



# A Degenerate Tripartite DNA-Binding Site Required for Activation of ComA-Dependent Quorum Response Gene Expression in *Bacillus subtilis*

Kevin L. Griffith and Alan D. Grossman\*

Department of Biology,  
Massachusetts Institute of  
Technology, Cambridge,  
MA 02139, USA

Received 15 April 2008;  
received in revised form  
9 June 2008;  
accepted 11 June 2008  
Available online  
18 June 2008

In *Bacillus subtilis*, the transcription factor ComA activates several biological processes in response to increasing population density. Extracellular peptide signaling is used to coordinate the activity of ComA with population density. At low culture densities, when the concentration of signaling peptides is lowest, ComA is largely inactive. At higher densities, when the concentration of signaling peptides is higher, ComA is active and activates the transcription of at least nine operons involved in the development of competence and in the production of degradative enzymes and antibiotics. We found that ComA binds a degenerate tripartite sequence consisting of three DNA-binding determinants or “recognition elements.” Mutational analyses showed that all three recognition elements are required for transcription activation *in vivo* and for specific DNA binding by ComA *in vitro*. Degeneracy of the recognition elements in the ComA-binding site is an important regulatory feature for coordinating transcription with population density (i.e., promoters containing an optimized binding site have high activity at low culture density and are no longer regulated in the normal-density-dependent manner). We found that purified ComA forms a dimer in solution, and we propose a model for how two dimers of ComA bind to an odd number of DNA-binding determinants to activate transcription of target genes. This DNA–protein architecture for transcription activation appears to be conserved for ComA homologs in other *Bacillus* species.

© 2008 Elsevier Ltd. All rights reserved.

**Keywords:** quorum sensing; transcription; response regulator; DNA binding; *Bacillus subtilis*

Edited by J. Karn

## Introduction

Bacteria often coordinate physiological processes using quorum or diffusion sensing.<sup>1–3</sup> Cells communicate with one another using small diffusible signaling molecules that are secreted into the environment and sensed by neighboring cells. In Gram-negative bacteria, signaling molecules are typically acylated homoserine lactone derivatives. In contrast, signaling molecules used by Gram-positive bacteria are typically peptides (see reviews by Waters and Bassler<sup>4</sup> and Auchtung and Grossman<sup>5</sup>). Responding to population density enables bacteria to coordinate responses when sufficient numbers of cells are present.

In *Bacillus subtilis*, quorum sensing contributes to a variety of physiological processes, including the development of genetic competence, the decision to sporulate, and the production of degradative enzymes and antibiotics.<sup>5–8</sup> The ComX–ComP–ComA signaling pathway controls the quorum response in *B. subtilis*. ComX pheromone is a farnesylated 10-amino-acid peptide that is secreted into the growth medium and accumulates extracellularly as culture density increases.<sup>9–12</sup> ComX binds to its cognate receptor kinase ComP, resulting in autophosphorylation of ComP at a conserved histidine residue.<sup>13</sup> As with other two-component systems, phosphorylated ComP donates phosphate to its cognate response regulator, ComA, on a conserved aspartate. ComA is present and appears to be expressed continuously during exponential growth (unpublished observations) and, once phosphorylated, ComA~P functions to activate transcription of target genes.<sup>6–8,14,15</sup>

\*Corresponding author. E-mail address: adg@mit.edu.

Abbreviations used: RE1, recognition element 1; RE2, recognition element 2; RE3, recognition element 3; PCR, polymerase chain reaction; BSA, bovine serum albumin.

At least four other density-dependent signaling pathways influence the activity of ComA. PhrC (also known as the competence- and sporulation-stimulating factor), PhrF, PhrH, and PhrK are pentapeptides that are secreted into the growth medium. The pentapeptides are transported back into the cell through the oligopeptide permease Opp (aka Spo0K), where they bind to and inhibit the activity of their cognate Rap proteins RapC, RapF, RapH, and RapK. RapC, RapF, RapH, and, presumably, RapK inhibit ComA binding to its target sites.<sup>16–21</sup> Thus, the activity of ComA is highly regulated, resulting in little or no ComA-dependent activation of target genes at low culture densities when the concentration of signaling peptides is low. At higher culture densities, when the concentration of signaling peptides increases, ComA becomes activated and stimulates transcription of target genes (reviewed by Lazazzera<sup>22</sup> and Pottathil and Lazazzera<sup>23</sup>).

The combined work from several groups led to the identification of 20 genes in nine operons whose expression appears to be directly regulated by ComA.<sup>14,16,24–29</sup> Sequence alignments, mutational analyses, and DNA footprinting studies led to the model that the binding site for ComA is an inverted repeat containing two 6-bp recognition elements separated by a 4-bp spacer.<sup>14,26,28,29</sup> Several genes regulated by ComA have a single inverted repeat, whereas others have multiple inverted repeats.

We describe experiments indicating that the ComA-binding site contains three distinct sequence recognition elements. For simplicity, we refer to these as recognition element 1 (RE1), recognition element 2 (RE2), and recognition element 3 (RE3) throughout. RE1 and RE2 comprise the inverted repeat previously characterized as part of the ComA-binding site. RE3 is a previously unrecognized sequence downstream of RE1 and RE2 with a consensus sequence identical with that of RE1. We analyzed the relative contributions of all three sequence elements (RE1, RE2, and RE3) and the spacing between the three elements in transcription activation by ComA. Based on our results, we conclude that: (1) all three recognition elements are critical for activation by ComA *in vivo* and for DNA binding by ComA *in vitro*; (2) the spacing between recognition elements is important for transcription activation by ComA; (3) there is some sequence-dependent information in the spacer regions; and (4) the overall sequence context and the degeneracy of the binding site are critical for the population-density-dependent regulation of genes controlled by ComA.

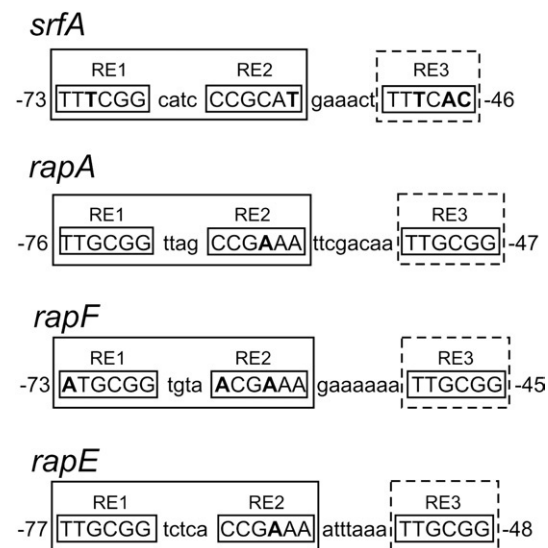
## Results

### A third potential recognition element in ComA-binding sites

The previously proposed consensus ComA-binding site (5'-TTGCGGnnnnnCCGCAA) is an inverted repeat comprising two 6-bp half-sites (consensus 5'-TTGCGG) separated by a 4-bp spacer.<sup>14,26,28,29</sup> For

characterized promoter regions, the ComA-binding site is located upstream of the –35 recognition element for RNA polymerase. We refer to the promoter distal half-site of the inverted repeat (5'-TTGCGG) as “RE1,” and we refer to the promoter-proximal half-site of the inverted repeat (5'-CCGCAA) as “RE2” (Fig. 1).

During the course of analyzing ComA function, DNA binding, and *in vivo* target genes, we noticed that all nine targets known or thought to be directly activated by ComA contained a conserved sequence with a consensus identical with RE1 (consensus 5'-TTGCGG) located downstream of the inverted repeat (Fig. 1; data not shown). We refer to this sequence element as “RE3.” There is no recognizable fourth recognition element upstream of RE1 or downstream of RE3 that would be expected if two complete inverted repeats were required for transcription activation by ComA. This arrangement of three putative DNA-binding determinants, consisting of an inverted repeat and a half-site, is found for some LysR-type transcriptional regulators,<sup>30,31</sup> but otherwise seems to be unusual for transcriptional



**Fig. 1.** Alignment of ComA–DNA binding sites used in this study. The promoter-proximal ComA-binding sites of target genes used in this study are shown. The numbering represents the position of the binding site relative to the start of transcription, as determined by primer extension for each gene (unpublished results). *rapA* and *rapF* have two transcription start sites separated by 2 bp. The most abundant (upstream) transcript was used in each case to determine the position of the ComA-binding site. Three DNA sequence determinants, referred to as RE1–RE3, make up a single ComA-binding site. *srfA*, *rapE*, and *rapA* have two ComA-binding sites, and only the promoter-proximal site is shown. RE1 and RE2 form a palindrome comprising two half-site sequences (consensus 5'-TTGCGG) separated by a 4-bp spacer, with the exception being *rapE*, which has a 5-bp spacer. The inverted repeat formed by RE1 and RE2 is depicted as a rectangle with solid lines, while the newly identified RE3 is depicted as a dashed rectangle. Nucleotides shown in bold represent mismatches from the consensus sequence.

activators. Results described below demonstrate that a single ComA-binding site includes all three recognition elements.

### RE3 can function in *srfA* activation

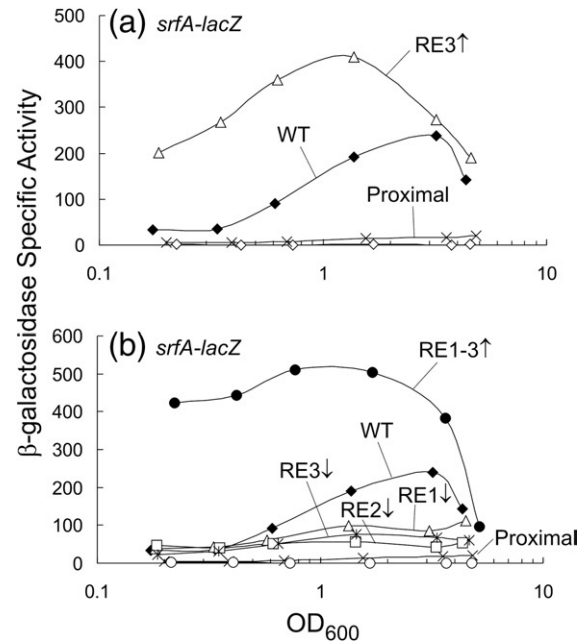
*srfA* is one of the most widely characterized target operons of ComA due to its involvement in the production of the antimicrobial agent surfactin and in the development of genetic competence (i.e., the ability to take up DNA).<sup>32–35</sup> The *srfA* promoter has two RE1–RE2 inverted repeats separated by 28 bp. ComA binds to both inverted repeats in the *srfA*-regulatory region.<sup>14,26</sup> Deletion of the promoter-distal inverted repeat greatly reduces transcription from the *srfA* promoter.<sup>26</sup> However, transcription is restored by compensatory mutations in the promoter-proximal inverted repeat that make it closer to consensus.<sup>26</sup>

A third sequence element, RE3, is present downstream of each of the previously characterized inverted repeats (RE1+RE2) in the *srfA*-regulatory region. We characterized the downstream ComA-binding site (including the inverted repeat and RE3) and found that RE3 was important for transcription. As observed previously,<sup>26</sup> removal of the upstream inverted repeat reduced transcription of a *srfA*–*lacZ* transcriptional fusion by ~8-fold (Fig. 2a). However, three substitutions in the promoter-proximal RE3 (5'-TTTCAC to 5'-TTGCGG), allowing it to match the consensus sequence, compensated for the lack of the upstream inverted repeat (Fig. 2a). Furthermore, expression was greater than that of the complete wild-type promoter containing both inverted repeats (Fig. 2a). This increased expression was dependent on ComA (data not shown) and was most obvious at low culture density, where there was a >5-fold increase in  $\beta$ -galactosidase-specific activity relative to that of wild type (Fig. 2a). From these results, we conclude that RE3 can function to promote transcriptional activation of *srfA*, and this activation depends on ComA.

### Analysis of a mutant *srfA* promoter containing consensus sequences in all three recognition elements

To determine the relative contribution of all three recognition elements to transcription activation, we modified the *srfA*–*lacZ* promoter fusion containing only the promoter-proximal ComA-binding site such that all three recognition elements matched the consensus sequence (5'-TTGCGG for RE1 and RE3, and the reverse complement 5'-CCGCAA for RE2). We also made constructs in which each of the consensus recognition elements was individually replaced with 5'-GCATAT (containing changes away from consensus at every position) and measured the effects on expression.

Expression of the *srfA*–*lacZ* fusion, with all three recognition elements matching the consensus sequence for ComA binding, was quite high (~11-fold greater than wild type), especially at low culture densities (Fig. 2b). Replacement of any of the con-



**Fig. 2.** Role of RE3 in transcription activation of *srfA*. Cultures containing *PsrfA*–*lacZ* fusions were grown in defined minimal medium, and samples were removed throughout growth for determination of  $\beta$ -galactosidase-specific activity.  $\beta$ -Galactosidase-specific activity is plotted as a function of cell density ( $OD_{600}$ ). Mutation to a consensus recognition element (5'-TTGCGG for RE1 and RE3; 5'-CCGCAA for RE2) is depicted as an up arrow, while the down arrow represents mismatches from consensus in all six positions of a single recognition element (5'-GCATAT for RE1 and RE3; 5'-ATATGC for RE2). (a) KG125 wild type (filled diamonds); KG102 promoter-proximal binding site only (X); KG160 promoter-proximal binding site with RE3 consensus (open triangles); KG150 *comA*-null mutant with the wild-type reporter (open diamonds). (b) KG125 wild-type (filled diamonds) and KG102 promoter-proximal site only (X) are the same as in (a) for comparison; KG158 promoter-proximal site only with RE1–RE3 consensus (filled circles); KG780 promoter-proximal site only with consensus RE2 and RE3 and nonconsensus RE1 (5'-GCATAT) (open triangles); KG567 promoter-proximal site only with consensus RE1 and RE3 and nonconsensus RE2 (5'-ATATGC) (open squares); KG565 promoter-proximal site only with consensus RE1 and RE2 and nonconsensus RE3 (5'-GCATAT) (asterisk); and KG464 promoter-proximal site only with consensus RE1–RE3 in  $\Delta comA$  background (open circles).

sensus recognition elements with 5'-GCATAT caused a significant decrease in transcription, and the magnitudes of the decrease were similar for mutations in each element (Fig. 2b). These results indicate that, in the context of three consensus recognition elements, each one is equally important for transcription activation of the *srfA* promoter.

### All three recognition elements contribute to DNA binding by ComA *in vitro*

To determine whether RE3 is important for ComA to bind DNA, we measured the binding of purified ComA (with a his<sub>6</sub> tag on the amino terminus) to





RE2–RE3 spacer, indicating that conformational changes occur on the DNA when bound by ComA.<sup>14,36</sup>

We are not able to directly compare the concentrations of ComA needed for binding DNA *in vitro* to those needed for transcription activation *in vivo* because the fraction of purified his<sub>6</sub>-ComA that is active *in vitro* is not known. In addition, the DNA templates used for *in vitro* binding assays are short linear fragments, whereas the templates *in vivo* are contained in supercoiled chromosomal DNA coated with many DNA-binding proteins.

### Relative contribution of RE1, RE2, and RE3 to transcription activation of *rapA*

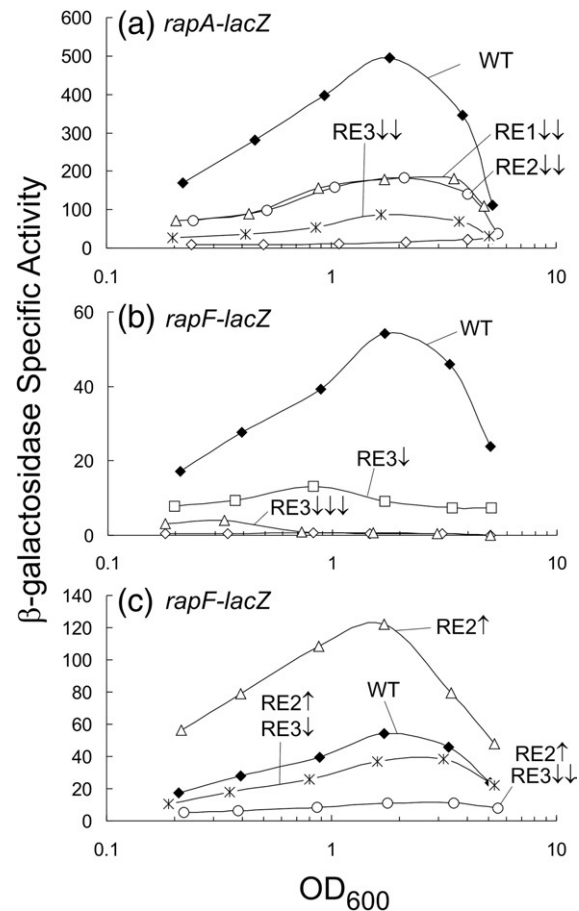
*rapA* has two potential ComA-binding sites,<sup>29</sup> and the promoter-proximal site (Fig. 1) is close to consensus in all three recognition elements, containing only a single change in RE2 (5'-CCGAAA) away from consensus (5'-CCGCAA). In contrast to *srfA*, which requires both the proximal and the distal ComA-binding sites (Fig. 2 and Nakano and Zuber<sup>26</sup>), the promoter-proximal ComA-binding site from *rapA* was sufficient for regulated expression of a *rapA-lacZ* transcriptional fusion (Fig. 4a; data not shown). Transcription of *rapA-lacZ* was relatively high at low culture density (Fig. 4a). There was a small but reproducible 2- to 2.5-fold increase in  $\beta$ -galactosidase-specific activity as the culture density increased (Fig. 4a). In contrast, there was an ~10-fold increase in  $\beta$ -galactosidase-specific activity in cultures containing a *srfA-lacZ* fusion (Fig. 2).

Like *srfA*, we found that each of the recognition elements contributes to transcription of *rapA-lacZ*. Substitutions were made on each recognition element so that two positions varied away from consensus (5'-TTTCTGA). Mutations in RE1 and RE2 reduced  $\beta$ -galactosidase-specific activity by ~2- to 2.5-fold (Fig. 4a). Mutations in RE3 had a larger effect, reducing  $\beta$ -galactosidase-specific activity by ~5-fold compared to wild type (Fig. 4a). Based on these results, we conclude that all three recognition elements are required for optimal expression of *rapA* and that, in the context of this promoter, RE3 is most important.

### RE3 is required for transcriptional activation of *rapF*

*rapF* has a single ComA-binding site in its regulatory region (Fig. 1). Transcription of *rapF-lacZ* increased as the culture density increased and maximal  $\beta$ -galactosidase-specific activity occurred near the end of exponential growth (Fig. 4b). Like the other target genes tested, transcription of *rapF* was dependent on ComA, as very little  $\beta$ -galactosidase activity was observed in a *comA*-null mutant (Fig. 4b).

We found that RE3 is required for transcription activation of *rapF-lacZ*. A single G-to-T mutation at position 3 of RE3 (5'-TTGCGG to 5'-TTTCGG) caused a significant decrease in  $\beta$ -galactosidase-specific activity (Fig. 4b). A triple mutation in RE3 (5'-TTGCGG to 5'-GTGTCTG) away from consensus re-



**Fig. 4.** Roles of recognition elements in transcription activation of *rapA* and *rapF*. Cultures containing *PrapA-lacZ* (a) or *PrapF-lacZ* (b and c) fusions were grown in defined minimal medium, and aliquots were taken throughout growth for determination of  $\beta$ -galactosidase-specific activity. Each arrow represents a single base substitution: up arrows indicate substitutions toward consensus, and down arrows indicate substitutions away from consensus. Underlined nucleotides represent mismatches from the consensus sequence. (a) KG112 wild-type *PrapA-lacZ* (filled diamonds); KG544 RE1 5'-TTTCTGA and RE2 consensus (triangles); KG513 RE2 5'-TCGAAA (circles); KG545 RE3 5'-TTTCTGA and RE2 consensus (asterisk); KG148 wild-type reporter  $\Delta comA$  (open diamonds). (b) KG277 wild-type *PrapF-lacZ* (filled diamonds); KG556 one mismatch in RE3 from consensus (5'-TTTCTGG) (squares); KG266 three mismatches in RE3 from consensus (5'-GTGTCTG) (triangles); KG239 wild-type reporter  $\Delta comA$  (open diamonds). (c) KG277 wild-type *PrapF-lacZ* (filled diamonds) is the same as in (b); KG555 one mismatch in RE2 toward consensus (5'-CCGAAA) (triangles); KG566 one mismatch in RE2 toward consensus (5'-CCGAAA) and one mismatch in RE3 from consensus (5'-TTTCTGG) (asterisk); KG557 one mismatch in RE2 toward consensus (5'-CCGAAA) and two mismatches in RE3 from consensus (5'-TTTCTGT) (circles).

duced the  $\beta$ -galactosidase-specific activity of *rapF-lacZ* to levels similar to those in a *comA*-null mutant (Fig. 4b). The decrease in transcription correlates with the severity of mismatches away from consensus and indicates that RE3 is required for transcriptional activation of *rapF*.

### Compensatory substitutions within the ComA-binding site restore transcription activation of *rapF*

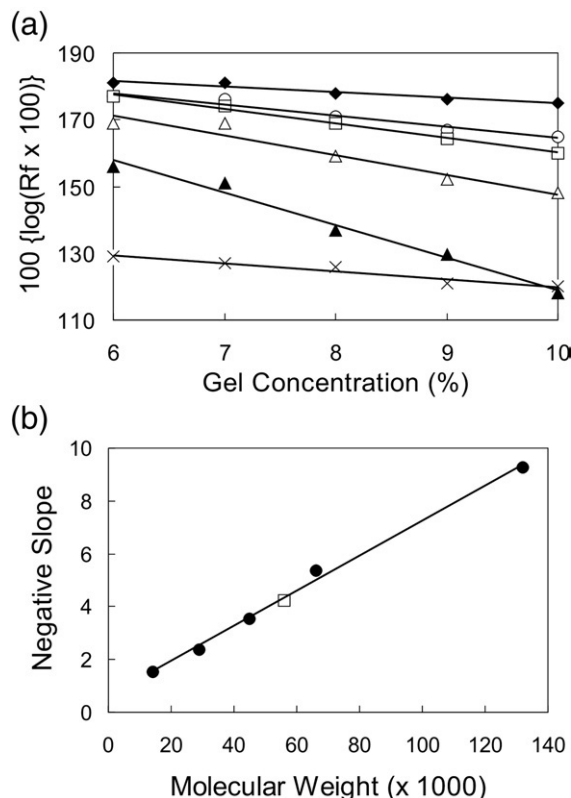
We found that substitutions in RE2 toward consensus could compensate for mutations in RE3 away from consensus in the *rapF*-regulatory region. The single mutation away from consensus in RE3 (5'-TTGCGG to 5'-TTTCGG), described above, was significantly suppressed by an A-to-C substitution toward consensus in position 1 of RE2 (5'-ACGAAA to 5'-CCGAAA), as expression of *rapF-lacZ* was restored to near-wild-type levels (Fig. 4c). An additional mutation in RE3 (5'-TTGCGG to 5'-TTTCGT) away from consensus (again in the context of the A1C mutation in RE2) further decreased *rapF-lacZ* expression to ~3- to 5-fold below that of wild type (Fig. 4c). In the context of the wild-type RE3, the RE2 A1C mutation caused an ~2- to 3-fold increase in expression throughout the growth cycle, as compared to wild type (Fig. 4c).

In combination, our results indicate that substitutions in one recognition element toward the consensus sequence can compensate for substitutions in the other element(s) away from consensus. The different nucleotide combinations within the three binding determinants influence transcription activation *in vivo* by affecting expression at low culture density and the amount of induction that occurs when cultures are grown to high density.

### Determination of the oligomeric state of ComA in solution

We found that ComA interacts with itself in a yeast two-hybrid assay (data not shown). Furthermore, when native ComA and his<sub>6</sub>-ComA were overexpressed together in *Escherichia coli* from compatible plasmids, the two forms of ComA appeared to stably associate with each other. Purification of his<sub>6</sub>-ComA by Ni-affinity chromatography resulted in the recovery of both his<sub>6</sub>-ComA and untagged ComA (data not shown). These results indicate that ComA is able to interact with itself and that it probably does not function as a monomer.

ComA is thought to function as a dimer based on its binding to an inverted repeat.<sup>14,26</sup> Based on our findings that three recognition elements are required for ComA-mediated transcriptional activation *in vivo* and contribute to DNA binding *in vitro*, we sought to determine the oligomeric state of ComA. We used a method based on mobility in gels with different polyacrylamide concentrations.<sup>37</sup> Briefly, purified his<sub>6</sub>-ComA and a set of protein standards of predetermined molecular masses were subjected to native gel electrophoresis in different concentrations of polyacrylamide. The migration distance of each protein was plotted against the acrylamide concentration, with the slope of the line representing the retardation coefficient (Fig. 5a). A standard curve was generated from the retardation coefficients of the protein standards, and the molecular mass of ComA was interpolated from the graph (Fig. 5b).



**Fig. 5.** Determination of the oligomeric state of ComA. Purified his<sub>6</sub>-ComA was separated by native gel electrophoresis in different concentrations of polyacrylamide along with protein standards of known molecular mass. The migration distance ( $R_f$ ) was determined by measuring the distance that each protein traveled in the gel and by dividing this value by the distance traveled by the bromophenol blue dye. (a) Representative experiment showing the migration distances of purified his<sub>6</sub>-ComA and protein standards in gels with different polyacrylamide concentrations.  $\alpha$ -Lactalbumin (filled diamonds); chicken egg albumin (circles); his<sub>6</sub>-ComA (squares); BSA monomer (open triangles); BSA dimer (filled triangles); and carbonic anhydrase from bovine erythrocytes (X). (b) Representative standard curve of the slope of the migration distances of each protein standard. The negative slope of each protein standard was determined from the experiment in (a) and plotted as a function of known molecular mass for that particular protein:  $\alpha$ -lactalbumin (14.2 kDa); carbonic anhydrase (29 kDa); egg albumin (45 kDa); BSA monomer (66 kDa); and BSA dimer (132 kDa). Protein standards (filled circles) and his<sub>6</sub>-ComA (square). The molecular mass of his<sub>6</sub>-ComA was 53 kDa based on the average of three independent experiments. The theoretical molecular mass of his<sub>6</sub>-ComA was 25 kDa. We conclude that his<sub>6</sub>-ComA functions as a dimer in solution.

Based on the average of three individual experiments, the molecular mass of his<sub>6</sub>-ComA appeared to be 53 kDa. Since the theoretical molecular mass of his<sub>6</sub>-ComA is 25 kDa, it appears that ComA functions as a dimer in solution.

The question remains: How does ComA bind to the tripartite ComA-binding site to activate transcription? We postulate that two dimers of ComA

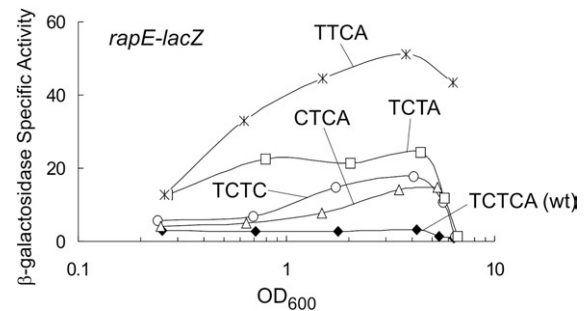


occupy a single binding site consisting of RE1, RE2, and RE3. Two possible models for the binding configuration seem most plausible, and results described below support the first model. (1) One dimer of ComA binds RE1 and RE2, and a second dimer binds RE3 and, perhaps, nonspecific sequences downstream of RE3. Alternatively, (2) a dimer of ComA could bind RE2 and RE3 (themselves an inverted repeat), and a second dimer could bind RE1 and, perhaps, nonspecific sequences upstream of RE1. It is also possible that the fourth ComA subunit in these putative complexes is not bound to DNA. In either model, it seems most likely that the spacing between the two recognition elements that are bound by a single ComA dimer would be most severely restricted in length, while the spacer separating two dimers of ComA might be more accommodating to alterations in length. We tested the effects of altering the lengths of the spacer between RE1 and RE2 and also between RE2 and RE3.

### Spacer length separating RE1 and RE2 is important for transcription activation

The 4-bp spacer length separating RE1 and RE2 is conserved among the known ComA-binding sites, except for *rapE*, which has a 5-bp spacer (see below). We tested the importance of the spacer distance in the context of *rapF* and found that any deviation from 4 bp severely disrupted transcription. We created insertions and deletions of varying lengths (−1 bp, −2 bp, −3 bp, +1 bp, +2 bp, and +3 bp) between RE1 and RE2 of the *rapF*-regulatory region. We also introduced half-helical and full-helical turns of DNA (i.e., 5, 6, 10, or 11 consecutive adenine nucleotides) within the spacer. All of the mutations severely reduced transcription of *rapF* *in vivo* (data not shown), indicating that a 4-bp spacer separating RE1 and RE2 is the optimal length for the transcription activation of *rapF* by ComA and that changes are not well-tolerated.

Since *rapE* has an unusual 5-bp spacer (TCTCA) separating RE1 and RE2 (Fig. 1), we sought to determine whether this atypically long spacer had an effect on transcription. A fragment of the *rapE* promoter containing only the promoter-proximal ComA-binding site and encoding the first 10 codons of *rapE* was fused to *lacZ* and used to monitor transcription. Expression of *rapE-lacZ* was low and relatively constant throughout growth, with little obvious increase in  $\beta$ -galactosidase-specific activity at high culture density (Fig. 6). Moreover, the low level of  $\beta$ -galactosidase-specific activity was reduced to background levels in a *comA*-null mutant, indicating that transcription of *rapE* is dependent on *comA* (data not shown). We constructed four single nucleotide deletions in the spacer separating RE1 and RE2 (resulting in spacer sequences TTCA, TCTA, TCTC, and CTCA) and measured the effects on *rapE* expression. Removal of a single nucleotide significantly increased transcription of *rapE* in all cases, albeit to different extents, depending on the sequence (Fig. 6). All of the four nucleotide spacers, with the exception of TCTA, allowed the density-dependent increase in transcription typically

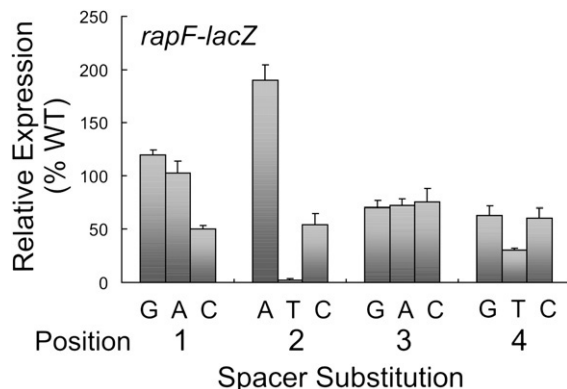


**Fig. 6.** Role of the spacer separating RE1 and RE2 in transcription activation of *rapE*. Cultures containing *PrapE-lacZ* fusions were grown in defined minimal medium, and aliquots were taken throughout growth for determination of  $\beta$ -galactosidase-specific activity. KG522 wild-type TCTCA spacer (filled diamonds); KG841 TTCA spacer (asterisk); KG521 TCTA spacer (squares); KG268 TCTC spacer (circles); KG852 CTCA spacer (triangles).

observed with other genes activated by ComA. The spacer sequence TTCA had the largest effect on transcription (i.e., a 6-fold increase in  $\beta$ -galactosidase-specific activity was observed at low culture density, and expression at high density was increased by ~17-fold relative to wild type) (Fig. 6). The spacer sequence TCTA also increased  $\beta$ -galactosidase-specific activity at low culture density by 6-fold compared to wild type, but resulted in the smallest induction in expression (~1.5-fold) at high culture density of all the spacer mutants tested (Fig. 6). The spacer sequences TCTC and CTCA had similar effects, increasing  $\beta$ -galactosidase-specific activity by ~6-fold at high culture density compared to wild type, while no effect was observed on expression at low culture density (Fig. 6). From this, we conclude that a 4-bp spacer separating RE1 and RE2 is optimal for transcriptional activation of target genes by ComA. Moreover, the sequence of the spacer affects both the level of expression observed at low culture density and the amount of induction during the response to high culture density.

### DNA sequence determinants in the spacer between RE1 and RE2

The sequence of the spacer separating RE1 and RE2 is not well conserved among the known ComA-binding sites [consensus N(a/t)T(g/c)].<sup>29</sup> However, the results with the spacer mutations in *rapE* (above) indicate that there might be additional sequence-specific information in this region. To determine whether sequence-specific information does exist in the spacer separating RE1 and RE2, we altered each base in the ComA-binding site of *rapF-lacZ* to the other three bases and measured effects on  $\beta$ -galactosidase-specific activity. In general, substitutions at positions 1 and 3 were well tolerated and had little or no effect on expression of *rapF*—with the exception being a T-to-C substitution at position 1, which decreased transcription by ~2-fold (Fig. 7). Substitutions at positions 2 and 4 had more dramatic effects on the transcription of *rapF*. Specifically, a T at either



**Fig. 7.** Effect of the spacer sequence separating RE1 and RE2 on transcription activation of *rapF*. Cultures containing *PrapF-lacZ* fusions were grown in defined minimal medium, and aliquots were taken throughout growth for determination of  $\beta$ -galactosidase-specific activity. The time point containing the maximal  $\beta$ -galactosidase activity that typically peaks at an  $OD_{600}$  of 1–1.5 is shown. The wild-type RE1–RE2 spacer is 5'-TGTA and is defined as 100%. Substitutions in the spacer: position 1: KG331 (T-G), KG282 (T-A), KG283 (T-C); position 2: KG284 (G-A), KG285 (G-T), KG305 (G-C); position 3: KG311 (T-G), KG312 (T-A), KG286 (T-C); position 4: KG325 (A-G), KG310 (A-T), KG287 (A-C). The data are an average of three independent experiments, with standard deviation shown.

position had a negative effect, virtually eliminating transcription when present at position 2 and reducing it by >2-fold when present at position 4 (Fig. 7). Interestingly, a G-to-A substitution at position 2 had a stimulatory effect on transcription, increasing it by ~2-fold relative to wild type (Fig. 7).

Taken together, our results indicate that there is sequence-specific information in the spacer region between RE1 and RE2. However, it appears that this information is largely context-specific; analyses of other spacer regions indicate that the effects of specific nucleotide changes differ among the various regulatory regions (unpublished results). A more detailed combinatorial analysis would be necessary to make general rules about the sequence information in the RE1–RE2 spacer region.

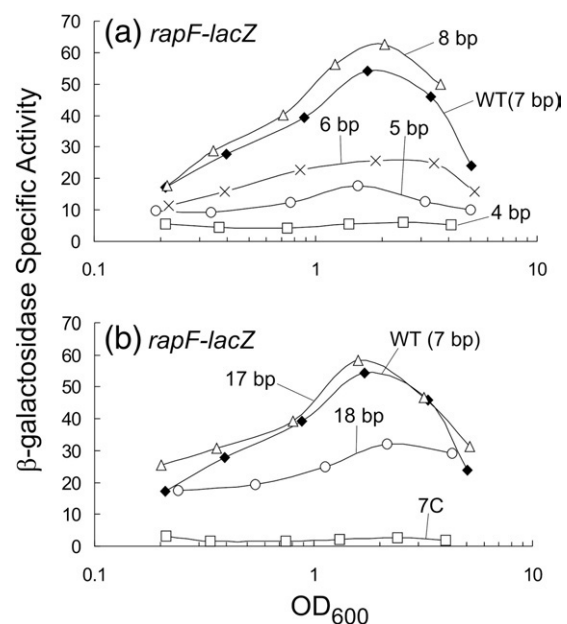
### Role of the spacer separating RE2 and RE3 in transcription activation

Of the known ComA-binding sites present within target gene promoters, the spacer region separating RE2 from RE3 varies in length from four to eight nucleotides, with a median length of 7 bp (Fig. 1; data not shown). We sought to determine the role of the spacer separating RE2 and RE3 in transcription by altering the length of the spacer in the *rapF*-regulatory region. We found that a 7-bp or a 17-bp spacer (an added helical turn of DNA) is optimal for transcription. Moreover, the sequence composition is important, as A/T-rich residues are required for transcription activation by ComA.

We altered the spacer separating RE2 and RE3 in *rapF-lacZ* from 7 bp to 8 bp, 6 bp, 5 bp, and 4 bp and monitored  $\beta$ -galactosidase activity throughout growth. An eight-nucleotide spacer (GAAAAAAA, with an extra A) had no effect on transcription of *rapF-lacZ* (Fig. 8a). Removal of additional nucleotides within the spacer to 6 bp, 5 bp, and 4 bp resulted in a progressive decrease in transcription of *rapF-lacZ*. For example, removal of one or two nucleotides (resulting in GAAAAA or GAAAA) decreased  $\beta$ -galactosidase-specific activity by ~2- or ~3-fold, respectively, while a 4-bp spacer (GAAA) reduced expression of *rapF-lacZ* to levels of a *comA*-null mutant (Figs. 4b and 8a).

To further refine the spacer length separating RE2 and RE3 and to investigate any helical phasing that might exist, we introduced a half-helical turn (5 or 6 adenine nucleotides) and a full-helical turn (10 or 11 adenine nucleotides) of DNA within this region of the ComA-binding site. Addition of a full-helical turn of DNA (10 adenine nucleotides) for a spacer length of 17 bp had no effect on the transcription of *rapF-lacZ*, and addition of 11 adenine nucleotides had a small effect, reducing  $\beta$ -galactosidase activity by ~2-fold compared to wild type (Fig. 8b). Addition of a half-helical turn of DNA reduced the transcription of *rapF* by 60–80% of wild type (data not shown).

To determine whether the spacing requirements separating RE2 and RE3 identified for *rapF* are also



**Fig. 8.** Effects of the spacer separating RE2 and RE3 on transcription activation of *rapF*. Cultures containing *PrapF-lacZ* fusions were grown in defined minimal medium, and aliquots were taken throughout growth for determination of  $\beta$ -galactosidase-specific activity. (a) KG277 wild-type 7-bp spacer G(A)<sub>6</sub> (filled diamonds); KG531 8-bp spacer G(A)<sub>7</sub> (triangles); KG324 6-bp spacer G(A)<sub>5</sub> (X); KG532 5-bp spacer G(A)<sub>4</sub> (circles); KG314 4-bp spacer G(A)<sub>3</sub> (squares). (b) KG277 wild-type 7-bp spacer G(A)<sub>6</sub> (filled diamonds) the same as in (a); KG309 17-bp spacer G(A)<sub>16</sub> (triangles); KG472 18-bp spacer G(A)<sub>17</sub> (circles); KG473 7-bp spacer entirely comprising cytosines (C)<sub>7</sub> (squares).



true for other target genes of ComA, we altered this region of the ComA-binding site in *rapA* and *rapC* and determined the effects on transcription. *rapA* has an 8-bp spacer separating RE2 and RE3 (TTCGACAA). Changing this spacer to 7 bp (TTCACAA) caused a small increase in the transcription of *rapA-lacZ* (data not shown). In contrast, *rapC* normally has a 7-bp spacer (ACAAAGA). Changing that to 8 bp caused a small decrease in the transcription of *rapC-lacZ* (data not shown). These results indicate that a 7-bp spacer separating RE2 and RE3 is optimal for activation by ComA, at least for the three promoters tested (*rapF*, *rapA*, and *rapC*).

To determine whether the same helical phasing exists between RE2 and RE3 as was determined for *rapF*, we introduced 9 and 10 consecutive adenine nucleotides between RE2 and RE3 of the *rapA* and *rapC* ComA-binding sites, respectively, to yield a total spacer length of 17 bp. Like *rapF*, introduction of DNA to the optimal 17-bp spacer length had no effect on transcription of *rapA* or *rapC* (data not shown). Taken together, we conclude that a 7-bp or a 17-bp spacer separating RE2 and RE3 is optimal for transcription activation of target genes by ComA and that the helical phasing is important.

The base composition of the spacer separating RE2 and RE3 is conserved among the known ComA-binding sites. The average A/T composition within this region of all known ComA-binding sites is 74%, which is unusually high even for the low G+C *B. subtilis*. We found that changing the base composition of the A/T-rich spacer to G/C caused a decrease in the transcription of *rapF-lacZ*. The seven adenine nucleotides were replaced with cytosines, and the  $\beta$ -galactosidase-specific activity of *rapF-lacZ* was reduced to levels similar to those in a *comA*-null mutant (Figs. 4b and 8b).

Taken together, our findings indicate that, within the context of a minimal ComA-binding site, the optimal spacer length separating RE2 and RE3 appears to be 7 bp or 17 bp. Finally, the composition of the spacer appears to be critical for transcription activation of *rapF* and, by inference, other known target genes.

### Estimation of the number of ComA-binding sites present in the *B. subtilis* genome

Based on the mutagenesis of the regulatory regions of several ComA-dependent target genes, we propose a refined consensus ComA-binding site (5'-TTGCGGnnnnCCGCAA- $n_{(6-8 \text{ or } 17-18)}$ -TTGCGG). Although our results indicate that the optimal RE2-RE3 spacer length is 7 bp or 17 bp, spacer lengths of 6 bp, 8 bp, and 18 bp are still functional for activation by ComA and are included in our consensus sequence. The identification of a revised ComA-binding site should aid in our understanding of how ComA functions to activate the transcription of target genes and in the identification of additional target genes, should they exist. The nine known target genes contain an average of 3.6 mismatches from the consensus sequence. Search of the *B. subtilis* genome using this refined sequence as query and allowing for 3 and 4

mismatches revealed 37 and 208 hits, respectively. This is significantly less than the number of predicted sites (~700) expected with the old consensus sequence.

The ComA-dependent regulatory regions of all known target genes were identified in our search, except for *yvfH* and *pel*, which contain a 4-bp and a 5-bp RE2-RE3 spacer, respectively. We excluded sites containing a spacer of <6 bp in our refined consensus sequence because the mutagenesis with *rapF* showed that a spacer of this length was deleterious for transcription (Fig. 8a). In addition to a small spacer, the *yvfH* ComA-binding site has six mismatches away from consensus, with half of them residing in RE3. Our mutagenesis of the *srfA*-regulatory region indicates that three mismatches in a single recognition element virtually eliminate transcription activation by ComA, unless additional upstream sites are present (Fig. 2). With no obvious upstream regulatory sequences present in the *yvfH* promoter, we presume that the combination of a degenerate RE3 and a suboptimal RE2-RE3 spacer explains why ComA has such a small effect (<2-fold) on transcription activation of *yvfH*.<sup>29</sup> *pel*, on the other hand, has additional upstream regulatory sequences that presumably compensate for an unusually small RE2-RE3 spacer (unpublished data).

Without considering the effects of additional regulatory elements (e.g., those present in *pel*), we may have underestimated the number of potential ComA-binding sites. Nonetheless, our genomic analysis revealed a total of 33 additional genes (many of which have no known function) that contain a putative ComA-binding site located within 500 bp upstream of the coding sequence. None of these genes was found to be affected by ComA using DNA microarray analyses in two independent studies.<sup>28,29</sup> It is possible that the ComA-binding site for some of these genes is nonfunctional for activation of transcription because it is not in the proper position relative to the binding site of RNA polymerase. Alternatively, some of these genes may require regulatory proteins, in addition to ComA, to activate transcription, as is the case with *degU*.<sup>24</sup> Another possibility is that some genes are negatively regulated by transcription factors under the growth conditions examined. We suspect that there are regulatory regions corresponding to each of these possibilities.

### Discussion

In this work, we found that the promoter regions for genes activated by ComA contain three recognition elements: RE1, RE2, and RE3. RE1 and RE2 comprise an inverted repeat, and the RE3 consensus sequence is identical with that of RE1. Each of these sequence elements is required for ComA-dependent transcriptional activation of target genes. Each element influences binding by purified  $\text{his}_6$ -ComA *in vitro*, and previous footprinting studies<sup>14,36</sup> indicate some protection of all three elements by ComA. The simplest interpretation of these findings is that

the ComA-binding site is composed of all three recognition elements.

### Model for ComA-binding DNA

In addition to the three recognition elements, the sequences separating them are important for transcription activation of target genes by ComA, presumably because they function to properly position the recognition elements for ComA–DNA binding. As a result, permutations in the length of the spacers typically have deleterious effects on transcription (Figs. 6 and 8).

The effects of altering spacers length strongly support a model in which a dimer of ComA binds to RE1 and RE2 and a second dimer binds to RE3 (Fig. 9). The strict 4-bp spacing requirement separating RE1 and RE2 is consistent with a single dimer of ComA occupying these recognition elements. Moreover, the flexibility in the spacer separating RE2 from RE3 (i.e., spacer length of 6–8 bp or 17–18 bp) indicates that another dimer of ComA probably binds RE3 and interacts with the dimer bound at RE1 and RE2. In this model, the RE2–RE3 spacer serves as a flexible bridge allowing the two dimers to interact, resulting in cooperative binding observed in the gel mobility assays (Fig. 3). Replacement of the A/T-rich spacer separating RE2 and RE3 with G-C base pairs virtually eliminated transcription activation of *rapF* (Fig. 8b), presumably because the DNA could not bend properly to allow for functional alignment of the recognition elements.

### Prevalence of a tripartite ComA-regulatory sequence in other *Bacillus* species

Homologs of ComA are present in other *Bacillus* species, including *Bacillus licheniformis* ATCC 14580, *Bacillus amyloliquefaciens* FZB42, and *Bacillus pumilus* SAFR-032. In *B. licheniformis*, ComA directly reg-

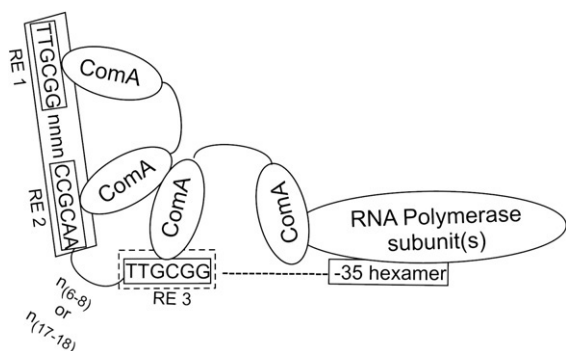
ulates transcription of *lchA*, which is involved in the production of the lipopeptide lichenysin A.<sup>38,39</sup> The *lchA*-regulatory region of *B. licheniformis* resembles that of *srfA* in *B. subtilis*, with two inverted repeats (5'-TTTCGGtatacACGCAT and 5'-ATTTCGGcatcCCGCAT) separated by 17 bp. Mutational analyses of the promoter-proximal inverted repeat revealed the importance of RE1 and RE2 in the transcription of *lchA* by ComA.<sup>38</sup> The existence of RE3 was not known at the time of that analysis; however, closer examination of the *lchA* promoter region reveals a putative RE3 (5'-TTTCAC) located 6 bp downstream of RE2 in the promoter-proximal ComA-binding site.

In *B. amyloliquefaciens*, ComA was postulated to directly regulate transcription of *degQ*.<sup>40</sup> Analysis of the *degQ* promoter reveals a well-conserved RE1 and RE2 inverted repeat (5'-TTGCGGtgatcACGCAG), with a putative RE3 (5'-TTTCGG) positioned 17 bp downstream of RE2. It appears that ComA likely utilizes a similar tripartite binding site to activate transcription of target genes in other *Bacillus* species.

### Degeneracy of the ComA binding site is required for normal cell density-dependent regulation

We found that degeneracy of the ComA-binding site is important for the regulation of genes in a population-density-dependent manner. ComA-binding sites in gene-regulatory regions average 3.6 mismatches away from consensus. Promoters with a near-consensus ComA-binding site(s) have elevated transcription at low culture density compared to promoters with a degenerate site. This elevated activity, in turn, depresses the magnitude of the response (fold induction) observed at high culture density, thus lessening the ability to coordinate transcription with population density. On the other hand, promoters with a degenerate site have low transcription at low culture density and respond to increased population density with significantly increased transcription, resulting in a larger induction ratio.

These trends are most obvious when comparing transcription of *rapA* and *srfA*. *rapA* has a near-consensus binding site with high expression at low culture density and a modest 2.5-fold induction as the population density increases (Fig. 4a). In contrast, *srfA* has two degenerate binding sites, resulting in low expression at low culture density and in an ~10-fold increase in expression at high culture density (Fig. 2) (i.e., *srfA* is regulated in a population-density-dependent manner). Mutations in the promoter-proximal ComA-binding site of the *srfA*-regulatory region toward the consensus sequence increased expression of *srfA* at low culture density so much that no further increase in transcription was observed at high culture density (Fig. 2b). This result could indicate that there is no regulation by population density and that expression is always at a high level. Alternatively, there could still be some regulation by culture density, but we are technically unable to go to a low-enough density to see the effect. In either case, the normal pattern of cell density regulation is abolished, and although a better binding site



**Fig. 9.** Model for ComA-binding DNA. We propose that two dimers of ComA bind DNA, with one dimer occupying RE1 and RE2 and a second dimer occupying RE3 and nonspecific sequence downstream. Protein-protein interactions between the two dimers probably help stabilize the complex, resulting in cooperative binding observed in the gel shift assays (Fig. 3). The A/T-rich tract separating RE2 and RE3 probably facilitates DNA bending and the proper positioning of the recognition elements for ComA–DNA binding.

enhances ComA–DNA binding and transcription activation of target genes, it is detrimental to the regulation of genes in a population-density-dependent manner.

The affinity of ComA for its DNA-binding site provides a mechanism for controlling the temporal expression of regulon genes and for fine-tuning the response to population density. Transcription of the ComA regulon is dependent on the concentration of ComA ~ P. At low culture density, ComA is predominately in the nonphosphorylated inactive state. The small amounts of ComA ~ P present at low culture density probably bind to high-affinity sites (e.g., the regulatory region of *rapA*), resulting in enhanced transcription of target genes at low population densities (Fig. 4a). In contrast, we presume that low-affinity degenerate sites (e.g., those present in *srfA*) are largely unoccupied by the small amounts of ComA ~ P present at low culture densities, resulting in low levels of transcription (Fig. 2).

Degeneracy in transcription factor binding sites is analogous to degeneracy in bacterial promoter sequences. There exists tremendous variation in the sequences of bacterial promoters recognized by a given form of RNA polymerase. Many weak promoters require activator proteins that stimulate transcription initiation, often by recruiting RNA polymerase to the promoter.<sup>41–43</sup> In the well-studied example of the *lac* operon promoter, mutations in the promoter toward consensus can bypass the need for the activator CAP cAMP, thereby reducing some of the regulation normally associated with the activator. Thus, sequence degeneracy in certain DNA-binding sites is critical to their regulatory function. Definitions of consensus sequences can be misleading and often fail to capture the importance of weaker binding sites for regulation.

## Materials and Methods

### Bacterial strains and growth media

Routine cloning was performed in *E. coli* strain DH5 $\alpha$ . *B. subtilis* strains (Table 1) were all derived from the parental strain JH642 (*trpC2 pheA1*).<sup>44</sup> Liquid cultures of *B. subtilis* were grown in S7 defined minimal medium salts<sup>45</sup> containing 50 mM 4-morpholinepropanesulfonic acid instead of 100 mM (S7<sub>50</sub>) and supplemented with 1% glucose, 0.1% glutamate, tryptophan (40  $\mu$ g/ml), phenylalanine (40  $\mu$ g/ml), and threonine (120  $\mu$ g/ml), where appropriate. *B. subtilis* was grown on solid medium containing Spizizen minimal salts<sup>46</sup> supplemented with 1% glucose, 0.1% glutamate, and appropriate individual amino acids as described above. LB agar plates were used for routine cloning and growth of *B. subtilis* and *E. coli*. The following concentrations of antibiotics were used: ampicillin, 100  $\mu$ g/ml; neomycin, 2.5  $\mu$ g/ml; chloramphenicol, 5  $\mu$ g/ml.

### Oligonucleotides

All oligonucleotides used in this study were synthesized by Integrated DNA Technologies, and sequences are available upon request.

### Cloning and mutagenesis

Transcriptional fusions to *lacZ* were first created by amplifying the promoter of interest from *B. subtilis* genomic DNA using polymerase chain reaction (PCR) with *Taq* DNA polymerase (Roche). Fusions are indicated in the strain table, and numbers in the regulatory regions represent the number of base pairs from the start of the open reading frame. EcoRI and BamHI restriction enzyme recognition sites were engineered into the 5' and 3' ends of each PCR product, respectively. PCR products were digested with EcoRI and BamHI restriction enzymes (NEB) and ligated into pKS2,<sup>9</sup> which was also digested with the same two enzymes. Ligation reactions were transformed into strain DH2 $\alpha$  and plated on LB with ampicillin. Plasmid DNA was isolated from transformants by the alkaline lysis method in accordance with the manufacturers' instructions (Qiagen and Invitrogen). Clones were verified by DNA sequencing (MIT Biopolymers Laboratory and MGH Sequencing Facility). Plasmid DNA was transformed into *B. subtilis* strain JH642 and plated on LB with neomycin. All *lacZ* fusions contained the first 10 amino acids of the coding sequence of the gene of interest, followed by a termination codon. Mutations in the ComA-binding sites were created by add-on PCR or PCR SOEing,<sup>47</sup> where appropriate.

An overexpression construct was made to express ComA as a his<sub>6</sub>-tagged fusion protein (his<sub>6</sub>-ComA) for purification in *E. coli*. Briefly, an N-terminal his<sub>6</sub> tag was introduced between codons 1 and 2 of *comA* by add-on PCR of *B. subtilis* genomic DNA using *Taq* polymerase and the following primers: 5'-GCTTAGTGGGT**ACCA**AGGAGATATACATATGcatcaccatcaccatcacAAAAAGATACTAGTGATTGA-3' and 5'-TGCTACGAG**CA**TGCTTAAAGTACACCGTCTGA-3', where KpnI and SphI restriction enzymes are in bold, the ribosome-binding site is underlined, the his<sub>6</sub> tag is in lowercase, and the termination codon is in italics. The PCR product was digested with KpnI and SphI and ligated into pBAD-Ap18, which was also digested with the same two restriction enzymes. The ligation reaction was transformed into strain DH5 $\alpha$  and plated on LB with ampicillin. Plasmid DNA was isolated, and the correct identity was verified by sequencing.

### Growth conditions and assay of $\beta$ -galactosidase activity

Overnight cultures were grown as light lawns on minimal medium plates incubated at 37 °C. Three milliliters of Spizizen salts was used to flood each plate, and OD<sub>600</sub> was determined using a spectrophotometer. Shaker flasks containing S7<sub>50</sub> minimal medium were inoculated to OD<sub>600</sub> ~ 0.02 and incubated with vigorous aeration at 37 °C. One-milliliter aliquots were taken at specified times throughout the growth cycle and placed in a 2.2-ml 96-well polypropylene block (Qiagen), which was stored at –20 °C until time to assay  $\beta$ -galactosidase activity. A second aliquot was taken to determine OD<sub>600</sub>.

$\beta$ -Galactosidase-specific activity was determined as described,<sup>48</sup> with some modifications. Briefly, cells in the 96-well blocks were thawed to room temperature, and 20  $\mu$ l of toluene was added to each well. Cells were permeabilized directly in the blocks by vigorous pipetting up and down using a multichannel pipettor. Permeabilized cells were transferred to a second block containing 1 ml of Z-buffer.<sup>48</sup> A 100- $\mu$ l aliquot of the cell suspension was transferred to a microtiter plate, and the  $\beta$ -galactosidase assay was initiated with the addition of 20  $\mu$ l of freshly prepared Ortho-nitrophenyl- $\beta$ -D-galactopyranoside (4 mg/ml) and terminated with the addition of 40  $\mu$ l of 1 M Na<sub>2</sub>CO<sub>3</sub>. Cell debris



**Table 1.** Strains used

| Strain                   | Genotype <sup>a</sup>  |
|--------------------------|--|
| <i>srfA-lacZ fusions</i> |  |
| KG102                    | <i>amyE::[srfA (−372 to +10)-lacZ neo]</i>   |
| KG125                    | <i>amyE::[srfA (−434 to +10)-lacZ neo]</i>   |
| KG150                    | <i>amyE::[srfA (−434 to +10)-lacZ neo]ΔcomA::cat</i>   |
| KG158                    | <i>amyE::[srfA (−372 to +10; −362T-G; −349T-A; −340T-G; −338A-G; −337C-G)-lacZ neo]</i>            |
| KG160                    | <i>amyE::[srfA (−372 to +10; −340T-G; −338A-G; −337C-G)-lacZ neo]</i>                              |
| KG464                    | <i>amyE::[srfA (−372 to +10; −362T-G; −349T-A; −340T-G; −338A-G; −337C-G)-lacZ neo]ΔcomA::cat</i>  |
| KG565                    | <i>amyE::[srfA (−372 to +10; −362T-G; −349T-A; −342 to −337GCATAT)-lacZ neo]</i>                   |
| KG567                    | <i>amyE::[srfA (−372 to +10; −362T-G; −354 to −349ATATGC; −340T-G; −338A-G; −337C-G)-lacZ neo]</i> |
| KG780                    | <i>amyE::[srfA (−372 to +10; −364 to −359GCATAT; −349T-A; −340T-G; −338A-G; −337C-G)-lacZ neo]</i> |
| <i>rapA-lacZ fusions</i> |  |
| KG112                    | <i>amyE::[rapA (−126 to +10)-lacZ neo]</i>   |
| KG148                    | <i>amyE::[rapA (−126 to +10)-lacZ neo]ΔcomA::cat</i>   |
| KG513                    | <i>amyE::[rapA (−126 to +10; −96C-T)-lacZ neo]</i>   |
| KG544                    | <i>amyE::[rapA (−126 to +10; −93A-C; −101G-A; −104G-T)-lacZ neo]</i>                               |
| KG545                    | <i>amyE::[rapA (−126 to +10; −93A-C; −80A-T; −77G-A)-lacZ neo]</i>                                 |
| <i>rapF-lacZ fusions</i> |  |
| KG239                    | <i>amyE::[rapF (−731 to +10)-lacZ neo]ΔcomA::cat</i>   |
| KG266                    | <i>amyE::[rapF (−108 to +10; −83T-G; −80C-T; −79G-C)-lacZ neo]</i>                                 |
| KG277                    | <i>amyE::[rapF (−108 to +10)-lacZ neo]</i>   |
| KG282                    | <i>amyE::[rapF (−108 to +10; −100T-A)-lacZ neo]</i>  |
| KG283                    | <i>amyE::[rapF (−108 to +10; −100T-C)-lacZ neo]</i>  |
| KG284                    | <i>amyE::[rapF (−108 to +10; −99G-A)-lacZ neo]</i>   |
| KG285                    | <i>amyE::[rapF (−108 to +10; −99G-T)-lacZ neo]</i>   |
| KG286                    | <i>amyE::[rapF (−108 to +10; −98T-C)-lacZ neo]</i>   |
| KG287                    | <i>amyE::[rapF (−108 to +10; −97A-C)-lacZ neo]</i>   |
| KG305                    | <i>amyE::[rapF (−108 to +10; −99G-C)-lacZ neo]</i>   |
| KG309                    | <i>amyE::[rapF (−108 to +10; −84[+10A])-lacZ neo]</i>  |
| KG310                    | <i>amyE::[rapF (−108 to +10; −97A-T)-lacZ neo]</i>   |
| KG311                    | <i>amyE::[rapF (−108 to +10; −98T-G)-lacZ neo]</i>   |
| KG312                    | <i>amyE::[rapF (−108 to +10; −98T-A)-lacZ neo]</i>   |
| KG314                    | <i>amyE::[rapF (−108 to +10; Δ[−86 to −84])-lacZ neo]</i>  |
| KG324                    | <i>amyE::[rapF (−108 to +10; Δ−84A)-lacZ neo]</i>  |
| KG325                    | <i>amyE::[rapF (−108 to +10; −97A-G)-lacZ neo]</i>   |
| KG331                    | <i>amyE::[rapF (−108 to +10; −100T-G)-lacZ neo]</i>  |
| KG472                    | <i>amyE::[rapF (−108 to +10; −84[+11A])-lacZ neo]</i>  |
| KG473                    | <i>amyE::[rapF (−108 to +10; −90G-C; [−89 to −84]A-C)-lacZ neo]</i>                                |
| KG531                    | <i>amyE::[rapF (−108 to +10; −84[+1A])-lacZ neo]</i>   |
| KG532                    | <i>amyE::[rapF (−108 to +10; Δ[−85 to −84])-lacZ neo]</i>  |
| KG555                    | <i>amyE::[rapF (−108 to +10; −96A-C)-lacZ neo]</i>   |
| KG556                    | <i>amyE::[rapF (−108 to +10; −81G-T)-lacZ neo]</i>   |
| KG557                    | <i>amyE::[rapF (−108 to +10; −96A-C; −81G-T; −78G-T)-lacZ neo]</i>                                 |
| KG566                    | <i>amyE::[rapF (−108 to +10; −96A-C; −81G-T)-lacZ neo]</i>   |
| <i>rapE-lacZ fusions</i> |  |
| KG268                    | <i>amyE::[rapE (−112 to +10; Δ−98A)-lacZ neo]</i>  |
| KG521                    | <i>amyE::[rapE (−112 to +10; Δ−99C)-lacZ neo]</i>  |
| KG522                    | <i>amyE::[rapE (−112 to +10)-lacZ neo]</i>   |
| KG841                    | <i>amyE::[rapE (−112 to +10; Δ−101C)-lacZ neo]</i>   |
| KG852                    | <i>amyE::[rapE (−112 to +10; Δ−102T)-lacZ neo]</i>   |

<sup>a</sup> All strains are derived from JH642 and contain *trpC2* and *pheA1* alleles (not indicated). The position of DNA relative to the start of the coding sequence and alterations in the ComA-binding site are indicated inside parentheses.

was pelleted in the microtiter plate by centrifugation at 3000g for 10 min. A 100-μl aliquot of each supernatant was transferred to a new plate using a multichannel pipettor. A<sub>420</sub> was determined using a SpectraMax plate reader (Molecular Dynamics), and data analysis was performed using Microsoft Excel. β-Galactosidase-specific activity was calculated as follows: 1000 × {(ΔA<sub>420</sub>/min/ml)/OD<sub>600</sub> of culture}.

### Purification of his<sub>6</sub>-ComA

A fresh overnight culture of strain DH5α containing pBAD-his<sub>6</sub>-ComA was diluted 1:200 into LB with ampi-

cillin (300 μg/ml) and grown to an OD<sub>600</sub> of 0.5–0.8 at 37 °C, with vigorous aeration. L-Arabinose (Sigma) was added (final concentration, 0.2%) to induce expression from pBAD. Cells were harvested 4–6 h later by centrifugation at 5000g for 5 min at 4 °C. The cell pellet from 1 L of culture was resuspended in 10 ml of sonication buffer (20 mM Tris pH 8, 0.3 M NaCl, 5% glycerol, 5 mM imidazole, 5 mM β-mercaptoethanol, and 5 mM MgCl<sub>2</sub>), and cells were lysed by sonication (eight cycles of 20 s on and 40 s off, at settings 4 and 5). The culture was cleared by centrifugation at 10,000 rpm for 20 min at 4 °C, and the cell extract was passed over 2 ml of Ni-NTA (Qiagen). After 10 washes with 10 ml of sonication buffer, his<sub>6</sub>-ComA was eluted from the column in 10 ml of sonication buffer with

increasing concentrations of imidazole (15 mM, 50 mM, 120 mM, and 300 mM). The fractions were analyzed for purity by SDS-PAGE, followed by Coomassie staining. Typically, the fraction eluted in 120 mM imidazole was >95% pure (data not shown) and was dialyzed to remove imidazole against five buffer changes of 2 L of dialysis buffer (20 mM Tris pH 8, 0.3 M NaCl, 5% glycerol, 10 mM  $\beta$ -mercaptoethanol, and 5 mM MgCl<sub>2</sub>). Dialyzed protein was concentrated to >10 mg/ml using Centricon-10 (Amicon). Glycerol was added to a final concentration of 40%, and the protein concentration was determined by Bradford assay using bovine serum albumin (BSA) as protein standard. Purified his<sub>6</sub>-ComA was stored at -20 °C until further use.

### Gel mobility shift assays

DNA for the gel mobility shift assays was prepared by annealing two complementary oligonucleotides containing the promoter-proximal ComA-binding site in the *srfA*-regulatory region from -73 bp to -46 bp from the start of the annotated coding sequence. Briefly, one of the oligonucleotides from each pair was labeled on its 5' end using [<sup>32</sup>P]γ-ATP (NEN) and T4 polynucleotide kinase (NEB). A 1.3-fold molar excess of its complement was added to the mixture and heated to 95 °C for 5 min, followed by slow cooling to room temperature to facilitate annealing of the oligonucleotides. Duplex DNA was purified away from unincorporated label using a G-25 Centrispin-10 column (Princeton Separations). DNA templates contained bases 5'-TCA preceding the ComA binding sequence and bases TC-3' following it.

*In vitro* binding reactions contained 10 mM Hepes (pH 7.6), 2 mM MgCl<sub>2</sub>, 0.1 mM ethylenediaminetetraacetic acid, 0.2 M KCl, 10% glycerol, 5 mM DTT, 5–10 nM labeled DNA, 10 nM poly(dI-dC), and purified his<sub>6</sub>-ComA. Protein-DNA complexes were allowed to equilibrate at 37 °C for 30 min prior to the addition of 5 μl of 5× agarose gel loading dye. Samples were loaded into the wells of 15% polyacrylamide gels containing 5% glycerol and electrophoresed into the gel at 300 V. Once the loading dye had entered the gel, the voltage was reduced to 120 V and gels were run for 5–6 h at 4 °C. Gels were dried and analyzed using a PhosphorImager (Molecular Dynamics).

### Determination of the oligomeric state of ComA

The oligomeric state of ComA was determined using native gels as previously described.<sup>37</sup> Briefly, purified his<sub>6</sub>-ComA and native protein standards, including bovine milk α-lactalbumin (molecular mass, 14.2 kDa), carbonic anhydrase from bovine erythrocytes (molecular mass, 29 kDa), chicken egg albumin (molecular mass, 45 kDa), BSA monomer (66 kDa), and BSA dimer (132 kDa), were subjected to electrophoresis in native gels containing 6%, 7%, 8%, 9%, and 10% acrylamide. The relative mobility (*R<sub>f</sub>*) of each protein was determined by dividing its migration distance from the top of the gel to the center of the protein band by the migration distance of the bromophenol blue tracking dye from the top of the gel. A standard curve was generated by plotting 100×{log(*R<sub>f</sub>*×100)} versus the gel concentration. The negative slopes generated from the standard curve were plotted against the known molecular masses of the protein standards. The approximate native molecular mass of his<sub>6</sub>-ComA was estimated. The experiments were performed three times, with an average native molecular mass of 53 kDa for his<sub>6</sub>-ComA.

### BLAST searches of the *B. subtilis* genome

BLAST pattern searches of the *B. subtilis* genome were performed using the Subtilist website†. The previously proposed consensus sequence (5'-TTGCGGnnnnCCG-CAA) and the refined consensus sequence (5'-TTGCGGnnnn-CCGCAA-*n*<sub>(6–8 or 17–18)</sub>-TTGCGG) were used as query sequences. Using the previously proposed sequence as query and allowing for two and three mismatches, 372 and 3210 hits, respectively, were revealed, compared to just 37 and 208 hits using the refined consensus sequence as query and allowing for three and four mismatches, respectively. All previously known ComA-binding sites present in target gene promoters were identified in both BLAST searches, except for *yofH* and *pel*, which have a 4-bp and a 5-bp spacer separating RE2 and RE3, respectively.

### Acknowledgements

We thank C.A. Lee and R.B. Weart for comments on the manuscript. This work was supported, in part, by National Institutes of Health Ruth L. Kirschstein NRSA grant GM071224 to K.L.G. and Public Health Service grant GM50895 from the National Institutes of Health to A.D.G.

### References

1. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. (1994). Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J. Bacteriol.* **176**, 269–275.
2. Redfield, R. J. (2002). Is quorum sensing a side effect of diffusion sensing? *Trends Microbiol.* **10**, 365–370.
3. Winans, S. C. & Bassler, B. L. (2008). *Chemical Communication among Bacteria*. ASM Press, Washington, DC.
4. Waters, C. M. & Bassler, B. L. (2005). Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.*, 319–346.
5. Auchtung, J. M. & Grossman, A. D. (2008). Extra-cellular peptide signaling and quorum responses in development, self-recognition, and horizontal gene transfer in *Bacillus subtilis*. In *Chemical Communication Among Microbes* (Winans, S. C. & Bassler, B. L., eds.), pp. 13–30, ASM Press, Washington, DC.
6. Lazazzera, B., Palmer, T., Quisel, J. & Grossman, A. D. (1999). Cell density control of gene expression and development in *Bacillus subtilis*. In *Cell-Cell Signaling in Bacteria* (Dunny, G. M. & Winans, S. C., eds), pp. 27–46, ASM Press, Washington, DC.
7. Msadek, T. (1999). When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends Microbiol.* **7**, 201–207.
8. Tortosa, P. & Dubnau, D. (1999). Competence for transformation: a matter of taste. *Curr. Opin. Microbiol.* **2**, 588–592.
9. Magnuson, R., Solomon, J. & Grossman, A. D. (1994). Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell*, **77**, 207–216.

† <http://genolist.pasteur.fr/Subtilist/>

10. Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandic-Mulec, I. & Dubnau, D. (2001). Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. *J. Bacteriol.* **183**, 451–460.
11. Ansaldi, M., Marolt, D., Stebe, T., Mandic-Mulec, I. & Dubnau, D. (2002). Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Mol. Microbiol.* **44**, 1561–1573.
12. Bacon Schneider, K. P. T. M. & Grossman, A. D. (2002). Characterization of comQ and comX, two genes required for the production of ComX pheromone in *Bacillus subtilis*. *J. Bacteriol.* **184**, 410–419.
13. Weinrauch, Y., Penchev, R., Dubnau, E., Smith, I. & Dubnau, D. (1990). A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems. *Genes Dev.* **4**, 860–872.
14. Roggiani, M. & Dubnau, D. (1993). ComA, a phosphorylated response regulator protein of *Bacillus subtilis*, binds to the promoter region of *srfA*. *J. Bacteriol.* **175**, 3182–3187.
15. Grossman, A. D. (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu. Rev. Genet.* **29**, 477–508.
16. Solomon, J. M., Lazazzera, B. A. & Grossman, A. D. (1996). Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes Dev.* **10**, 2014–2024.
17. Core, L. A. P. M. (2003). TPR-mediated interaction of RapC with ComA inhibits response regulator–DNA binding for competence development in *Bacillus subtilis*. *Mol. Microbiol.* **49**, 1509–1522.
18. Bongiorno, C., Ishikawa, S., Stephenson, S., Ogasawara, N. & Perego, M. (2005). Synergistic regulation of competence development in *Bacillus subtilis* by two Rap–Phr systems. *J. Bacteriol.* **187**, 4353–4361.
19. Auchtung, J. M., Lee, C. A. & Grossman, A. D. (2006). Modulation of the ComA-dependent quorum response in *Bacillus subtilis* by multiple Rap proteins and Phr peptides. *J. Bacteriol.* **188**, 5273–5285.
20. Hayashi, H., Kensuke, T., Kobayashi, K., Ogasawara, N. & Ogura, M. (2006). *Bacillus subtilis* RghR (YvaN) represses rapG and rapH, which encode inhibitors of expression of the *srfA* operon. *Mol. Microbiol.* **59**, 1714–1729.
21. Smits, W. K., Bongiorno, C., Veengin, J. W., Hamoen, L. W., Kuipers, O. P. & Perego, M. (2007). Temporal separation of distinct differentiation pathways by a dual specificity Rap–Phr system in *Bacillus subtilis*. *Mol. Microbiol.* **65**, 103–120.
22. Lazazzera, B. A. (2000). Quorum sensing and starvation: signals for entry into stationary phase. *Curr. Opin. Microbiol.* **3**, 177–182.
23. Pottathil, M. & Lazazzera, B. A. (2003). The extracellular Phr peptide–Rap phosphatase signaling circuit of *Bacillus subtilis*. *Front. Biosci.* **8**, 32–45.
24. Msadek, T., Kunst, F., Klier, A. & Rapoport, G. (1991). DegS–DegU and ComP–ComA modulator–effector pairs control expression of the *Bacillus subtilis* pleiotropic regulatory gene *degQ*. *J. Bacteriol.* **173**, 2366–2377.
25. Mueller, J. P., Bukusoglu, G. & Sonenshein, A. L. (1992). Transcriptional regulation of *Bacillus subtilis* glucose starvation-inducible genes: control of *gsiA* by the ComP–ComA signal transduction system. *J. Bacteriol.* **174**, 4361–4373.
26. Nakano, M. M. & Zuber, P. (1993). Mutational analysis of the regulatory region of the *srfA* operon in *Bacillus subtilis*. *J. Bacteriol.* **175**, 3188–3191.
27. Jiang, M., Grau, R. & Perego, M. (2000). Differential processing of propeptide inhibitors of Rap phosphatases in *Bacillus subtilis*. *J. Bacteriol.* **182**, 303–310.
28. Ogura, M., Yamaguchi, H., Yoshida, K., Fujita, Y. & Tanaka, T. (2001). DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B. subtilis* two-component regulatory systems. *Nucleic Acids Res.* **29**, 3804–3813.
29. Comella, N. & Grossman, A. D. (2005). Conservation of genes and processes controlled by the quorum response in bacteria: characterization of genes controlled by the quorum sensing transcription factor ComA in *Bacillus subtilis*. *Mol. Microbiol.* **57**, 1159–1174.
30. Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* **47**, 597–626.
31. Jourlin-Castelli, C., Mani, N., Nakano, M. M. & Sonenshein, A. L. (2000). CcpC, a novel regulator of the LysR family required for glucose repression of the *citB* gene in *Bacillus subtilis*. *J. Mol. Biol.* **295**, 865–878.
32. Nakano, M. M., Marahiel, M. A. & Zuber, P. (1988). Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *J. Bacteriol.* **170**, 5662–5668.
33. Jaacks, K. J., Healy, J., Losick, R. & Grossman, A. D. (1989). Identification and characterization of genes controlled by the sporulation regulatory gene *spo0H* in *Bacillus subtilis*. *J. Bacteriol.* **171**, 4121–4129.
34. van Sinderen, D., Withoff, S., Boels, H. & Venema, G. (1990). Isolation and characterization of *comL*, a transcription unit involved in competence development of *Bacillus subtilis*. *Mol. Gen. Genet.* **224**, 396–404.
35. Nakano, M. M., Magnuson, R., Meyers, A., Curry, J., Grossman, A. D. & Zuber, P. (1991). *srfA* is an operon required for surfactin production, competence development, and efficient sporulation in *Bacillus subtilis*. *J. Bacteriol.* **173**, 1770–1778.
36. Zhang, Y., Nakano, S., Choi, S. Y. & Zuber, P. (2006). Mutational analysis of the *Bacillus subtilis* RNA polymerase alpha C-terminal domain supports the interference model of Spx-dependent repression. *J. Bacteriol.* **188**, 4300–4311.
37. Bryan, J. K. (1977). Molecular weights of protein multimers from polyacrylamide gel electrophoresis. *Anal. Biochem.* **78**, 513–519.
38. Yakimov, M. M. & Golyshin, P. N. (1997). ComA-dependent transcriptional activation of lichenysin A synthetase promoter in *Bacillus subtilis* cells. *Biotechnol. Prog.* **13**, 757–761.
39. Yakimov, M. M., Kroger, A., Slepak, T. N., Giuliano, L., Timmis, K. N. & Golyshin, P. N. (1998). A putative lichenysin A synthetase operon in *Bacillus licheniformis*: initial characterization. *Biochim. Biophys. Acta*, **141**–153.
40. Koumoutsis, A., Chen, X. H., Vater, J. & Borriess, R. (2007). DegU and YczE positively regulate the synthesis of bacillomycin D by *Bacillus amyloliquefaciens* strain FZB42. *Appl. Environ. Microbiol.* **73**, 6953–6964.
41. Rosenberg, M. & Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**, 319–353.
42. Reznikoff, W. S. & Abelson, J. N. (1978). The *lac* promoter. In *The Operon* (Miller, J. H. & Reznikoff,



- W.S., eds), pp. 221–243, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
43. Busby, S. & Ebright, R. H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell*, **79**, 743–746.
  44. Perego, M., Spiegelman, G. B. & Hoch, J. A. (1988). Structure of the gene for the transition state regulator, *abrB*: regulator synthesis is controlled by the *spo0A* sporulation gene in *Bacillus subtilis*. *Mol. Microbiol.* **2**, 689–699.
  45. Vasantha, N. & Freese, E. (1980). Enzyme changes during *Bacillus subtilis* sporulation caused by deprivation of guanine nucleotides. *J. Bacteriol.* **144**, 1119–1125.
  46. Harwood, C. R. & Cutting, S. M. (1990). *Molecular Biological Methods for Bacillus*. John Wiley and Sons, Chichester, England.
  47. Horton, R. M., Cai, Z. L., Ho, S. N. & Pease, L. R. (1990). Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques*, **8**, 528–535.
  48. Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor Press, Cold Spring Harbor, NY.