MicroReview

The galactose regulon of Escherichia coli

Michael J. Weickert¹ and Sankar Adhya²*
¹Somatogen Inc., 5797 Central Avenue, Boulder, Colorado 80301, USA.
²Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

Summary

Galactose transport and metabolism in Escherichia coli involves a multicomponent amphibolic pathway. Galactose transport is accomplished by two different galactose-specific transport systems. At least four of the genes and operons involved in galactose transport and metabolism have promoters containing similar regulatory sequences. These sequences are recognized by at least three regulators, Gal repressor (GalR), Gal isorepressor (GalS) and cAMP receptor protein (CRP), which modulate transcription from these promoters. The negative regulators, GalR and GalS, discriminate between utilization of the high-affinity (regulated by GalS) and low-affinity (regulated by GalR) transport systems, and modulate the expression of genes for galactose metabolism in an overlapping fashion. GalS is itself autogenously regulated and CRP dependent, while the gene for GalR is constitutive. The gal operon encoding the enzymes for galactose metabolism has two promoters regulated by CRP in opposite ways; one (P₁) is stimulated and the other (P₂) inhibited by CRP. Both promoters are strongly repressed by GalR but weakly by GalS. All but one of the constituent promoters of the gal regulon have two operators. The gal regulon has the potential to co-ordinate galactose metabolism and transport in a highly efficient manner, under a wide variety of conditions of galactose availability.

Introduction

The sugar D-galactose is of importance to Escherichia coli, not only as an energy source, but as a building block in complex polysaccharide formation. First, energy is produced by catabolism of galactose to the glycolytic intermediate, glucose-1-phosphate. Second, two of the intermediates of the galactose metabolic pathway, UDP-galactose and UDP-glucose are required for biosynthetic glycosylation reactions. This makes the galactose pathway, also known as the Leloir pathway, an amphibolic one. Many of the genes involved in the transport and metabolism of galactose have been identified and characterized. Since the last review of the regulation of transcription of the gal operon, which was known to encode the first three enzymes of galactose metabolism (Adhya, 1987) new regulatory and structural genes of galactose transport and metabolism have been identified and sequenced. In addition, it has been shown that most of these genes, including the gal operon, are modulated by common regulatory mechanisms and form a regulon. In this review, we will summarize more recent information about the galactose genes and their products, examine how these genes are regulated, and discuss possible mechanisms for co-ordination of transcription of these genes and their implications. The map positions of the galactose genes on the E. coli chromosome are shown in Fig. 1.

Galactose-transport genes

D-galactose is primarily transported by two specific transport systems, one with high-affinity \( K_m = 1 \mu M \) and the other with low-affinity \( K_m = 50-450 \mu M \) for the sugar. The high-affinity galactose transport is accomplished by the action of the three proteins of the methyl-D-galactoside permease system encoded by the mgl operon (Robbins et al., 1976), while low-affinity galactose transport is facilitated by galactose permease, the galP gene product (Riordan and Kornberg, 1977). Both are discussed in greater detail below. Galactose is also transported with much less efficiency by four other transport systems (see Sihavy et al., 1978 for a review of E. coli sugar transport). These additional systems, which transport galactose because of their broad specificity, can be divided into two groups — the proton-linked transport systems and the sugar-binding protein transport complexes. The proton-linked transport proteins belong to a large family of transport proteins that generally have 12 putative transmembrane domains and similar structural motifs (Henderson, 1991; Henderson et al., 1992; Griffith et al., 1992). Three members of this family of proteins can transport...
Galactose non-specifically: LacY permease; MelB permease: and AraE, the low-affinity L-arabinose transport protein. Arabinose and galactose transports are also accomplished by a sugar-binding high-affinity transport complex, encoded by the genes of the araFGH operon (Silhavy et al., 1978). There are several similarities between arabinose transport and galactose transport that may account for the low-efficiency galactose transport by the arabinose transport system as discussed below.

High-affinity galactose transport

High-affinity galactose transport is accomplished by a galactose-binding protein (GBP) and two membrane-associated transport proteins (Robbins et al., 1976). The complex is called MeGal or the P-MG transport system. The genes for these three proteins are organized, as mentioned before, into the mgl operon, named for the ability of this complex to also facilitate the transport of methyl-β-D-galactopyranosides. The operon, which encodes GBP (mglB), and the two membrane-associated proteins (mglA and mglC), is similar to the araFGH high-affinity arabinose transport operon in function, organization and structure of its gene products (Hogg et al., 1991; Scripture et al., 1987; Vyas et al., 1991). In addition to high-affinity galactose transport, GBP is required for galactose chemotaxis (Ordal and Adler, 1974). Binding to galactose, GBP interacts with the chemotactic signal protein Trg to mediate chemotaxis against a galactose concentration gradient (Kondoh et al., 1979; Hazelbauer et al., 1981).

Low-affinity galactose transport

When present at high extracellular concentrations, galactose is transported into E. coli preferentially by a galactose-specific proton-linked transport protein, galactose permease, (GalP). The galP gene (Riordan and Kornberg, 1977) has recently been sequenced and the deduced GalP amino acid sequence is 64% identical to AraE, the L-arabinose proton-linked transport protein (Roberts, 1992; Henderson et al., 1992). There are four known members — AraE, LacY, MelB and GalP — of the proton-symporter family. AraE is the only member of the family, other than GalP, that transports galactose, though with much less affinity than GalP. Like most proton-linked sugar transport proteins, sequence comparison of GalP indicates that it contains 12 putative membrane-spanning α-helices in two groups of six, with a large hydrophilic loop between helices 6 and 7 (Roberts, 1992; Henderson et al., 1992). The similarities between corresponding helices in the two groups of six suggest a duplication event in the evolution of the galP gene (Roberts, 1992; Griffith et al., 1992).

Galactose metabolism genes

Up to six genes are involved in the initial steps of β-galactose metabolism: galK (galactokinase), galT (galactose-1-phosphate uridylyltransferase), galE (uridine-diphosphogalactose-4-epimerase), galU (UTP: α-D-glucose-1-phosphate uridylyltransferase), and pgm (phosphoglucomutase). They have been reviewed previously by Adhya (1987). A mutarotase function, encoded by a galM gene, may be involved in the intracellular interconversion of D-α- and D-β-galactose (G. Bouffard, K. Rudd and S. Adhya, in preparation; Maskell, 1992). The complete sequence of the galM gene has been determined. The deduced amino acid sequence of mutarotase is similar to the ones of Streptococcus thermophilus, Lactobacillus helveticus, and Acinetobacter calcoaceticus (Poolman et al., 1990; Mollet and Pilloud, 1991; Gatz and Hillen, 1986). galM turns out to be a fourth gene of the gal operon, composed of galE, T, K and M, in that order.

The sequence of the galU gene has also recently been determined by Ueguchi and Ito (1992) and by A. C. Weissborn, M. K. Rumley and E. P. Kennedy (personal communication). The galU gene product is similar to glucose-1-phosphate uridylyltransferase, called CelA, from Acetobacter xylinum in amino acid sequence and is over 54% identical to a hypothetical second glucose-1-phosphate uridylyltransferase, most likely the galF gene product, in Salmonella typhimurium LT2 (A. C. Weissborn et al., personal communication). The pgm gene is yet to be cloned and sequenced.

Transcriptional regulators of the galactose genes

Of the genes of galactose uptake and metabolism discussed above, galP, mglBCA and galETKM are part of the gal regulon. The galU and pgm genes do not appear to be subject to galactose regulation (M. J. Weickert and S. Adhya, unpublished results). The known regulators of the
and are members of a large family of related transcriptional regulators, the GalR-LacI family. The common functional feature of these regulators is that they act as transcriptional repressors, although some of them may also activate transcription at some promoters. CRP, which is not a member of the GalR-LacI family of regulators, also functions as a symmetric dimer by binding to DNA sites with dyad symmetry located upstream of the RNA polymerase in the promoters. CRP can activate or repress transcription. CRP acts through one or both of the following mechanisms: by directly contacting RNA polymerase and by inducing structural change in the DNA (for reviews, see Adhya and Garges, 1990; Bell et al., 1990; Eschenlauer and Reznikoff, 1991; Ishihama, 1993; Kolb et al., 1993; Zhou et al., 1993).

**Fig. 2. DNA sequences involved in or potentially involved in regulation of transcription at gal regulon promoters.**

A. CRP-cAMP binding sites in gal regulon promoters. Nucleotides enclosed within boxed region are homologous to the CRP-cAMP binding site consensus sequence (Ebright et al., 1984; Barber and Zhurkin, 1990).

B. Comparison of half sites for the symmetric operator sequences and potential operators. The 5' antisense (left top half) sequence and the 3' sense (right bottom half) sequence are aligned in a 5' to 3' orientation (see galOP example). Boxed nucleotides are identical to *", the consensus gal operator of Majumdar and Adhya (1987). The filled circles above the consensus sequence indicate bases believed to make direct contact with GalR.

Regulon are galR, galS and Crp, which encode Gal repressor (GalR), Gal isorepressor (GalS) and cyclic AMP receptor protein (CRP), respectively.

The gal operon has two promoters; P1 and P2. They have served as a model system for examining both positive and negative regulation of transcription. All three regulators, CRP, GalR and GalS, which modulate gal regulon transcription, work at these two promoters. cAMP and CRP, acting as a complex, stimulate (positive control) transcription at P1, whereas they inhibit transcription from P2. Both GalR and GalS negatively regulate transcription of P1 and P2 (reviewed in Adhya, 1987; Tokeson et al., 1991; Golding et al., 1991; Weickert and Adhya, 1992a; Choy and Adhya, 1992).

GalR and GalS are over 53% identical and 85% similar and are members of a large family of related transcriptional regulatory proteins in bacteria, called the GalR-LacI family (Weickert and Adhya, 1992b). These regulators contain an amino-terminal domain through which they bind specific operators with dyad symmetry. When associated with DNA, these regulators are virtually symmetric dimers. Additional conserved domains are involved in substrate (inducer) binding, dimerization, and for some, like LacI, tetramerization. The most common functional feature of these regulators is that they act as transcriptional repressors, although some of them may also activate transcription at some promoters.

CRP, which is not a member of the GalR-LacI family of regulators, also functions as a symmetric dimer by binding to DNA sites with dyad symmetry located upstream of the RNA polymerase in the promoters. CRP can activate or repress transcription. CRP acts through one or both of the following mechanisms: by directly contacting RNA polymerase and by inducing structural change in the DNA (for reviews, see Adhya and Garges, 1990; Bell et al., 1990; Eschenlauer and Reznikoff, 1991; Ishihama, 1993; Kolb et al., 1993; Zhou et al., 1993).

**Similarities among gal regulon promoters**

The promoter regions of the gal and mgl operons, and of the galP, galU, galR and galS genes have been sequenced. The transcription start sites have been determined for the gal and mgl operons (Irani et al., 1989; Hogg et al., 1991; Weickert and Adhya, 1992a), and the galR and galS genes (Weickert and Adhya, 1993). Although the mgl, galR and galS genes, unlike the gal operon, contain only one promoter each, their promoter regions and that of galP contain sequences homologous to GalR- and GalS-binding sites (gal operators) and CRP-cAMP-binding sites (Fig. 2) present in the gal operon with more or less similar arrangements (Fig. 3): (i) all of them contain a CRP-cAMP-binding site (Fig. 2A) near -40 with respect to the start site of transcription (Fig. 3) and act as type II promoters (Ishihama, 1993). CRP binding to the -40 region activates transcription from the P1 promoter of the gal operon as well as from the mgl, galS and galP promoters. Transcription from these promoters is reduced three- to 20-fold upon disruption of the crp gene (galP; Irani et al., 1989; galS; Weickert and Adhya, 1993; galP and mgl; M. J. Weickert and S. Adhya, unpublished results).

Like O2 in the gal operon, a gal operator is located at the -60 region in mgl, galR and galS (Figs 2B and 3). Similar also to the gal operon, galR and galS but not mgl have a second gal operator sequence (O1) in the cognate protein-coding region. GalR has been shown to negatively regulate, like the gal operon promoters, the galP promoter (Wilson, 1974). GalS negatively regulates the promoter of its own gene, galS (Weickert and Adhya, 1993) and the mgl promoter (Weickert and Adhya, 1992a). It is not clear whether repression of galP and galS genes requires a DNA looping by the use of bipartite operators much like the mechanism implicated in the gal operon (Haber and Adhya, 1988; Mandal et al., 1990; Choy and Adhya, 1992).

Interestingly, all of the gal regulon promoters, except the...
How gal operon transcription is regulated

The unusual spacing between the –10 and –35 hexamers of the galP₁ promoter (21 bp) and the relatively low homology of the –35 region with the consensus –35 hexamer, has a profound effect on RNA polymerase binding to the gal promoter. RNA polymerase binding at galP₁ does not require specific contacts in the –35 region (Chan and Busby, 1989; Chan et al., 1990). Chemical and DNase footprinting of the galP₁ promoter indicate that RNA polymerase fails to protect the –35 region while upstream sequences, up to –55, are protected. The DNA in the promoter region is probably distorted by these compensatory upstream contacts (Chan and Busby, 1990; A. Majumdar and S. Adhya, in preparation). In addition, the galP₁ promoter is partially dependent upon three phased poly-adenine stretches upstream of the CRP site, but on the same face of the DNA helix as the CRP site and –10 region (Lavigne et al., 1992). These poly-adenine regions are naturally curved, and their presence increases the rate of RNA polymerase occupation and isomerization at P₁, especially in the presence of CRP–cAMP. Bent sequences alone can also activate P₁ transcription when they substitute for the CRP–cAMP site (Bracco et al., 1989). Note that one of the predominant features of CRP–cAMP binding to DNA is the induction of a pronounced DNA bend (Wu and Crothers, 1984).

In the normal gal operon, CRP–cAMP binding switches transcription initiation between start sites from P₁ and P₂ entirely to P₁ (Irani et al., 1989; Goodrich and McClure, 1992; Choy and Adhya, 1993). In the absence of CRP–cAMP, initiation occurs in approximately equal proportions from these two overlapping promoters while binding of CRP–cAMP switches initiation predominantly (>95%) to P₁. CRP–cAMP bound to the DNA promotes transcription initiation preferentially at P₁, by both increasing RNA polymerase binding to P₁ and elevating the rate of isomerization to open complex at P₁.

Negative regulation of the gal operon by GalR in vivo requires simultaneous DNA binding by dimers of GalR to palindromic operator sites (OE and O₁) flanking the promoter region. In experiments where one of the gal operators was replaced with a lac operator, there was no repression (Haber and Adhya, 1988; Mandal et al., 1990; Choy and Adhya, 1992; Brenowitz et al., 1990). However, when both operators were replaced by lac operators, LacI repressor was able to fully repress the operon. This requirement for the binding of the same type of repressor to each operator suggests that an interaction between the
two repressor dimers, bound to their cognate sites, is required for full repression, and occurs by forming a loop of intervening DNA. Tetramerization by GalR dimers bound to operators to form a DNA loop, unlike that of LacI, has not yet been demonstrated in vitro and may reflect the requirement of an additional factor for the association to occur. Unlike LacI, GalR does not contain a mini-leucine zipper at the carboxyl terminus of the protein, which facilitates LacI tetramerization (Alberti et al., 1991). GalR binds non-co-operatively to the two operator sites (Brenowitz et al., 1990) and alters the DNA conformation (Wartell and Adhya, 1988; Majumdar and Adhya, 1989) at least in part by inducing a DNA bend (Kuhnke et al., 1989; Zwiebel et al., 1989). RNA polymerase, CRP–cAMP, and GalR can all simultaneously occupy their respective binding sites demonstrating that steric hindrance cannot account for the regulation seen at the gal operon promoter (Adhya and gat, 1989; Majumdar and Adhya, 1989) at least in part by inducing a DNA bend (Kuhnke et al., 1989; Zwiebel et al., 1989). RNA polymerase, CRP–cAMP, and GalR can all simultaneously occupy their respective binding sites demonstrating that steric hindrance cannot account for the regulation seen at the gal operon promoter (Adhya and gat, 1989; Majumdar and Adhya, 1989). RNA polymerase, CRP–cAMP, and GalR can all simultaneously occupy their respective binding sites demonstrating that steric hindrance cannot account for the regulation seen at the gal operon promoter (Adhya and gat, 1989; Majumdar and Adhya, 1989). RNA polymerase, CRP–cAMP, and GalR can all simultaneously occupy their respective binding sites demonstrating that steric hindrance cannot account for the regulation seen at the gal operon promoter (Adhya and gat, 1989; Majumdar and Adhya, 1989). RNA polymerase, CRP–cAMP, and GalR can all simultaneously occupy their respective binding sites demonstrating that steric hindrance cannot account for the regulation seen at the gal operon promoter (Adhya and gat, 1989; Majumdar and Adhya, 1989).

The gal operon is induced by D-galactose or the non-metabolizable analogue, D-fucose. Induction of the gal operon requires inducer binding to GalR; in the presence of a high concentration of inducer, the repressor loses its affinity for DNA (Majumdar and Adhya, 1984; Majumdar et al., 1987). In mutants lacking Gal repressor, the addition of galactose or fucose stimulates two- to fourfold greater gal expression (Tokeson et al., 1991). This phenomenon, termed ultrainduction, is another negative-control mechanism that operates at the level of transcription, and is mediated by the GalS (Golding et al., 1991; Weickert and Adhya, 1992a). GalS, like GalR, can be titrated specifically by multicycle gal operators to which the Gal repressor itself binds (Tokeson et al., 1991). GalS is believed to repress transcription, although not efficiently, in a manner analogous to GalR (Weickert and Adhya, 1992a). The cause of the weak repression of the gal operon by GalS is unknown.

**Implications for gal regulon promoters**

The striking similarities between the gal regulon promoters suggest that many of the mechanisms involved in regulating the gal operon promoter P, by CRP–cAMP that are discussed above are likely to operate in the other promoters. The other gal regulon promoters may be somewhat less complicated than the gal operon promoter segment because there is no evidence for the presence of a second promoter. The promoter switching that occurs between galP and P is by CRP–cAMP is probably necessary to maintain a minimal level of the galE and galT gene products under different growth conditions because of their importance in glycosylation reactions. There is no indication that the other genes in the gal regulon are similarly required in the absence of the CRP–cAMP complex.

It is possible that the internal gal operator, O, is coincidental in all other gal regulon promoters. The mgl promoter lacks an internal operator entirely, yet it is regulated at least four- to fivefold (Weickert and Adhya, 1992a). It has been shown that occupation of both O and O, and DNA looping brings about complete repression of P and P, by the Lac repressor (Choy and Adhya, 1992). Occupation of O alone by a repressor brings about 75% repression of P, but none of P. The cause of this repression without DNA looping needs to be investigated. It may parallel the repression of the mgl promoter by only one operator located at the region of O.

**Co-ordination of gal regulon transcription by a two-repressor system**

Understanding the co-ordination of gal regulon promoters is complicated by the unusual presence of an isorepressor, GalS. In addition to apparent cross-talk at the gal operators, the isorepressor is the primary negative regulator of two other gal regulon promoters (Weickert and Adhya, 1992a; 1993). With the identification of the isorepressor, it is clear that the repressor and isorepressor recognize the same or nearly the same operator sequence yet regulate different sets of promoters differently. This greatly complicates the potential mechanisms for control of transcription. The level at which such differential regulation occurs remains to be worked out. Figure 4 summarizes the spectrum of various protein–DNA interactions that contribute to modulation of transcription of gal regulon genes.

Several observations are critical to an understanding of the gal regulon co-ordination. (i) The gal transport systems are negatively regulated by different repressors. The different affinities of each transport system for galactose suggests that induction should occur at different galactose concentrations for each. If so, we would expect that GalS will be induced at lower galactose concentrations than GalR, since the mgl operon is the more efficient transport system at low galactose concentrations. (ii) The galS gene is autoregulated and its expression is CRP–
cAMP dependent, while galR expression appears to be constitutive (Weickert and Adhya, 1993). Autoregulation of gene expression (reviewed in Maloy and Stewart, 1993) creates a condition-dependent concentration of the regulator. We do not know why the galR gene expression is constitutive, although it carries two operators, one external and one internal.

These two observations lead to the following suggestions. In the absence of galactose, GalS concentration is low, and may facilitate induction of the mgl operon at lower galactose concentrations simply by its own scarcity or by a greater sensitivity to galactose than GalR. The constitutivity of GalR synthesis coupled with the inducibility of GalS synthesis could result in a varying proportion of GalR to GalS during changes in galactose concentration and metabolism (through CRP-cAMP). If these highly related proteins can form heterodimers in the same fashion as eukaryotic transcription factors (Turner and Tjian, 1989; Gentz et al., 1989) additional DNA and inducer binding specificities could be created. The proportion of heterodimers would be profoundly influenced by the relative proportions of the available repressor species. In experiments using purified GalR, the free energies of dimerization and DNA binding were nearly identical (BreNowitz et al., 1990). Since dimerization is essential for DNA binding, it was suggested that an additional level of regulation might operate at the dimerization interface of GalR. The potential existence of an additional protein that may influence the dimerization increases the potential for regulation at this level.

A heterodimeric species is likely to have different induction kinetics and DNA binding specificities. If each protein has subtly different DNA-binding specificities, heterodimeric species allow for the orientation of the subunits depending upon the sequence of the binding site. This is especially important if protein–protein contacts between the repressor species and other components of the nucleoprotein complex contribute to regulation.

Though the promoters of the gal regulon are likely to have similar regulatory mechanisms, the two-repressor system creates the potential for uncoordinated control of these genes, although they involve related functions and sequences. This would allow a more precise and physiological control over genes whose expression must be carefully modulated to ensure efficient cell growth under conditions that vary considerably. If heterodimeric repressor species can be identified for the gal regulon, it seems likely that heterodimers of other repressor species will also be found playing a role in modulating transcription.

Acknowledgements

We would like to thank Susan Garges, Hyon Choy and Michael Brenowitz for interesting discussions, and Peter Henderson and Audrey Weissborn for providing data before publication.

References

Adhya, S. (1987) In Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. Neidhardt, F.C. et al. (eds). Washington, D.C.: American Society for Microbiology, pp. 1503–1512.
Adhya, S. (1989) Annu Rev Genet 23: 227–250.
Adhya, S., and Garges, S. (1990) J Biol Chem 265: 10797–10800.
Adhya, S., and Majumdar, A. (1987) In RNA Polymerase and the Regulation of Transcription. Reznikoff, W.S. et al. (eds). New York: Elsevier Science Publishing Co. Inc., pp. 129–135.
Alberti, S. et al. (1991) New Biol 3: 57–62.
Barber, A.M., and Zhurkin, V.B. (1990) J Biomol Struct Dyn 8: 213–232.
