Journal of Cellular Biochemistry 51:83–90 (1993)

Transcriptional Regulation of the *bgl* Operon of *Escherichia coli* Involves Phosphotransferase System-Mediated Phosphorylation of a Transcriptional Antiterminator

Orna Amster-Choder and Andrew Wright

Department of Molecular Biology, Tufts University Health Sciences Campus, Boston, Massachusetts 02111

Key words: β-glucosides, operon expression, two-component regulatory systems, antiterminator proteins, *E. coli* K-12

The *bgl* operon of *Escherichia coli* is regulated by a novel two-component regulatory system which consists of a sensory protein, BglF, present in the cytoplasmic membrane and a regulatory protein, BglG, present in the cytoplasm. The sensor is a phosphotransferase system (PTS) transport protein which controls the activity of the regulatory protein according to substrate availability. The regulator is an RNA binding protein which controls operon expression by transcriptional antitermination. In the absence of substrate, BglF phosphorylates BglG thus inactivating it. In the presence of substrate, BglF removes the phosphate from BglG thus allowing it to function as an antiterminator. In this review we focus on the interactions between the sensor and regulator of the *bgl* system and discuss other systems that are regulated in a similar way.

ELEMENTS INVOLVED IN REGULATING *bgl* OPERON EXPRESSION

The *bgl* operon of *E. coli* K-12 contains genes involved in the uptake and catabolism of aromatic β-glucosides. In *E. coli* K-12 the operon is cryptic [1]. However, in certain wild type *E. coli* strains, operon expression is inducible by β-glucosides [2]. A variety of mutations, including insertions upstream of the *bgl* promoter, activate the operon in *E. coli* K-12 [3–5]. Once activated by mutation, operon expression becomes inducible, transcription occurring from a unique, CAP-cAMP-dependent promoter [6]. The basis for crypticity of the operon in *E. coli* K-12 is not fully understood though it seems to be related in part to sequences upstream of the promoter [4]. Since the operon is completely intact and potentially functional in *E. coli* K-12, it seems likely that operon expression must occur in wild type cells, perhaps through the action of an unidentified inducer or under specific growth conditions. Alternatively, it is possible that the organism can switch between the non-activated (cryptic) and activated states by reversible DNA rearrangements [3].

The structure of the *bgl* operon is shown in Figure 1. Three essential genes, *bgl*G, *bgl*F, and *bgl*B, were identified by genetic analysis [7] and their presence confirmed by DNA sequence analysis [8]. A key feature of the operon, central to its regulation, is the presence of rho-independent terminator coding sequences flanking the first gene, *bgl*G (see below) [8,9]. The BglF protein was shown to be a member of the PTS family of transport proteins, sharing significant homology with the II^{Glc} and III^{Glc} PTS proteins involved in glucose transport in *E. coli* and with other PTS family members as well [8,10,11]. These proteins function by coupling phosphorylation of the substrate with its transport. The BglB protein was shown to be essential for hydrolysis of the phosphorylated β-glucoside yielding

Received September 21, 1992; accepted September 21, 1992.

Address reprint requests to Andrew Wright, Department of Molecular Biology, Tufts University Health Sciences Campus, 136 Harrison Ave., Boston, MA 02111.

Orna Amster-Choder's present address is Department of Molecular Biology, The Hebrew University, Hadassah Medical School, Jerusalem, Israel.

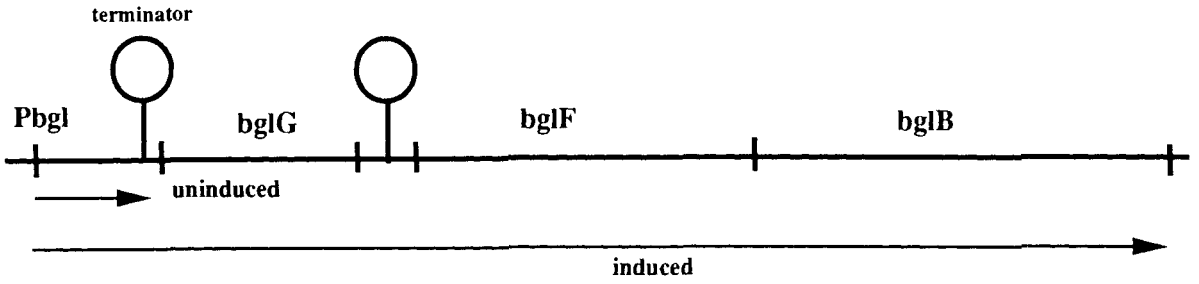


Fig. 1. Schematic representation of the *bgl* operon.

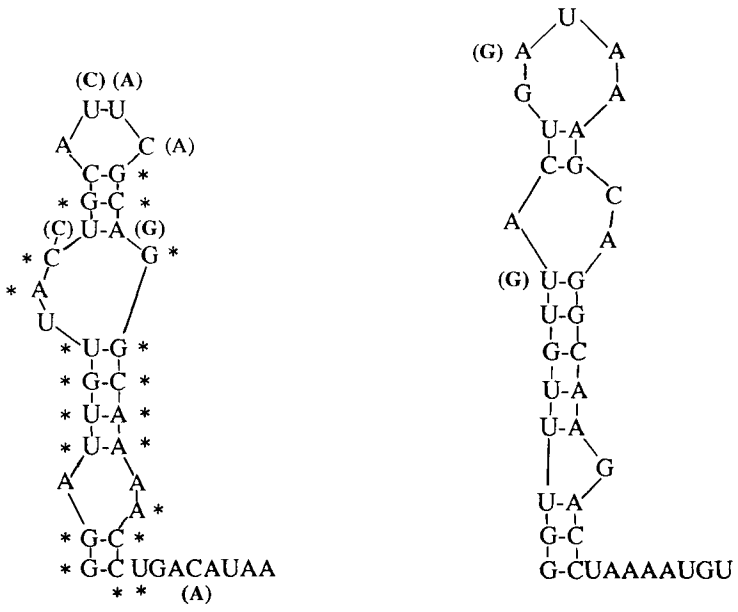
glucose-6-phosphate. BglF and BglB are necessary and sufficient for β-glucoside uptake and hydrolysis; however, production of these proteins requires BglG, which functions as a positive regulator of operon expression [9].

Genetic and biochemical analysis of the activated *bgl* operon in *E. coli* uncovered a novel form of regulation. First it was noted that substitution of the *lac* promoter for the *bgl* promoter had no effect on operon regulation; i.e., the operon was still inducible by β-glucosides [9]. The properties of a series of *bgl-lacZ* fusions with endpoints at different sites within the operon indicated that one major site of regulation was in or near the first rho-independent terminator downstream of the promoter [9]. Mutations reducing the efficiency of this terminator led to constitutive expression of a *bglG-lacZ* fusion, as did deletion of the complete terminator. Thus the terminator itself is involved in regulating operon expression. Further analysis indicated that expression of a *bgl-lacZ* fusion containing the first terminator region was absolutely dependent on the presence of BglG, thus defining BglG as a positive regulator of *bgl* operon expression. BglG-dependent expression of the *bgl-lacZ* fusion was constitutive in the absence of BglF but required inducer when BglF was present [9]. Thus, in addition to being involved in β-glucoside transport, BglF is a negative regulator of operon expression. Two classes of *bgl* mutants, with mutations in BglF and BglG, respectively, were identified which gave rise to constitutive expression of the *bgl-lacZ* fusion [7]. All but one of the mutations (79 total) were located in the *bglF* gene leading to loss of BglF function. The single exception was a mutation in *bglG*, resulting in a BglG protein with altered function. This mutation, *bglG33*, and another, *bglG4* [1], were insensitive to the negative effect of BglF. The *bglG33* and *bglG4* mutations, which occur at different sites in BglG (Diaz-Torres and Wright, unpublished observations), most likely define sites that allow interaction between BglG and BglF (see below). Based on these results it was concluded that BglF is a negative regulator of operon expression which exerts its effect through a direct interaction with BglG [7].

POSITIVE REGULATION

Transcriptional Antitermination via Protein-RNA Interaction

From the evidence indicated above it was clear that BglG regulated operon expression by acting as a transcriptional antiterminator. Subsequently it was demonstrated that the protein functions by binding to the *bgl* leader RNA just upstream of the first terminator [12]. The minimal binding site in the leader RNA was estimated to be 32 nucleotides long. This sequence actually extends into the leading stem of the terminator hairpin, thus binding of the protein to its target would physically block formation of the terminator leading to antitermination. The 32 nucleotide target sequence has the potential to form a stem-loop structure with two bulged regions (Fig. 2). Recent experiments have confirmed the presence of such a structure (Diaz-Torres and Wright, unpublished observations). Thus there are two alternative secondary structures in the *bgl* leader region, one in the BglG binding site, the other in the terminator. Binding of BglG to its target would thus preclude formation of the terminator. We have obtained evidence that BglG binds to an almost identical target sequence in the second terminator which lies in the intercistronic region between the *bglG* and *bglF* genes. Thus the same mechanism must be used for readthrough at this site as well. Such a system is reminiscent of attenuation control in the leader region of the *trp* operon, where the position of a regulatory ribosome determines RNA secondary structure



Bgl G1 and (G2) sites SacY and (SacT) sites

Fig. 2. RNA targets for antiterminator proteins. Shown are the RNA binding sites for BglG, SacY, and SacT. The structure on the left shows both BglG binding sites; the G1 site is present upstream of the *bgl*G gene, whereas the G2 site is downstream of the *bgl*G gene. For those positions where differences are observed, the G2 sequence is indicated in parentheses. The SacY and SacT sites, on the right, differ at only two positions. The asterisks on the BglG structure indicate residues that are conserved in all four structures. The structure on the left has been confirmed by nuclease analysis, but that on the right is theoretical.

which in turn determines the fate of the transcription complex [13]. It is not yet known how BglG binds to its RNA target nor is anything known about its binding affinity. In this regard it is interesting that overproduction of a truncated form of *bgl* leader RNA, lacking the complete terminator, strongly inhibited BglG antitermination in vivo. Under these in vivo conditions, BglG appeared to be able to bind tightly to its target sequence despite the fact that the latter was not in its normal mRNA context.

NEGATIVE REGULATION

Regulation of Transcription Antitermination by Reversible Protein Phosphorylation

Transcription from the *bgl* promoter is constitutive, but in the absence of β-glucosides in the growth medium most transcripts terminate at the first rho-independent terminator which is located upstream of the first gene in the operon [9]. Thus, under these conditions, BglG is not functioning as a transcriptional antiterminator. The presence of β-glucosides outside the cell is the signal for induction of expression by transcription antitermination. The protein that senses this signal and controls the activity of BglG as an antiterminator is BglF. This protein, an integral membrane protein, as deduced from its amino-acid sequence [8,10], interacts with the β-glucosides in the medium since it functions in their uptake and phosphorylation. BglF was shown to be related by function [14] and by amino-acid homology [11] to a group of transport proteins that are components of the phosphoenolpyruvate (PEP)-dependent PTS of *E. coli*. The phosphate is transferred from PEP to BglF, also termed enzyme II^{bgl}, via phosphorylated protein intermediates, and then to the sugar [15]. By analogy with other PEP-dependent phosphotransferases, BglF seems to contain two sites involved in phosphate transfer; one acting as an acceptor of phosphate from HPr, the heat-stable phosphoryl carrier of the PTS, the other acting to transfer the phosphate group to the sugar [16]. Site-directed mutagenesis, carried out to elucidate which residues are involved in these two phosphorylation events, identified a histidine as essential for the first event and both a cysteine and a second histidine as essential to the second [17,18]. Other resi-

dues, examined by this analysis, were shown to affect the kinetics of these events.

The way BglF carries out its other function, i.e., transfer of the signal to the transcriptional antiterminator, BglG, also involves transient phosphorylation events [15]. In the absence of β-glucosides, BglF phosphorylates BglG, thus blocking its action as a transcriptional antiterminator. Upon addition of inducer, BglF dephosphorylates BglG, allowing it to function as a positive regulator of gene expression. This mechanism was suggested originally based on in vitro studies of BglG phosphorylation by BglF, and on the properties of BglG mutant proteins, which allow constitutive expression of the operon and which show inability or decreased ability to be phosphorylated by BglF in vitro [15]. Later, it was demonstrated that both the phosphorylated and non-phosphorylated forms of BglG existed in vivo [17,19]. The activity of BglF as both BglG kinase and phosphatase was reinforced by demonstrating that the degree of phosphorylation of BglG in the cell is proportional to the level of BglF produced, and that rapid dephosphorylation can occur upon addition of β-glucosides to the growth medium [19].

It has been established that the flow of phosphate from HPr through BglF to substrate occurs by transfer to histidine 547, then either to histidine 306 or to cysteine 24, and finally to substrate [17,18]. In the absence of substrate, phosphate is transferred from histidine 547 to BglG. An unexpected finding was that a mutation at histidine 547 failed to block BglG phosphorylation. BglG phosphorylation in this case was found to be due to enzyme III^{glc}. Thus a second PTS protein can regulate BglG activity. It is likely that this interaction is important from a metabolic point of view when more than one carbon source is available to the cell. As expected, a mutation at histidine 306 of BglF made the *bgl* operon uninducible because it blocks the flow of phosphate to substrate, resulting in permanent phosphorylation of BglG.

Accumulating evidence obtained from studies on other sensory transduction systems (e.g., chemotaxis, nitrogen utilization, porin expression) suggests that transient protein phosphorylation is central to the mechanism of many information processing systems in bacteria [20,21]. A combination of genetic and biochemical approaches has shown that stimulus-response coupling involves two families of signal transduction proteins. One component of each regulatory pair is involved in sensing the stimulus and functions as a histidine-kinase that activates the second component, which acts as a response regulator, by phosphorylating it on an aspartate residue. Some regulators are DNA binding proteins which upon phosphorylation, in response to environmental information, can activate gene expression [21, and references therein]. Sequence motifs were detected in these proteins that may act as independent transmitter or receiver modules in mediating protein-protein communication, since they retain their functional identities in many protein hosts. Signal propagation is assumed to be a natural consequence of the recognition and association of these matched modules, accompanied by a conformational change in one or both of the interacting elements due to phosphorylation. Many pairs of proteins involved in various sensory processes, for which phosphorylation has not yet been demonstrated, are classified as members of these two families based on the fact that they contain sequences resembling the transmitter-receiver modules [21,22]. The β-glucoside utilization system, although a two-component regulatory system induced by an external stimulus, shares no homology with the proteins of the other group of two-component systems. The fact that the kinase BglF, by phosphorylating BglG, blocks its action as a positive regulator is in marked contrast to the way the other pairs of proteins are regulated. BglF is phosphorylated on a histidine, as are the sensors from the other group, but it phosphorylates a histidine residue on BglG rather than an aspartate residue, as is the case for the other response regulators (Amster-Choder and Wright, unpublished data). Thus, the system involved in *bgl* regulation defines a new group of two-component regulators. Other candidates for this group are discussed under the Related Systems section.

The basis for recognition of the positive regulator BglG by its kinase BglF is not completely understood yet. Studies with a series of hybrids containing various portions of BglG and complementary portions of SacY (see Related Systems), a similar antiterminator from *Bacillus subtilis,* suggest that regions of BglG, apart from the phosphorylation site, are required for phosphorylation by BglF (Amster-Choder, Bascom, and Wright, unpublished data). SacY can substitute efficiently for BglG in *bgl* operon expression in *E. coli.* However, it is not negatively regulated by BglF indicating that it lacks determinants

required for BglF recognition. Whether this portion of BglG contains a site for kinase recognition that is separate from the phosphorylation site, or is required for proper folding and presentation of the protein to its kinase, remains to be determined.

Modulation of Dimerization of BglG by Reversible Phosphorylation

How does transient phosphorylation of BglG modulate its activity as a transcriptional antiterminator? Three pieces of evidence obtained recently indicate that BglG exists in two configurations in vivo, an active non-phosphorylated form, which is a dimer, and a phosphorylated inactive form, which is a monomer [23]. First, a fusion protein, which contains a portion of BglG but lacks antitermination activity, exerted a dominant negative effect on wild type BglG in vivo. This suggested that an oligomeric form of BglG might function in antitermination. Second, determination of molecular weights on native gels showed that BglG exists as a dimer while phosphorylated BglG is a monomer. Third, the chimera formed by replacing the dimerization domain of λ repressor with BglG behaved like intact λ repressor in its ability to repress λ gene expression, evidence that it dimerizes in vivo. The efficiency of repression by the λ-BglG hybrid, which reflects the efficiency of its dimerization, was found to depend on the level of BglF and on β-glucoside availability. It is generally assumed that phosphorylation leads to changes in the conformation or configuration of proteins that affect their activity, but the nature of such changes has never been shown directly before. The *bgl* system is the first example of phosphorylation and dephosphorylation modulating a configurational change, in this case a change in the oligomeric state of BglG, which controls the activity of this protein.

Modulation of dimerization has been reported to be involved in transcription regulation by many DNA binding proteins. BglG is the first RNA binding protein reported to regulate transcription depending on its dimeric state. One question was whether the failure of monomeric phosphorylated BglG to antiterminate transcription is due to its inability to bind RNA. By directly assaying the ability of the monomers and dimers to bind RNA, it was possible to determine that while dimeric BglG binds to its RNA target, monomeric phosphorylated BglG does not [23].

A MODEL FOR *bgl* REGULATION

A model for *bgl* regulation consistent with the reported observations is shown in Figure 3. Expression of the operon is positively and negatively regulated by two of its gene products, BglG and BglF, respectively. BglG is a sequence-specific RNA binding protein whose ability to inhibit termination of the operon transcription is modulated by BglF, a membrane bound kinase-phosphatase protein that senses the presence of β-glucosides [7,12,15,19]. According to our model, the ability of BglG to bind and antiterminate depends on dimer formation, a process that is controlled by phosphorylation [23]. In the absence of signal (β-glucosides), BglF transfers phosphate to BglG, interfering with its dimerization, preventing it from binding RNA and thus blocking its action as an antiterminator. A terminator structure is formed on the RNA which leads to dissociation of the RNA polymerase from the transcript. Addition of β-glucosides stimulates dephosphorylation of BglG by BglF; non-phosphorylated BglG can dimerize and bind its RNA target, thus preventing the formation of the terminator and enabling RNA polymerase to transcribe through the operon.

The secondary structure formed by the target RNA (Fig. 2) is an important feature in the model. The free energy calculated for this structure, whose existence was confirmed by enzymatic probing of RNA and analysis of compensatory mutations as described above, suggests that it is not very stable. It is suggested that BglG stabilizes this structure by binding to it. This type of protein-RNA recognition explains why mutations in the RNA target recognized by BglG exhibit a drastic effect, completely eliminating binding, since they result in production of RNA species with different secondary structures that are not recognized by BglG (Diaz-Torres and Wright, unpublished data). The two RNA targets, which partially overlap the two terminators flanking the *bgl*G gene, have exactly the same predicted secondary structure. They differ from each other at four positions, two within the predicted loop and the other two are replacement of one Watson-Crick base pair by another within the stem (Fig. 2).

The reason why BglG needs to dimerize in order to act as an antiterminator is not understood. Many dimeric proteins that regulate transcription bind to symmetric sequences on DNA, i.e., sequences that contain two binding sites,

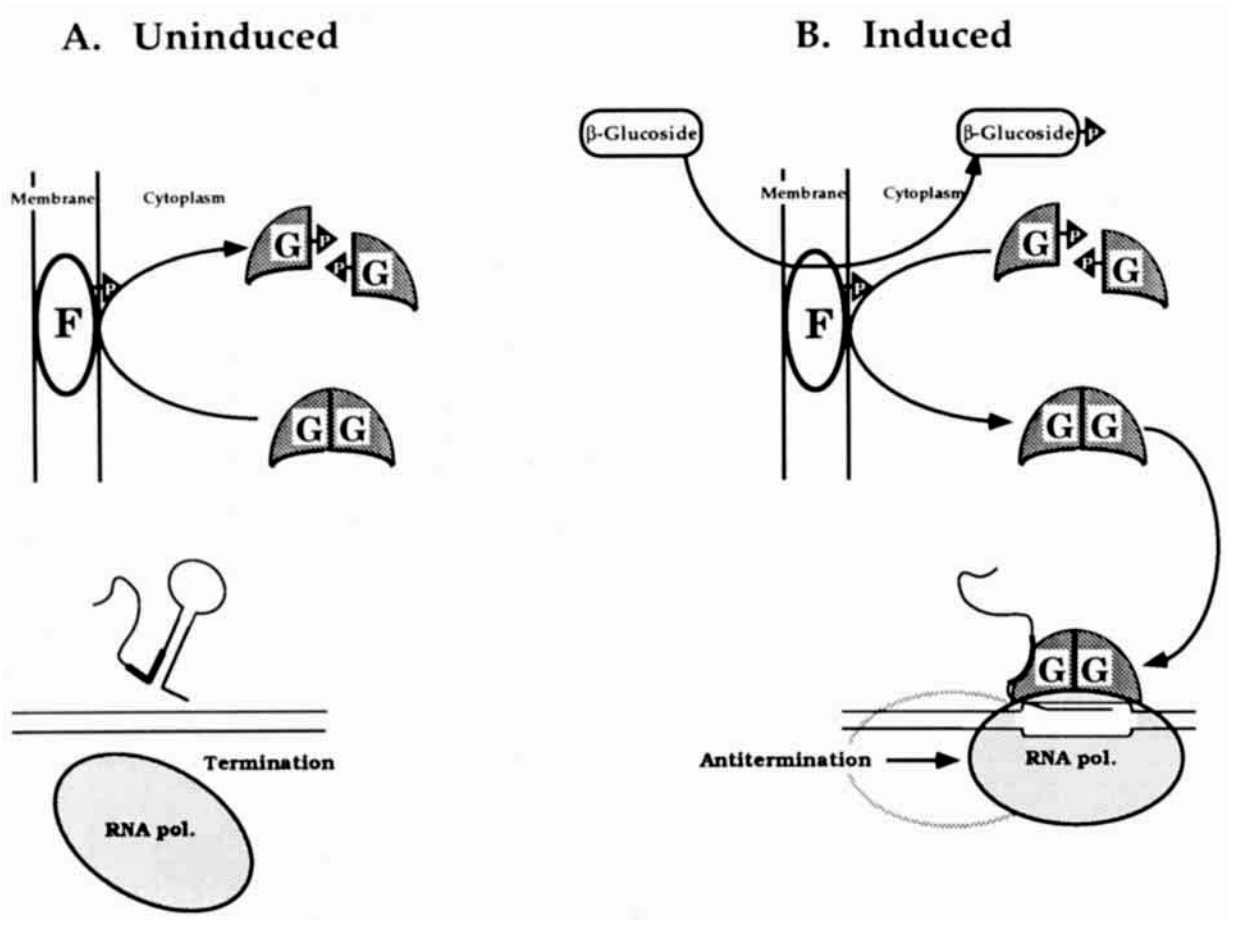


Fig. 3. Model of *bgl* operon regulation: (A) uninduced; (B) induced.

each binding a monomer. The RNA sequence to which BglG binds contains only one binding site. Thus the need for BglG dimers to bring about antitermination might reflect a concomitant interaction with RNA and RNA polymerase or other cellular proteins. Contacts between various transcription factors and the β- and α-subunits of RNA polymerase have been shown to be crucial for regulation of gene expression in many cases [24]. Since BglG is acting in such close proximity to the transcription complex, it is quite possible that it actually contacts it directly or does so indirectly by interacting with one of the host Nus proteins which are known to be required for various termination and antitermination events [25]. The need for dimers might alternatively reflect a type of interaction between BglG and the RNA transcript in a way that differs from dimer-DNA interaction.

RELATED SYSTEMS

Genes involved in sucrose utilization in *B. subtilis,* namely *sac*B and *sac*PA, appear to be regulated by mechanisms that are similar to that which controls *bgl* operon expression in *E. coli.* They are each preceded by rho-independent terminator sites that block transcription in the absence of sucrose which acts as inducer [26–28]. These sites contain sequences that are very similar to those in the *bgl* operon (Fig. 2). Expression of *sac*B and *sac*PA is controlled by proteins encoded by the *sac*XY [29] operon and the *sac*T gene [28], respectively. SacY and SacT are homologous to one another [28] (48% identity) and to BglG (35% identity), while SacX is a PTS membrane protein [30]. Genetic evidence indicates that SacY and SacT act as transcriptional antiterminators at these sites. SacX regulates the activity of SacY presumably by phosphorylation and dephosphorylation, thus it is the *B. subtilis* analogue of BglF. As expected, *sac*X$^-$ mutants are constitutive for *sac*B expression [29]. Interestingly, they are also constitutive for *sac*PA expression, indicating cross-talk between the two systems. SacT activity is also regulated. However, in this case, its regulatory protein has yet

to be identified. A SacT mutant has been isolated which expresses the *sac*PA and *sac*B genes constitutively indicating again that there is cross-talk between the two systems [31]. In cells with PTS defects, SacY acts in a fully constitutive way, turning on the *sac* genes [30]. This is the phenotype expected if SacY is negatively regulated by phosphorylation. Surprisingly, in the same mutant background SacT is non-functional. Thus, SacT apparently requires an intact PTS for its activity [30]. A possibility suggested to explain SacT behavior is that it may possess a second site that is either phosphorylated or which binds a ligand thus becoming activated. As stated in the previous section, we have recently found that SacY can function efficiently in *E. coli* to allow *bgl* operon expression, but it is not subject to regulation by BglF. Thus, these proteins combine independent regulatory specificity with a common RNA binding specificity. The analysis of hybrids constructed between the various positive regulatory genes should allow identification of regions that are recognized by their individual PTS counterparts.

The phytopathogenic bacterium *Erwinia chrysanthemi* contains a group of three genes, organized as an operon (*arb*), which mediate β-glucoside metabolism and which are highly homologous to the genes of the *bgl* operon [32]. The ArbG protein which regulates expression of the other two genes, *arb*F and *arb*B, is 61% identical to BglG. While it almost certainly acts as an antiterminator in *E. chrysanthemi* it does not substitute for BglG in *E. coli*. Whether it fails to bind the RNA target sequences in the *bgl* operon or is non-functional for some other reason is not known.

SIMILARITIES TO EUKARYOTIC SYSTEMS

Although the *bgl* and *sac* systems do not resemble other known bacterial regulatory systems, they do resemble eukaryotic regulatory systems in various ways. One example is phosphorylation of the c-Myb protein by casein kinase II at a site near the Myb DNA binding domain [33]. This phosphorylation negatively regulates the function of c-Myb as a transcriptional activator by preventing specific binding of this protein to DNA. Another example is the Tat protein of HIV which functions as a transactivator of HIV genome expression by binding to the RNA encoded by the TAR sequence element, thus increasing transcriptional initiation and stabilizing elongation [34]. There are a variety of other known cases where phosphorylation either stimulates or inhibits the ability of transcription factors to bind to DNA.

CONCLUDING REMARKS

Two-component regulatory systems like those discussed in this review are likely to be widespread among bacteria. Given the present state of knowledge of the *bgl* and *sac* systems, it should now be possible to determine precisely how the sensor (PTS protein) and regulator protein interact and also how PTS proteins like enzyme III^{glc} interact with the regulatory proteins. Understanding the rules for recognition between the various pairs of proteins should allow a clearer understanding of how such regulatory systems are linked to global regulation in the cell.

REFERENCES

1 Prasad I, Schaefler S J Bacteriol 120 638–650, 1974
2 Hall BG Mol Biol Evol 5 456–467, 1988
3 Reynolds AE, Felton J, Wright A Nature (London) 293 625–629, 1981
4 Lopilato J, Wright A In Drlica K, Riley M (eds) "The Bacterial Chromosome" Washington, DC American Society for Microbiology, 1990, pp 435–444
5 Schnetz K, Rak B EMBO J 7 3271–3277, 1988
6 Reynolds AE, Mahadevan S, LeGrice SFJ, Wright A J Mol Biol 191 85–95, 1986
7 Mahadevan S, Reynolds AE, Wright A J Bacteriol 169 2570–2578, 1987
8 Schnetz K, Toloczyki C, Rak B J Bacteriol 169 2579–2590, 1987
9 Mahadevan S, Wright A Cell 50 485–494, 1987
10 Bramley HF, Kornberg HL J Gen Microbiol 133 563–573, 1987
11 Bramley HF, Kornberg HL Proc Natl Acad Sci USA 84 4777–4780, 1987
12 Houman F, Diaz-Torres M, Wright A Cell 62 1153–1163, 1990
13 Landick R, Yanofsky C In Neidhardt F, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger E (eds) "*E coli* and *S typhimurium*, Cellular and Molecular Biology" Washington, DC American Society for Microbiology, pp 1276–1301
14 Fox CF, Wilson G Proc Natl Acad Sci USA 59 988–995, 1968
15 Amster-Choder O, Houman F, Wright A Cell 58 847–855, 1989
16 Saier M Microbiol Rev 53 109–120, 1989
17 Schnetz K, Sutrina SL, Saier M, Rak B J Biol Chem 265 13464–13471, 1990
18 Amster-Choder O, Wright A Science 249 540–542, 1990
19 Schnetz K, Rak B Proc Natl Acad Sci USA 87 5074–5078, 1990
20 Bourret RB, Borkovich KA, Simon MI Annu Rev Biochem 60 401–441, 1991

21 Stock JB, Stock AM, Mottonen JM Nature (London) 344 395–400, 1990
22 Kofoid EC, Parkinson JS Proc Natl Acad Sci USA 85 4981–4985, 1988
23 Amster-Choder O, Wright A Science 257 1395–1398, 1992
24 Igarashi K, Ishihama A Cell 65 1015–1022, 1991
25 Whalen W, Das A New Biol 11 975–991, 1990
26 Shimotsu H, Henner D J Bacteriol 168 380–388, 1986
27 Steinmetz M, LeCoq D, Djemia HB, Gay P Mol Gen Genet 191 138–144, 1983
28 Debarbouille M, Fouet A, Amauud M, Klier A, Rapoport G J Bacteriol 172 3966–3973, 1990
29 Crutz AM, Steimetz S, Aymerich S, Richter R, LeCoq D J Bacteriol 172 1043–1050, 1990
30 Arnaud M, Vary P, Zagorek M, Klier A, Debarbouille M, Postma P, Rapaport G J Bacteriol 174 3161–3170, 1992
31 Steinmetz M, LeCoq D, Aymerich S J Bacteriol 171 1519–1523, 1989
32 Hassouni M, Henrissat B, Chippaux M, Barras F J Bacteriol 174 765–777, 1992
33 Luschner B, Christenson E, Litchfield DW, Krebs EG, Eisenman R Nature 344 517–522, 1990
34 Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel A Science 252 1167–1171, 1991