The promoter architectural landscape controlled by the bacterial transcriptional activator PhoP

Igor Zwir[4,5], Xualin Tu[4], Henry Huang[4] and Eduardo A. Groisman[1,2,3,4,*]

Department of Molecular Microbiology, Howard Hughes Medical Institute, Washington University School of Medicine, Campus Box 8230, 660 S. Euclid Avenue, St. Louis, Missouri, 63110, USA

[1]Section of Microbial Pathogenesis, Yale School of Medicine, 295 Congress Avenue, 354D, New Haven, CT 06536, USA; [2]Yale Microbial Diversity Institute, PO Box 27389, West Haven, CT 06516, USA; [3]Howard Hughes Medical Institute; [4]Department of Molecular Microbiology, Washington University School of Medicine, Campus Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110, USA; and [5]Department of Computer Science and Artificial Intelligence, University of Granada, E-18071 Granada, Spain

[*]Corresponding author. E-MAIL eduardo.groisman@yale.edu; Tel. (+1) 203 737 7940; Fax (+1) 203 785 4067

SUMMARY

A major challenge in molecular genetics has been to understand how *cis*-regulatory information is integrated to determine differential gene expression. The interaction between a transcription factor and its target sequences *in vitro* can be well characterized, but the *in vivo* promoter output resulting from the insertion of these sequences in the context of other promoter elements is poorly understood. To better comprehend this, we constructed synthetic promoters using a battery of *cis*-acting features identified in promoters controlled by the transcription factor PhoP, which governs the expression of ~5% of the genes in the bacterium *Salmonella enterica* serovar Typhimurium. These promoters contain different quantities of binding sites, which may be placed at various orientations, locations, and hence rotational phasing along the DNA with respect to the RNA polymerase (RNAP) binding site, and might have distinct functional roles, as well as RNAP binding sites of different qualities. Systematic analysis of the ranges of these features, as well as of their organization in the promoter regions, reveals that only a few combinations are functional in natural promoters. These are the promoter architectures that define the rules for programming gene expression of PhoP-activated genes, encoding distinct mechanisms used by PhoP to interact with the RNAP and to differentially control ancestral vs. horizontally-acquired genes. Genomic analysis also indicates that certain promoter architectures are species-specific.

INTRODUCTION

Bacterial transcriptional activators stimulate gene transcription by binding to specific DNA sequences on target promoters, where they make specific contacts with RNA polymerase (RNAP) and/or alter the local DNA structure [1]. These activators may bind to a single or to multiple sites at a given promoter, in both possible orientations and at various distances from the site bound by RNAP [2,3]. Activator binding to a promoter may require co-factors that increase the sensitivity of the regulation [4,5] and/or overcome the silencing effects of nucleoid-associated proteins [6]. Although we understand through atomic resolution the interactions that certain transcriptional activators establish with promoter sequences [7,8], it is still unclear which particular arrangement of *cis*-acting regulatory features, such as the number, orientation, location and sequence recognized by an activator (i.e., the promoter architecture), can be used by a given transcriptional activator to promote gene transcription. This merits especial interest with regard to activators that control multiple targets because genes co-regulated by a particular activator protein are often expressed in distinct fashions, and this could be due to the corresponding promoters having different architectures. Moreover, it is not clear if these architectures are unique for each regulated promoter or if there are classes of architectures shared by several promoters.

PhoP is a transcriptional activator that governs virulence and Mg^{2+} homeostasis in several enteric species [9]. PhoP regulates expression of ~5% of the genes in the Gram-

negative pathogen *Salmonella enterica* serovar Typhimurium [10]. The regulation is direct for some of these genes where the PhoP protein binds to the corresponding promoter regions [11] and indirect for other genes where PhoP alters the levels and/or activity of regulatory proteins [10] or regulatory RNAs [12]. Analysis of the genes directly activated by the PhoP protein has revealed that the majority exhibits a limited phylogenetic distribution and appears to have been acquired by horizontal gene transfer. Moreover, the encoded products differ in their biochemical functions and physiological roles [10] (Table S1), indicating that they are produced in distinct amounts when the bacterium experiences PhoP-inducing conditions.

The PhoP protein binds to its DNA targets *in vivo* only when phosphorylated [11,13]. This process is controlled by the PhoQ protein, which advances the phosphorylation of PhoP (PhoP-P) when the bacterium experiences low extracytoplasmic Mg^{2+} [14] or antimicrobial peptides [10,15]. PhoP-P binds to a hexanucleotide direct repeat separated by 5 nucleotides in its target promoters designated PhoP box [16], and this could reflect the fact that PhoP-P binds its targets as a dimer [17]. Among the relatively small number of PhoP-activated promoters experimentally defined to date, the PhoP box can be found at various distances and in both possible orientations with respect to the -10 hexamer sequence recognized by RNAP [18,19], implying that PhoP utilizes different mechanisms to promote gene transcription [1,5]. Supporting this notion, *in*

vitro experiments revealed that the C-terminal domain of the α subunit of RNAP is required for transcription of some PhoP-activated promoters but not of others [20,21].

Here, we use a combination of computational, genetic and molecular approaches to define the rules that govern PhoP-dependent transcriptional activation. We examined the promoter DNA sequences for: the presence and number of PhoP binding sites; the location of the PhoP binding sites defined as the distance between these sites and the -10 hexamer sequence recognized by RNAP; the orientation of each PhoP binding site relative the RNAP binding site, where direct orientation refers to the 5'-3' PhoP binding site sequence pointing towards the -10 hexamer sequence, and reverse orientation refers to the PhoP binding site sequence pointing away from that sequence; and the sequence elements characteristic of promoters recognized by the σ^{70} form of RNAP, including the -10 hexamer, the -35 hexamer, and the distance between the two hexamers. Our analysis demonstrates that the PhoP protein utilizes at least five promoter architectures to drive transcription of its activated genes. These architectures are composed of specific (as opposed to arbitrary) combinations of *cis*-acting regulatory elements. In addition, we establish that the different promoter architectures correspond to different mechanisms by which PhoP controls expression of its activated genes, and different architectures are used to differentially control ancestral vs. horizontally-acquired genes. Genomic analysis using a combination of chromatin immunoprecipitation (ChIP-chip) and custom expression microarray experiments (Nimblegen tiling arrays) in different

gamma/enterobacteria indicates that certain promoter architectures are species-specific, and their changes through evolutionary turnover events may represent a significant source of inter-species variation.

RESULTS

Exploring the architectural diversity of PhoP-activated promoters

Transcriptomic and chromatin immunoprecipitation (ChIP) experiments have allowed the identification of a large number of genes directly activated by the PhoP protein [19]. However, except for a limited number of cases [18], the particular sequences recognized by the PhoP protein at the corresponding promoters, as well as the location of these sequences relative to the transcription start sites, have remained unknown. Therefore, to define the scope of PhoP-activated promoters, we determined the transcription start sites for 14 genes known to be directly activated by the PhoP protein by carrying out S1 mapping experiments with RNA harvested from isogenic wild-type and *phoP Salmonella* strains grown under PhoP-inducing (i.e., 10 μM Mg^{2+}) or –repressing (i.e., 10 mM Mg^{2+}) conditions. Most PhoP-activated genes had a single start site (Fig. S1A), which was detected in the wild type strain in low Mg^{2+} but not in the PhoP-mutant regulation of the growth condition. However, few genes had additional start sites that are PhoP-

independent, as observed also in the *phoP* mutant whether grown in high or low Mg^{2+} and in the wild-type strain grown under noninducing conditions (Fig. S1A).

To define the sequences recognized by the PhoP-P protein at its activated promoters, we carried out DNase footprinting analysis of the promoter regions corresponding to 15 PhoP-activated genes utilizing phosphorylated PhoP-His6. PhoP-P protected a single site in some promoters and two sites in others (Fig. S1B). In the latter case, PhoP-P exhibited differential affinity for the two sites (Fig. S1B). Then, we conducted a bioinformatics analysis of 20 PhoP-activated promoters corresponding to the promoters discussed above as well as to additional promoters whose transcription start sites and PhoP-binding sites had been previously reported [22,23,24,25]. This analysis entailed inspecting the promoter DNA sequences for the following features (Fig. S2): 1) the number of PhoP boxes (i.e., sequences resembling the hexanucleotide direct repeat (T/G)GTTTA separated by five nucleotides [16]); 2) the location of the PhoP box(es) defined as the distance between the PhoP box and the -10 hexamer sequence recognized by RNAP; 3) the orientation of each PhoP box relative to the -10 hexamer sequence recognized by RNAP, where direct orientation refers to the half PhoP box sequence 5' (T/G)GTTTA 3' pointing towards the -10 hexamer sequence, and reverse orientation where the half PhoP box sequence is pointing away from the -10 hexamer sequence; and 4) the presence of sequence elements characteristic of promoters recognized by the σ^{70} form of RNAP, including the -10 hexamer, the -35 hexamer, the distance between the

two hexamers as well as the extended -10 hexamer [1,5,26]. To facilitate the comparison of the 20 promoters, we aligned the corresponding sequences with respect to the RNAP -10 hexamer using the 5′ most edge of this element as a referential landmark instead of the typically used transcription start site (Fig. S2) [7,27].

We determined that 7 of the 20 promoters harbor a single PhoP box, which is always in the direct orientation (Fig. S2). For some of these promoters, the PhoP box is located 12 nt upstream of the -10 hexamer at the position normally occupied by the -35 hexamer. This location suggests that PhoP activates transcription from these promoters by interacting with the σ subunit of RNAP [5,20,21]. For the other PhoP-activated promoters harboring a single PhoP box, the PhoP box is located upstream of the -35 hexamer sequence (Fig. S2), implying that their activation involves interactions with a subunit of RNAP different from the first group [1,5]. The remaining 13 promoters harbor two PhoP boxes, which can be found in the direct or the reverse orientations (Fig. S2). For promoters harboring two PhoP boxes, one of the PhoP boxes was invariably located at the site normally occupied by the -35 hexamer or upstream of this region (Fig. S2). By contrast, the location of the second PhoP box varied among this group of promoters: in some cases, it was located further upstream of the first PhoP box, but in other cases, it was found downstream of the first PhoP box, including instances where it overlapped with the -10 hexamer or was present downstream of the

transcription start site (Fig. S2). These findings indicate that some of the PhoP boxes might function as repression sites despite being part of PhoP-activated promoters.

In sum, PhoP-activated promoters may harbor one or two PhoP boxes, which can be present in either orientation and at various distances from the -10 hexamer for RNAP. Thus, we seeked to define the combinations of PhoP box location and orientation that enable PhoP-dependent gene transcription. In some cases, we modified natural PhoP-activated promoters, but in others, we engineered sets of synthetic promoters with the same PhoP box sequence so as to focus on the *cis*-acting regulatory elements (other than a match to the PhoP box consensus sequence) that govern the functionality of PhoP-activated promoters. The PhoP-dependent promoters were fused to a promoterless *gfp* gene in a medium copy number plasmid. We then measured the fluorescence produced by either wild-type *Salmonella* or an isogenic *phoP* mutant harboring the constructed plasmids. Promoters labeled with the suffix "4" (e.g., *phoP4*) are derivatives of the original promoters including only the DNA fragment corresponding both to the PhoP box(es) identified in DNase footprinting assays and to the RNAP -10 and -35 hexamers identified in the S1 mapping experiments. For example, a derivative termed *phoP4* with as little as 52 nt upstream and 5 nt downstream of the start site displayed a behavior that was virtually identical to that of the original *phoP* plasmid (Fig. 1A), indicating that all the sequence information necessary for Mg^{2+}-regulated PhoP-dependent transcription was present in the 58 bp DNA fragment.

The PhoP box mediates gene activation in an orientation-dependent manner

Classical promoter enhancers retain the ability to activate gene transcription when placed in the opposite orientation [28]. To test whether a PhoP box that is normally in one orientation can function when present at the same position but in the opposite orientation, we first made derivates of the natural *phoP* and *rstA* promoters where the 17 nucleotides corresponding to their single PhoP box in the direct orientation were placed in the opposite orientation (Fig. S3). The wild-type strain harboring the resulting plasmids produced no fluorescence (Fig. 1A), in contrast to the fluorescence displayed by the strains harboring the plasmids with the original *phoP* and *rstA* promoters (Fig. 1A). These data demonstrate that the PhoP box functions in an orientation-dependent manner, at least when located 12 nt or 23 nt upstream of the -10 hexamer sequence of RNAP.

We then examined whether the PhoP box from the *phoP* promoter, which is normally present in the direct orientation, can functionally replace one of the two PhoP boxes in the *mgtC* and *pagK* promoters, which are normally present in the opposite orientation (Fig. S3). We determined that both engineered derivatives of the *mgtC* and *pagK* promoters were functional, indicating that a given PhoP box can function in both possible orientations (Fig. 1B). However, GFP expression became constitutive (i.e.,

PhoP-independent) when the PhoP box in the engineered *pagK* promoter was placed in the direct orientation (Fig. 1C). These data establish that a given PhoP box can function in both possible orientations, and that PhoP box functionality is dictated by a combination of orientation and distance from the -10 hexamer site of RNAP.

The PhoP box mediates gene activation in a location-dependent manner

The location of a transcription factor binding site may restrict its interaction with RNAP [5] and thus its ability to activate gene transcription [27,29]. To explore the range of distances over which a directly oriented PhoP box can promote gene activation, we engineered plasmids carrying GFP fusions to promoters harboring the PhoP box from the *phoP* promoter at various distances from the RNAP -10 hexamer. The PhoP box mediated gene transcription when located 12, 23 or 33 nt upstream of the RNAP -10 hexamer (Fig. 1D). Similar fluorescence values were produced by strains carrying constructs with the *phoP* PhoP box 12 or 23 nt upstream of the RNAP -10 hexamer (Fig. 1D), the normal respective locations of the PhoP box in the *phoP* and *rstA* promoters (Fig. S2). The *phoP* PhoP box could still drive *gfp* transcription when located 33 nt upstream of the -10 hexamer even though there are no examples of natural PhoP-activated promoters with a directly oriented PhoP box at this location (Fig. 1D). Yet, the

strain with the latter construct produced only half as much fluorescence as the strains with the former two plasmids (Fig. 1D).

The two promoters just described share the orientation and sequence of the PhoP box but not its location, which curiously is one or two integral turns of a DNA helix away from the position of the PhoP box in the other promoters. This suggests that the PhoP-P protein would be positioned on the same face of the DNA when promoting transcription from these three promoters. If being on the same face of the DNA is critical for PhoP-activated gene transcription, then altering the distance by a number of nucleotides different from 10-11 (corresponding to one turn of a DNA helix) should abolish gene transcription. As predicted, there was little fluorescence in strains in which *gfp* transcription was driven by a promoter with a directly oriented PhoP box located 7, 17 or 28 nt upstream of the -10 hexamer sequence (Fig. 1D). Similarly low levels of fluorescence were displayed by strains where the PhoP box was at positions 25, 26, or 27 nt upstream from the -10 hexamer sequence (Fig. 1D). Finally, when the PhoP box was located 21, 22 or 24 nt upstream from the -10 hexamer, the expression was ~25% of that observed when the PhoP box was at position 23 (Fig. 1D).

Next, we examined the locations at which the PhoP box could promote gene transcription when present in the reverse orientation (Fig. 1E). We created artificial promoters analogous to those described above, except that the PhoP box was present in the reverse orientation at various distances from the -10 hexamer sequence, some of

which correspond to locations normally occupied by a reverse-oriented PhoP box in natural promoters (Fig. S2). The PhoP box promoted transcription when placed 27, 37 or 48 nt upstream of the RNAP -10 hexamer, and the expression increased as the distance of the PhoP box to the -10 hexamer increased (Fig. 1E). In contrast to promoters harboring the PhoP box in the direct orientation, the PhoP box in the three promoters with the PhoP box in the opposite orientation is separated by half-integral turns of the DNA helix from the RNAP -10 hexamer. Moreover, consistent with the notion that being on a particular face of the DNA is necessary for PhoP-promoted gene transcription, there was little expression when the PhoP box was present 28, 29, 30, 31 or 32 nt upstream of the -10 hexamer sequence (Fig. 1E).

These results demonstrate that the PhoP box can function from more than one location up to a certain distance from the RNAP -10 hexamer, and spaced by integral and half-integral turns of the DNA helix from this promoter sequence. At any given position, the PhoP box is active only in one of the two possible orientations: the PhoP box conferred high levels of expression when present 48 nt upstream of the -10 hexamer in the reverse orientation but no expression when in the direct orientation (Fig. 1E); the converse was true when the PhoP box was present 12 nt upstream of the -10 hexamer sequence (Fig. 1E).

A second PhoP box can mediate activation or repression in PhoP-activated promoters

Some transcriptional activators bind to more than one site in a promoter [30,31]. A particular binding site can mediate activation, repression or play no role in gene transcription [5,32]. Of the 20 PhoP-activated promoters, 14 harbor more than one PhoP box (Fig. S1B), which can be located at different positions with respect to each other and to the RNAP -10 hexamer (Fig. S2). For example, both PhoP boxes in the *ybjX* promoter are located upstream of the RNAP -10 hexamer; one PhoP box is located upstream, but the other overlaps the RNAP -10 hexamer in the *ugtL* [24] promoter. In addition, two of the three PhoP boxes in the *mgtC* promoter are located downstream of the transcription start site (Fig. S2). This raises a question as to the specific role(s) played by the individual PhoP boxes in the PhoP-activated promoters. We therefore examined the fluorescence produced by strains harboring a plasmid with a promoterless *gfp* gene driven by wild-type and PhoP box mutant derivatives of these PhoP-activated promoters (Fig. S3).

The *ybjX*, *virK*, and *mig-14* promoters harbor two PhoP boxes, one of which is located 12 nt from the RNAP -10 hexamer and spaced at integral turns of the DNA helix from this hexamer (Fig. S2). We determined that mutation of the PhoP box at this location abolished expression of the corresponding promoter::fusions in all the promoters (Fig. 2A-C). By contrast, mutation of the second PhoP box in the *ybjX*, *virK*, and *mig-14* promoters, which are located 30, 26, and 8 nt upstream of the first PhoP box, respectively, reduced but did not eliminate GFP expression (Fig. 2A-C). If having a

second PhoP box upstream of a PhoP box located 12 nt upstream of the -10 hexamer increases gene transcription, then adding a second PhoP box to a promoter that normally has a single PhoP box should result in increased levels of expression. As hypothesized, adding a second PhoP box 26 nt upstream of the first PhoP box, which is the location of the second PhoP box in the *virK* promoter, improved the transcription levels of the *phoP* promoter (*phoP4/2*, Fig. 2D).

The *ugtL*, *mgtC*, and *pcgL* promoters also harbor at least two PhoP boxes (Fig. S2), one of which was required for activation ([24], Fig. 2E-F). A second PhoP box in these promoters is located downstream, overlapping, and further upstream of the RNAP -10 hexamer (Fig. S2). Because binding sites involved in repression can be located anywhere as long as they block or limit transcription initiation or elongation [5], we hypothesized that a second PhoP box in the *ugtL*, *mgtC*, and *pcgL* promoters can function as a repression site. This is the case of a second PhoP box in the *ugtL* promoter, which overlaps the RNAP -10 hexamer, and it has been demonstrated that PhoP binds to this site to repress transcription [24]. We then examined the presence of a second and a third PhoP box located downstream of the transcription start site in the *mgtC* promoter. To test the role of these PhoP boxes, we first made a shortened derivative of the PhoP-activated *mgtC* promoter (*mgtC4*, Fig. 2E), which lacks the downstream PhoP boxes. The shortened construct produced >18000-fold higher GFP activity than that of the full-length promoter fragment, indicating that these are PhoP boxes functioning as

repression sites. To explore other locations over which a PhoP box can function as a repression site, we mutated a second PhoP box in the *pcgL* promoter (Fig. S3), which is positioned 75 nt upstream of the RNAP -10 hexamer (Fig. S2). Mutation of this PhoP box resulted in derepressing activity *(pcgL4/dn*, Fig. 2F*)*.

In sum, PhoP-activated promoters must have at least one PhoP box involved in activation, whereas the other may function either in activation or repression. The PhoP boxes involved in transcription activation are located at a certain orientation, distance and face with respect to the RNAP binding site to promote interactions between the PhoP protein and the RNAP. A second PhoP box located upstream and at a relatively close distance from an activation PhoP box may also contribute to promoter activation. In contrast, a second PhoP box required for repression is located further upstream, may overlap or lie downstream of the RNAP binding site to modulate transcription in PhoP-activated genes.

The PhoP box mediates gene activation by promoting different interactions between the PhoP protein and the RNAP

Promoters controlled by inducible systems typically lack strong RNAP binding sites because, otherwise, these promoters would be constitutively expressed. This is the case of PhoP-activated genes; sequence analysis of the RNAP binding sites identified from

the S1 mapping assay in the 20 promoters revealed that most of these sites display a weak resemblance to the canonical RNAP motif [5], as judged by the presence of conserved -10 and/or -35 elements and the distance between them (Table S2). Furthermore, analysis of the distance between the PhoP box and the RNAP binding site revealed that 10 of the 20 examined promoters have one PhoP box completely overlapping the -35 element (Fig. S2, Table S2). As expected, the -35 sequences of these promoters weakly resemble the -35 consensus [5] because the PhoP protein recognizes a different DNA sequence at the same location.

In contrast to those promoters described above, the remaining 10 promoters harbor PhoP boxes that do not overlap the RNAP binding site (Fig. S1B and Fig. S2). Therefore, how do the quality of the RNAP -10 and -35 hexamers affect the expression levels of these promoters, where the PhoP protein is likely to require interaction with the α subunit of the RNAP [5,33]? To answer these questions, we engineered GFP expression plasmids carrying promoters without overlapping between the PhoP box and the RNAP binding site and harboring either weaker or stronger -10 or/and -35 hexamer motifs than those of the wild type promoters (Fig. S3).

We first made derivates of the natural *phoP* promoter, where the PhoP box in the direct orientation was placed 11 nt upstream of its original site in the natural location of the *rstA* promoter, and where it was placed in the opposite orientation, 48 and 37 nt upstream of its original location in the natural location of the *mgtC* and *pagK* promoters,

respectively. Second, we examined whether the – 10 hexamer could functionally be replaced by either a less conserved or a perfect consensus sequence [34]. We determined that replacing the -10 element ("CATAAT") in the *rstA* promoter-based construct by a less conserved sequence ("TATGTT") reduced the GFP levels 4-fold (*phoP4*+11-10, Fig. 3A). By contrast, introducing the consensus -10 sequence in the *mgtC* promoter-based construct increased fluorescence of the strain with the wild-type plasmid >3-fold (*phoP4*/*revC*, Fig. 3B). Third, we explored the effect of introducing the -35 consensus sequence [34] in the *pagK* promoter-based construct, which increased >30-fold the GFP levels produced by the strain with the original promoter plasmid (*phoP4*/*revK*, Fig. 3C). This means that the RNAP -10 hexamer does affect the transcription levels of PhoP regulated genes, but the RNAP -35 hexamer has a larger effect on transcription.

If having a better RNAP increases gene transcription, then adding the consensus -10 sequence and the -35 sequence from a promoter where the PhoP box does not overlap a normally weak RNAP binding site should result in increased levels of expression of the corresponding promoter. However, it might also result in increased constitutive expression, suggestive of why the majority of the PhoP-activated promoters lack strong RNAP binding sites. To test this hypothesis we engineered plasmids carrying GFP fusions to promoters harboring a relatively weak RNAP binding site from the *phoP* promoter as well as a relatively strong RNAP binding site from the *mgtC* promoter.

Both constructs shared the same PhoP box, in the reverse orientation, and located 48 nt upstream of the RNAP -10 hexamer. As hypothesized, we determined that gene transcription mediated by the strong RNAP exhibited >4-fold higher GFP levels than those from the weak one (Fig. 3D). We then compared the same constructs in noninducing conditions (i.e., high Mg^{2+}), and determined that ~1/5 of the expression produced by the strain harboring the plasmid with the RNAP from the *mgtC* promoter was constitutively produced, whereas that with the RNAP from the *phoP* promoter produced nonsignificant levels of fluorescence in noninducing conditions. This suggests that a strong RNAP increases the constitutive levels of PhoP-activated promoters, but their expression still depends on the PhoP protein (i.e., ~4/5 of the total expression) (Fig. 3E). Moreover, exceptionally strong effects of the RNAP in gene transcription such as those exhibited by the *mgtC* promoter are controlled by the PhoP protein via repression PhoP boxes (Fig. 2E).

Consistent with not overlapping the -35 hexamer, the PhoP protein may recruit RNAP to the promoter by directly interacting with the RNAP α-CTD, as is the case in the PhoP-activated *ugtL* promoter [33], where the PhoP box is 30 nt upstream of the -10 hexamer. To explore this possibility, we performed *in vitro* transcription assays comparing wild-type and mutant RNAP (α Δ-CTD). We determined that the α-CTD subunit of RNAP is required to promote *in vitro* transcription of two genes with PhoP boxes that do not overlap the RNAP binding site (Fig. S4): the *rstA* gene, with a PhoP

box in the direct orientation, located only 23 nt upstream of the RNAP -10 hexamer, as well as the *pagC* gene, with a PhoP box in the reverse orientation, located 47 nt upstream of the same hexamer. These data suggests that PhoP interacts with the RNAP α-CTD when the PhoP box does not overlap the RNAP binding site, despite its orientation and relative distance to that site. In contrast, the PhoP protein may contact the σ subunit [20,21] of the RNAP in promoters with a PhoP box overlapping the -35 hexamer, as transcription of the *phoP* and *mgtA* promoters could still take place *in vitro* using an RNAP lacking the corresponding α subunit of the C-terminal domain (α-CTD) [1,5].

DISCUSSION

The fundamental mechanism governing inducible gene expression operates at the level of transcription initiation whereby DNA binding proteins recognize specific sequences in the promoters to activate or repress gene transcription by RNAP. Although the interaction between these proteins and their target sequences *in vitro* can be characterized thoroughly, how a transcription factor recognizes short sequences with different specificity in the context of a large promoter region to control transcription initiation *in vivo* is poorly understood. Here, we report that the PhoP protein recognizes its functional PhoP boxes within the context of several *cis*-acting features, which are precisely organized in the promoters, to produce distinct outputs of the PhoP gene

expression program *in vivo*. We applied a combination of computational, genetic and molecular approaches to identify key *cis*-acting features and their organization in each of 20 PhoP-activated promoters. We then constructed synthetic promoters based on these natural promoters to make the following determinations. First, the PhoP boxes can function in either the direct or the reverse orientation with respect to the RNAP -10 hexamer, where each orientation has a different range of functional locations. Second, PhoP-activated promoters may have more than one PhoP box, which can activate or repress transcription. PhoP boxes involved in transcription activation are located at specific distances from the RNAP -10 hexamer, to allow interactions between the PhoP protein and the RNAP. PhoP boxes required for repression are located further upstream or may overlap or lie downstream of the RNAP binding site. Third, PhoP-activated promoters have relatively weak RNAP binding sites, where the PhoP protein interacts either with the σ or the α subunit of the RNAP, depending on the location of the PhoP box with respect to the RNAP binding site.

Nonarbitrary architectural organization of PhoP-activated promoters

Five main architectures were identified by grouping promoters displaying similar organizations of *cis*-acting elements (Fig. 4), which are responsible for the diverse regulatory program carried out by the PhoP protein. Architecture I includes promoters harboring a single PhoP box in the direct orientation with respect to the RNAP, located

12 nt upstream of the RNAP -10 hexamer, i.e., one turn of the DNA helix away from this hexamer. Architecture II promoters harbor a single PhoP box in the direct orientation but are located 23 nt (two turns) upstream of the RNAP -10 hexamer. Architecture III promoters harbor two PhoP boxes, one in a direct orientation, with location and phasing similar to the single PhoP box in architecture I and II promoters, and another PhoP box located > 45 nt upstream of this site. Architecture IV promoters harbor two PhoP boxes, one at similar orientation, location and phasing to the single PhoP box in architecture I promoters, and the other located < 30 nt upstream of this site. Architecture V promoters also harbor two PhoP boxes, one in the reverse orientation with respect to the RNAP 27, 37, or 48 nt upstream of the RNAP -10 hexamer, separated by half-integral turns of the DNA helix from this hexamer. The other PhoP box either overlaps the RNAP -10 hexamer or is located downstream of the transcription start site. These data demonstrate that the architectures are specific and nonarbitrary combinations of *cis*-acting elements in natural PhoP-activated promoters, because each of the five organizations of these elements is shared by more than one promoter, instead of being randomly arranged for each individual promoter.

Promoter architectures indicate that PhoP utilizes different mechanisms to promote gene transcription

Most transcription factors activate transcription by making contact with either the α subunit (specifically, the C-terminal domain [CTD]) or the σ^{70} subunit (the most commonly used σ factor) of RNAP, both of which can bind DNA [35]. The PhoP protein appears to interact with the α-CTD in promoter architectures having a PhoP box that does not overlap with the RNAP binding site. In addition, the PhoP protein, like the PhoB and VanR regulatory proteins [20,21], is also predicted to contact the σ subunit of RNAP in promoter architectures having a PhoP box completely overlapping the RNAP -35 hexamer. Some promoters (e.g., *mig-14*) may combine both of these mechanistic variants [1,5], where one PhoP box, overlapping the RNAP -35 hexamer, interacts with the σ subunit, and the other, located upstream of this site, interacts with the α-CTD. This strongly supports the idea that the PhoP protein can interact with RNAP in several different ways. For example, transcription of the *phoP* promoter with a PhoP box in the direct orientation and overlapping the RNAP -35 hexamer is α-CTD independent, but the same PhoP box is also functional in the same orientation when located 11 nt upstream of that position (Fig. 2D), the location of the PhoP box in the α-CTD dependent *rstA* promoter (Fig. S4A). Despite the relocation of the PhoP box, and thus the "rewiring" of the PhoP-RNAP interaction, the functionality of the *phoP* promoter is preserved, and similar transcription levels from the respective promoters are produced [27].

Analysis of cAMP-receptor (CRP) activated promoters revealed that activation occurs when the distance between the CRP box and the RNAP -10 hexamer correspond to integral or near-integral turns of the DNA helix (Fig. S5A) [27]. However, no CRP activation was observed when its site was located at half-integral turns distant, which are rotated by roughly 180° relative to RNAP bound at the RNAP -10 hexamer (Fig. S5B), presumably due to a lack of contact between CRP and RNAP [27]. This indicates that the interactions between CRP and RNAP may involve a canonical set of contacts that promote transcription when they bind DNA at integral turn distances. Similar to CRP, the PhoP protein is able to activate transcription when its PhoP boxes are located at integral turns away from the RNAP -10 hexamer (e.g., architecture II promoters). This activation, like that of CRP, requires the α-CTD of RNAP (Fig. S4A). Unlike CRP, PhoP is also able to activate transcription when its PhoP box is in the reverse orientation and located at half-integral turns away from the RNAP -10 hexamer (e.g., architecture V promoters, Fig. S5C). At these half-integral turn distances, there is no activation if the PhoP box is in the direct orientation. Thus, depending on the orientation and location of the PhoP box, PhoP-RNAP interactions must differ from those established by CRP-RNAP.

The structure of the CRP dimer (Fig. S5E) [27,36,37,38] and the regulatory domain PhoP protein (Fig. S5F) are mirror symmetric [17]. However, it has been postulated that the DNA-binding domain is asymmetrical in the PhoP protein (Fig. S5F) [17]. If PhoP,

like the PhoB protein [20], has a head-to-tail orientation of the DNA binding domains when bound to DNA, consistent with the tandem arrangement of the PhoP box [16], then this asymmetry may be the reason for the different, orientation-dependent interactions between PhoP and RNAP. In sum, the ability of the PhoP protein to interact with the σ subunit of the RNAP, as well as with the α-CTD when the PhoP box is either at 0° degrees in the direct orientation or rotated by roughly 180° relative to the RNAP binding site in the reverse orientation, differentiates this protein from others.

Promoter architectures combine activation and repression binding sites that modulate gene transcription

Two of the identified promoter architectures have at least one PhoP box acting as an activation site and another functioning to repress transcription (Fig. 2E-F). The net transcription using these architectures reflects a balance between the binding of PhoP to the activation and repression sites. Similar, but not equal, forms of regulation have been reported where different activators and repressors bind to different sites in the same promoter to modulate transcription [1,32,39]. For example, the CytR-repressed promoters in *E. coli* depend on activation by the CRP protein [31,40], and some FNR-activated promoters are also repressed by the IHF and the Fis proteins [41,42]. Yet, the same protein can act as an activator in one context and as a repressor in another [2]. For example, the *E. coli* GalR [32] and MerR [43] proteins have dual activation and

repression behaviors depending on the promoter with which they interact, or the condition under which they were induced, respectively. The λcI protein activates or represses its targets by contacting distinct subunits of the RNAP [3]. Interestingly, for the same RNAP promoter and under the same inducing conditions, PhoP binds at one site to activate and at another site to repress transcription. We demonstrated that a clear function of these repression sites is to modulate transcription levels, especially in the presence of a strong RNAP binding site, as was shown for the *mgtC* promoter. Presumably, this is because an inducible system like PhoP/PhoQ has to maintain a balance between the expression levels of its targets when induced, and the level of constitutive transcription under noninducing conditions.

Promoter architectures differ between ancestral and horizontally-acquired genes

The PhoP protein regulates genes that are expressed at different levels; some of these genes are ancestral (e.g., architecture I and II promoters (Fig. 4, Fig. S6A and Fig. S6C)), but others were horizontally-acquired (e.g., architecture III-V promoters (Fig. 4, Fig. S6B, and Fig. S6C)). (Ancestral versus horizontally-acquired genes were distinguished based on the Conservation Scores (CS) of their ORFs, and their corresponding GC-content (Fig. S6)). Therefore, what changes, if any, might a newly appearing, horizontally-acquired gene require to be appropriately regulated by a transcription factor? The high conservation of PhoP box sequences within gamma/enterobacterias, as

well as of PhoP itself [33], suggests that the major effect of differential regulation between species is not due to differences in sequences of the binding site targets [19], as was previously suggested for other regulators [44], but rather is due to the context where they are located (i.e., promoter architecture). Thus, we reason that a variety of promoter architectures associated with different mechanisms used by PhoP to control its target genes may facilitate the integration of new members from different origins to the PhoP regulon.

Several PhoP-activated promoters harbor more than one PhoP box. However, it is not yet clear why two PhoP boxes are needed for activation instead of a single high affinity target for PhoP. A possible explanation would be that one of the two PhoP boxes may have roles other than interacting with RNAP to facilitate transcription initiation. Promoters with more than one PhoP box are typically associated with horizontally-acquired genes (Fig. S6B-C) whose levels of transcription are often influenced by nucleoid-associated proteins that interfere with the binding of RNAP or a transcriptional activator [6,25]. This is the case of the promoter architecture V *pagC* and *ugtL* genes, which are silenced by H-NS and require the PhoP-activated SlyA protein to overcome this silencing effect [25]. Similarly, the promoter architecture IV *virK* gene harbors two activation PhoP boxes and is also silenced by the H-NS protein [45]. The first PhoP box is absolutely required for transcription from this promoter, while the second has a lesser effect on transcription (Fig. 2B). In contrast to the first PhoP box, the

location of the second PhoP box differs from that of the PhoP boxes utilized by PhoP for transcription activation, indicating that PhoP binding to the second PhoP box may have a different role, such as overcoming H-NS-mediated repression.

Promoter architectures are species specific

Characterization of interspecies differences in gene regulation is central to understanding the molecular basis of both differential gene expression and evolution. The evolution of a regulatory system depends on changes in three nonexclusive components of these systems: 1) the regulatory protein, and/or 2) the co-evolution of the target genes (i.e., orthologs), and/or 3) alterations in their regulation, including the affinity of the regulatory protein to its binding sites and/or the *cis*-acting features composing the promoter architecture. Therefore, we explored the possible changes in the components of the PhoP system across the gamma/enterobacteria.

First, evaluation of the DNA-binding domain of the PhoP protein, like other regulators of two-component systems [46], revealed that it has not changed significantly throughout the gamma/enterobacteria (*e-value* <1E-5, expected number of false positives in a reciprocal BLAST search [47]). Nonetheless, the *Yersinia* PhoP protein can bind to the *Salmonella mgtA* and *ugtL* promoters, but cannot recruit RNAP and promote transcription of the *ugtL* gene [33]. Because the *ugtL* gene is horizontally-acquired in

Salmonella, changes in the regulation can be due to a change in the regulatory protein and/or in the target gene.

Second, analysis of genes directly regulated by PhoP in the gamma/enterobacteria revealed that few are relatively well conserved (Fig. S6A-B), indicating that most genes have a high chance of being gained or lost during evolution [48]. Moreover, a subset of these genes is unlikely to be true orthologs because the level of amino acid identity between the corresponding products is considerably lower than the median identity between species placed at the opposite boundaries of the gamma/enterobacteria evolutionary tree [33,49]. The diversity of these genes suggests that there must be multiple ways (i.e., mechanisms) utilized by the PhoP protein to include them functionally in its regulon.

Third, it has been proposed that the affinity of a transcription factor for its binding sites in target promoters, and their corresponding binding site classes of motifs, is the major cause of divergence between species [44]. However, we demonstrated that this was not the case for the PhoP-activated promoters because most of the submotifs present in *Salmonella* are also present in most of the gamma/enterobacteria, with a remarkably low rate of change [19]. Our previous computational analysis anticipated that the binding sites are not the major cause of divergence between species in the PhoP system, but rather, that changes in the context where they are located (i.e., promoter

architecture) through evolutionary turnover events represent a significant source of interspecies variation [19].

Here, divergence in the promoter architectures utilized by PhoP to control its target genes in *Salmonella* [19], *Klebsiella pneumoniae* (Table S7), and *Yersinia* pestis KIM [48] was explored by means of genome-wide searches using a combination of promoter binding by the PhoP protein, measured by a ChIP-chip assay and identified by using the *D&C* method [19], and gene expression, measured by custom expression microarray experiments of wild-type and *phoP* mutated strains (Nimblegen tiling arrays). We first established that several *cis*-acting features of the PhoP regulated promoters are conserved among PhoP-activated genes in the three species. For example, the single PhoP box required for activation in the direct orientation with respect to the RNAP -10 hexamer is present in the *phoP* and *slyB* promoters in all three species, and also in the *Salmonella* and *Klebsiella mgtA* and *ompX* promoters, as demonstrated by significant gene expression levels, ChIP peaks (i.e., Log_2-ratio of the ChIP-chip signal intensity detected using a 500 bp sliding window), and PhoP box motifs in their promoter regions [19,48] (Table S7). In addition, ~30% of the PhoP-activated genes in *Klebsiella*, harboring significant expression and ChIP peaks, exhibit a putative PhoP box in the reverse orientation with respect to their corresponding ORFs (Table S7), like the PhoP-activated promoters in *Salmonella*. Furthermore, we identified additional PhoP boxes in several *Yersinia* [33] and *Klebsiella* (Table S7) PhoP-activated promoters. Those additional PhoP

boxes in *Yersinia* are low-affinity sites footprinted by the PhoP protein [33]. Some of these sites either overlap the RNAP -10 hexamer or are located downstream of the transcription start site [19], as in the *Salmonella* promoters (Fig. 4). These PhoP boxes were shown to repress transcription when bound by PhoP in *Salmonella* (Fig. 2E) [24].

Even though the individual *cis*-features are apparently well conserved throughout the evaluated species, their organization as promoter architectures can vary. We reason that the relatively rapid change of the PhoP-activated genes in the gamma/enterobacteria may require promoter architectures flexible enough across these species to allow them to be incorporated into the PhoP regulon and to be precisely controlled by the PhoP protein. Therefore, we compared the organization of the *cis*-acting features corresponding to PhoP-activated promoters among the *Salmonella*, *Klebsiella*, and *Yersinia* genomes, using as key criteria the orientation and location of the PhoP box for differentiating promoter architectures as well as their associated mechanisms.

We then determined that the PhoP boxes required for activation in the direct orientation, located 12 nt and 23 nt upstream of the RNAP -10 hexamer, and separated by integral turns of the DNA helix from that hexamer, occur both in *Salmonella* (Fig. 1) and *Yersinia* [33] PhoP-activated promoters. However, none of the *Yersinia* PhoP-activated promoters apparently harbor a PhoP box in the reverse orientation, located 27, 37 or 48 nt upstream of the RNAP -10 hexamer, and separated by half-integral turns of

the DNA helix from that hexamer [33], which is the PhoP box location in *Salmonella* promoters displaying architecture V (Fig. 1E and Fig. 4). In particular, the PhoP box in the *Yersinia* xenolog is in the direct orientation to that of the *Salmonella mgtC* gene (Fig. S5, CS = 63%, *e-value* <2E-65 [47]) and located 34 nt upstream from the RNAP -10 hexamer [48], a location not observed in any of the 20 natural promoters evaluated in *Salmonella* (Fig. S2). Moreover, regardless of the conservation of the gene product, the *Klebsiella* promoters of the *pmrD* (CS = 44%, *e-value* < 4E-10 [47]) and *rstA* (Fig. S5, CS = 80%, *e-value* <1E-104E-65 [47]) genes harbor a single PhoP box in the reverse orientation with respect to their corresponding ORFs (Table S7), differing from the PhoP box in the direct orientation in the orthologous *Salmonella* promoters (Fig. S2). This suggests that PhoP may use different mechanisms to interact with RNAP at these *Salmonella* and *Klebsiella* promoters.

Overall, we determined that divergence of promoter architectures across the gamma/enterobacteria far exceeds the interspecies variations in the PhoP protein, or the PhoP box affinity, as deduced by conservation of the PhoP box submotifs. This raises the question of why changing the promoter architectures is the preferred strategy to regulate genes. One possibility is that the *cis*-acting features constitute a flexible kind of "molecular Lego" [29,50], which can be assembled in different combinations to effect the desired expression levels. This should be contrasted with an alternative strategy of regulation based on altering complex (e.g., allosteric) protein-protein interactions. A

notable advantage of encoding gene control by modifying promoter architectures, as opposed to manipulating the regulatory proteins, is evolvability. Unlike altering the regulatory proteins, each promoter architecture controls the expression of a given subset of genes and hence can be programmed and reprogrammed with minimal pleiotropic effects. This may be of particular value when genes, possibly together with *cis*-acting elements, are horizontally-acquired, and their expression needs to be modified in order to integrate them functionally into the regulon.

MATERIALS AND METHODS

DNaseI footprinting assays

Primers A2, A4, B2, C2, D2, E2, F2, G2, H2, I2, I4, J2, K2, L2, M2 and N2 were labeled with 2 units of T4 polynucleotide kinase and 10 picomol of γ-^{32}P-ATP using 10 x polynucleotide kinase buffer in total volume of 25 μl at 37° C for 1 and 2 hrs and unincorporated $^{32}\beta$ was removed using ProbeQuant G-50 microcolumns (GE Healthcare) (Table S4). Then 14028s PCR fragments to be used as probes were generated by using chromosomal DNA as template for each pair of labeled and unlabeled primers A1 and A2, A3 and A4, B1 and B2, C1 and C2, D1 and D2, E1 and E2, F1 and F2, G1 and G2, H1 and H2, I1 and I2, I3 and I4, J1 and J2, K1 and K2, L1 and L2, M1 and M2 respectively. Probes generated from primers A3 and A4 and I3 and I4 were used with SlyA-H6 protein. DnaseI footprinting assays were carried out as described in [51] using different concentrations of PhoP-His6 for each gene and 0.02 units of DNaseI (Epicentre).

SI nuclease-protection assay

SI nuclease-protection assay was performed as described [52] with RNA harvested from early exponential (OD600, 0.3-0.4) phase cultures of wild-type (14028s) and *phoP* (MS8953s) *Salmonella* grown in 30 ml of N-minimal medium, pH 7.4, containing either

10 mM $MgCl_2$ or 10 μM $MgCl_2$. Total RNA was isolated using SV Total RNA Isolation System (Promega, Madison, WI) according to the manufacturer specifications. Chromosomal DNA was used as template. Probes used for *mig-14*, *pagK*, *yobG*, *pagP*, *ybjX*, *slyB*, *pdgL*, *pagD*, and *virK* were the same probes used for DNaseI footprinting. Probes for *yrbL*, *ompX*, *pipD* and *mgtC* were generated by labeling C2, J4, K3, M4 and using PCR as described above with labeled and unlabeled primers C3 and C2, J3 and J4, K1 and K3 and M3 and M4, respectively (Table S5).

Constructions of GFP reporter plasmids

The promoter region (i.e. intergenic region between two ORFs) was amplified using PCR. A list of the promoter-specific primers used in the PCR reactions is shown in (Table S6). The PCR fragment was digested with *Bam*HI and *Xho*I, purified, and then introduced to the cloning site of pMS201 (GFP reporter vector plasmid, a gift from Alon, U. [53]). pMS201 was transformed into EG13918 (Table S3). Sequences of promoter region were verified by nucleotide sequencing.

GFP assay: bacterial strains, plasmids and growth conditions

Bacterial strains and plasmids used in this study are listed in Table S3. *Salmonella enterica* serovar Typhimurium strains used in this study are derived from strain 14028s. Single colonies were used to inoculate 2 ml cultures and grown for 16 hours in N-

minimal media at pH 7.7 [54] with 10 mM Mg^{2+} and kanamycin (25 μg/ml) at 37°C with shaking at 250 rpm. The cultures were diluted 1:50 into 2 ml of the same medium as described above and grown an additional 4 hours at 37°C with shaking at 250 rpm. Then 1 ml of culture was removed and spun down at 13,000 rpm (centrifuge 5417, Eppendorf, Hamburg, Germany) for 2 minutes. The supernatant was carefully removed with a pipette. The bacterial pellet was washed with 1 ml of N-mimimal media with 2.5 mM Mg^{2+}. After spinning down the cells, the cell pellet was resuspended in 200 μl of N-minimal media with 2.5 mM Mg^{2+}. The concentrated cultures were diluted 1:50 into N-minimal media without Mg^{2+} or with 10 mM Mg^{2+}, at a final volume of 150 μl per well in flat-bottomed 96-well plates (PerkinElmer 6005225, Waltham, MA). The final Mg^{2+} concentrations in these two conditions were 50 μM and 10 mM, which induce and repress the PhoP/PhoQ system, respectively. The cultures were covered by a 50-μl layer of mineral oil (Sigma M-3516, St. Louis, MO) to prevent evaporation. The GFP assay was performed as described [55]. The cultures were grown in a Wallac Victor^3 multiwell fluorimeter (PerkinElmer) set at 37°C and assayed with an automatically repeating protocol of shaking (1 mm orbital, fast speed, 30 s), fluorescence readings (filters F485, F535, 0.5 s, CW lamp energy 10,000), and absorbance (OD) measurements (600 nm, P600 filter, 0.1 s). Time between repeated measurements was 6 min. Background fluorescence of cells bearing a promotorless GFP vector was

subtracted from the values obtained with the cells harboring plasmids with PhoP-activated promoters (or engineered derivatives) fused to a promoterless *gfp* gene.

Data preprocessing

The raw data corresponding to the GFP and OD signals were used to calculate the promoter activity as *[dGi(t)/dt]/ODi(t)]* [56]. The activity signal, ODs and background were smoothed by a shape-preserving interpolant fitting algorithm (Piecewise Cubic Hermite Interpolating Polynomial, Matlab 7.1) that finds values of an underlying interpolating function at intermediate points not described in the experimental assays. Then, we applied a polynomial fit (sixth order) on each expression signal. This smoothing procedure captured the dynamic well, while removing the noise inherent in the differentiation of noisy signals [56]. Observed ODs were standardized (linear regressions $R^2 > 0.99$) by using india ink from actual ODs measured by spectrophotometer. Observed fluorescence was standardized using dilutions of flurorescein and expressed as fluorecin concentrations.

Identifying the *cis*-regulatory features that determine distinct promoter architectures

We examined the promoters' DNA sequences for the presence of the *cis*-acting features as described in [19,57]. These features were encoded as Fuzzy Sets (or probabilistic distributions) because they are not crisp, but exist in a range of possible values [18].

Then, we grouped promoters sharing a common set of features using the GPS method as described [18].

ACKNOWLEDGMENTS

This work was supported by grant AI49561 from the National Institutes of Health to EAG, who is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Barnard A, Wolfe A, Busby S (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. Curr Opin Microbiol 7: 102-108.

2. Roy S, Semsey S, Liu M, Gussin GN, Adhya S (2004) GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant. J Mol Biol 344: 609-618.

3. Li M, McClure WR, Susskind MM (1997) Changing the mechanism of transcriptional activation by phage lambda repressor. Proc Natl Acad Sci U S A 94: 3691-3696.

4. Pedersen H, Sogaard-Andersen L, Holst B, Valentin-Hansen P (1991) Heterologous cooperativity in Escherichia coli. The CytR repressor both contacts DNA and the cAMP receptor protein when binding to the deoP2 promoter. J Biol Chem 266: 17804-17808.

5. Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. Nat Rev Microbiol 2: 57-65.

6. Dorman CJ (2007) H-NS, the genome sentinel. Nat Rev Microbiol 5: 157-161.

7. Liu M, Tolstorukov M, Zhurkin V, Garges S, Adhya S (2004) A mutant spacer sequence between -35 and -10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. Proc Natl Acad Sci U S A 101: 6911-6916.

8. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U (2006) Plasticity of the cis-regulatory input function of a gene. PLoS Biol 4: e45.

9. Groisman EA (1998) The ins and outs of virulence gene expression: Mg2+ as a regulatory signal. Bioessays 20: 96-101.

10. Groisman EA, Mouslim C (2006) Sensing by bacterial regulatory systems in host and non-host environments. Nat Rev Microbiol 4: 705-709.

11. Shin D, Groisman EA (2005) Signal-dependent binding of the response regulators PhoP and PmrA to their target promoters in vivo. J Biol Chem 280: 4089-4094.

12. Cromie MJ, Shi Y, Latifi T, Groisman EA (2006) An RNA sensor for intracellular Mg(2+). Cell 125: 71-84.

13. Shin D, Lee EJ, Huang H, Groisman EA (2006) A positive feedback loop promotes transcription surge that jump-starts Salmonella virulence circuit. Science 314: 1607-1609.

14. Groisman EA (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol 183: 1835-1842.

15. Bader MW, Sanowar S, Daley ME, Schneider AR, Cho U, et al. (2005) Recognition of antimicrobial peptides by a bacterial sensor kinase. Cell 122: 461-472.

16. Kato A, Tanabe H, Utsumi R (1999) Molecular characterization of the PhoP-PhoQ two-component system in Escherichia coli K-12: identification of extracellular Mg2+-responsive promoters. J Bacteriol 181: 5516-5520.

17. Bachhawat P, Stock AM (2007) Crystal structures of the receiver domain of the response regulator PhoP from Escherichia coli in the absence and presence of the phosphoryl analog beryllofluoride. J Bacteriol 189: 5987-5995.

18. Zwir I, Shin D, Kato A, Nishino K, Latifi T, et al. (2005) Dissecting the PhoP regulatory network of Escherichia coli and Salmonella enterica. Proc Natl Acad Sci U S A 102: 2862-2867.

19. Harari O, Park SY, Huang H, Groisman EA, Zwir I (2010) Defining the plasticity of transcription factor binding sites by Deconstructing DNA consensus sequences: the PhoP-binding sites among gamma/enterobacteria. PLoS Comput Biol 6: e1000862.

20. Blanco AG, Sola M, Gomis-Ruth FX, Coll M (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. Structure 10: 701-713.

21. Depardieu F, Courvalin P, Kolb A (2005) Binding sites of VanRB and sigma70 RNA polymerase in the vanB vancomycin resistance operon of Enterococcus faecium BM4524. Mol Microbiol 57: 550-564.

22. Soncini FC, Vescovi EG, Groisman EA (1995) Transcriptional autoregulation of the Salmonella typhimurium phoPQ operon. J Bacteriol 177: 4364-4371.

23. Lejona S, Aguirre A, Cabeza ML, Garcia Vescovi E, Soncini FC (2003) Molecular characterization of the Mg2+-responsive PhoP-PhoQ regulon in Salmonella enterica. J Bacteriol 185: 6287-6294.

24. Shi Y, Latifi T, Cromie MJ, Groisman EA (2004) Transcriptional control of the antimicrobial peptide resistance ugtL gene by the Salmonella PhoP and SlyA regulatory proteins. J Biol Chem 279: 38618-38625.

25. Perez JC, Latifi T, Groisman EA (2008) Overcoming H-NS-mediated transcriptional silencing of horizontally acquired genes by the PhoP and SlyA proteins in Salmonella enterica. J Biol Chem 283: 10773-10783.

26. Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, et al. (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. Cell 125: 1069-1082.

27. Gaston K, Bell A, Kolb A, Buc H, Busby S (1990) Stringent spacing requirements for transcription activation by CRP. Cell 62: 733-743.

28. Kadener S, Fededa JP, Rosbash M, Kornblihtt AR (2002) Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. Proc Natl Acad Sci U S A 99: 8185-8190.

29. Hunziker A, Tuboly C, Horvath P, Krishna S, Semsey S (2010) Genetic flexibility of regulatory networks. Proc Natl Acad Sci U S A 107: 12998-13003.

30. Holst B, Sogaard-Andersen L, Pedersen H, Valentin-Hansen P (1992) The cAMP-CRP/CytR nucleoprotein complex in Escherichia coli: two pairs of closely linked binding sites for the cAMP-CRP activator complex are involved in combinatorial regulation of the cdd promoter. Embo J 11: 3635-3643.

31. Valentin-Hansen P, Sogaard-Andersen L, Pedersen H (1996) A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. Mol Microbiol 20: 461-466.

32. Roy S, Garges S, Adhya S (1998) Activation and repression of transcription by differential contact: two sides of a coin. J Biol Chem 273: 14059-14062.

33. Perez JC, Groisman EA (2009) Transcription factor function and promoter architecture govern the evolution of bacterial regulons. Proc Natl Acad Sci U S A 106: 4319-4324.

34. Wade JT, Struhl K, Busby SJ, Grainger DC (2007) Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. Mol Microbiol 65: 21-26.

35. Hochschild A, Dove SL (1998) Protein-protein contacts that activate and repress prokaryotic transcription. Cell 92: 597-600.

36. Pedersen H, Valentin-Hansen P (1997) Protein-induced fit: the CRP activator protein changes sequence-specific DNA recognition by the CytR repressor, a highly flexible LacI member. Embo J 16: 2108-2118.

37. Rhodius VA, West DM, Webster CL, Busby SJ, Savery NJ (1997) Transcription activation at class II CRP-dependent promoters: the role of different activating regions. Nucleic Acids Res 25: 326-332.

38. Weber IT, Steitz TA (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 A resolution. J Mol Biol 198: 311-326.

39. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, et al. (2006) A bottom-up approach to gene regulation. Nature 439: 856-860.

40. Shin M, Kang S, Hyun SJ, Fujita N, Ishihama A, et al. (2001) Repression of deoP2 in Escherichia coli by CytR: conversion of a transcription activator into a repressor. Embo J 20: 5392-5399.

41. Browning DF, Beatty CM, Wolfe AJ, Cole JA, Busby SJ (2002) Independent regulation of the divergent Escherichia coli nrfA and acsP1 promoters by a nucleoprotein assembly at a shared regulatory region. Mol Microbiol 43: 687-701.

42. Browning DF, Cole JA, Busby SJ (2000) Suppression of FNR-dependent transcription activation at the Escherichia coli nir promoter by Fis, IHF and H-NS: modulation of transcription initiation by a complex nucleo-protein assembly. Mol Microbiol 37: 1258-1269.

43. Ansari AZ, Bradner JE, O'Halloran TV (1995) DNA-bend modulation in a repressor-to-activator switching mechanism. Nature 374: 371-375.

44. Ni L, Bruce C, Hart C, Leigh-Bell J, Gelperin D, et al. (2009) Dynamic and complex transcription factor binding during an inducible response in yeast. Genes Dev 23: 1351-1363.

45. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, et al. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science 313: 236-238.

46. Alm E, Huang K, Arkin A (2006) The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comput Biol 2: e143.

47. Janky R, van Helden J (2007) Discovery of conserved motifs in promoters of orthologous genes in prokaryotes. Methods Mol Biol 395: 293-308.

48. Perez JC, Shin D, Zwir I, Latifi T, Hadley TJ, et al. (2009) Evolution of a bacterial regulon controlling virulence and Mg(2+) homeostasis. PLoS Genet 5: e1000428.

49. Deng W, Burland V, Plunkett G, 3rd, Boutin A, Mayhew GF, et al. (2002) Genome sequence of Yersinia pestis KIM. J Bacteriol 184: 4601-4611.

50. Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A 100: 5136-5141.

51. Kato A, Latifi T, Groisman EA (2003) Closing the loop: the PmrA/PmrB two-component system negatively controls expression of its posttranscriptional activator PmrD. Proc Natl Acad Sci U S A 100: 4706-4711.

52. Kox LF, Wosten MM, Groisman EA (2000) A small protein that mediates the activation of a two-component system by another two-component system. Embo J 19: 1861-1872.

53. Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 100: 11980-11985.

54. Snavely MD, Gravina SA, Cheung TT, Miller CG, Maguire ME (1991) Magnesium transport in Salmonella typhimurium. Regulation of mgtA and mgtB expression. J Biol Chem 266: 824-829.

55. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, et al. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. Science 292: 2080-2083.

56. Ronen M, Rosenberg R, Shraiman BI, Alon U (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. Proc Natl Acad Sci U S A 99: 10555-10560.

57. Harari O, del Val C, Romero-Zaliz R, Shin D, Huang H, et al. (2009) Identifying promoter features of co-regulated genes with similar network motifs. BMC Bioinformatics 10 Suppl 4: S1.

FIGURE LEGENDS

Figure 1. Gene transcription of PhoP-activated promoters is determined by the location, orientation and phasing of the PhoP box. Promoter activity measured by $[dG_i(t)/dt]/OD_i(t)]$ where $G_i(t)$ is GFP fluorescence and $OD_i(t)$ is the optical density. Preprocessing was performed as described in Materials and Methods. Each construct is a shortened derivative of the corresponding PhoP-activated promoters that includes the most upstream PhoP box and the RNAP binding site (Fig. S3). (**A**) The *phoP* (blue line) promoter compared with its shortened derivative *phoP4* (green line). The *phoP4* (green line) and *rstA4* (cyan line) promoters compared with their derivatives *phoP4*/*rev* (red line) and *rstA4*/*rev* (orange line), respectively, where the PhoP box orientation was reversed in their original locations. (**B**) The *phoP4*/*revK* (blue line) and the *phoP4*/*revC* (green line) promoters, where the orientation of the PhoP box was reversed and moved -37 bp and -47 bp from the RNAP -10 hexamer, respectively (i.e., the position of the PhoP box in the natural *pagK* and *mgtC* promoters). (**C**) The *phoP4+revK'* promoter (blue line) compared with its derivative (red line), where the PhoP box was reversed to the direct orientation. The experiments were recreated under high Mg^{2+} concentrations, which is the condition where the PhoP protein is noninduced (green and cyan lines). (**D**) The PhoP box moved upstream/downstream from its normal location and orientation in the *phoP4* promoter at 7, 12, 17, 21, 22, 23 24, 25, 26, 27, 28, 33, 36, and 48

bp from the RNAP -10 hexamer. (**E**) The oppositely oriented PhoP box in the *phoP4* promoter was moved upstream/downstream at 12, 27, 28, 29, 30, 31, 32, 36, 37, 47, and 48 bp from the RNAP -10 region.

Figure 2. Regulatory role of the PhoP boxes in different promoters *in vivo*. Promoter activity was measured as in Fig. 1. Mutageneses performed in each promoter construction are detailed in Fig. S3. The PhoP box sequence, location and orientation in each promoter region is detailed in Fig. S2. is described in Fig. S3. All plasmid constructs were evaluated on strain EG13918. (**A**) The *ybjX4* promoter (blue line) compared with its derivatives *ybjX4 /mut_dn* (red line) and *ybjX4 /mut_up* (green line), where two PhoP boxes are individually mutated (green cylinders). (**B**) The *virK4* promoter (blue line) compared with its derivatives *virK4 /mut_dn* (red line) and *virK4/mut_up* (green line), where two PhoP boxes are individually mutated (green cylinders). (**C**) The *mig-14_4* (blue line) compared with its derivatives *mig-14_4 /mut_up* (red line) and *mig-14_4/mut_dn* (green line), where two PhoP boxes are individually mutated. (**D**) The *phoP4* promoter (blue line) compared with its derivative *phoP4/2* (green line), where a second PhoP box upstream of the single PhoP box was added. (**E**) The *mgtC* promoter (green line) compared with the shortened derivative *mgtC4* (blue line). (**F**) The *pcgL4* (blue line) compared with its derivatives *pcgL4/mut_dn* (red line)

and *pcgL4/mut_up* (green line), where two PhoP boxes are individually mutated (green cylinders).

Figure 3. The RNAP binding site in the PhoP regulated promoters. Promoter activity was measured as in Fig. 1, and the corresponding constructs detailed in Fig. S3. Promoters with PhoP boxes located upstream of the RNAP -35 region were used to prevent overlapping. (**A**) The *phoP4+11* promoter (blue line) compared with its derivative (green line), where the canonical -10 region of *phoP4* was replaced by a weaker consensus sequence from the PhoP-activated *rstA* promoter. (**B**) The *phoP4+revC* promoter (blue line) compared with its derivative (green line), where the canonical -10 region of *phoP4* was replaced by a stronger -10 consensus sequence from the PhoP-activated *mgtC* promoter. (**C**) The *phoP4+revK* promoter (blue line) compared with its derivative (green line), where the canonical -35 region of *phoP4* was replaced by a stronger -35 consensus sequence. (**D**) The *phoP4+revC* promoter (blue line), where the -10 and -35 RNAP regions are those of the PhoP-activated *phoP* promoter, is compared to a variant of the *mgtC4* promoter (green line), where the -10 and -35 RNAP regions are those of the PhoP-activated *mgtC* promoter, but the PhoP box is that of the *phop4* promoter. (**E**) The experiments performed in (**D**) were recreated under high Mg2+ concentrations, which is the condition where the PhoP protein is noninduced.

Figure 4. Promoter architectures of PhoP-activated genes. Architectures of 20 PhoP-activated promoters where the following c*is*-regulatory features are highlighted: -10 and -35 regions (red), PhoP box(es) (blue), PhoP box orientation (green line), role of the PhoP box (+: activation, -: repression), distance and phasing (top panel in scale). From the top: Architecture I; Architecture II; Architecture III; Architecture IV; and Architecture V.

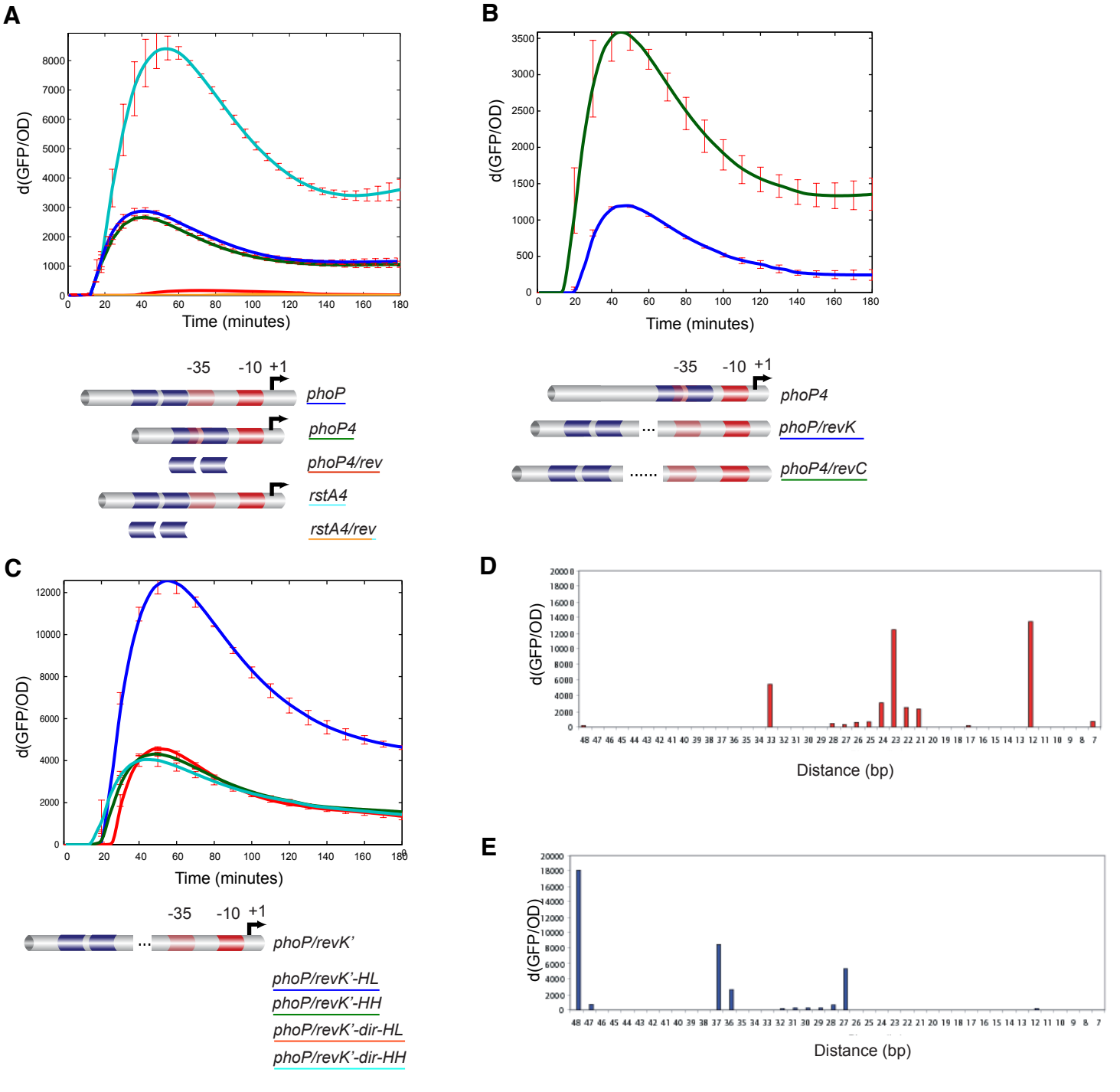
A
d(GFP/OD)
Time (minutes)
-35
-10
+1
phoP
phoP4
phoP4/rev
rstA4
rstA4/rev
B
d(GFP/OD)
Time (minutes)
-35
-10
+1
phoP4
phoP/revK
phoP4/revC
C
d(GFP/OD)
Time (minutes)
-35
-10
+1
phoP/revK'
phoP/revK'-HL
phoP/revK'-HH
phoP/revK'-dir-HL
phoP/revK'-dir-HH
D
d(GFP/OD)
Distance (bp)
E
d(GFP/OD)
Distance (bp)

Figure 1

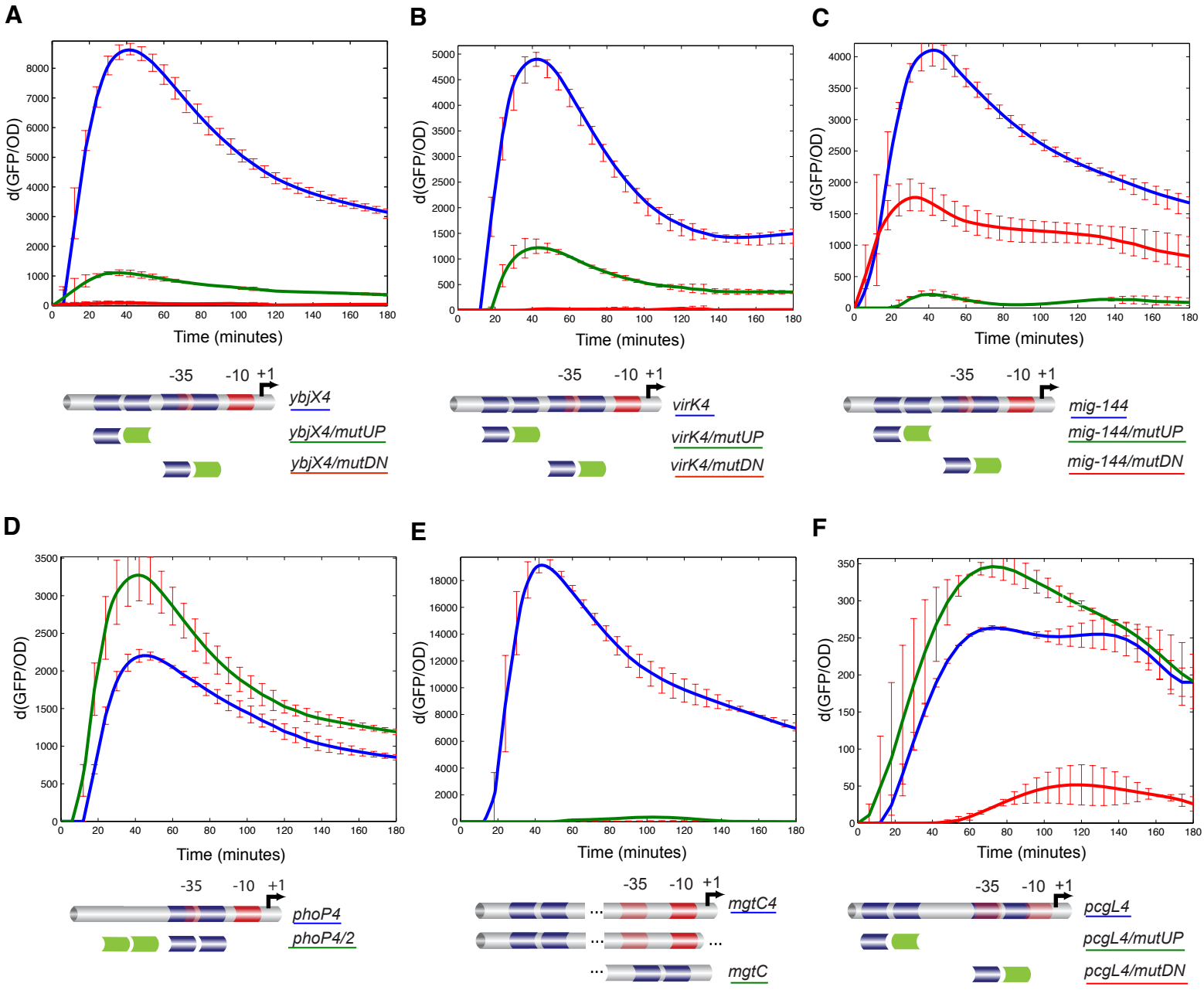
A
B
C
D
E
F
d(GFP/OD)
Time (minutes)
-35
-10
+1
ybjX4
ybjX4/mutUP
ybjX4/mutDN
virK4
virK4/mutUP
virK4/mutDN
mig-144
mig-144/mutUP
mig-144/mutDN
phoP4
phoP4/2
mgtC4
mgtC
pcgL4
pcgL4/mutUP
pcgL4/mutDN

Figure 2

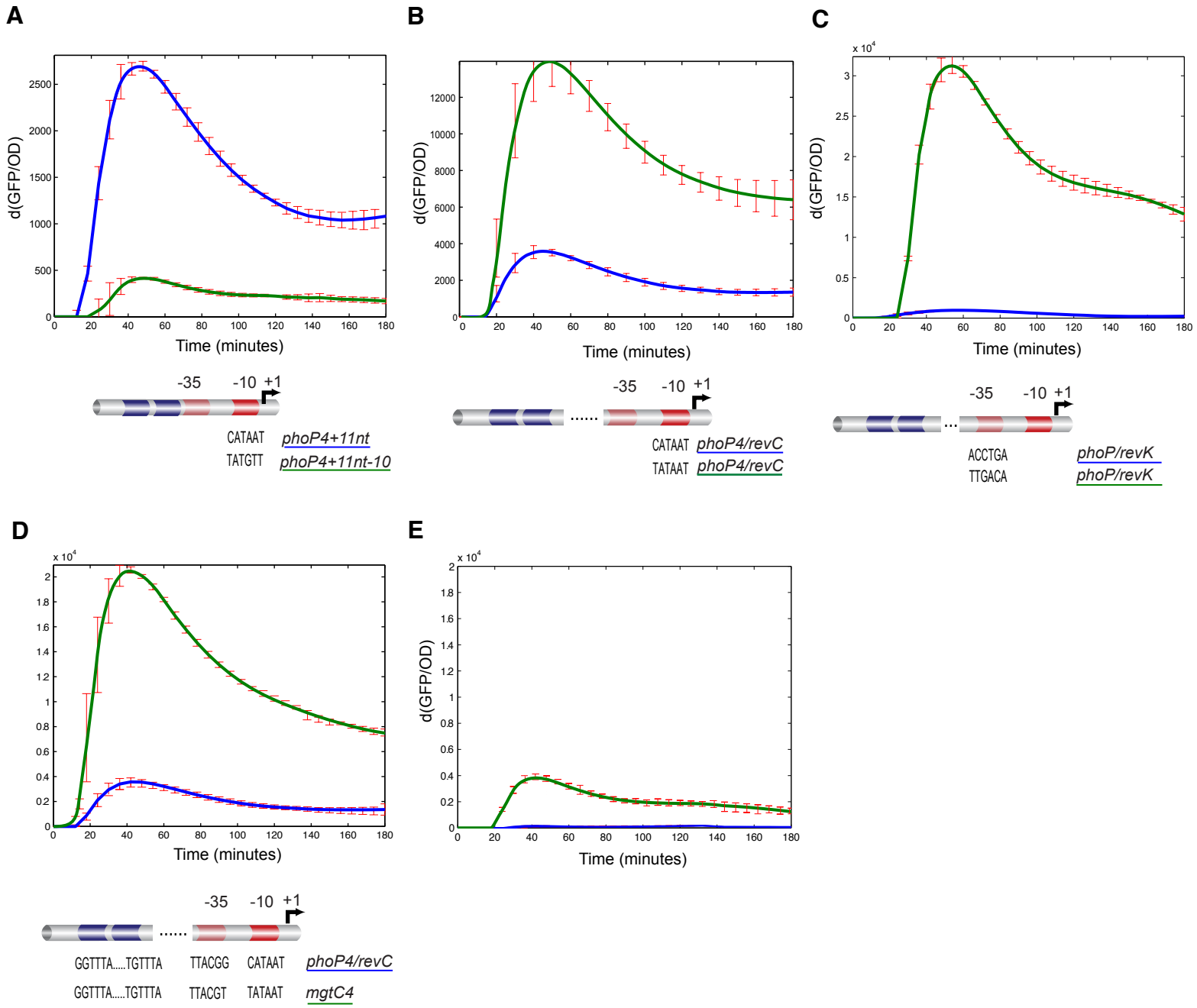
A
d(GFP/OD)
Time (minutes)
-35
-10
+1
CATAAT phoP4+11nt
TATGTT phoP4+11nt-10
B
d(GFP/OD)
Time (minutes)
-35
-10
+1
CATAAT phoP4/revC
TATAAT phoP4/revC
C
x 10^4
d(GFP/OD)
Time (minutes)
-35
-10
+1
ACCTGA phoP/revK
TTGACA phoP/revK
D
x 10^4
Time (minutes)
-35
-10
+1
GGTTTA.....TGTTTA TTACGG CATAAT phoP4/revC
GGTTTA.....TGTTTA TTACGT TATAAT mgtC4
E
x 10^4
d(GFP/OD)
Time (minutes)

Figure 3

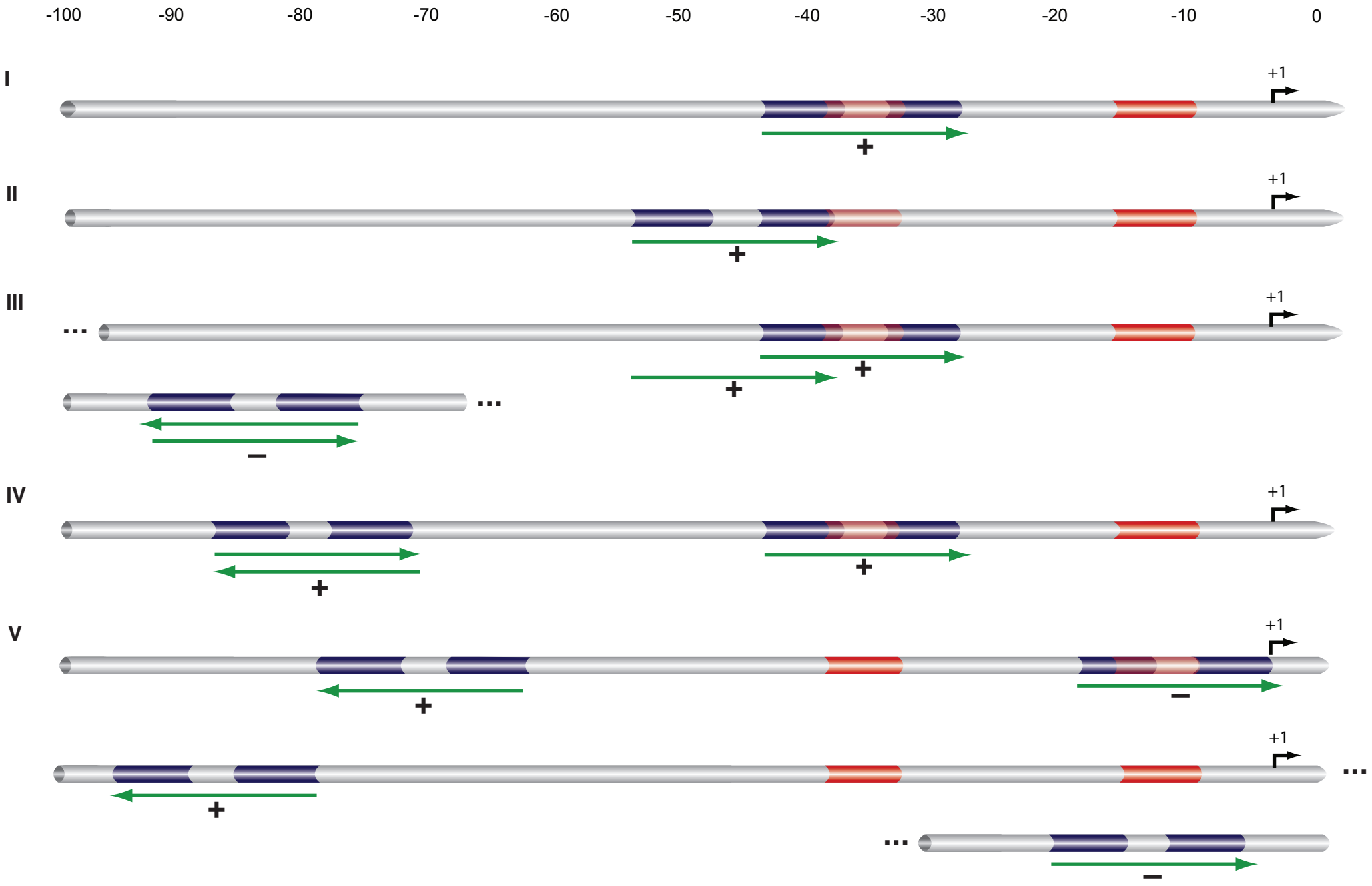

Figure 4