Characterization of AgaR and YihW, Members of the DeoR Family of Transcriptional Regulators, and GlpE, a Rhodanese Belonging to the GlpR Regulon, Also a Member of the DeoR Family

by

William Keith Ray

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

APPROVED:

_____

Timothy J. Larson, Chairman

_____ _____

Jiann–Shin Chen David R. Bevan

_____ _____

Elizabeth A. Grabau Ann M. Stevens

August 2, 1999

Blacksburg, Virginia

Key Words: transcriptional control, carbon metabolism, sulfurtransferase, iron–sulfur cluster biosynthesis

Characterization of AgaR and YihW, Members of the DeoR Family of Transcriptional Regulators, and GlpE, a Rhodanese Belonging to the GlpR Regulon, Also a Member of the DeoR Family

by

William Keith Ray

Timothy J. Larson, Chairman

Biochemistry

(ABSTRACT)

AgaR, a protein in *Escherichia coli* thought to control the metabolism of *N*–acetylgalactosamine, is a member of the DeoR family of transcriptional regulators. Three transcriptional promoters within a cluster of genes containing the gene for AgaR were identified, specific for *agaR*, *agaZ* and *agaS*, and the transcription start sites mapped. Transcription from these promoters was specifically induced by *N*–acetylgalactosamine or galactosamine, though K–12 strains lacked the ability to utilize these as sole sources of carbon. The activity of these promoters was constitutively elevated in a strain in which *agaR* had been disrupted confirming that the promoters are subject to negative regulation by AgaR. AgaR–His_6, purified using immobilized metal affinity chromatography, was used for DNase I footprint analysis of the promoter regions. Four operator sites bound by AgaR were identified. A putative consensus binding sequence for AgaR was proposed based on these four sites. *In vivo* and *in vitro* analysis of the *agaZ* promoter indicated that this promoter was activated by the cAMP–cAMP receptor protein (CRP). Expression from the *aga* promoters was less sensitive to catabolite repression in revertants capable of *N*–acetylgalactosamine utilization, suggesting that these revertants have mutation(s) that result in an elevated level of inducer for AgaR.

A cluster of genes at minute 87.7 of the *E. coli* genome contains a gene that encodes another member of the DeoR family of transcriptional regulators. This protein, YihW, is more similar to GlpR, transcriptional regulator of *sn*–glycerol 3–phosphate metabolism in *E. coli*, than other members of the DeoR family. Despite the high degree of similarity, YihW lacked the ability to repress P_{glpK}, a promoter known to be controlled by GlpR. A variant of YihW containing substitutions in the putative recognition helix to more closely match the recognition helix of GlpR was also unable to repress P_{glpK}. Transcriptional promoters identified in this cluster of genes were negatively regulated by YihW.

Regulation of genes involved in the metabolism of *sn*–glycerol 3–phosphate in *E. coli* by GlpR has been well characterized. However, the function of a protein (GlpE) encoded by a gene cotranscribed with that for GlpR was unknown prior to this work. GlpE was identified as a single–domain, 12–kDa rhodanese (thiosulfate:cyanide sulfurtransferase). The enzyme

was purified to near homogeneity and characterized. As shown for other characterized rhodaneses, kinetic analysis revealed that catalysis occurs via an enzyme–sulfur intermediate utilizing a double–displacement mechanism requiring an active–site cysteine. K_m (SSO_3^{2-}) and K_m (CN^-) were determined to be 78 mM and 17 mM, respectively. The native molecular mass of GlpE was 22.5 kDa indicating that GlpE functions as a dimer. GlpE exhibited a k_{cat} of 230 s^{-1}. Thioredoxin, a small multifunctional dithiol protein, served as sulfur–acceptor substrate for GlpE with an apparent K_m of 34 μM when thiosulfate was near its K_m, suggesting thioredoxin may be a physiological substrate.

This work is dedicated to the memory of my aunt, Minnie Ruth Ray Simpson.

Acknowledgements

I first want to express my sincere appreciation and respect for my advisor, Dr. Timothy J. Larson. I want to thank him for his guidance, inspiration and patience over the past four years. I also want to thank my committee, Dr. Jiann–Shin Chen, Dr. David R. Bevan, Dr. Elizabeth A. Grabau and Dr. Ann M. Stevens for their support. I want to especially thank Dr. Bevan for his help with numerous computer problems. I also want to thank Dr. Walter G. Niehaus, Jr. for providing me a solid background in biochemistry, both in the laboratory and in the classroom.

Thanks to the members of the Larson laboratory, past and present, who have made the past four years an enjoyable and memorable experience, especially Ali Bhattacharya, Janet Donahue, Farzana Ahmed, Bing Yang, Bekir Col, Andy Ewens and Ben Potters. I would also like to thank Dr. E. M. Gregory and Paul McBee for making biochemistry seem so interesting. Without their enthusiasm for the subject, I would not have charted this course for myself.

Most importantly, I want to thank my family, especially my mother, Gaynell Ray, for their unconditional love and support. I share this achievement with Jesse Duncan, my best friend, now and forever. Without him this would not have been possible nor even worthwhile.

Last, but by no means least, I want to thank the LTG. They are always in my heart. And this was called civilization …

List of Abbreviations

A	absorbance
AU	absorbance units
aga	*N*–acetylgalactosamine
Aga	*N*–acetylgalactosamine
Ap	ampicillin
(d)ATP	(deoxy)adenosine 5'–triphosphate
bp	base pairs
cAMP	cyclic adenosine 3',5'–monophosphate
Cm	chloramphenicol
CRP	cAMP receptor protein (catabolite activator protein)
dNTP	deoxyribonucleoside triphosphate
DTNB	5,5'–dithiobis (2–nitrobenzoic acid)
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetate
EF–Tu	elongation factor Tu
glp	*sn*–glycerol 3–phosphate
glycerol–P	*sn*–glycerol 3–phosphate
IMAC	immobilized metal affinity chromatography
IPTG	isopropylthio–β–D–galactopyranoside
K_m	Michaelis constant
Km	kanamycin
kb	kilobase pairs
kDa	kilodaltons
LB	Luria–Burtani broth
Man	Mannose
MCS	multiple–cloning site
MW	molecular weight
NAD(H)	oxidized (reduced) form of nicotinamide adenine dinucleotide
NADP(H)	oxidized (reduced) form of nicotinamide adenine dinucleotide phosphate
ONPG	*o*–nitrophenyl–β–D–galactopyranoside
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
PNPP	*p*–nitrophenyl–phosphate
psi	pounds per square inch
PTS	bacterial phosphotransferase system
RBS	ribosome–binding site (Shine–Dalgarno sequence)
rpm	rotations per minute
SDS	sodium dodecyl sulfate
Sp	spectinomycin
Tc	tetracycline
Tris	tris(hydroxymethyl)aminomethane

Table of Contents

List of Figures

List of Tables

LITERATURE REVIEW:

The DeoR Family of Transcriptional Regulators

In eubacteria the expression of metabolic genes is often regulated at the transcriptional level by proteins that bind to specific DNA sequences called operators. The operators are usually located at or near the site of transcription initiation (1). Binding of the regulatory protein(s) to the operators may repress or activate transcription initiation and thus repress or activate the expression of the genes transcribed from this site. The regulatory protein may interact with a metabolic intermediate of the pathway it controls. Binding of this metabolic intermediate or inducer results in a conformational change of the regulatory protein that alters the affinity of the protein for the operators. In the case of transcriptional repressors, the affinity for the operator sites is reduced, thus allowing expression of the metabolic genes (2).

Distinct families of transcriptional regulators have been identified based on amino acid sequence similarities (3–5). Members of specific families share common DNA–binding, inducer–binding and/or oligomerization motifs. Amino acid residues important for a specific protein function are often highly or even absolutely conserved among members of a family. These sequence similarities can thus define a signature sequence that is shared by all family members. One such family of transcriptional regulators is the DeoR family (6). DeoR, one of the more well–studied members of the family, regulates the expression of genes involved in the metabolism of deoxyribonucleosides in *Escherichia coli* (7). The DeoR family of proteins has increased considerably in size recently due to the systematic sequencing of bacterial genomes including those of *E. coli* (8), *Haemophilus influenzae* (9) and *Bacillus subtilis* (10). The genomes of these organisms are predicted to encode 12, 4 and 6 members of the DeoR family, respectively (BLAST search results, (11)). The DeoR family is distinct from the LacI–GalR family (3) and other characterized families of regulatory proteins (4,5). Members of the DeoR family of proteins contain from 250 to 270 amino acid residues and are typically 25 to 35% identical to one another. They contain a highly–conserved region near the amino–terminus that includes a helix–turn–helix DNA–binding motif (6). This highly conserved region serves as the signature sequence for the family. The signature sequence is:

———— recognition helix ————

$$RX_3[LIM]X_3[LIV]X_{18}\ X_2[TN][IVA]\mathbf{R}[RKANQ][DE][LIVF]$$

where X can be any amino acid residue and brackets indicate alternative residues at a particular position. This sequence has been modified from that suggested by Saier et al. (12) to include a number of putative proteins recently identified with significant sequence similarity to other DeoR family members (BLAST search results, (11)). The invariant arginine residue (shown in bold above) within the recognition helix has been shown to be essential for binding to DNA in GlpR, the regulator of *sn*–glycerol 3–phosphate metabolism in *E. coli* (6).

To date there are over 30 members of this family of proteins in a variety of bacterial species including Gram–negatives such as *E. coli* and *Haemophilus influenzae*,

Gram–positives such as *Bacillus subtilis* and *Lactococcus lactis*, and the *Mycoplasma capricolum* (BLAST search results, (11)). Due to the dramatic pace of sequencing projects presently underway, new members are being identified frequently.

A function has been identified for approximately 20 members of the DeoR family based on either experimental evidence or sequence similarity to a protein from another species with an experimentally–defined function. Those with a defined function regulate the metabolism of a diverse group of compounds including glycerol (13), glucitol (14) and galactitol (15) in *E. coli*, inositol in *B. subtilis* (16), lactose in *L. lactis* (17), *Staphylococcus aureus* (18) and *Streptococcus mutans* (19) and opines in *Agrobacterium tumefaciens* (20,21). All but two members of the family with a defined function serve as repressors of transcription. The known exceptions are the fucose operon activators, FucR, of *H. influenzae* (9) and *E. coli* (22). One member of previously undefined function in *E. coli* has recently been proposed to regulate the metabolism of *N*–acetylgalactosamine (AgaR, YhaW) (12). The functions of the remaining family members, six of which are found in *E. coli*, are not known. The six regulators of undefined function are encoded throughout the *E. coli* genome [YihW at minute 87.7, SgcR at minute 97.5 (23), YgbI at minute 61.6, YciT at minute 28.9, YjfQ at minute 95.2 and YdjF at minute 40.0]. The regulatory genes are generally transcribed divergently from or are embedded within other clusters of genes of undefined function. Some of these nearby genes are likely to be controlled by the trancriptional regulators and may encode enzymes responsible for as yet unidentified metabolic pathways. Figure 1 is a phylogenetic tree illustrating the evolutionary relationships among the presently–identified members of the DeoR family.

The carboxy–terminal regions of proteins in the DeoR family are responsible for protein oligomerization and binding of the inducer (24,25). Where identified, the inducer for the regulators is a phosphorylated intermediate in the metabolic pathway they control. DeoR is induced by deoxyribose–5–phosphate (26), FucR is induced by fuculose–1–phosphate (27), GlpR is induced by *sn*–glycerol 3–phosphate (13) and LacR, the regulator of lactose metabolism in *L. lactis,* is induced by tagatose–6–phosphate (25).

A highly conserved region near the carboxy–terminus had been identified as a putative inducer–binding site (17); however, recent evidence suggests that residues throughout the carboxy–terminus are important for binding the inducer. Residues important for inducer response in LacR include lysines 72, 80 and 213 and aspartic acid 210 (25). Expression of variants of LacR possessing substitutions at these residues in a *lacR* strain resulted in constitutive repression of phospho–β–galactosidase activity, the gene product of *lacB*. These strains also exhibited a significantly slower growth rate when grown on lactose (approximately 2 to 3–fold increase in generation time) due to constitutive repression of genes involved in lactose metabolism. Gel mobility shift assays were also used with the purified variants to show that binding to operator regions by these variants had become insensitive to the presence of the inducer, tagatose–6–phosphate (25).

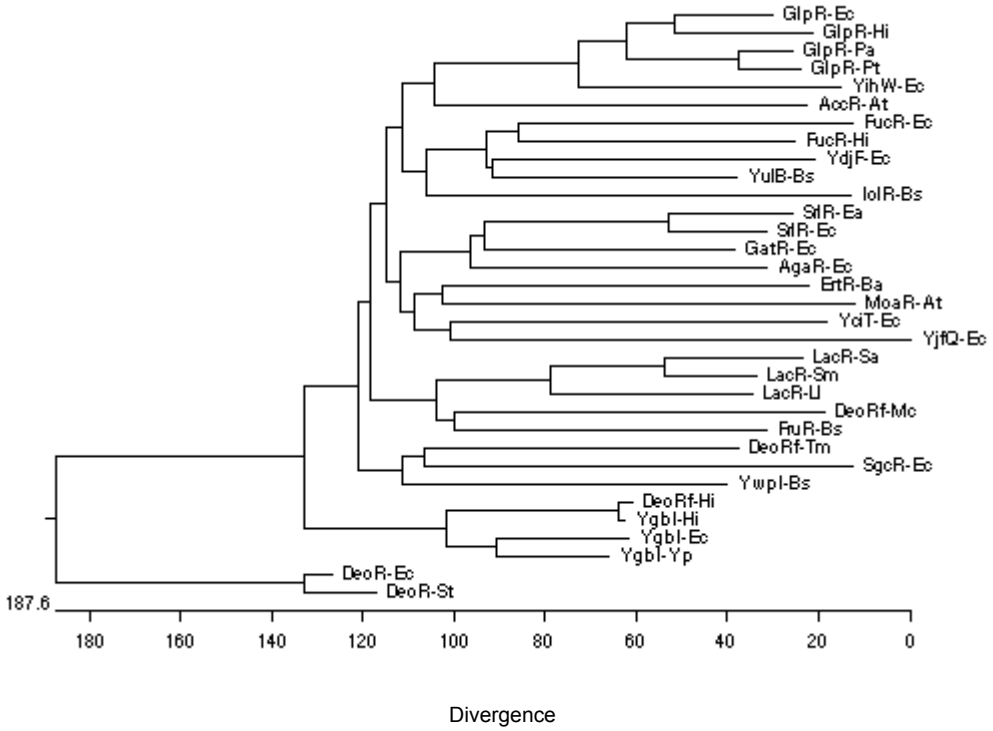


Figure 1. Phylogenetic tree of the DeoR family. Protein names are given when one has been assigned, otherwise the proteins are designated as DeoRf for member of the DeoR family. Abbreviations for species are: Ec, *Escherichia coli*; Hi, *Haemophilus influenzae*; Pa, *Pseudomonas aeruginosa*; Pt, *Pseudomonas tolaasii*; At, *Agrobacterium tumefaciens*; Bs, *Bacillus subtilis*; Ea, *Erwinia amylovora*; Ba, *Brucella abortus*; Sa, *Staphylococcus aureus*; St, *Streptococcus mutans*; Ll, *Lactococcus lactis*; Mc, *Mycoplasma capricolum*; Tm, *Thermotoga maritima*; Yp, *Yersinia pestis*. The sequence alignment was performed using the Clustal method with PAM250 residue weight table utilizing Lasergene.

In GlpR, a number of residues including threonine 101 and aspartic acid 207 (corresponding to aspartic acid 210 of LacR) have been shown to be important for inducer response using similar criteria (24). Most of these amino acid residues are conserved among the members of the DeoR family. The conserved residues probably function similarly in all members of this family. Figure 2 is a sequence alignment of some representative members of the DeoR family. Residues experimentally determined to play a role in DNA binding or inducer response are highlighted.

Oligomerization is believed to be important for efficient repression by GlpR and DeoR. Under native conditions, GlpR is tetrameric (13) and DeoR is hexameric or octameric (7,28). For some of the promoters controlled by these repressors it has been shown that widely–spaced operators are important for efficient repression (29–33). This is generally believed to be due to the repressor oligomer binding simultaneously to the widely–spaced operators resulting in looping of the DNA around the site of transcription initiation. Looping of DNA is utilized as a means of efficient repression by a number of other oligomeric

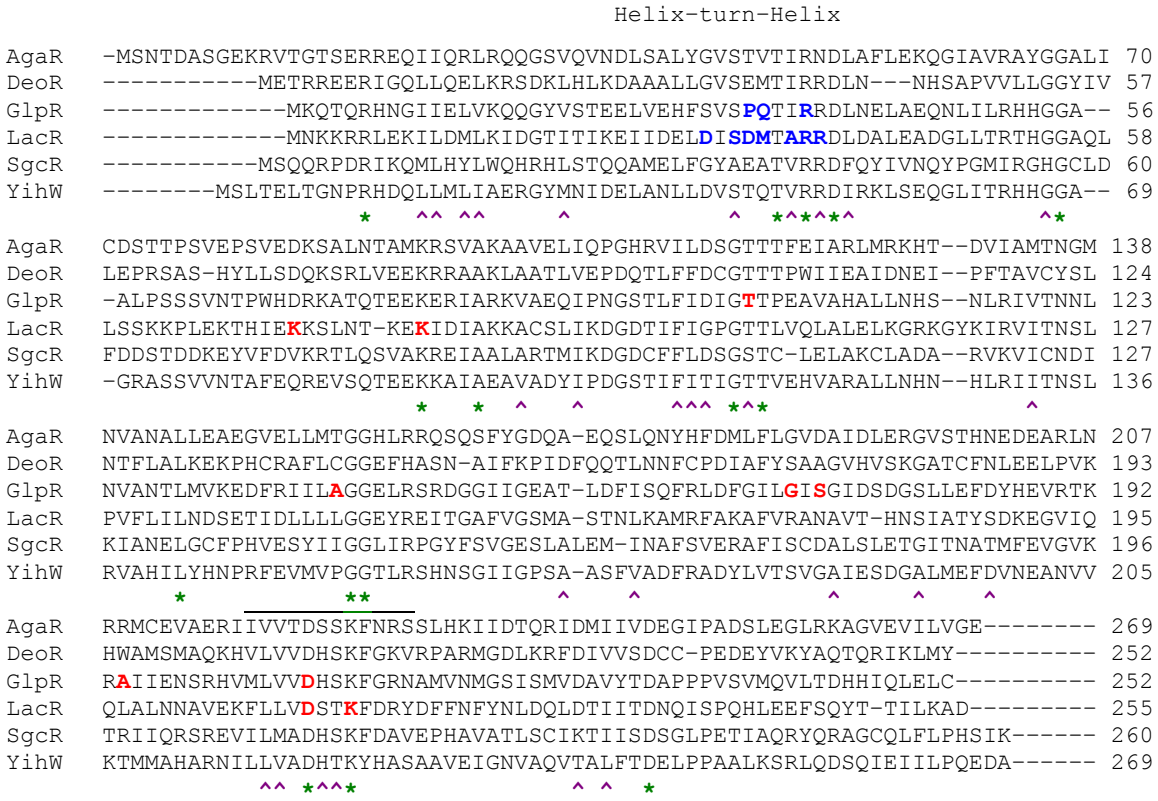


Figure 2. Alignment of AgaR with five other representative members of the DeoR family. With the exception of LacR all are from *E. coli*. AgaR, in this work, is shown to regulate the metabolism of *N*–acetylgalactosamine, as had been previously predicted (12). DeoR and GlpR are the repressors of the *deo* and *glp* systems, responsible for metabolism of deoxyribonucleosides (29) and glycerol (13), respectively. LacR is the *lac* repressor from *L. lactis* which regulates expression of genes involved in lactose metabolism (17). SgcR (23) and YihW are putative proteins of unknown function. Completely conserved and highly conserved residues are indicated with "*" and "^", respectively. Positions in GlpR or LacR where substitutions were found to cause a defect in induction are **red** (24,25). The residues within the helix–turn–helix motif shown to play a role in DNA binding are **blue** (6,39). Previously proposed inducer binding region is overscored (17).

repressors including LacI, GalR, NagC and the lambda repressor (34–37). The nature of the interactions resulting in oligomerization for proteins of the DeoR family has not been elucidated, but residues important for oligomerization are generally believed to be in the carboxy terminal region of these proteins (6).

The molecular details of the transcriptional regulation mediated by members of the DeoR family has been elucidated for only a small number of the proteins. Table 1 illustrates a comparison of the recognition helices of the better–studied members of this family and the corresponding operator sequences recognized by the proteins. Extensive characterization of the molecular details of regulation has only been determined for DeoR and GlpR and the operator sequences shown are consensus sequences determined by comparison of a number of operator sites verified by footprinting studies (28,38). Note the striking similarities between

both the recognition helices and the consensus operator sequences for DeoR and GlpR. Despite these striking similarities, variants of GlpR possessing amino acid substitutions in the recognition helix to match that of DeoR even more closely (P33E/Q34M), were not able to recognize *deo* operators. In fact, these variants had also lost the ability to recognize *glp* operators and instead conferred a strong negatively–dominant phenotype to the chromosomally–encoded repressor. These results indicated that the variants retained the ability to oligomerize and also showed that the putative recognition helix was important for DNA–binding. It also suggested that more than just the primary structure of the proteins is important for operator recognition (6).

TABLE 1

Comparison of recognition helices and operator sites for DeoR proteins

Protein	Operator half–site[a]	Recognition Helix
GlpR	WATGTTCGWT	PQTIRRDLNN
DeoR	TATGTTAGAA	EMTIRRDLNE
IolR	WRAYCAADARD	KNTVRRDINK
GutR	WAWCTTTCAWW	GTTIRKDLVI
LacR	TGTTTGATAT	DMTARRDLDA

[a] The sequence of the left half of the palindromic consensus operator is given for GlpR, DeoR and LacR. Sequences shown for IolR and GutR are found as direct repeats. D is A, G or T; R is A or G; W is A or T; and Y is C or T. References are given in the text.

Footprinting has also been used to identify site(s) recognized by IolR of *B. subtilis* (40), GutR of *E. coli*, and LacR of *L. lactis*. The footprinting analysis using GutR identified a relatively large section of the *gutABD* regulatory region protected by GutR (over 100 bp) (41). The putative operator half–site sequence reported in Table 1 is derived from a repeating sequence within this protected region. The operator half–site sequence given for LacR from *L. lactis* was tentatively identified by determining the effects of flanking regions on the promoter activity for the *lacABCDFEGX* operon (42) and more recently by footprinting (39). DeoR, GlpR and LacR bind to palindromic sequences that in some cases are found as direct repeats and in others as widely–separated sites (28,38,42). IolR and GutR bind to direct repeats (40,41). There appears to be no common theme among DeoR family members for binding to DNA except for the possibility of DNA looping.

Systematic sequencing of the genomes of a number of organisms as described above has increased the knowledge of gene organization and distribution but has left one major question unanswered: What is the function of the many new proteins revealed by sequence analysis? The long–range goal of this research is to determine the function of and characterize the regulation mediated by the seven proteins of unknown function belonging to the DeoR family encoded by the *E. coli* genome. This group of proteins includes AgaR, which has been proposed to regulate *N*–acetylgalactosamine metabolism (12). In this work I verify that AgaR functions to negatively regulate the expression of a cluster of genes involved in *N*–acetylgalactosamine metabolism and characterize the regulation mediated by AgaR (Section I). I have also shown that another member of the DeoR family of undefined function, YihW, negatively regulates the transcription of genes adjacent to the gene encoding YihW (Section II). I also show that though YihW is closely related to GlpR (Figure 1), YihW does not function as an isoform of GlpR (Section II).

The objective of this research was to develop a standard approach to be used for the elucidation of the function of the proteins belonging to the DeoR family. This approach is based on previous work in this laboratory with GlpR (6,13,24). As mentioned above, two important types of GlpR variants have been generated: i) those that do not bind DNA operator sites, and ii) those that remain bound to operator sites even in the presence of the inducer, glycerol–P. These variants contain amino acid substitutions at highly conserved positions of the DeoR family (Figure 2). Important for the proposed work is the fact that both types of repressor variant, when encoded by multicopy plasmids, are dominant to the chromosomally–encoded wild–type repressor since members of the DeoR family require oligomerization to efficiently repress transcription. Thus, the first type of variant (negatively dominant) confers a constitutive phenotype due to formation of inactive mixed oligomers (6), and the second type confers a noninducible, or growth–negative, phenotype (24). Phenotypic testing of cells harboring either negatively dominant or noninducible variants of the repressors will facilitate determination of the function of these proteins. A difference in growth phenotype on a particular carbon source will indicate that the repressor is involved in the regulation of the corresponding metabolic pathway. Important for this objective is the ability to create variants of the repressor proteins that confer the negatively dominant or noninducible phenotypes. In Section I, I show that both negatively dominant and noninducible variants of AgaR could be constructed corresponding to analogous variants previously described for GlpR. Similar variants have also been tentatively identified for YihW (Section II).

It is also the goal of this research to assign activities to the enzymes encoded by genes controlled by the DeoR family of transcriptional repressors. Comparsion of the amino acid sequences to sequences of proteins of defined function combined with the knowledge of the carbon source metabolized by the proteins will allow for the assignment of activities. It is interesting to note that although the regulation mediated by GlpR has been well–characterized, the functions of several proteins encoded by genes belonging to the GlpR regulon remain unknown. In this work, I report that one of those proteins, GlpE, functions as a rhodanese (thiosulfate:cyanide sulfurtransferase). I also describe the purification and characterization of GlpE and discuss possible physiological roles for this enzyme (Section III).

SECTION I:
Characterization of AgaR and the Regulation of *N*–Acetylgalactosamine Metabolism

Introduction

The *aga* gene cluster, a group of 10 genes located at minute 70.7 of the *Escherichia coli* genome, has been proposed to encode proteins involved in the metabolism of *N*–acetylgalactosamine. Divergently transcribed from this gene cluster is a gene, *agaR*, proposed to encode a transcriptional regulator of this gene cluster (12). The protein encoded by *agaR* (AgaR) belongs to the DeoR family of transcriptional regulators (12). Figure 3(a) illustrates the organization of the *aga* gene cluster in *E. coli*.

Within this cluster of genes are five that encode homologues of the "splinter group" of bacterial phosphotransferase system (PTS) proteins that includes the mannose (*man*) enzyme II complex of *E. coli* (12). These genes are *agaB* and *agaV* encoding enzyme IIB^{Man} homologues; *agaC* and *agaW*, enzyme IIC^{Man} homologues; and *agaD*, an enzyme IID^{Man} homologue. The bacterial phosphotransferase system is a group of proteins involved in sugar uptake and phosphorylation. For a review see Saier and Reizer (43).

Also included in this cluster are three genes encoding proteins having significant sequence similarity to either deacetylases (*agaA*), isomerizing deaminases (*agaI*) or aldolases (*agaY*). The remaining two genes include one encoding a protein previously proposed to be a carbohydrate kinase (*agaZ*) and one encoding a protein having limited sequence similarity to the carboxy–terminal region of glucosamine–6–phosphate synthases (*agaS*). It is the carboxy–terminal domain of the glucosamine–6–phosphate synthases that possesses the ketose–aldose isomerase activity (12). These enzymes, such as the protein encoded by *glmS*, function in the synthesis of *N*–acetylglucosamine and glucosamine for incorporation into peptidoglycan and lipopolysaccharides (45).

Figure 4 illustrates the proposed metabolic pathway for the utilization of *N*–acetylgalactosamine. This pathway is based on the metabolism of the similar sugar derivative *N*–acetylglucosamine and the sequence similarities of proteins encoded by genes within the *aga* cluster to proteins of defined function (12) (J. W. Lengeler, personal communication). Also shown in Figure 4 is the parallel pathway for the catabolism of galactitol (dulcitol) (15) (J. W. Lengeler, personal communication). The pathways for *N*–acetylgalactosamine and galactitol converge at the intermediate tagatose–6–phosphate.

The phosphorylation of tagatose–6–phosphate has been found to be catalyzed, but relatively slowly, by phosphofructokinase I (encoded by *pfkA*) which is suggested to function more generally as a phosphoketose kinase (15,46). Phosphofructokinase II (encoded by *pfkB*) can also phosphorylate tagatose–6–phosphate and can substitute for phosphofructokinase I in galactitol metabolism when overexpressed (15). Despite having no sequence similarity to known sugar or sugar–phosphate kinases, the proteins encoded by *gatZ* and *agaZ* (which are 53% identical to each other) were proposed to encode enzymes that would also catalyze this

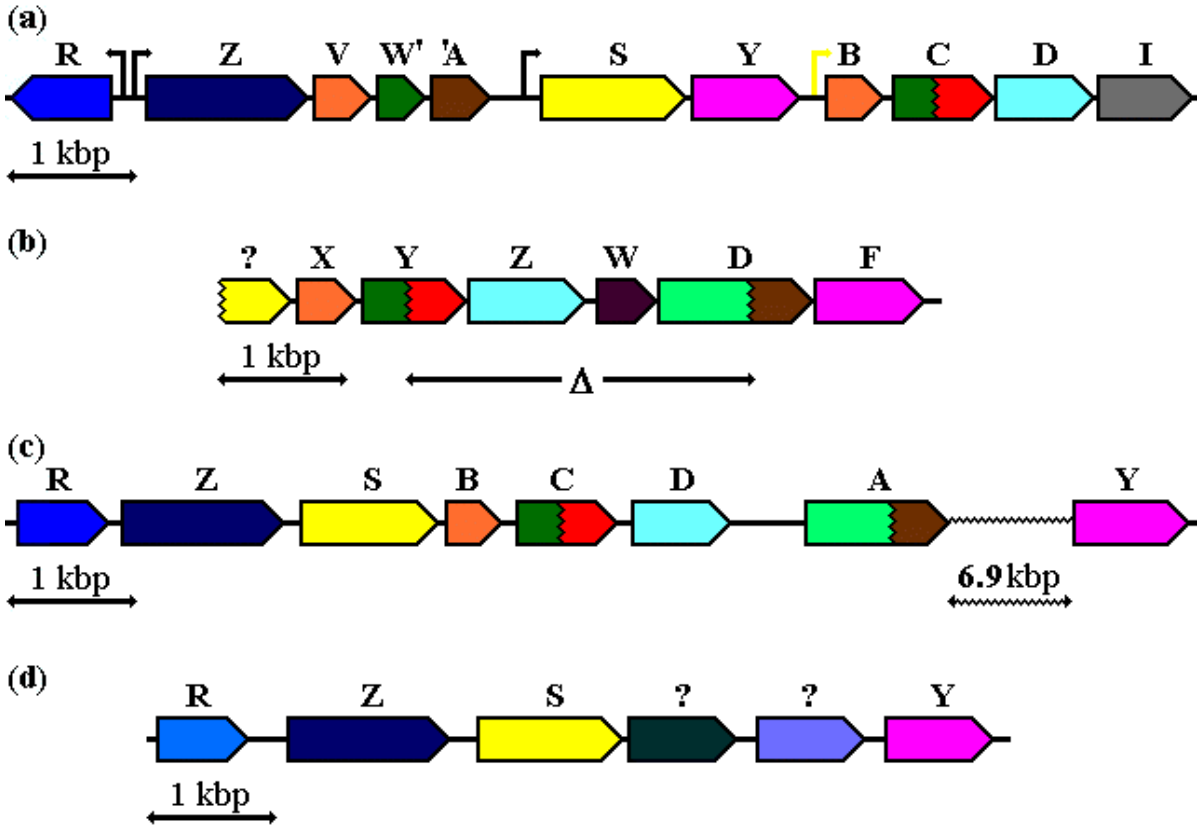


Figure 3. The *aga* gene cluster of *E. coli* and similar clusters identified in other bacteria. Genes or portions of genes possessing sequence similarity are represented using the same color. (a) The *aga* gene cluster of *E. coli* (12). Transcriptional promoters verified in this study are shown by black arrows. The yellow arrow adjacent to *agaB* represents a transcriptional promoter predicted by the *E. coli* Genome Sequencing Project that has not yet been verified. The apostrophes adjacent to *agaW* and *agaA* indicate that these open reading frames are truncated at the C– and N–terminal coding regions, respectively. (b) The mannose operon of *Vibrio furnissii*. The *manX–manD* genes are believed to have been incorporated into the *E. coli* genome by a horizontal transfer event. Genes or portions of genes believed to later have been deleted after incorporation into *E. coli* are indicated by Δ (44). (c) The *aga* gene cluster of *Yersinia pestis*. (d) The *aga* gene cluster of *Streptomyces coelicolor*. The protein encoded by *agaR* of *S. coelicolor* belongs to the GntR family of transcriptional regulators.

phosphorylation (12). Recently, it was shown that no increase in tagatose–6–phosphate kinase activity was associated with overexpression of GatZ, and a strain deleted for both phosphofructokinase I and II (*pfkA* and *pfkB*) remained galactitol negative when GatZ was overexpressed from a plasmid. It was shown, however, that GatZ is required for galactitol utilization (15). Recent evidence has shown that *agaZ* and *gatZ* encode subunits of a heteromultimeric tagatose–1,6–bisphosphate aldolase, the second subunits of which are encoded by *agaY* and *gatY*, respectively (J. W. Lengeler, personal communication).

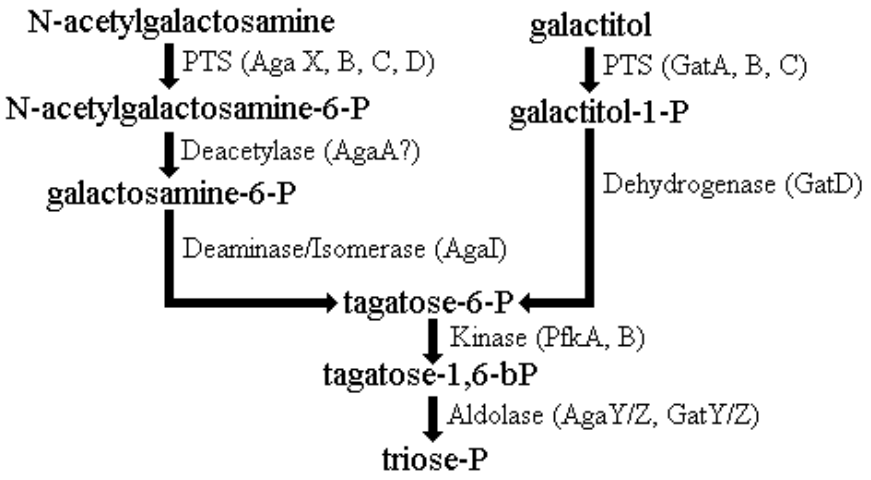


Figure 4. Metabolic pathways for *N*–acetylgalactosamine and galactitol. Metabolic pathway suggested for *N*–acetylgalactosamine by Saier et al. (12) with revisions suggested by J. W. Lengeler et al. (personal communication and (15)) and the metabolic pathway for galactitol as proposed by J. W. Lengeler et al. (personal communication and (15)). The deacetylase encoded by *agaA* is a truncated form possessing only the carboxy terminal region and may not be functional (see Figure 3). The *aga* cluster does not possess a gene encoding a homologue of the enzymes IIA^{Man} of the PTS. A coding region located in the 2.4 to 4.1 minute region of the *E. coli* genome encodes a homologue of the IIA^{Man} enzyme and was suggested by Saier et al. (12) to function as the IIA^{Aga}. The reading frame was tentatively given the name *agaX* (12). Recent corrections in the sequence of this region show that *agaX* (*yadI*) encodes a full–length homologue and not a truncated version as previously reported by Saier et al. (12). The role played by the protein encoded by *agaS* in the metabolism of *N*–acetylgalactosamine is unclear (12).

Metabolism of galactitol in many strains of *E. coli* has been found to be temperature–sensitive (46). A heat–labile tagatose–1,6–bisphosphate aldolase encoded by *gatY* and *gatZ* is believed to be responsible for this phenotype. Derepression of *agaY*, encoding a protein 54% identical to GatY, and *agaZ* was shown to complement the temperature–sensitive aldolase encoded by *gatY*/*gatZ* in revertants capable of metabolizing galactitol at higher temperatures (J. W. Lengeler et al., personal communication and (15)). The genetic locus specific for temperature–resistant growth on galactitol was previously named *kba* for ketose–bisphosphate aldolase and mapped to the *aga* gene cluster (15,46).

Oddly, the *aga* gene cluster contains two PTS enzyme IIB^{Man} homologues, *agaB* and *agaV*. The cluster also possesses two PTS enzyme IIC^{Man} homologues, *agaC* and *agaW* (12). It has been proposed that *agaV* and *agaW* as well as *agaA* are genes that originated in *Vibrio furnissii* and were incorporated into the *E. coli* genome as the result of a horizontal transfer event (44). The sequence identities between genes within the region of the *aga* cluster believed to have been incorporated from *V. furnisii* and genes encoding a putative deacetylase and the mannose PTS enzymes IIB and IIC in *V. furnisii* are striking (approximately 70 to 80%) (44). Figure 3(b) illustrates the organization of the mannose operon in *V. furnisii*. Genes flanking this region of the mannose operon of *V. furnisii* incorporated into the *aga* cluster have sequence similarity to the genes *agaS* and *agaY* (44). These similarities suggest

the transfer may have resulted from an homologous recombination event. It is interesting to note that during the course of my work I found that many strains of *E. coli* lack the ability to utilize *N*–acetylgalactosamine as a sole source of carbon without prior selection. The disruption of the *aga* gene cluster by the horizontal transfer of genes from *V. furnisii* may be a factor contributing to the apparent crypticity of the *aga* genes.

A search of bacterial genomic sequences revealed an *aga* gene cluster in *Yersinia pestis*, a Gram–negative bacterium closely related to *E. coli* (Figure 3(c)). An *aga* cluster has also been identified in the genome of *Streptomyces coelicolor*, a less closely related Gram–positive bacterium. In these gene clusters the *agaZ* gene is immediately followed by the *agaS* gene, an organization that is predicted to have existed in ancestral *E. coli* before the transfer of genes from *V. furnisii* (44). Only one set of PTS enzyme IIB^{Man}/IIC^{Man} homologues are present in the *aga* gene cluster of *Yersinia pestis*. These facts support the hypothesis that a horizontal transfer event has led to the present organization of the *aga* gene cluster in *E. coli*.

The protein encoded by *agaW* is not a full–length PTS enzyme IIC^{Man} homologue. The protein is homologous only to the amino terminus of these proteins. The protein encoded by *agaA* possesses significant sequence similarity to deacetylases, most notably NagA which functions in the metabolism of *N*–acetylglucosamine. However, AgaA is also truncated, possessing only the carboxy terminus (12). A deletion of part of this region (shown in red in Figure 3(b)) after incorporation into the *E. coli* genome is believed to be the cause of the apparent truncation of *agaA* and *agaW* in *E. coli* (44). In contrast, a full–length protein is encoded by *agaA* of *Y. pestis* (Figure 3(c)). There is some question whether the apparently truncated protein encoded by *agaA* in *E. coli* could be functional (12) and this may also play a role in the inability of many strains to utilize *N*–acetylgalactosamine as a sole source of carbon without previous selection.

In this study I have verified the proposal by Saier et al. (12) that the *aga* gene cluster is involved in the metabolism of *N*–acetylgalactosamine and characterized the regulation of the expression of the genes within this cluster. Utilizing transcriptional fusions of three promoters identified in this cluster to *lacZ*, it was shown that the expression of genes within the *aga* gene cluster is specifically induced by the presence of *N*–acetylgalactosamine or galactosamine in the growth medium. By comparing the activities of these transcriptional fusions in strains with and without a functional gene for the repressor, it was also confirmed that AgaR functions as a repressor to regulate the expression of this gene cluster. I have mapped the transcription start sites for each of the three promoters (see Figure 3(a)) identified in this gene cluster. AgaR, purified to near homogeneity, has been used for DNase I footprint analysis to identify specific sites recognized by AgaR. Four operator sites have been identified within the *aga* promoter regions and from these four sites a consensus sequence recognized by AgaR has been proposed. It has also been shown that the promoter specific for *agaZ* requires the cAMP–cAMP receptor protein (CRP) complex, which positively regulates many metabolic genes in *E. coli* (47), for full activation. The site recognized by this complex has been identified by footprint analysis. I have also attempted to determine why K–12 strains of *E. coli* lack the ability to utilize *N*–acetylgalactosamine as a sole source of carbon and what mutation(s) occur in revertants that gain this ability.

Experimental Procedures

Materials and reagents. Unless explicitly listed below, the reagents used were purchased from either Sigma Chemical Company or Fisher Scientific. DuPont–New England Nuclear supplied [α–^{35}S]dATP and [γ–^{33}P]ATP. Restriction endonucleases and reagents for PCR and cloning were supplied by New England Biolabs. Imidazole and ammonium sulfate, both supplied by Sigma Chemical Company, were enzyme grade. cAMP receptor protein (CRP) used was that previously purified in this laboratory (33).

Synthetic oligonucleotides were prepared using an Applied Biosystems DNA Synthesizer (model 381A) using reagents supplied by Cruachem. Alternatively, some oligonucleotides were purchased and used without additional purification from DNAgency, Gibco BRL or Genosys, Inc. Synthetic oligonucleotides used in the study of AgaR are listed in Table 2. Position numbers given in Table 2 correspond to those shown in either Figure 5 (nucleotide sequence of *agaR* and the *agaR*/*agaZ* intergenic region) or Figure 6 (nucleotide sequence of the *agaA*/*agaS* intergenic region) unless noted otherwise.

Bacterial strains. Strains of *E. coli* used or constructed are described in Table 3. All strains except BL21(DE3) are K–12 derivatives. BL21(DE3), a B strain, was used for overexpression of AgaR or variants containing a hexahistidine tag expressed from derivatives of pGZ117 (a derivative of pT7–7) (48). DH5αZ1 (49) was used as a host during plasmid construction. It was also used as the host strain for β–galactosidase assays of transcriptional fusions when present in multicopy on pSP417 derivatives and for overexpression of AgaR or variants expressed from plasmids derived from the pZ family of expression plasmids (49).

TL524 was used as the host to prepare lysogens of the λ phages carrying the transcriptional fusions transferred from the pSP417 derivatives by homologous recombination. TL524 was also used for selection of revertants capable of utilizing *N*–acetylgalactosamine as a sole source of carbon (KR201 to KR205). TL524 is a tetracycline–sensitive derivative of TL504 that was obtained as described by Maloy and Nunn (59). TL504 is a *ΔlacZ* derivative of the wild–type strain MG1655 described previously (55). KR161, KR162 and KR164 are single λ lysogens of TL524 containing *agaR'–lacZ*, *agaZ'–lacZ* and *agaS'–lacZ* transcriptional fusions derived from pKR161, pKR162 and pKR164, respectively. The method for preparing the λ lysogens and identification of single lysogens was as described previously (31,60).

KR1161, KR1162 and KR1164 are derivatives of the lysogenic strains KR161, KR162 and KR164, respectively, in which the *agaR* gene has been disrupted. KR1524 is a derivative of TL524 in which the *agaR* gene has been disrupted. The construction of these strains is described below.

LE392 (56) was used for propagation of phage λDD704 (8,61). CAG12072 (CGSC# 7440) and CAG12127 (CGSC # 7441) contain Tn*10* and Tn*10kan* insertions, respectively, at minute 71.8, near the *aga* gene cluster (57,58). These strains were used for P1 transduction mapping studies. P1 transduction was performed using a previously published method (62).

TABLE 2

Oligonucleotides used in the study of AgaR

Name	Sequence (5' → 3')[a]	Position[b]
329890	TGCGTgAATtcGCAAGGTGGGCTTG	537–561
336849	ATGGTGTctAgAATGCCCGCCGAAC	1495–1471
372696	AAGGACTTcaTATGAGTAATACCGcaG	620–646
258920	TTCTCcgTCgaCGACCAGAATCACTT	1443–1418
372667	GATCCGCAACGATCTGG	771–787
386252	CATCTTCATTATGTGTGC	1237–1220
372597	TTCCGGGAtCACCACTTTTGAGATTG	969–994
331644	AAGTGGTGaTCCCGGAATCGAGGATA	985–960
366895	GATGTTTCAaTTgCCTCTCCAGTTCAG	373–399
386247	AGCGTCGGaATTcCTCAgAGTAAGTCC	648–622
WKR1	AAACGCGCaAttGTCGTTGCGCTG	23–46[c]
WKR2	GATGGCGGAatTCTTCTTCAGTCC	578–555[c]
WKR3	GTGACGATCgcCAACGATCTGGCGTT	766–791
WKR4	CAGATCGTTGgcGATCGTCACAGTAG	786–761
WKR5	CAAGGTGGAaTTcTTTGAATCCTGCC	4–29
WKR6	GCTCAGGGGAGTAAACAGGAGACAAG	2781–2805[d]
Bing7	CCCGGGATGCATGCGGCCGC	[e]
agaC–A	TTGCTCTTCCATG…	[f]
agaC–C	TTGCTCTTCGTTA…	[f]
agaV–A	TTGCTCTTCCATG…	[f]
agaV–C	TTGCTCTTCGTTA…	[f]

[a] Lower case letters indicate base substitutions to create restriction sites or mutations.

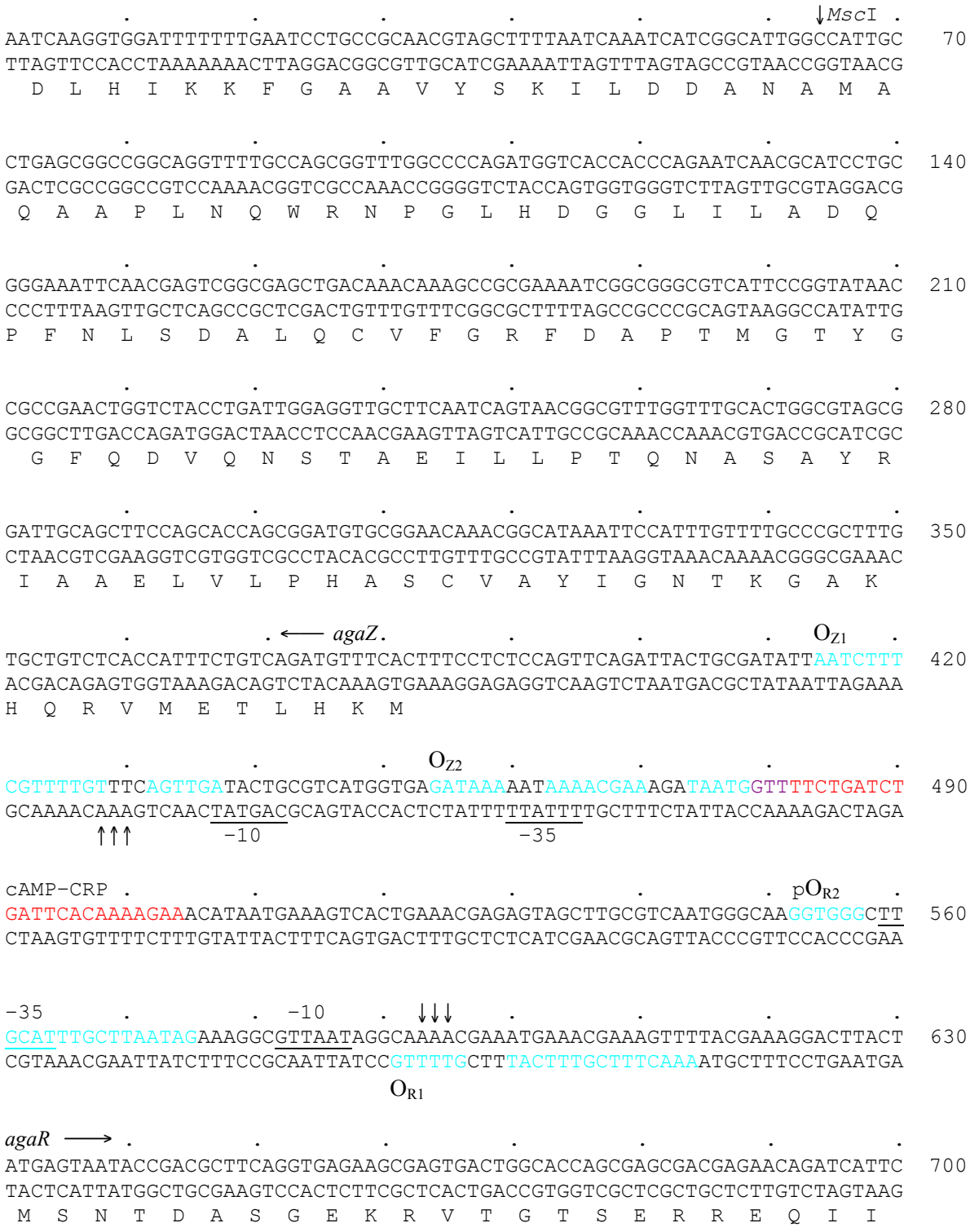
[b] Unless indicated otherwise, position numbers correspond to those given in Figure 5.

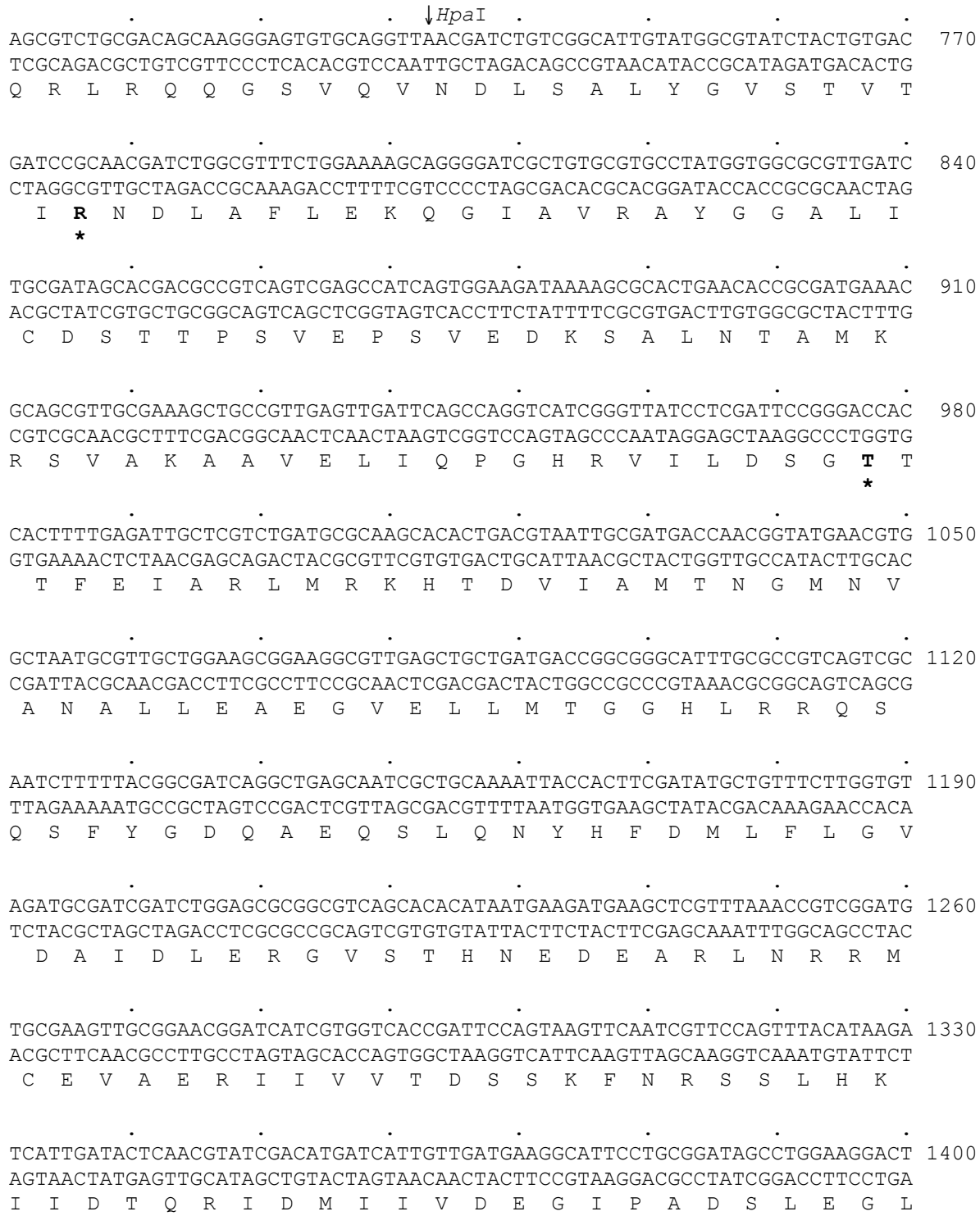
[c] Position numbers correspond to those given in Figure 6.

[d] Position numbers correspond to the complete sequence of transposon Tn10 (GenBank accession number J01830). This oligonucleotide corresponds to a sequence located near the 3' terminus of the *tetA* gene contained within the tetracycline–resistance cassette of pCP16 and is directed out of the cassette (50).

[e] This oligonucleotide is complementary to the multiple–cloning site of pSP417 (51) from *Sma*I to *Not*I and is directed away from the *lacZ* gene.

[f] These pairs of oligonucleotides (ORFmers™), purchased from Genosys, Inc., are designed to amplify the open reading frames indicated. The 5' sequences are shown. The full–length sequences are proprietary information but match the coding sequence for the given ORF. A corresponds to the amino–terminal primer and C corresponds to the carboxy–terminal primer.


↓MscI
70
D L H I K K F G A A V Y S K I L D D A N A M A
140
Q A A P L N Q W R N P G L H D G G L I L A D Q
210
P F N L S D A L Q C V F G R F D A P T M G T Y G
280
G F Q D V Q N S T A E I L L P T Q N A S A Y R
350
I A A E L V L P H A S C V A Y I G N T K G A K
agaZ
OZ1
420
H Q R V M E T L H K M
OZ2
490
−10
−35
cAMP-CRP
pOR2
560
−35
−10
630
OR1
agaR
700
M S N T D A S G E K R V T G T S E R R E Q I I


↓HpaI
AGCGTCTGCGACAGCAAGGGAGTGTGCAGGTTAACGATCTGTCGGCATTGTATGGCGTATCTACTGTGAC 770
TCGCAGACGCTGTCGTTCCCTCACACGTCCAATTGCTAGACAGCCGTAACATACCGCATAGATGACACTG
Q R L R Q Q G S V Q V N D L S A L Y G V S T V T
GATCCGCAACGATCTGGCGTTTCTGGAAAAGCAGGGGATCGCTGTGCGTGCCTATGGTGGCGCGTTGATC 840
CTAGGCGTTGCTAGACCGCAAAGACCTTTTCGTCCCCTAGCGACACGCACGGATACCACCGCGCAACTAG
I R N D L A F L E K Q G I A V R A Y G G A L I
*
TGCGATAGCACGACGCCGTCAGTCGAGCCATCAGTGGAAGATAAAAGCGCACTGAACACCGCGATGAAAC 910
ACGCTATCGTGCTGCGGCAGTCAGCTCGGTAGTCACCTTCTATTTTCGCGTGACTTGTGGCGCTACTTTG
C D S T T P S V E P S V E D K S A L N T A M K
GCAGCGTTGCGAAAGCTGCCGTTGAGTTGATTCAGCCAGGTCATCGGGTTATCCTCGATTCCGGGACCAC 980
CGTCGCAACGCTTTCGACGGCAACTCAACTAAGTCGGTCCAGTAGCCCAATAGGAGCTAAGGCCCTGGTG
R S V A K A A V E L I Q P G H R V I L D S G T T
*
CACTTTTGAGATTGCTCGTCTGATGCGCAAGCACACTGACGTAATTGCGATGACCAACGGTATGAACGTG 1050
GTGAAAACTCTAACGAGCAGACTACGCGTTCGTGTGACTGCATTAACGCTACTGGTTGCCATACTTGCAC
T F E I A R L M R K H T D V I A M T N G M N V
GCTAATGCGTTGCTGGAAGCGGAAGGCGTTGAGCTGCTGATGACCGGCGGGCATTTGCGCCGTCAGTCGC 1120
CGATTACGCAACGACCTTCGCCTTCCGCAACTCGACGACTACTGGCCGCCCGTAAACGCGGCAGTCAGCG
A N A L L E A E G V E L L M T G G H L R R Q S
AATCTTTTTACGGCGATCAGGCTGAGCAATCGCTGCAAAATTACCACTTCGATATGCTGTTTCTTGGTGT 1190
TTAGAAAAATGCCGCTAGTCCGACTCGTTAGCGACGTTTTAATGGTGAAGCTATACGACAAAGAACCACA
Q S F Y G D Q A E Q S L Q N Y H F D M L F L G V
AGATGCGATCGATCTGGAGCGCGGCGTCAGCACACATAATGAAGATGAAGCTCGTTTAAACCGTCGGATG 1260
TCTACGCTAGCTAGACCTCGCGCCGCAGTCGTGTGTATTACTTCTACTTCGAGCAAATTTGGCAGCCTAC
D A I D L E R G V S T H N E D E A R L N R R M
TGCGAAGTTGCGGAACGGATCATCGTGGTCACCGATTCCAGTAAGTTCAATCGTTCCAGTTTACATAAGA 1330
ACGCTTCAACGCCTTGCCTAGTAGCACCAGTGGCTAAGGTCATTCAAGTTAGCAAGGTCAAATGTATTCT
C E V A E R I I V V T D S S K F N R S S L H K
TCATTGATACTCAACGTATCGACATGATCATTGTTGATGAAGGCATTCCTGCGGATAGCCTGGAAGGACT 1400
AGTAACTATGAGTTGCATAGCTGTACTAGTAACAACTACTTCCGTAAGGACGCCTATCGGACCTTCCTGA
I I D T Q R I D M I I V D E G I P A D S L E G L

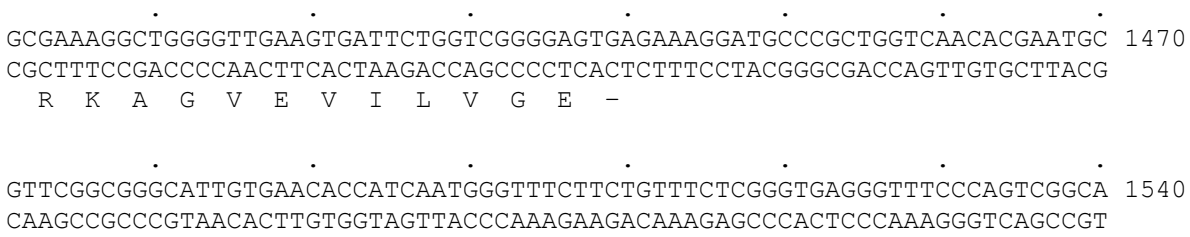


Figure 5. Nucleotide sequence of *agaR* and *agaR–agaZ* promoter region. Sequence is from accession number AE000394 of the Entrez database available at NCBI. Open reading frames are labeled. The footprinted regions protected by AgaR are shown in **blue** (O_{R1}, O_{Z1} and O_{Z2}) as is the putative operator site not verified experimentally but having sequence similarity to the proposed AgaR consensus binding sequence (pO_{R2}). The sequence protected by cAMP–CRP is shown in **red**. Where the site recognized by cAMP–CRP overlaps O_{Z2} is shown in **violet**. Transcription start sites are labeled with ↓. Putative –35 and –10 sequences are labeled and underlined. Sites of R49A and T116I substitutions to create negatively dominant and noninducible variants of AgaR, respectively, are bold and marked with an asterisk (*). Restriction sites utilized to construct pKR513 are also shown.

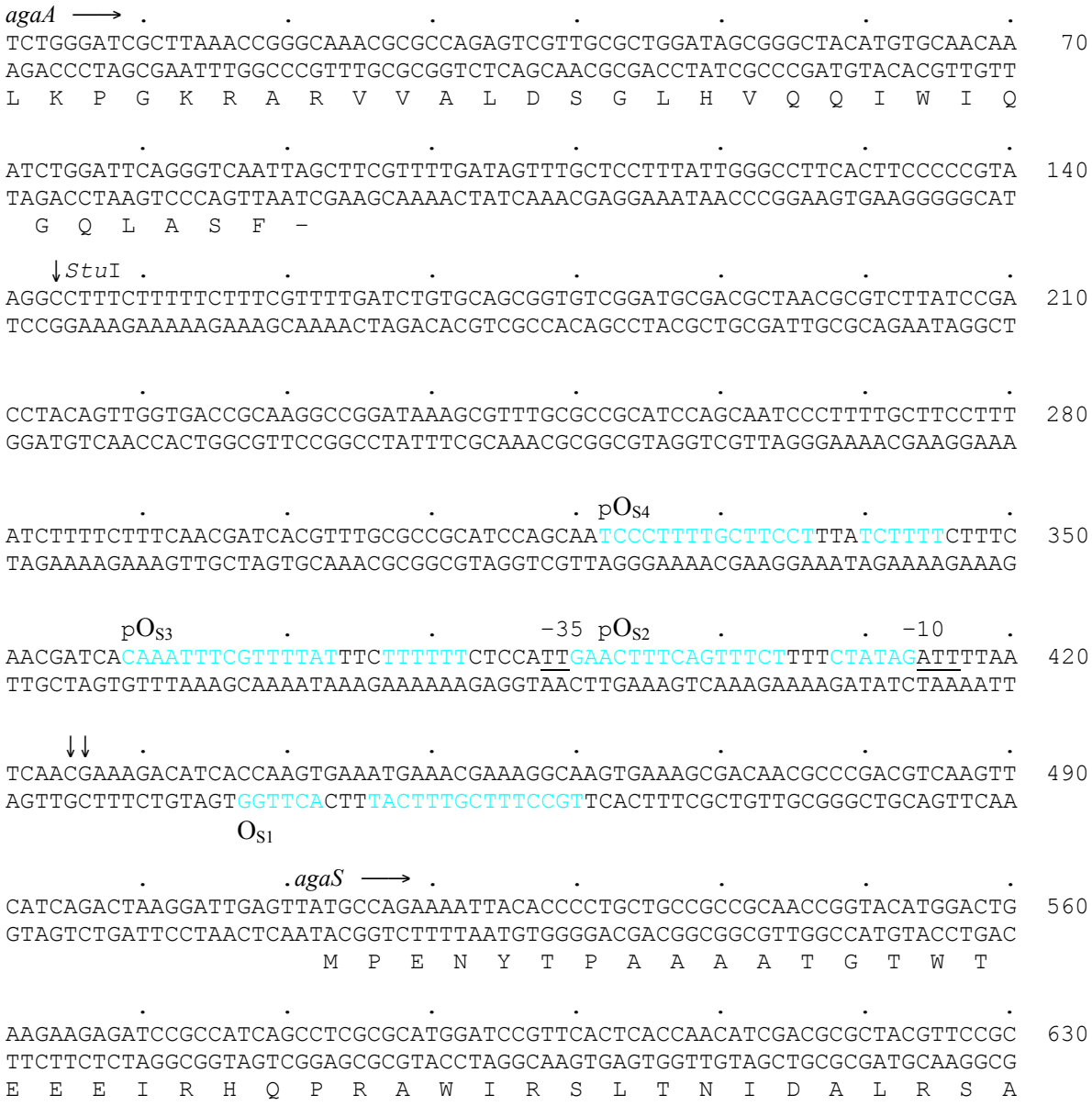


Figure 6. Nucleotide sequence of the *agaA–agaS* promoter region. Sequence is from accession numbers AE000394 and AE000395 of the Entrez database available at NCBI. Open reading frames are labeled. The footprinted region protected by AgaR is shown in **blue** (O_{S1}) as are the putative operator sites not verified experimentally but having sequence similarity to the proposed AgaR consensus binding sequence (pO_{S2}, pO_{S3} and pO_{S4}). The transcription start site is labeled with ↓. Putative –35 and –10 sequences are labeled and underlined. The *Stu*I site described in the text is indicated.

TABLE 3

Strains of *E. coli* used or constructed in the study of AgaR

Strain[a]	Genotype	Derivation or reference
BL21(DE3)	F$^-$ *hsdS gal (λcIts857 ind1 Sam7nin5 lacUV5–T7gene1)*	(52)
MC4100	*F$^-$ araD139 Δ(argF–lac)U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	(53)
DH5α	*(φ80dlacZΔM15) endA1 recA1 hsdR17 supE44 thi1 gyrA relA1 Δ(lacZYA–argF)U169*	(54)
DH5α Z1	DH5α *(λatt lacI^q tetR$^+$* Sp^r*)*	(49)
MG1655	wild–type isolate (F$^-$ λ$^-$)	(8)
TL504	MG1655 *Δ(lacZYA–argF)U169 zah–735*::Tn*10*	(55)
TL524	TL504 Tc^s	Tc^s selection[b]
KR161	TL524 Φ(*agaR'–lacZ*)	λKR161 lysogen of TL524
KR1161	KR161 *agaR*::Tc^r	[c]
KR162	TL524 Φ(*agaZ'–lacZ*)	λKR162 lysogen of TL524
KR1162	KR162 *agaR*::Tc^r	[c]
KR164	TL524 Φ(*agaS'–lacZ*)	λKR164 lysogen of TL524
KR1164	KR164 *agaR*::Tc^r	[c]
KR1524	TL524 *agaR*::Tc^r	[c]
KR201 to 205	TL524 Aga$^+$	Selection for growth on *N*–acetylgalactosamine[d]
KR1201	KR201 *agaR*::Tc^r	[c]
KR201–161	KR201 Φ(*agaR'–lacZ*)	λKR161 lysogen of KR201
KR201–162	KR201 Φ(*agaZ'–lacZ*)	λKR162 lysogen of KR201
KR201–164	KR201 Φ(*agaS'–lacZ*)	λKR164 lysogen of KR201

Strain[a]	Genotype	Derivation or reference
KR203–161	KR203 Φ(*agaR'–lacZ*)	λKR161 lysogen of KR203
KR203–162	KR203 Φ(*agaZ'–lacZ*)	λKR162 lysogen of KR203
KR203–164	KR203 Φ(*agaS'–lacZ*)	λKR164 lysogen of KR203
KR205–161	KR205 Φ(*agaR'–lacZ*)	λKR161 lysogen of KR205
KR205–162	KR205 Φ(*agaZ'–lacZ*)	λKR162 lysogen of KR205
KR205–164	KR205 Φ(*agaS'–lacZ*)	λKR164 lysogen of KR205
KR301 to 302	TL4100 (MC4100 *recD*::Tn*10*) Aga^+	Selection for growth on *N*–acetylgalactosamine[d]
KR391/KR392	KR301/KR302 *sfsB(nlpA)*3198::Tn*10kan* Aga^+	P1(CAG12127) x KR301/KR302 Km^r selection
KR303	TL4100 (MC4100 *recD*::Tn*10*) Gat^+ (T^r)	Selection for growth on galactitol at 37°C[d]
KR501, 503 and 505	KR201 *sfsB(nlpA)*3198::Tn*10kan* Aga^-	P1(CAG12127) x KR201 Km^r selection
KR502 and 504	KR201 *sfsB(nlpA)*3198::Tn*10kan* Aga^+	P1(CAG12127) x KR201 Km^r selection
LE392	(F^- λ^-) *lacY1 glnV44 galK2 galT22 tyrT58 metB1 hsdR514 trpR55*	(56)
CAG12072	(λ^-) *sfsB(nlpA)*203::Tn*10 rph*–1	(57,58)
CAG12127	(λ^-) *sfsB(nlpA)*3198::Tn*10kan rph*–1	(57,58)

[a] All strains are derivatives of *E. coli* K–12 except BL21(DE3) which is derived from *E. coli* B.

[b] The selection for tetracycline sensitive strains was as described by Maloy and Nunn (59). This strain was constructed by Ali Bhattacharya.

[c] The series of selections used for isolating *agaR*::Tc^r strains is described in the text.

[d] Selection for revertants capable of utilizing *N*–acetylgalactosamine or galactitol as a sole source of carbon is described in the text.

Plasmids and Phages. Plasmids used or constructed in this study are listed in Table 4. Construction of individual sets of plasmids is discussed below. Phage λDD704 (8,61) was obtained from the *E. coli* Genome Sequencing Project and contains the *aga* region from strain MG1655, the wild–type isolate used for the sequencing project (8). λKR161, λKR162 and λKR164 contain the *agaR'–lacZ*, *agaZ'–lacZ* and *agaS'–lacZ* transcriptional fusions derived from pKR161, pKR162 and pKR164, respectively. These phages are the result of the transfer of the transcriptional fusions from the plasmids to λRS45 by homologous recombination. λRS45 and the selection procedure has been previously described (31,63).

Growth media and conditions. Cultures were grown in either Luria–Bertani (LB) broth (62) or the minimal A and B salts described by Clark and Maaløe (64) with additions as described for specific experiments. All cultures were grown with aeration at 37°C unless indicated otherwise. Antibiotics were included at 100 μg/ml ampicillin, 50 μg/ml kanamycin, 10 μg/ml tetracycline, 25 μg/ml chloramphenicol or 20 μg/ml spectinomycin where appropriate. Cells harboring pMS421 (65) were grown in the presence of 50 μg/ml spectinomycin. Overexpression of AgaR or variants was induced by the addition of 0.5 mM IPTG for strains harboring derivatives of pT7–7 (66) or 100 ng/ml tetracycline for strains harboring derivatives of the pZ family of expression plasmids (49).

General molecular biological techniques. Cloning (e.g. PCR, agarose gel electrophoresis, ligation), preparation and transformation of competent cells and transduction were performed using standard procedures (62,68) except for transformation using pKO3 and derivatives (see below). Products from PCR reactions were routinely purified, either directly or from a slice of agarose following agarose gel electrophoresis, using the GFX™ PCR DNA and Gel Band Purification kit supplied by Amersham Pharmacia Biotech, Inc. or the QIAquick Gel Extraction kit from Qiagen. Lysate prepared as described by Silhavy et al. (62) from phage λDD704 grown on LE392 was routinely used as template for PCR reactions to amplify various regions of the *aga* gene cluster. Typically, 1 to 3 μl of a 1/20 dilution of the high–titer lysate was added to each PCR reaction. Plasmid DNA was isolated using kits supplied by either Qiagen or Promega. Chromosomal DNA was isolated using a previously published procedure (62).

Construction of promoter–*lacZ* transcriptional fusions. Regions of the *aga* gene cluster believed to contain transcriptional promoters were amplified by PCR using either 366895 and 386247 (P_R and P_Z) or WKR1 and WKR2 (P_S). These primer pairs introduced an *Mfe*I site at one end and an *Eco*RI site at other end of the amplified product. Both enzymes produce the same cohesive terminus. These fragments were then cloned into the *Eco*RI site of pSP417 (51). Restriction digestion of the resulting plasmids using various restriction enzymes with sites within both the vector and the insert was used to determine the orientation of the insert. The orientation of the insert was also confirmed by sequence analysis. The resulting plasmids, pKR161, pKR162 and pKR164, contain P_R–*lacZ*, P_Z–*lacZ* and P_S–*lacZ* transcriptional fusions, respectively.

TABLE 4

Plasmids used or constructed in the study of AgaR

Plasmid	Insert Vector	Description [a]
pCP16	(50)[b]	Ap^r, Tc^r
pZE21–*luc*	(49)	P_{N25}–*luc* ColE1, Km^r
pZS24–*luc*	(49)	$P_{lac/ara}$–*luc* pSC101, Km^r
pSP417	(51)	Promoterless *lacZ* ColE1, Ap^r
pT7–7	(53)	T7 promoter, MCS, ColE1, Ap^r
pAB126	(Ali Bhattacharya, personal communication)	*plsX–His*$_6$ ColE1, Ap^r
pKO3	(67)	T^s ori, Cm^r
pMS421	(65)	*lacI*^{q}, pSC101, Sp^r
pKR211	329890/336849 (PCR)[c], *Eco*RI/*Xba*I pZE21–*luc*, *Xba*I/*Eco*RI	P_{N25}–*agaR* ColE1, Km^r
pKR311	pKR211, *Xho*I/*Avr*II pZS24–*luc*, *Avr*II/*Xho*I	P_{N25}–*agaR* pSC101, Km^r
pKR215	329890/336849 (PCR)[d], *Eco*RI/*Xba*I pZE21–*luc*, *Xba*I/*Eco*RI	P_{N25}–*agaR* (T116I) ColE1, Km^r
pKR315	pKR215, *Xho*I/*Avr*II pZS24–*luc*, *Avr*II/*Xho*I	P_{N25}–*agaR* (T116I) pSC101, Km^r
pKR216	329890/336849 (PCR)[d], *Eco*RI/*Xba*I pZE21–*luc*, *Xba*I/*Eco*RI	P_{N25}–*agaR* (R49A) ColE1, Km^r
pKR316	pKR216, *Xho*I/*Avr*II pZS24–*luc*, *Avr*II/*Xho*I	P_{N25}–*agaR* (R49A) pSC101, Km^r
pKR310 (pZS21–*luc*)	pZE21–*luc*, *Xho*I/*Avr*II pZS24–*luc*, *Avr*II/*Xho*I	P_{N25}–*luc* pSC101, Km^r
pKR161	366895/386247 (PCR), *Mfe*I/*Eco*RI pSP417, *Eco*RI	P_R–*lacZ* ColE1, Ap^r
pKR162	386247/366895 (PCR), *Eco*RI/*Mfe*I pSP417, *Eco*RI	P_Z–*lacZ* ColE1, Ap^r
pKR163	WKR2/WKR1 (PCR), *Eco*RI/*Mfe*I pSP417, *Eco*RI	P_S (rev)–*lacZ* ColE1, Ap^r

Plasmid	Insert Vector	Description [a]
pKR164	WKR1/WKR2 (PCR), *Mfe*I/*Eco*RI	P_S–*lacZ*
	pSP417, *Eco*RI	ColE1, Ap^r
pKR186	372696/336849 (PCR), *Nde*I	P_{T7}–*agaR*
	pT7–7, *Nde*I/*Sma*I	ColE1, Ap^r
pKR411	372696/258920 (PCR), *Nde*I/*Sal*I	P_{T7}–*agaR*His_6
	pAB126, *Sal*I/*Nde*I	ColE1, Ap^r
pKR415	372696/258920 (PCR)[d], *Nde*I/*Sal*I	P_{T7}–*agaR* (T116I)His_6
	pAB126, *Sal*I/*Nde*I	ColE1, Ap^r
pKR416	372696/258920 (PCR)[d], *Nde*I/*Sal*I	P_{T7}–*agaR* (R49A)His_6
	pAB126, *Sal*I/*Nde*I	ColE1, Ap^r
pKR511	WKR5/386252 (PCR), *Eco*RI	*agaZ′*–*agaR′*
	pT7–7, *Sma*I/*Eco*RI	ColE1, Ap^r
pKR512	pCP16, *Hin*dIII(filled–in)/*Sma*I	*agaR*::Tc^r, ColE1
	pKR511, *Hpa*I	Tc^r, Ap^r
pKR513	pKR512, *Msc*I/*Bam*HI	*agaR*::Tc^r, T^s ori
	pKO3, *Bam*HI/*Sma*I	Tc^r, Cm^r
pKR611	329890/336849 (PCR)[e], *Eco*RI/*Xba*I	P_{N25}–*agaR*
	pZE21–*luc*, *Xba*I/*Eco*RI	ColE1, Km^r
pKR613	329890/336849 (PCR)[e], *Eco*RI/*Xba*I	P_{N25}–*agaR*
	pZE21–*luc*, *Xba*I/*Eco*RI	ColE1, Km^r

[a] Where relevant, the expression promoter, origin of replication and antibiotic resistance(s) are indicated.

[b] For previously constructed plasmids, instead of giving insert and vector fragments, references are shown.

[c] Unless indicated otherwise, the template used for PCR was λDD704 high–titer phage lysate.

[d] See text for derivation of template used for PCR.

[e] Chromosomal DNA isolated from either strain KR201 (pKR611) or KR203 (pKR613) was used as template for PCR reactions. These strains are Aga^+ derivatives of TL524.

Construction of plasmids expressing AgaR or variants. pAB126 (a derivative of pT7–7) contains the *plsX* gene (unrelated to this project) followed by an additional coding region to express a variant of the PlsX protein tagged with a hexahistidine tail at the carboxy terminus. pAB126 is a derivative of pGZ117 which expresses GlpR with a polyhistidine tag at the carboxy terminus. Construction of pGZ117 has been described elsewhere (48). The *agaR* gene was PCR amplified using λDD704 phage as template and primers 372696 and 258920. The *agaR* gene was then cloned into the *Nde*I and *Sal*I sites of pAB126, replacing *plsX* with *agaR*, giving pKR411. This plasmid was used to overexpress AgaR tagged with a hexahistidine tail at the carboxy terminus for purification by immobilized metal affinity chromatography (IMAC).

To overexpress AgaR without a hexahistidine tag, pKR211 was constructed. The *agaR* gene was amplified by PCR using primers 329890 and 336849 which introduced an *Eco*RI site at the 5' end and an *Xba*I site at the 3' end of the gene. The amplified fragment was cloned into the *Eco*RI and *Xba*I sites of pZE21–*luc* replacing the luciferase gene (49). This plasmid contains the high–copy number origin of replication ColE1 also found in pSP417 (51).

To overexpress AgaR from a plasmid compatible with the pSP417 derivatives containing the transcriptional fusions and at a more physiological level, pKR311 was constructed. This plasmid contains the low–copy number pSC101 origin of replication and kanamycin–resistance cassette from pZS24–*luc* but contains the *agaR* gene controlled by the tetracycline–inducible promoter P_{N25} from pKR211 (*Xho*I to *Avr*II). To serve as a control for this plasmid, pKR310 was also constructed in which the luciferase gene controlled by the P_{N25} promoter from pZE21–*luc* was transferred using the same restriction sites to pZS24–*luc* (49). Using the standard nomenclature for the pZ family of expression plasmids, this plasmid would be designated pZS21–*luc* (49).

Pairs of overlapping mutagenic oligonucleotides (WKR3/WKR4 for R49A and 372597/331644 for T116I) were used with the oligonucleotide pairs described above to PCR amplify the *agaR* gene as two separate fragments. The separate fragments were combined, heat–denatured, annealed and then extended using T7 DNA polymerase (supplied with the Sequenase Version 2.0 DNA sequencing kit from USB). These reactions were then used as templates for a second round of PCR reactions. The resulting products were then cloned into either pAB126 or pZE21–*luc* as for the wild–type *agaR* described above. This method of site–directed mutagenesis had been desribed previously (69). Plasmids were constructed that encoded variants of AgaR containing either an R49A or T116I substitution, expressed either with (derivatives of pAB126) or without (derivatives of pZE21–*luc*) a polyhistidine tag. The mutations were confirmed by sequence analysis. Based on previous work with GlpR, these variants were expected to be either negative–dominant (R49A) or noninducible (T116I) (see Figures 2 and 5) (6,24).

Construction of *agaR*::Tc^r strains. In order to construct *agaR* disruption strains, pKR513 was constructed (Figure 7). This plasmid contains a portion of the *agaZ* gene and most of the *agaR* gene. In this plasmid, *agaR* has been disrupted at the *Hpa*I site by a DNA fragment conferring tetracycline resistance. The plasmid is a derivative of pKO3, specifically designed for engineering gene knockouts (67). pKR513 therefore contains a

temperature–sensitive origin of replication. Selection for growth at 43°C of cells carrying pKR513 on LB agar plates supplemented with tetracycline and chloramphenicol results in isolates in which the plasmid has been incorporated into the chromosome by homologous recombination. The plasmid pKO3 also carries the *sacB* gene, encoding levansucrase, which is detrimental to *E. coli* in the presence of sucrose (67). Therefore, a second selection for growth in the presence of sucrose and tetracycline was used to isolate cells in which the plasmid DNA was lost by a second recombination event. Loss of plasmid genes was verified by checking for chloramphenicol–sensitivity. P1 transductional mapping and PCR analysis also confirmed the location of the tetracycline–resistance cassette at the *agaR* locus (see Results).

Transformation of cells with derivatives of pKO3, including pKR513 described above, utilized a modified procedure since the plasmids contain a temperature–sensitive origin of replication (67). Heat shock was performed at 37°C instead of 42°C and cells were incubated following heat shock at 30°C instead of 37°C.

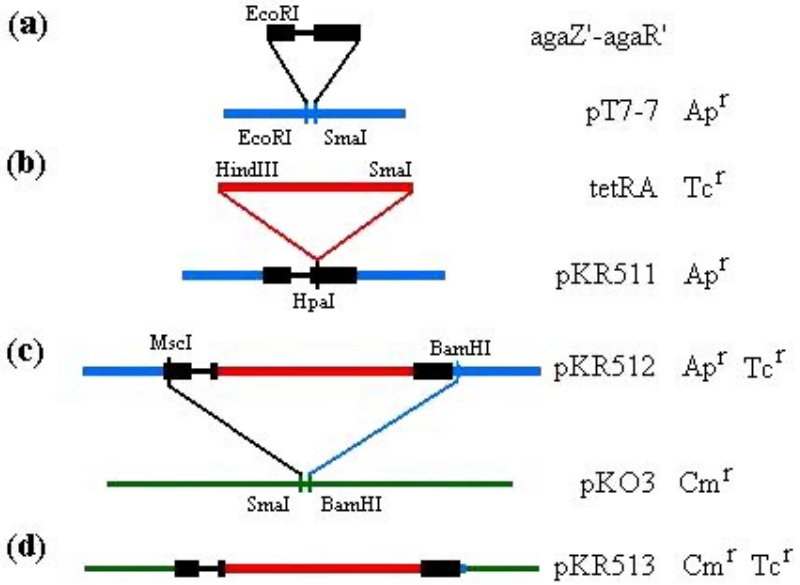


Figure 7. Construction of pKR513. (a) A portion of the *agaZ* and *agaR* reading frames (black) was amplified by PCR, digested with *Eco*RI and cloned into the *Eco*RI and *Sma*I sites of pT7–7 (blue) (53) yielding pKR511. (b) The tetracycline–resistance cassette from pCP16 (red) (50) was isolated using *Hin*dIII and *Sma*I. The fragment was made blunt–ended by treatment with T7 DNA polymerase and dNTPs and cloned into the *Hpa*I site within the *agaR* reading frame yielding pKR512. (c) The disrupted *agaR* gene was removed from pKR512 using *Msc*I (located within the *agaZ* reading frame) and *Bam*HI (located in the MCS of pT7–7). This fragment was cloned into the *Bam*HI and *Sma*I sites of pKO3 (green) (67). (d) The resulting plasmid, pKR513, was used to create *agaR*::Tc^r strains using the series of selections described in the text, as pioneered by Church et al. (67).

Selection for Aga$^+$ revertants. A culture (750 μl) of TL524 grown in LB to an A_{600} of 3.9 was pelleted and resuspended in 150 μl LB, and then plated on agar containing minimal media (64) supplemented with 10 mM *N*–acetylgalactosamine and 2 μg/ml thiamine. After incubation at 37°C for several days, approximately 20 Aga$^+$ revertants were obtained. A similar method was used for selection of revertants of other strains capable of utilizing either *N*–acetylgalactosamine, galactosamine or galactitol at either 37°C or 30°C.

β–Galactosidase assays. β–galactosidase activity was determined using the method of Miller (70) using logarithmically–growing cultures. Specific activities are expressed in Miller units (70) and are the mean of assays performed in duplicate utilizing two independently–isolated colonies unless noted otherwise.

DNA sequencing. Sequencing of DNA was performed using the dideoxynucleotide chain termination method described by Sanger (71). The Sequenase Version 2.0 DNA sequencing kit supplied by USB, the SequiTherm EXCEL™ DNA sequencing kit supplied by Epicentre Technologies and the cycle sequencing kit supplied by Pharmacia Biotech were all used as described by their respective manufacturers. Sequencing reactions, as well as footprinting reactions and S1 nuclease protection assays, were electrophoresed on 6% polyacrylamide gels containing 7 M urea using Tris–Borate–EDTA (TBE) buffer. The M13 24–mer forward primer (universal –40 primer) and the plasmid pSAD2 (a pUC derivative) (72,73) supplied with the Epicentre Technologies sequencing kit were used for sequencing reactions to serve as size markers for both footprinting and S1 nuclease mapping experiments described below.

Overproduction and purification of AgaR–His$_6$. A 50 ml LB overnight culture of BL21(DE3) pKR411 pMS421 containing 100 μg/ml ampicillin and 50 μg/ml spectinomycin was added to 1 l fresh LB containing the same concentrations of ampicillin and spectinomycin. The culture was grown with constant shaking at 37°C for two hours and then expression of AgaR–His$_6$ was induced by the addition of 0.5 mM IPTG. After three more hours of growth, the culture was harvested and washed with 1/2 volume of wash buffer (100 mM sodium phosphate buffer (pH 7.4), 500 mM NaCl and 1 mM EDTA). The cells were pelleted and stored at –70°C.

The cell pellet was thawed on ice and resuspended in 25 ml resuspension buffer (100 mM sodium phosphate buffer (pH 7.4), 500 mM NaCl, 0.1 mM EDTA, 1 mM PMSF, 1 mM β–mercaptoethanol, 10 mM imidazole and 15% glycerol). The cells were lysed by two passes through a French pressure cell at 16,000 to 20,000 psi. Cell debris was removed by centrifugation for 5 minutes at 3000 g. The supernatant was transferred to a new precooled tube and further clarified by centrifugation for 15 minutes at 8000 g. A final centrifugation of the supernatant for 90 minutes at 100,000 g (40,000 rpm in a type 65 rotor) in a Beckman model L5–65 ultracentrifuge was used to remove membrane material. The supernatant was stored at –70°C. All centrifugation steps were carried out at 4°C.

The soluble fraction was thawed and loaded onto a 2 ml IMAC column charged with $NiCl_2$ as described by the manufacturer. The column material used was Chelating Sepharose

Fast Flow provided by Pharmacia Biotech. The column had been preequilibrated at room temperature using buffer A (100 mM sodium phosphate buffer (pH 7.4) containing 500 mM NaCl and 10 mM imidazole). Flow was maintained at 0.5 ml/min. The column was washed with 60 ml buffer A followed by 25 ml buffer A containing 50 mM imidazole. The column was developed using a 60 ml linear gradient from 50 to 500 mM imidazole in buffer A. Protein concentration was monitored using the absorbance at 280 nm. AgaR–His_6 eluted at approximately 150 to 250 mM imidazole. These fractions were pooled and protein was concentrated by the addition of 0.45 g ammonium sulfate per ml. The solution was incubated on ice 30 minutes and protein was pelleted by centrifugation for 30 minutes at 10,000g. Protein was resuspended in 1.5 ml of a solution containing 20 mM Tris–HCl (pH 7.5), 1 mM EDTA, 0.1 mM dithiothreitol, 500 mM KCl, 15% glycerol. This resuspension was stored in aliquots at –70°C.

Determination of protein concentrations. Protein concentrations were determined by the method of Bradford (74) with bovine serum albumin as the standard. Coomassie protein assay reagent and the bovine serum albumin standard were purchased from Pierce and the manufacturer's microassay method was used.

Estimation of subunit and native molecular mass of AgaR–His_6. Subunit molecular mass of purified AgaR with the hexahistidine tag was determined by comparison to markers on a sodium dodecyl sulfate–15% polyacrylamide gel. SDS–PAGE was performed as described by Laemmli (75). Protein bands were visualized using Fast Stain from Zoion as described by the manufacturer.

Molecular exclusion chromatography was used to determine the native molecular mass of AgaR. Samples of 0.1 ml containing from 40 to 80 μg protein were applied to a Waters glass Protein–Pak™ 300SW (8 x 300 mm) column equilibrated at ambient temperature with 20 mM Tris–HCl (pH 7.2), 1 mM EDTA, 100 mM NaCl at a flow rate of 0.8 ml/min. Void volume and total volume were determined using blue dextran (average molecular weight of 2,000,000) and vitamin B_{12}, respectively. Proteins used for generation of a standard curve were bovine serum albumin, ovalbumin, soybean trypsin inhibitor, myoglobin and ribonuclease.

Gel mobility shift assays. Gel mobility shift assays of AgaR were performed essentially as described previously for GlpR except electrophoresis of the reactions was performed at room temperature (76). Radioactive DNA fragments used were generated by PCR reactions containing 200 μM of each dNTP except dATP. Nonradioactive dATP was included at 20 μM and [α–^{35}S]dATP was included at 1 μM. The resulting products were purified using either agarose gel electrophoresis or polyacrylamide gel electrophoresis. Gel mobility shift assays typically contained from 1,000 to 10,000 cpm of the radioactive fragment. Radioactive bands were visualized using autoradiography and/or a Packard Instant Imager.

DNase I footprinting. Dnase I footprinting of AgaR and cAMP–CRP complex was performed essentially as described previously (31,33). DNA fragments were radiolabeled by treating one of two PCR primers specific for the fragment with T4 polynucleotide kinase and [γ–^{33}P]ATP. The labeled oligonucleotide was purified from unincorporated [γ–^{33}P]ATP using TE Micro Select–D G–25 spin columns purchased from 5 Prime→3 Prime, Inc. PCR using the labeled primer and a nonlabeled primer was then used to amplify the fragment of DNA. The fragment was then purified using agarose gel electrophoresis. The promoter region specific for *agaS* was digested with *Stu*I prior to purification. Each footprint reaction contained approximately 500 cpm per base of radioactively–labeled DNA fragment. After incubation for 30 minutes at 30°C with varying concentrations of purified AgaR–His_6, the reactions were partially digested using 1.7 ng DNase I incubated at 37°C for 4 minutes. DNA was precipitated using ethanol, resuspended in PAGE loading buffer, and then electrophoresed as described above for sequencing reactions. Radioactive bands were visualized using autoradiography after drying the polyacrylamide gel.

Isolation of RNA and S1 nuclease mapping. Total RNA was isolated using the RNAqueous™ kit supplied by Ambion as described by the manufacturer. RNA was purified from 5 ml cultures of DH5αZ1 harboring either pKR161, pKR162 or pKR164 grown overnight in minimal (A and B salts) media supplemented with 100 μg/ml ampicillin, 2 μg/ml thiamine, 0.2% casamino acids and 10 mM *N*–acetylgalactosamine. RNA concentrations were quantitated by measuring the absorbance at 260 nm.

Labeled single–stranded DNA probe was prepared by asymmetric PCR. First, the promoter regions cloned into pSP417 were amplified by PCR using primer Bing7, complementary to the multiple cloning site of pSP417, and either 366895, 386247 or WKR1. These double–stranded products were purified on an agarose gel and subsequently used as templates for the asymmetric PCR reactions. The reactions contained 200 μM of each dNTP except dATP. Nonradioactive dATP was included at 20 μM and [α–^{35}S]dATP was included at 1 μM. The resulting products were purified using agarose gel electrophoresis.

S1 nuclease protection assays were performed using the kit supplied by Ambion essentially as described by the manufacturer except 5 μg sonicated salmon sperm DNA (5 μl of 1 μg/μl) was added just prior to the final precipitation step as carrier. Three reactions were performed for each of three probes. One reaction contained 30 μg total RNA from the strain carrying the plasmid corresponding to the probe used. The other two reactions contained 30 μg yeast RNA, only one of which was treated with S1 nuclease. Each reaction contained approximately 800,000 cpm of the corresponding probe. The reactions were analyzed using polyacrylamide gel electrophoresis as described for sequencing and the results were visualized using autoradiography after drying the polyacrylamide gel.

Preliminary sequence data. Preliminary sequence data for *Yersinia pestis* was obtained from The Institute for Genomic Research website at http://www.tigr.org.

Results

Identification of transcriptional promoters in the *aga* gene cluster. Based on the organization of the *aga* gene cluster (Figure 3(a)), two regions were considered likely to contain transcriptional promoters. One was the *agaR–agaZ* (approximately 275 bp) intergenic region since the two genes are divergently transcribed. The other was the *agaA–agaS* intergenic region based on its relative size (over 400 bp). Both of these regions were cloned into the multiple cloning site of the promoter–probe vector, pSP417 (51), which contains a promoterless *lacZ* gene. The *agaR–agaZ* intergenic region was cloned into pSP417 in both orientations since the region would be expected to have both a transcriptional promoter specific for *agaR* (P_R) and one for *agaZ* (P_Z). When DH5αZ1 cells containing the plasmids with the transcriptional fusions were assayed for β–galactosidase activity (70), P_R–*lacZ* (pKR161), P_Z–*lacZ* (pKR162) and P_S–*lacZ* (pKR164), yielded specific activities in Miller units of 7470, 1520 and 35000, respectively, confirming the presence of transcriptional promoters in these regions. For comparison, the promoter–probe vector without an insertion in the multiple cloning site (pSP417) or a derivative containing the *agaA–agaS* intergenic region in the opposite orientation (pKR163) yielded only 50 Miller units.

Expression of the *aga* genes is induced by *N*–acetylgalactosamine. In order to study the transcriptional regulation of the *aga* gene cluster using single–copy gene fusions, the plasmid–borne fusions described above were transferred to phage λRS45 by homologous recombination (63). Subsequently, λ lysogens in which the λ genome had inserted itself into the *E. coli* chromosome were selected. Since the λ genome can insert itself more than once into the chromosome, single copy lysogens were identified by comparing the β–galactosidase specific activities of several independently–isolated lysogens (63). This procedure was used to obtain strains containing a single copy of either P_R–*lacZ* (KR161), P_Z–*lacZ* (KR162) or P_S–*lacZ* (KR164) on the chromosome. The native *lacZ* gene had been deleted from the parent strain (TL524) used to create the lysogens so β–galactosidase activity measured in these strains resulted exclusively from the transcriptional fusion. These strains were used for several experiments designed to determine details of the regulation of the transcriptional promoters.

Table 5 lists the specific activities of β–galactosidase in Miller units measured for these strains grown under various conditions. Addition of either *N*–acetylgalactosamine or galactosamine to the growth medium resulted in elevated levels of β–galactosidase activity. The increases in activity indicate the promoters within the *aga* gene cluster are induced by the presence of *N*–acetylgalactosamine or galactosamine. Induction ranged from approximately 4 to 7 fold for promoters specific for *agaR* and *agaZ* to 20 to 60 fold for the promoter specific for *agaS*. Addition of *N*–acetylglucosamine did not result in induction (data not shown).

The presence of galactitol in the growth medium did not result in induction of the P_S–*lacZ* transcriptional fusion in KR164. The addition of galactitol to cultures of KR303, a derivative of MC4100 selected for the ability to utilize galactitol at 37°C, lysogenized

TABLE 5

Induction of *aga* promoters by *N*–acetylgalactosamine and galactosamine

Addition to growth media[a]	Transcriptional fusion		
	P_R–*lacZ* (KR161)	P_Z–*lacZ* (KR162)	P_S–*lacZ* (KR164)
No addition	121 ± 3	107 ± 8	80 ± 5
N–Acetylgalactosamine	600 ± 65	435 ± 67	1,600 ± 130
Galactosamine	700 ± 40	760 ± 53	5,100 ± 710

[a] Cells growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine and 0.2% casamino acids. *N*–acetylgalactosamine or galactosamine (10 mM) was added where indicated.

with λKR162 (P_Z–*lacZ*) resulted in no increase in β–galactosidase activity compared to cultures containing no galactitol. The metabolic pathway of galactitol converges with that of *N*–acetylgalactosamine at tagatose–6–phosphate (12,15). These results suggest that the inducer molecule for AgaR may be an intermediate in the metabolic pathway for *N*–acetylgalactosamine prior to the convergence with the metabolic pathway for galactitol (See Figure 4 and below).

Construction of *agaR*::Tc^r strains. In order to verify that AgaR was responsible for the regulation of the genes within the *aga* cluster, a plasmid was constructed (pKR513) in order to create strains in which the *agaR* gene had been disrupted. The *agaR* gene disrupted by a DNA fragment containing genes which confer tetracycline resistance was cloned into plasmid pKO3 designed for creating gene disruptions and deletions (67). (See Experimental Procedures and Figure 7 for details concerning the construction of this plasmid.) The parent plasmid pKO3 confers chloramphenicol resistance and possesses a temperature–sensitive origin of replication (67). Selection at 43°C in the presence of tetracycline and chloramphenicol allows only for the growth of cells in which the plasmid (pKR513) has incorporated itself by homologous recombination into the chromosome at the *agaR* gene. Subsequently, selection for growth on medium containing sucrose and tetracycline results in strains in which the plasmid DNA excises itself. This occurs since the plasmid carries the *sacB* gene, the product of which, levansucrase, is detrimental to growth of *E. coli* in the presence of sucrose (67). The loss of plasmid DNA was confirmed by checking that isolates had become chloramphenicol–sensitive.

Plasmid pKR513 was used to create an *agaR*::Tc^r derivative of TL524. To confirm the gene disruption, chromosomal DNA was isolated from TL524 and the *agaR*::Tc^r

derivative (TL1524). Two sets of PCR reactions were performed using these chromosomal DNAs as templates and oligonucleotide WKR6. This oligonucleotide corresponds to a sequence near the 3' terminus of the *tetA* gene contained within the tetracycline–resistance cassette used to disrupt *agaR*. One set of reactions utilized oligo 329890. Use of this primer with WKR6 should result in a 530 bp product if the disruption is present on the chromosome (Figure 8, lanes 4 and 5). The other set of reactions utilized oligo *agaV*–C. This primer corresponds to the 3' terminus of the *agaV* gene which is not present in pKR513. Primers *agaV*–C and WKR6 were thus used in conjunction to confirm the presence of the tetracycline–resistance cassette at the correct chromosomal location. The expected product (2.46 kb) was observed only when chromosomal DNA isolated from TL1524 was used as template (Figure 8, lanes 2 and 3).

The tetracycline–resistance cassette was also mapped by P1 transduction to the *aga* locus (approximately minute 70.7). Selection for kanamycin–resistant transductants utilizing a P1 lysate grown on CAG12127 (containing a Tn*10kan* at minute 71.8) resulted in isolates that became tetracycline sensitive with a frequency of over 50%.

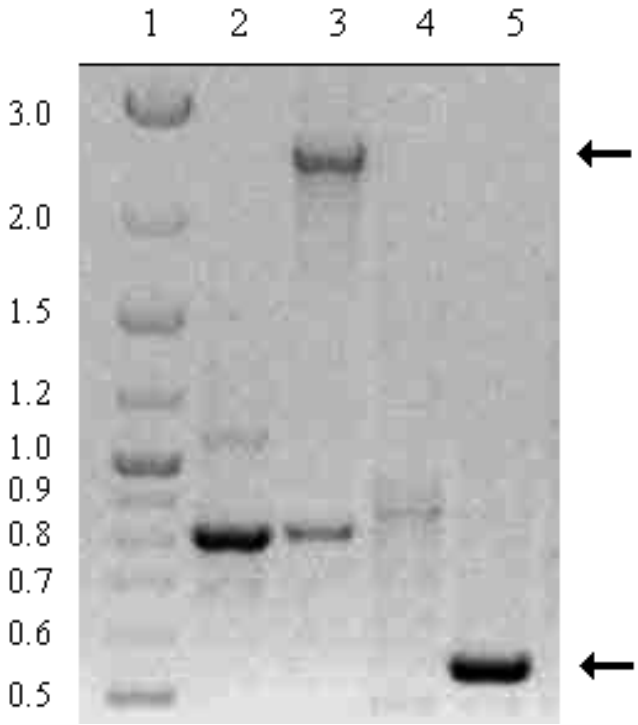


Figure 8. Verification of *agaR*::Tc^r strain. Agarose gel electrophoresis and ethidium bromide staining of PCR reactions. (1) Markers of the indicated sizes, in kb. (2 and 3) PCR reactions utilizing the primers WKR6 and *agaV*–C and chromosomal DNA from either TL524 (2) or TL1524 (3). (4 and 5) PCR reactions utilizing the primers WKR6 and 329890 and chromosomal DNA from either TL524 (4) or TL1524 (5). Assuming the presence and correct chromosomal location of the Tc^r cassette, the products for the two sets of reactions should be 2.46 and 0.53 kb, respectively. Arrows indicate positions of expected products. (Colors have been reversed for easier visualization.)

Expression of the *aga* genes is negatively regulated by AgaR. Plasmid pKR513 was also used to create derivatives of the P_R–*lacZ*, P_Z–*lacZ* and P_S–*lacZ* lysogens described above (KR161, KR162 and KR164, respectively) in which the *agaR* gene had been disrupted (KR1161, KR1162 and KR1164, respectively). Table 6 lists the specific activities of β–galactosidase for the parent strains and the derivatives in which the *agaR* gene has been disrupted. These strains were grown in a medium containing glucose so the promoters would be uninduced. Comparison of the activities of each promoter in the presence or absence of the repressor, AgaR, indicates that AgaR negatively regulates the transcription from each of these promoters. The promoters have a constitutively elevated level of expression in strains containing the disrupted *agaR* gene.

For P_R–*lacZ* and P_S–*lacZ*, the level of expression in induced cells, i.e. grown in the presence of galactosamine, is similar to the expression observed in the absence of functional repressor. However, expression from P_Z–*lacZ* in induced cells is almost 10 times greater than expression from this transcriptional fusion in the absence of repressor. This indicated that some other form of regulation in addition to repression by AgaR functions to regulate expression from this promoter.

TABLE 6

Repression of *aga* promoters by AgaR and activation of P_Z by cAMP–CRP

Genotype / Growth conditions[a]	Transcriptional fusion		
	P_R–*lacZ* (KR161 / KR1161)	P_Z–*lacZ* (KR162 / KR1162)	P_S–*lacZ* (KR164 / KR1164)
agaR$^+$ / Glucose	84 ± 6	8 ± 2	17 ± 1
agaR$^+$ / Glucose + cAMP	ND[b]	55 ± 5	ND
ΔagaR / Glucose	900 ± 70	83 ± 3	8,800 ± 210
ΔagaR / Glucose + cAMP	ND	1,100 ± 80	ND
agaR$^+$ / Galactosamine[c]	700 ± 40	760 ± 53	5,100 ± 710

[a] Cells growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine, 0.2% casamino acids and 0.2% glucose. cAMP (0.5 mM) was added where indicated.

[b] Not determined.

[c] Values from Table 5 illustrating promoter activities when induced by the presence of galactosamine.

P_Z is positively regulated by cAMP–CRP. A sequence similar to that recognized by the cAMP–cAMP receptor protein (CRP) complex was identified in the *agaR–agaZ* intergenic region (47,77). cAMP is elevated in cells grown in the absence of glucose. CRP, when complexed with cAMP activates many transcriptional promoters that express metabolic genes in *E. coli* (47,78). Therefore, the possibility that P_Z was regulated by this global regulatory protein was investigated. Strains containing the P_Z–*lacZ* fusion with or without a functional *agaR* gene were grown in the presence of exogenous cAMP with glucose as the carbon source. As shown in Table 6, addition of cAMP resulted in about a 10–fold increase in β−galactosidase specific activity in both the strain with a functional *agaR* gene and the strain in which the *agaR* gene had been disrupted. The P_Z–*lacZ* fusion strain with the disrupted *agaR* gene grown in the presence of cAMP resulted in a level of β−galactosidase specific activity comparable to that observed in the presence of galactosamine. This indicated that the promoter P_Z requires induction of AgaR and activation by cAMP–CRP for full activity.

Expression and purification of AgaR–His_6. Purified AgaR was obtained by constructing a plasmid that overexpressed the protein tagged with a polyhistidine tail at the carboxy terminus. Introduction of the tag changes the carboxy terminus of the protein from VILVGE to VILVVDLVPRGSHHHHHH. The possible effects this may have on the protein are discussed below. This tag had been constructed such that it could be removed using thrombin (48). Digestion of AgaR–His_6 with thrombin, however, revealed that AgaR possesses an internal thrombin site. A one–step purification scheme utilizing immobilized metal affinity chromatography (IMAC) was used to obtain nearly homogeneous protein (Figure 9(a)). The size of the purified protein as visualized by SDS–PAGE (36 kDa) was slightly larger than that predicted based on the deduced amino acid sequence of the protein (31 kDa). This phenomenon was previously observed for GlpR and also for AgaR without the polyhistidine tag. This anomalous mobility during SDS–PAGE may be attributed to the relatively acidic isoelectric point of these proteins (5.79 for GlpR, 5.16 for AgaR, 5.92 for AgaR–His_6) as determined for other acidic proteins (79).

As described previously, GlpR, DeoR and possibly all members of this family of proteins, oligomerize. Oligomerization appears to be required for efficient repression. The carboxy terminal region of this family appears to be involved in oligomerization. To ensure that the polyhistidine tag introduced at the carboxy terminus does not interfere with the ability of AgaR to oligomerize and to determine the native size of AgaR, gel filtration chromatography was used (Figure 9(b)). The native size of AgaR as determined by comparison to a standard curve generated as described in Experimental Procedures was 126 kDa. This indicates that the polyhistidine tag does not interfere with the ability of the protein to oligomerize and that under native conditions AgaR is a tetramer.

AgaR recognizes specific sequences within the promoter regions. In order to determine the specific sequences recognized by AgaR within the promoter regions, DNase I footprint analysis of these regions was performed (32,33,80,81). Varying amounts of

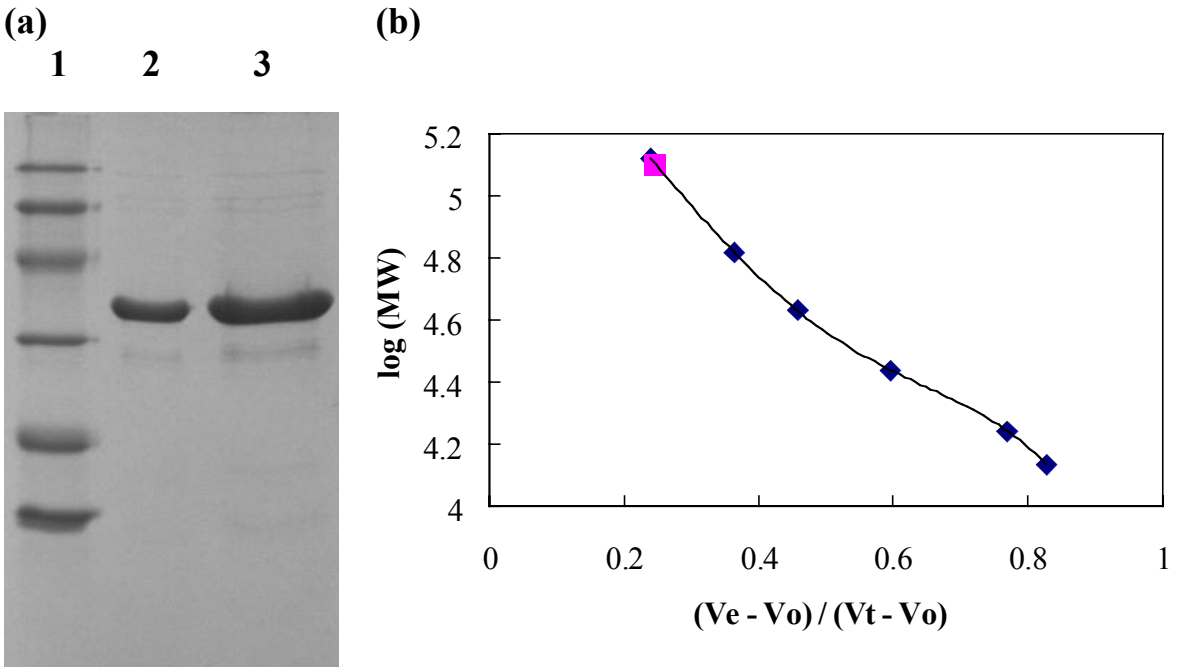
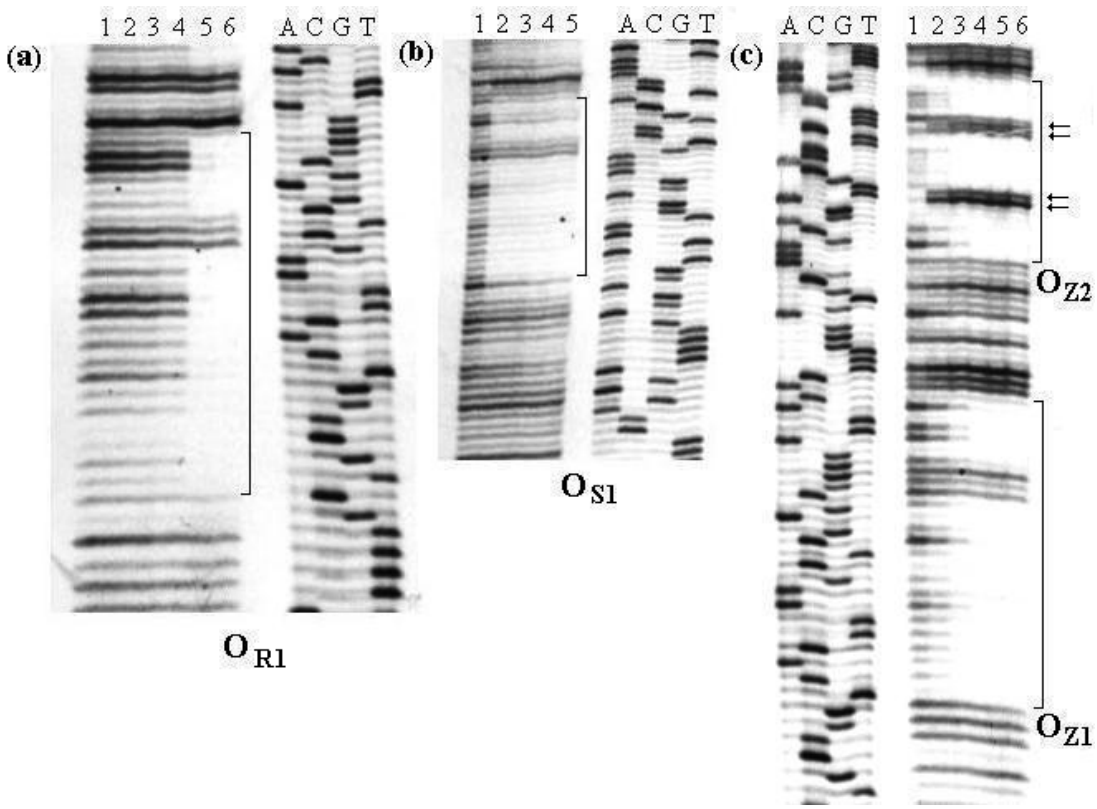


Figure 9. Estimation of the subunit and native molecular masses of purified AgaR–His$_6$. (a) SDS–PAGE illustrating purified AgaR–His$_6$. Lane 1 contains molecular mass markers corresponding to 97.4 kDa (phosphorylase b), 66.2 kDa (bovine serum albumin), 45 kDa (ovalbumin), 31 kDa (carbonic anhydrase), 21.5 kDa (trypsin inhibitor) and 14.4 kDa (lysozyme). Lanes 2 and 3 contain 2 and 5 μg, respectively, of purified AgaR–His$_6$. (b) Gel filtration chromatography was used to estimate the native molecular mass of AgaR–His$_6$ (■). Proteins used for generation of the standard curve in order of decreasing molecular mass are bovine serum albumin (132 and 66 kDa), ovalbumin (43 kDa), soybean trypsin inhibitor (27.5 kDa), myoglobin (17.6 kDa) and ribonuclease (13.7 kDa) (◆). V_e, V_o, and V_t correspond to the elution volume, void volume and total volume, respectively.

purified AgaR–His$_6$ were incubated with a promoter region radiolabeled at one end. The DNA was then subjected to limited digestion using DNase I. Digested DNA was electrophoresed on a standard DNA sequencing gel with sequencing reactions to serve as size markers.

Figure 10(a, b and c) illustrates the results of the footprinting experiments using purified AgaR–His$_6$. Three operator sites were identified in the *agaR–agaZ* intergenic region and one was identified in the *agaA–agaS* intergenic region. Footprint analysis using cell extract from cultures overexpressing AgaR without a polyhistidine tag (from plasmid pKR211) was shown to specifically protect the same sequences indicating that the footprints were not artifacts caused by the polyhistidine tag. By comparing the concentration of AgaR–His$_6$ required to observe the footprints in Figure 10, it can be concluded that AgaR has roughly equal affinity for operators specific for P_S and P_Z, but an approximately two to three fold lower affinity for the operator identified near P_R. Results from gel mobility shift assays were consistent with these conclusions (data not shown).



(d)

O_{R1}	A A A C T T T C G T T T C A T t t c G T T T T G
O_{Z1}	A A T C T T T C G T T T T G T t t c A G T T G A
cO_{Z2}	T A **T C T** T T C G T T T T **A T t** t t T A T C T C
O_{S1}	T G C C T T T C G T T T C A T t t c A C T T G G
Consensus	W A N C T T T C G T T T Y A T t t c N N T T K N
pO_{R2}	G A T A A T T C G T T T A C G t t c G G G T G G
pO_{S2}	G A A C T T T C A G T T T C T t t t C T A T A G
pO_{S3}	C A A A T T T C G T T T T A T t t c T T T T T T
pO_{S4}	T C C C T T T T G C T T C C T t t a T C T T T T

Figure 10. Identification of sites within the *agaR–agaZ* and *agaA–agaS* intergenic regions bound by AgaR. (a) Footprint analysis of the *agaR–agaZ* intergenic region labeled at the *agaR* end. Reactions 1–6 contained 0, 2, 4, 8, 16 and 32 nM AgaR tetramers, respectively. (b) Footprint analysis of the *agaA–agaS* intergenic region labeled at the *agaS* end. Reactions 1–5 contained 0, 6, 12, 24 and 48 nM AgaR tetramers, respectively. (c) Footprint analysis of the *agaR–agaZ* intergenic region labeled at the *agaZ* end. Reactions 1–6 contained 0, 3, 6, 12, 24 and 48 nM AgaR tetramers, respectively. Hypersensitive sites identified within O_{Z2} are labeled with arrows. Sequencing reactions A, C, G and T as described in Experimental Procedures were used as size markers. (d) The sequences of the four operator sites recognized by AgaR identified by footprint analysis are shown. cO_{Z2} corresponds to the complementary sequence of O_{Z2}. Positions corresponding to hypersensitive sites are bold and underlined. pO_{R2} corresponds to a putative operator in the P_R region and pO_{S2}, pO_{S3} and pO_{S4} correspond to putative operators in the *agaA–agaS* intergenic region not verified but exhibiting sequence similarity to the proposed consensus sequence. The proposed consensus sequence based on the four verified operator sequences is also shown where N is any nucleotide, W is either A or T, Y is either C or T and K is either G or T. Uppercase letters correspond to bases that were or are predicted to be protected from cleavage by AgaR. Sequences are written 5' to 3' except that given for pO_{R2}. Nucleotides not matching consensus are shown in red.

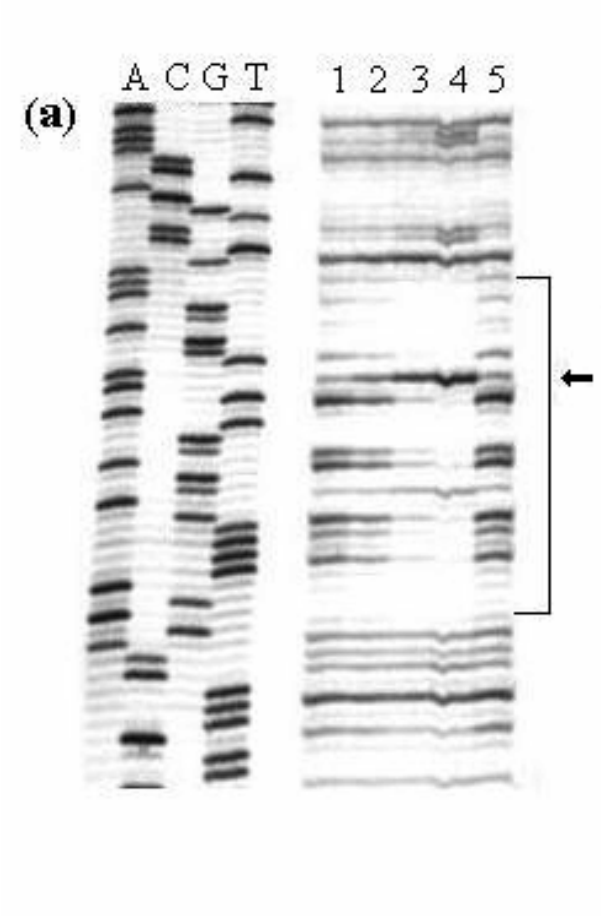
A comparison of the sequences footprinted by AgaR is shown in Figure 10(d) along with a proposed consensus sequence based on these four footprints. There are three putative operators within the *agaA–agaS* intergenic region and one near P_R that are similar to the proposed consensus sequence for AgaR. These sequences are also shown. Confirming these sites as operators proved difficult since the sites are relatively far from either end of the DNA fragments used for footprinting.

It is also interesting to note that binding of AgaR–His_6 to O_{Z2} caused enhanced cleavage by DNase I at specific sites (Figure 10(c)). Increased sensitivity to DNase I cleavage is likely due to bending of the DNA by AgaR, consistent with the observation that members of the DeoR family cause formation of DNA loops when bound to operator sites.

Footprint analysis of the *agaR–agaZ* intergenic region using cAMP–CRP. As described earlier (Table 6), it was shown that the promoter specific for *agaZ* required cAMP–CRP for full activation. Purified CRP had been obtained previously in this lab (33). This protein was thus readily available for DNase I footprint analysis of the *agaR–agaZ* intergenic region. Figure 11(a) illustrates the results of the footprinting experiment using CRP. CRP in the presence of cAMP was found to specifically protect the site in the *agaR–agaZ* intergenic region found to possess sequence similarity to other known CRP binding sites (Figure 11(b)). A hypersensitive site is also observed in the footprint identified for the cAMP–CRP complex in Figure 11. The cAMP–CRP complex is known to bend DNA (82) and the presence of the hypersensitive site is consistent with bending of the P_Z promoter region upon binding of the complex.

Mapping the transcription start sites for P_R, P_Z and P_S. S1 nuclease protection assays and the single–stranded DNA probes described in Experimental Procedures were used to map the transcription start sites for the three promoters identified in the *aga* gene cluster. The results are shown in Figure 12. The nucleotide sequences of these promoter regions

given in Figures 5 and 6 summarize the information obtained from this study concerning the organization of these promoters. A pictorial summary of the control elements identified in the *aga* promoter regions is shown in Figure 13.



(b)

```
[1]  t t c t t t T G T G A a t c a g a T C A G A a a a c c
[2]        a a a T G T G A t n t a n a T C A C A t t t
```

Figure 11. Identification of the cAMP–CRP binding site in the *agaR–agaZ* intergenic region. (a) Footprint analysis of the *agaR–agaZ* intergenic region labeled at the *agaZ* end. Reactions 1–5 contained 0, 5, 50, 250 and 250 nM CRP, respectively. Reactions 1–4 also contained 0.5 mM cAMP. Hypersensitive site is labeled with an arrow. Sequencing reactions A, C, G and T as described in Experimental Procedures were used as size markers. (b) Comparison of the sequence footprinted by cAMP–CRP in the *agaR–agaZ* intergenic region [1] and the consensus sequence recognized by cAMP–CRP [2]. Positions most important for recognition by cAMP–CRP are in capitalized (28).

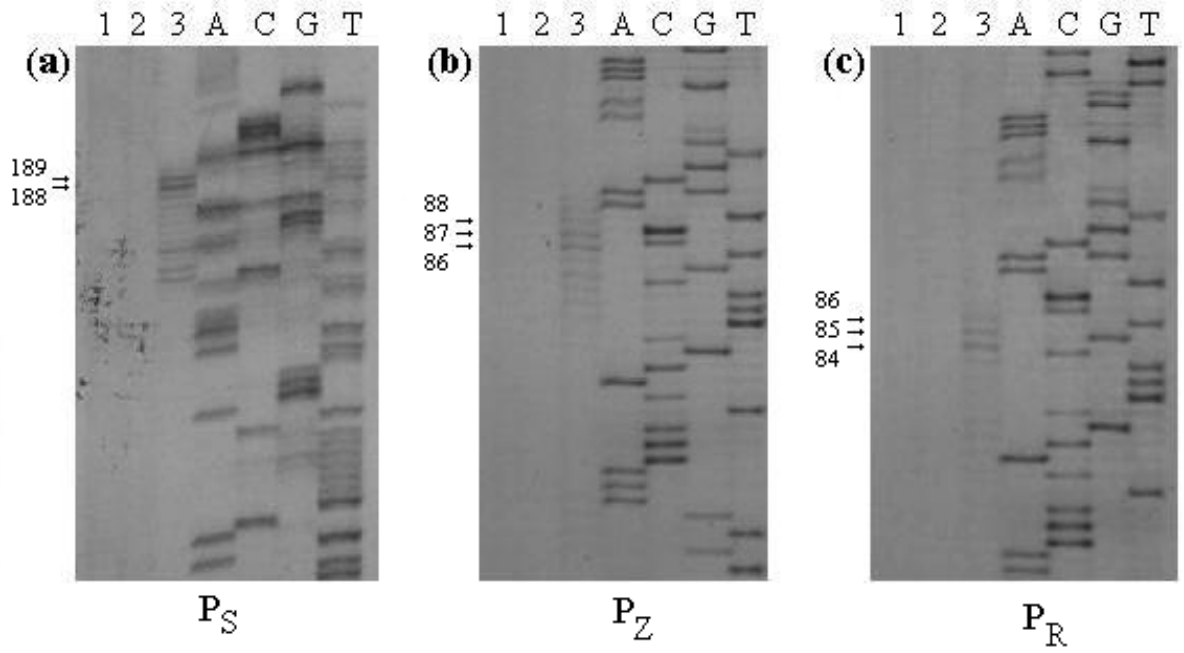


Figure 12. Mapping of the transcription start sites of *aga* promoters. Each set of assays performed used 800,000 cpm of a single–stranded DNA probe generated as described in Experimental Procedures. Lanes 1 and 2 contained 30 μg yeast RNA without (1) or with (2) treatment with S1 nuclease. Lane 3 contained 30 μg total RNA isolated from cells harboring the transcriptional fusion corresponding to the probe used in the assay. Sequencing reactions A, C, G and T as described in Experimental Procedures were used as size markers. (a) P_S. (b) P_Z. (c) P_R. Arrows indicate transcription initiation sites and are labeled with the corresponding fragment sizes as determined by comparison to the sequencing reactions. These sites are indicated in the nucleotide sequences of the promoter regions shown in Figures 5 and 6.

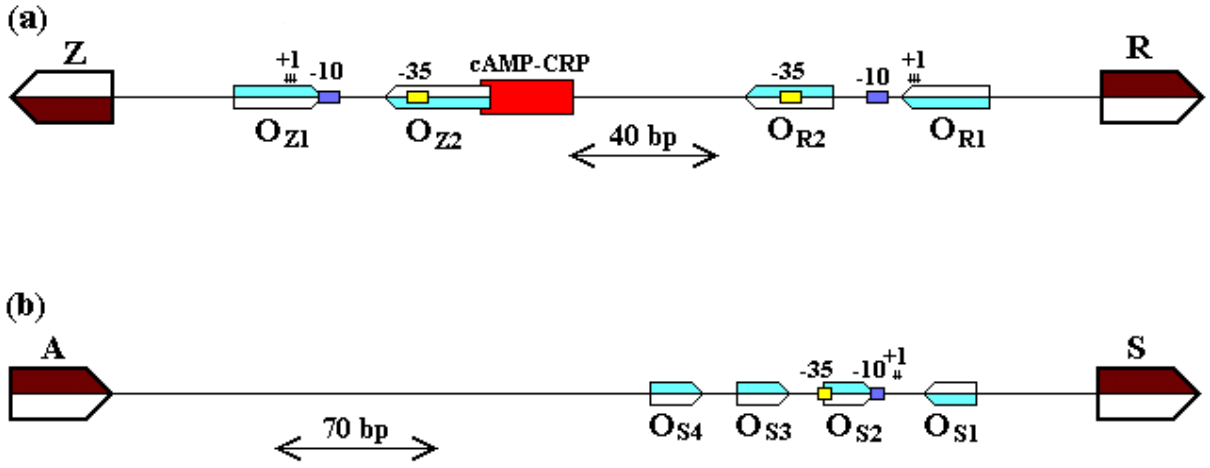


Figure 13. Organization of the P_Z, P_R and P_S promoter regions. (a) *agaR–agaZ* intergenic region containing P_Z and P_R. (b) *agaA–agaS* intergenic region containing P_S. Double–headed arrows indicate the scales used to represent the intergenic regions. Reading frames (not to scale) are shown in dark red. Transcription start sites are labeled (+1) and marked with small arrows. Putative –35 and –10 sequences are represented by yellow and violet boxes, respectively, and labeled. Verified (O_{Z1}, O_{Z2}, O_{R1} and O_{S1}) or proposed (O_{R2}, O_{S2}, O_{S3} and O_{S4}) operator sites recognized by AgaR are labeled and represented by blue arrows to indicate directionality of the operators based on sequences given in Figure 10(d). The red box represents the cAMP–CRP binding site. The position of the color within the arrows representing open reading frames and operator sites corresponds to the strand on which the sequences are located.

Many *E. coli* strains cannot grow on *N*–acetylgalactosamine without previous selection. During the course of this study, it was found that commonly used derivatives of *E. coli* K–12 including TL524 (a derivative of MG1655), DH5αZ1 (a derivative of DH5α) and MC4100 (53) cannot utilize *N*–acetylgalactosamine as the sole source of carbon without previous selection. BL21(DE3), a derivative of *E. coli* B, was able to grow on *N*–acetylgalactosamine. As described in Experimental Procedures, however, revertants gaining the ability to utilize *N*–acetylgalactosamine were readily isolated by plating cells on minimal medium containing *N*–acetylgalactosamine as the only source of carbon (strains KR201–205 and KR301–302). TL524 but not MC4100 was able to utilize galactitol without previous selection at 37°C. Using the same procedure outlined for isolating Aga^+ derivatives of TL524, a revertant of TL4100 (a derivative of MC4100) able to utilize galactitol at 37°C was also obtained (KR303). It was found that selection for the ability to utilize galactitol did not concomitantly result in the ability to utilize *N*–acetylgalactosamine nor vice versa. Though galactosamine can induce the *aga* system when included in the growth medium (Table 5), no revertants capable of utilizing this sugar derivative as a sole source of carbon could be isolated.

Selection for tetracycline–resistant or kanamycin–resistant transductants utilizing P1 lysates grown on either CAG12072 or CAG12127 (containing a Tn*10* or Tn*10kan* insertion, respectively, at minute 71.8) and Aga^+ revertants as recipient strains resulted in isolates which had become Aga^- with a frequency ranging from 20 to 90%. (The Aga phenotype was rather unstable and thus proved difficult to score.) P1 lysates grown on isolates retaining the Aga^+ phenotype were used to attempt to cotransduce the Aga^+ phenotype with either the Tn*10* or Tn*10kan* into TL524, but none of the antibiotic–resistant transductants gained the ability to utilize *N*–acetylgalactosamine. These results (which have been duplicated by researchers in another laboratory (J. W. Lengeler, personal communication)) suggest that there may be another mutation necessary for reversion to Aga^+ that is not located in the *aga* gene cluster.

To determine if the reversion to Aga^+ is due to a change in the regulation of the expression of the *aga* genes, single–lysogens with P_R–*lacZ*, P_Z–*lacZ* and P_S–*lacZ* fusions were isolated using Aga^+ derivatives of TL524 as recipients. If the reversion had occurred because of a loss of repressor activity, it was anticipated that the uninduced level of β–galactosidase in the Aga^+ derivatives would be constitutively elevated. However, Table 7 shows that the uninduced level of activity is the same as that observed in the Aga^- parent strain. It was also found that disruption of *agaR* as described above did not result in Aga^- to Aga^+ conversion.

It was determined instead that the *aga* promoters were induced by *N*–acetylgalactosamine to a greater extent in Aga^+ strains than in the Aga^- parent (Table 7). The level of induction observed in Aga^+ strains in the presence of glycerol was not significantly different from that observed in the Aga^- parent without glycerol in the growth medium (see Table 5). However, the results in Table 7 show that there is some difference between the parent strain and Aga^+ revertants affecting the regulation of all three promoters. Two possibilities to explain these differences are: i) an increase in the sensitivity of AgaR to the inducer or ii) an increase in the uptake or level of the inducer. It was found that overexpression of AgaR, even from a low–copy number plasmid (pKR311), prevented Aga^+ strains from growing on *N*–acetylgalactosamine. To determine if the Aga^+ strains

possessed an altered form of repressor, plasmids were constructed carrying the *agaR* gene from Aga^+ strains. Overexpression of AgaR from these plasmids (pKR611 and pKR613) also resulted in an Aga^- phenotype when introduced into an Aga^+ host. This argued against the possibility that the reversion from Aga^- to Aga^+ was due to a change in the sensitivity of AgaR to its inducer. Therefore, increased accumulation of inducer seems the more likely explanation for the Aga^- to Aga^+ reversion.

TABLE 7

Induction of *aga* promoters in Aga^- and Aga^+ Strains

Parent strain of lysogen	Transcriptional fusion		
	P_R–*lacZ* (λKR161)	P_Z–*lacZ* (λKR162)	P_S–*lacZ* (λKR164)
– *N*–acetylgalactosamine [a]:			
TL524 (Aga^-)	99 ± 2	26 ± 2	19 ± 12
KR201 (Aga^+)	95 ± 5	26 ± 2	65 ± 13
KR203 (Aga^+)	98 ± 3	22 ± 3	34 ± 3
+ *N*–acetylgalactosamine:			
TL524 (Aga^-)	260 ± 7	120 ± 10	290 ± 45
KR201 (Aga^+)	830 ± 60	590 ± 15	11,000 ± 1,500
KR203 (Aga^+)	1,030 ± 20	710 ± 30	12,200 ± 250

[a] Cells growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine, 0.05% casamino acids and 0.4% glycerol. *N*–acetylgalactosamine (10 mM) was added where indicated.

Phenotypes of R49A and T116I variants of AgaR. Previous work with GlpR in this laboratory has shown that certain residues conserved in the DeoR family are important for binding to either operator DNA or inducer. Plasmids expressing variants of AgaR containing substitutions corresponding to those in GlpR that confer either a negatively dominant (R49A) (6) or noninducible (T116I) (24) phenotype were constructed. As mentioned previously, the arginine located within the recognition helix is absolutely conserved. Substitution of this residue in GlpR was shown to prevent binding to operator DNA (6) but this variant of GlpR retains the ability to oligomerize, and it is thus negatively dominant to chromosomally–encoded GlpR. The threonine is a highly conserved residue (T101 in GlpR) that when substituted results in a variant that is less sensitive to the presence of the inducer,

glycerol–P. This is believed to be due to a lower affinity for binding of the inducer (24) thus resulting in a noninducible phenotype, i.e. unable to utilize glycerol–P.

KR164, possessing a single copy P_S–*lacZ* transcriptional fusion and a wild–type *agaR* gene, was used to test the phenotype of the negatively dominant variant, AgaR–R49A. Promoter activity was determined in this strain harboring a low–copy number plasmid that expresses either AgaR (pKR311), AgaR–R49A (pKR316) or no variant of AgaR (pKR310 (vector)). Cells harboring the vector control plasmid (pKR310) exhibited the expected pattern of induction in the presence of *N*–acetylgalactosamine (Table 8(a)). Cells expressing AgaR–R49A (pKR316) were found to have a constitutively elevated level of β–galactosidase activity independent of the presence of inducer. The constitutively elevated level of expression from P_S is consistent with AgaR–R49A being a negatively dominant variant of AgaR. Expression of wild–type AgaR from the low–copy number plasmid pKR311 resulted in a very low level of activity that was almost completely insensitive to the presence of inducer. Interestingly, this is the phenotype expected from a noninducible variant of the protein. This is consistent with the observation that expression of AgaR from this plasmid caused Aga^+ revertants to become Aga^-.

To distinguish between a noninducible variant and wild–type AgaR, which appears noninducible in the system described above, a different experimental system was designed. The low–copy number plasmids expressing AgaR and AgaR–T116I, pKR311 and pKR315, respectively, were introduced into DH5αZ1 cells harboring the P_S–*lacZ* transcriptional fusion on the compatible multi–copy plasmid pKR164. As shown in Table 8(b), P_S–*lacZ* expression from the multi–copy plasmid was constitutively elevated in cells harboring the vector control (pKR310). Cells expressing wild–type AgaR exhibited a much lower level of activity and a 14–fold increase in the activity was observed when the medium contained *N*–acetylgalactosamine. AgaR–T116I, however, was insensitive to induction by *N*–acetylgalactosamine, consistent with a noninducible phenotype. Since substitution of this highly conserved threonine results in a noninducible variant of both GlpR (24) and AgaR, this threonine is likely to play a role in inducer binding for all members of the DeoR family. However, it should be noted that in some members of this family a serine is present at this position and work with GlpR has shown that introduction of a serine residue at this position (GlpR–T101S) also resulted in a noninducible phenotype (24).

Search for the inducer molecule for AgaR. Gel mobility shift assays were used in an attempt to identify the inducer for AgaR. Based on the proposed pathway for the metabolism of *N*–acetylgalactosamine, likely candidates are *N*–acetylgalactosamine–6–phosphate or galactosamine–6–phosphate. As described previously, growth with galactitol does not result in induction of the promoters in the *aga* cluster even in strains able to utilize galactitol at elevated temperatures (data not shown), conditions in which the tagatose–6–phosphate aldolase encoded by *agaY/Z* functions to metabolize galactitol (J. W. Lengeler, personal communication). The metabolic pathways for the utilization of the two sugars converge at tagatose–6–phosphate. Therefore, any compound in the pathway at or following the point of convergence may be ruled out as the inducer for AgaR (see Figure 4). (However, there is evidence that in some strains selected for growth on

Table 8

AgaR–R49A is negatively dominant and AgaR–T116I is noninducible

(a)

Growth conditions[a]	Repressor variant		
	None (pKR310)	AgaR (pKR311)	AgaR–R49A (pKR316)
– *N*–acetylgalactosamine	60 ± 10	1.5 ± 0.1	24,500 ± 1,300
+ *N*–acetylgalactosamine	1,600 ± 60	2.7 ± 1.2	24,800 ± 800

(b)

Growth Conditions[b]	Repressor variant		
	None (pKR310)	AgaR (pKR311)	AgaR–T116I (pKR315)
– *N*–acetylgalactosamine	21,600 ± 3,000	72 ± 3	54 ± 7
+ *N*–acetylgalactosamine	22,000 ± 11,000	1,040 ± 5	54 ± 10

[a] Cells (KR164 harboring one of the plasmids indicated above) growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine, 0.2% casamino acids and 50 μg/ml kanamycin. Cultures were supplemented with 10 mM *N*–acetylgalactosamine where indicated.

[b] Cells (DH5αZ1 harboring pKR164 and one of the plasmids indicated above) growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine, 0.2% casamino acids, 100 μg/ml ampicillin, 20 μg/ml spectinomycin and 50 μg/ml kanamycin. Cultures were supplemented with 10 mM *N*–acetylgalactosamine where indicated.

galactitol at 37°C, the presence of galactitol does result in elevated expression of *aga* genes (J. W. Lengeler, personal communication). The nature of this inconsistency has yet to be determined.)

Unfortunately, neither *N*–acetylgalactosamine–6–phosphate nor galactosamine–6–phosphate is commercially available. A report of an *N*–acetylglucosamine–6–phosphate deacetylase from *Vibrio cholerae* that can also hydrolyze *N*–acetylglucosamine–6–sulfate (83) prompted me to test the sulfurylated compounds, which are commercially available, for an effect on binding of operator DNA by AgaR.

Gel mobility shift assays containing a radiolabeled DNA fragment with the *agaR–agaZ* promoter region and 10 nM AgaR–His_6 tetramers were supplemented with varying amounts of either *N*–acetylglucosamine–6–sulfate, glucosamine–6–sulfate or galactosamine–6–sulfate and electrophoresed on a non–denaturing polyacrylamide gel as described in Experimental Procedures. (*N*–acetylgalactosamine–6–sulfate was not immediately available.) The presence of these sugar derivatives, even at concentrations as high as 10 mM, had no effect on the shift caused by binding of AgaR–His_6 to the radiolabeled DNA fragment (Figure 14). That these compounds do not cause induction, however, does not rule out the possibility that the corresponding phosphorylated sugar derivative would. As mentioned previously, the inducers for members of the DeoR family, for which one has been identified, are all phosphorylated compounds. Therefore, it is reasonable to assume that the phosphate group is critical in inducer binding and the orientation and charge distribution of the sulfate group may not correctly mimic that of the phosphate group.

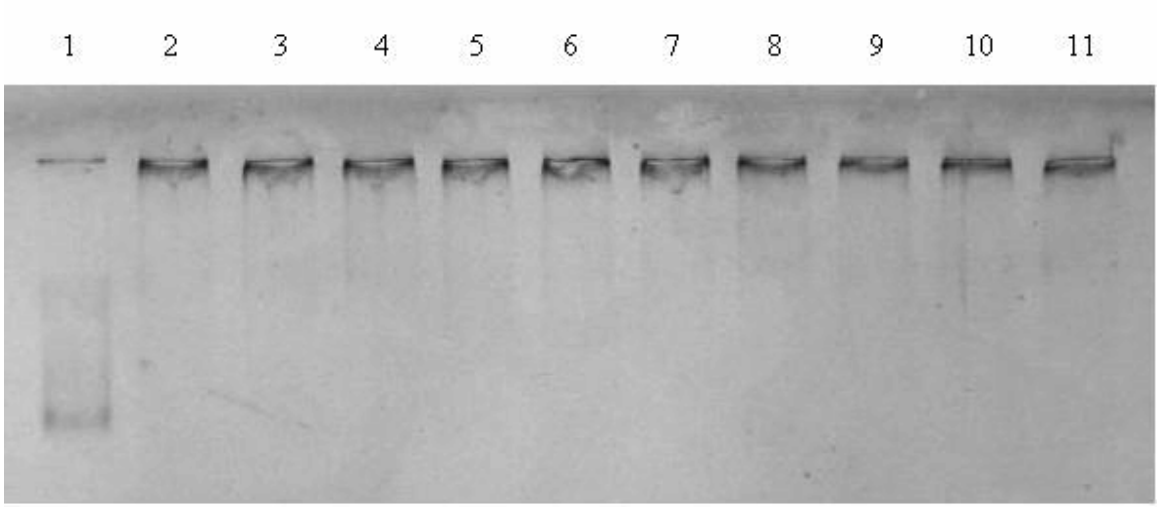


Figure 14. Testing of the effect of sugar–sulfates on DNA operator binding of AgaR–His_6. Each assay contained approximately 200,000 cpm of radiolabeled *agaR–agaZ* intergenic region. All assays but (1) also contained 10 nM AgaR–His_6 tetramer. (2) Control without any added sugar derivative. (3–5) Supplemented with 0.1, 1 and 10 mM D–galactosamine–6–sulfate, respectively. (6–8) Supplemented with 0.1, 1 and 10 mM D–glucosamine–6–sulfate, respectively. (9–11) Supplemented with 0.1, 1 and 10 mM *N*–acetylglucosamine–6–sulfate, respectively.

Discussion

In this study, it has been confirmed that the *aga* gene cluster is involved in the metabolism of *N*–acetylgalactosamine as proposed previously by Saier et al. (12). It has also been shown that these genes, including the gene encoding AgaR, are negatively regulated by AgaR. These conclusions are drawn from experiments utilizing transcriptional fusions of promoters identified in the *aga* cluster to *lacZ* (encoding β–galactosidase). The three promoters identified in the *aga* gene cluster, P_R, P_Z and P_S, were shown to be targets of repression by AgaR using both *in vivo* comparison of the promoter activities in *agaR*$^+$ versus *agaR*::Tc^r backgrounds and DNase I footprint analysis of the promoter regions with purified AgaR–His_6.

AgaR–His_6 was found to exist as a tetramer under nondenaturing conditions. The operators in the promoter regions, including those identified by DNase I footprint analysis and those proposed based on sequence similarity to the experimentally verified operators, were found to be neither directly repeated nor palindromic sequences. The operators are instead arranged to form pseudo–palindromic binding sites since the operators appear to be roughly evenly spaced and are found in both orientations within the same promoter regions (see Figure 13). The relative distances between the operators in these promoter regions and the fact that AgaR is a tetramer suggest that, like GlpR and DeoR, AgaR binds several operators simultaneously and may result in a DNA loop around the site of transcription initiation (29–33). The presence of hypersensitive sites within O_{Z2} is consistent with bending of the operator regions due to binding of AgaR. Such bending may facilitate DNA looping that seems to be a common theme among the DeoR family.

It appears that AgaR is the main factor regulating expression from P_R and P_S, but P_Z requires both induction of AgaR by the presence of either *N*–acetylgalactosamine or galactosamine in the growth medium and activation by the cAMP–CRP complex. The site bound by cAMP–CRP determined by DNase I footprint analysis is centered at –60.5 nucleotides relative to the transcription start site for P_Z (Figure 5). The position of the site is indicative of a class I cAMP–CRP binding site since the site does not overlap the –35 sequence for this promoter (84). The elements identified in this study responsible for regulation of the expression from the *aga* promoters are summarized in Figure 13.

Of the three transcriptional promoters identified and mapped in the *aga* gene cluster, only two (P_R and P_Z) were predicted by the *E. coli* Genome Sequencing Project. It is interesting to note that P_S, the one not predicted, was found to be a very strong promoter. The relative strength of this promoter can be explained by the optimal spacing (17 bp) and the close match to consensus of its –35 (TTGAAC versus the TTGACA consensus) and –10 (TAGAAT versus the TATAAT consensus) sequences. Sequence analysis of this region by the *E. coli* Genome Sequencing Project had also predicted a transcriptional promoter in the *agaY–agaB* intergenic region (see Figure 3). The existence of this promoter has not been investigated. However, no sequences resembling the AgaR binding sites identified in this study are present in this region suggesting that if a promoter in this region does exist it may only be minimally sensitive to repression by AgaR.

That there is no bacterial phosphotransferase system (PTS) enzyme IIA^{Man} homologue in the *aga* gene cluster indicates that other gene(s) elsewhere on the chromosome may be

required for *N*–acetylgalactosamine uptake and/or utilization (12). Such genes might also be expected to be negatively regulated by AgaR. A reading frame (*yadI*) tentatively identified as *agaX* by Saier et al. (12) in the 2.4 to 4.1 minute region of the *E. coli* genome encodes a PTS enzyme IIA^{Man} homologue but no apparent AgaR binding sites are located upstream of this gene (12).

N–acetylgalactosamine, in addition to being a utilizable carbon source in *E. coli*, is also a component of lipopolysaccharides that coat the surface of cells in some strains of *E. coli* (85,86). Therefore, it is likely that some strains of *E. coli* may possess the ability to synthesize as well as utilize *N*–acetylgalactosamine as they do *N*–acetylglucosamine (45,87,88). In *B. subtilis*, *N*–acetylgalactosamine is synthesized from *N*–acetylglucosamine via a UDP–N–acetylglucosamine–4–epimerase (89). This same activity may be catalyzed by GalE (UDP–glucose–4–epimerase) or a homologous protein in *E. coli*. Eukaryotic and prokaryotic UDP–glucose–4–epimerases share a striking degree of sequence similarity (BLAST search results, (11)) and at least one eukaryotic UDP–glucose–4–epimerase has been shown to utilize UDP–glucose and UDP–*N*–acetylglucosamine equally well (90). It was also shown that the reactions catalyzed by the porcine UDP–glucose–4–epimerase were readily reversible (90).

The transcriptional regulator, NagC, represses expression of the *nag* regulon, encoding proteins involved in the utilization of *N*–acetylglucosamine (36,87). Recent experimental evidence suggests that NagC may also function to activate expression of *glmS*, encoding a glucosamine–6–phosphate synthase, thereby coordinately regulating the synthesis and degradation of *N*–acetylglucosamine (45). Analogously, the gene encoding a UDP–*N*–acetylglucosamine–4–epimerase to function in the synthesis of *N*–acetylgalactosamine for incorporation into lipopolysaccharides may be positively regulated by AgaR.

It is interesting to note that the protein encoded by *agaS* displays sequence similarity to the carboxy–terminal domain of the glucosamine–6–phosphate synthase encoded by *glmS*, but the significance of this similarity is presently unknown. It is the carboxy–terminal domain of the glucosamine–6–phosphate synthases that possesses the ketose–aldose isomerase activity and the protein encoded by *agaS* has already been tentatively identified as a tagatose–6–phosphate ketose/aldose isomerase. The promoter specific for this gene is very strongly repressed by AgaR, therefore, the gene product would be expected to be involved in the degradation not the synthesis of *N*–acetylgalactosamine. The physiological role played by this protein in the metabolism of *N*–acetylgalactosamine remains to be elucidated.

The reasons for the Aga^- phenotype of K–12 strains and the mutations necessary for reversion to Aga^+ are still unclear. One possible explanation is that the insertion of genes from *V. furnissii* has interfered with the ability of these strains to import *N*–acetylgalactosamine. The sequence similarity between the PTS enzymes encoded by *agaB* and *agaV* or *agaC* and the truncated *agaW* may result in some sort of competition between the two sets of PTS proteins. By PCR analysis using chromosomal DNA isolated from both an Aga^- parent and two Aga^+ revertants, it was determined that no gross change occurs in the organization of the *aga* cluster in the region affected by the horizontal transfer of DNA from *V. furnissii*. It would be interesting to determine if the Aga^+ phenotype is common to B

strains such as BL21(DE3), and if these strains possess the genes originating in *V. furnisii*. The horizontal transfer event may have occurred after the divergence of the K–12 and B subspecies of *E. coli*.

It was found that selection for the ability to utilize *N*–acetylgalactosamine does not lead to constitutive expression from *aga* promoters, but the global regulation of the promoters has been altered (Table 7). The results thus far tend to indicate that the reversion results in an increased ability to accumulate the inducer for AgaR (this study and J.W. Lengeler, personal communication). The accumulation may be due to increased uptake of *N*–acetylgalactosamine or decreased degradation of the inducer molecule or perhaps a combination of both. Evidence suggests that there are at least two mutations necessary for the reversion to Aga^+ and not all of the mutations necessary for the reversion are located within the *aga* cluster. Determination of the nature and locations of these mutations will require further analysis.

As mentioned previously, the long–term objective of this research is to determine the functions of all members of the DeoR family in *E. coli*. Important for this objective was the ability to construct plasmids that express either a negatively dominant or noninducible variant of AgaR. The variants of AgaR containing substitutions R49A or T116I were shown to be negatively dominant or noninducible, respectively (Table 8). The substitutions R49A and T116I are analogous to substitutions previously shown in GlpR to give rise to negatively dominant and noninducible variants, respectively (6,24). In strains possessing the ability to utilize *N*–acetylgalactosamine as a sole source of carbon, co–expression of the negatively dominant or noninducible variants with chromosomally–encoded wild–type AgaR resulted in Aga^+ or Aga^- phenotype, respectively. Thus, the proposed phenotypic screening would have allowed identification of the metabolic pathway regulated by AgaR. However, as described above, the strains of *E. coli* K–12 available in our laboratory lacked the ability to utilize *N*–acetylgalactosamine without previous selection. This shows that cryptic genes or pathways may complicate the proposed screening process. We therefore propose to combine the phenotypic screening with microarray technology to determine changes in mRNA levels in strains expressing negatively dominant versus noninducible variants of the repressors of unknown function. Using microarray methods (91–93), it should be possible to identify all genes regulated by the repressors which will provide information regarding the metabolic pathways controlled by the repressors.

SECTION II:
Characterization of YihW

Introduction

YihW is a member of the DeoR family with a presently undefined function. The *yihW* gene is located at minute 87.7 of the *E. coli* genome within a cluster of uncharacterized genes (Figure 15). One of the most striking features of YihW is its sequence similarity to GlpR, the transcriptional repressor regulating the expression of genes involved in the metabolism of glycerol–P (see Figures 1 and 2). Most proteins in the DeoR family exhibit 25–35% identity to one another but YihW is 44% identical (69% similar) to GlpR. Because of this striking sequence similarity we hypothesized that YihW may function to regulate the same promoters controlled by GlpR, similarly to the repression of *gal* genes involved in galactose metabolism mediated by the isorepressors GalR and GalS (94). Furthermore, we hypothesized that YihW may also regulate promoters within the *yih* gene cluster.

In this study I have attempted to determine the metabolic pathway encoded by the genes in the *yih* cluster. I have verified the existence of transcriptional promoters between the divergently transcribed genes *yihU* and *yihV* and shown that these promoters are negatively–controlled by YihW. I have also shown that, despite its high degree of identity with GlpR, YihW or even variants that have substitution(s) to make the recognition helix more closely resemble that of GlpR cannot regulate transcription from a promoter known to be controlled by GlpR.

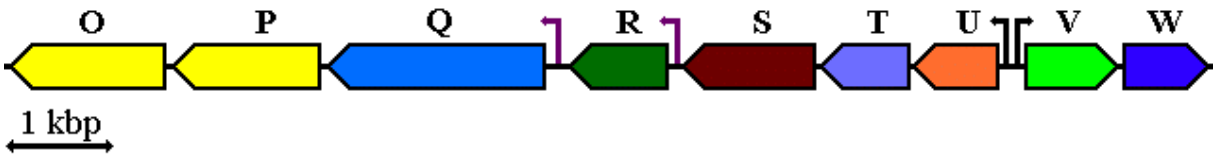


Figure 15. The *yih* gene cluster at minute 87.7 of the *E. coli* genome. Genes O and P encode putative permeases or symporters. Q and R encode a putative α–glycosidase and 1 or 4–epimerase, respectively. S encodes a protein having sequence similarity to phosphomannose isomerases and was tentatively identified as *manC* (*mni*). T, U and V encode a putative aldolase, a dehydrogenase and a kinase, respectively. W encodes the transcriptional repressor. The locations of the transcriptional promoters identified in this work are shown by black arrows. Purple arrows indicate positions of promoters predicted by *the E. coli* Genome Sequencing Project but not yet verified.

The kinase encoded by *yihV* was overexpressed but the overexpressed protein was found predominantly within insoluble inclusion bodies. Efforts to identify the substrate for this kinase proved inconclusive. Other attempts to elucidate the metabolic pathway encoded by the *yih* gene cluster using various techniques also proved inconclusive. However, the discovery that one of the genes within the cluster (*yihS*) probably encodes a protein possessing D–mannose isomerase activity previously described by Wu et al. (95,96) may help future researchers design experiments to determine the metabolic pathway encoded by the *yih* gene cluster.

Experimental Procedures

Unless explicitly described below, the experimental procedures and materials used for this section are the same as for Section I.

Materials and reagents. Synthetic oligonucleotides used in the study of YihW are listed in Table 9. Position numbers given in Table 9 correspond to those shown in Figure 16.

Bacterial strains. Strains used or constructed are described in Table 10. DH5αZ1 (49) was used as a host during plasmid construction. It was also used as the host strain for β–galactosidase assays of transcriptional fusions present in multicopy on pSP417 derivatives and for overexpression of YihW or variants expressed from plasmids derived from the pZ family of expression plasmids (49).

Phages and plasmids. Phage λDD876 (61) was obtained from the *E. coli* Genome Sequencing Project and contains the *yih* region from strain MG1655, the wild–type isolate used for the sequencing project (8). λKR150, λKR151 and λKR152 contain the $P_{K/R}$–*lacZ*, P_K–*lacZ* and P_{dH}–*lacZ* transcriptional fusions derived from pKR150, pKR151 and pKR152, respectively. Phage M111 is an M13mp19 derivative containing the same insert as pKR101 described below. M113 to M116, derived from M111, contain *yihW* with mutations introduced using the Kunkel method for site–directed mutagenesis (97) corresponding to pKR103 to pKR106. Propagation of phages and isolation of replicative form M13 DNA was performed utilizing previously published methods (68).

Plasmids used or constructed in this study are listed in Table 11. DNA fragments were amplified by PCR using chromosomal DNA isolated from MG1655 as template. Plasmids expressing variants of YihW contain inserts derived from M113 to M116 described above. Naturally occurring restriction sites used for cloning are shown in Figure 16.

The steps utilized in the construction of the plasmid (pKR117) for creating a *yihW* gene knockout strain (KR2524) are as follows: (a) The *yihV* and *yihW* reading frames were amplified by PCR, digested with *Eco*RI and *Xba*I and cloned into the *Eco*RI and *Xba*I sites of pZE12–*luc* (49) yielding pKR114. (b) The *Eco*RI to *Mfe*I region upstream of *yihV* was removed, thereby destroying the *Eco*RI site (which can be recognized by *Apo*I) yielding pKR115. (c) The kanamycin–resistance cassette from pCP15 (50) was isolated using *Hin*dIII and *Eco*RI. The *Hin*dIII site was made blunt–ended by treatment with T7 DNA polymerase and dNTPs prior to digestion with *Eco*RI. This fragment was then cloned into the *Apo*I and *Sma*I sites just upstream and within the *yihW* gene, respectively, yielding pKR116. (d) The disrupted *yihW* gene was removed from pKR116 by digestion with *Hpa*I (sites within the *yihV* gene and just downstream of *yihW*). This fragment was cloned into the *Sma*I site of pKO3 (67). The resulting plasmid, pKR117, was used to create a *yihW*::Km^r strain (KR2524) using the series of selections described by Church et al. (67) and for the construction of *agaR* knockout strains in Section I.

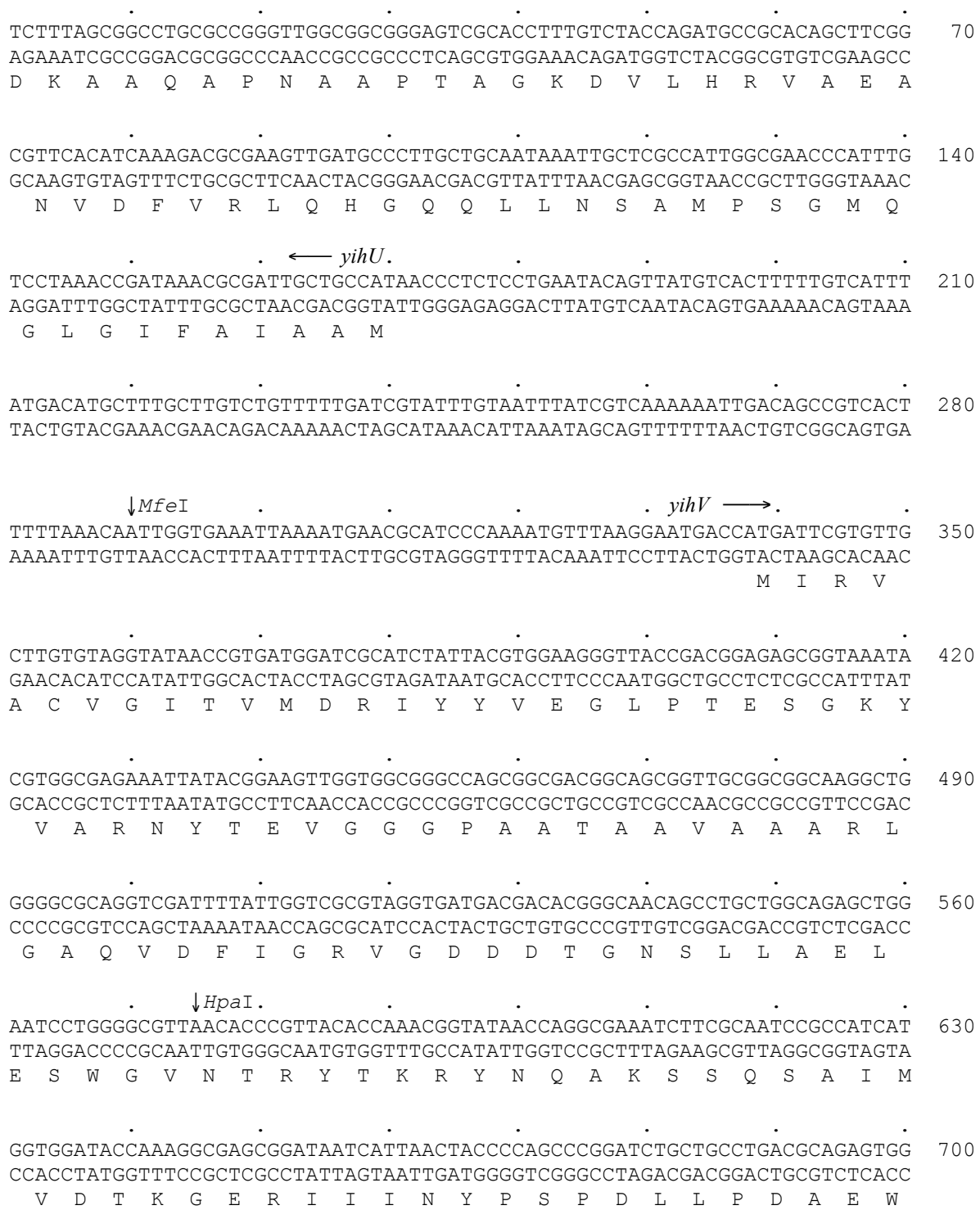
TABLE 9

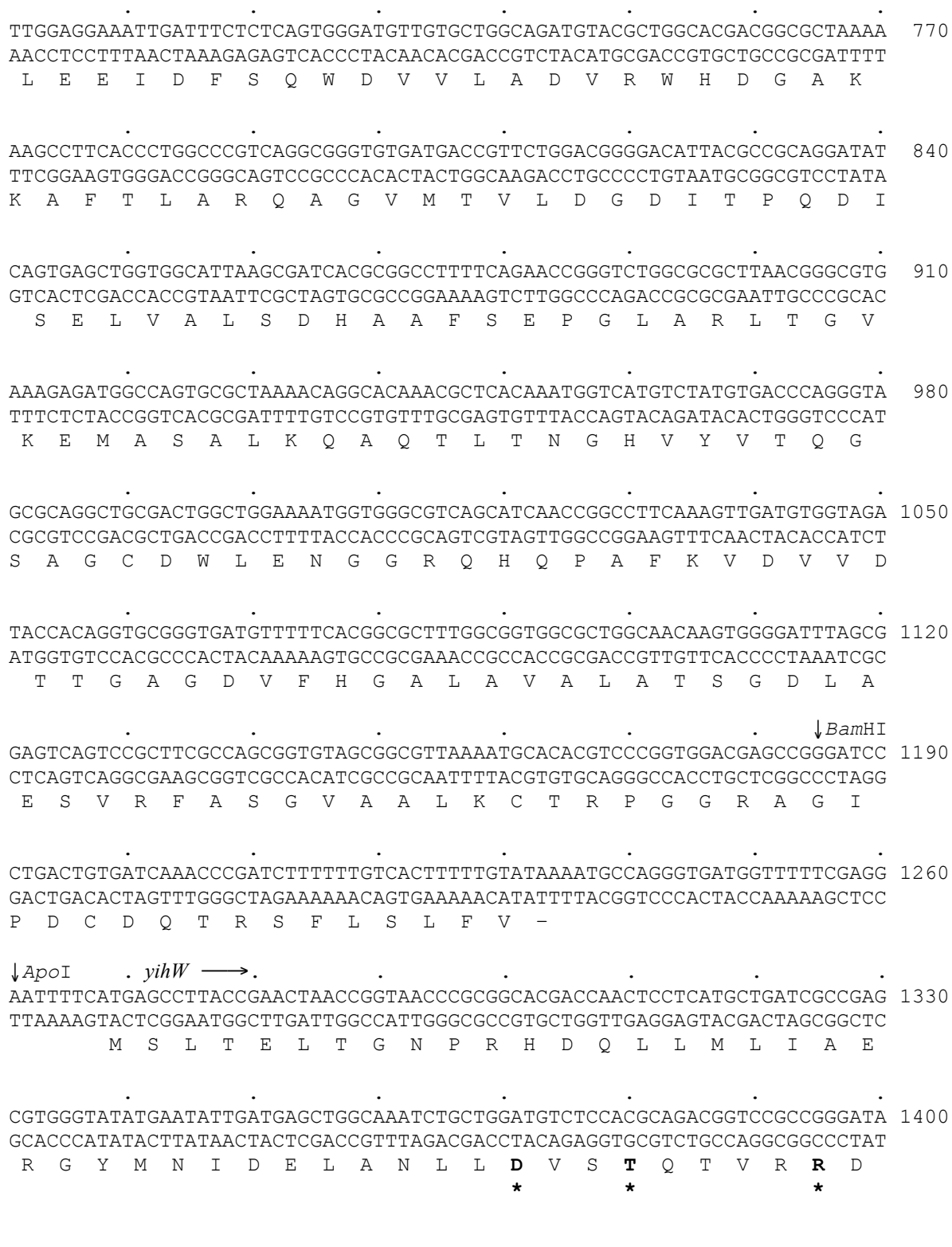
Oligonucleotides used in the study of YihW

Name	Sequence (5' → 3')[a]	Position[b]
069886	GCACAGCTTCGGaaTTCACATCAAAGACGC	59–88
207093	CAATTGGTGAAtTcAAAATGAACGCATCCC	288–305
240410	CGGGTTACCGtcTAGaTCGGTAAGGCTCATG	1298–1267
207118	AGCCGGaATtCCTGACTGTGATCAAACCC	1180–1208
183282	TCGCGCCtCtAGAAATTAACTGTAGCGGG	2139–2111
336908	GGAATTcatATGAGCCTTACCGAAC	1259–1283
372516	CTcCGCGTCgaCCTGGGGAAGAATG	2053–2029
069736	GCAAATCTGCTGtcTGTCTCCcCGCAGACGGTC	1358–1390
069796	GGATGTCTCCcCGCAGACGGTC	1369–1390
082794	GACGGTCCGCgcGGATATTCGTA	1384–1406
077757	CACCATTGGTATCACTGTTGAGCA	1573–1596
083332	CGAGCGTGGGTATATGA	1327–1343
225250	ATTCCTGATGGTTCAAC	1547–1563
094406	GAGAGCGATGGCGCGT	1808–1823
279425	GATATCAGTGAGCTGGTG	836–853
299586	CGAGAAATTATACGGAAG	426–443
269216	CGTGCTGATTATCTGGTA	1772–1789
279293	CATGGTGGATACCAAAGG	628–645

[a] Lower case letters indicate base substitutions to create restriction sites or mutations.

[b] Position numbers correspond to those given in Figure 16.


TCTTTAGCGGCCTGCGCCGGGTTGGCGGCGGGAGTCGCACCTTTGTCTACCAGATGCCGCACAGCTTCGG 70
AGAAATCGCCGGACGCGGCCCAACCGCCGCCCTCAGCGTGGAAACAGATGGTCTACGGCGTGTCGAAGCC
D K A A Q A P N A A P T A G K D V L H R V A E A
CGTTCACATCAAAGACGCGAAGTTGATGCCCTTGCTGCAATAAATTGCTCGCCATTGGCGAACCCATTTG 140
GCAAGTGTAGTTTCTGCGCTTCAACTACGGGAACGACGTTATTTAACGAGCGGTAACCGCTTGGGTAAAC
N V D F V R L Q H G Q Q L L N S A M P S G M Q
←— yihU.
TCCTAAACCGATAAACGCGATTGCTGCCATAACCCTCTCCTGAATACAGTTATGTCACTTTTTGTCATTT 210
AGGATTTGGCTATTTGCGCTAACGACGGTATTGGGAGAGGACTTATGTCAATACAGTGAAAAACAGTAAA
G L G I F A I A A M
ATGACATGCTTTGCTTGTCTGTTTTTGATCGTATTTGTAATTTATCGTCAAAAAATTGACAGCCGTCACT 280
TACTGTACGAAACGAACAGACAAAAACTAGCATAAACATTAAATAGCAGTTTTTTAACTGTCGGCAGTGA
↓MfeI yihV ——→.
TTTTAAACAATTGGTGAAATTAAAATGAACGCATCCCAAAATGTTTAAGGAATGACCATGATTCGTGTTG 350
AAAATTTGTTAACCACTTTAATTTTACTTGCGTAGGGTTTTACAAATTCCTTACTGGTACTAAGCACAAC
M I R V
CTTGTGTAGGTATAACCGTGATGGATCGCATCTATTACGTGGAAGGGTTACCGACGGAGAGCGGTAAATA 420
GAACACATCCATATTGGCACTACCTAGCGTAGATAATGCACCTTCCCAATGGCTGCCTCTCGCCATTTAT
A C V G I T V M D R I Y Y V E G L P T E S G K Y
CGTGGCGAGAAATTATACGGAAGTTGGTGGCGGGCCAGCGGCGACGGCAGCGGTTGCGGCGGCAAGGCTG 490
GCACCGCTCTTTAATATGCCTTCAACCACCGCCCGGTCGCCGCTGCCGTCGCCAACGCCGCCGTTCCGAC
V A R N Y T E V G G G P A A T A A V A A A R L
GGGGCGCAGGTCGATTTTATTGGTCGCGTAGGTGATGACGACACGGGCAACAGCCTGCTGGCAGAGCTGG 560
CCCCGCGTCCAGCTAAAATAACCAGCGCATCCACTACTGCTGTGCCCGTTGTCGGACGACCGTCTCGACC
G A Q V D F I G R V G D D D T G N S L L A E L
↓HpaI.
AATCCTGGGGCGTTAACACCCGTTACACCAAACGGTATAACCAGGCGAAATCTTCGCAATCCGCCATCAT 630
TTAGGACCCCGCAATTGTGGGCAATGTGGTTTGCCATATTGGTCCGCTTTAGAAGCGTTAGGCGGTAGTA
E S W G V N T R Y T K R Y N Q A K S S Q S A I M
GGTGGATACCAAAGGCGAGCGGATAATCATTAACTACCCCAGCCCGGATCTGCTGCCTGACGCAGAGTGG 700
CCACCTATGGTTTCCGCTCGCCTATTAGTAATTGATGGGGTCGGGCCTAGACGACGGACTGCGTCTCACC
V D T K G E R I I I N Y P S P D L L P D A E W


TTGGAGGAAATTGATTTCTCTCAGTGGGATGTTGTGCTGGCAGATGTACGCTGGCACGACGGCGCTAAAA 770
AACCTCCTTTAACTAAAGAGAGTCACCCTACAACACGACCGTCTACATGCGACCGTGCTGCCGCGATTTT
L E E I D F S Q W D V V L A D V R W H D G A K
AAGCCTTCACCCTGGCCCGTCAGGCGGGTGTGATGACCGTTCTGGACGGGGACATTACGCCGCAGGATAT 840
TTCGGAAGTGGGACCGGGCAGTCCGCCCACACTACTGGCAAGACCTGCCCCTGTAATGCGGCGTCCTATA
K A F T L A R Q A G V M T V L D G D I T P Q D I
CAGTGAGCTGGTGGCATTAAGCGATCACGCGGCCTTTTCAGAACCGGGTCTGGCGCGCTTAACGGGCGTG 910
GTCACTCGACCACCGTAATTCGCTAGTGCGCCGGAAAAGTCTTGGCCCAGACCGCGCGAATTGCCCGCAC
S E L V A L S D H A A F S E P G L A R L T G V
AAAGAGATGGCCAGTGCGCTAAAACAGGCACAAACGCTCACAAATGGTCATGTCTATGTGACCCAGGGTA 980
TTTCTCTACCGGTCACGCGATTTTGTCCGTGTTTGCGAGTGTTTACCAGTACAGATACACTGGGTCCCAT
K E M A S A L K Q A Q T L T N G H V Y V T Q G
GCGCAGGCTGCGACTGGCTGGAAAATGGTGGGCGTCAGCATCAACCGGCCTTCAAAGTTGATGTGGTAGA 1050
CGCGTCCGACGCTGACCGACCTTTTACCACCCGCAGTCGTAGTTGGCCGGAAGTTTCAACTACACCATCT
S A G C D W L E N G G R Q H Q P A F K V D V V D
TACCACAGGTGCGGGTGATGTTTTTCACGGCGCTTTGGCGGTGGCGCTGGCAACAAGTGGGGATTTAGCG 1120
ATGGTGTCCACGCCCACTACAAAAAGTGCCGCGAAACCGCCACCGCGACCGTTGTTCACCCCTAAATCGC
T T G A G D V F H G A L A V A L A T S G D L A
↓BamHI
GAGTCAGTCCGCTTCGCCAGCGGTGTAGCGGCGTTAAAATGCACACGTCCCGGTGGACGAGCCGGGATCC 1190
CTCAGTCAGGCGAAGCGGTCGCCACATCGCCGCAATTTTACGTGTGCAGGGCCACCTGCTCGGCCCTAGG
E S V R F A S G V A A L K C T R P G G R A G I
CTGACTGTGATCAAACCCGATCTTTTTTGTCACTTTTTGTATAAAATGCCAGGGTGATGGTTTTTCGAGG 1260
GACTGACACTAGTTTGGGCTAGAAAAAACAGTGAAAAACATATTTTACGGTCCCACTACCAAAAAGCTCC
P D C D Q T R S F L S L F V –
↓ApoI . yihW ⟶ . 1330
AATTTTCATGAGCCTTACCGAACTAACCGGTAACCCGCGGCACGACCAACTCCTCATGCTGATCGCCGAG
TTAAAAGTACTCGGAATGGCTTGATTGGCCATTGGGCGCCGTGCTGGTTGAGGAGTACGACTAGCGGCTC
M S L T E L T G N P R H D Q L L M L I A E
CGTGGGTATATGAATATTGATGAGCTGGCAAATCTGCTGGATGTCTCCACGCAGACGGTCCGCCGGGATA 1400
GCACCCATATACTTATAACTACTCGACCGTTTAGACGACCTACAGAGGTGCGTCTGCCAGGCGGCCCTAT
R G Y M N I D E L A N L L D V S T Q T V R R D
* * *

```
        .         .         .         .         .         .         .
TTCGTAAATTAAGCGAGCAAGGCCTGATTACGCGCCATCACGGTGGCGCGGGTCGGGCTTCCAGCGTCGT 1470
AAGCATTTAATTCGCTCGTTCCGGACTAATGCGCGGTAGTGCCACCGCGCCCAGCCCGAAGGTCGCAGCA
I  R  K  L  S  E  Q  G  L  I  T  R  H  H  G  G  A  G  R  A  S  S  V  V

        .         .         .         .         .         .         .
TAATACGGCGTTCGAGCAGCGTGAGGTTTCGCAAACCGAGGAAAAAAAAGCGATTGCCGAAGCGGTGGCA 1540
ATTATGCCGCAAGCTCGTCGCACTCCAAAGCGTTTGGCTCCTTTTTTTTCGCTAACGGCTTCGCCACCGT
  N  T  A  F  E  Q  R  E  V  S  Q  T  E  E  K  K  A  I  A  E  A  V  A

        .         .         .         .         .         . ↓SmaI   .
GACTATATTCCTGATGGTTCAACAATATTTATCACCATTGGTACGACTGTTGAGCATGTTGCCCGGGCGT 1610
CTGATATAAGGACTACCAAGTTGTTATAAATAGTGGTAACCATGCTGACAACTCGTACAACGGGCCCGCA
 D  Y  I  P  D  G  S  T  I  F  I  T  I  G  T  T  V  E  H  V  A  R  A
                                          *
        .         .         .         .         .         .         .
TACTTAACCATAATCATTTGCGGATAATCACCAACAGCCTGCGTGTGGCGCATATTCTTTACCACAACCC 1680
ATGAATTGGTATTAGTAAACGCCTATTAGTGGTTGTCGGACGCACACCGCGTATAAGAAATGGTGTTGGG
L  L  N  H  N  H  L  R  I  I  T  N  S  L  R  V  A  H  I  L  Y  H  N  P

        .         .         .         .         .         .         .
GCGCTTTGAAGTGATGGTGCCCGGCGGTACGTTGCGCTCTCATAATAGCGGGATCATTGGCCCTTCAGCG 1750
CGCGAAACTTCACTACCACGGGCCGCCATGCAACGCGAGAGTATTATCGCCCTAGTAACCGGGAAGTCGC
  R  F  E  V  M  V  P  G  G  T  L  R  S  H  N  S  G  I  I  G  P  S  A

        .         .         .         .         .         .         .
GCGTCCTTTGTGGCTGATTTTCGTGCTGATTATCTGGTAACAAGCGTTGGGGCGATTGAGAGCGATGGCG 1820
CGCAGGAAACACCGACTAAAAGCACGACTAATAGACCATTGTTCGCAACCCCGCTAACTCTCGCTACCGC
 A  S  F  V  A  D  F  R  A  D  Y  L  V  T  S  V  G  A  I  E  S  D  G

        .         .         .         .         .         .         .
CGTTGATGGAGTTTGATGTAAACGAAGCTAACGTGGTGAAAACGATGATGGCGCACGCGAGAAATATTCT 1890
GCAACTACCTCAAACTACATTTGCTTCGATTGCACCACTTTTGCTACTACCGCGTGCGCTCTTTATAAGA
A  L  M  E  F  D  V  N  E  A  N  V  V  K  T  M  M  A  H  A  R  N  I  L

        .         .         .         .         .         .         .
GCTGGTCGCCGATCACACTAAGTATCATGCTTCGGCGGCGGTTGAAATTGGTAACGTGGCACAGGTCACT 1960
CGACCAGCGGCTAGTGTGATTCATAGTACGAAGCCGCCGCCAACTTTAACCATTGCACCGTGTCCAGTGA
  L  V  A  D  H  T  K  Y  H  A  S  A  A  V  E  I  G  N  V  A  Q  V  T

        .         .         .         .         .         .         .
GCGCTCTTTACCGACGAGCTGCCGCCCGCTGCGCTAAAATCACGCTTACAAGACAGCCAAATTGAAATCA 2030
CGCGAGAAATGGCTGCTCGACGGCGGGCGACGCGATTTTAGTGCGAATGTTCTGTCGGTTTAACTTTAGT
 A  L  F  T  D  E  L  P  P  A  A  L  K  S  R  L  Q  D  S  Q  I  E  I

        .         .         .       ↓HpaI     .         .         .
TTCTTCCCCAGGAAGACGCGTAGATTTTTGTGACCGTTAACCTGGCTTCATACCTTGCCACATAGCCAAA 2100
AAGAAGGGGTCCTTCTGCGCATCTAAAAACACTGGCAATTGGACCGAAGTATGGAACGGTGTATCGGTTT
I  L  P  Q  E  D  A  –
```

```
         .          .          .          .          .          .          .
CCCATCCTTTCCCGCTACAGTTAATTTCTTGTGGCGCGAAAGGAGGCAAAAATGCTCTATATCTTTGATT 2170
GGGTAGGAAAGGGCGATGTCAATTAAAGAACACCGCGCTTTCCTCCGTTTTTACGAGATATAGAAACTAA
```

Figure 16. Nucleotide sequence of the *yihU–yihV* intergenic region, *yihV* and *yihW*. Sequence is from accession number AE000464 of the Entrez database available at NCBI. Open reading frames are labeled. *yihU* encodes a putative dehydrogenase. *yihV* encodes a putative carbohydrate kinase. *yihW* encodes a transcriptional repressor belonging to the DeoR family. Sites of R43A and T106I substitutions to create negative dominant and noninducible variants of YihW, respectively, are bold and indicated with an asterisk (*****). Also shown in bold and marked with an asterisk are the residues within the recognition helix that were substituted to create the recognition helix of GlpR (D35S and T38P). Restriction sites utilized for cloning are shown.

TABLE 10

Strains of *E. coli* used or constructed in the study of YihW

Strain[a]	Genotype	Derivation or reference
BL21(DE3)	*hsdS gal (λcIts857 ind1 Sam7nin5 lacUV5–T7gene1)*	(52)
MC4100	*F⁻ araD139 Δ(argF–lac)U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	(53)
NZ45	MC4100 ϕ(*glpK–lacZ*)hyb λp*lac*Mu *glpEGR*::Km^r *(λatt lacI^q tetR$^+$* Sp^r*) recA1 srl*::Tn*10*	(76)
GZ577	MC4100 *glpEGR*::Km^r *(λatt lacI^q tetR$^+$* Sp^r*) recA1 srl*::Tn*10*	(76)
KR101	GZ577 Φ($P_{K/R}$–*lacZ*)	λKR150 lysogen of GZ577
DH5α	*(φ80dlacZΔM15) endA1 recA1 hsdR17 supE44 thi1 gyrA relA1 Δ(lacZYA–argF)U169*	(54)
DH5α Z1	DH5α *(λatt lacI^q tetR$^+$* Sp^r*)*	(49)
KR102	DH5αZ1 Φ($P_{K/R}$–*lacZ*)	λKR150 lysogen of DH5αZ1
MG1655	wild–type isolate (F$^-$ λ$^-$)	(8)
TL504	MG1655 *Δ(lacZYA–argF)U169 zah–735*::Tn*10*	(55)
TL524	TL504 Tc^s	Tc^s selection[b]
KR2524	TL524 *yihW*::Km^r	[c]
CAG18636	(λ$^-$) *zih3088*::Tn*10kan rph*–1	(57)
CAG18495	(λ$^-$) *zih35*::Tn*10 rph*–1	(57,58)
CAG18601	(λ$^-$) *zih3166*::Tn*10kan rph*–1	(57,58)

[a] All strains are derivatives of *E. coli* K–12 except BL21(DE3) which is derived from *E. coli* B.

[b] The selection for tetracycline sensitive strains was as described by Maloy and Nunn (59). This strain was constructed by Ali Bhattacharya.

[c] The series of selections used for isolating the *yihW*::Km^r strain is as described in Section I.

TABLE 11

Plasmids used or constructed in the study of YihW

Plasmid	Insert Vector	Description
pCP15	(50)[b]	Ap^r, Km^r
pZE12–*luc*	(49)	P_{lac}–*luc* ColE1, Ap^r
pZE21–*luc*	(49)	P_{N25}–*luc* ColE1, Km^r
pZS24–*luc*	(49)	$P_{lac/ara}$–*luc* pSC101, Km^r
pSP417	(51)	Promoterless *lacZ* ColE1, Ap^r
pT7–7	(53)	T7 promoter, MCS, ColE1, Ap^r
pKO3	(67)	T^s ori, Cm^r
pGZ158	(76)	P_{lac}–*glpR* ColE1, Ap^r
pKR101	207118/183282 (PCR)[a], *Eco*RI/*Xba*I pZE12–*luc*, *Xba*I/*Eco*RI	P_{lac}–*yihW* ColE1, Ap^r
pKR201	pKR101, *Eco*RI/*Xba*I pZE21–*luc*, *Xba*I/*Eco*RI	P_{N25}–*yihW* ColE1, Km^r
pKR301	pKR201, *Xho*I/*Avr*II pZS24–*luc*, *Avr*II/*Xho*I	P_{N25}–*yihW* pSC101, Km^r
pKR103	M113, *Eco*RI/*Xba*I pZE12–*luc*, *Xba*I/*Eco*RI	P_{lac}–*yihW* (T38P) ColE1, Ap^r
pKR104	M114, *Eco*RI/*Xba*I pZE12–*luc*, *Xba*I/*Eco*RI	P_{lac}–*yihW* (D35S, T38P) ColE1, Ap^r
pKR105	M115, *Eco*RI/*Xba*I pZE12–*luc*, *Xba*I/*Eco*RI	P_{lac}–*yihW* (T106I) ColE1, Ap^r
pKR205	pKR105, *Eco*RI/*Xba*I pZE21–*luc*, *Xba*I/*Eco*RI	P_{N25}– *yihW* (T106I) ColE1, Km^r
pKR305	pKR205, *Xho*I/*Avr*II pZS24–*luc*, *Avr*II/*Xho*I	P_{N25}– *yihW* (T106I) pSC101, Km^r
pKR106	M116, *Eco*RI/*Xba*I pZE12–*luc*, *Xba*I/*Eco*RI	P_{lac}– *yihW* (R43A) ColE1, Ap^r
pKR206	pKR106, *Eco*RI/*Xba*I pZE21–*luc*, *Xba*I/*Eco*RI	P_{N25}– *yihW* (R43A) ColE1, Km^r

Plasmid	Insert Vector	Description
pKR306	pKR206, *Xho*I/*Avr*II	P_{N25}– *yihW* (R43A)
	pZS24–*luc*, *Avr*II/*Xho*I	pSC101, Km^r
pKR114	069886/183282 (PCR), *Eco*RI/*Xba*I	P_{lac}–*yihVW*
	pZE12–*luc*, *Xba*I/*Eco*RI	ColE1, Ap^r
pKR115	*Eco*RI/*Mfe*I pKR114	pKR114 – *Eco*RI/*Mfe*I
pKR116	pCP15, *Eco*RI/*Hin*dIII(filled–in)	*yihW*::Km^r , ColE1
	pKR115, *Sma*I/*Apo*I	Km^r, Ap^r
pKR117	pKR116, *Hpa*I/*Hpa*I	*yihW*::Km^r, T^s ori
	pKO3, *Sma*I	Km^r, Cm^r
pKR150	pKR114, *Eco*RI/*Bam*HI	$P_{K/R}$–*lacZ*
	pSP417, *Bam*HI/*Eco*RI	ColE1, Ap^r
pKR151	pKR150, *Eco*RI/*Mfe*I	P_K–*lacZ*
	pSP417, *Eco*RI	ColE1, Ap^r
pKR152	pKR150, *Mfe*I/*Eco*RI	P_{dH}–*lacZ*
	pSP417, *Eco*RI	ColE1, Ap^r
pKR176	336908/183282 (PCR), *Nde*I	P_{T7}–*yihW*
	pT7–7, *Nde*I/*Sma*I	ColE1, Ap^r
pKR801	069886/240410 (PCR), *Eco*RI/*Xba*I	P_{N25}–*yihV*
	pZE21–*luc*, *Xba*I/*Eco*RI	ColE1, Km^r
pKR802	*Eco*RI/*Mfe*I pKR801	P_{N25}–*yihV*
		ColE1, Km^r

[a] For previously constructed plasmids, instead of giving insert and vector fragments, references are given.

[b] Chromosomal DNA isolated from strain MG1655 was used as template for these PCR reactions.

Kinase assays. Two types of assays were used to attempt to identify the substrate for the putative kinase encoded by *yihV*. Both assays utilized sonicated cells overexpressing the protein YihV using pKR802 in DH5αZ1. (a) A coupled enzyme assay utilizing pyruvate kinase and lactate dehydrogenase was used to detect ADP in the presence of ATP and various possible substrates (98). The buffer used was 50 mM potassium phosphate (pH 7.5), 8 mM $MgCl_2$ and 2 mM DTT. The assay is based on the quantitation of NADH oxidation, and activity is quantitated by measuring the decrease in absorbance at 340 nm. (b) Possible substrates were also tested by incubation of sonicated cells overexpressing YihV with [γ–^{33}P]ATP using the same buffer system utilized for (a). After incubation at 37°C, thin layer chromatography in a system containing *n*–butanol:formic acid:water (33:50:17) was used to separate radiolabeled product from ATP.

Use of Biolog plates. The 96–well microtiter indicator plates commercially available from Biolog, Inc., were utilized to screen bacteria for the ability to utilize a wide variety of carbon sources (99). The ability of a bacterial culture innoculated into the wells of these plates to oxidize a compound is indicated by formation of a colored tetrazolium dye. In this study, the ES plates, designed to distinguish *E. coli* from *Salmonella typhimurium*, were used following the procedure outlined by the manufacturer.

Results

Promoters in *yih* gene cluster are negatively regulated by YihW. Using the promoter–probe vector pSP417, described in Section I, two or possibly three transcriptional promoters were identified in the *yih* gene cluster (Table 12 and Figure 15). The two promoters identified are specific for *yihU* (P_{dH}) and *yihV* (P_K). The 3' end of the fragment containing P_K is located 50 nucleotides upstream of the *yihV* start codon and part of this promoter may not be included in this fragment. A longer fragment ($P_{K/R}$) containing most of the *yihV* coding region conferred over 15 times more promoter activity. This may be due to the the longer fragment possessing the complete sequence for the promoter specific for *yihV* or to the presence of a third promoter within the *yihV* coding region that is specific for *yihW*. It was found that coexpression of YihW from a low–copy number plasmid (pKR301) dramatically reduced the activity of these promoters (Table 12) indicating that these promoters are negatively regulated by YihW.

TABLE 12

Repression of promoters in the *yih* cluster by YihW[a]

Additional plasmid	Transcriptional fusion		
	P_K–*lacZ* (pKR151)	$P_{K/R}$–*lacZ* (pKR150)	P_{dH}–*lacZ* (pKR152)
None	1300	20700	19000
pKR301 (*yihW*)	10	500	40

[a] Cells growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine, 0.2% casamino acids and 0.2% maltose and the appropriate antibiotics. Expression of YihW in cells containing pKR301 was induced by the addition of 100 ng/ml tetracycline.

Phenotypic testing of variants of YihW. Plasmids were constructed that express YihW or several variants of the protein. One variant (R43A) was expected to lack the ability to bind to operator DNA and another was expected to be noninducible (T106I), based on previous studies with GlpR, LacR and the work described above with AgaR. Strain DH5αZ1 harboring pKR150 ($P_{K/R}$–*lacZ*) containing a low–copy number compatible plasmid expressing either YihW, YihW (R43A) or YihW (T106I) or a vector control was grown in minimal A and B salts medium containing 0.2% maltose and appropriate antibiotics. Expression of repressor variants was induced by the addition of 100 ng/ml tetracycline. Cells expressing either YihW or YihW (T106I) were found to have approximately 600

Miller units of β–galactosidase activity while cells not expressing a variant of YihW or those expressing YihW (R43A) had approximately 14,000 Miller units. This indicated that YihW (R43A) could not bind DNA, as shown for the corresponding substitution in AgaR (see Section I). Both YihW and YihW (T106I), however, could efficiently repress transcription from $P_{K/R}$ located on the multi–copy plasmid. It should be noted that overexpression of YihW or variants could not be verified using SDS–PAGE using any of the plasmids described in this Section, but these differences in phenotype at least confirm that YihW and YihW (T106I) were being expressed from pKR301 and pKR305, respectively.

Attempts to determine metabolic pathway encoded by *yih* cluster. Strain GZ570 (76) or DH5αZ1 (49) containing plasmids expressing wild–type YihW or the two variants described above (pKR101, pKR105 and pKR106) were screened for utilization of a number of carbon sources. The goal was to identify the substrates whose metabolism is regulated by YihW. It was anticipated that the noninducible variant may confer a growth–negative phenotype for a particular compound, while the negatively dominant variant may allow constitutive utilization or decryptify a system that is strongly repressed or noninducible. Because of the similarity between YihW and GlpR, initial experiments focused on three–carbon compounds including glycerol, dihydoxyacetone, hydroxyacetone and 1,2–propanediol. No reproducible phenotypic differences were observed using these compounds.

The 96–well microtiter indicator plates commercially available from Biolog, Inc., provide a means to screen many other carbon sources (99). The wells of these plates contain an array of compounds utilized by various bacterial species. The ability of a bacterial strain to oxidize a compound is indicated by formation of a colored tetrazolium dye. These plates were found to effectively distinguish between strains expressing inducible versus noninducible variants of GlpR. The plates were used to screen for differences in utilization of various compounds by strain GZ570 expressing wild–type YihW, the noninducible variant of YihW or the variant lacking the ability to bind DNA. Since YihW functions in a negative fashion to control transcription, it was anticipated that one or more wells would indicate efficient oxidation of the substrate when the negatively dominant variant of YihW was expressed, while the corresponding well would remain colorless when the noninducible variant was expressed. However, no reproducible difference in the ability to oxidize the carbon sources within the plates was observed for cells expressing the noninducible versus the negatively dominant variant of YihW. It is highly likely that the carbon source metabolized by the *yih* gene cluster was not present in the plates.

Another attempt to determine the pathway encoded by this gene cluster involved the overexpression of the putative kinase encoded by *yihV*. It was found that plasmid pKR802 but not pKR801 overexpresses YihV. pKR802 is derived from pKR801 but the fragment upstream of the *Mfe*I site has been removed (see Figure 16). The putative kinase, when overexpressed using pKR802, constituted a significant portion of the total protein, however it was found that essentially all of the protein was located within insoluble inclusion bodies. Despite this, two different kinase assays described in Experimental Procedures were used to attempt to determine the substrate for YihV. Using the coupled–enzyme assay (pyruvate kinase – lactate dehydrogenase), sonicated cells from cultures overexpressing YihV were

compared to those not overexpressing the protein. The significant background level of NADH oxidation detected using cell extracts not overexpressing the kinase and the likely fact that only a very small percentage of the overexpressed kinase was active, prevented the conclusive determination of the substrate for the kinase. However, there was a small but duplicatable increase in fructose–6–phosphate kinase activity in cell extracts overexpressing the kinase. No significant increase in kinase activity was observed using ribose, fructose or tagatose as substrates. However, no increase in fructose–6–phosphate kinase activity could be detected in cells overexpressing YihV using the radioactive kinase assay described in Experimental Procedures. This assay was, however, able to detect the phosphorylation of glycerol by glycerol kinase (GlpK).

YihW is not an isorepressor of GlpR. As mentioned previously, one of the most striking features of the protein encoded by *yihW* is its high degree of sequence similarity to GlpR. For this reason, the possibility that YihW is an isoform of GlpR was investigated. Despite the high degree of similarity between the two proteins, especially in the helix–turn–helix region (see below), YihW did not repress a transcriptional fusion of a GlpR–regulated promoter (P_{glpK}) to *lacZ* (Table 13). It was also found that a variant of YihW with two amino acid substitutions in the recognition helix to more closely match the recognition helix of GlpR (YihW (D35S, T38P)) only exhibited 2–fold repression of this same transcriptional fusion. The recognition helices of GlpR, YihW and the variant are:

GlpR	S V S P Q T I R R D L
YihW (D35S, T38P)	**S** V S **P** Q T V R R D I
	↑ ↑
YihW	D V S T Q T V R R D I

TABLE 13

YihW is not an isoform of GlpR[a]

Plasmid	Description	β–galactosidase activity
None	——	450 ± 27
pZE12–*luc*	Vector	510 ± 30
pKR101	YihW	500 ± 25
pKR104	YihW (D35S, T38P)	270 ± 20
pGZ158	GlpR	<1

[a] Cultures of NZ45 (*ΔglpR* ϕ(*glpK–lacZ*)) harboring the plasmids indicated above growing logarithmically were harvested and assayed for β–galactosidase specific activity. Cells were grown in minimal A and B salts medium containing 2 μg/ml thiamine, 0.2% casamino acids and 0.2% maltose. Cultures of cells containing a plasmid were supplemented with 100 μg/ml ampicillin. Expression plasmid–encoded proteins was induced by the addition of 0.5 mM IPTG.

Discussion

In this study, YihW was shown to not be an isoform of GlpR. It does, however, negatively regulate promoters within the *yih* gene cluster. The metabolic pathway encoded by the *yih* cluster has not been determined. However, this preliminary work should facilitate the future screening of possible carbon sources. This may be accomplished either by looking for growth phenotypes in cells expressing negatively dominant YihW versus the noninducible variant or looking for induction of one of the transcriptional promoters identified. Optimization of the overexpression of the kinase, YihV, may also allow elucidation of this enzyme's substrate.

The genes within the *yih* cluster encode proteins having sequence similarity to permeases or symporters (*yihO* and *yihP*), α–glycosidases (*yihQ*), 1 and 4–epimerases (*yihR*), phosphomannose isomerases (*yihS*), aldolases (*yihT*), dehydrogenases (*yihU*) and sugar kinases, including ribokinase from *E. coli* and fructokinase from *Klebsiella pneumoniae* (*yihV*) (BLAST search results, (11)). The dehydrogenase encoded by *yihU* is 36% identical to β–hydroxyisobutyrate dehydrogenase (MmsB) from *P. aeruginosa* (100), an enzyme involved in the catabolism of branched–chain amino acids (101). This pathway for the degradation of isoleucine and valine is not known to exist in *E. coli* (102). However, MmsB is only one of a relatively large family of β–hydroxyacid dehydrogenases (BLAST search results, (11)).

Two separate reports from Wu et al. describe a mutant of *E. coli* that had gained the ability to utilize D–lyxose (95,96). The strain was found to constitutively express a D–mannose isomerase. The gene encoding this enzyme was mapped (95) and the protein purified and characterized (96). On the current *E. coli* linkage map, the location of this locus (*manC*, *mni*) is given as 87.5 minutes (103). The protein described by Wu et al. (96) had a similar subunit size and amino acid composition to the one that is encoded by *yihS*. Given this information and the fact that the protein encoded by *yihS* displays sequence similarity to phosphomannose isomerases, it is reasonable to assume that *yihS* encodes the protein described by Wu et al. (96). This enzyme was found to convert D–mannose to D–fructose and D–lyxose to D–xylulose. The K_m values determined for D–mannose and D–lyxose were 80 mM and 300 mM, respectively (96). Surprisingly, D–mannose–6–phosphate was not a substrate for this enzyme. Other sugars tested that could not serve as substrates were: D– and L–arabinose, D– and L–fucose, D–galactose, D–galacturonate, D–glucuronate, D– and L–glucose, D–glucose–6–phosphate, L–lyxose, L–mannose, L–rhamnose, D– and L–xylose, D–talose, D–ribose, D–mannose–1–phosphate and *N*–acetylmannosamine (F. Stevens, personal communication).

As mentioned previously, the long–term objective of this research is to determine the functions of all members of the DeoR family in *E. coli*. Important for this objective is the ability to construct plasmids that express either a negatively dominant or noninducible variant of YihW and other members of the DeoR with no defined function. Plasmids expressing variants of YihW containing substitutions R43A and T106I have been constructed and are expected to result in a negatively dominant or noninducible phenotype, respectively. Preliminary results from this study suggest that YihW (R43A) is impaired in DNA operator binding as is expected for a negatively dominant variant. The substitutions R43A and T106I

in YihW are analogous to substitutions previously shown in GlpR and AgaR to give rise to negatively dominant and noninducible variants, respectively (6,24) (see also Section I). Phenotypic screening combined with microarray technology to determine changes in mRNA levels in strains expressing the negatively dominant versus noninducible variants of YihW should allow for the identification of all genes regulated by YihW as well as the metabolic pathway controlled by this repressor.

The genes within the *yih* cluster encode proteins that would probably be necessary to metabolize a disaccharide since one enzyme (YihQ) appears to be a glycosidase. However, the presence of a gene encoding a β–hydroxyacid dehydrogenase suggests the disaccharide may contain an acid–derivative of a sugar. The other genes within the cluster encode proteins having sequence similarity to enzymes that function in the metabolism of various sugars.

SECTION III:
Characterization of a 12–kDa Rhodanese Encoded by *glpE*

Introduction

Genes of known function belonging to the *glp* regulon of *Escherichia coli* encode proteins that are responsible for the metabolism of *sn*–glycerol 3–phosphate (glycerol–P) and its precursors, glycerol and glycerophosphodiesters (26,104). Figure 17 illustrates the metabolic pathway of glycerol–P and its precursors. The genes comprising this regulon belong to five operons. Transcription of all but the *glpEGR* operon is negatively regulated by the *glp* repressor, GlpR, a member of the DeoR family of transcriptional regulators (26,104).

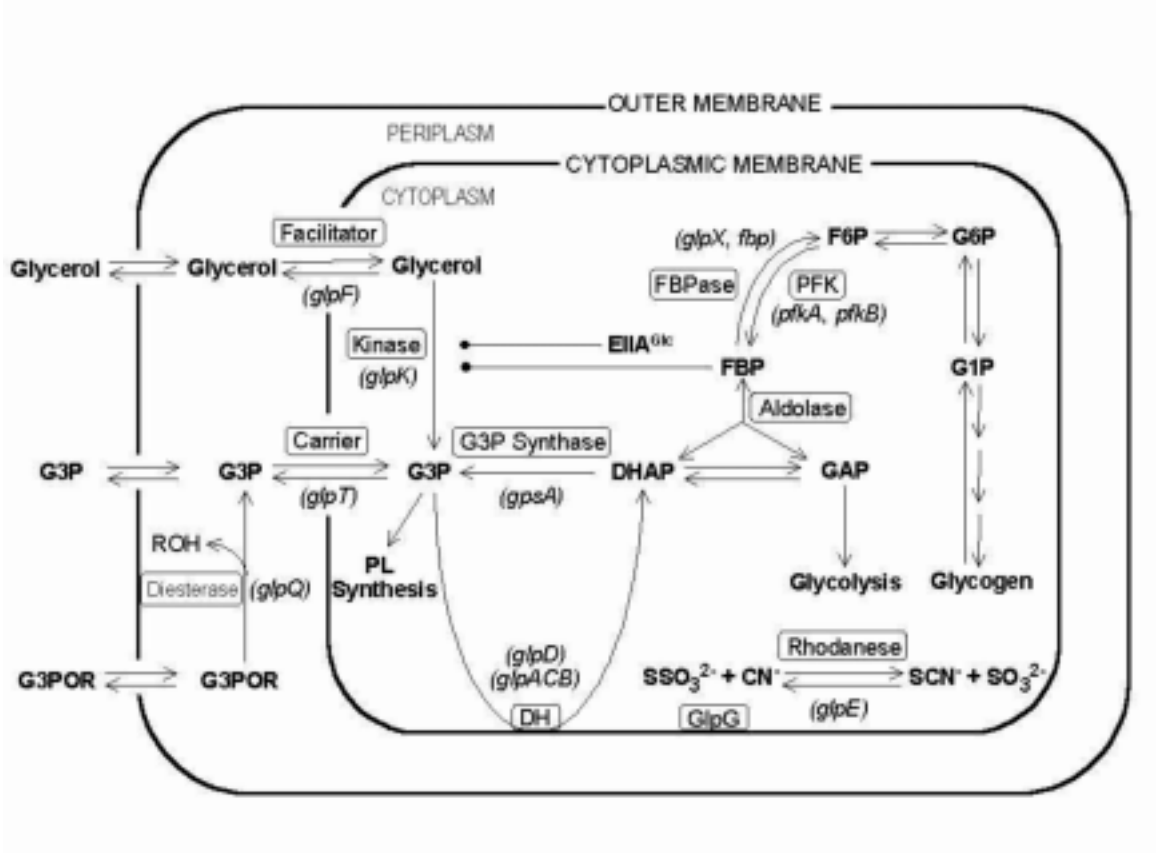


Figure 17. Glycerol–3–phosphate metabolism in *E. coli*. Figure courtesy of Janet Donahue. Abbreviations used in this figure are: G3P, glycerol–3–phosphate; G3POR, glycerophosphodiesters; PL, phospholipid; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde–3–phosphate; FBP, fructose–1,6–bisphosphate; F6P, fructose–6–phosphate; G6P, glucose–6–phosphate; G1P, glucose–1–phosphate; PFK, phosphofructokinase; $EIIA^{Glc}$, phosphotransferase system enzyme IIA specific for glucose; *gpsA*, gene encoding glycerol–3–phosphate synthase.

Operon *glpACB*, encoding the subunits of the anaerobic glycerol–P dehydrogenase is located near minute 51. Divergently transcribed from *glpACB* is *glpTQ*. The genes *glpT* and *glpQ* encode a glycerol–P permease and a periplasmic glycerophosphodiesterase, respectively. The regulation of these divergent operons by GlpR has been previously characterized. Repression involves widely separated operators recognized by GlpR. Tetrameric GlpR when bound to these operators is believed to result in looping of the DNA around the transcription initiation sites for these operonsand thus prevent the interaction of RNA polymerase with the transcription initiation sites (32,33,105). The *glpFKX* operon, encoding a glycerol diffusion facilitator, glycerol kinase and a fructose–1,6–bisphosphatase, respectively, maps to minute 89 of the *E. coli* genome. Regulation of this operon by GlpR has also been characterized (81).

The genes *glpE* and *glpG* together with the gene encoding the repressor, *glpR*, form a complex operon at minute 77 divergently transcribed from *glpD* (60). *glpD* encodes the aerobic glycerol–P dehydrogenase and regulation of its expression by GlpR also appears to involve the formation of a DNA loop (31). Expression of *glpEGR*, however, is not regulated by GlpR (32). The functions of the proteins encoded by *glpE* and *glpG* and their roles in glycerol–P metabolism have yet to be determined. However, the gene products have been partially characterized (6,60) and a variant of GlpE, the gene product of *glpE* containing a hexahistidine tag, has been purified (76).

GlpE exhibits sequence similarity to thiosulfate:cyanide sulfurtransferases (EC 2.8.1.1), enzymes that have traditionally been called rhodaneses (106). These enzymes catalyze the transfer of the sulfane sulfur from thiosulfate to cyanide forming thiocyanate and sulfite:

$$SSO_3^{2-} + CN^- \rightarrow SO_3^{2-} + SCN^-$$

The physiological role of rhodaneses is still in question. Proposed roles include cyanide detoxification (107), sulfur metabolism (108) and mobilization of sulfur for iron–sulfur cluster biosynthesis (109–111).

The well–characterized rhodanese from bovine liver is approximately twice the molecular weight of GlpE and functions as a monomer. The bovine liver enzyme, however, contains two structurally similar domains, though the domains have only limited sequence similarity to each other (112). An active site cysteine residue, present in only the C–terminal domain, functions in forming a stable enzyme–sulfur intermediate wherein the sulfane sulfur from thiosulfate forms a persulfide linkage with the cysteine sulfur. In the second step of the reaction, a thiophilic acceptor substrate accepts the sulfur reforming the sulfur–free form of the enzyme. Catalysis thus occurs utilizing a double–displacement mechanism (113).

In addition to the bovine liver enzyme, other rhodanese–like proteins have been described from both eukaryotic and prokaryotic sources, some containing two domains like the bovine liver enzyme and others containing only a single domain. Rhodaneses display sequence similarity to mercaptopyruvate sulfurtransferases and together the two enzymes appear to constitute a sulfurtransferase family of proteins (114). A variant of the rat rhodanese in which the arginine and lysine residues immediately following the active–site cysteine were replaced with a glycine and serine as found in the rat mercaptopyruvate sulfurtransferase was constructed. This variant was found to have lost most of its rhodanese

activity but possessed significantly higher mercaptopyruvate sulfurtransferase activity than the wild–type rat rhodanese (115). Conversely, it has also been shown that variants of the rat mercaptopyruvate sulfurtransferase that have amino acid substitution(s) to match more closely the sequence of the rat rhodanese have increased rhodanese activity and decreased mercaptopyruvate sulfurtransferase activity (115). These results are consistent with the finding that the lysine (K249) following the active–site cysteine as well as an arginine upstream of the active–site (R186) are important for thiosulfate binding and recognition in the bovine liver rhodanese (116). These residues have also been shown to mediate the anion–specific inhibition observed in the bovine liver rhodanese (117).

Figure 18 illustrates an alignment of the active sites of several members of this family of sulfurtransferases with the putative active site of GlpE. Shown are both the carboxy and amino terminal domains of the well–characterized bovine liver rhodanese as well as the two domains of the rat rhodanese and mercaptopyruvate sulfurtransferase.

(a)

```
GlpE   G A F M R D N D F D T P V M V M C Y H G N S S K G A A Q Y L L Q Q  81
PspE   R I A T A V P D K N D T V K V Y C N A G R Q S G Q A K E I L S E M  83
```

(b)

```
RhdB   M F E A K K V D L T K P L I A T C R K G - - - V T A C H I A L A A 260
RhdR   I F Q D K K V D L S Q P L I A T C R K G - - - V T A C H I A L A A 258
MPST   L F Q E K K V D L S K P L V A T C G S G - - - V T A C H V V L G A 261
SseA   I F F G R G V S Y D K P I I V S C G S G - - - V T A A V V L L A L 304
CysA   V Y G E A G L D T D K D T I A Y C R I G E R S S H T W F V L R E L 249
RhdA1  I L E A K G V T P D K D V I V T C S T G R E A S L Q Y L V L K H L 290
RhdA2  R L E E L G I T P D K E I V T H C Q T H H R S G L T Y L I A K A L 246
```

(c)

```
RhdB   Y V G S L G I S N D T H V V V Y D G D D L G S F Y A P R V W W M F 116
RhdR   Y V G N L G I S N D T H V V V Y D G D D L G S F Y A P R V W W M F 114
MPST   Y A G S L G V S A A T H V V I Y D G S D Q G L Y S A P R V W W M F 117
SseA   A M R E L G V N Q D K H L I V Y D E G N L - - F S A P R A W W M L 161
CysA   L L S A K G I G N D D T V I L Y G G N N - - N W F A A Y A Y W Y F  98
RhdA1  I E G H L P G A V N I A D A A Y R G P N - - G F L P V Q I W D P E 103
RhdA2  L F G E L G H R P E A V Y V V Y D D E G G G - - W A G R F I W L L 102
```

Figure 18. Sequence alignment of the putative active site of GlpE with that of other members of the sulfurtransferase family. The active site cysteine or the corresponding residue in the N–terminus of the two–domain proteins is highlighted in red. The lysine important for binding of thiosulfate and anion inhibitors of the bovine liver rhodanese is highlighted in blue. The residues that play a role in substrate specificity for thiosulfate versus mercaptopyruvate in the rat enzymes are highlighted in violet and green, respectively. (a) Two of the single–domain members of the sulfurtransferase family, GlpE and PspE, a phage–shock protein from *E. coli* (118). (b) C–terminal domains of several two–domain members of the sulfurtransferase family. RhdB, bovine liver rhodanese (112); RhdR, rat rhodanese (123); MPST, rat mercaptopyruvate sulfurtransferase (115). The remaining sequences shown are rhodanese–like proteins from various bacterial sources. SseA, *E. coli* (119); CysA, *S. erythraea* (121); RhdA1, *Synechococcus* PCC7942 (122); RhdA2, *A. vinelandii* (120). (c) N–terminal domains of the same two–domain members shown in (b).

Also included are several bacterial rhodanese–like proteins. PspE, like GlpE, is a single–domain rhodanese–like protein found in *E. coli*. The gene encoding PspE is a member of an operon that is induced by phage–shock (118). SseA is a two–domain rhodanese–like protein from *E. coli*. Overexpression of SseA results in enhanced serine–sensitivity (119). CysA, RhdA1 and RhdA2 correspond to two–domain rhodanese–like proteins from *Saccharopolyspora erythraea*, *Synechococcus* sp. strain PCC 7942 and *Azotobacter vinelandii* (120), respectively. Disruption of the gene encoding CysA in *S. erythraea* results in cysteine auxotrophy (121). Expression of RhdA in *Synechococcus* is induced by sulfur–starvation and the protein is believed to be located in the periplasmic space (122).

Rhodaneses and mercaptopyruvate sulfurtransferases are also distantly related to tyrosine and dual–specificity phosphatases such as the human Cdc25a phosphatase. These relationships had been proposed based on a combination of structural and sequence comparisons (106). Comparison of the crystal structure of bovine liver rhodanese to that of the mammalian Cdc25a phosphatase has confirmed this relationship. The active site of the members of this superfamily of transfer proteins contains a cysteine which during the course of the reaction becomes modified forming a persulfide in the sulfurtransferases or a phosphocysteine in the phosphatases. It has also been shown that two structural motifs, designated CH2A and CH2B, located on either side of the active site, are conserved in this superfamily, even in domains which no longer possess the active–site cysteine. The role of these motifs is unclear but it is suggested that they function, at least in the phosphatases, in protein–protein interactions mediating substrate recognition (124).

Two rhodaneses and a mercaptopyruvate sulfurtransferase have been previously described in *E. coli*, but the genes for these enzymes have not been mapped nor has a sequence for the proteins been determined (125–127). The most well–characterized of the two rhodaneses is "accessible" to the substrates thiosulfate and cyanide even in intact cells and is released from cells by freeze–thaw treatment. Based on these results it was suggested this rhodanese is located in the periplasmic space. The enzyme was purified and characterized. It exhibits many similarities to the well–characterized rhodanese from bovine liver but was found to be only about half the molecular weight (126).

For this study a new purification scheme for GlpE, one without the use of a hexahistidine tag, was developed. GlpE has been shown to possess rhodanese activity and has been kinetically characterized and compared to both the bovine liver rhodanese and the "accessible" rhodanese previously described in *E. coli*. It has also been shown that GlpE is cytoplasmic though most methods to isolate periplasmic proteins result in the release of GlpE from intact cells. This finding indicates that GlpE is a member of a previously described group D class of proteins that seem to be uniquely associated with the inner membrane (128–130). This group of proteins includes thioredoxin which has been previously shown to act as a sulfur–acceptor substrate for the bovine liver rhodanese (131). In this study I show that thioredoxin can serve as a sulfur–acceptor substrate for GlpE with much higher affinity than cyanide. These findings are considered in a discussion of the possible physiological role of GlpE in light of the fact that both thioredoxin and rhodaneses have been implicated in iron–sulfur cluster biosynthesis.

Experimental Procedures

Materials and reagents. Unless explicitly listed below, the reagents used for the study of GlpE were purchased from either Sigma Chemical Company or Fisher Scientific. Synthetic oligonucleotides were prepared using an Applied Biosystems DNA Synthesizer (model 381A) from phosphoramidites supplied by Cruachem. Oligonucleotides were purified using cartridges also supplied by Cruachem as described by the manufacturer. Reagents for preparing growth media not supplied by either Fisher or Sigma were purchased from Difco. Restriction endonucleases and reagents for PCR and cloning were supplied by New England Biolabs. Ammonium thiosulfate and DTNB were purchased from Aldrich. Ferric nitrate and the sodium salt of mercaptopyruvate were purchased from ICN. Thioredoxin from *E. coli* was supplied by Promega. Thioredoxin reductase purified from *E. coli* was a generous gift from C. H. Williams, Jr., at the Department of Veterans Affairs, Department of Biological Chemistry, Ann Arbor, Michigan (132). NifS from *Azotobacter vinelandii* was kindly supplied by D. R. Dean at the Department of Biochemistry–Fralin Biotechnology Center, Virginia Tech, Blacksburg, Virginia (133).

Bacterial strains and plasmids. The bacterial strains and plasmids used or constructed for the study of GlpE are listed in Tables 14 and 15, respectively. BL21(DE3) harboring pGZ105, a previously described derivative of pT7–7, was used to overexpress GlpE for purification. Plasmids pGZ154 and pGZ132 were used to overexpress GlpE in DH5αZ1, GZ570 and KR1570. DH5αZ1 and GZ570 are derivatives of DH5α and MC4100, respectively, that contain a spectinomycin–resistance cassette carrying $lacI^q$ and *tetR* inserted at the λ*att* site allowing transcriptional control of genes expressed from derivatives of the pZ family of expression plasmids (49).

Plasmids pGZ154 and pGZ132 contain *glpE* controlled by the tetracycline–inducible P_{N25} transcriptional promoter. The *glpE* gene was amplified from pGZ105 by PCR using the primers 5'–acgAAttcccGctagCaat–3' and 5'–tcactagtttgacagcttatc–3', where uppercase letters indicate mismatches used to create restriction sites. The amplified product was cloned into the *Eco*RI and *Xba*I sites of either pZA31–*luc* to create pGZ154 or pZE21–*luc* to create pGZ132. The plasmid pGZ154 has a p15A origin of replication and confers chloramphenicol resistance and pGZ132 has a ColE1 origin of replication and confers kanamycin resistance (49). Standard molecular biological techniques were used for cloning and preparation and transformation of competent *E. coli* cells (62,68). The plasmid $pATCBO_A2+6$ was used for constitutive expression of the periplasmic marker alkaline phosphatase. pSH60 was used for overexpression of GlpD.

TL524 is a tetracycline–sensitive derivative of TL504 that was obtained by selection on plates containing chlortetracycline hydrochloride and fusaric acid as described previously (59). TL504 is a *ΔlacZ* derivative of MG1655 described previously (55). BP331 is a single λ lysogen of TL524 containing a *glpE'–lacZ* transcriptional fusion derived from pBY331 with 156 bp of the *glpE* control region (250 to 406) (32). The methods for preparing the λ derivative containing the *glpE'–lacZ* transcriptional fusion and the subsequent selection for lysogens were as described previously (31,60).

TABLE 14

Strains of *E. coli* used or constructed in the study of GlpE

Strain[a]	Genotype	Derivation or reference
BL21(DE3)	*hsdS gal (λcIts857 ind1 Sam7nin5 lacUV5–T7gene1)*	(52)
GZ544[b]	BL21(DE3) *glpEG*::Km^r	P1(DA142) x BL21(DE3) Km^r selection
MC4100	F$^-$ *araD139 Δ(argF–lac)U169 rpsL150 deoC1 relA1 rbsR ptsF25 flbB5301*	(53)
NZ41	MC4100 *recB21 recC22 sbcB15 glpEGR*::Km^r	(134)
GZ570[b]	MC4100 (*λatt lacI^q tetR*$^+$ Sp^r)	P1(DH5αZ1) x MC4100 Sp^r selection
DA142	MC4100 *glpEG*::Km^r *malT*::Tn*10*	(135)
DH5α	*(φ80dlacZΔM15) endA1 recA1 hsdR17 supE44 thi1 gyrA relA1 Δ(lacZYA–argF)U169*	(54)
DH5α Z1	DH5α (*λatt lacI^q tetR*$^+$ Sp^r)	(49)
MG1655	*E. coli* K–12 wild–type isolate (F$^-$ λ$^-$)	(8)
TL504	MG1655 *Δ(lacZYA–argF)U169 zah–735*::Tn*10*	(55)
TL524[c]	TL504 Tc^s	Tc^s selection
BP331[d]	TL524 Φ(*glpE'–lacZ*)	λBP331 lysogen of TL524
BP1331[d]	BP331 *glpEGR*::Km^r	P1(NZ41) x BP331 Km^r selection
BP342[d]	TL524 Φ(*glpG'–lacZ*)	λBP342 lysogen of TL524
BP1342[d]	BP342 *glpEGR*::Km^r	P1(NZ41) x BP342 Km^r selection
AW405	*gal–1 gal–2 his–4 ara lac xyl thr leu thi rpsL* resistant to bacteriophages T1, T5 and T6	(136)

Strain[a]	Genotype	Derivation or reference
PB104	AW405 *mscL*::Cm^r *recA*::Tn*10*	(137)
KR1570	GZ570 *mscL*::Cm^r	P1(PB104) x GZ570 Cm^r selection

[a] All strains are derivatives of *E. coli* K–12 except BL21(DE3) and derivatives which are derived from *E. coli* B.

[b] These strains were constructed by Gang Zeng.

[c] The selection for tetracycline sensitive strains is as described by Maloy and Nunn. This strain was constructed by Ali Bhattacharya.

[d] These strains were constructed by Ben Potters.

TABLE 15

Plasmids used or constructed in the study of GlpE

Plasmid	Description	Derivation or reference
pT7–7	T7 promoter and strong RBS upstream of a MCS ColE1 ori Ap^r	(66)
pGZ105	*glpE* in *Nde*I/*Hin*dIII of pT7–7	(6)
pZE21–*luc*	ColE1 ori Km^r P_{N25}–*luc*	(49)
pZA31–*luc*	p15A ori Cm^r P_{N25}–*luc*	(49)
pGZ154	*glpE* in *Eco*RI/*Xba*I of pZA31–*luc*	*glpE* from pGZ105 into *Eco*RI/*Xba*I of pZA31–*luc*[a]
pGZ132	*glpE* in *Eco*RI/*Xba*I of pZE21–*luc*	*glpE* from pGZ105 into *Eco*RI/*Xba*I of pZE21–*luc*[a]
$pATCBO_A2+6$	ϕ(*glpT'–phoA*) ColE1 ori Ap^r	(33)
pSH60	*glpD* P15A ori Ap^r	(138)
pB10b	$lacI^q$ Ap^r P_{lacUV5} pBR322 ori	(137)
p522b	pB10b with P_{lacUV5}–*mscL*	(137)
pSP417	promoterless *lacZ* Ap^r ColE1 ori	(51)
pBY342	*glpG* promoter region in *Eco*RI/*Xba*I of pSP417	(60)
pBY331	*glpE* promoter region in *Eco*RI/*Xba*I of pSP417	(60)

[a] See text for PCR primers used to amplify *glpE* from pGZ105 and introduce *Eco*RI and *Xba*I restriction sites. These plasmids were constructed by Gang Zeng.

BP1331 is a derivative of BP331 that carries a kanamycin–resistance cassette introduced by P1 transduction from NZ41 that results in a *glpEGR* deletion (24). GZ544 is a derivative of BL21(DE3) that carries a kanamycin–resistance cassette introduced by P1 transduction from DA142 that results in a *glpEG* deletion. KR1570 is a derivative of GZ570 that carries a chloramphenicol–resistance cassette introduced by P1 transduction from PB104 that results in an *mscL* deletion. P1 transduction was performed as described previously (62).

Growth media and conditions. Cultures overexpressing GlpE were grown in Luria–Bertani (LB) broth (62). Antibiotics were included at 100 μg/ml ampicillin, 50 μg/ml kanamycin, 25 μg/ml chloramphenicol or 20 μg/ml spectinomycin where appropriate. Overexpression of GlpE in cells carrying pGZ105 was induced in mid–log phase by the addition of 0.5 mM IPTG and in cells carrying pGZ154 or pGZ132 by 0.1 μg/ml tetracycline. Cultures were grown an additional 2 to 4 hours after induction and then harvested in late log phase as described below. BP331 and BP1331 were grown in minimal M9 media (68) with variations described in the text. For anaerobic growth, cells were grown in sealed tubes filled to the top with growth media. All cultures were grown at 37°C.

Protein purification. BL21(DE3) harboring pGZ105 was grown and induced in approximately 500 ml LB as described above. Extraction of periplasmic contents was performed using a modified freeze–thaw method (139). Cells were harvested by centrifugation and washed with 1/2 volume of 25 mM Tris–acetate (pH 8.6), 10 mM ammonium thiosulfate. Cells were repelleted and stored at −70°C. The frozen pellet was thawed on ice, resuspended in 1/50 the original volume of buffer A (50 mM Tris–HCl (pH 7.2), 3 mM EDTA) and incubated on ice for 30 minutes. Cells were pelleted and the supernatant decanted and saved. The incubation was repeated twice more and the supernatants were pooled giving the periplasmic extract (P_{ext}).

The extract was loaded onto a 10 x 100 mm pre–packed Waters quaternary methylamine(Q)–polymethacrylate anion exchange column (Protein–Pak™ Q15HR 1000A) equilibrated at room temperature with buffer A. Flow through the column was maintained at 1.5 ml/min using a Waters 650 Advanced Purification System. Protein concentration in the column eluate was followed by continuously monitoring the absorbance at 280 nm. After the sample had been loaded the column was washed for 2 h with buffer A. The column was then developed with a 45 ml gradient from 0 to 105 mM NaCl in buffer A followed by a 112.5 ml gradient from 105 to 255 mM NaCl in buffer A. Fractions containing rhodanese activity, eluted between 155 and 190 mM NaCl, were pooled (Q_1).

The column was washed with buffer A containing 1 M NaCl for 1 h and then reequilibrated in buffer A. The pooled fractions (Q_1) were diluted 2–fold with buffer A and reloaded onto the column. Protein was eluted and concentrated using buffer A containing 750 mM NaCl. Fractions containing rhodanese activity were pooled (Q_2). The second Q column eluate was brought to 15% glycerol, divided into aliquots and stored at −70°C.

Protein concentrations were determined by the method of Bradford (74) with bovine serum albumin as the standard. Coomassie protein assay reagent and the bovine serum albumin standard were purchased from Pierce and the manufacturer's microassay method was used.

Estimation of subunit and native molecular mass. Subunit molecular mass of GlpE was determined by comparison to markers on a sodium dodecyl sulfate–15% polyacrylamide gel. SDS–PAGE was performed as described by Laemmli (75) except a modified sample loading buffer was used. The buffer contained 62.5 mM Tris–HCl (pH6.8), 2% SDS, 10% glycerol, 0.02% bromophenol blue and 1.5 M β–mercaptoethanol and was prepared immediately prior to use. Protein bands were visualized using Fast Stain from Zoion as

described by the manufacturer. Subunit molecular mass of GlpE was verified using mass spectrometry (Kim Harich, personal communication).

Molecular exclusion chromatography was used to determine the native molecular mass of GlpE. Samples of 0.1 ml containing from 40 to 80 μg GlpE or various standards were applied to a Waters glass Protein–Pak™ 300SW (8 x 300 mm) column. The column had been equilibrated at ambient temperature with 20 mM Tris–HCl (pH 7.2), 1 mM EDTA, 100 mM NaCl at a flow rate of 0.8 ml/min. Void volume and total volume were determined using blue dextran (average molecular mass, 2,000 kDa) and vitamin B_{12}, respectively. Protein standards used for generation of a standard curve were bovine serum albumin, ovalbumin, soybean trypsin inhibitor, myoglobin and ribonuclease. The standard curve was generated by fitting the data to a fourth order polynomial using Microsoft Excel.

Sulfurtransferase activity measurements. Two methods were used to measure the sulfurtransferase activity of GlpE.

(a) The assay used to measure sulfurtransferase activity during purification of GlpE and for characterization of the enzyme was essentially the same as that described previously to quantify thiocyanate formation (108,126). Briefly, the protein sample was added to 0.1 ml 500 mM Tris–acetate (pH 8.6), 0.05 ml 500 mM ammonium thiosulfate and brought to 0.45 ml using deionized water. The reaction was initiated by the addition of 0.05 ml 500 mM KCN and terminated, usually after 0.5 to 2 minutes (up to 30 minutes for cell extracts not overexpressing GlpE), by the addition of 0.25 ml 15% formaldehyde. Color was developed by the rapid addition of 0.75 ml of ferric nitrate reagent (100 g $Fe(NO_3)_3 \cdot 9H_2O$ and 200 ml 65% HNO_3 per 1500 ml) and the absorbance at 460 nm was determined. Assays were performed at ambient temperature (approximately 22°C). Assays containing whole cells or visible precipitated protein were centrifuged prior to measuring the absorbance. One unit of enzyme is defined as the amount that catalyzes the production of 1 μmol of thiocyanate per minute corresponding to an absorbance change at 460 nm of 2.8 in this system.

When assaying partially oxidized samples of GlpE, the protein was diluted to a working concentration in 100 mM Tris–acetate (pH 8.6), 6 mM cysteine, or cysteine was included in the assay mixture at 6 mM. In instances where the cysteine might be expected to interfere with the results, as when assaying GlpE inactivated by DTNB, no cysteine was included. As described for the previously characterized *E. coli* rhodanese, addition of cysteine did not result in a significant increase in activity of freshly prepared enzyme, but only of partially oxidized samples (126).

(b) Assays to measure the ability of GlpE to transfer sulfur from thiosulfate to thioredoxin were performed essentially as described previously for the bovine liver rhodanese (131). Each assay was prepared in a 1 ml cuvette and contained sufficient reagents such that the final concentrations in 1 ml would be 50 mM potassium phosphate (pH 8.2), 0.1 units/ml thioredoxin reductase and 50 μM NADPH. Sufficient thioredoxin was also added to give the desired concentration in a final volume of 1 ml. Deionized water was added to a volume of 1 ml less the amount to be subsequently added. Cuvettes containing all reagents except NADPH were used as a blank. NADPH was added and the reaction mixtures were allowed to equilibrate for at least 1 hour. Periodic measurements of the absorbance at 340 nm were used to ensure mixtures had reached equilibrium. Once equilibrium had been reached, either

purified GlpE (4 μM monomer, 48 μg/ml), ammonium thiosulfate (80 mM) or both were added bringing the final volume of each reaction to 1 ml. The absorbance at 340 nm was periodically measured to quantitate the rate of NADPH oxidation. Assays to determine the ability of NifS from *A. vinelandii* to transfer sulfur from cysteine to thioredoxin utilized the same procedure except 4 μM NifS monomer (36 μg/ml) was substituted for GlpE and 0.5 mM cysteine was substituted for ammonium thiosulfate.

Chemical inactivation of GlpE. GlpE (60 μg) was incubated overnight at ambient temperature in 0.2 ml 100 mM Tris–acetate (pH 8.6), 150 mM NaCl with varying molar ratios of DTNB. Remaining sulfurtransferase activity was determined by method (a) described above and compared to GlpE incubated without DTNB.

Fractionation procedures. Two fractionation procedures were used to determine the cellular location of GlpE. DH5αZ1 harboring pATCBO$_A$2+6 and pGZ154 was grown and overexpression of GlpE was induced using tetracycline as described above. Cells were harvested and either of the two fractionation procedures described below were performed to extract periplasmic contents.

(a) Spheroplasts were prepared by incubation with lysozyme and EDTA essentially as described previously by Kaback (140). Cells were washed twice with 1/10 volume 10 mM Tris–HCl (pH 8) and pelleted by centrifugation. The cell pellet was resuspended in 30 mM Tris–HCl (pH 8), 20 % sucrose, 10 mM EDTA at an A_{600} of approximately 20. The cell suspension was brought to 0.5 mg/ml lysozyme using a freshly–prepared stock solution of 25 mg/ml in 10 mM Tris–HCl (pH 8). After incubation at room temperature for 30 minutes with gentle agitation, spheroplasts were removed by centrifugation for 5 minutes at 10000 g. The supernatant was decanted and saved as the periplasmic fraction. Pellets were resuspended in the same volume of 30 mM Tris–HCl (pH 8), 20 % sucrose, 10 mM EDTA and sonicated 3 x 10 seconds at a setting of 50% using a Fisher Sonic Dismembrator (model 300). Rhodanese, alkaline phosphatase and glucose–6–phosphate dehydrogenase activities were then determined for both the periplasmic and cytoplasmic fractions. Total activities reported are the sum of the activities measured in the periplasmic and cytoplasmic fractions.

(b) Freeze–thaw treatment was performed as described above for the purification of GlpE. After isolation of the periplasmic extract the cell pellet was resuspended in freeze/thaw extraction buffer and sonicated as described above. Again, rhodanese, alkaline phosphatase and glucose–6–phosphate dehydrogenase activities were determined for both the periplasmic and cytoplasmic fractions. Total activities reported are the sum of the activities measured in the periplasmic and cytoplasmic fractions.

Other enzyme assays. β–galactosidase activity was determined using the method of Miller (70) utililizing ONPG as substrate and measuring the change in absorbance at 420 nm. Alkaline phosphatase activity was determined using a previously described modification (33) of the method of Schneider and Beck (141) utilizing the substrate PNPP and continuously monitoring the increase in absorbance at 410 nm. Glucose–6–phosphate dehydrogenase activity was determined as described previously (142) by following the increase in absorbance at 340 nm due to reduction of NADP.

Results

Identification of GlpE as a rhodanese. Comparison of GlpE to other proteins in the data bases revealed that GlpE possesses sequence similarity to the family of sulfurtransferases that includes thiosulfate:cyanide sulfurtransferases or rhodaneses (106). (Refer to Figure 18 for a sequence alignment of the putative active site of GlpE with the active sites of other representative members of this family.) To determine if GlpE possesses rhodanese activity, rhodanese specific activities of strains that overexpress GlpE were compared to those of similar strains that did not overexpress the protein. Overexpression of GlpE using pGZ154, pGZ132 or pGZ105 in the appropriate host strain typically resulted in a 10 to 1000 fold increase in rhodanese specific activity depending on the expression system (see Table 16). The higher level of rhodanese specific activity corresponded to an increased production of GlpE as visualized by SDS–PAGE (Figure 19). The level of rhodanese activity determined for cells not overexpressing GlpE (0.03 to 0.04 units/mg) agreed well with a previously published report (0.035 units/mg) (119).

Purification of GlpE. Previous studies suggested there are at least two enzymes responsible for rhodanese activity in *E. coli*, one apparently isolated from the periplasm using freeze–thaw treatment and the other released from cells only after sonication (125,126). During initial attempts to purify GlpE, it was discovered that most methods traditionally used for isolating periplasmic proteins from intact cells also resulted in the release of the overexpressed GlpE. This apparent periplasmic localization of GlpE and the use of an overexpression plasmid facilitated purification of the enzyme.

A 500 ml culture of BL21(DE3) harboring pGZ105 grown and induced as described in Experimental Procedures was used as starting material for the purification. The purification scheme is described in detail in Experimental Procedures and summarized in Table 17 with results from a typical purification. Extraction of the periplasmic contents was performed using a modified freeze–thaw method. Addition of Triton X–100 to the extraction buffer did not appreciably increase the efficiency of the extraction, but storage of the cell pellet at –70°C for a week or more increased the percentage of total GlpE released by the modified freeze–thaw method.

The periplasmic extract was applied to an anion exchange column. Three peaks containing rhodanese activity were observed during anion exchange chromatography of the periplasmic extract. The relative protein concentrations of the three peaks varied from preparation to preparation but two of the three peaks eluting near 160 and 185 mM NaCl typically contained most of the total rhodanese activity. The third peak eluted during the 1 M NaCl wash of the column. The major protein found in all three peaks exhibited similar mobility during SDS–PAGE (approximately 12 kDa). The proteins were therefore tentatively identified as various forms of GlpE, each having a different affinity for the anion exchange column. Rhodaneses are known to assume a variety of forms, including a sulfur–free form, a persulfide–containing form (113,143) and possibly oxidized, less active forms (126,144,145). Addition of cysteine to samples applied to the column had no significant effect on the relative amounts of each peak. Addition of ammonium thiosulfate was not tried but may have proven useful in favoring the persulfide form of GlpE over the others.

TABLE 16

Overexpression of GlpE and corresponding increase in rhodanese activity

Strain [Plasmid]	Rhodanese specific activity[a]	Fold overexpression
	Units/mg	
DH5αZ1 [pZE21–*luc*]	0.03	(1)
DH5αZ1 [pGZ154]	0.6	20
DH5αZ1 [pGZ132]	0.7	23
BL21(DE3) [pT7–7]	0.04	(1)
BL21(DE3) [pGZ105]	21	530

[a] Determined utilizing crude sonicated cell extract.

TABLE 17

Purification of the *glpE*–encoded rhodanese from *E. coli*

Purification step	Volume	Protein	Activity	Specific activity	Yield
	ml	mg	units	units/mg	%
1. Whole cells	500	120	1300	11	(100)
2. Extract (P_{ext})	30	9.6	860	90	66
3. Q column (Q_1)	32	1.6	350	220	27
4. Q column (Q_2)	2	1.3	350	270	27

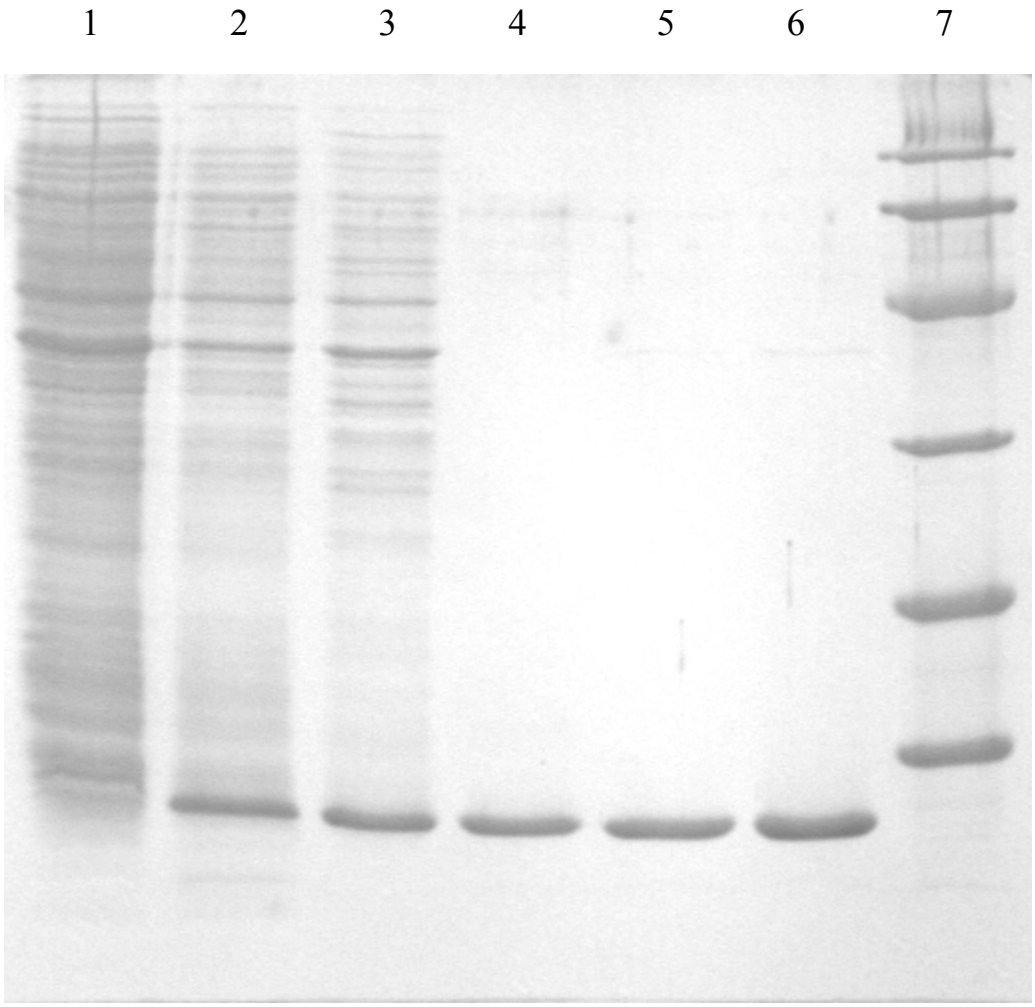


Figure 19. SDS–PAGE illustrating the overexpression and purification of GlpE. (1) 10 μg total cellular protein from BL21(DE3) harboring pT7–7. (2) 10 μg total cellular protein from BL21(DE3) harboring pGZ105 (whole cells). Both cultures were induced with IPTG as described in Experimental Procedures. (3) 10 μg protein released from intact cells using the modified freeze–thaw method (P_{ext}). (4) 2 μg protein from the pooled fractions following the first anion exchange chromatographic step (Q_1). (5,6) 2 and 4 μg protein, respectively, from the pooled fractions following the second anion exchange chromatographic step (Q_2). (7) Molecular mass markers corresponding to 97 kDa (phosphorylase b), 66 kDa (bovine serum albumin), 45 kDa (ovalbumin), 30 kDa (carbonic anhydrase), 21 kDa (trypsin inhibitor) and 14.4 kDa (lysozyme).

Fractions eluted between 150 and 190 mM NaCl encompassing two of the three peaks were pooled. The resulting protein was greater than 95% homogeneous when visualized on a 15% SDS–PAGE gel using the loading buffer described in Experimental Procedures (See Figure 19). Use of a less reducing buffer such as that described by Laemmli (75) resulted in a doublet. The pooled fractions were diluted 2–fold using buffer A and reapplied to the anion exchange column. Buffer A containing 750 mM NaCl was used to elute and concentrate the

purified protein. GlpE was not quantitatively precipitated using ammonium sulfate even near 100% saturation. Acidification by addition of either unbuffered cysteine or hydrochloric acid to pH 3.5 was also found useful in concentrating GlpE without significant loss of activity but only from solutions that contained greater than 0.5 mg/ml total protein.

When stored for extended periods at either 4ºC or –70°C, GlpE was found to be stabilized by 750 mM NaCl or $(NH_4)_2SO_4$. Though stabilized by the presence of salts, GlpE still lost activity, presumably due to oxidation as described for other rhodaneses (126,144,145). Addition of micromolar concentrations of cysteine, dithiothreitol or β–mercaptoethanol was found to reactivate oxidized enzyme. However, millimolar concentrations of the reducing agents (except for cysteine) resulted in an apparent decrease in enzyme activity. This may be due to competition between these agents and cyanide for the sulfur transferred from thiosulfate by GlpE. Similar phenomena have been reported for the well–characterized bovine liver rhodanese (146,147) and the previously described "accessible" rhodanese from *E. coli* (126).

Molecular mass determination of GlpE. The estimated molecular mass of the GlpE monomer is 12 kDa based on the migration of the fully–reduced protein (see Experimental Procedures for the loading buffer used) in an SDS–15% polyacrylamide gel (Figure 19). This value agrees well with that calculated from the deduced amino acid sequence (12.1 kDa) and a previously published report (12.1 kDa) (6). This value was also confirmed by mass spectrometry (Kim Harich, personal communication). SDS–PAGE of an oxidized sample using loading buffer without reducing agent resulted in a series of lighter bands corresponding to roughly 24, 36, 48 and 60 kDa in addition to the major band observed near 12 kDa. Addition of cysteine or DTT to samples applied to the gel resulted in only the band corresponding to 12 kDa.

Gel filtration chromatography was used to determine the native molecular mass of GlpE. GlpE migrated as a dimer (22.5 kDa) under a variety of buffer conditions (Figure 20). Addition of cysteine or DTT to the buffer, though able to activate oxidized enzyme in assays, did not appreciably change the migration of the enzyme through the gel filtration column. In addition to the major peak corresponding to the dimeric form of GlpE, minor peaks were observed when the oxidized sample described above was applied to the column. These peaks corresponded to higher order polymeric forms similar to those observed by SDS–PAGE under non–reducing conditions.

Kinetic properties of GlpE. The well–characterized bovine liver rhodanese utilizes a double–displacement (ping pong) mechanism during catalysis (113,143). Data from activity measurements of purified GlpE with varying concentrations of thiosulfate at several fixed concentrations of cyanide were fit to the equation describing this type of mechanism:

$$1/v_o = \frac{K_m^{SSO_3^{2-}}}{V}\frac{1}{[SSO_3^{2-}]} + \frac{K_m^{CN^-}}{V}\frac{1}{[CN^-]} + \frac{1}{V}$$

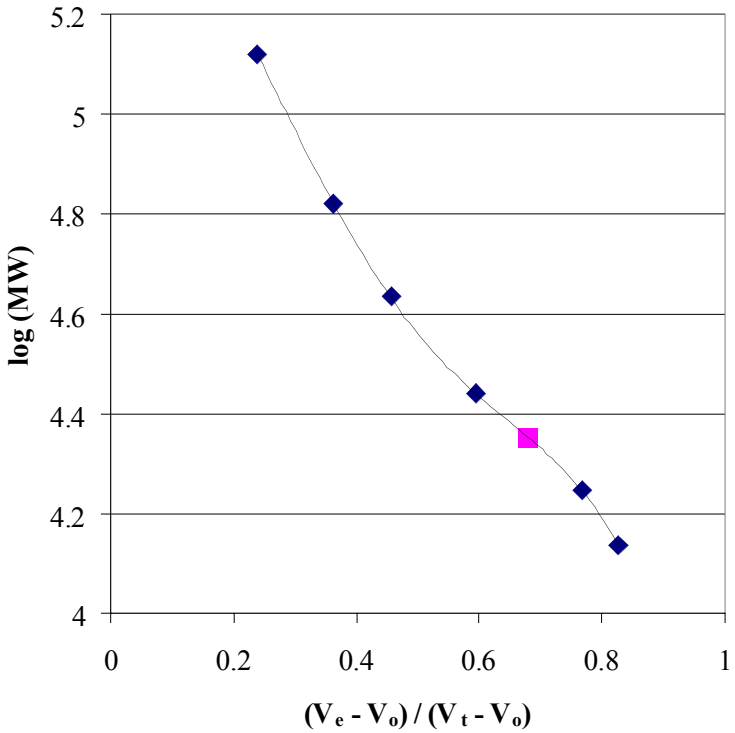
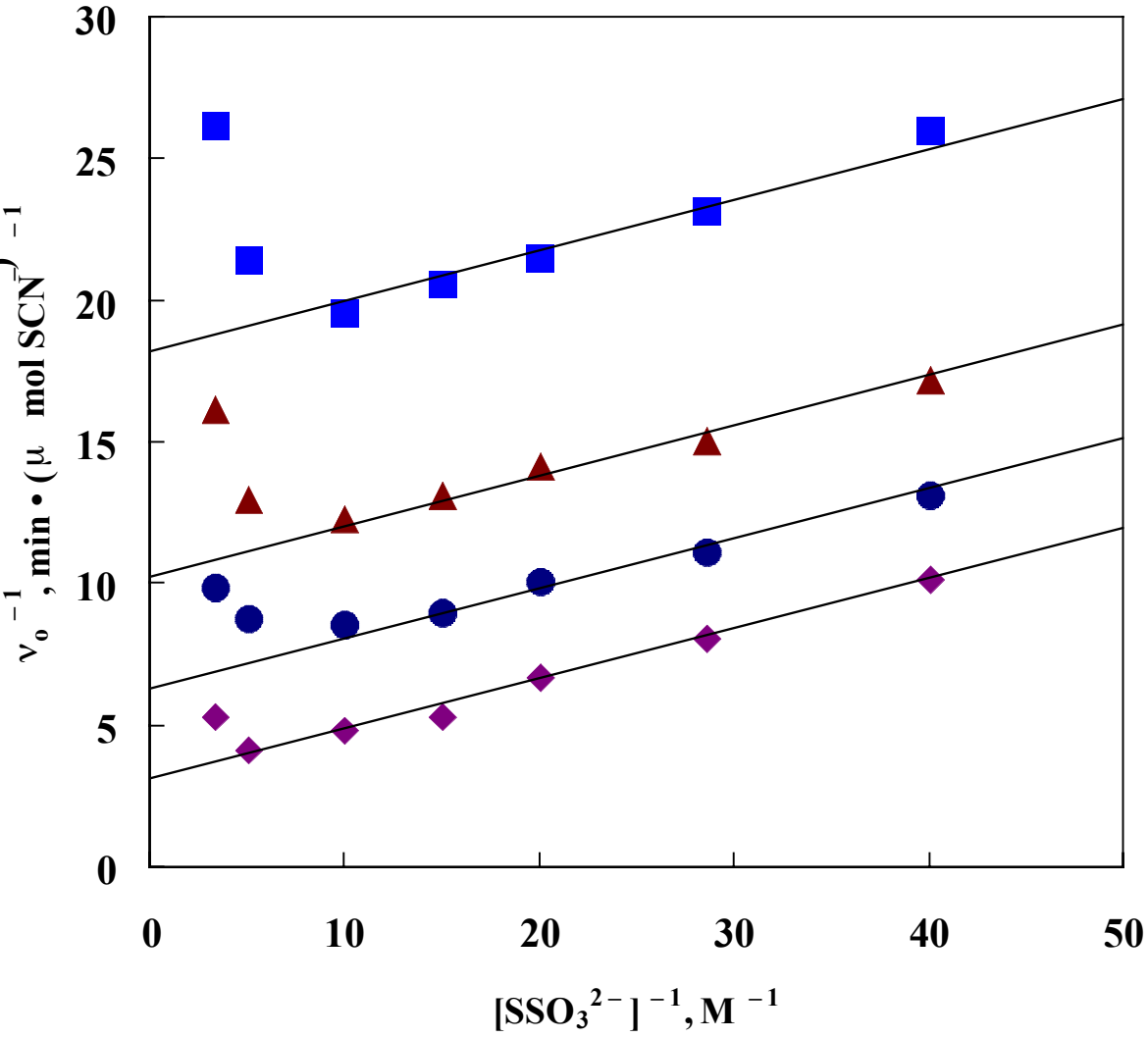


Figure 20. Estimation of the native molecular mass of GlpE. Gel filtration chromatography was used to estimate the native molecular mass of GlpE (■). Proteins used for generation of the standard curve in order of decreasing molecular mass are bovine serum albumin (132 and 66 kDa), ovalbumin (43 kDa), soybean trypsin inhibitor (27.5 kDa), myoglobin (17.6 kDa) and ribonuclease (13.7 kDa) (◆). V_e, V_o, and V_t correspond to the elution volume, void volume and total volume, respectively.

Figure 21(a) shows the data and the fit obtained using a BASIC program designed for linear regression of multivariable functions. Figure 21(b) is a secondary double–reciprocal plot of the apparent maximum velocities obtained from 21(a) versus cyanide concentration. The regression of the data to the equation describing a double–displacement mechanism yielded K_m values for thiosulfate and cyanide of 78 mM and 17 mM, respectively. The k_{cat} based on GlpE functioning as a dimer was determined to be 230 s^{-1}.

Cysteine, thioglycerol, dithiothreitol and β–mercaptoethanol could not act as sulfur donors for GlpE as evidenced by lack of thiocyanate formation when these substances were used in place of thiosulfate in standard rhodanese assays. However, as described above, dithiothreitol and β–mercaptoethanol may serve as sulfur acceptors at millimolar concentrations. As described in the Introduction, mercaptopyruvate sulfurtransferases are related to rhodaneses (114,115). However, GlpE exhibited less than 1% of the thiocyanate formation when mercaptopyruvate was used as substrate versus the same concentration of thiosulfate.

(a)



(b)

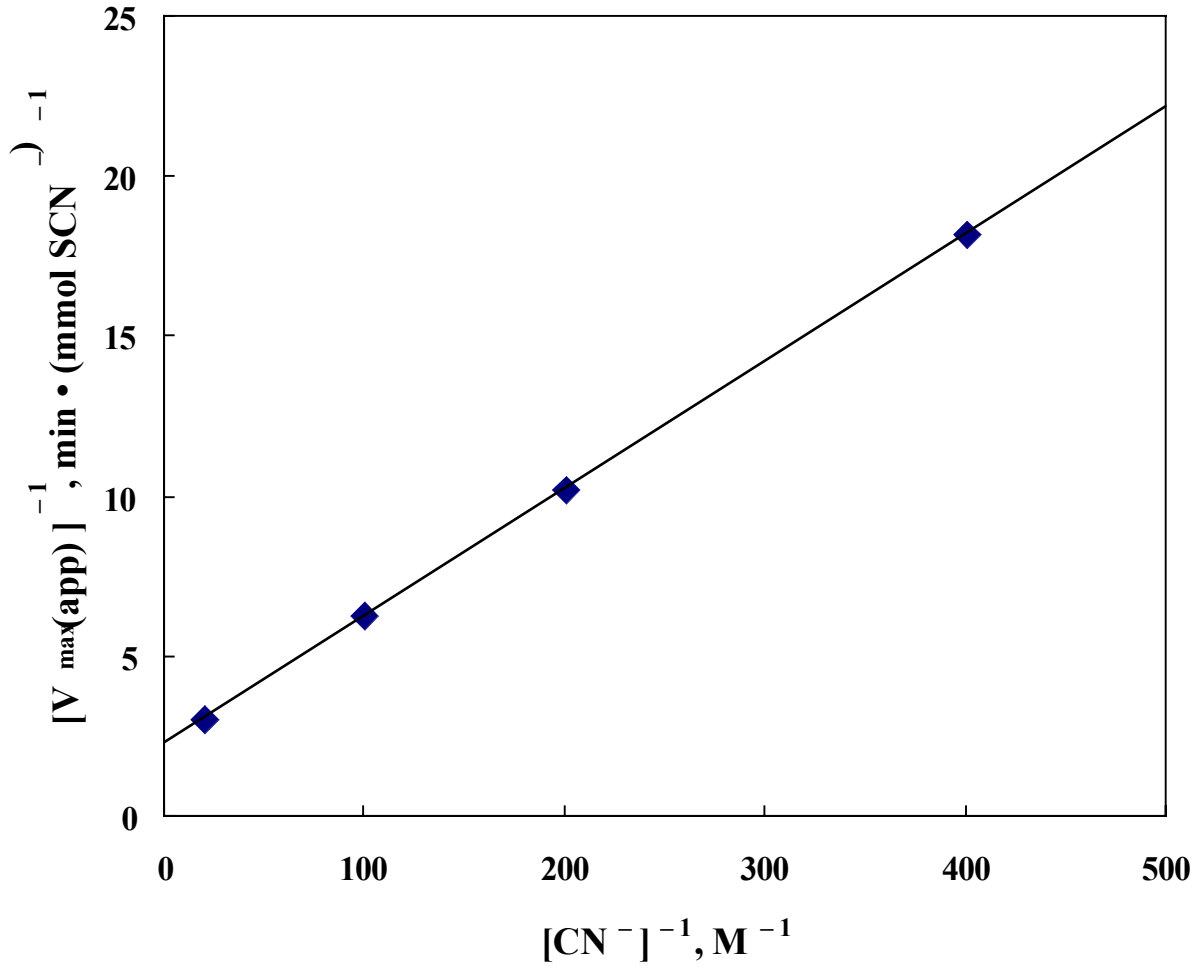


Figure 21. Kinetic characterization of the thiosulfate:cyanide sulfurtransferase reaction catalyzed by GlpE. Each assay, performed as described in Experimental Procedures, contained 10 µl (0.75 µg) purified GlpE in 100 mM Tris–acetate (pH 8.6), 6 mM cysteine. (a) Double–reciprocal plot of the rate of thiocyanate formation versus thiosulfate concentration at various fixed concentrations of cyanide. 2.5 mM (■), 5 mM (▲), 10 mM (●) and 50 mM (◆). (b) Secondary double–reciprocal plot of apparent V_{max} from (a) versus cyanide concentration. Data, excluding the points exhibiting substrate inhibition, were fit to the equation describing a double–displacement mechanism.

GlpE exhibited a broad pH optimum near 7, but some discrepancies were observed using various buffering systems. Tris–acetate at pH 8.6, the buffering system used in the standard rhodanese assay, yielded the highest activity. Several previously characterized rhodaneses including the bovine liver enzyme (148) and the "accessible" rhodanese of *E. coli* have been shown to be specifically inhibited by certain anions (126). No specific inhibition by the anions acetate, sulfate or phosphate was observed for GlpE. GlpE activity in the presence of these anions was similar to the activity observed in the presence of other anions, such as chloride or citrate, at a comparable ionic strength. Addition of various salts including sodium phosphate, sodium acetate, sodium chloride, sodium sulfate or potassium phosphate up to 0.3 I resulted in not more than a 20% loss of activity (data not shown). Substrate inhibition by thiosulfate was observed particularly at low cyanide concentrations (Figure 21(a)). This type of inhibition had been previously described for the bovine liver rhodanese and the "accessible" rhodanese of *E. coli* (126,148).

Chemical inactivation of GlpE by cysteine–specific reagents. Catalysis by other characterized rhodaneses requires a cysteine residue. The cysteine forms a persulfide linkage to the sulfane sulfur from thiosulfate in an enzyme–sulfur intermediate (113,143). GlpE contains two cysteine residues, one of which has been identified as the putative active site of GlpE based on sequence alignment with other characterized rhodaneses (see Figure 18). GlpE was incubated with the cysteine–specific modifying reagent, DTNB, to determine if a cysteine residue is important for activity. GlpE incubated with DTNB in a 2 to 1 molar ratio resulted in an over 90% loss of activity (Figure 22) indicating that at least one of the cysteine residues present in GlpE plays an important role in catalysis. Alignment of GlpE with other sulfurtransferases (Figure 18) indicates the active–site cysteine is most likely C65.

Cellular location of GlpE. Two rhodanese activities have been described previously in *E. coli*. One is released from cells only after sonication (125). The other is "accessible" to substrates when intact cells are added to assay mixtures and is released from cells by freeze–thaw treatment (125,126). Release of a protein from cells by freeze–thaw treatment may indicate a periplasmic localization. Like the "accessible" enzyme, GlpE is released from cells by freeze–thaw treatment. GlpE, however, has no N–terminal targeting sequence that would direct it to the periplasm. Also, GlpE is sensitive to oxidation and the periplasm is known to be an oxidizing environment while the cytoplasm is a more reductive environment. To determine the cellular location of GlpE, the periplasmic contents of cells overexpressing GlpE from pGZ154 were isolated using either the modified freeze–thaw method or the preparation of spheroplasts by incubation with lysozyme and EDTA. Rhodanese activities in the periplasmic and cytoplasmic fractions were compared. Plasmid pATCBO$_A$2+6 (33) was introduced into DH5αZ1 harboring pGZ154 to provide constitutive expression of alkaline phosphatase which served as a periplasmic marker. Glucose–6–phosphate dehydrogenase activity was measured as a cytoplasmic marker. In cells that overexpress GlpE, most of the rhodanese activity was released from cells by the modified freeze–thaw method. Lack of significant glucose–6–phosphate dehydrogenase activity in the periplasmic extract indicated that there was little contamination by cytoplasmic contents (Table 18). Other methods that

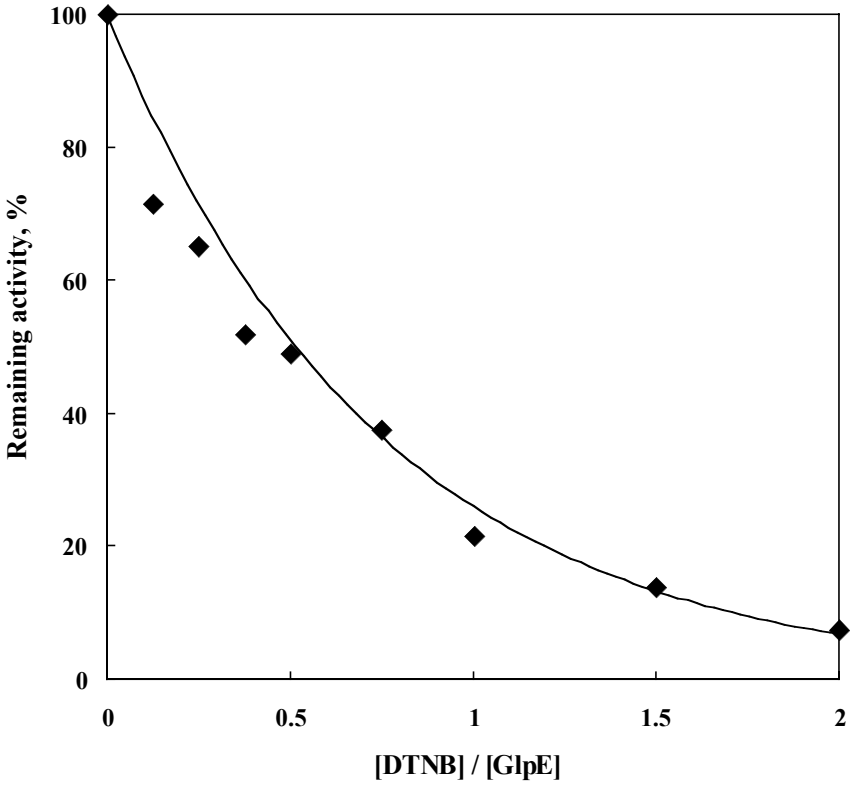


Figure 22. Inactivation of GlpE by the cysteine–specific modifying reagent DTNB. GlpE (60 μg) was incubated overnight at ambient temperature in 0.2 ml 100 mM Tris–acetate (pH 8.6), 150 mM NaCl with varying molar ratios of DTNB. Remaining sulfurtransferase activity was determined and compared to GlpE incubated without DTNB.

have been used for isolating periplasmic proteins including extraction with Triton X–100 (149), treatment with chloroform (150) or osmotic shock (151) also resulted in the almost quantitative release of rhodanese activity from cells overexpressing GlpE (data not shown). However, as shown in Table 18, preparation of spheroplasts by incubation with lysozyme and EDTA without a subsequent osmotic shock released the periplasmic marker alkaline phosphatase but failed to release a significant percentage of the total rhodanese or glucose–6–phosphate dehydrogenase activities. Preparation of spheroplasts, a gentler method for isolating periplasmic contents, is considered a more stringent method for determining cellular location (128,140). These results demonstrate that GlpE is a cytoplasmic protein.

Thioredoxin acts as a sulfur–acceptor substrate for GlpE. There are a number of other proteins known to be cytoplasmic but which are released by methods such as osmotic shock that are usually highly selective for the release of periplasmic proteins. This unique group of proteins, previously called group D proteins (128), includes elongation factor Tu (EF–Tu) (152), DnaK (130) and the small dithiol protein thioredoxin (129). These proteins are generally believed to be associated with the inner membrane (128–130,152).

TABLE 18

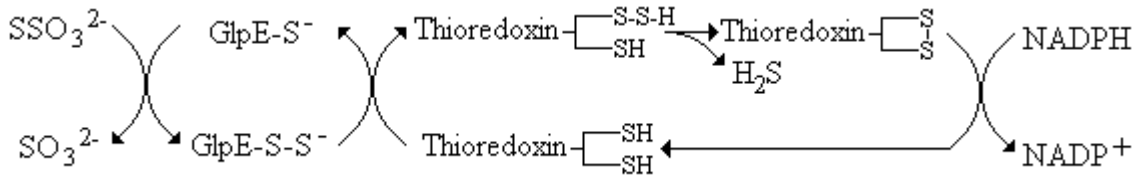
Localization of GlpE to the cytoplasm

Fractionation Method	Alkaline Phosphatase	Glucose–6–P dehydrogenase	Rhodanese
	% Activity released		
Spheroplasts[a]	69	5	5
Freeze/thaw[b]	76	<1	97

[a] Total enzyme activities recovered by the preparation of spheroplasts from 33 ml logarithmically growing culture: 0.03 units alkaline phosphatase; 0.3 units glucose–6–phosphate dehydrogenase; 1 unit rhodanese.

[b] Total enzyme activities recovered by the freeze/thaw fractionation method from 25 ml logarithmically growing culture: 12 units alkaline phosphatase; 0.1 units glucose–6–phosphate dehydrogenase; 10 units rhodanese.

Interestingly, another connection between thioredoxin and rhodanese has been postulated. It has been shown that reduced dithiols such as dithiothreitol (146,147), dihydrolipoate (153) and thioredoxin (131) can act as sulfur–acceptor substrates for the bovine liver rhodanese. Thioredoxin acts as a substrate with an affinity near that determined for cyanide. Dihydrolipoate also functions as a sulfur–acceptor substrate for the "accessible" *E. coli* rhodanese (126). The apparent K_m of the bovine liver rhodanese for thioredoxin isolated from *E. coli* was determined to be 18.5 μM while the K_m for cyanide is 63 μM. This observation led to the testing of thioredoxin as a sulfur–acceptor substrate for GlpE. The assay employed was essentially the same as that used for the bovine liver rhodanese (131). The basis for the assay is that when thioredoxin, which contains the active–site motif WCGPC, accepts a sulfur from rhodanese, a persulfide is formed at the N–terminal cysteine within the active site. The C–terminal cysteine within the active site then reacts with the persulfide yielding H_2S and oxidized thioredoxin. Thioredoxin reductase acts to reduce the disulfide bond of oxidized thioredoxin with the concomitant oxidation of NADPH:



A mixture containing thioredoxin, thioredoxin reductase and NADPH was allowed to equilibrate. Then either GlpE, ammonium thiosulfate or both were added. The oxidation of NADPH was quantitated by measuring the decrease in absorbance at 340 nm. Figure 23(a) shows that addition of either GlpE or ammonium thiosulfate alone resulted in no significant oxidation of NADPH. However, addition of both resulted in a marked linear rate of NADPH oxidation. This indicated that thioredoxin could act as a sulfur acceptor substrate for GlpE.

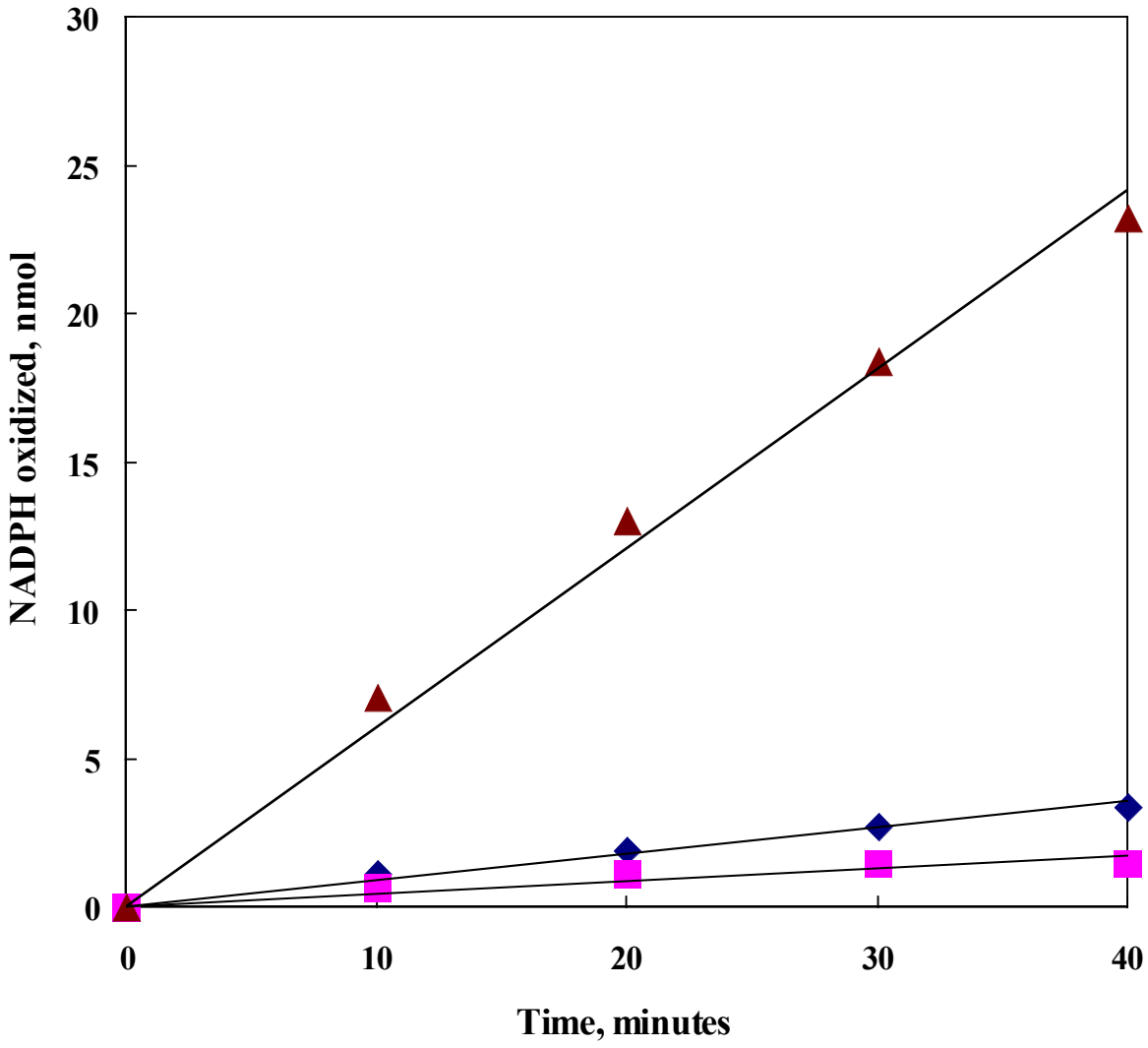
To quantitate the affinity of GlpE for thioredoxin, thioredoxin concentration was varied with ammonium thiosulfate concentration fixed at 80 mM. The reciprocal of the rates of NADPH oxidation are plotted in Figure 23(b) versus the reciprocal of the thioredoxin concentrations. Regression of the data yields an estimate of the apparent K_m for thioredoxin of 34 μM and an apparent maximum rate over 14 times greater than that observed with the bovine liver rhodanese. The rate corresponds to a turnover number of approximately 0.5 mole NADPH oxidized per mole GlpE dimer per minute. It is interesting to note that GlpE and the bovine liver rhodanese have a similar affinity for thioredoxin while GlpE has a much lower affinity for cyanide.

NifS from *A. vinelandii* is essential for nitrogenase activity. NifS catalyzes the desulfurization of cysteine to produce alanine and sulfur in the absence of reductant and is believed to play a vital role in the biosyntheis of iron–sulfur clusters in nitrogenase (133). Both NifS, in the presence of ferrous iron, cysteine and DTT (154), and thioredoxin, in the presence of ferrous iron, thioredoxin reductase, NADPH and sulfide (155), have been found to catalyze the formation of iron–sulfur clusters in SoxR, a transcriptional regulator sensitive to oxidative stress. In order to determine if NifS could also act as a sulfur–donor for thioredoxin, the above assay was utilized containing 16.6 μM thioredoxin. NifS at 4 μM was added in place of GlpE and cysteine at 0.5 mM replaced ammonium thiosulfate. As shown in Figure 24, addition of cysteine alone resulted in substantial NADPH oxidation (33 nmol/hr). However, addition of both cysteine and NifS resulted in a significantly higher rate (62 nmol/hr) of NADPH oxidation. When corrected for the oxidation due to cysteine (29 nmol/hr), NifS catalyzes sulfur transfer to thioredoxin at a rate similar to that observed with a comparable concentration of GlpE (20 nmol/hr).

The possibility that NifS may transfer sulfur from cysteine to GlpE was also tested. NifS and cysteine were included in standard rhodanese assays containing GlpE but no thiosulfate. No thiocyanate formation was observed indicating that GlpE could not accept sulfur from NifS.

Transcriptional control of *glpE*. To provide some insight into the physiological function of GlpE the regulation of transcription of *glpE* was examined. For this purpose, a transcriptional fusion of the *glpE* promoter to *lacZ* in single copy on the chromosome was used. Growth under a variety of conditions resulted in only small changes in the promoter activity as determined by assays of β–galactosidase activity. A two–fold decrease in activity was observed when cells were grown anaerobically. There appeared to be an approximate two–fold activation when thiosulfate was included as a sulfur source versus sulfate, but this phenomenon was only observed when magnesium chloride was inadvertently left out of the

(a)



(b)

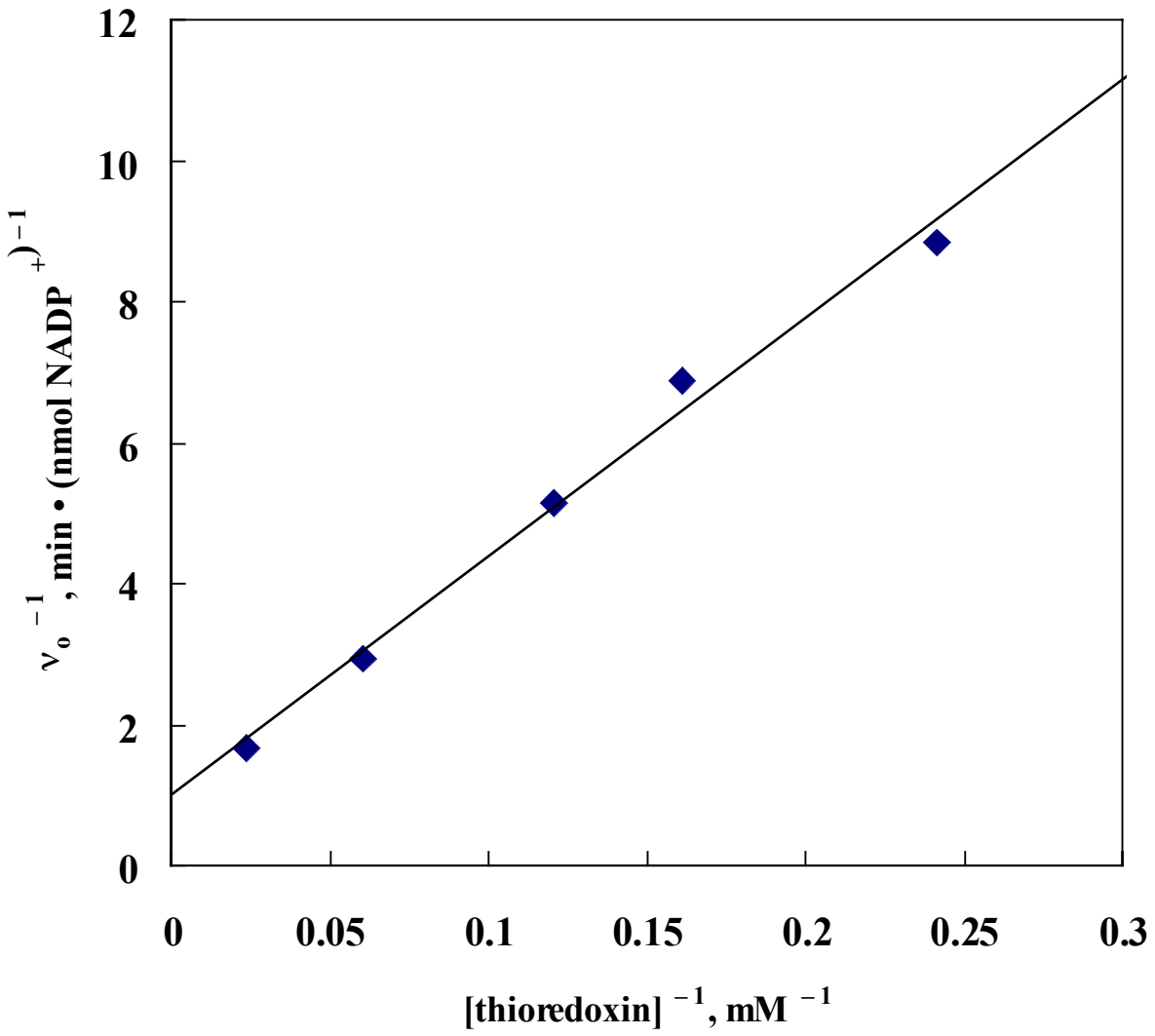


Figure 23. Thioredoxin acts as a sulfur–acceptor substrate for GlpE. (a) Assay mixtures containing 16.6 μM thioredoxin, 0.1 units/ml thioredoxin reductase and 50 μM NADPH were preequilibrated as described in Experimental Procedures. At time 0, either 4 μM GlpE (■), 80 mM ammonium thiosulfate (◆) or both (▲) were added. The absorbance at 340 nm was periodically measured to quantitate the rate of NADPH oxidation. (b) Double–reciprocal plot of the rate of NADPH oxidation measured as described in (a) versus thioredoxin concentration.

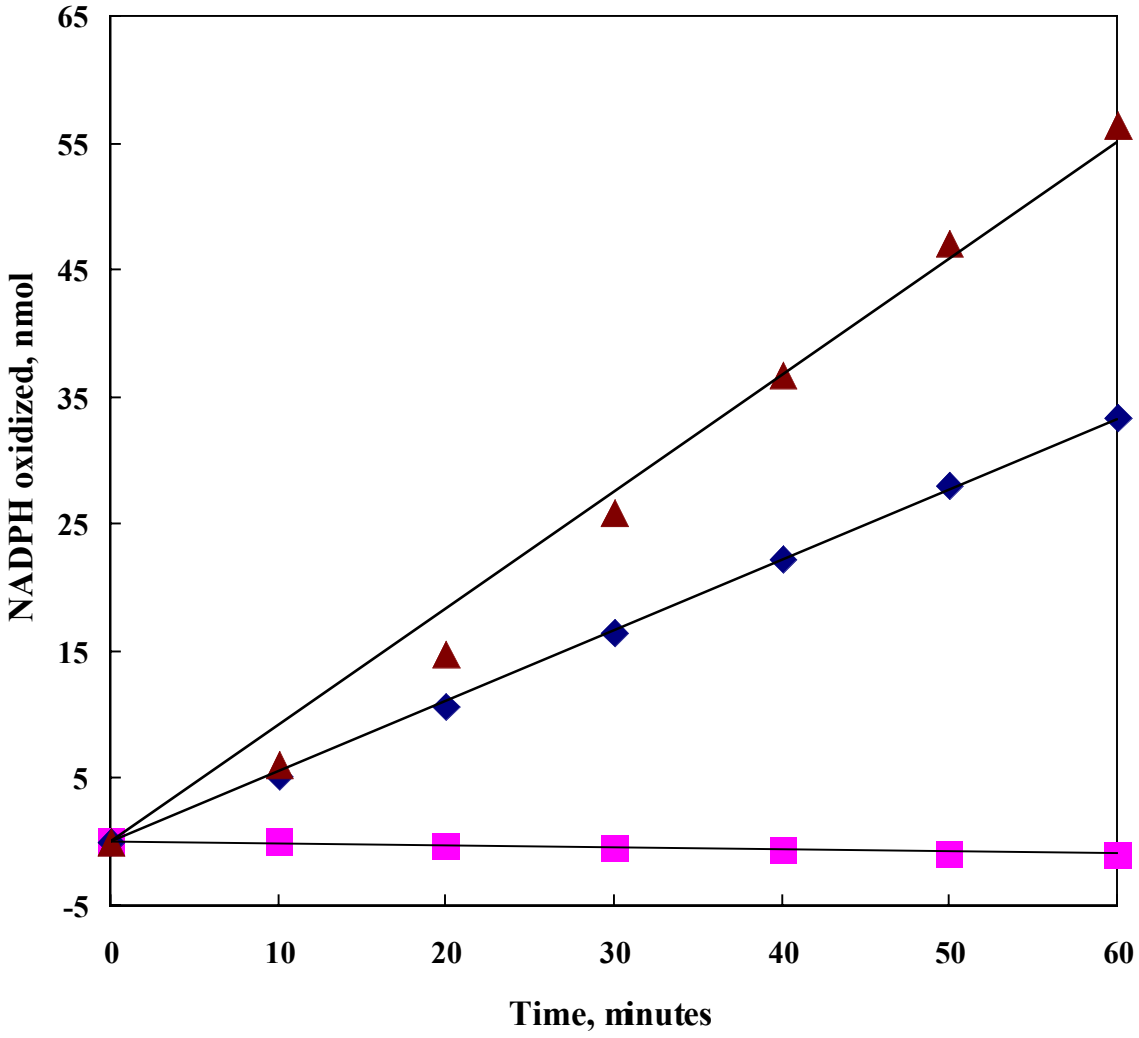


Figure 24. Thioredoxin acts as a sulfur–acceptor substrate for NifS. Assay mixtures containing 16.6 μM thioredoxin, 0.1 units/ml thioredoxin reductase and 50 μM NADPH were preequilibrated as described in Experimental Procedures. At time 0, either 4 μM NifS (■), 0.5 mM cysteine (◆) or both (▲) were added. The absorbance at 340 nm was periodically measured to quantitate the rate of NADPH oxidation.

medium used for this study. Addition of glycerol did not activate the expression of *glpE* and deletion of the *glpEGR* operon did not have a significant effect on the expression of *lacZ* from this promoter (Ben Potters, personal communication). This result is consistent with earlier work indicating that the promoter specific for *glpE* is not regulated by GlpR (60). Previous work with this transcriptional fusion on a multicopy plasmid (from which this lysogen was derived) indicated that the promoter was subject to catabolite–repression (60). Experiments to confirm this type of regulation for this promoter using the lysogenic strain (BP331) were inconclusive (Ben Potters, personal communication).

Discussion

Comparison of GlpE to other sulfurtransferases. GlpE appears to have some of the properties of a previously characterized *E. coli* rhodanese. This previously described rhodanese is referred to as the "accessible" rhodanese because of its accessibility to substrates when intact cells are added to assay mixtures (126). In cells overexpressing GlpE, typically 50% of the total rhodanese activity can be measured using intact cells (data not shown). Both GlpE and the previously described "accessible" rhodanese are released from cells by freeze–thaw treatment. The previously described enzyme had an estimated molecular mass of 14 kDa (126). The molecular mass of GlpE is 12.1 kDa (6). Expression of both the "accessible" rhodanese and GlpE is subject to catabolite repression and repressed during anaerobic growth (60,126). However, addition of cysteine not only activated oxidized "accessible" rhodanese, but also resulted in a shift from dimeric to monomeric form (126). Though GlpE is activated by cysteine, it does not exhibit this shift to a monomeric form in the presence of cysteine or the reducing agent DTT. These enzymes also have somewhat different kinetic characteristics. GlpE has a K_m (SSO_3^{2-}) of 78 mM and a k_{cat} of 230 per second based on the dimeric molecular weight. The "accessible" rhodanese has a K_m (SSO_3^{2-}) of 5 mM and a k_{cat} of 260 per second based on the estimated monomeric molecular weight (126).

GlpE also does not exhibit the anion–specific inhibition described for the "accessible" rhodanese and other characterized rhodaneses. K_i values for sulfate and acetate versus thiosulfate determined for the "accessible" rhodanese were 36 and 45 mM, respectively (126). For the bovine liver rhodanese these values were 40 and 167 mM (148). In the bovine liver rhodanese, this specific inhibition by anions has been linked to the two positively–charged residues located near the active–site cysteine (see Figure 18) and an arginine residue (R186) upstream of the active site (117). None of these residues are conserved in GlpE.

These basic residues have also been implicated in the recognition of thiosulfate as substrate. Substitution of R186 in the bovine liver rhodanese with isoleucine increases the K_m for thiosulfate from 3.7 mM to 73.2 mM. Substitution of K249 with alanine results in a protein which has almost no activity with thiosulfate as substrate but which can utilize a number of organic thiosulfonates almost as well as the wild–type protein (116). That these basic residues are not conserved in GlpE may provide a clue to the lower affinity for thiosulfate observed with GlpE versus the "accessible" rhodanese and the bovine liver rhodanese. The physiological sulfur donor for GlpE, based on this low affinity, is probably not thiosulfate.

The results of SDS–PAGE analysis of purified protein under reducing and nonreducing conditions as well as the results from gel filtration chromatography are consistent with GlpE forming a dimer that in the active form possesses no disulfide bonds. Since GlpE is related to the well–characterized rhodanese from bovine liver, this dimeric structure without a disulfide linkage seems the most likely native form for GlpE. The enzyme isolated from bovine liver is approximately twice the molecular weight of GlpE but folds into two almost identical domains, despite the two halves possessing only limited sequence similarity (112). The active site cysteine lies in a pocket between the two domains (156). The two domains interact by a number of hydrophobic, ionic and hydrogen–bonding interactions,

but the enzyme in the active form possesses no disulfide linkages (112,157). It has been suggested that the larger molecular weight sulfurtransferases resulted from a gene duplication event involving a gene for an ancestral low molecular weight rhodanese. Divergent evolution eventually gave rise to the superfamily of transfer proteins described by Hofmann, Bucher and Kajava (106).

Localization of GlpE to the cytoplasm. A small number of proteins in *E. coli* have been found to be released by most methods of isolating periplasmic proteins but have been determined to be located in the cytoplasm. This so–called group D class of proteins (128) includes thioredoxin (129), DnaK (130) and elongation factor Tu (EF–Tu) (152). The anomalous results obtained when attempting to determine cellular location of the group D class of proteins may explain the unexpected apparent periplasmic subcellular localization of some proteins observed by Link et al. (158).

The release of a cytoplasmic protein by osmotic shock has been suggested to require peripheral association with the inner surface of the inner membrane (128). This association with the inner membrane has been verified for these proteins by careful isolation of membranes (129,130,152). That GlpE is released from intact cells by most methods of isolating periplasmic proteins but not by production of spheroplasts indicates it also is a member of this so–called group D class (128) and may associate with the inner membrane. Analysis of the amino acid sequence of GlpE shows no striking hydrophobic region but GlpG, encoded by the gene immediately downstream of the gene encoding GlpE, is known to be an integral membrane protein (76). Association with GlpG could account for the peripheral association of GlpE with the membrane. The implications of this localization for the physiological role of GlpE are discussed below.

For thioredoxin, disruption of the gene encoding the mechanosensitive channel MscL greatly reduces its release during osmotic shock (159). Disruption of *mscL*, the gene encoding this channel, had no significant effect on the release of overexpressed GlpE during osmotic shock of strain GZ570 (*mscL*$^+$) versus strain KR1570 (*mscL*::Cm^r) (data not shown). In both cases, GlpE overexpressed using pGZ132 was almost quantitatively released by osmotic shock. Osmotic shock did not, however, release a significant percentage of GlpE overexpressed from pGZ132 in PB104 (137), the original *mscL*::Cm^r strain (data not shown). The reason for the inconsistency of these results has yet to be determined but may be related to the resistance of the parent strain (AW405) to the T1, T5 and T6 bacteriophages (136). This resistance often results from the absence of specific outer membrane proteins that act as receptors for the bacteriophages.

Physiological Role of GlpE. The physiological role of rhodaneses is still in question. Cyanide detoxification had been proposed as one possible role, but given the low affinity of GlpE for cyanide (K_m (CN^-) = 17 mM) this seems unlikely for GlpE. Overexpression of GlpE did not provide cells with increased resistance to cyanide (Farzana Ahmed, personal communication).

During previous work with GlpR, *ΔglpEGR* and *ΔglpEG* strains had been constructed. To date no phenotype associated with these deletions except for that expected for a *ΔglpR* strain has been identified. Rhodanese activity in these *ΔglpE* strains is comparable to parent

strains without the deletion under a variety of growth conditions (BL21(DE3) versus GZ544 for example, data not shown). Disruption of rhodanese–like genes in various bacteria has also been shown to have no significant effect on the total rhodanese activity (120–122). The level of expression of the rhodaneses may be coordinately regulated as has been shown for the thioredoxin family of proteins (160). The existence of several rhodanese–like enzymes in *E. coli* probably complicates the search for a phenotype. Staining of native gels for rhodanese activity as described by Whitehouse et al. (161) following electrophoresis of crude extracts of either wild–type versus *ΔglpEGR* strains showed no apparent differences.

Nothing has thus far implicated GlpE in sulfur metabolism in *E. coli*. Growth on various sulfur sources did not have a significant effect on the expression of *glpE*. It is of interest to note that the only phenotype identified to date for a rhodanese–like enzyme is that of CysA from *S. erythraea*. Disruption of the gene for this enzyme and the open reading frame immediately downstream results in cysteine auxotrophy but does not have a significant effect on the total rhodanese specific activity. Cysteine biosynthesis in this organism is different from that established for *E. coli* and is believed to involve thiosulfate as an intermediate. The researchers suggest CysA may function in the reverse direction to synthesize thiosulfate from elemental sulfur and sulfite (121). Thiosulfate is not an intermediate in aerobic cysteine biosynthesis from sulfate in *E. coli* or *Salmonella typhimurium*, a bacterium closely related to *E. coli*. However, thiosulfate has been suggested to be an intermediate in the anaerobic biosynthesis of cysteine from sulfate in *S. typhimurium* (162,163).

Rhodaneses have also been proposed to mobilize sulfur for the formation of iron–sulfur clusters. The iron–sulfur clusters of succinate dehydrogenase (164), mitochondrial NADH dehydrogenase (111), and ferredoxins from spinach (110) and *Clostridium pateurianum* (109) have each been shown to be reconstituted by incubation with bovine liver rhodanese, thiosulfate and ferric iron ions in the presence of dihydrolipoate. It has also been shown that addition of spinach apoferredoxin to rhodanese assays has an apparent inhibitory effect on the measured rhodanese activity, suggesting some of the sulfur is tranferred to the ferredoxin (153).

That rhodaneses function in iron–sulfur cluster biosynthesis is an attractive prospect for GlpE since the gene is divergently transcribed from *glpD* which encodes aerobic glycerol–P dehydrogenase. The *E. coli* aerobic glycerol–P dehydrogenase is thought to contain an essential iron–sulfur cluster (165,166), a hypothesis that has been verified for eukaryotic mitochondrial glycerol–P dehydrogenases (167,168). In fact *glpE* and *glpD* may share some regulatory components since both are repressed during anaerobic growth, subject to catabolite repression and separated by less than 200 nucleotides (60,169,170).

Previously, cells overexpressing GlpD were found to have a lower specific activity of glycerol–P dehydrogenase than what would be expected from the overexpression of the polypeptide observed by SDS–PAGE (138). The possibility that GlpE may be necessary for activation of GlpD was tested, but no increase in GlpD specific activity was observed in cells overexpressing both GlpD and GlpE (data not shown). However, comparison of the *glp* genes in other organisms shows that the organization of the *glpD–glpEGR* divergent operons is not highly conserved and the relevance of the association of the *glpE*–encoded rhodanese with glycerol–P metabolism in *E. coli* has yet to be determined.

Recent studies have indicated that thioredoxin is capable of serving as a sulfur–acceptor substrate for bovine liver rhodanese (131). I have shown that GlpE, like the bovine liver rhodanese, can transfer the sulfur from thiosulfate to thioredoxin with relatively high affinity. GlpE has an apparent K_m for thioredoxin of 34 μM when thiosulfate is present near its K_m. Thus the affinity for thioredoxin is almost 1000 times higher than that determined for cyanide. This interaction between GlpE and thioredoxin seems even more significant when the possible association of both proteins with the inner membrane is considered. A second thioredoxin (TrxC), also identified as a member of the group D class of proteins, has been recently characterized in *E. coli* (171). *E. coli* also possesses three isoforms of glutaredoxin, a protein structurally and functionally similar to thioredoxin (172–174). It is very likely that these proteins may also accept sulfur from GlpE, perhaps with a higher affinity.

Mitochondrial and bacterial membranes are known to be the richest source of respiratory chain enzymes containing iron–sulfur clusters. The well–characterized bovine liver rhodanese is localized within the mitochondria (108) and in this study I provide evidence that suggests a membrane–association for the rhodanese encoded by *glpE*. As mentioned previously, the association of the thioredoxins (TrxA and TrxC) of *E. coli* with the inner membrane has already been shown (129,171). Recently, novel thioredoxins were described in both rat liver and *Saccharomyces cerevisiae* that are localized within the mitochondria (175,176). The co–localization of the thioredoxins and rhodaneses with proteins rich in iron–sulfur clusters is consistent with the proposed roles of thioredoxins and rhodaneses in the formation of the iron–sulfur clusters.

It is unclear why NifS from *A. vinelandii* would also interact with thioredoxin with such high affinity unless NifS, thioredoxin and rhodanese are involved in some type of sulfur–transport chain as envisioned by Flint (177). Flint identified a NifS–like protein (IscS) in *E. coli* that could participate in the biosynthesis of the iron–sulfur cluster of dihydroxy–acid dehydratase. However, crude extract was more efficient at reconstituting the iron–sulfur cluster versus the purified NifS–like protein, suggesting the involvement of other factors (177). It is interesting that a multiprotein complex from extracts of bovine liver was found to possess both rhodanese and cysteine desulfhydrase activity and was implicated in the formation of iron–sulfur clusters (125). It was also shown that a partially–purified rhodanese isolated from *Thiobacillus novellus* produced sulfite in the absence of cyanide, but the purified enzyme required the presence of cyanide for sulfite production. The physiological sulfur–acceptor substrate was found to have a molecular mass of approximately 10 to 20 kDa (178), a value similar to thioredoxin.

The true physiological functions of GlpE specifically and rhodaneses in general have yet to be elucidated. It seems, however, as suggested by Colnaghi et al. (120), that the rhodanese structure provides an effective solution to a basic problem in sulfur metabolism, that of interconversion from one form of sulfur to another. Therefore, though some organisms may possess a number of rhodanese–like enzymes, the true physiological function of each may be unrelated. Nevertheless, the identification of GlpE as a rhodanese brings us one step closer to meeting our long–term objective of defining a function for all members of the DeoR family in *E. coli* and the genes regulated by these transcriptional repressors.

LITERATURE CITED

1. Gralla, J.D. and Collado–Vides, J. (1996) Organization and function of transcription regulatory elements. *Escherichia coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F.C.; Curtiss, R.; Ingraham, J.L.; Lin, E.C.C.; Low, K.B.; Magasanik, B.; Reznikoff, W.S.; Riley, M.; Schaechter, M. and Umbarger, H.E., eds) pp. 1232–1245, ASM Press, Washington, DC.

2. Choy, H. and Adhya, S. (1996) Negative control. *Escherichia coli and Salmonella. Cellular and Molecular Biology.* (Neidhardt, F.C.; Curtiss, R.; Ingraham, J.L.; Lin, E.C.C.; Low, K.B.; Magasanik, B.; Reznikoff, W.S.; Riley, M.; Schaechter, M. and Umbarger, H.E., eds) pp. 1287–1299, ASM Press, Washington, DC.

3. Weickert, M.J. and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J.Biol.Chem.* 267, 15869–15874.

4. Sulavik, M.C.; Gambino, L.F. and Miller, P.F. (1995) The MarR repressor of the multiple antibiotic resistance (*mar*) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. *Mol.Med.* 1, 436–446.

5. Sawers, A.; Kaiser, M.; Sirko, A. and Freundlich, M. (1997) Transcriptional activation by FNR and CRP: reciprocity of binding–site recognition. *Mol.Microbiol.* 23, 835–845.

6. Zeng, G.; Ye, S. and Larson, T.J. (1996) Repressor for the *sn*–glycerol 3–phosphate regulon of *Escherichia coli* K–12. Primary structure and identification of the DNA–binding domain. *J.Bacteriol.* 178, 7080–7089.

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

8. Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado–Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K–12. *Science* 277, 1453–1462.

9. Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.–F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.–I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O. and Venter, J.C. (1995) Whole–genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.

10. Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F. and Danchin, A. (1997) The complete genome sequence of the gram–positive bacterium *Bacillus subtilis*. *Nature* 390, 249–256.

11. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J.Mol.Biol.* 215, 403–410.

12. Reizer, J.; Ramseier, T.M.; Reizer, A.; Charbit, A. and Saier, M.H., Jr. (1996) Novel phosphotransferase genes revealed by bacterial genome sequencing: A gene cluster encoding a putative *N*–acetylgalactosamine metabolic pathway in *Escherichia coli*. *Microbiology* 142, 231–250.

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

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

15. Nobelmann, B. and Lengeler, J.W. (1996) Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *J.Bacteriol.* 178, 6790–6795.

16. Yoshida, K.–I.; Seki, S.; Fujimura, M.; Miwa, Y. and Fujita, Y. (1995) Cloning and sequencing of a 36–kb region of the *Bacillus subtilis* genome between the *gnt* and *iol* operons. *DNA Res.* 2, 61–69.

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

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

19. Rosey, E.L. and Stewart, G.C. (1992) Nucleotide and deduced amino acid sequences of the *lacR, lacABCD*, and *lacFE* genes encoding the repressor, tagatose 6–phosphate gene cluster, and sugar–specific phosphotransferase system components of the lactose operon of *Streptococcus mutans*. *J.Bacteriol.* 174, 6159–6170.

20. Kim, H. and Farrand, S.K. (1997) Characterization of the *acc* operon from the nopaline–type Ti plasmid pTiC58, which encodes utilization of agrocinopines A and B and susceptibility to agrocin 84. *J.Bacteriol.* 179, 7559–7572.

21. Lyi, S.M.; Jafri, S. and Winans, S.C. (1999) Mannopinic acid and agropinic acid catabolism region of the octopine–type Ti plasmid pTi15955. *Mol.Microbiol.* 31, 339–347.

22. Chen, Y.M.; Zhu, Y. and Lin, E.C.C. (1987) The organization of the *fuc* regulon specifying L–fucose dissimilation in *Escherichia coli* K–12 as determined by gene cloning. *Mol.Gen.Genet.* 210, 331–337.

23. Reizer, J.; Charbit, A.; Reizer, A. and Saier, M.H., Jr. (1996) Novel phosphotransferase system genes revealed by bacterial genome analysis: operons encoding homologues of sugar–specific permease domains of the phosphotransferase system and pentose catabolic enzymes. *Genome Sci.Technol.* 1, 52–75.

24. Zhao, N.; Hill, T. K.; Garrison, B. R.; Kok, A. L.; Bandara, A. B. P. A.; Zeng, G. and Larson, T. J. Characterization of noninducible *glp* repressor variants of *Escherichia coli* K–12 altered in *sn*–glycerol 3–phosphate binding affinity. 1999. (personal communication).

25. van Rooijen, R.J.; Dechering, K.J.; Niek, C.; Wilmink, J. and de Vos, W.M. (1993) Lysines 72, 80 and 213 and aspartic acid 210 of the *Lactococcus lactis* LacR repressor are involved in the response to the inducer tagatose–6–phosphate leading to induction of *lac* operon expression. *Protein Engineering* 6, 201–206.

26. Lin, E.C.C. (1996) Dissimilatory pathways for sugars, polyols, and carboxylates. *Escherichia coli and Salmonella. Cellular and Molecular Biology.* (Neidhardt, F.C.; Curtiss, R.; Ingraham, J.L.; Lin, E.C.C.; Low, K.B.; Magasanik, B.; Reznikoff, W.S.; Riley, M.; Schaechter, M. and Umbarger, H.E., eds) pp. 307–342, ASM Press, Washington, DC.

27. Bartkus, J.M. and Mortlock, R.P. (1986) Isolation of a mutation resulting in constitutive synthesis of L–fucose catabolic enzymes. *J.Bacteriol.* 165, 710–714.

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

29. Hammer, K. and Dandanell, G. (1989) The *deoR* Repressor from *E. coli* and its action in regulation–at–a–distance. *Nucleic Acids and Mol.Biol.* 3, 79–91.

30. Dandanell, G. and Hammer, K. (1985) Two operator sites separated by 599 base pairs are required for *deoR* repression of the *deo* operon of *Escherichia coli. EMBO J.* 4, 3333–3338.

31. Yang, B. and Larson, T.J. (1996) Action at a distance for negative control of transcription of the *glpD* gene encoding *sn*–glycerol 3–phosphate dehydrogenase of *Escherichia coli*. *J.Bacteriol.* 178, 7090–7098.

32. Yang, B.; Gerhardt, S.G. and Larson, T.J. (1997) Action at a distance for *glp* repressor control of *glpTQ* transcription in *Escherichia coli*. *Mol.Microbiol.* 24, 511–521.

33. Larson, T.J.; Cantwell, J.S. and van Loo–Bhattacharya, A.T. (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. Oehler, S.; Eismann, E.R.; Kramer, H. and Müller–Hill, B. (1990) The three operators of the *lac* operon cooperate in repression. *EMBO J.* 9, 973–979.

35. Mandal, N.; Su, W.; Haber, R.; Adhya, S. and Echols, H. (1990) DNA looping in cellular repression of transcription of the galactose operon. *Genes Dev.* 4, 410–418.

36. Plumbridge, J. and Kolb, A. (1993) DNA loop formation between Nag repressor molecules bound to its two operator sites is necessary for repression of the *nag* regulon of *Escherichia coli* in vivo. *Mol.Microbiol.* 10, 973–981.

37. Revet, B.; Von Wilcken–Bergmann, B.; Bessert, H.; Barker, A. and Müller–Hill, B. (1999) Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Curr.Biol.* 9, 151–154.

38. Zhao, N.; Oh, W.; Trybul, D.; Thrasher, K.S.; Kingsbury, T.J. and Larson, T.J. (1994) Characterization of the interaction of the *glp* repressor of *Escherichia coli* K–12 with single and tandem *glp* operator variants. *J.Bacteriol.* 176, 2393–2397.

39. van Rooijen, R. J. Characterization of the *Lactococcus lactis* lactose genes and regulation of their expression. Ph.D. Dissertation. 1993. Department of Biophysical Chemistry, Netherlands Institute for Dairy Research, Ede, The Netherlands.

40. Yoshida, K.I.; Shibayama, T.; Aoyama, D. and Fujita, Y. (1999) Interaction of a repressor and its binding sites for regulation of the *Bacillus subtilis iol* divergon. *J.Mol.Biol.* 285, 917–929.

41. van der Woude, M.W.; Hale, W.B. and Low, D.A. (1998) Formation of DNA methylation patterns: nonmethylated GATC sequences in *gut* and *pap* operons. *J.Bacteriol.* 180, 5913–5920.

42. van Rooijen R.J.; Gasson, M.J. and de Vos W.M. (1992) Characterization of the *Lactococcus lactis* lactose operon promoter: contribution of flanking sequences and LacR repressor to promoter activity. *J.Bacteriol.* 174, 2273–2280.

43. Saier, M.H., Jr. and Reizer, J. (1994) The bacterial phosphotransferase system: new frontiers 30 years later. *Mol.Microbiol.* 13, 755–764.

44. Charbit, A. and Autret, N. (1998) Horizontal transfer of chromosomal DNA between the marine bacterium *Vibrio furnissii* and *Escherichia coli* revealed by sequence analysis. *Microb.Comp.Genomics* 3, 119–132.

45. Plumbridge, J.A.; Cochet, O.; Souza, J.M.; Altamirano, M.M.; Calcagno, M.L. and Badet, B. (1993) Coordinated regulation of amino sugar–synthesizing and degrading enzymes in *Escherichia coli* K–12. *J.Bacteriol.* 175, 4951–4956.

46. Lengeler, J.W. (1977) Analysis of mutations affecting the dissimilation of galactitol (dulcitol) in *Escherichia coli* K–12. *Mol.Gen.Genet.* 152, 83–91.

47. Busby, S.; Kolb, A. and Buc, H. (1995) The *E. coli* cyclic AMP receptor protein. *Nucleic Acids and Mol.Biol.* 9, 177–191.

48. Zeng, G. and Larson, T.J. (1996) Use of engineered thrombin cleavage site for determination of translational reading frames. *BioTechniques* 21, 822–824.

49. Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1–I_2 regulatory elements. *Nucleic Acids Res.* 25, 1203–1210.

50. Cherepanov, P.P. and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: Tc^r and Km^r cassettes with the option of Flp–catalyzed excision of the antibiotic– resistance determinant. *Gene* 158, 9–14.

51. Podkovyrov, S.M. and Larson, T.J. (1995) A new vector–host system for construction of *lacZ* transcriptional fusions where only low–level gene expression is desirable. *Gene* 156, 151–152.

52. Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high–level expression of cloned genes. *J.Mol.Biol.* 189, 113–130.

53. Casadaban, M.J. (1976) Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J.Mol.Biol.* 104, 541–555.

54. Liss, L.R. (1987) New M13 host, DH5αF' competent cells. *Focus* 3, 13–14.

55. Podkovyrov, S.M. and Larson, T.J. (1996) Identification of promoter and stringent regulation of transcription of the *fabH*, *fabD* and *fabG* genes encoding fatty acid biosynthetic enzymes of *Escherichia coli*. *Nucleic Acids Res.* 24, 1747–1752.

56. Borck, K.; Beggs, J.D.; Brammar, W.J.; Hopkins, A.S. and Murray, N.E. (1976) The construction *in vitro* of transducing derivatives of phage lambda. *Mol.Gen.Genet.* 146, 199–207.

57. Singer, M.; Baker, T.A.; Schnitzler, G.; Deischel, S.M.; Goel, M.; Dove, W.; Jaacks, K.J.; Grossman, A.D.; Erickson, J.W. and Gross, C.A. (1989) A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. *Microbiol.Rev.* 53, 1–24.

58. Nichols, B.P.; Shafiq, O. and Meiners, V. (1998) Sequence analysis of Tn*10* insertion sites in a collection of *Escherichia coli* strains used for genetic mapping and strain construction. *J.Bacteriol.* 180, 6408–6411.

59. Maloy, S.R. and Nunn, W.D. (1981) Selection for loss of tetracycline resistance by *Escherichia coli*. *J.Bacteriol.* 145, 1110–1111.

60. Yang, B. and Larson, T.J. (1998) Multiple promoters are responsible for transcription of the *glpEGR* operon of *Escherichia coli* K–12. *Biochim.Biophys.Acta* 1396, 114–126.

61. Kohara, Y.; Akiyama, K. and Isono, K. (1987) The physical map of the whole *E. coli* chromosome. Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* 50, 495–508.

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

63. Simons, R.W.; Houman, F. and Kleckner, N. (1987) Improved single and multicopy *lac*–based cloning vectors for protein and operon fusions. *Gene* 53, 85–96.

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

65. Waldburger, C. and Susskind, M.M. (1994) Probing the informational content of *Escherichia coli* σ^{70} region 2.3 by combinatorial cassette mutagenesis. *J.Mol.Biol.* 235, 1489–1500.

66. Tabor, S. and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc.Natl.Acad.Sci.U.S.A.* 82, 1074–1078.

67. Link, A.J.; Phillips, D. and Church, G.M. (1997) Methods for generating precise deletions and insertions in the genome of wild–type *Escherichia coli*: application to open reading frame characterization. *J.Bacteriol.* 179, 6228–6237.

68. Sambrook, J.; Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

69. Higuchi, R. (1990) Recombinant PCR. *PCR Protocols. A Guide to Methods and Applications*. (Innis, M.A.; Gelfand, D.H.; Sninsky, J.J. and White, T.J., eds) pp. 177–183, Academic Press,Inc., San Diego, CA.

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

71. Sanger, F.; Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain–termination inhibitors. *Proc.Natl.Acad.Sci.U.S.A.* 74, 5463–5467.

72. DeLange, A.M. and McFadden, G. (1987) Efficient resolution of replicated poxvirus telomeres to native hairpin structures requires two inverted symmetrical copies of a core target DNA sequence. *J.Virol.* 61, 1957–1963.

73. DeLange, A.M.; Reddy, M.; Scraba, D.; Upton, C. and McFadden, G. (1986) Replication and resolution of cloned poxvirus telomeres *in vivo* generates linear minichromosomes with intact viral hairpin termini. *J.Virol.* 59, 249–259.

74. 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.

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

76. Zeng, G. Characterization of the *glpEGR* operon of *Escherichia coli.* Ph.D. Dissertation. 1996. Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA.

77. Zhang, X. and Ebright, R.E. (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.

78. Busby, S. and Ebright, R.E. (1997) Transcription activation at class II CAP–dependent promoters. *Mol.Microbiol.* 23, 853–859.

79. Sumper, M.; Berg, E.; Mengele, R. and Strobel, I. (1990) Primary structure and glycosylation of the S–layer protein of *Haloferax volcanii.* *J.Bacteriol.* 172, 7111–7118.

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

81. Weissenborn, D.L.; Wittekindt, N. and Larson, T.J. (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.

82. Schultz, S.C.; Shields, G.C. and Steitz, T.A. (1991) Crystal structure of a CAP–DNA complex: The DNA is bent by 90 degrees. *Science* 253, 1001–1007.

83. Yamano, N.; Matsushita, Y.; Kamada, Y.; Fujishima, S. and Arita, M. (1996) Purification and characterization of *N*–acetylglucosamine 6–phosphate deacetylase with activity against *N*–acetylglucosamine from *Vibrio cholerae* non–O1. *Biosci.Biotechnol.Biochem* 60, 1320–1323.

84. Zhou, Y.; Merkel, T.J. and Ebright, R.H. (1994) Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP). II. Role at class I and class II CAP–dependent promoters. *J.Mol.Biol.* 243 , 603–610.

85. Kogan, G.; Jann, B. and Jann, K. (1992) Structure of the *Escherichia coli* 0104 polysaccharide and its identity with the capsular K–9 polysaccharide. *FEMS Microbiol.Lett.* 70, 135–140.

86. Morris, J.A.; Stevens, A.E. and Sojka, W.J. (1977) Preliminary characterization of cell–free K–99 antigen isolated from *Escherichia coli* B41. *J.Gen.Microbiol.* 99, 353–357.

87. Plumbridge, J.A. (1990) Induction of the *nag* regulon of *Escherichia coli* by *N*–acetylglucosamine and glucosamine: role of the cyclic AMP–catabolite activator protein complex in expression of the regulon. *J.Bacteriol.* 172, 2728–2735.

88. Plumbridge, J. and Vimr, E. (1999) Convergent pathways for utilization of the amino sugars *N*–acetylglucosamine, *N*–acetylmannosamine, and *N*–acetylneuraminic acid by *Escherichia coli*. *J.Bacteriol.* 181, 47–54.

89. Saier, M.H., Jr. and Reizer, J. (1992) Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *J.Bacteriol.* 174, 1433–1438.

90. Estrela, A.I.; Pooley, H.M.; de Lencastre, H. and Karamata, D. (1991) Genetic and biochemical characterization of *Bacillus subtilis* 168 mutants specifically blocked in the synthesis of the teichoic acid poly(3–O–β–D–glucopyranosyl–*N*–acetylgalactosamine 1–phosphate): *gneA*, a new locus, is associated with UDP–*N*–acetylglucosamine 4–epimerase activity. *J.Gen.Microbiol.* 137, 943–950.

91. Ramsay, G. (1998) DNA chips: state–of–the art. *Nat.Biotechnol.* 16, 40–44.

92. Duggan, D.J.; Bittner, M.; Chen, Y.; Meltzer, P. and Trent, J.M. (1999) Expression profiling using cDNA microarrays. *Nat.Genet.* 21, 10–14.

93. Brown, P.O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat.Genet.* 21, 33–37.

94. Weickert, M.J. and Adhya, S. (1993) The galactose regulon of *Escherichia coli*. *Mol.Microbiol.* 10, 245–251.

95. Stevens, F.J. and Wu, T.T. (1976) Growth on D–lyxose of a mutant strain of *Escherichia coli* K–12 using a novel isomerase and enzymes related to D–xylose metabolism. *J.Gen.Microbiol.* 97, 257–265.

96. Stevens, F.J.; Stevens, P.W.; Hovis, J.G. and Wu, T.T. (1981) Some properties of D–mannose isomerase from *Escherichia coli* K–12. *J.Gen.Microbiol.* 124, 219–223.

97. Kunkel, T.A.; Roberts, J.D. and Zakour, R.A. (1987) Rapid and efficient mutagenesis without phenotypic selection. *Methods Enzymol.* 154, 368–382.

98. Konings, C.H. (1988) A kinetic procedure for the estimation of arginine in serum using arginine kinase. *Clin.Chim.Acta* 176, 185–193.

99. Bochner, B.R. (1989) Sleuthing out bacterial identities. *Nature* 339, 157–158.

100. Plunkett, G., III; Burland, V.; Daniels, D.L. and Blattner, F.R. (1993) Analysis of the *Escherichia coli* genome III. DNA sequence of the region from 87.2 to 89.2 minutes. *Nucleic Acids Res.* 21, 3391–3398.

101. Steele, M.I.; Lorenz, D.; Hatter, K.; Park, A. and Sokatch, J.R. (1992) Characterization of the *mmsAB* operon of *Pseudomonas aeruginosa* PAO encoding methylmalonate–semialdehyde dehydrogenase and 3–hydroxyisobutyrate dehydrogenase. *J.Biol.Chem.* 267, 13585–13592.

102. McFall, E. and Newman, E.B. (1996) Amino acids as carbon sources. *Escherichia coli and Salmonella. Cellular and Molecular Biology.* (Neidhardt, F.C.; Curtiss, R.; Ingraham, J.L.; Lin, E.C.C.; Low, K.B.; Magasanik, B.; Reznikoff, W.S.; Riley, M.; Schaechter, M. and Umbarger, H.E., eds) pp. 358–379, ASM Press, Washington, DC.

103. Berlyn, M.B. (1998) Linkage map of *Escherichia coli* K–12, edition 10: the traditional map. *Microbiol.Mol.Biol.Rev.* 62, 814–984.

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

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

106. Hofmann, K.; Bucher, P. and Kajava, A.V. (1998) A model of cdc25 phosphatase catalytic domain and cdk–interaction surface based on the presence of a rhodanese homology domain. *J.Mol.Biol.* 282, 195–208.

107. Vennesland, B.; Castric, P.A.; Conn, E.E.; Solomonson, L.P.; Volini, M. and Westley, J. (1982) Cyanide metabolism. *Fed.Proc.* 41, 2639–2648.

108. Westley, J. (1981) Thiosulfate: cyanide sulfurtransferase (rhodanese). *Methods Enzymol.* 77, 285–291.

109. Bonomi, F.; Pagani, S. and Kurtz, D.M., Jr. (1985) Enzymic synthesis of the 4Fe–4S clusters of *Clostridium pasteurianum* ferredoxin. *Eur.J.Biochem* 148, 67–73.

110. Pagani, S.; Bonomi, F. and Cerletti, P. (1984) Enzymic synthesis of the iron–sulfur cluster of spinach ferredoxin. *Eur.J.Biochem* 142, 361–366.

111. Pagani, S. and Galante, Y.M. (1983) Interaction of rhodanese with mitochondrial NADH dehydrogenase. *Biochim.Biophys.Acta* 742, 278–284.

112. Ploegman, J.H.; Drent, G.; Kalk, K.H.; Hol, W.G.; Heinrikson, R.L.; Keim, P.; Weng, L. and Russell, J. (1978) The covalent and tertiary structure of bovine liver rhodanese. *Nature* 273, 124–129.

113. Volini, M. and Westley, J. (1966) The mechanism of the rhodanese–catalyzed thiosulfate–lipoate reaction. Kinetic analysis. *J.Biol.Chem.* 241, 5168–5176.

114. Nagahara, N.; Okazaki, T. and Nishino, T. (1995) Cytosolic mercaptopyruvate sulfurtransferase is evolutionarily related to mitochondrial rhodanese. Striking similarity in active site amino acid sequence and the increase in the mercaptopyruvate sulfurtransferase activity of rhodanese by site–directed mutagenesis. *J.Biol.Chem.* 270, 16230–16235.

115. Nagahara, N. and Nishino, T. (1996) Role of amino acid residues in the active site of rat liver mercaptopyruvate sulfurtransferase. cDNA cloning, overexpression, and site–directed mutagenesis. *J.Biol.Chem.* 271, 27395–27401.

116. Luo, G.X. and Horowitz, P.M. (1994) The sulfurtransferase activity and structure of rhodanese are affected by site–directed replacement of Arg–186 or Lys–249. *J.Biol.Chem.* 269, 8220–8225.

117. Lijk, L.J.; Torfs, C.A.; Kalk, K.H.; de Maeyer, M.C. and Hol, W.G. (1984) Differences in the binding of sulfate, selenate and thiosulfate ions to bovine liver rhodanese, and a description of a binding site for ammonium and sodium ions. An X–ray diffraction study. *Eur.J.Biochem* 142, 399–408.

118. Brissette, J.L.; Weiner, L.; Ripmaster, T.L. and Model, P. (1991) Characterization and sequence of the *Escherichia coli* stress–induced *psp* operon. *J.Mol.Biol.* 220, 35–48.

119. Hama, H.; Kayahara, T.; Ogawa, W.; Tsuda, M. and Tsuchiya, T. (1994) Enhancement of serine–sensitivity by a gene encoding rhodanese–like protein in *Escherichia coli*. *J.Biochem (Tokyo)* 115, 1135–1140.

120. Colnaghi, R.; Pagani, S.; Kennedy, C. and Drummond, M. (1996) Cloning, sequence analysis and overexpression of the rhodanese gene of *Azotobacter vinelandii*. *Eur.J.Biochem* 236, 240–248.

121. Donadio, S.; Shafiee, A. and Hutchinson, C.R. (1990) Disruption of a rhodanese–like gene results in cysteine auxotrophy in *Saccharopolyspora erythraea*. *J.Bacteriol.* 172, 350–360.

122. Laudenbach, D.E.; Ehrhardt, D.; Green, L. and Grossman, A. (1991) Isolation and characterization of a sulfur–regulated gene encoding a periplasmically localized protein with sequence similarity to rhodanese. *J.Bacteriol.* 173, 2751–2760.

123. Weiland, K.L. and Dooley, T.P. (1991) Molecular cloning, sequencing and characterization of cDNA to rat liver rhodanese, a thiosulphate sulphurtransferase. *Biochem J.* 275, 227–231.

124. Fauman, E.B.; Cogswell, J.P.; Lovejoy, B.; Rocque, W.J.; Holmes, W.; Montana, V.G.; Piwnica–Worms, H.; Rink, M.J. and Saper, M.A. (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell* 93, 617–625.

125. Volini, M.; Ogata, K. and Alexander, K. (1983) Molecular forms of the rhodaneses. *Frontiers in Biochemical and Biophysical Studies of Proteins and Membranes.* (Liu, T.–Y.; Sakakibara, S.; Schechter, A.N.; Yagi, K.; Yajima, H. and Yasunobu, K.T., eds) pp. 183–192, Elsevier Science Publishing, Inc., New York, NY.

126. Alexander, K. and Volini, M. (1987) Properties of an *Escherichia coli* rhodanese. *J.Biol.Chem.* 262, 6595–6604.

127. Vachek, H. and Wood, J.L. (1972) Purification and properties of mercaptopyruvate sulfur transferase of *Escherichia coli*. *Biochim.Biophys.Acta* 258, 133–146.

128. Beacham, I.R. (1979) Periplasmic enzymes in gram–negative bacteria. *Int.J.Biochem.* 10, 877–883.

129. Lunn, C.A. and Pigiet, V.P. (1982) Localization of thioredoxin from *Escherichia coli* in an osmotically sensitive compartment. *J.Biol.Chem.* 257, 11424–11430.

130. El Yaagoubi, A.; Kohiyama, M. and Richarme, G. (1994) Localization of DnaK (chaperone 70) from *Escherichia coli* in an osmotic–shock sensitive compartment of the cytoplasm. *J.Bacteriol.* 176, 7074–7078.

131. Nandi, D.L. and Westley, J. (1998) Reduced thioredoxin as a sulfur–acceptor substrate for rhodanese. *Int.J.Biochem Cell Biol.* 30, 973–977.

132. Williams, C.H., Jr. (1995) Mechanism and structure of thioredoxin reductase from *Escherichia coli*. *FASEB J.* 9, 1267–1276.

133. Zheng, L.; White, R.H.; Cash, V.L.; Jack, R.F. and Dean, D.R. (1993) Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc.Natl.Acad.Sci.U.S.A.* 90, 2754–2758.

134. Zhao, N. Studies on the operator–repressor–effector interactions in the *glp* regulon of *Escherichia coli.* Ph.D. Dissertation. 1993. Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA.

135. Austin, T. D. Characterization of the *glpD* and *glpEGR* operons of *Escherichia coli* K–12. Ph.D. Dissertation. 1991. Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA.

136. Armstrong, J.B.; Adler, J. and Dahl, M.M. (1967) Nonchemotactic mutants of *Escherichia coli*. *J.Bacteriol.* 93, 390–398.

137. Sukharev, S.I.; Blount, P.; Martinac, B.; Blattner, F.R. and Kung, C. (1994) A large–conductance mechanosensitive channel in *E. coli* encoded by *mscL* alone. *Nature* 368, 265–268.

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

139. Johnson, B.H. and Hecht, M.H. (1994) Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology (N.Y.)* 12, 1357–1360.

140. Kaback, H. R. (1971) Bacterial membranes. *Methods Enzymol.* 22, 99–120.

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

142. Fraenkel, D.G. and Levisohn, S.R. (1967) Glucose and gluconate metabolism in an *Escherichia coli* mutant lacking phosphoglucose isomerase. *J.Bacteriol.* 93, 1571–1578.

143. Westley, J. and Heyse, D. (1971) Mechanisms of sulfur transfer catalysis. Sulfhydryl–catalyzed transfer of thiosulfonate sulfur. *J.Biol.Chem.* 246, 1468–1474.

144. Horowitz, P.M. and Hua, S. (1995) Rhodanese conformational changes permit oxidation to give disulfides that form in a kinetically determined sequence. *Biochim.Biophys.Acta* 1249, 161–167.

145. Dungan, J.M. and Horowitz, P.M. (1993) Thermally perturbed rhodanese can be protected from inactivation by self–association. *J.Protein Chem.* 12, 311–321.

146. Pecci, L.; Pensa, B.; Costa, M.; Cignini, P.L. and Cannella, C. (1976) Reaction of rhodanese with dithiothreitol. *Biochim.Biophys.Acta* 445, 104–111.

147. Aird, B.A. and Horowitz, P.M. (1988) Acceptor substrate–potentiated inactivation of bovine liver rhodanese. *J.Biol.Chem.* 263, 15270–15276.

148. Wang, S.F. and Volini, M. (1973) The interdependence of substrate and protein transformations in rhodanese catalysis. I. Enzyme interactions with substrate, product, and inhibitor anions. *J.Biol.Chem.* 248, 7376–7385.

149. Thorstenson, Y.R.; Zhang, Y.; Olson, P.S. and Mascarenhas, D. (1997) Leaderless polypeptides efficiently extracted from whole cells by osmotic shock. *J.Bacteriol.* 179, 5333–5339.

150. Ames, G.F.; Prody, C. and Kustu, S. (1984) Simple, rapid, and quantitative release of periplasmic proteins by chloroform. *J.Bacteriol.* 160, 1181–1183.

151. Birdsell, D.C. and Cota–Robles, E.H. (1967) Production and ultrastructure of lysozyme and ethylenediaminetetraacetate–lysozyme spheroplasts of *Escherichia coli*. *J.Bacteriol.* 93, 427–437.

152. Jacobson, G.R. and Rosenbusch, J.P. (1976) Abundance and membrane association of elongation factor Tu in *E. coli*. *Nature* 261, 23–26.

153. Pagani, S.; Bonomi, F. and Cerletti, P. (1983) The inhibition of rhodanese by lipoate and iron–sulfur proteins. *Biochim.Biophys.Acta* 742, 116–121.

154. Hidalgo, E. and Demple, B. (1996) Activation of SoxR–dependent transcription in vitro by noncatalytic or NifS–mediated assembly of [2Fe–2S] clusters into apo–SoxR. *J.Biol.Chem.* 271, 7269–7272.

155. Ding, H. and Demple, B. (1998) Thiol–mediated disassembly and reassembly of [2Fe–2S] clusters in the redox–regulated transcription factor SoxR. *Biochemistry* 37, 17280–17286.

156. Ploegman, J.H.; Drent, G.; Kalk, K.H. and Hol, W.G. (1979) The structure of bovine liver rhodanese. II. The active site in the sulfur–substituted and the sulfur–free enzyme. *J.Mol.Biol.* 127, 149–162.

157. Ploegman, J.H.; Drent, G.; Kalk, K.H. and Hol, W.G. (1978) Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. *J.Mol.Biol.* 123, 557–594.

158. Link, A.J.; Robison, K. and Church, G.M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K–12. *Electrophoresis* 18, 1259–1313.

159. Ajouz, B.; Berrier, C.; Garrigues, A.; Besnard, M. and Ghazi, A. (1998) Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells. *J.Biol.Chem.* 273, 26670–26674.

160. Miranda–Vizuete, A.; Rodriguez–Ariza, A.; Toribio, F.; Holmgren, A.; Lopez–Barea, J. and Pueyo, C. (1996) The levels of ribonucleotide reductase, thioredoxin, glutaredoxin 1, and GSH are balanced in *Escherichia coli* K–12. *J.Biol.Chem.* 271, 19099–19103.

161. Whitehouse, D.B.; Pilz, A.J.; Porta, G. and Hopkinson, D.A. (1988) Rhodanese isozymes in human tissues. *Ann.Hum.Genet.* 52, 1–10.

162. Kredich, N.M. (1996) Biosynthesis of cysteine. *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology.* (Neidhardt, F.C.; Curtiss, R.; Ingraham, J.L.; Lin, E.C.C.; Low, K.B.; Magasanik, B.; Reznikoff, W.S.; Riley, M.; Schaechter, M. and Umbarger, H.E., eds) pp. 514–527, ASM Press, Washington, DC.

163. Filutowicz, M.; Wiater, A. and Hulanicka, D. (1982) Delayed inducibility of sulphite reductase in *cysM* mutants of *Salmonella typhimurium* under anaerobic conditions. *J.Gen.Microbiol.* 128, 1791–1794.

164. Bonomi, F.; Pagani, S.; Cerletti, P. and Cannella, C. (1977) Rhodanese–mediated sulfur transfer to succinate dehydrogenase. *Eur.J.Biochem* 72, 17–24.

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

166. Robinson, J.J. and Weiner, J.H. (1980) The effect of amphipaths on the flavin–linked aerobic glycerol–3–phosphate dehydrogenase from *Escherichia coli*. *Can.J.Biochem* 58, 1172–1178.

167. Cole, E.S.; Lepp, C.A.; Holohan, P.D. and Fondy, T.P. (1978) Isolation and characterization of flavin–linked glycerol–3–phosphate dehydrogenase from rabbit skeletal muscle mitochondria and comparison with the enzyme from rabbit brain. *J.Biol.Chem.* 253, 7952–7959.

168. Garrib, A. and McMurray, W.C. (1986) Purification and characterization of glycerol–3–phosphate dehydrogenase (flavin–linked) from rat liver mitochondria. *J.Biol.Chem.* 261, 8042–8048.

169. Choi, Y.–L.; Kawase, S.; Kawamukai, M.; Sakai, H. and Komano, T. (1991) Regulation of *glpD* and *glpE* gene expression by a cyclic AMP–cAMP receptor protein (cAMP–CRP) complex in *Escherichia coli*. *Biochim.Biophys.Acta* 1088, 31–35.

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

171. Miranda–Vizuete, A.; Damdimopoulos, A.E.; Gustafsson, J. and Spyrou, G. (1997) Cloning, expression, and characterization of a novel *Escherichia coli* thioredoxin. *J.Biol.Chem.* 272, 30841–30847.

172. Vlamis–Gardikas, A.; Åslund, F.; Spyrou, G.; Bergman, T. and Holmgren, A. (1997) Cloning, overexpression, and characterization of glutaredoxin 2, an atypical glutaredoxin from *Escherichia coli*. *J.Biol.Chem.* 272, 11236–11243.

173. Åslund, F. and Beckwith, J. (1999) The thioredoxin superfamily: redundancy, specificity, and gray–area genomics. *J.Bacteriol.* 181, 1375–1379.

174. Lillig, C.H.; Prior, A.; Schwenn, J.D.; Åslund, F.; Ritz, D.; Vlamis–Gardikas, A. and Holmgren, A. (1999) New thioredoxins and glutaredoxins as electron donors of 3'–phosphoadenylylsulfate reductase. *J.Biol.Chem.* 274, 7695–7698.

175. Pedrajas, J.R.; Kosmidou, E.; Miranda–Vizuete, A.; Gustafsson, J.A.; Wright, A.P. and Spyrou, G. (1999) Identification and functional characterization of a novel mitochondrial thioredoxin system in *Saccharomyces cerevisiae*. *J.Biol.Chem.* 274, 6366–6373.

176. Spyrou, G.; Enmark, E.; Miranda–Vizuete, A. and Gustafsson, J. (1997) Cloning and expression of a novel mammalian thioredoxin. *J.Biol.Chem.* 272, 2936–2941.

177. Flint, D.H. (1996) *Escherichia coli* contains a protein that is homologous in function and N–terminal sequence to the protein encoded by the *nif*S gene of *Azotobacter vinelandii* and that can participate in the synthesis of the Fe–S cluster of dihydroxy–acid dehydratase. *J.Biol.Chem.* 271, 16068–16074.

178. Fukumori, Y.; Hoshiko, K. and Yamanaka, T. (1989) Purification and some properties of thiosulphate–cleaving enzyme from *Thiobacillus novellus*. *FEMS Microbiol.Lett.* 53, 159–163.

William Keith Ray

Personal

Date of Birth: June 21, 1971

Place of Birth: Richlands, Virginia

Education

1995–1999

Virginia Polytechnic Institute and State University, Blacksburg, VA

Ph.D., biochemistry

- Date of completion: August 1999.
- Committee chair: Timothy J. Larson, Ph.D.

1989–1994

Virginia Polytechnic Institute and State University, Blacksburg, VA

B.S., chemical engineering, minor in chemistry

Honors

- Horace E. and Elizabeth F. Alphin agricultural sciences graduate student scholarship, 1998.
- Bruce M. Anderson award for outstanding first year graduate student in the department of biochemistry, 1996.

Publications

Ray, W. and Larson T. (1999) Characterization of AgaR and the Regulation of N–Acetylgalactosamine Metabolism in *Escherichia coli*. In preparation.

Ray, W.; Zeng, G.; Potters, M.; Mansuri, A. and Larson, T. (1999) Characterization of a 12kDa Rhodanese Encoded by *glpE* of *Escherichia Coli*. In preparation.

Niehaus, W.; White, R.; Richardson, S.; Bourne, A. and Ray, W. (1995) Polyethylene sulfonate: a tight–binding inhibitor of 6–phosphogluconate dehydrogenase of *Cryptococcus neoformans*. *Arch Biochem Biophys*. 324(2), 325–330.

Experience

- Gene amplification, cloning and overexpression.
- Characterization of DNA–binding proteins.
- Enzyme purification and characterization.