Mol Gen Genet (1987) 208:247–253

MGG

Sequence of the *mglB* gene from *Escherichia coli* K12: Comparison of wild-type and mutant galactose chemoreceptors

Annette Scholle, Jörg Vreemann, Volker Blank, Annette Nold, Winfried Boos, and Michael D. Manson*
Department of Biology, University of Konstanz, D-7750 Konstanz, Federal Republic of Germany

Summary. The *mglB* gene of *Escherichia coli* codes for a galactose-binding protein (GBP) that serves both as the galactose chemoreceptor and as the recognition component of the β-methylgalactoside transport system. The *mglB*551 mutation eliminates the chemotactic function of GBP without altering its transport or substrate-binding properties. To investigate the interaction between GBP and Trg, the chemotactic signal transducer for galactose, we sequenced the *mglB* genes from wild-type and *mglB*551 mutant strains. The mutation causes the replacement of Gly_{74} of GBP by Asp. This residue is located in alpha-Helix III at the tip of the P domain in the GBP tertiary structure farthest removed from the substrate-binding cleft between the P and Q domains. We conclude that Helix III must be part of, or at least adjacent to, the recognition site for Trg. Our sequence also included part of the *mglA* gene, which is immediately distal to *mglB*. The amino acid sequence deduced for the beginning of the MglA protein showed homology with a family of polypeptides that contain an ATP-binding site and are components of binding-protein-dependent transport systems.

Key words: Chemoreceptors – DNA Sequence – MglB – Trg – MglA

Introduction

The periplasmic galactose-binding protein (GBP) of *Escherichia coli* (Anraku 1967; Boos 1969) was the first specific chemoreceptor to be identified in bacteria (Hazelbauer and Adler 1971). This protein has a dual function, since it also serves as the recognition component of the β-methylgalactoside (Mgl) transport system. Although methylgalactosides are actually rather poor substrates, the system does transport galactose and glucose with high affinity (on the order of 1 μ*M*).

The synthesis of *mgl* gene products is repressed by glucose or methyl-1-thio-β-D-galactopyranoside and is induced by the non-metabolizable galactose analog D-fucose (Lengeler et al. 1971). The first structural gene in the *mgl* operon, *mglB*, codes for GBP (Ordal and Adler 1974a, b). It is synthesized as a precursor from which the initial 23 residues, the leader peptide, are cleaved during export into the periplasmic space (Scripture and Hogg 1983). Two further genes, *mglA* and *mglC*, code for proteins that are cytoplasmic membrane components of the transport system (Rotman and Guzman 1982; Harayama et al. 1983). In the closely related bacterium *Salmonella typhimurium*, the *mgl* operon appears to contains an additional gene, *mglE*, whose product is presumably also required for transport (Müller et al. 1985). It is not known whether *mglE* exists in *E. coli*. Thus, Mgl appears to conform to the pattern for binding-protein-dependent transport systems, which generally consist of one or several soluble periplasmic components (the binding proteins themselves) and three inner-membrane transport components (Ames 1986).

Ordal and Adler (1974a, b) found that *mglA* and *mglC* mutants lack galactose transport but that they retain galactose chemotaxis. In contrast, most *mglB* mutations lead to a decrease or loss of both transport and chemotaxis. The *mglB*551 mutation, however, causes a nearly complete loss of galactose taxis but leaves the galactose-binding and transport functions of GBP intact (Ordal and Adler 1974b). Strains containing *mglB*551 produce normal amounts of GBP, but on isoelectric focussing gels the mutant protein exhibits a shift of 0.4 pH units relative to wild-type GBP, from a pI value of 5.5 to 5.1 (Boos et al. 1981; Engelhardt-Altendorf 1983).

When substrate binds to GBP the protein undergoes a conformational change (Boos 1972; Zukin et al. 1977). In this altered conformation GBP can interact with the Trg protein (Kondoh et al. 1979; Hazelbauer et al. 1981) to initiate a chemotactic signal. Trg, one of four chemotactic signal transducers identified in *E. coli* (Hazelbauer and Harayama 1983), is localized in the cytoplasmic membrane and mediates chemotaxis to galactose and ribose, the latter via the ribose-binding protein (RBP).

Since *mglB*551 mutants are specifically impaired in chemotaxis, GBP coded by this allele may be defective in its interaction with Trg. In order to identify a domain of GBP critical for this interaction, we sequenced the wild-type and mutant *mglB* genes and located the *mglB*551 mutation within the DNA sequence.

Materials and methods

Reagents and enzymes. α-[^{32}P]dATP and deoxyadenosine 5′-α-[^{35}S]thiotriphosphate were purchased from Amersham Corp. D-[^{14}C]galactose was obtained from New England

* *Present address:* Department of Biology, Texas A & M University, College Station, TX 77843, USA

Offprint requests to: M. Manson

Nuclear. Dideoxynucleotides, restriction endonucleases, T_4 DNA ligase and the Klenow fragment of *E. coli* DNA polymerase were purchased from Boehringer Mannheim. All other chemicals were of reagent grade.

Media. Cells were grown in minimal medium A (MMA) supplemented with a carbon source or in LB medium (Miller 1972). Our LB medium contained glucose only when the addition of the sugar is indicated. Galactose swarm plates contained 0.3% Bacto-Agar (Difco Laboratories), motility salts [100 m*M* potassium phosphate, pH 7.0, 10 m*M* $(NH_4)_2SO_4$, 1 m*M* $MgSO_4$, 0.5% NaCl, 0.5 µg/ml $FeCl_3$], 1 µg/ml thiamine and 50 µ*M* galactose. When used, tetracycline and ampicillin were present at 5 and 50 µg/ml, respectively.

Bacterial strains. MM318 is a derivative of the *E. coli* K12 strain MC4100 (Casadaban 1976). MM318 is *flbB*$^+$ and carries a Tn*10* insertion in the *cir* gene, which is 70% cotransducible with *mgl* (Middendorf et al. 1984). Strains MM319 and MM320 are isogenic with MM318 except that they carry the *mglB*551 allele from K12 strain AW551 and the *mglB*502 allele from K12 strain LA6022 (Brass et al. 1983), respectively. Strain LA5709 was used for plasmid isolation and propagation. LA5709 is Mgl$^-$ and does not synthesize GBP; it is therefore unable to transport galactose in the micromolar concentration range (Müller et al. 1985). M13 phage isolation and propagation were done in KK2186, a strain isogenic with JM103 (Messing et al. 1981) except that it has been cured of phage P1.

Isolation of galactose taxis-positive revertants. Spontaneous revertants were isolated on swarm plates by a procedure similar to that described for maltose chemotaxis mutants (Manson and Kossmann 1986). Cells from a 5 ml overnight culture of strain MM319 in LB medium were concentrated in 50 µl after washing twice with 0.9% NaCl and applied with an Eppendorf pipette in a 5 cm long "ditch" on a galactose-tetracycline swarm plate. Revertants were identified by their ability to form swarm rings extending out from the streak.

Preparative techniques. DNA manipulations were carried out as described by Maniatis et al. (1982) or as recommended by the supplier of restriction enzymes (Boehringer Mannheim). Shock fluid containing GBP was prepared by cold osmotic shock (Neu and Heppel 1965) from cells grown overnight in 1 l of LB medium containing 1 m*M* D-fucose.

SDS-polyacrylamide slab gel electrophoresis. Electrophoresis on sodium dodecyl sulfate (SDS)-polyacrylamide (12%) gels was done by the method of Laemmli (1970). Samples were heated at 100° C for 5 min before loading on the gel.

Isoelectric focussing. Crude shock fluid was analyzed by isoelectric focussing on analytical Ampholine (LKB Instruments, Inc.) polyacrylamide gel electrophoresis plates (pH range 4.0–6.5) according to the recommendations of the manufacturer.

In vivo galactose-binding assay. The galactose-binding activity of whole cells was measured with strain LA5709 carrying various plasmids. Cells were grown overnight at 37° C in LB medium containing 0.5% glucose and 50 µg/ml ampicillin and then washed twice with 0.9% NaCl. About 10^9 washed cells were used to inoculate 5 ml LB medium containing 50 µg/ml ampicillin and 1 m*M* D-fucose. These cultures were grown at 37° C for 5–6 h to a cell density of about 2×10^9/ml. The cells were then washed with 0.9% NaCl and resuspended in 5 ml MMA. D-[^{14}C]galactose (specific activity 321 mCi/mmol) was added to 100 µl of the cell resuspension to a final concentration of 0.3 µ*M*. After 2 min incubation at room temperature the entire sample was filtered onto nitrocellulose membrane filters (0.45 µm pore size; Millipore Corp.) and washed quickly with 10–15 ml MMA. The radioactivity on the dried filters was counted in a toluene-based scintillation fluid (Lipoluma; Lumac) in a liquid scintillation counter.

Reconstitution of high-affinity galactose transport in calcium-permeabilized cells. We used a modified version (Dahl and Manson 1985) of the standard procedure for reconstitution of binding-protein-dependent transport systems in whole cells (Brass et al. 1983; Brass and Manson 1984). This method is based on the permeabilization of the bacterial outer membrane by treatment with calcium. Cells were grown in MMA medium containing 0.5% glycerol, 0.2% casaminoacids, 1 µg/ml thiamine, 1 m*M* galactose, 1 m*M* $MgSO_4$ and 0.5 µg/ml $FeCl_3$. One milligram of lyophilized, powdered crude shock fluid, isolated from plasmid-containing strains that overproduce GBP, was used for reconstitution of 1×10^9 cells. The final GBP concentration was about 10 mg/ml (ca. 300 µ*M*).

The initial rate of galactose transport was assayed by a procedure essentially identical to that described previously for maltose transport (Brass et al. 1983). D-[^{14}C]galactose (specific activity 321 mCi/mmol) was used at 0.1 µ*M*.

Sequencing and sequence analysis. Appropriate DNA restriction fragments were cloned into M13mp8 and M13mp9 vectors (Messing and Vieira 1982) and sequenced by the dideoxy chain termination method of Sanger et al. (1977), as described in the laboratory handbooks "M13 cloning 'dideoxy' sequencing manual" and "M13 cloning and sequencing handbook" of Bethesda Research Laboratories (BRL) and Amersham, respectively. The "universal" 15 bp *lacZ* primer (200 ng/100 µl; BRL) and several custom-synthesized oligonucleotide primers were used. DNA fragments were separated on 5.5% denaturing acrylamide/bisacrylamide (19:1) sequencing gels ($0.3 \times 20 \times 40$ cm). DNA sequences were analyzed with a Macintosh Computer (Apple Computer, Inc.) using the DNA Inspector II computer program of Textco.

Results

Construction of mgl *plasmids*

Phage λp*mgl'lacZ* (Middendorf et al. 1984) is an *att*$^-$ Lac$^+$ transducing phage derived from λp*lac*Mu1 (Bremer et al. 1984). The phage contains an *mglA'-'lacZ* translational fusion constructed in strain AW551, which carries the *mglB*551 mutation. This phage will be referred to as λp*mglB*551. It was used to lysogenize the *mglB*$^+$ strain MM318. The phage integrates and excises from the *mgl* region of the host chromosome via homologous recombination. Phages spontaneously released from several indepen-

dent Lac$^+$ lysogens were obtained from the supernatants of overnight cultures of the lysogens. Recombinant *mglB*$^+$ phages were identified by using the released phages to make Lac$^+$ lysogens of strain MM320 (GBP$^-$) and testing the lysogens for galactose chemotaxis on swarm plates. One such phage was designated λp*mglB*$^+$.

The 6.9 kb *Bam*HI restriction fragments from phages λp*mglB*551 and λp*mglB*$^+$ were subcloned into the *Bam*HI site of plasmid pBR322 (Bolivar et al. 1977). The resulting 11.2 kb plasmids were called pB23-551 and pB23-wt. These plasmids contain pBR322 DNA, from 2.8 to 3.7 kb of λ DNA and 3.0 to 3.9 kb of chromosomal DNA, and 117 bases from the S end of phage Mu. (About 0.9 kb at the junction of λ and chromosomal DNA could come from either source.) The 3.5 kb *Sma*I-*Bam*HI DNA fragments from the two pB23 plasmids were subcloned into the *Bam*HI-*Pvu*II sites of pBR322 to generate the 5.8 kb plasmids pVB2-551 and pVB2-wt. These plasmids contain only pBR322 and chromosomal DNA. The 4.3 kb plasmids pVB1-551 and pVB1-wt were constructed by subcloning the 2.2 kb *Eco*RI-*Sma*I fragments from the two pB23 plasmids into the *Eco*RI-*Pvu*II sites of pBR322.

The two pB23 plasmids and the two pVB2 plasmids contain the entire *mglB* gene. The single *Eco*RI restriction site in *mglB* is located at the extreme 3′ end of the gene (Rotman and Guzman 1982), 12 bp before the stop codon (see Fig. 3). In the two pVB1 plasmids the *mglB* gene is fused at its *Eco*RI site to an open reading frame of 14 nucleotide triplets from pBR322 DNA (in the direction of the *bla* gene) until a TAA stop codon is reached. Thus, the pVB1 plasmids code for a GBP lacking the 3 carboxy-terminal residues of wild-type GBP (Ser Lys$^+$ Lys$^+$) and having 14 new residues coded by pBR322 (Leu Lys$^+$ Thr Lys$^+$ Gly Pro Arg$^+$ Asp$^-$ Arg$^+$ Pro Ile Phe Leu Gly), leading to a net gain of 11 amino acids and 1 positive charge. Analysis on an SDS-polyacrylamide gel (data not shown) indicated that GBP coded by the two pVB1 plasmids (M_r of 34000) and GBP coded by the two pB23 plasmids (M_r of 32500) differ by about 1500 daltons.

On an isoelectric focussing gel (Fig. 1) GBP coded by the *mglB*551 and wild-type plasmid series had the expected pI values of pH 5.1 and pH 5.5, respectively. GBP from the pVB1-551 and pVB1-wt plasmids had isoelectric points nearly identical to those of the corresponding intact proteins, although the predicted C-termini of GBP from the pVB1 plasmids contain one more positive charge.

Galactose-binding assay and galactose transport in reconstituted cells

There was no significant difference in the galactose-binding activity of whole cells of strain LA5709 containing the pVB1-wt, pVB1-551, pVB2-wt and pVB2-551 plasmids. 50 ± 5 pmol galactose were bound by 1×10^8 cells, corresponding to 3×10^5 molecules per cell. Since the K_d of GBP for galactose is about 1 μ*M*, and the concentration used in the binding test was 0.3 μ*M*, the number of GBP molecules per cell must be on the order of 10^6, at least 40-fold higher than the amount found in induced wild-type cells (Koman et al. 1979).

We could not measure transport of galactose at micromolar concentrations in plasmid-bearing strains that overproduce GBP because the large amount of galactose bound in the periplasm created very high background levels of radioactivity. We therefore measured transport activity in cells of strain MM320 (GBP$^-$) reconstituted with 300 μ*M* GBP. (The concentration of GBP in the periplasm of induced wild-type cells is roughly 600 μ*M*, estimated from the data of Koman et al. 1979). At a galactose concentration of 0.1 μ*M*, cells reconstituted with GBP from plasmids pVB2-wt, pVB2-551, pVB1-wt and pVB1-551 had initial transport rates of 20, 25, 10 and 12 pmol/min per 2.5×10^8

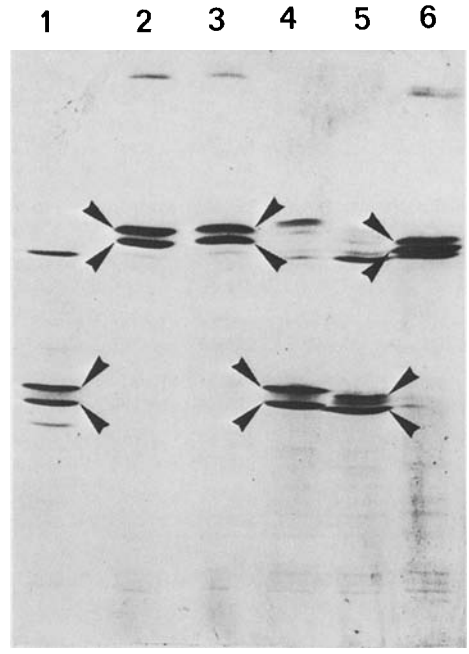

Fig. 1. Isoelectric focussing of plasmid-coded galactose-binding protein (GBP). Samples of crude shock fluid containing approximately equal amounts of protein were applied to each lane. Shock fluids were prepared from strain LA5709 containing the indicated plasmids. The gel was stained with Coomassie Blue. The pH range ran from 6.5 at the top to 4.0 at the bottom. On such gels GBP forms at least two bands in the absence of added substrate. These bands are indicated by *arrows* for each lane. The pI values given in the text are for the midpoint between the two bands. Lane 1, pB23-551; lane 2, pB23-wt; lane 3, pVB2-wt; lane 4, pVB2-551; lane 5, pVB1-551; lane 6, pVB1-wt

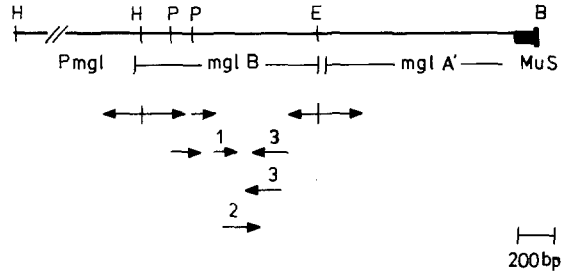

Fig. 2. Sequencing strategy. Above is shown a ca. 3.5 kb region of the *Escherichia coli* chromosome extending from a *Hpa*I site ca. 1 kb in front of the *mgl* operon to the *Bam*HI site at the *mglA′-′lacZ* fusion joint. The *thick black bar* indicates the 117 bp of DNA from the S end of phage Mu. The extents of *mglB* and the *mglA′* fragment are shown below the chromosome. The *length* of the *arrows* delineates sequences that could be read off individual gels. *Arrows* with numbers refer to sequences obtained using synthesized primers; the *two arrows* for Primer 3 represent two different sequencing runs. Relevant restriction sites are indicated: B, *Bam*HI; E, *Eco*RI; H, *Hpa*I; P, *Pvu*II. Not shown is a *Sma*I site that lies about 0.1 kb to the right of the leftmost *Hpa*I site

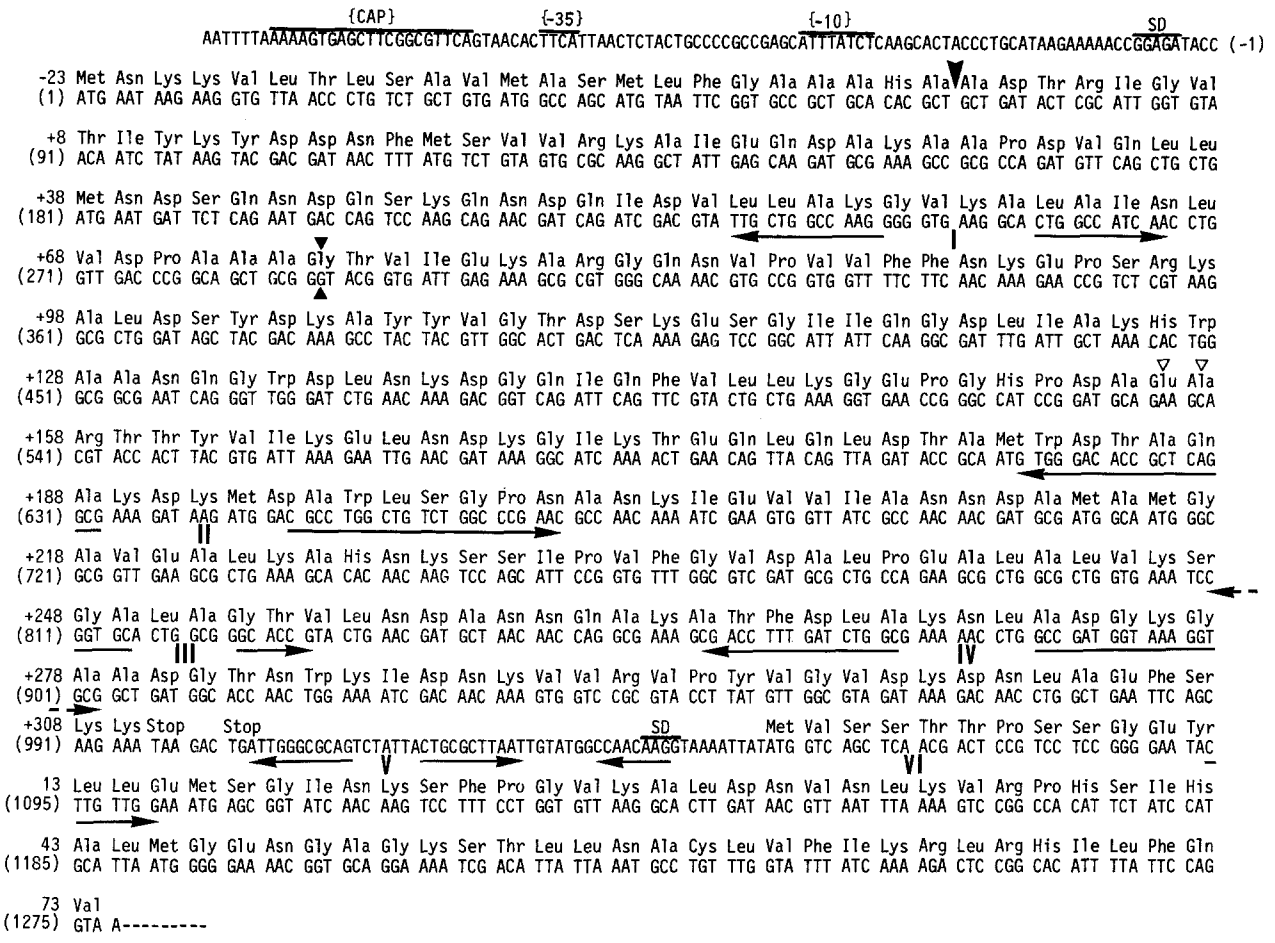

Fig. 3. Sequence of 1384 nucleotides from *mglB* and the surrounding chromosomal DNA from *Escherichia coli* K12. The nucleotide numbering (*in parentheses*) begins with the first base of the ATG start codon for *mglB*; the preceding sequence is designated by negative numbers reading backward from this point. The sequence after nucleotide 18 was determined for *mglB*$^+$ and *mglB*551; the sequence 5′ to this point was determined only for *mglB*$^+$. The amino acid sequence is numbered (without parentheses) divergently from the leader peptide cleavage site (*large arrowhead*); the leader peptide is assigned negative numbers. A second open reading frame, starting with an ATG triplet, begins at nucleotide 1059 and extends for 220 nucleotides to the end of our sequence. The amino acids of this reading frame are numbered 1 through 73. Probable Shine-Dalgarno (SD) sequences are indicated for *mglB* and *mglA*. We have also indicated possible locations for the −35 and −10 sites of the *mgl* promoter as well as a sequence having moderate homology to the consensus sequence for a CAP binding site. The *closed triangles* identify the nucleotide and amino acid altered by the *mglB*551 mutation (*mglB*$^+$, GGT=Gly; *mglB*551, GAT=Asp). The *open triangles* identify the two residues that differ from the protein sequence determined for *E. coli* strain B/r (Mahoney et al. 1981). Dyad symmetries in the *mgl* sequence are indicated by *diverging arrows* below the nucleotide sequence and are identified by Roman numerals. These sequences could form stem-loop structures in the *mgl* mRNA, with the indicated ΔG values: I, −14.0 kcal; II, −25.4 kcal; III, −12.6 kcal; IV, −17.6 kcal; V, −16.6 kcal; VI, −9.4 kcal. Values of ΔG were calculated according to Tinoco et al. (1973)

cells, respectively. Transport in strain MM320 without added GBP was negligible. For comparison, calcium-treated cells of the GBP$^+$ strain MM318 had a rate of 30 pmol/min per 2.5×10^8 cells. Thus, GBP coded by all four plasmids functions in high-affinity galactose transport, although GBP from the two pVB1 plasmids may be slightly less active.

Analysis of galactose taxis-positive revertants

To test whether the *mglB*551 mutation was reversible we looked for spontaneous revertants of strain MM319 on galactose swarm plates. Many revertants were found that formed swarms of the same morphology and size as the GBP$^+$ strain MM318. GBP from two independently isolated revertants had the same isoelectric point as wild-type GBP (data not shown), but we do not know whether the proteins from the revertants are identical to each other or to the wild-type protein.

Sequencing of the mglB *region*

To locate the *mglB*551 mutation we sequenced the wild-type and mutant genes. The sequencing strategy is shown in Fig. 2. The DNA sequence and the amino acid sequence predicted from it are presented in Fig. 3. The coding region of *mglB* comprises an open reading frame of 996 nucleotides and ends with a TAA stop codon at nucleotides 997–999 followed after one intervening triplet by a second stop codon (TGA). Within *mglB* codon usage frequencies corresponded to those for highly expressed genes (Grosjean and Fiers 1982), and none of the seven "modulating codons"

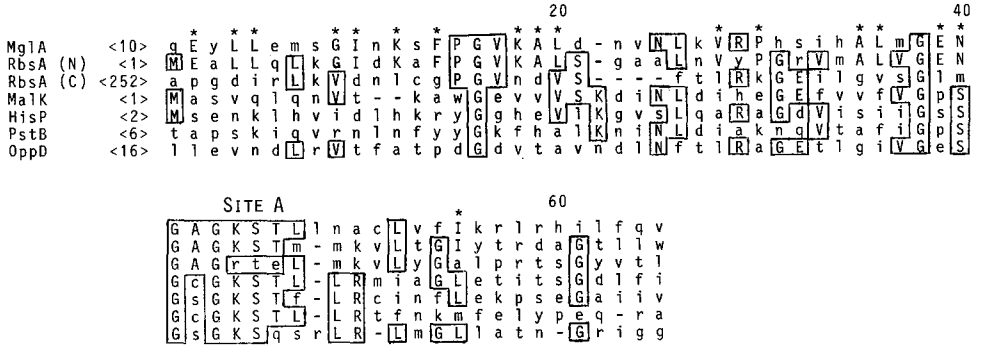

Fig. 4. Homologous sequences in a series of transport proteins. Data for all proteins except MglA are taken from Higgins et al. (1986). The sequences are aligned to aid visualization of the homology. Except for RbsA the sequences all come from the N-termini of the respective proteins; the RbsA protein apparently arose by a tandem duplication event, so that the homologous sequences appear in both the N-terminal and C-terminal halves of the polypeptide. Numbers in brackets give the first residue shown for each protein. Amino acids appearing at a given position in at least three of the polypeptides are given by upper-case letters and are *boxed*. Amino acids appearing in the same position only in MglA and the N-terminal portion of RbsA are given in upper-case letters and are marked by *asterisks*. All other residues are given in lower case letters. Site A is a highly conserved region that is thought to be involved in nucleotide binding

were present. The only difference in the wild-type and mutant DNA sequences was at nucleotide 290, resulting in the conversion of the GGT triplet that codes for Gly_{74} in the wild-type protein to GAT, which codes for Asp.

Distal to *mglB* a stretch of 62 nucleotides (including the *mglB* stop codons and a probable Shine-Dalgarno sequence) was followed by an ATG triplet that initiates an open reading frame of 220 nucleotides extending to the end of our sequence. We believe these nucleotides correspond to the beginning of the *mglA* gene.

Discussion

The amino acid sequence deduced from our DNA sequence for the *mglB* gene from *E. coli* K12 agrees with the 309 residue sequence determined for GBP from *E. coli* strain B/r (Mahoney et al. 1981) except at residues 156 and 157, where we find Glu-Ala instead of Lys-Glu. This is a region of the gene that we sequenced from both strands. Also, the DNA sequence for the *mglB* gene from *S. typhimurium* predicts that residues 156–157 of GBP from that species should be Glu-Ala (D. Benner, personal communication). Thus, we believe that our assignment of amino acid residues for *E. coli* K12 is likely to be correct, whatever the situation may be in *E. coli* B/r. At nucleotide 160 we found C instead of G as reported previously for *E. coli* K12 (Scripture and Hogg 1983). This exchange would convert a GCG triplet into a GCC triplet, both of which code for Ala.

Since we do not know where transcription of the *mgl* operon begins it is impossible to define promoter elements unambiguously. Nevertheless, in Fig. 3 we have indicated possible locations of the −35 and −10 regions. Also shown is a 21 nucleotide stretch that could serve as a binding site for CAP (cyclic AMP catabolite gene activator protein). This assignment is based on the moderate homology of this region with the published consensus sequence for CAP binding sites (Ebright et al. 1984).

The presumptive ATG start codon of *mglA* (nucleotides 1059–1061) is preceded by a Shine-Dalgarno sequence (AAGG; nucleotides 1046–1049). Based on this partial sequence, MglA is one of a class of polypeptides characteristic of binding-protein-dependent transport systems (Fig. 4). These polypeptides contain an ATP binding site (Higgins et al. 1986). Also, like another protein of this group, MalK (Bavoil et al. 1980; Shuman 1981), MglA is rather loosely associated with the cytoplasmic membrane (Müller et al. 1985). MglA shows more homology with the N-terminus of RbsA than with MalK or the corresponding components from other transport systems (Fig. 4).

Another feature of the partial *mglA* sequence is the rather frequent (6 out of 73 triplets) occurrence of modulating codons (Grosjean and Fiers 1982). These codons may be rate-limiting for translation and are generally a feature of weakly expressed genes. Their abundance in *mglA* relative to *mglB* may contribute to the overproduction of GBP relative to the MglA protein (Müller et al. 1985).

Within the *mglB* coding region we identified four inverted repeats that could form stable stem-loop structures ($\Delta G < -10$ kcal/mol) in the mRNA (Fig. 3). Three of these dyad symmetries, which occur late in the gene, could potentially be involved in coupling the synthesis and export of GBP in a manner similar to that suggested for maltose-binding protein (MBP) and its structural gene, *malE* (Randall et al. 1980; Duplay et al. 1984). Another hairpin loop can be formed in the mRNA in the region immediately following the termination codons for *mglB*; its function, if any, in the regulation of *mgl* gene expression is unknown.

Sequences at nucleotides 1041–1048 and 1094–1101 could also generate a stem-loop structure ($\Delta G = -9.4$ kcal), which overlaps the proposed Shine-Dalgarno sequence for *mglA*. Mutations creating hairpin loops with ΔG values of −12.2 kcal and −8.4 kcal (*lamB*$^+$, −2 kcal) reduce *lamB* expression 25- and 4-fold, respectively, presumably because their stems include the Shine-Dalgarno sequence of the gene and limit ribosomal access (Hall et al. 1982). The same phenomenon might operate in the wild-type *mgl* operon to decrease expression of *mglA* relative to *mglB*.

GBP coded by the two pVB1 plasmids is exported efficiently into the periplasmic space, although the last 3 amino acids of the wild-type protein are replaced by 14 different amino acids. This result suggests that other amino acid sequences fused to the extreme carboxy-terminal end of

GBP might also be exported. The *Eco*RI site present at the 3′ end of *mglB* should prove useful in generating such fusion proteins. GBP coded by the pVB1 plasmids can also bind galactose and support its uptake into the cytoplasm. Thus, the extreme C-terminal end of GBP need not be intact in order for the protein to function in substrate binding and transport.

The sequences of the $mglB^+$ and *mglB*551 genes were identical except at nucleotide 290. Here a G to A transition converts the wild-type GGT codon into GAT, leading to a substitution of Gly_{74} by Asp. This result was consistent with the observations that the *mglB*551 mutation reverts readily (this work; Engelhardt-Altendorf 1983) and that the mutant protein has an isoelectric point 0.4 pH units more acidic than wild-type GBP. This change is probably caused by the introduction of the acidic aspartate residue in an exposed region of the protein. An alternative explanation, that the loss of a glycine residue greatly disrupts the local secondary structure of the protein, is unlikely, since the X-ray diffraction patterns of GBP from wild-type and *mglB*551 strains are indistinguishable (F. Quiocho, personal communication).

Based on the published three-dimensional structures for *E. coli* GBP (Vyas et al. 1983), *S. typhimurium* GBP (Mowbray and Petsko 1983) and *E. coli* arabinose-binding protein (ABP; Gilliland and Quiocho 1981), residue 74 can be placed within or adjacent to alpha-Helix III. Helix III is located at the surface of the P domain at the point farthest from the substrate-binding site, which is in the cleft between the P and Q domains of the protein. The conformation of regions of GBP far from the cleft is altered by substrate binding (Zukin et al. 1977). Thus, in wild-type GBP bound substrate could affect the conformation or accessibility of Helix III to facilitate its recognition by Trg, and the change introduced by the *mglB*551 mutation might prevent the GBP-Trg interaction.

We have recently found that the chemotactic defects caused by certain mutations in *malE* can be overcome by extragenic suppressors (Manson and Kossmann 1986). The suppressing mutations map in *tar*, whose protein product, the Tar signal transducer, mediates maltose and aspartate chemotaxis. In analogous attempts to select mutations that suppress the chemotactic defect caused by the *mglB*551 mutation no suppressing mutations in *trg* were found (Engelhardt-Altendorf 1983; Park and Hazelbauer, personal communication). The change in Helix III introduced by the *mglB*551 mutation may so disturb the GBP-Trg interaction that it cannot be restored by a compensating change in Trg.

Gly_{74} lies within one of the two amino acid sequences showing strong homology among GBP, ABP, RBP and MBP (Argos et al. 1981; Duplay et al. 1984). In GBP these regions extend from residues 35 to 92 (Sequence A) and 205 to 256 (Sequence B). RBP also interacts with Trg, but it shows no greater similarity to GBP within Sequence A than does MBP, which interacts with Tar, or ABP, which does not serve as a chemoreceptor. In fact, Helix III shows a high level of homology between ABP and GBP, both at the level of primary and tertiary structure (Argos et al. 1981; Mowbray and Petsko 1983). Thus, the location of the site of the amino acid substitution caused by the *mglB*551 mutation does not by itself explain why GBP is a chemoreceptor and ABP is not.

The amino acid substitutions in MBP that specifically disrupt its interaction with Tar are located within the first 60 N-terminal residues of the protein (Manson and Kossmann 1986; M. Kossmann, unpublished data). This region lies within an N-terminal extension of MBP that probably has no direct counterpart in the other three well-characterized sugar-binding proteins of *E. coli* (Duplay et al. 1984). We conclude that the MBP-Tar and the GBP-Trg interactions may be rather dissimilar at the molecular level.

Acknowledgments. We thank C.F. Higgins and F. Quiocho for communication of results prior to publication, D. Benner and K. Eiglmeier for advice on the dideoxynucleotide method of DNA sequencing, and E. Bremer and J.M. Brass for helpful suggestions for improving the manuscript. N. Müller pointed out the utility of the *Eco*R1 site of *mglB* for creating protein fusions. We are especially grateful to M.W. Schwarz for providing custom-synthesized primers for DNA sequencing. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 156).

References

Ames GL (1986) Bacterial periplasmic transport systems: Structure, mechanism, and evolution. Annu Rev Biochem 55:397–425

Anraku Y (1967) The reduction and restoration of galactose transport in osmotically shocked cells of *Escherichia coli*. J Biol Chem 242:793–800

Argos P, Mahoney WC, Hermondson MA, Hanei M (1981) Structural predictions of sugar binding proteins functional in chemotaxis and transport. J Biol Chem 256:4357–4361

Bavoil P, Hoffnung M, Nikaido H (1980) Identification of a cytoplasmic membrane-associated component of the maltose transport system of *E. coli*. J Biol Chem 255:8366–8369

Bolivar F, Rodriguez RL, Green PJ, Betlach MC, Heyneker HL, Boyer HW (1977) Construction of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95–113

Boos W (1969) The galactose binding protein and its relationship to the β-methylgalactoside permease from *Escherichia coli*. Eur J Biochem 10:66–73

Boos W (1972) Transport properties of the galactose-binding protein of *Escherichia coli*: Substrate-induced conformational change. J Biol Chem 247:917–924

Boos W, Steinacher I, Engelhardt-Altendorf D (1981) Mapping of *mglB*, the structural gene of the galactose-binding protein of *Escherichia coli*. Mol Gen Genet 184:508–518

Brass JM, Manson MD (1984) Reconstitution of maltose chemotaxis in *Escherichia coli* by addition of maltose binding-protein to calcium-treated cells of maltose regulon mutants. J Bacteriol 157:93–99

Brass JM, Ehmann U, Bukau B (1983) Reconstitution of maltose transport in *Escherichia coli*: conditions effecting import of maltose-binding protein into the periplasm of calcium treated cells. J Bacteriol 155:97–106

Bremer E, Silhavy TJ, Weisemann JM, Weinstock GM (1984) λ*plac*Mu: a transposable derivative of phage λ for creating *lacZ* protein fusions in a single step. J Bacteriol 158:1084–1093

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

Dahl M, Manson MD (1985) Interspecific reconstitution of maltose transport and chemotaxis in *Escherichia coli* with maltose-binding protein from various enteric bacteria. J Bacteriol 164:1057–1063

Duplay P, Bedouelle H, Fowler A, Zabin I, Saurin W, Hofnung M (1984) Sequences of the *malE* gene and its product, the maltose-binding protein of *Escherichia coli* K12. J Biol Chem 259:10606–10613

Ebright RH, Cossart P, Gicquel-Sanzey B, Beckwith J (1984) Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. Nature 311:232–235

Engelhardt-Altendorf D (1983) Mechanismus der Signalübertragung während der Chemotaxis bei *Escherichia coli*. Doctoral thesis, University of Konstanz, FRG

Gilliland GL, Quiocho FA (1981) Structure of the L-arabinose-binding protein from *Escherichia coli* at 2.4 Å resolution. J Mol Biol 146:341–362

Grosjean H, Fiers W (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18:199–209

Hall MN, Gabay J, Débarbouillé, Schwartz M (1982) A role for secondary structure in the control of translation initiation. Nature 295:616–618

Harayama S, Bollinger J, Iino T, Hazelbauer GL (1983) Characterization of the *mgl* operon of *Escherichia coli* by transposon mutagenesis and molecular cloning. J Bacteriol 153:408–415

Hazelbauer GL, Adler J (1971) Role of the galactose-binding protein in chemotaxis of *Escherichia coli* toward galactose. Nature New Biol 230:101–104

Hazelbauer GL, Harayama S (1983) Sensory transduction in bacterial chemotaxis. Int Rev Cytol 81:33–70

Hazelbauer GL, Engstrom P, Harayama S (1981) Methyl-accepting chemotaxis protein III and transducer gene *trg*. J Bacteriol 145:43–49

Higgins CF, Hiles ID, Salmond GPC, Gill DR, Downie JA, Evans IJ, Holland IB, Gray L, Buckel SD, Bell AW, Hermondson MA (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature 323:448–450

Koman A, Harayama S, Hazelbauer GL (1979) Relation of chemotactic response to the amount of receptor: Evidence for different efficiencies of signal transduction. J Bacteriol 138:739–747

Kondoh H, Ball CB, Adler J (1979) Identification of a methyl accepting chemotaxis protein for the ribose and galactose chemoreceptors of *Escherichia coli*. Proc Natl Acad Sci USA 76:260–264

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lengeler J, Hermann KO, Unsöld HJ, Boos W (1971) The regulation of the β-methylgalactoside transport system of the galactose binding protein of *Escherichia coli* K12. Eur J Biochem 19:457–470

Mahoney WC, Hogg RW, Hermodson MA (1981) The amino acid sequence of the D-galactose-binding protein from *Escherichia coli* K12. J Biol Chem 256:4350–4356

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Manson MD, Kossmann M (1986) Mutations in *tar* supress defects in maltose chemotaxis caused by specific *malE* mutations. J Bacteriol 165:34–40

Messing J, Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene 19:269–276

Messing J, Crea R, Seeburg PH (1981) A system for shotgun DNA sequencing. Nucleic Acids Res 9:309–321

Middendorf A, Schweizer H, Vreemann J, Boos W (1984) Mapping of markers in the *gyrA-his* region of *Escherichia coli*. Mol Gen Genet 197:175–181

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Mowbray SL, Petsko GA (1983) The X-ray structure of the periplasmic galactose-binding protein from *Salmonella typhimurium* at 3.0 Å resolution. J Biol Chem 258:7991–7997

Müller N, Heine HN, Boos W (1985) Characterisation of the *Salmonella typhimurium mgl* operon and its gene products. J Bacteriol 163:37–45

Neu HC, Heppel LA (1965) The release of enzymes of *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J Biol Chem 254:7529–7533

Ordal GW, Adler J (1974a) Isolation and complementation of mutants in galactose taxis and transport. J Bacteriol 117:509–516

Ordal GW, Adler J (1974b) Properties of mutants in galactose taxis and transport. J Bacteriol 117:517–526

Randall LL, Josefsson L-G, Hardy SJS (1980) Novel intermediates in the synthesis of maltose-binding protein in *Escherichia coli*. Eur J Biochem 107:375–379

Rotman B, Guzman RJ (1982) Identification of the *mglA* gene product in the β-methylgalactoside transport system of *Escherichia coli*. J Biol Chem 257:9030–9034

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Scripture JB, Hogg RW (1983) The nucleotide sequences defining the signal peptides of the galactose-binding protein and the arabinose-binding protein. J Biol Chem 258:10853–10855

Shuman HA (1981) The use of gene fusions to study bacterial transport proteins. J Membr Biol 61:1–11

Tinoco I Jr, Borer PN, Dengler B, Levine M, Uhlenbeck OC, Crothers DM, Gralla J (1973) Improved estimation of secondary structure in ribonucleic acids. Nature New Biol 246:40–41

Vyas NK, Vyas MN, Quiocho FA (1983) The 3 Å resolution structure of a D-galactose-binding protein for transport and chemotaxis in *Escherichia coli*. Proc Natl Acad Sci USA 80:1792–1796

Zukin RS, Hartig PR, Koshland DE Jr (1977) Use of a distant reporter group as evidence for a conformational change in a sensory receptor. Proc Natl Acad Sci USA 74:1932–1936

Communicated by J. Lengeler

Received October 13, 1986