninducible repressors.

LITERATURE REVIEW

GLYCEROL-P METABOLISM AND *glp* GENE ORGANIZATION IN *E. coli*

sn-Glycerol-3-phosphate (glycerol-P) is a direct precursor for the biosynthesis of phospholipids, which are essential components of all biological membranes (1,2). When glycerol or its precursors are in excess, they can be used as carbon and energy sources for many microorganisms (3). Thus, organisms must coordinate the levels of the catabolic activities with the levels of the phospholipid biosynthetic activities to maintain the cellular concentration of glycerol-P at a level allowing an optimal rate of phospholipid synthesis. The proteins that are responsible for dissimilating glycerol, glycerol-P and glycerophosphodiesters are encoded by the *glp* regulon in *E. coli* (Fig. 1) (3,4). This regulon is composed of five operons positioned at three different regions on the linkage map of *E. coli* (Fig. 2) (4). All of the *glp* genes are negatively controlled by the *glp* repressor

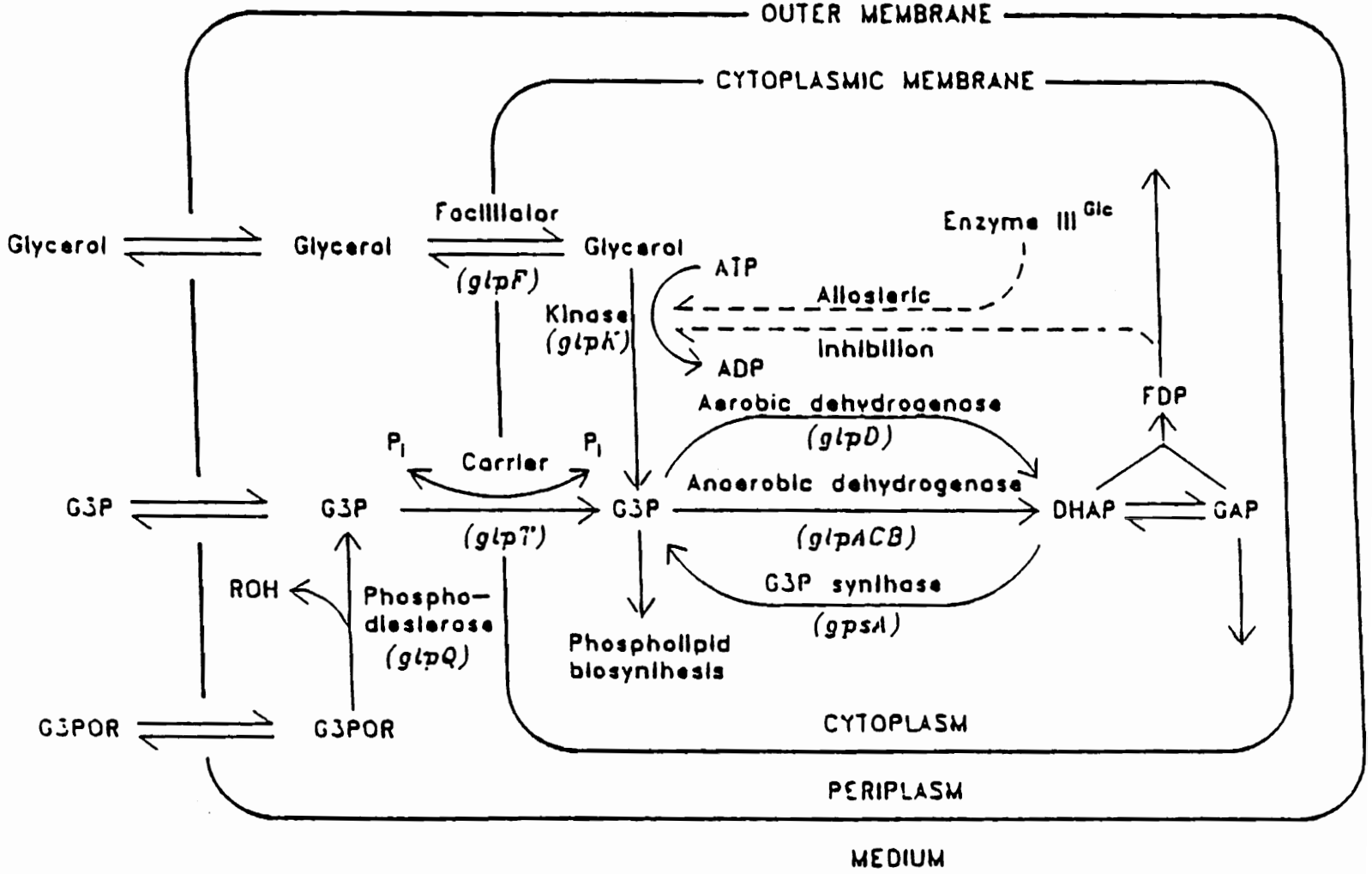

Figure 1. Glycerol-P metabolism in *E. coli*.

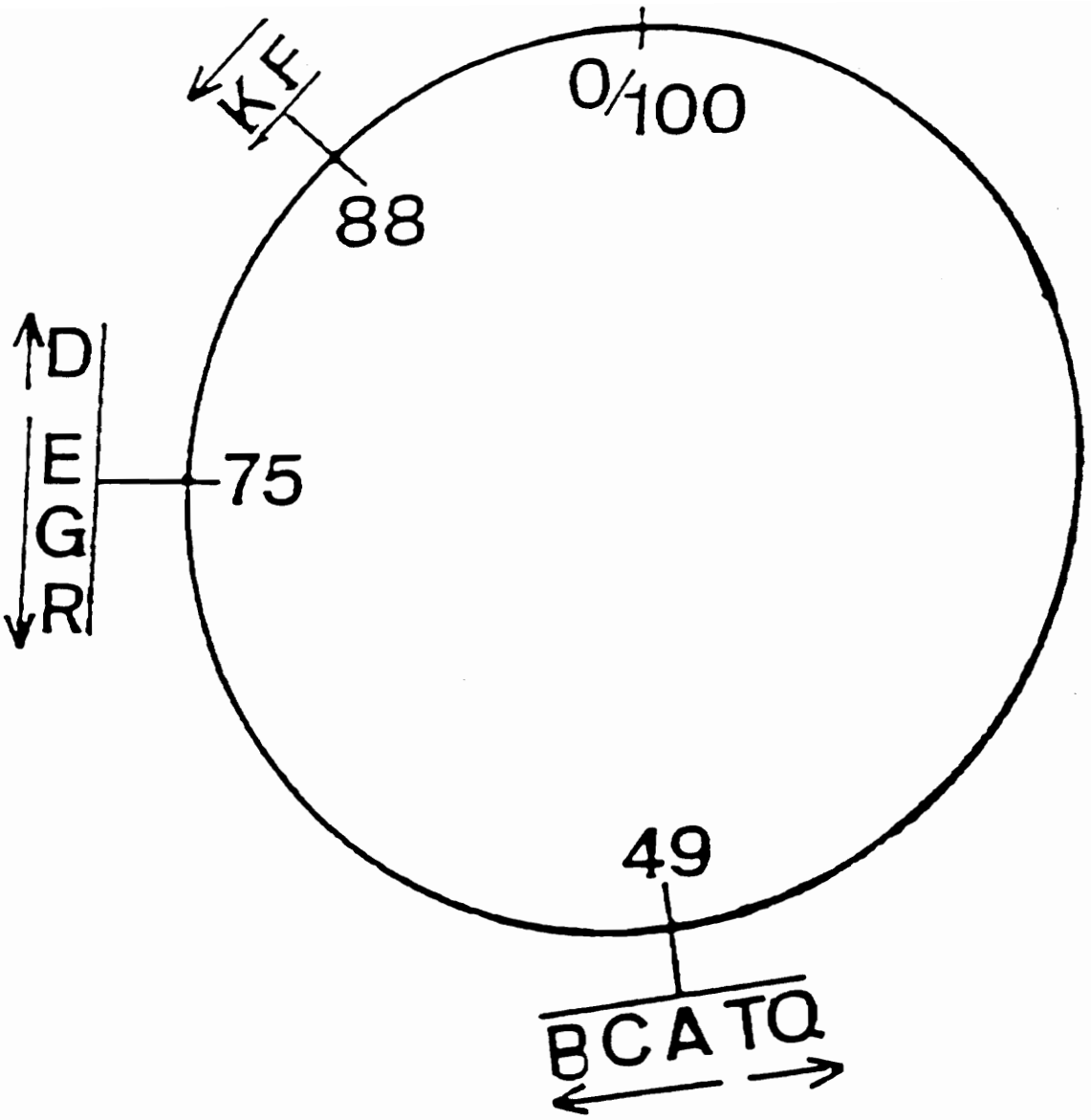

Figure 2. Genetic organization of the *glp* regulon of *E. coli*.

(GlpR) which requires the GlpE and GlpG proteins to function effectively (3-7). The inducer for the regulon is glycerol-P (3-5). When exogenous sources of glycerol-P are rich, the *glp* regulon is induced and the excess glycerol-P can be catabolized as carbon source; when the glycerol-P or its precursors are inadequate, the *glp* regulon is repressed, so that the cell can use almost all the glycerol-P available for synthesis of glycerophospholipid. This regulation prevents a futile cycle involving the catabolic enzyme and the biosynthetic enzyme (glycerol-P synthase encoded by *gpsA*) catalyzing the reverse reaction.

Glycerol and glycerophosphodiesters can be utilized by *E. coli* after conversion to glycerol-P. Exogenous glycerol enters the periplasmic space, and is then transported into the cytoplasm by glycerol diffusion facilitator protein which is encoded by *glpF* (8,9). The glycerol in the cytoplasm is phosphorylated by glycerol kinase, encoded by *glpK*, and thus trapped inside the cell as glycerol-P (10). GlpF is a membrane protein with a deduced MW of 29,780 (11). Glycerol kinase consists of four identical 55 kDa subunits. Each subunit contains a binding site for glycerol and FDP (12). FDP, a key metabolite in glycolysis, is a noncompetitive allosteric inhibitor for glycerol

kinase (13). When glucose is plentiful, *E. coli* will preferentially use glucose as carbon and energy source, and the inhibition of the glycerol kinase activity by FDP and nonphosphorylated form of enzyme III^{Glc} will avoid the overconsumption of glycerol. The maximum inhibition by FDP is about 80% with a Ki of 0.5 mM (13). For mutants lacking the kinase, glycerol can neither serve as a source of carbon and energy nor act as an inducer for the remaining genes of the *glp* regulon (3). So the true inducer is glycerol-P. Recently, a new gene, *glpX*, located directly downstream of the *glpK* gene was identified. GlpX has a calculated MW of 35,769, but its function remains unknown (14). *glpFKX* comprises an operon that is located at min 88 of the linkage map with *glpF* promotor proximal (9,14,15).

Extracellular glycerophosphodiesters (the deacylated products of phospholipids) are hydrolyzed to glycerol-P and alcohol in the periplasmic space (14). This reaction is catalyzed by a specific phosphodiesterase, the product of *glpQ* (16). The native enzyme is a dimer composed of two identical subunits (17). The periplasmic glycerol-P is actively transported into the cytoplasm by the glycerol-P permease, GlpT (18). This permease has a Km for glycerol-P of 12 μM (3). It is a glycerol-P/Pi antiporter that

mediates the influx of glycerol-P at the expense of cytoplasmic Pi (19,20). The apparent size of permease (44 or 33 kDa depending on the conditions for solubilization of the proteins prior to electrophoresis) (21) is smaller than that deduced from the nucleotide sequence (50,251) (22), probably due to the high content of hydrophobic amino acid residues in permease (21). The *glpTQ* operon maps at min 49 (23).

The cytoplasmic glycerol-P is converted to dihydroxyacetone phosphate by either the aerobic (GlpD) or the anaerobic dehydrogenase (GlpACB) (24). GlpA and GlpC are the catalytic subunits with molecular masses of 62 kDa and 44 kDa respectively while GlpB serves as membrane anchor and accepts the reducing equivalents from GlpAC in the electron transport chain (24,25). GlpB is a 43 kDa protein (24). Mutants lacking anaerobic glycerol-P dehydrogenase do not grow anaerobically on either glycerol or glycerol-P with fumarate as the hydrogen acceptor (26). The *glpACB* operon is adjacent to and is transcribed divergently from *glpTQ* (23). Aerobic dehydrogenase is a membrane associated enzyme consisting of two identical subunits (27). Each subunit has a MW of 56,747, as deduced from the nucleotide sequence of *glpD* (28). The protein has noncovalently bound

FAD as its coenzyme (28). The purified enzyme has a Km for glycerol-P of 0.8 mM (3). Cells missing aerobic glycerol-P dehydrogenase grow on neither glycerol nor glycerol-P in the presence of molecular oxygen (3). The *glpD* gene alone comprises an operon located at min 75 (4). This operon is adjacent to and is transcribed divergently from the *glpEGR* operon (6,29-31).

Results of early studies showed that *glp* repressor is a tetramer with four identical subunits under native conditions (6). The calculated subunit MW based on the nucleotide sequence of the *glpR* gene is 28,046 (25). Each subunit contains a binding site for glycerol-P exhibiting a dissociation constant of approximately 30 μM in the absence of DNA (32). The other two products of the *glpEGR* operon, GlpE and GlpG, are required for the repression of the *glp* regulon by GlpR (7). The apparent MW's of GlpE and GlpG are 13,000 and 26,000, respectively (31). This three component regulatory system is unique in *E. coli*. The mechanism by which GlpE and GlpG function in the regulation is not clear yet.

TRANSCRIPTIONAL REGULATION OF THE *glp* REGULON IN *E. coli*

The *glp* operons are controlled transcriptionally in response to three types of environmental stimuli: the presence of the inducer glycerol-P or its precursors, glycerol and glycerophosphodiesters; the presence of glucose (catabolite repression); and the presence of oxygen versus alternate terminal electron acceptors (4).

GlpR is the repressor for the *glp* regulon. Its affinity for operator sites in the control regions of the *glp* operons is decreased when the cytoplasmic level of glycerol-P increases (3-5). The degree of repression varies for the different *glp* operons with *glpFKX* controlled most tightly (11,33). The *glpR* gene is the distal gene in the *glpEGR* operon which is subject to autoregulation by three to four fold (7). The autoregulation is weaker than repression of the other operons in the *glp* regulon, allowing the appropriate amount of repressor present to exert its negative regulatory function in the absence of inducer.

Transcription of the *glp* operons is positively regulated by the cAMP-

cAMP receptor protein (CRP) complex (34). Catabolite repression occurs when cells are growing in the presence of glucose. Growth on glucose causes a decrease in the level of cAMP. The *glpTQ* and *glpFKX* operons are the most sensitive operons to catabolite repression (4,11).

The respiratory state of the cell affects the expression of the divergent *glpACB-glpTQ* operons and the *glpD* gene (4). Anaerobic glycerol-P dehydrogenase activity is maximal during anaerobic growth with fumarate present as electron acceptor (4). Under anaerobic conditions, the FNR protein has a positive influence on the transcription of genes involved in anaerobic respiration, including the *glpACB* operon (26). The FNR protein is structurally homologous to CRP and presumably activates transcription via a similar mechanism (35-37), in response to anaerobiosis. The signal for activation of FNR is thought to be the binding of a redox-sensitive metal cation (possibly ferrous iron) to a cluster of four cysteine residues near the amino terminus of FNR (38-40).

Aerobic glycerol-P dehydrogenase activity is maximal under well-oxygenated growth conditions (4). Anaerobiosis causes repression of the *glpD* operon as well as the other members of the aerobic respiratory control

modulon (*arc*) (41). The *arcA*/*arcB*-encoded two component regulatory system is responsible for anaerobic repression of these genes (41,42).

STRUCTURE AND FUNCTION OF REPRESSORS AND THEIR OPERATOR SITES

In 1961, Jacob and Monod proposed in their operon model that control genes encode repressors that turn off the structural gene(s). The isolation of nonsense mutations in the λ repressor *cI* gene (1962) and in the *lac* repressor *I* gene (1965, 1966) provided the first convincing genetic evidence for the nature of repressors. The isolation of *lac* repressor by Gilbert and Müller-Hill in 1966 brought the final proof (43). The repressor protein had dual properties: it prevented transcription by binding to a DNA sequence called the operator; and it recognized the small-molecule inducer. Thus, negative regulation of the operon is accomplished by the repressor protein, the operator site on the bacterial chromosome, and inducer (44). The binding sites on the repressor protein for DNA and inducer are separate but mutually

interacting: binding of inducer alters the affinity of repressor for operator, and vice versa (44). The interplay between the DNA and inducer binding sites is mediated by a conformational change in the repressor protein and is the substance of the regulatory mechanism (44). The operator has a feature common to many recognition sites for regulatory proteins: it is palindromic. Therefore, active repressors are usually multimeric. The operator usually overlaps the promoter region or lies between the promoter and structural genes. The binding of repressor will prevent RNA polymerase binding to the promoter or the formation of open binary complex by RNA polymerase at the promoter region. The binding of inducer changes the conformation of the repressor to a form that releases the operator.

LacR is a tetramer with four identical subunits, each with a MW of 38,000 (43). Each tetramer contains four binding sites for IPTG (inducer) (45). The binding sites for operator and inducer can be identified within the repressor subunit by mutations in the *lacI* gene that inactivate them. These mutations are found throughout the *lacI* gene. Three types of mutations were identified. Repressor-recessive mutations (*lacI*$^{-}$) that result from the inability of monomer aggregation are located between residues 220 and 290 (46).

Negative dominant mutations (*lacI*^{d}) are found at the 5' end of the gene (47,48). Repressors encoded by *lacI*^{d} alleles have normal oligomerization to form the tetramer, but cannot bind to the operator DNA, indicating that the DNA binding domain is at the N-terminus of the repressor. The third type of mutations are noninducible mutations (*lacI*^{s}), which are clustered into two regions, approximately between residues 80-120 and 190-280 (48). These are the inducer binding sites (48). The above genetic mapping results were supported by the protein chemical mapping of LacI. When the *lac* repressor is treated with trypsin, it is cleaved preferentially at amino acid 59. The C-terminal domain, known as the trypsin resistant core, contains amino acid residues 60-360. It retains the inducer binding and oligomerization abilities, but cannot bind to the operator DNA (49). The amino-terminal domain retains the DNA binding ability but with lowered affinity compared to the intact repressor (50). The *lac* repressor also nonspecifically binds to DNA other than operator with much lower affinity (equilibrium association constant for operator is 1×10^{13} M in the absence of inducer; for nonspecific DNA, the association constant is 1×10^{6} M) (44). The *lac* repressor-inducer complex is altered only in the specific recognition of operator, and not in the

binding to the backbone of the DNA (44).

The nucleotide sequence of the *lac* operator was deduced by Gilbert and Maxam in 1973 (52). The inverted repeats of the *lac* operator are not quite identical. Base substitutions at the right half site that result in a fully symmetric operator (ideal operator) increase the binding affinity of the repressor; while the mutations that results in asymmetry of the operator decrease the affinity (51). Further studies showed that in addition to the primary operator (O_1), there are two other operators in the *lac* operon, O_2 and O_3 (54). O_2, in the *lacZ* gene, is 401 bp downstream from O_1 with about 0.2 of the affinity of O_1 ; O_3, in the *lacI* gene, is 92 bp upstream from O_1 and the binding affinity is 0.06-0.001 of binding to O_1 (53). However, O_2 and O_3 are necessary for obtaining complete repression (about 6700 fold repression). The *lac* repressor tetramer could bind to O_1 and O_2 or O_2 and O_3 simultaneously, with intervening DNA forming a loop (54). The spacing between the operators, the formation of tetramers and supercoiling of the DNA play important roles in the loop formation.

λ repressor is an acidic protein whose monomer has a MW of 26,228 (55). Each monomer has two distinct domains connected by a string of 40

amino acids (56). These monomers are in concentration-dependent equilibrium with dimers (Kd of 10^{-8} M) (55). When treated with protease, the two domains of the repressor monomer are released as two separate fragments. The N-terminal domain is comprised of residues 1-92, and has the operator binding ability although the affinity is lower than that of the repressor; the C-terminal domain, residues 132-236, is responsible for forming the dimer (57). Thus, the λ repressor is similar to *lac* repressor in the functional organization although there is no overall amino acid sequence similarity. There are two operator regions on the λ phage chromosome. Each operator has three contiguous sites for repressor binding. λ repressor binds to subsequent sites within each operator in a cooperative manner . The presence of a dimer at site 1 greatly increases the affinity with which a second dimer can bind to site 2 (57). If site 1 is inactive, then repressor binds cooperatively to site 2 and site 3 (57). This interaction occurs directly between repressor dimers and not via conformational change (57). When repressor concentration is low, repressor dimers interact at O_{L1}, O_{L2}, and O_{R1}, O_{R2}, turn off the lytic gene expression by blocking access of RNA polymerase to the corresponding promoters (P_L and P_R). The binding of repressor dimers at O_{R1}-O_{R2} also activates the transcription of its own product

presumably by protein-protein interaction between repressor and RNA polymerase bound at P_{RM} which is adjacent to and transcribed divergently from P_R. When repressor concentration is high, it binds to O_{R3} which overlaps P_{RM}, and thus turns off its own gene expression. This autoregulation allows maintenance of λ repressor at an appropriate level (57).

The *gal* repressor has two identical 38 kDa subunits (58). Each monomer contains two domains. Like other repressors, the amino terminal domain binds to DNA and the carboxyl domain binds to inducer and is responsible for monomer aggregation (59).

The *deo* repressor is an octamer of identical subunits. The inducer for DeoR is deoxyribose-5-phosphate (60). The DNA-binding motif is located within the N-terminal domain of the repressor (61). Three *deo* operators are present within the *deo* operon. Two of them ($DeoO_1$ and $DeoO_2$) overlap the Pribnow boxes of the two *deo* promoters while the third ($DeoO_E$) is located 279 bp upstream of $DeoO_1$ (62). Two operator sites are required for efficient repression of either promotor. The distance between the operator sites can be varied from 46 bp to several kb and efficient DeoR repression is still maintained (63). Simultaneous DeoR binding to three operator sites,

resulting in formation of double loops, has been observed in electron micrographs (64). Recent studies (62) showed that a 16 bp palindromic sequence constitutes the *deo* operator. Positions outside this palindrome can be changed without any major effect on DeoR binding. Most of the central 6-8 bp of the palindrome (positions ± 1, ±2, ±3) can be substituted with any other nucleotide without significant effect on repressor binding; while changes at position ±4 and ±5 have a varied effect depending on the nucleotide. Positions ±6, ±7 and ±8 are most critical for repressor binding.

GlpR is a tetramer of identical subunits. It negatively controls the *glp* regulon by binding to the operator sites in the *glp* operons. A total of thirteen operators were identified in the *glp* operons (11). They match more or less well the consensus operator: 5'-WATGTTCGWT AWCGAACATW-3' (W = A or T) (11). Tandemly repeated repressor binding sites are present in the control regions of the *glpACB*, *glpD* and *glpFKX* operons (11,14,33,65). A single operator overlapping the CRP site was found in the *glpTQ* operon (33). The *glp* operons exhibit differential sensitivity for repressor. The difference may be due to differences in the degree of match with consensus, the number of operators and the position of operators with respect to the promoter regions. Previous study suggested that *glpACB*

operators are involved in control of *glpTQ* expression perhaps via formation of a DNA loop (33). DNA repression loop formation is also proposed for the *glpFKX* operon (11,14).

In this study, the critical positions in *glp* operator sites for repressor binding were identified and the cooperativity between two adjacent operators for repressor binding was investigated (Part I).

STRUCTURE AND FUNCTION OF THE HTH DNA BINDING PROTEIN

DNA-binding proteins play a central role in biology. Structural sequence comparison made it possible to classify the DNA-binding proteins into groups that use related structural motifs for recognition (66). The proteins that have the helix-turn-helix (HTH) DNA recognition motif are called HTH proteins (66). Among them, are LacR, GlpR, DeoR, CRP and FNR (66). The second helix in the HTH motif, the recognition helix, plays an important role in recognizing small, specific target sites among a large

excess of other DNA sequences (66).

The specificity of protein-DNA interactions is primarily due to complex networks of hydrogen bonds and van der Walls interactions that require both the close contact and proper orientation of the amino acid residues and nucleotide bases. To distinguish a specific DNA sequence from other nonspecific DNA sequence, proteins must make multiple specific contacts with individual base pairs. This is achieved by forming multiple hydrogen bonds between amino acid side chain(s) and nucleotide(s) as well as van der Walls interactions between amino acid side chains and the edge of nucleotide base pairs.

Sequence comparisons showed the most highly conserved residue in the HTH motif include a glycine in the turn and several hydrophobic residues (66).

There are several lines of evidence that suggest that the recognition helix is the major determinant of binding specificity. One of them is the conversion of the DNA-binding specificity of bacteriophage 434 repressor (434R) to P22 repressor (P22R) by construction of an appropriate hybrid $434R^{P22R}$ (67). This hybrid repressor was constructed by replacing five

amino acids of 434R with those derived from P22 repressor, which presumably face the outside surface of the helix in the recognition helix of 434 repressor (68). This indicated that the outside surface of the recognition helix is responsible for recognition of the specific sequence of DNA. Another example involves CRP and FNR proteins. CRP and FNR are both transcriptional activators and the consensus DNA site for CRP binding is similar to that for FNR (35). Moreover, the amino acid sequences of these two proteins are homologous, especially in the DNA-binding motif. Substitution of three amino acids at the outside surface of the FNR putative recognition helix by those found at the equivalent position in CRP results in a CRP-like activity of FNR (36). Altered FNR interacts with DNA such that activation of CRP-dependent gene expression occurs under anaerobic conditions and catabolic repression is lost. Furthermore, 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 (37).

The first helix in the HTH motif also plays an important role in the recognition of DNA operator sites. The structures of several repressor-operator complexes have been determined at high resolution (66). Both

helixes in the HTH motif make critical contacts with the specific DNA (66).

LacR of *L. lactis* is the repressor protein of lactose operon, and DeoR, GutR and FucR are the repressor proteins of the deoxyribonucleoside, glucitol and galactitol, and fucose operons of *E. coli*. The high overall homology of amino acids between these catabolic regulatory proteins suggested that these proteins are evolutionarily related (69). Amino acid sequence comparisons also identified two highly conserved regions for all four proteins (5). One is the HTH DNA-binding motif which is located in the N-terminal domain while the other in the C-terminal domain is probably the inducer-binding site (Fig. 6) (69).

Early studies on *glp* repressor resulted in identification of a potential recognition helix of the DNA-binding domain (25). This was done by comparing the amino acid sequence of *glp* repressor with that of the *deo* repressor (*deo* repressor binds to a very similar operator sequence). Seven of ten amino acids in the putative recognition helixes of the two repressors are identical, and there is only one nucleotide difference found (position 4) when comparing the two operator half sites. Mutations in the putative recognition helix greatly decrease the binding affinity of mutant repressor for

glp operator. Unexpectedly, an altered GlpR (PQ→ EM) which has almost the same recognition helix as that of the *deo* repressor cannot bind to either of the operators (25). This implicates the involvement of the first helix or even the residues outside the HTH motif in the recognition of operator DNA.

The monomer of repressor usually consists of two domains: one is the DNA-binding domain which is responsible for the recognition of operator DNA; the other domain is responsible for the inducer binding and oligomerization. The alterations in the repressor which result in either an increased operator binding affinity or a lowered inducer binding affinity will confer a noninducible phenotype upon the corresponding regulon.

Cozzarelli et al. isolated a mutant strain, ECL89, which was generated by chemical mutagenesis (5). This strain has a noninducible phenotype with regard to the members of the *glp* regulon so that it cannot grow on either glycerol or glycerol-P. In order to analyze the molecular basis for the noninducible phenotype of this strain, the nucleotide sequence of the *glpR12* gene from this strain was determined. The inducer binding activity was measured for both the wild-type and noninducible repressors (Part II).

MATERIALS AND METHODS

***Materials*:** Oligonucleotides were synthesized using an Applied Biosystems model 381A DNA synthesizer and purified by Oligonucleotide Purification Cartridges (Cruachem) as recommended by the manufacturer. In some cases, mixtures of phosphoramidite bases were coupled to generate multiple base substitutions at the target site. The sequences of the synthetic oligonucleotides are listed in Tables 1, 2 and 3. Restriction endonucleases were from U.S. Biochemical, Promega, New England Biolabs or American Allied Biochemical, Inc. T4 DNA ligase was obtained from New England Biolabs. Magic minipreps DNA purification kit, Magic PCR preps DNA purification kit, gel drying kit, pGEM-T vector system and pGEM3Z were from Promega. Sequenase version 2.0 DNA sequencing kit, X-gal and IPTG were purchased from U.S. Biochemical. GeneAmp PCR kit and GeneAmp PCR system 9600 were from Perkin Elmer. Geneclean II kit for DNA extraction from agarose gels was provided by BIO 101. [α-^{35}S] dATP and [^{3}H] glycerol-P were supplied by Du Pont-New England Nuclear. Yeast

Table 1. Oligonucleotides used for creation of *glp* operators

Operator	Nucleotide sequence[1]	Serial#
glp consensus	GATCCTATGTTCGATATCGAACATACTGCA	275818
	GATACAAGCTATAGCTTGTATG	256166
7,7′	GATCCTAT(AC)TTCGATATCGAA(TG)ATACTGCA	275566
	GATA(TG)AAGCTATAGCTT(AC)TATG	256452
4,4′ A,T (*deo* operator)	GATCCTATGTTAGATATCTAACATACTGCA	275847
	GATACAATCTATAGATTGTATG	214634
1	GATCCTATGTTCGA(ACG)ATCGAACATACTGCA	85167
	GATACAAGCT(TGC)TAGCTTGTATG	122501
2	GATCCTATGTTCG(TCG)TATCGAACATACTGCA	93932
	GATACAAGC(AGC)ATAGCTTGTATG	122533
3	GATCCTATGTTC(ATC)ATATCGAACATACTGCA	46144
	GATACAAG(TAG)TATAGCTTGTATG	122891
4	GATCCTATGTT(ATG)GATATCGAACATACTGCA	257659
	GATACAA(TAC)CTATAGCTTGTATG	289734
5	GATCCTATGT(ACG)CGATATCGAACATACTGCA	244944
	GATACA(TGC)GCTATAGCTTGTATG	225657
6	GATCCTATG(ACG)TCGATATCGAACATACTGCA	12829
	GATAC(TGC)AGCTATAGCTTGTATG	104690
7	GATCCTAT(ATC)TTCGATATCGAACATACTGCA	45034
	GATA(TAG)AAGCTATAGCTTGTATG	122853
8	GATCCTA(ACG)GTTCGATATCGAACATACTGCA	263184
	GAT(TGC)CAAGCTATAGCTTGTATG	289266
9	GATCCT(TCG)TGTTCGATATCGAACATACTGCA	45130
	GA(AGC)ACAAGCTATAGCTTGTATG	122917
glp Tandem	ATCGAACATATATGTTCGAT	261968
deo Tandem	ATCTAACATATATGTTAGAT	262093
glp Tandem (-2)	ATCGAACATATGTTCGAT	260262

glp Tandem (+2)	ATCGAACATATATATGTTCGAT	249305
glp Tandem (+4)	ATCGAACATATCGATATGTTCGAT	249705
glp Tandem (+6)	ATCGAACATATCGCGATATGTTCGAT	259967

[1] The top strand is written 5′ → 3′ for the duplexes. Parentheses indicate that oligonucleotides with a mixture of the indicated bases were synthesized.

Table 2. Synthetic oligonucleotide primers used for sequencing of the *glpR* gene

Serial #	Nucleotide sequence (5′ → 3′)	Position in sequence[1]
202884	GATTCGCTCAATGCGCG	1545 to 1561
117273	GCCCGCAGACTATTCGCC	1685 to 1702
290516	GCCCGCAAAGTGGCGG	1834 to 1849
117265	CATTGGCGAAGCGACGCTC	2043 to 2061
267777	GCCGTCGCTATCGATG	2127 to 2112
117241	GCCATTATTGAGAACTCGCGCC	2170 to 2191
258981	GCTGTGCTGATCCTGCA	2340 to 2356

[1] The numbers represent the positions in *glpEGR* sequence (25) in which the *glpR* S.D. sequence is located between 1579 to 1583 (Fig. 4).

Table 3. Synthetic oligonucleotide primers used for PCR

Serial#	Nucleotide sequence (5′ → 3′)[1]	Position in sequence[2]
	*Eco*RI	
82445	ATGC<u>GAATTC</u>TCGGCGATCAGG	1074 to 1095
	*Sal*I	
82930	TCTC<u>GTCGAC</u>CGTATGCAACTG	2529 to 2508
	*Eco*RV	
28423	ATC<u>GATATC</u>GGCA*T*CACGCCGG	1879 to 1900
	*Eco*RV	
28463	ATC<u>GATATC</u>GGCACCACGCCGG	1879 to 1900

[1] Mismatched bases are written in italics.

[2] The numbers represent the positions in *glpEGR* sequence (25) in which the *glpR* S.D. sequence is located between 1579 to 1583 (Fig. 4).

extract, tryptone and bacto agar were from Difco Laboratories. Antibiotics, amino acids, protein molecular weight standards, ONPG and MOPS were from Sigma Chemical Company. Chloroform, 2-propanol, isoamylalcohol, phenol, agarose, acrylamide and other chemicals were from Fisher Scientific Company. Fast Stain for SDS gels was from Zoion Research, Inc.

***Bacterial strains and growth media*:** The bacterial strains utilized were *E. coli* K-12 derivatives and are shown in Table 4. Preparation of phage lysates and transductions with P1*vir* were done in this laboratory according to the method described by Silhavy et al. (70). Strain DH5αF' was used as host during construction of recombinant plasmids. LB (70) medium was used for growth of the cells. Ampicillin (100 μg/ml) was added where appropriate. The lactose phenotype of strains carrying *glp-lacZ* fusions was determined on LB + Amp + X-gal (40 μg/ml) medium. Glycerol-P (0.02%) was added to the medium where indicated.

***Isolation of DNA*:** Plasmid DNA was isolated using Magic minipreps DNA purification system according to the protocol provided by the manufacturer. The plasmid DNA isolated from 1.5 ml of overnight culture was dissolved

Table 4. Strains of *E. coli* K-12 used in this study

Strain	Genotype	Source/Derivation
DH5αF′	F′ ϕ80d*lacZΔM15 recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 Δ(lacZ-argF)U169*	Bethesda Res.Labs.
SY102	MC4100 *glpR2 ϕ(glpD-lacZ)hyb* λ*plac*Mu *recA1 srl*::Tn*10*	S. Ye (25)
ECL89	HfrC *glpR12 (glpR^{n}) phoA8 fhuA22 ompF627 fadL701 relA1 pit-10 spoT1*	E. C. C. Lin (5)
MC4100	F^{-} *araD139 (argF-lac)U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	M. J. Casadaban (71)
TS100	MC4100 *glpR2*	T. J. Silhavy (21)
SH305	MC4100 *ΔglpD102 recA1 srl*::Tn*10*	H. Schweizer (30)
TST3	MC4100 *malT*::Tn*10*	T. J. Silhavy (71)
TL681	MC4100 *ϕ(glpA101-lacZ)* λp1(209) *ΔglpD102 sdh-9*	T. J. Larson (23)
GD4	MC4100 *glpR2 ϕ(glpD-lacZ)hyb* λp1(209) *zih-730*::Tn*10*	G. Sweet
GD6	TS100 *ϕ(glpD-lacZ)hyb* λp1(209)	P1(GD4)-->TS100 (Lac^{+} selection)
TJS1	ECL89 *malT*::Tn*10*	P1(TST3)-->ECL89 (Tet^{r} selection)
TJS51	GD6 *glpR12 malT*::Tn*10*	P1(TJS1)-->GD6 (Tet^{r} λimm selection)
TJS52	TJS51 *malT*$^{+}$	P1(TL681)-->TJS51 (Mal^{+} selection)
WO331[1]	TJS52 *recA1 srl*::Tn*10*	P1(SH305)-->TJS52 (Tet^{r} selection)

[1] This strain had been constructed in this laboratory.

in 40 μl dH_2O. Isolation of chromosomal DNA was carried out as described by Silhavy et al. (70).

Preparation of competent cells (70)**:** A fresh overnight culture was diluted 100 fold into 50 ml of fresh prewarmed LB medium and grown at 37°C to A_{600} = 0.6-0.8. Cells were centrifuged at 4°C and washed with 25 ml of ice cold 0.1 M $MgCl_2$. After centrifugation again, the pellet was resuspended in 15 ml ice cold T-salts (75 mM $CaCl_2$, 6 mM $MgCl_2$) and incubated on ice for 20 to 30 minutes. Then, the suspension was centrifuged as above, and the pellet was resuspended in 1.5 ml ice cold T-salts and held on ice for 20 minutes. The cells were then competent and ready for use. For storage, glycerol was added to a final concentration of 15%, and the aliquoted cells were kept at -70°C.

***Transformation*:** Transformation of *E. coli* cells was done according to a standard protocol (70). Plasmid DNA (1 μl) from minipreps or 2-7 μl DNA from a ligation mixture was mixed with 100 μl *E. coli* competent cells on ice for 30 to 45 minutes and the cells were subject to a heat shock at 42°C for

2 minutes. LB (1 ml) was added to the mixture and incubation continued at 37°C for 1 hour. One tenth or all (if the DNA from a ligation mixture was transformed) of the cells was spread on a selective medium plate.

***Agarose gel electrophoresis*:** Horizontal gels containing 0.8-1.0% agarose were used to separate the restriction fragments and purify the PCR products. Low melting temperature gels were used if DNA extraction was performed after electrophoresis. The high melting temperature agarose gels were run at room temperature while the low melting temperature gels were run at 4°C. Electrophoresis was done at constant voltage (10-12 v/cm) with TBE buffer (0.05 M Tris-HCl, 0.05 M boric acid, 0.002 M EDTA, pH 8.0) and EtBr (0.5 μg/ml) in both cases. A vertical DNA polyacrylamide (8%) gel was used when 41 bps needed to be distinguished from 61 bps.

***Polymerase Chain Recation (PCR)*:** The polymerase chain reaction was used to amplify DNA fragments, to create desired mutations in the resulting products, or to incorporate appropriate restriction site(s) for cloning. The reaction was carried out using the GeneAmp PCR kit on a GeneAmp PCR System 9600 according to the manufacturer's instructions. The conditions

for PCR were varied with the template and primers used. Chromosomal DNA (1 μg) or 100 ng plasmid DNA and 100 pmol of each primer were used in 100 μl reaction mixtures. The cycles were set up as follows for amplification of WO331 chromosomal DNA: Initial melting of template at 94°C for 3 min, followed by 30 cycles including melting at 94°C for 30 s; annealing at 57°C for 40 s and extension at 72°C for 1 min 20 s. The conditions for amplifying plasmid DNA fragments in this study were: Denaturation of template DNA at 94 °C for 3 min, then 30 cycles including melting at 94°C for 30 s; annealing at 50 °C for 30 s and extension at 72°C for 30 s. Ten microliters of each reaction mixture was run on a 0.8% agarose gel to check the size and purity of the product after the reaction was done.

***Endonuclease digestion*:** Restriction enzyme digestion took place in a 20 μl reaction mixture containing the appropriate digestion buffer. In the cases of double digests, half of the buffer for each enzyme was used. The reaction mixtures were incubated at the appropriate temperature for 3 hours to overnight in the case of complete digestion. Partial digestion of pSY2-6 with 2 U *Eco*RV was achieved after incubating the digestion mixture at 37°C for

15 min.

***DNA extraction and ligation*:** Restriction fragments or PCR products were extracted from low melting temperature agarose gel using the Geneclean II kit or Magic PCR Preps DNA purification kit and eluted in 5-8 μl (for the Geneclean kit) or 8-15 μl dH_2O (for Magic PCR preps). Ligations were done in 15 μl reaction mixtures containing 10 U T4 DNA ligase and the appropriate molar ratio of vector and insert DNA (1:3-4 for blunt end ligation and 1:1 for sticky end ligation). The ligation mixtures were incubated at 16-18°C overnight and then used for transformation.

***DNA sequencing*:** Double stranded plasmids were used as templates for the dideoxynucleotide chain termination method (72). Sequencing reactions were made with the T7 primer or synthetic primers according to the protocol in the sequencing kit. Reaction mixtures (3.5 μl) were analyzed on wedge-shaped gels containing 6% polyacrylamide and 7 M urea. Gels were prerun for 15 to 30 minutes before loading the samples. Electrophoresis was carried out in TBE buffer at constant temperature (~50°C) by application of constant power (~60 W) for 1.5 h (to read up to 250 nucleotides) or 5 h (to read up

to 350 nucleotides). After electrophoresis, the gels were fixed, dried and then exposed to film for 18-36 h.

***Preparation and cloning of operator DNA*:** Operator duplexes were prepared with *Bam*HI and *Pst*I sites present on either end for cloning. An *Eco*RV site was present at the center of symmetry to aid in identification of recombinant plasmids conatining operator DNA and for insertion of an additional duplex to generate plasmids with tandemly repeated operators (Table 1). For cloning of single operator DNA, plasmid vector pGEM3Z was digested with *Bam*HI and *Pst*I and purified following electrophoresis on a low melting temperature agarose gel. For construction of tandem operators, the plasmid containing the single *glp* consensus operator was digested with *Eco*RV. Equimolar amounts of the oligonucleotides to be cloned were mixed (the oligonucleotides for construction of tandem operators were self-annealed) in 10 mM Tris-HCl (pH 8.0), 50 mM NaCl and 1 mM EDTA and heated to 85 °C. The annealing mixture was allowed to cool to room temperature during several hours. A forty-fold molar excess of annealed oligonucleotide was mixed and ligated to 150 ng of cleaved vector

DNA in 15 μl. An aliquot of the ligation mixture was used for transformation of strain DH5αF'. The presence of the desired insert was verified phenotypically (white colony on LB + Amp + X-gal medium), by restriction analysis and by nucleotide sequence analysis.

***Construction of recombinant DNA*:** The plasmids constructed in this study are listed in Table 5. PCR products were cloned into pGEM-T directly for construction of pNZ101, pNZ101-A and pNZ101-B. pNZ80, pNZ81, pNZ83, pNZ102 and pNZ103 were consructed by substitution of appropriate restriction fragments.

***Assay of β-galactosidase activity*:** The β-galactosidase activity was determined in duplicate (or triplicate) as described by Miller (73) with some modifications. A single colony was inoculated into 1.5 ml LB and aerated overnight at 37 °C. The overnight culture was diluted 30 to 40 fold into 1.5 ml LB medium and grown until the A_{578} reached 0.28-1.0. The cells were cooled on ice for ten minutes to stop growth. Then, the cells were harvested in microfuge tubes and resuspended in 1 ml Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH

Table 5. Plasmids constructed in this study

Plasmid	Insert (Position in the sequence[1])	Source of insert (or PCR tempelate)	vector[2]/ (site used)
pNZ80	*Bgl*II-*Nsi*I (272-1192)	pDA910[3]	pSH79 (*Bam*HI-*Nsi*I)
pNZ101	PCR product (1074-2529)	WO331 DNA	pGEM-T
pNZ101-A (T101→I)	PCR product[4] (1879-2529)	pNZ80	pGEM-T
pNZ101-B (P147→R)	PCR product[5] (1879-2529)	pNZ101	pGEM-T
pNZ102	*Eco*RV-*Sal*I (1884-2523)	pNZ101	pNZ80 (*Eco*RV-*Sal*I)
pNZ81	*Eco*RV-*Sal*I (1884-2523)	pNZ101-A	pNZ80 (*Eco*RV-*Sal*I)
pNZ83	*Eco*RV-*Sal*I (1884-2523)	pNZ101-B	pNZ80 (*Eco*RV-*Sal*I)
pNZ103	*Eco*RV-*Hinc*II (1884-2529+77)[6]	pNZ101	pSY2-6 (*Eco*RV-*Eco*RV)[7]

[1] The numbers represent the positions in *glpEGR* sequence (25) in which the *glpR* S.D. sequence is located between nucleotides 1579 to 1583 (Fig. 4).

[2] The vector in pNZ80, pNZ81, pNZ83 and pNZ102 was Bluescript KS M13+ (Stratagene); in pNZ101, pNZ101-A and pNZ101-B, the vector was pGEM-T; in pNZ103 and pSY2-6, the vector was pSY223 (25). pSY223 is an expression vector with the *tac* promoter, derived from pKK223 (Pharmacia).

[3] pDA910 was constructed by D. Austin and contains an engineered *Bgl*II site between the divergent *glpD* and *glpEGR* promoters at position 272 (7); pSH79 was constructed by H. Schweizer (25) and contains the 3 Kb *Eco*RI-*Sal*I fragment of pSH21 (6).

[4] *glpR*$^+$ DNA was used as PCR template, and oligonucleotide 28423 which contained the mutation resulting in T101 → I was used together with oligonucleotide 82930 as PCR primers (Table 3).

[5] *glpR12* DNA which contained two mutations resulting in T101 →I, P147 → R was used as PCR template, and oligonucleotide 82463 which perfectly matches *glpR*$^+$ DNA in the region encoding amino acid 101 was used together with oligonucleotide 82930 as PCR primers (Table 3).

[6] The *Hinc*II site is located at position 77 of pGEM-T vector.

[7] pSY2-6 was partially digested with *Eco*RV and the resulting fragment with deletion of 0.8 kb (the 3′ end of *glpR*) was used as the vector in this case.

7.0). The A_{578} was measured for each sample. Appropriate volumes of cells (in duplicate or triplicate) were diluted into Z buffer to a final volume of 800 μl and then permeabilized with 1 drop of 0.1% sodium dodecyl sulfate (SDS) and 2 drops of chloroform followed by a 10 s vortex. The cells were preincubated for 10 minutes at 28°C. Reactions were started by adding 160 μl of ONPG (4 mg/ml in 0.1 M phosphate buffer, pH 7.0) and stopped by adding 0.4 ml of 1 M Na_2CO_3 when appropriate yellow color (A_{420} = 0.4-0.8) developed. The reaction mixtures were centrifuged at room temperature for 2 min, and the A_{420} of each supernatant was read versus a blank containing no cells. The specific activity of β-galactosidase was expressed in Miller units:

$$\text{Specific activity} = \frac{A_{420} \times 800}{A_{578} \times \text{Cell volume assayed (ml)} \times \text{Reaction time (min)}}$$

***Overproduction of the* glp *repressor*:** Overproduction of the *glp* repressor was achieved by the use of pSY2-6 (wt) and pNZ103 (noninducible phenotype) where the *glpR* genes are cloned downstream from the *tac* promoter in pSY223 (25). An overnight culture of SY102 (0.5 ml)

harboring either pSY2-6, or pNZ103 or pSY223 was inoculated into 50 ml LB and incubated at 37°C. When the A_{600} reached 0.6-1.0, IPTG was added to the cells to a final concentration of 0.5 mM and incubation continued for 2 h. Cells were harvested and resuspend in TNED (20 mM Tris-HCl, 50 mM NaCl, 0.1 mM EDTA and 0.2 mM DTT, pH 7.5). Serine protease inhibitor PMSF (final concentration 0.4 mM) was added and cells were lysed using a French press. Then the suspension was centrifuged at 15,000 rpm at 4°C for 0.5 h and the supernatant fraction (crude extract) containing GlpR was collected. Protein concentrations were measured and the samples were analyzed by SDS gel electrophoresis. Glycerol-P (inducer) binding activity was determined.

***Determination of protein concentration*:** Protein concentrations were determined by the Bradford method (74) using BSA as the standard.

***SDS gel electrophoresis*:** Protein samples were analyzed by electrophoresis on 10% SDS-polyacrylamide gels using the buffer system of Laemmli (75) at a constant voltage (~200 V) for 3-4 hours. Protein bands were visualized with Fast Stain and gels were dried using a gel drying kit.

***Glycerol-P binding assay*:** Appropriately diluted crude extracts (50 μl) were mixed with 20 μl of 0.25 M MOPS (pH 6.5), 10 μl of 4 mg/ml casein and 20 μl of 0.5 mM ^{3}H-glycerol-P (110 cpm/pmol) in a 1.5 ml microfuge tube and incubated at room temperature for 5 min. Then, 0.3 ml saturated $(NH4)_2SO_4$ was added and samples were held on ice for 5 min. The reaction mixtures were centrifuged at 4°C for 5 min. The pellets were washed with 0.5 ml saturated $(NH_4)_2SO_4$ and centrifuged for another 2 min at 4°C. The final pellets were resuspended in 0.1 ml water, and the radioactivities were determined by scintillation counting.

PART I

in vivo STUDIES ON

THE glp REPRESSOR-OPERATOR INTERACTION

RESULTS

In order to identify the critical position(s) of the *glp* operator for repressor binding, synthetic *glp* single operator and various *glp* operators substituted singly or symmetrically at positions 1-9 bases removed from the center of operator symmetry were cloned into the multicopy plasmid pGEM3Z. The cooperative binding of *glp* repressor at two adjacent operator sites was studied using synthetic *glp* tandem consensus operators cloned into the same vector. The binding affinity of repressor for each operator was estimated in *vivo* using strain WO331 [ϕ(*glpD-lacZ*) *hyb recA1 glpR*^{n}]. The *glpR*^{n} allele (noninducible repressor) confers a Lac$^-$ phenotype on this strain which is changed to Lac$^+$ upon introduction of a multicopy plasmid harboring a *glp* operator. Thus, the level of β-galactosidase activity obtained from the *glpD-lacZ* fusion (titration level) is proportional to the binding affinity of repressor for the operator carried by the plasmid.

***Binding of* glp *repressor to consensus and altered operators*:** The

binding affinity of *glp* repressor for various operators was estimated as described above. The pGEM3Z vector without a *glp* operator was used as the negative control. Strain WO331 harboring the vector yielded 36 Miller units, while WO331 harboring the *glp* single consensus operator gave 115 Miller units (Table 6). Substitutions at critical positions of the *glp* operator were expected to decrease markedly the binding affinity for *glp* repressor. Decreased affinity would be reflected by a decreased β-galactosidase specific activity (compared with that of the *glp* single consensus operator). Since highly conserved positions of the *glp* operator are most likely the important positions, *glp* operators with symmetrical alterations at positions 4 and 7 (these are the most highly conserved positions) were constructed first. The results (Table 6) showed that these altered operators cannot derepress the chromosomal *glpD-lacZ* fusion as indicated by the low activity of β-galactosidase (<36 Milller units). Thus, the symmetrical changes at these two positions abolished specific binding of *glp* repressor to these operators. Considering that single alterations in operator sequence might have smaller effects, and that differential influences resulting from substitution at each position might be easier to compare, *glp* operators were constructed with all

Table 6. Binding of *glp* repressor to consensus and altered operators[1]

Operator name	Sequence[2]	β-Galactosidase[3]
pGEM3Z vector	no operator sequence	36 (8)
Consensus, single	TATGTTCGAT ATCGAACATA	115 (11)
4,4′ A,T (*deo*)	TATGTTAGAT ATCTAACATA	29 (4)
7,7′ A,T	TATATTCGAT ATCGAATATA	17 (2)
7,7′ A,G	TATATTCGAT ATCGAAGATA	20 (2)
7,7′ C,G	TATCTTCGAT ATCGAAGATA	24 (2)
7,7′ C,T	TATCTTCGAT ATCGAATATA	18 (2)
1 A	TATGTTCGAA ATCGAACATA	132 (2)
1 C	TATGTTCGAC ATCGAACATA	45 (3)
1 G	TATGTTCGAG ATCGAACATA	70 (3)
2 C	TATGTTCGCT ATCGAACATA	55 (2)
2 G	TATGTTCGGT ATCGAACATA	61 (3)
2 T	TATGTTCGTT ATCGAACATA	93 (2)
3 A	TATGTTCAAT ATCGAACATA	34 (3)
3 C	TATGTTCCAT ATCGAACATA	48 (3)
3 T	TATGTTCTAT ATCGAACATA	37 (3)
4 A	TATGTTAGAT ATCGAACATA	29 (2)
4 G	TATGTTGGAT ATCGAACATA	29 (2)
4 T	TATGTTTGAT ATCGAACATA	25 (2)
5 A	TATGTACGAT ATCGAACATA	22 (2)

5	G	TATGTGCGAT ATCGAACATA	23 (2)
5	C	TATGTCCGAT ATCGAACATA	30 (3)
6	A	TATGATCGAT ATCGAACATA	39 (5)
6	G	TATGGTCGAT ATCGAACATA	62 (6)
6	C	TATGCTCGAT ATCGAACATA	138 (4)
7	A	TATATTCGAT ATCGAACATA	25 (2)
7	C	TATCTTCGAT ATCGAACATA	34 (2)
7	T	TATTTTCGAT ATCGAACATA	98 (6)
8	A	TAAGTTCGAT ATCGAACATA	32 (2)
8	G	TAGGTTCGAT ATCGAACATA	31 (3)
8	C	TACGTTCGAT ATCGAACATA	44 (2)
9	G	TGTGTTCGAT ATCGAACATA	81 (5)
9	C	TCTGTTCGAT ATCGAACATA	62 (6)
9	T	TTTGTTCGAT ATCGAACATA	73 (5)

[1] The plasmids carrying operator variants had been constructed by other students in our laboratory. The results of β-galactosidase specific activities are from this study and previous studies in this laboratory.

[2] The mutated bases in the sequence are underlined.

[3] The numbers in parentheses are the numbers of independent duplicate or triplicate β-galactosidase activity assays.

possible single substitutions at positions 1-9. The ability of the various *glp* operators to titrate chromosomally-encoded *glp* repressor is listed in Table 6. The substitutions at positions 1-9 impaired binding affinity of *glp* operator for repressor to different degrees (except 1A and 6C). Substitutions at positions 3, 4, 5 and 8 with any other base all caused a great decrease in binding affinity while substitutions at other positions influenced binding of repressor differentially, depending on the base substituted. Since the β-galactosidase specific activity reflects the binding affinity of operator carried by the plasmid for *glp* repressor, the effect of the substitutions on binding affinity could be defined as follows: β-galactosidase activity >110, normal; 90-110, mild; 60-89, moderate; < 60, severe effect on binding. Using these criteria, the results in Table 6 can be summarized as shown in Table 7. It is apparent that all substitutions at positions 3, 4, 5 and 8 had a severe effect on repressor binding, suggesting that these are the critical positions for binding of *glp* repressor. Substitutions at position 7 with C or A; at position 6 with A; at position 2 with C and position 1 with C also had a severe effect on the binding affinity. Substitutions at position 9, position 6 with G, position 2 with G and position 1 with G had a moderate effect. Operator 7T, 6C, 2T and 1A yielded high levels of β-galactosidase activity (> 90),

Table 7. Substitution effect on operator for binding of *glp* repressor[1]

	9 A	8 T	7 G	6 T	5 T	4 C	3 G	2 A	1 T
Normal				C					A
Mild			T					T	
Moderate	G T C			G				G	G
Severe		C A G	C A	A	C G A	G A T	C T A	C	C

[1] The sequence of the half-site *glp* operator is given on the top. The effects of mutations on the operator binding affinity were estimated according to the data given in Table 6 as follows: β-galactosidase < 60, severe; 60-89, moderate; 90-110, mild; >110, normal (no effect).

indicating these bases substitutions are tolerable for binding of repressor. Based on comparison of native operators (11) and results of this study, a modified *glp* operator consensus sequence can be deduced (Fig. 3). The modified half-site consensus (WATKYTCGWW) is derived both from the frequency of occurrence of the bases in the natural operators and from the present data showing which substitutions had mild or no effect on repressor binding.

***Binding of* glp *repressor to tandem operators and tandem operators with various insertions*:** Tandemly repeated operators occur in the native *glpD*, *glpACB* and *glpFKX* operons (11). Cooperative binding of *glp* repressor for control of *glpACB-glpTQ* and *glpFKX* operons has been suggested (11). In order to determine if repressor binds cooperatively to two adjacent operators, a *glp* tandem operator (with two directly repeated 20-bp consensus operators) was constructed. In addition, *glp* tandem operators with insertions of different sizes between the operators were constructed. A *deo* tandem operator was also constructed. The binding affinity of each operator construct was measured in vivo using strain WO331 as described above. The

Operator	Position	Sequence (5′ → 3′)	Match with consensus (%)
O_D1	-10	TATGTTCGAT AACGAACATT	100
O_D2	+11	TATGagCtTT AACGAAAgTg	75
O_A1	-60	AATGTTCaAA ATgacGCATg	75
O_A2	-28	AcTTTcgaAT TAtGAGCgaA	70
O_A3	-8	TATGCgCGAA ATCaAACAaT	85
O_A4	+33	AATGgTaaAA AACGAACtTc	75
O_T	-51	TgTGTgCGgc AAttcACATT	65
O_F1	-89	AtgGCgCGAT AACGctCATT	75
O_F2	-68	TATGacgagg cACacACATT	55
O_F3	-47	TAaGTTCGAT ATttctCgTT	70
O_F4	-27	TtTGCTCGTT AACGAtAAgT	85
O_K1	+993	cgcGgTCGTA ATgGAtCAcg	60
O_K2	+1035	gcaGCgCGAA TTtGAGCAaA	70

Frequency of occurrence: $A_{11}A_{15}T_{18}G_{23}T_{11}T_{14}C_{18}G_{15}A_{13}T_{16}$

$T_9\ T_5\ C_4\ T_3\ C_7\ G_{10}A_3\ A_8\ T_{11}A_7$

$C_4\ C_3\ A_3\quad A_6\ C_2\ T_3\ C_2\ G_2\ G_2$

$G_2\ G_3\ G\quad G_2\quad G_2\ T\quad C$

Operator half-site consensus: 5′- W A T K Y T C G W W -3′

(K = G or T, W = A or T, Y = C or T)

Figure 3. Comparison of the operators of the *glp* regulon: Operator sequence is based on comparison of native operators

and on results of this study. The positiom number refers to the position of the first base listed relative to the start of transcription. Bases matching the consensus are indicated in upper case letters. This Figure is modified from (11).

results are shown in Table 8. pGEM3Z was used as the negative control. The results showed that the *glp* single consensus operator binds repressor specifically, yielding 115 Miller units. However, the *glp* tandem operator had a much higher binding affinity (1440 Miller units). This result indicates that *glp* tandem operator binds repressor cooperatively. The insertion of 2 bases between the tandem operators had no effect on the cooperativity as indicated by the high β-galactosidase activity. During construction of $(glp)_2$+2, $(glp)_2$+2 (8'G) was obtained by a spontaneous mutation. The latter also exhibited full cooperativity for binding of repressor (Table 8). The deletion of 2 bp or insertion of 4 bp resulted in decreased cooperativity, while insertion of 6 bp eliminated the cooperativity between tandem operators (the ratio of Miller units from $(glp)_2$+6 and *glp* is 2). *glp* repressor was unable to bind to the $(deo)_2$ operator (β-galactosidase activity was almost the same as that from pGEM3Z vector) although a tandem symmetrical DNA sequence was provided.

Table 8. Binding of *glp* repressor to tandem operators

Operator name[1]	Sequence	β-Galactosidase[2]
pGEM3Z vector	no operator sequence	36 (8)
Consensus, single	5′-TATGTTCGATATCGAACATA-3′	115 (11)
deo tandem	5′-TATGTTAGATATCTAACATA- TATGTTAGATATCTAACATA-3′	36 (3)
glp tandem	5′-TATGTTCGATATCGAACATA- TATGTTCGATATCGAACATA-3′	1440 (6)
glp tandem (-2)	5′-TATGTTCGATATCGAACAT- ATGTTCGATATCGAACATA-3′	990 (6)
glp tandem (+2)	5′-TATGTTCGATATCGAACATAT- ATATGTTCGATATCGAACATA-3′	1500 (2)
glp tandem (+2; 8′G)	5′-TATGTTCGATATCGAACGTAT- ATATGTTCGATATCGAACATA-3′	1530 (2)
glp tandem (+4)	5′-TATGTTCGATATCGAACATATC- GATATGTTCGATATCGAACATA-3′	780 (4)
glp tandem (+6)	5′-TATGTTCGATATCGAACATATCG- CGATATGTTCGATATCGAACATA-3′	220 (4)

[1] Plasmids carrying *glp* tandem (+2) and *glp* tandem (+2; 8′G) were constructed in this study, and the others had been constructed in this laboratory.

[2] The numbers in parentheses are the numbers of independent duplicate or triplicate β-galactosidase activity assays.

DISCUSSION

The *glp* repressor negatively controls the members of the *glp* regulon by binding to the *glp* operator sites. A total of thirteen *glp* operator sites have been identified in the previous studies (Fig. 3) (11). Sequence comparison revealed a 10 base pair operator half-site consensus sequence (11). Each operator matches more or less well the consensus sequence. The frequency of occurence for each base at positions 1-10 is also shown in Fig. 3. Results in this study showed that *glp* repressor binds specifically to the *glp* operator consensus. Substitutions at positions 1-9 resulted in a decreased binding affinity (except 1A and 6C) as indicated by the lowered β-galactosidase activity. The effect was mild, moderate or severe depending on the position where the substitution was introduced and on the specific base used. From Table 6, it is apparent that substitutions at positions 3, 4, 5, or 8 resulted in almost complete loss of binding affinity of mutant operator for *glp* repressor, suggesting that these positions are most critical for *glp* repressor binding. This is consistent with the results of comparison of

the native operators (Fig. 3) which showed that positions 3, 4, 5, 7, and 8 are highly conserved. Substitution of C,G with A,T at position 4,4' changes the *glp* operator consensus to the consensus for the *deo* operator, and the resulting single (*deoO*) or tandem operator [$(deoO)_2$] was unable to bind to *glp* repressor. Studies on the DNA specificity of the *deo* operator site by Hammer et al. (62) showed that substition of A,T at 4,4' with C,G (*deoO* → *glpO*) also eliminated the specific binding of mutant operator for *deo* repressor (62). Thus, the base differences at positions 4 and 4' distinguish the *glp* operator from *deo* operator and allow the *glp* operons and *deo* operons to be controlled by their respective repressor. Substitutions at position 7 (A,C), 2 (C) or 1(C) also had a large effect on the operator binding affinity while substitutions at position 1 (G), 2 (T,G), 7 (T) and 9 had mild to moderate influence on the operator for repressor binding. Position 6 is less conserved compared with other positions in the *glp* operator. The substitution effect at this position was quite heterogeneous. The substitution with C increased the binding affinity for *glp* repressor while 6G or 6A had moderate and severe effects on repressor binding, respectively. The result suggested that positions 3, 4, 5, 7 and 8 are most important for

repressor binding while positions 1, 2, 6 and 9 also contribute to the specificity of the *glp* operator. The magnitude of the effect caused by a specific base substitution was consistent with the frequency of occurrence of that base at that position in most cases (Fig. 3).

Even though positions 1-9 of the *glp* operator are important for binding of *glp* repressor, individual native operators match the consensus operator to different degrees. The degree of repression by GlpR is also significantly affected by the position of the operator site relative to the promotor elements of the operon and the number of the operator sites in each operon. Multiple operator sites may provide the basis for cooperative binding which may greatly increase the responsiveness to control by repressor. Cooperative binding was observed in phage λ (tandem operators O_R1/O_R2), *lac*, *gal* and *deo* operons (widely seperated operators) (11,53,54,63,64,75,76). Studies on the regulation of the *glp* regulon by GlpR suggested that cooperativity also exists for repressor binding to tandem operators in the *glpTQ-glpACB* and *glpFKX* operons (11,34). In the present study, direct evidence for cooperative binding of repressor to tandem *glp* operators is given. The binding affinity ratio of *glp* tandem or *glp* tandem

+2 to the *glp* single consensus operator is 13, indicating repressor binds to adjacent operators cooperatively. Insertion of 4 bp or deletion of 2 bp between two *glp* operators resulted in a decrease in cooperativity while a 6 bp insertion (approximately one-half a helical turn) eliminated the cooperativity (the binding affinity ratio of *glp* tandem + 6 to *glp* single consensus is 2, implying that [$(glp)_2$ + 6] functions like two independent operators). $(glp)_2$ -2 can be considered as two single consensus operators with the deletion of 2 bases at position 10 and 9 of one operator. The decrease of the binding affinity in $(glp)_2$ - 2 may be caused by the decreased space between two bound repressors or the impaired binding affinity of the operator with the deletion. [$(glp)_2$ + 2, 8'G] was obtained by spontaneous mutation in a plasmid carrying [$(glp)_2$ + 2]. Since position 8 is critical for repressor binding, we expected that this operator would have a lowered cooperativity. The result showed that the cooperative binding of repressor to this operator was as strong as that of the tandem operator, suggesting that cooperativity may override absolute sequence specificity. This may explain the tight control of *glp* operons by *glp* repressor although the operator sequences do not perfectly match the *glp* operator consensus.

PART II

STUDIES ON THE

glp *REPRESSOR-EFFECTOR INTERACTION*

RESULTS

The *glp* operons are negatively controlled by the *glpR*-encoded repressor. The inducer for the regulon is glycerol-P. Strain ECL89 has a noninducible phenotype with regard to the members of the regulon, presumably conferred by an altered form of *glp* repressor. In order to determine the molecular basis for the noninducible phenotype, *glpR12* DNA from ECL89 was amplified by using the polymerase chain reaction, cloned into pGEM-T and then sequenced. Individual mutations were separated by PCR and subcloning techniques. The inducer binding affinity was measured for both the wild-type and noninducible repressors.

***Sequencing of* glpR12 *DNA from ECL89*:** *glpR12* DNA from ECL89 was amplified by PCR using two synthetic primers (Table 3) in which *Eco*RI and *Sal*I restriction sites were incorporated to facilitate subsequent cloning. The PCR product was cloned into pGEM-T and sequenced using appropriate

primers (Table 2). The nucleotide sequences of *glpR* (wild-type) and *glpR12* (noninducible repressor) along with their deduced amino acid sequences are shown in Figure 4. Six mutations which resulted in amino acid substitutions at positions 45 (E→D), 46 (Q→E), 87 (E→D), 88 (Q→E), 101 (T→I) and 147 (P→R) were identified. The sequence of *glpR12* encoding the helix-turn-helix DNA-binding motif was identical to that of the wild-type *glpR* gene. Substitution of a DNA fragment containing mutations that resulted in amino acid substitutions T101I and P147R into the wild-type *glpR* gene conferred a noninducible phenotype on a strain harboring this recombinant plasmid (Table 9), suggesting that one or both of these mutations were responsible for noninducibility. To eliminate the possibility that either of these two mutations was generated by the PCR, another independent PCR reaction was carried out. The above two mutations in *glpR12* were verified by sequencing.

***Identification of the mutation responsible for the noninducible phenotype conferred by* glpR12:** In order to determine if both of the mutations are responsible for the noninducible phenotype, pNZ80, pNZ102,

```
S. D.
AGGGATTTATAAATGAAACAAACACAACGTCACAACGGTATTATCGAACTGGTTAAACAG
    glpR --> M  K  Q  T  Q  R  H  N  G  I  I  E  L  V  K  Q

CAGGGTTATGTCAGTACCGAAGAGCTGGTAGAGCATTTCTCCGTCAGCCCGCAGACTATT
 Q  G  Y  V  S  T  E  E  L  V  E  H  F  S  V  S  P  Q  T  I
                        Helix                Turn

                              (CG)
CGCCGCGACCTCAATGAGCTGGCGGAGCAAAACCTGATCCTGCGCCATCATGGCGGTGCG
 R  R  D  L  N  E  L  A  E  Q  N  L  I  L  R  H  H  G  G  A
 Helix                 (D)(E)

GCGCTGCCTTCCAGTTCGGTTAACACGCCGTGGCACGATCGCAAGGCCACCCAGACCGAA
 A  L  P  S  S  S  V  N  T  P  W  H  D  R  K  A  T  Q  T  E

                                    (CG)
GAAAAAGAGCGCATCGCCCGCAAAGTGGCGGAGCAAATCCCCAATGGCTCGACGCTGTTT
 E  K  E  R  I  A  R  K  V  A  E  Q  I  P  N  G  S  T  L  F
                                 (D)(E)

            (T)
ATCGATATCGGCACCACGCCGGAAGCGGTAGCGCACGCACTGCTCAATCACAGCAATTTG
 I  D  I  G  T  T  P  E  A  V  A  H  A  L  L  N  H  S  N  L
            (I)

CGCATTGTCACCAACAATCTCAACGTTGCTAACACGTTGATGGTAAAAGAAGATTTTCGC
 R  I  V  T  N  N  L  N  V  A  N  T  L  M  V  K  E  D  F  R

                                    (GC)
ATCATTCTCGCCGGTGGCGAATTACGCAGCCCGGATGGCGGGATCATTGGCGAAGCGACG
 I  I  L  A  G  G  E  L  R  S  P  D  G  G  I  I  G  E  A  T
                                 (R)

CTCGATTTTATCTCCCAGTTCCGCCTTGATTTCGGCATTCTGGGGATAAGCGGCATCGAT
 L  D  F  I  S  Q  F  R  L  D  F  G  I  L  G  I  S  G  I  D

AGCGACGGCTCGCTGCTGGAGTTCGATTACCACGAAGTTCGCACCAAACGCGCCATTATT
 S  D  G  S  L  L  E  F  D  Y  H  E  V  R  T  K  R  A  I  I

GAGAACTCGCGCCACGTTATGCTGGTTGTCGATCACTCGAAATTTGGCCGTAACGCGATG
 E  N  S  R  H  V  M  L  V  V  D  H  S  K  F  G  R  N  A  M

GTCAATATGGGCAGCATCAGCATGGTAGATGCCGTCTACACCGACGCCCCGCCGCCAGTA
 V  N  M  G  S  I  S  M  V  D  A  V  Y  T  D  A  P  P  P  V

AGCGTGATGCAGGTGCTGACGGACCACCATATTCAACTGGAGCTGTGCTGA
 S  V  M  Q  V  L  T  D  H  H  I  Q  L  E  L  C  *
```

Fig. 4. Nucleotide sequence of *glpR* and deduced amino acid sequence of *glp* repressor. S.D., Shine-Dalgarno sequence. The mutations in the *glpR12* gene which resulted in alterations of amino acids at position 45 (E → D), 46 (Q → E), 87 (E → D), 88 (Q → E), 101 (T→I) and 147 (P→R) are written in parentheses. The helix-turn-helix region of GlpR12 is identical to that of the wild-type *glp* repressor and is italilized.

pNZ81 and pNZ83 were constructed as described before (Table 5). These plasmids harbor the *glpEGR* operon, encoding either the wild-type, T101I and P147R, T101I or P147R forms of repressor. Strain SY102 [ϕ(*glpD-lacZ*) *glpR2*] was transformed using the above plasmids. The lactose phenotype of the transformants was tested on LB + Amp + X-gal + Glycerol-P medium (Table 9). SY102 harboring pNZ80 turned blue on this medium, indicating that wild- type *glp* repressor is inactivated by the inducer, glycerol-P. SY102 harboring pNZ102 (T101I, P147R) or pNZ81 (T101I) appeared white, while SY102 harboring pNZ83 (P147R) turned blue on the indicator plate. The results demonstrated that the noninducible phenotype is conferred by the amino acid substitution at position 101 (T→I).

Determination of the nature of the noninducibility: In order to determine the nature of the noninducibility of GlpR12, $GlpR^+$ and GlpR12 were overproduced and the inducer binding activity was measured in both cases. The overproduction of GlpR12 was not high enough to readily detect *in vitro* using pNZ102, probably due to the autoregulation of the *glpEGR* operon. Therefore, pNZ103 which contains only the *glpR12* gene adjacent

Table 9. Lactose phenotype of strain SY102 [*ϕ*(*glpD-lacZ*) *glpR2*] harboring various cloned alleles of *glpR*

SY102/Plasmid (*glpR* allele)	LacZ phenotype[1]
pNZ80 (*glpR*$^+$)	+
pNZ102 (*glpR* T101I, P147R)	-
pNZ81 (*glpR* T101I)	-
pNZ83 (*glpR* P147R)	+

[1] Phenotype was determined on LB + Amp + X-gal + glycerol-P medium.

to the *tac* promoter of pSY223 was constructed as mentioned before (Table 5) in order to overexpress the *glpR12*-encoded repressor. pSY2-6 was used to overexpress the wild-type repressor. Overproduction of repressor was achieved in strain SY102 (*glpR2 recA1*) harboring pSY2-6 or pNZ103 and crude extracts were prepared. The overproduction of $GlpR^+$ and GlpR12 was observed on the SDS gel (Fig. 5). Glycerol-P binding specific activity was determined for crude extracts containing $GlpR^+$ or GlpR12 (Table 10). The inducer binding activity of GlpR12 (T101I, P147R) was 27 fold lower than that of the wild-type GlpR. The apparent binding activity for the extract containing GlpR12 was at the limit of detection for the assay and was indistinguishable from that of the crude extract from the strain with the vector containing no *glp* repressor. The results strongly suggest that the noninducibility of GlpR12 is caused by a lowered affinity for inducer binding.

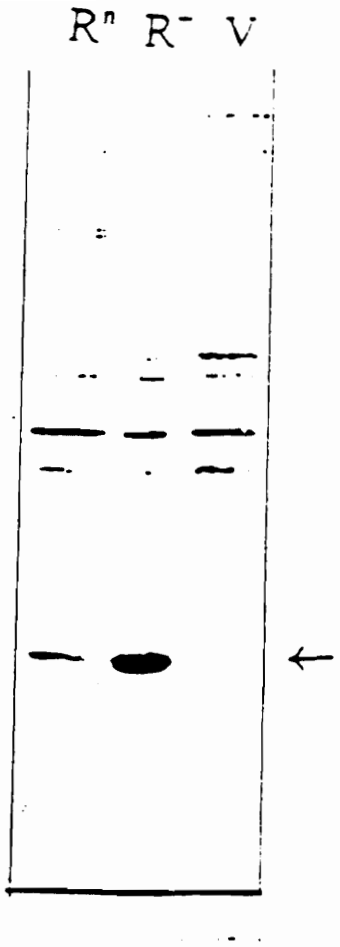

Figure 5. SDS gel: Overproduction of *glp* repressor variants.

Table 10. Specific activity of inducer binding for GlpR$^+$ and GlpR12 *in vitro*[1]

Plasmid/Repressor	Sp. Act. (unit/mg)[2]
pSY2-6/GlpR$^+$	357
pNZ103/GlpR (T101I, P147R)	13
pSY223 (vector, no GlpR)	18

[1] The assay was done using 0.1 mM [2-^{3}H]glycerol-P (110 cpm/pmol).

[2] Specific activity is expressed as unit per mg crude extract protein. One unit is defined as the amount of repressor that binds 1 pmol of glycerol-P. The specific activities are the mean numbers from six (pSY2-6) or seven (pNZ103 and pSY223) measurments.

DISCUSSION

Sequence analysis of *glpR*^{n} from strain ECL89 showed that there are two mutations in the carboxyl-terminal domain that cause amino acid substitutions at position 101 (T → I) and 147 (P → R). The putative helix-turn-helix DNA binding motif for *glpR*^{n} was identical to that of the wild-type repressor. Substitution of a DNA fragment containing these two mutations into the wild type *glpR* gene conferred a noninducible phenotype on a strain harboring this recombinant plasmid, suggesting that one or both of these mutations were responsible for the noninducible phenotype. Results from further experiments showed that the mutation which caused an amino acid substitution at position 101 (T → I) was sufficient to confer the noninducible phenotype (Table 9). Mutations resulting in noninducible repressor have been observed in other regulatory proteins such as LacI and GalR (45,76). The noninducible phenotype can be caused by a lowered affinity for the inducer and/or a tighter binding for the operator. Determination of *in vitro*

inducer binding specific activity in this study showed that the specific activity of inducer binding for $GlpR^+$ is 357 unit/mg while $GlpR^n$ yields 13 unit/mg which is even lowere than that of the vector (18 unit/mg), suggesting that the noninducibility of the *glp* regulon in strain ECL89 was mainly caused by a lowered affinity of *glp* repressor for the inducer.

Sequence comparisons show that GlpR is homologous to the DeoR family of repressors including DeoR, GutR, FucR of *E. coli* and LacR of Gram-positive bacteria (Fig. 6) (69). The overall amino acid sequences of these catabolic regulatory proteins are highly homologous especially in the helix-turn-helix DNA-binding motif and at several regions in the C-terminal domain. Since all of these repressors have a sugar-P as their inducers, part of the inducer binding site might be relatively conserved. A putative inducer binding site has been proposed by van Rooijen et al. (69) and is located at the C-terminal domain (between amino acid 203-213 of GlpR) (Fig. 6). Although the amino acid substitution T101I was not located in the putative inducer binding site, it is in a relatively conserved position. The Thr or Ser found in this position of the repressors may form a hydrogen bond to sugar-P. The substitution at this position with a hydrophobic residue (I) may cause

an allosteric change which decreases binding affinity of the repressor for the inducer. Alternatively, the loss of the proposed H-bond contact between T101 and the inducer may cause decreased binding affinity.

```
                                helix--t--helix
                                                                (DE)
GlpR      MKQTQRHNGIIELVKQQGYVSTEELVEHFSVSPQTIRRDLNELAEQNLILRHHGGAA-L--PSSS
GutR      MKPRQRQAAILEYLQKQGKCSVEELAQYFDTTGTTIRKDLVILEHAGTVIRTYGGVV-L--NKEE
FucR      MK-AARQQAIVDLLLNHTSLTTEALSEQLKVSKETIRRDLNELQTQGKILRNHGRAKYIHRQNQD
DeoR    METRREERIGQLLQELKRSDKLHLKDAAALLGVSEMTIRRDLNNHSAPVVLL---GGYI-VLEPRSA
SLacR     MNKHERLDEIAKLVNKKGTIRTNEIVEGLNVSDMTVRRDLIELENKGILTKIHGGAR-SNSTFQY
LLacR     MNKKRRLEKILDMLKIDGTITIKEIIDELDISDMTARRDLDALEADGLLTRTHGGAQ-LLSSKKP
                *   ... .     . ... . . .+  *.***  +     . .  *

                                    (DE)          (I)
GlpR    VNTPWHDRKATQTEEKERIARKVAEQIPNGSTLFIDIGTTPEAVAHALLNHS--NLRIVTNNLNVAN
GutR    SDPPIDHKTLINTHKKELIAEAAVSFIHDGDSIILDAGSTVLQMVPLLSRFN--NITVMTNSLHIVN
FucR    SGDPFHIRLKSHYAHKADIAREALAWIEEGMVIALDASSTCWYLARQLPD---INIQVFTNSHPICH
DeoR    SHYLLSDQKSRLVEEKRRAAKLAATLVEPDQTLFFDCGTTTPWIIEAIDNEIPF--TAVCYSLNTFL
SLacR   KEISHKEKHTRQIAEKRFIAKKAASLIEDGDTLFFGPGTTVELLAEEVNHHT---LTIITNCLPVYK
LLacR   LEKTHIEKKSLNTKEKIDIAKKACSLIKDGDTIFIGPGTTLVQLALELKGRKGYKIRVITNSLPVFL
              .      .*  +*. + . +  + ..... ..*   ..  .       . ..++ + .

                              (R)
GlpR    TLM-VKE-DFRIILAGGELRSPDGGIIGEATLDFISQFRLDFGILGISGIDSDGSLLEFDYHEVRTK
GutR    ALSELDN-EQTILMPGGTFRKKSASFHGQLAENAFEHFTFDKLFMGTDGIDLNAGVTTFN-EVYTVS
FucR    ELGKRERIQ--LISSGGTLERKYGCYVNPSLISQLKSLEIDLFIFSCEGIDSSGALWDSNAINADYK
DeoR    ALKE--KPHCRAFLCGGEFHASNAIFKPIDFQQTLNNFCPDIAFYSAAGVHVSKGATCFNLEELPVK
SLacR   ILLEKQTAHFRVYLIGGEMRHITEAFVGEMANAMLEKLRFSKMFFSSNAVN-KGAVMTSTLDEAYTQ
LLacR   ILNDSETIDL--LLLGGEYREITGAFVGSMASTNLKAMRFAKAFVRANAVT-HNSIATYSDKEGVIQ
         *      .      **. .     . .      .. . .  ..  ...    ..   .

GlpR    RAIIENSRHVMLVVDHSKFGRNAMVNMGSISMVDAVYTDAPPPVSVMQVLTDHHIQLELC      252
GutR    KAMCNAAREVILMADSSKFGRKSPNVVCSLESVDKLITDAGIDPAFRQALEEKGIDVIITGESNE 257
FucR    SMLLKRAAQSLLLIDKSKFNRSGEARIGHLDEVTHIISDERQVATSLVTA                243
DeoR    HWAMSMAQKHVLVVDHSKFGKVRPARMGDLKRFDIVVSDCCPEDEYVKYAQTQRIKLMY      252
SLacR   QLALSNSIEKYLLIDHTKVGKEDFTSFCQLNELTAVVMDYE-DEEKVETIKTYIEVVD       242
LLacR   QLALNNAVEKFLLVDSTKFDRYDFFNFYNLDQLDTIITDNQISPQHLEEFSQYTTILKAD     255
          . . .  *..* .*+.          +. .. .. *
```

Figure 6. Alignment of the amino acid sequences of the glp and related repressors. The amino acid sequences were aligned using the CLUSTAL program of PC/GENE. The postion of the presumtive helix-turn-helix motif is indicated at the top of the figure. Positions conserved in all six sequences are indicated by asterisks, positions containing five identities are indicated by plus signs, and well conserved positions are indicated by dots. GlpR, GutR, FucR and DeoR are the repressors of the glp, gut, fuc and deo systems of E. coli, respectively. SLacR and LLacR are the lac repressors of S. aureus and L. lactis, respectively. The altered amino acids found in GlpR12 are written in parentheses above the sequence of GlpR.

LITERATURE CITED

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

2. Raetz, C. R. H., and W. Dowhan. 1990. Biosynthesis and function of phospholipids in *Escherichia coli. J. Biol. Chem.* **265,** 1235-1238

3. Lin, E. C. C. 1976. Glycerol dissimilation and its regulation in bacteria. *Ann. Rev. Microbiol.* **30,** 535-578

4. Lin, E. C. C. 1987. Dissimilatory pathways for sugars, polyols, and carboxylates, p. 244-284. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds.) "*Escherichia coli* and *Salmonella typhimurium. Cellular and Molecular Biology*", American Society for Microbiology, Washington, D. C.

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

6. Schweizer, H., W. Boos, and T. J. Larson. 1985. Repressor for the *sn*-glycerol-3-phosphate regulon of *Escherichia coli* K-12: Cloning of the *glpR* gene and identification of its product. *J. Bacteriol.* **161,** 563-566

7. Austin, D. 1991. Characterization of the *glpD* and *glpEGR* operons of *Escherichia coli* K-12. Ph.D. thesis. Virginia Polytechnic Institute and State University, Blacksburg, Virginia

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

9. Sweet, G., C. Gandor, R. Voegele, N. Wittekindt, J. Beuerle, V. Truniger, E. C. C. Lin, and W. Boos. 1990. Glycerol facilitator of *Escherichia coli*. Cloning of *glpF* and identification of the *glpF* product. *J. Bacteriol.*

172, 424-430

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

11. Weissenborn, D. L., N. Wittekindt, and T. J. Larson. 1992. Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J. Biol. Chem.* **267,** 6122-6131

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

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

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

15. Lupski, J. R., Y. H. Zhang, M. Rieger, M. Minter, B. Hsu, B. G. Coi, T. Koeuth, and E. R. B. McCabe. 1990. Mutational analysis of the *Escherichia coli glpFK* region with *Tn5* mutagenesis and the polymerase chain reaction. *J. Bacteriol.* **173,** 101-107

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

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

18. Hayashi, S. I., J. P. Koch, and E. C. C. Lin. 1964. Active transport of

L-α-glycerophosphate in *Escherichia coli. J. Biol. Chem.* **239,** 3098-3105

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

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

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

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

23. Ehrmann, M., W. Boos, E. Ormseth, H. Schweizer, and T. J. Larson. 1987. Divergent transcription of the *sn*-glycerol-3-phosphate active transport (*glpT*) and anaerobic *sn*-glycerol-3-phosphate dehydrogenase (*glpA glpB glpC*) genes of *Escherichia coli* K-12. *J. Bacteriol.* **169,** 526-532

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

25. Ye, S. 1990. Structure and function of the repressor and operators of

the *sn*-glycerol-3-phosphate regulon of *Escherichia coli* K-12. Ph.D. thesis. Virginia Polytechnic Institute and State University, Blacksburg, Virginia

26. Kuritzkes, D. R., L. Y. Zhang, and E. C. C. Lin. 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

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

28. Austin, D., and T. J. Larson. 1991. Nucleotide sequence of the *glpD* gene encoding *sn*-glycerol 3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* **173,** 101-107

29. Choi, Y. L., S. Kawase, T. Nishida, H. Sakai, T. Komano, M. Kawamukai, R. Utsumi, Y. Kohara, and K. Akiyama. 1988. Nucleotide sequence of the *glpR* gene encoding the repressor for the glycerol-3-phosphate regulon of *Escherichia coli* K-12. *Nucleic Acids Res.* **16,** 7732

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

31. Schweizer, H., G. Sweet, and T. J. Larson. 1986. Physical and genetic structure of the *glpD-malT* interval of the *Escherichia coli* K-12 chromosome. Identification of two new genes of the *glp* regulon. *Mol. Gen. Genet.* **202,** 488-492

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

Chem. **262,** 15869-15874

33. Larson, T. J., J. S. Cantwell, and A. T. van Loo-Bhattacharya. 1992. Interaction at a distance between multiple operators controls the adjacent, divergently transcribed *glpTQ-glpACB* operons of *Escherichia coli* K-12. *J. Biol. Chem.* **267,** 6114-6121

34. de Crombrugghe, B., R. L. Perlman, H. E. Varmus, and I. Pastan. 1969. Regulation of inducible enzyme synthesis in *Escherichia coli* by cyclic adenosine 3',5'-monophosphate. *J. Biol. Chem.* **244,** 831-838

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

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

37. Zhang, X., and R. H. Ebright. 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 places the lac promoter in the FNR regulon. *J. Biol. Chem.* **265,** 12400-12403

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

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

40. Hennecke, H. 1990. Regulation of bacterial gene expression by metal-protein complexes. *Mol. Microbiol.* **4,** 1621-1628

41. Iuchi, S., S. T. Cole, and E. C. C. Lin. 1990. Multiple regulatory

elements for the *glpA* operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the *glpD* operon encoding aerobic glycerol-3-phosphate dehydrogenase in *Escherichia coli*. Further characterization of respiratory control. *J. Bacteriol.* **172,** 179-184

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

43. Gilbert, W., and B. Müller-Hill. 1970. The lactose repressor. p. 93-109. *In* J. R. Beckwith, and D. Zipser (eds.) "The lactose operon", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY

44. Barkley, M. D., and S. Bourgeois. 1980. Repressor recognition of operator and effectors. p. 177-220. In Miller, J. H., and W. S. Reznikoff (eds.) "The operon", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY

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

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

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

48. Müller-Hill, B. 1975. *lac* repressor and *lac* operator. *Prog. Biophys. Mol. Biol.* **30,** 227-252

49. Beyreuther, K. 1980. Chemical structure and functional organization of *lac* repressor from *Escherichia coli.* p. 123-154. *In* Miller, J. H., and W. S. Reznikoff (eds.) "The operon", Cold Spring Harbor Laboratory,

Cold Spring Harbor, New York, NY

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

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

52. Gilbert, W. and A. Maxam. 1973. The nucleotide sequence of the *lac* operator. *Proc. Natl. Acad. Sci.* **70,** 3581

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

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

55. Ptashne, M. 1980. λ repressor function and structure. p. 325-344. *In* Miller, J. H., AND W. S. Reznikoff (eds.) "The operon", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY

56. Sauer, R. T., and R. Anderegg. 1978. Primary structure of the λ repressor. *Biochemistry.* **17,** 1092-1100

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

58. Majumder, A., and S. Adhya. 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., M. Koenen, H. W. Griesser, and B. Müller-Hill. 1983. 72 residues of *gal* repressor fused to β-galactosidase repress the *gal* operon of *Escherichia coli. EMBO J.* **2,** 1271-1274

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

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

62. Hammer, K., L. Bech, P. Hobolth, and G. Dandanell. 1993. DNA specificity of *Escherichia coli deoP1* operator-DeoR repressor recognition. *Mol. Gen. Genet.* **237,** 129-133

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

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

65. Ye, S., and T. J. Larson. 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

66. Pabo, C. O. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61,** 1053-1095

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

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

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

70. Silhavy, T. J., M. L. Berman, and L. W. Enquist. 1988. *In* Experiments with gene fusions. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY

71. Sanger, F., S. Nicklen, and A. R. Coulson. 1987. DNA sequence with chain-termination inhibtors. *Proc. Natl. Acad. Sci. U.S.A.* **774,** 5463-5467

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

73. 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

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

75. Ptashne, M. 1986. *In* A Genetic Switch. Cell Press and Blackwell Scientific Publications. Cambridge, Mass.

76. Adhya, S. 1987. The galactose operon, p. 1503-1513. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds.) "*Escherichia coli* and *Salmonella typhimurium. Cellular and Molecular Biology*", American Society for Microbiology, Washington, D. C.

Vita

Ningyue Zhao was born in Shanghai, P. R. China on June 20, 1963. She recieved the Bachelor of Science degree in medicine from Shanghai Medical University in July of 1987. During 1987 - 1990, she served as an instructor in the department of Biology and Human genetics at Shanghai Medical University. She came to the United States to join her husband in October, 1990.

She entered the graduate program in Biochemistry and Nutrition at Virginia Polytechnic Institute and State University in August, 1991 and obtained the Master of Science degree in Dr. Larson's laboratory in September, 1993.

Ningyue Zhao