A Bacillus subtilis regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems

Yvette Weinrauch, Ruska Penchev, Eugenie Dubnau, Issar Smith, and David Dubnau

Department of Microbiology, The Public Health Research Institute, New York, New York 10016 USA

A Bacillus subtilis gene, required for genetic competence, was identified immediately upstream from the previously characterized gene comA. The comA gene product has been found to exhibit amino acid sequence similarity to the so-called effector class of signal-transduction proteins. DNA sequencing of the new determinant, named comP, revealed that the carboxy-terminal domain of the predicted ComP protein is similar in amino acid sequence to that of several sensor members of the bacterial two-component signal-transduction systems. The predicted amino-terminal domain contains several hydrophobic segments, postulated to be membrane-spanning. In vitro-derived comP disruptions are epistatic on the expression of all late competence genes tested, including comG, comC, comD, and comE, but not on expression of the early gene comB. Although comA has its own promoter, some transcription of comA, especially later in growth, occurs via readthrough from comP sequences. A roughly twofold epistatic effect of a comP disruption was noted on the downstream comA determinant, possibly due to interruption of readthrough transcription from comP to comA. Overexpression of comA fully restored competence to a comP mutant, providing evidence that ComA acts after ComP, and consistent with a role for the latter protein in activation of the former, possibly by phosphorylation. ComP probably is involved in transmitting information concerning the nutritional status of the medium, particularly the presence of nitrogen- and carbon-containing nutrients. ComP was also shown to play a role in sporulation, at least partly interchangeable with that of SpoIIJ, another putative sensor protein.

[Key Words: Genetic competence; signal transduction; sporulation; global regulation]

Received December 11, 1989; revised version accepted February 20, 1990.

Genetic competence in Bacillus subtilis requires the expression of several gene products that are responsible for the binding, processing, and uptake of transforming DNA [for review, see Dubnau 1989]. The manufacture of these proteins, which are encoded by so-called late competence genes, is controlled by an elaborate network of regulatory elements in response to environmental signals about which little is known. The competence regulon is expressed after the transition to stationary phase, and only in glucose minimal salts medium. It is therefore an example of a so-called late-growth-regulated system and is subject to nutritional control. At maximal competence, only 10–20% of the cells are competent, and these can be separated from the noncompetent cells and shown to be physiologically unique (Hadden and Nester 1968; Haseltine Cahn and Fox 1968). Thus, a nonterminal differentiation process results in the elaboration of two cell types. To investigate the regulation of this system, we have characterized a series of mutations that are epistatic on the expression of late competence genes and that appear to play a regulatory role.

One such regulatory element (comA) was defined by a Tn917lac insertion [Albano et al. 1987; Hahn et al. 1987; Guillen et al. 1989; Weinrauch et al. 1989] and was shown to specify a protein essential for the expression of late competence genes. The DNA sequence of comA revealed that its product closely resembled members of the effector (or response regulator) family of bacterial signal-transduction proteins [for review, see Stock et al. 1989]. Effector proteins usually serve to activate the expression of a set of target genes in response to environmental signaling. In several cases, the activities of effector proteins are known to be regulated by phosphorylation and dephosphorylation. The amino-terminal moieties of effector proteins contain conserved amino acid sequences that include the aspartate residue(s) that
appear to serve as phosphorylation systems. A second component of these signaling systems, the sensor, is often membrane-associated. Sensors can be autophosphorylated at a conserved histidine residue and can transfer this phosphate to their cognate effector. The various sensor proteins possess a conserved carboxyl-terminal transmitter domain, which includes the conserved histidine residue.

The participation of the putative effector ComA in competence development suggested that a sensor protein might also be involved. _comB_ is a second regulatory gene required for the development of competence (Guillen et al. 1989; Weinrauch et al. 1989). Inactivation of _comB_ produces a phenotype identical to that of _comA_ mutants, and the two genes are closely linked on the _B. subtilis_ chromosome. However, the predicted amino acid sequence of ComB revealed no similarity to known sensor proteins. We have now identified an additional regulatory gene for competence, _comP_, which appears to specify a membrane-localized sensor protein and is involved in regulating sporulation as well as competence. Evidence is presented indicating that _comP_ and _comA_ work together in transmitting environmental signals to the competence regulon and that these signals include information concerning the presence of nitrogen and carbon sources.

Results

Disruption of the open reading frame upstream of _comA_

DNA sequence determination upstream from _comA_ revealed part of an open reading frame (ORF) with a potential stem-loop terminator overlapping the _comA_ promoter. To determine whether these DNA sequences were required for competence, we replaced the wild-type upstream ORF in the chromosome by mutant versions using double-crossover events. Plasmid DNA (pBD441 and pBD442) containing a _Cm^R_ cassette in two different orientations inserted in the cloned copy of the upstream ORF was linearized and used to transform a _com^+_ strain. The _Cm^R_ cassette was inserted in the right-most _BcII_ site, indicated below in Figure 2. The resulting _Cm^R_ transformants exhibited levels of transformation ~200- to 400-fold lower than that of the isogenic wild-type strain, demonstrating that the upstream ORF encoded a product required for genetic competence. This gene was named _comP_. The strains carrying disruptions in _comP_ in two different orientations were designated BD1658 and BD1659.

Dependence of late _com_ genes on _comP_

It was shown previously that the products of the sporulation genes, _spoOA_ and _spoOH_, as well as the early competence genes, _comA, comB_, and _sin_, were required for the expression of several late expressing _com_ genes (Albano et al. 1987; Guillen et al. 1989). We therefore wished to determine whether _comP_ was also required for the expression of late _com_ genes. DNA from strains carrying the _Em^R_ late _com-lacZ_ fusions, _comC530, comD413, comE518, comG12_, and two early fusions, _comB138_ and _comA124_, was used to transform strain BD1658, which contains the _Cm^R_ disruption of _comP_ in the chromosome. This was possible because the _comP_ mutants exhibited some residual transformability. _Cm^R_ _Em^R_ transformants were grown through the one-step competence regimen, and _β_-galactosidase activity was measured as a function of growth stage. Figure 1 shows that the expression of _comB138_ was unaffected by the _comP_ disruption, whereas the expression of _comA124_ was decreased about twofold after _T_0. This twofold decrease was also observed in the strain carrying the _Cm^R_ cassette in _comP_ in the reverse orientation. In contrast, a complete blockage in expression was observed for all the late _com_ genes. These results showed that _comP_ was required for expression of late competence genes and that the disruption also caused a relatively small decrease in _comA_ expression. It is interesting that this effect on _comA_ expression is only apparent for the later time points. It appears that expression of _comA_ has two components; an early one that is not _comP_-dependent, and a later _comP_-dependent component.

Nucleotide sequence of _comP_

Figure 2 shows a genetic and physical map of the _comB-degQ-comP-comA_ region of the _B. subtilis_ chromosome (Guillen et al. 1989; Nakano and Zuber 1989; Weinrauch et al. 1989). The 1.4-kb _HindIII_ fragment that contained the _comA_ gene was sequenced previously (Dubnau 1989; Weinrauch et al. 1989) and found to contain part of _comP_. We isolated a 3-kb _EcoV-AsuII_ fragment from pED3, which included the entire _comP_ gene, and determined its nucleotide sequence on both strands. The sequence revealed one major ORF (Fig. 3), which extended from position 335 to a TAA stop codon at position 2642, encoding a protein of 749 amino acids with a deduced molecular weight of 86,763. The TTG start codon was preceded by the potential Shine-Dalgarno (Shine and Dalgarno 1974) sequence AGGTGGA.

The predicted amino acid sequence of the ComP protein was compared to the translated GenBank data base. Similarity was found to members of the sensor class of bacterial signal-transduction proteins. Figure 4 shows a comparison of ComP to the most closely related of these sensor proteins. All of the residues noted by Stock et al. (1989) to be completely conserved in the other proteins of this class are also conserved in ComP. Also noteworthy is the striking series of hydrophobic segments in the amino-terminal half of the predicted amino acid sequence (Fig. 5).

Transcription mapping

To localize the 3' end of the _comP_ transcript(s), a 1.8-kb _XmalIII-SstI_ probe was 3'-end-labeled at the _XmalIII_ site (Fig. 2). The probe was hybridized to RNA isolated from an isogenic _com^+_ strain (BD630) at _T_ -2, _T_ -1, and _T_ 0, and from a _com^+_ strain carrying _comP_ and _comA_ on a multicopy plasmid, at _T_ -1. The results are shown in Figure...
Weinrauch et al.

Figure 1. Epistatic effects of comP mutation on expression of com genes. comP was inactivated by insertion of a Cm\(^R\) cassette, as described in Methods. The effects of this insertion on expression of lacZ transcriptional fusion to the indicated com genes are shown. The fusions used were comA124, comB138, comC530, comD413, comE518, and comG12 (Albano et al. 1987). (O) com\(^+\); (■) comP disruption (orientation 1); (■) comP disruption (orientation 2). The time points refer to hours before or after the departure from exponential growth (defined as T\(_0\)).

6A. The comP mRNA appeared to have a major 3' terminus at a site ~450 residues downstream from the labeled XmaIII site. This would place the major comP terminator approximately at the position of a dyad symmetry element that overlaps the comA promoter (Fig. 3). Because the XmaIII–SstI probe encompassed the comA gene, together with the putative comA terminator (Weinrauch et al. 1989), we would expect any readthrough past the major comP termination site to generate a transcript of 1.6 kb (Fig. 2). In fact, such a potential readthrough transcript was observed in the T\(_{-1}\) and T\(_{1}\) samples, but not at T\(_{-2}\). These results suggested that readthrough transcription of the comA transcript, possibly from the comP promoter, may occur beginning

Figure 2. Physical and genetic map of the comP region of the B. subtilis chromosome. The locations of the genetic determinants are from Guillen et al. [1989] and Nakano and Zuber [1989]. The position of promoters and transcriptional terminators are indicated. A Cm\(^R\) chloramphenicol acetyltransferase (CAT) cassette was inserted in a BclI site in both orientations, as indicated. (BH) BamHI; (H) HindIII; (EV) EcoRV; (B) BclI; (X) XmaIII; (A) AsuII; (S) SstI.

862 GENES & DEVELOPMENT
Competence regulatory gene in B. subtilis

Figure 3. (Continued on following page.)
Figure 3. (Continued.)
at some time after $T_2$. The data in Figure 1 are consistent with this idea: Disruption of $comP$ by insertion of a Cm$^R$ cassette prevented the rise in $comA$ expression usually noted after about $T_{-2}$.

Previous attempts to localize the $comA$ start site by primer extension had failed, presumably due to the low abundance of $comA$ mRNA (Weinrauch et al. 1989). We therefore repeated these experiments, using $T_1$ RNA from a strain carrying $comA$ in multicopy (Fig. 6B). This attempt was successful, localizing the 5' termini (Fig. 3) of the $comA$ transcript within 2 bases of the approximate positions identified by S1 nuclease mapping in our previous publication (Weinrauch et al. 1989).

Complementation of the $comP$ defect by $comA$ in multicopy

A 3-kb EcoRV–AsuII fragment carrying $comP$ and a 1.4-kb HindIII fragment carrying $comA$ were each cloned into pBD445 to create pBD457 and pBD458, respectively. The vector plasmid that was derived from pIM13 (Monod et al. 1986; Projan et al. 1987) had a copy number of $\sim$150-200 and conferred resistance to kanamycin. We reported previously that a similar multicopy plasmid carrying $comA$ multicopy restored transformation to wild-type levels in multicopy have been reported to complement sensor defects (Weston and Kadner 1987; Slauch et al. 1988). The ability of the multicopy $comA$ strain to become competent in the absence of the $comP$ product provides an opportunity to explore further the role of the latter protein. If the $comP$ product is a sensor that transmits information to ComA, then the regulation of competence in the complemented strains may differ in some way from that in the wild type. This difference might offer a clue as to the nature of the signals normally detected by ComP. We have reported that the expression of late competence genes is repressed by the addition of glutamine and by the substitution of glycerol for glucose in competence medium (Albano et al. 1987). The data in Figure 7A confirm these results by using $comG12$, a transcriptional Tn917lacZ fusion to $comG$. Figure 7B shows the effects of the multicopy $comA$ plasmid. Two changes are apparent: First, the repressing effect of glutamine was reversed; second, the effect of glycerol substitution for glucose was largely reversed by the multicopy $comA$ plasmid. These observations will be discussed further below.

Sporulation phenotypes of $comP$ and $spolIJ$ mutants

We observed that the $comP$ disruption mutant grown on solid medium tryptose blood agar base (TBAB) appeared to be oligosporogenic, judging by colony morphology. In contrast, the $comA$ mutant appeared to be Spo$^+$. Quantitative spore tests, using heat resistance as a measure of sporulation, has confirmed these impressions, using a set of isogenic strains; the $comP$ mutant sporulated 10-20% as well as the wild type in nutrient sporulation medium (NSM), although it seemed to approach the wild-type level of sporulation upon more prolonged incubation (Fig. 8). This behavior was similar to that of the oligosporogenic $spolIJ$ mutant, which was depressed $\sim$25-fold in sporulation frequency in our genetic background (Fig. 8; Perego et al. 1989; Antoniewski et al. 1990). $spolIJ$ encodes a regulatory protein for sporulation that exhibits amino acid sequence similarity to the family of bacterial signal-transduction sensors (P. Stra-
Weinrauch et al.

Figure 4. Similarity of ComP to other sensor proteins. Amino acid residues identical to those in ComP are shaded. The positions of residues reported to be universally conserved in this family of proteins are indicated with asterisks (Stock et al. 1989). The asparagine residue at position 358 in FixL was depicted by Stock et al. (1989), as aligned with the conserved asparagine residues 8 positions downstream in the other proteins. The alignment is as plausible because it improves the fit near the conserved histidine residue 4 positions downstream from the conserved asparagine. Similarly, the G residue at position 477 in UhpB was depicted by Stock et al. (1989), as aligned with the conserved G residues 15 positions downstream in the other proteins. The alignments were obtained with FASTP (Lipman and Pearson 1985), followed by several minor manual adjustments.

gier, unpubl.; Perergo et al. 1989, Antoniewski et al. 1990). Interestingly, double mutants carrying both a Tn917 insertion inactivating spoIII and a comP disruption sporulated only ~0.1% as well as the wild-type strain (Fig. 8). Thus, it appears that comP is a regulatory gene for sporulation, as well as for competence. A strain carrying a Tn917 disruption of spoII is fully competent, and this mutation has no discernible epistatic effect on expression of the late competence genes (data not shown). The comP spoIII double mutant is no less competent than the comP mutant itself (data not shown). Thus, there is no evidence that spoIII plays a role in regulating competence.

Discussion

We have identified and partially characterized a new regulatory gene for both competence and sporulation that resembles sensor members of bacterial signal-transduction systems in its carboxy-terminal amino acid sequence. The similarity to this class of proteins is most striking in the cases of degS (Henner et al. 1988; Kunst et al. 1988), narX (Nohno et al. 1989), and uhpB (Friedrich and Kadner 1987). Applying the RDF2 program (Pearson and Lipman 1988) to test the significance of these comparisons yielded the highly significant values of 14.7, 12.6, and 12.5 SD, respectively, for the comparison of the ComP sequence with those of DegS, NarX, and UhpB. In addition, the several amino acid residues identified by Stock et al. (1989), as conserved in all known sensor proteins, appear to be present in ComP.
Dubnau, unpubl.), narX and uhpB are Escherichia coli regulatory genes for nitrate utilization and for sugar phosphate transport, respectively.

The hydrophobicity profile (Fig. 5) implies that ComP is organized in two domains: A large relatively hydrophobic amino-terminal moiety with several potential membrane-spanning segments, and a hydrophilic carboxy-terminal domain. These each comprise about half the protein. Similarity with other sensors is confined to the carboxy-terminal portion of ComP. DegS does not appear to possess a hydrophobic amino terminus and is a smaller protein than ComP. UhpB and NarX each contain hydrophobic amino termini. In the case of NarX, two potential membrane-spanning segments have been identified [Nohno et al. 1989]. This appears to be the case as well for the sensor proteins VirA, CpxA, and EnvZ, in which such inferences from amino acid sequences concerning membrane topology have been supported by biochemical and genetic data [Forst et al. 1987; Weber and Silverman 1988; Melchers et al. 1989]. In these proteins, the residues between the two membrane-spanning segments are believed to be external to the cell membrane. The UhpB protein has also been found to be membrane-localized [Weston and Kadner 1987], but UhpB [Friedrich and Kadner 1987], like ComP, is somewhat atypical in that more than two potential membrane-spanning segments appear to be present. The membrane topologies of these related proteins are likely to be more complex than those envisaged in the case of sensors with only two hydrophobic segments.

In Figure 9, we propose a working model for the topology of ComP. We have based this model on several arguments. First, the hydrophobicity profile (Fig. 5) has identified eight potential transmembrane segments of appropriate length (~20 residues) to span a typical bilayer. This conclusion was reached using the criterion of Kyte and Doolittle [1982] that the mean hydrophobicity must exceed 1.6 over about a 20-residue stretch. Segments IV and VIII exhibited somewhat marginal hydrophobicities by this criterion, as shown. The eight segments were also identified as membrane-spanning by the statistically based method of Klein et al. [1985] and by the method of Engelman et al. [1986]. Second, the procedure of Garnier et al. [1978] predicted locations for turn sequences exclusively outside the eight potential transmembrane segments (Fig. 5). Third, it has been pointed out that individual transmembrane segments tend to be organized with an excess of positive charges on the cytoplasmic side of the cell membrane [von Heijne 1986, 1989]. As shown in Figure 8, this asymmetry is respected by our model in all cases except for
Table 1. *Complementation by comA and comP in multicopy*

| Chromosomal genotype | Plasmid genotype | Leu⁺ transformation* |
|----------------------|------------------|----------------------|
| com⁺                 | —                | 1                    |
| comP                 | —                | 4 × 10⁻³             |
| comP comA            | —                | 3 × 10⁻³             |
| comP comP comA       | comP⁺            | 1.6                  |
| comP comA comA       | comA⁺            | 1.5 × 10⁻²           |

*Expressed as Leu⁺ transformants per colony-forming unit, normalized to the wild-type (com⁺) strain.

Although the suggested topology is certainly not correct in all details, we favor a model in which several hydrophobic segments serve to anchor ComP in the membrane, with its conserved carboxy-terminal domain located in the cytoplasm.

Various findings suggested strongly that the products of comP and comA (and possibly comB) operate as members of a single regulatory pathway. First, inactivation of any one of these genes, or of several different pairs of them, resulted in a several hundredfold decrease in transformability (Guillen et al. 1989, and Y. Weinrauch and D. Dubnau, unpubl.). Because the residual competence is roughly the same in these various mutants, it appears likely that the three genes do not operate in independent pathways. Second, the ability of overexpressed ComA protein not only to complement comA but also comP defects strongly implies that the comA protein functions at a later point than ComP and in the same pathway. Finally, the similarity of these two proteins to known effector (response regulator) and sensor (signal-transducing) proteins is consistent with the idea that ComP operates prior to ComA. We therefore suggest the following working hypothesis, based on similar suggestions for other two-component signal-transduction systems (for review, see Stock et al. 1989). ComP is an integral membrane protein that senses an extracellular signal, using its external or membrane-spanning moieties. Information is then relayed to the ComP cytosolic carboxy-terminal domain, via a conformational change that alters the activity of ComP-associated kinase. This, in turn, alters the phosphorylation state of ComA, activating this protein. We have argued elsewhere (Weinrauch et al. 1989) that ComA is a transcriptional activator, capable of increasing the transcription rate of one or more competence determinants. The ComA targets are unknown.

Our evidence implies that comA expression, itself, may be subject to regulation, at least partly at the level of transcriptional readthrough (Figs. 1 and 6). We have attempted to map the 5' end of the comP transcript by S1 nuclease and primer extension experiments (data not shown). Although these experiments were ambiguous with respect to the start site of comP transcription, they did suggest that the latter decreased after To. In spite of these potential complexities of unknown significance, transcription of both comP and of comA occurs during vegetative growth and, in the case of comA at least, in

Figure 7. Effects of ComA overexpression on expression of a transcriptional comG–lacZ fusion in various media. The effects of glutamine and glycerol substitution for glucose on comG12–lacZ expression are shown in the absence [A] and presence [B] of overproduced ComA protein. [A] Competence medium, [B] competence medium plus glutamine, [□] competence medium with glycerol and no glucose added.

Figure 8. Sporulation of isogenic wild-type (■), comP (△), spolII (○), and double mutant comP spolII (□) strains in NSM. The time scale refers to hours after the departure from exponential growth [T₀]. The dashed lines present total colony-forming units; the solid lines present heat-resistant colony-forming units.
Competence regulatory gene in *B. subtilis*

The data in Figure 7 provide a hint as to the nature of the extracellular signals detected by ComP. When the *comP* defect was complemented by overproduction of ComA, expression of a late competence gene was no longer inhibited by an excess of glutamine or by substitution of glycerol for glucose. This suggested that ComP may detect the level of nitrogen and carbon sources in the medium. It does not necessarily suggest that the nutritional status of the medium is directly sensed by the ComP protein. For instance, a molecule may be produced in response to changes in nutrient availability and may, in turn, bind to ComP, providing an essential signal to initiate competence. Although the overproduction of ComA bypasses regulation by glutamine and glycerol substitution, it does not completely bypass the nutritional control of competence. For instance, the strain carrying *comA* in multicopy does not express β-galactosidase from the *comG12* fusion in complex media (data not shown).

The involvement of *comP* in sporulation is of particular interest. The much larger effect of the *comP spoIIJ* double mutation on sporulation frequency than that of either mutation by itself suggests that the two sensors may ordinarily be able to substitute for one another. Perhaps both ComP and SpoIIJ are capable of participating in a sporulation-specific signal-transduction pathway, possibly involving the effector proteins SpoOA or SpoOF or both. The regulatory interrelationships among these [and other] factors promise to be quite complex. For instance, the products of *spoOE, spoOB*, and *spoOF* are involved in activating *spoOA* for sporulation [Hoch et al. 1985] but apparently not for competence [Albano et al. 1987]. Although ComA is required for competence, probably as a target for activation by ComP, it is not required for sporulation, at least under the conditions tested. Finally, it is worth noting that an additional pair of signal-transduction proteins, *degS* and *degU*, is involved in regulating competence as well as sporulation [Henner et al. 1988; Tanaka and Kawata 1988; Msadek et al. 1990; J. Hahn and D. Dubnau, unpubl.]. How are we to understand these complex requirements? The possible responses of *B. subtilis* to the onset of stationary phase are manifold: Sporulation, competence, motility, degradative enzyme production, surfactin synthesis, etc. Precisely which global system[s] will be derepressed in a given instance may depend critically on conditions. The presence of a certain constellation of signals may satisfy the requirements for initiation of competence, another for sporulation, and a third for both degradative enzyme production and sporulation. The several signal-transduction pathways may serve to funnel this information to suitable processing systems, resulting in the activation of appropriate responses.
Methods

Strains and plasmids

All *B. subtilis* strains were isogenic with BD630 [leu-8 metB5 hisH2]. *E. coli* JM109 was used as a host for recombinant M13 phage. The plasmid pBD439 has been described previously [Weinrauch et al. 1989]. It carries the entire *comA* gene, as well as 400 bp of upstream sequences containing the 5' end of the *comP* gene. A chloramphenicol resistance [Cm*] cassette was isolated on a 1.3-kb SalIIIA fragment from pED1 [Guillen et al. 1989] and inserted in two different orientations into one of the two BclI sites of pBD439 to inactivate the *com* ORF. The resultant plasmids pBD441 and pBD442 were then tested by restriction digest analysis to confirm the location of the Cm R cassette in the proper BclI site. A third plasmid, pBD444, resulted from the removal of the entire 1-kb BclI fragment containing the *comA* gene, part of the *comP* ORF, and the 5' end of a reading frame downstream from *comA* and known not to be required for competence [Weinrauch et al. 1989]. This 1-kb fragment was replaced by the Cm* cassette described above. Plasmid pBD445 was constructed using a BamHI–TaqI fragment containing the kanamycin resistance [Km*] determinant from pUB110 ligated to the *MboI–ClaI* fragment from pMI3 [Monod et al. 1986]. It therefore confers only Km* and replicates using the high-copy-number pIM13 replicon. A 4.5-kb EcoRV fragment, isolated from pED3 [Weinrauch et al. 1989] and containing *comP* and *comA*, was inserted into HindIII–MboI-cleaved pBD347 after filling in with the Klenow fragment of DNA polymerase I to give pBD459. pBD347 is a derivative of pMI3 in which the erythromycin resistance [Em*] determinant has been replaced by a Cm* gene from pC194 [Monod et al. 1986]. The strain, PY390, carrying a Tn917 insertion in the proper *comA* gene, as well as the *comP* gene, was then used as an insert to the *comA* gene. The strain, PY390, carrying a Tn917 insertion in the proper *comA* gene, was then used as an insert to the *comA* gene.

Media

Solid media for *B. subtilis* were TBAB [Difco Laboratories] and minimal medium [Anagnostopoulos and Spizizen 1961] with appropriate growth supplements. Chloramphenicol, kanamycin, or erythromycin (5 μg/ml) were added as required. Liquid media used for *B. subtilis* were VY [25 grams of veal infusion [Difco], 5 grams of yeast extract [Difco], 1000 ml of water] for growth of phage and recipient cultures in transduction experiments, and competence medium [Contente and Dubnau 1979b] for growth to competence. When added as supplements, amino acids were used at concentrations of 50 μg/ml. Liquid media used for *E. coli* experiments were LB broth or 2× yeast tryptone (2× YT) [Miller 1972]. Solid medium consisted of LB agar, supplemented with ampicillin (100 μg/ml) or X-gal (40 μg/ml), as required. For measurement of sporulation, strains were grown in NSM [Schaeffer et al. 1965].

Transformation and isolation of plasmid DNA

*B. subtilis* strains were screened for plasmid by the rapid plasmid isolation procedure described by Contente and Dubnau [1979a]. Large-scale preparation of *B. subtilis* plasmid DNA was by SDS–NaCl treatment of plasmid-bearing strains, as described by Guerry et al. [1973], followed by CsCl-ethidium bromide purification. Growth of cultures to competence and transformation was performed as described previously, by the one-step procedure [Albano et al. 1987].

Determination of sporulation frequency

Strains were grown for the indicated times in liquid NSM with shaking at 37°C. The inocula for these cultures were standardized so that equal numbers of exponentially growing cells were inoculated into each culture. The numbers of heat-resistant and total colony-forming units were measured by plating on TBAB agar, with and without incubation of the serial dilutions at 80°C for 30 min before plating.

SI nuclease protection and primer extension mapping

RNA was prepared from the wild-type *com* strain [BD630], which was grown in competence medium. Cells were harvested 1 and 2 hr before *T*₀ [the time of transition from the exponential to the stationary growth phase], and at 1 hr after *T*₀. RNA was also prepared at *T*₂₀ from strain BD1785, which contains the *comA* and *comP* genes on a multicopy plasmid [pBD459]. RNA was extracted, as described by Ulmanen et al. [1985]. SI nuclease mapping was carried out, as described previously [Berk and Sharp 1977; Favaloro et al. 1980; Guillen et al. 1989]. For primer extension analysis, the primer 5'-CCGTTCCTC-AAAAATGGTC-3' was used. This primer was complementary to a region of *comA* mRNA 74 bases from the presumed start site of transcription [Weinrauch et al. 1989]. The primer was 5' end-labeled with polynucleotide kinase, as described [Weinrauch et al. 1989]. A total of 5 ng of labeled primer was mixed with 50 μg of total RNA from BD1785, which carries *comA* and *comP* on a multicopy plasmid. The RNA was dissolved in 16 μl of buffer containing 50 mM Tris [pH 8.0], 30 mM KCl, 8 mM MgCl₂, 1 mM dithiothreitol, and 1 μl of RNasin [Promega Biotech]. The mixtures were heated at 80°C for 10 min and cooled slowly to 42°C. A total of 4 μl of a mixture of four deoxynucleotides triphosphates [each at 10 mM] and 1 μl of avian myeloblastosis virus reverse transcriptase [Life Sciences, Inc., St. Petersburg, FL] were added to the annealed mixture, which were then incubated at 42°C for 30 min. Analysis of the extended products was carried out by electrophoresis in 6% polyacrylamide/urea gels.

DNA sequencing

Sequencing of DNA was by the dideoxynucleotide chain-termination method [Sanger et al. 1977] with [α-³²P]dATP [sp. act. 500 Ci/m mole; Dupont, NEN Research Products, Boston, MA]. Sequencing reactions were performed with modified T7 DNA polymerase kits (Sequenase, U.S. Biochemical Corp.). A 3-kb EcoRV–AsuII fragment isolated from pED3 was cloned into M13mp18 in both orientations. Two HindIII fragments of 1.4 and 0.5 kb from within the 3-kb EcoRV–AsuII region were also subcloned into M13mp18. The nucleotide sequence was determined completely on both strands and across restriction sites used for cloning, by using the M13 universal primer and 17-mer primers made from known sequences. The sequence of *comP* was submitted to The Genetic Sequence Data Bank [GenBank] with the accession number M29851.

β-Galactosidase determination

Samples were taken at the indicated times during growth in competence medium, and β-galactosidase activity was determined as described previously [Gryczan et al. 1984]. β-Galactosidase activities are expressed as units per milligram of protein.

Computer analysis

Protein sequences were compared to a translated version of the GenBank data base by using the TFASTA and FASTP programs [Lipman and Pearson 1985; Pearson and Lipman 1988]. The
RDF2 randomization program was used to test the significance of sequence comparisons, with 100 uniform shuffles (Pearson and Lipman 1988). The hydropathy profile of the ComP protein was determined by the use of the scales of Kyte and Doolittle (1982) and Engelman et al. (1986), with a moving window of 19 residues. The profile determined with the Kyte–Doolittle scale was further analyzed by the procedure of Klein et al. (1985) by using a window of 17 residues and the linear inequality given in their paper. The amino acid sequence of the predicted ComP protein was analyzed for turns by using the alpha program in the MBIR package of sequence analysis programs [Department of Cell Biology, Baylor College of Medicine, Houston, TX]. This program is based on the method of Garnier et al. (1978).

Acknowledgments

We acknowledge, with gratitude, valuable discussions with F. Breidt, R. Breitling, J. Hahn, S. Mohan, and M. Roggiani. We thank P. Youngman for his kind gift of PY390. This work was supported by NIH grant RRNA-02990 and by National Science Foundation grant DBM-8502189.

References

Albano, M., J. Hahn, and D. Dubnau. 1987. Expression of competence genes in Bacillus subtilis. J. Bacteriol. 169: 3110–3117.

Anagnostopoulos, C. and J. Spizizen. 1961. Requirements for transformation in Bacillus subtilis. J. Bacteriol. 81: 741–746.

Antoniwski, C., B. Savelli, and P. Stragier. 1990. The spoIIJ gene, which regulates early developmental steps in Bacillus subtilis, belongs to a class of environmentally responsive genes. J. Bacteriol. 172: 86–93.

Berk, A.J. and P.A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12: 721–732.

Contente, S. and D. Dubnau. 1979a. Marker rescue transformation by linear plasmid DNA in Bacillus subtilis. Plasmid 8: 355–357.

—. 1979b. Characterization of plasmid transformation in Bacillus subtilis. Kinetic properties and the effect of DNA conformation. Mol. Gen. Genet. 167: 251–258.

Dubnau, D. 1989. The competence regulon of Bacillus subtilis. In Regulation of prokaryotic development. (ed. I. Smith, R. Slepecky, and P. Setlow) pp. 147–166. American Society for Microbiology, Washington, D.C.

Engelman, D.M., T.A. Steitz, and A. Goldman. 1986. Identifying nonpolar transblayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. 15: 321–353.

Favaloro, J., R. Treisman, and R. Kamen. 1980. Transcription of membrane proteins. J. Biol. Chem. 255: 138–147.

Fasal, J., R. Treisman, and R. Kamen. 1980. Transcription maps of polyoma virus-specific RNA analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. 65: 718–749.

Forst, S., D. Comeau, S. Norko, and M. Inouye. 1987. Localization and membrane topology of enoZ, a protein involved in osmoregulation of ompF and ompC in Escherichia coli. J. Bacteriol. 162: 16433–16438.

Friedrich, M.J. and R.J. Kadner. 1987. Nucleotide sequence of the uhp region of Escherichia coli. J. Bacteriol. 169: 3556–3563.

Garnier, J., D.J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97–120.

Gryczan, T.J., M. Israeli-Reches, and D. Dubnau. 1984. Induction of macrolide-lincosamide-streptogramin B resistance requires ribosomes able to bind inducer. Mol. Gen. Genet. 194: 357–361.

Guerry, P., D.J. Leblanc, and S. Falkow. 1973. General method for the isolation of plasmid deoxyribonucleic acid. J. Bacteriol. 116: 1064–1066.

Guilhen, N., Y. Weinrauch, and D. Dubnau. 1989. Cloning and characterization of the regulatory Bacillus subtilis competence genes, comA and comB. J. Bacteriol. 171: 5354–5361.

Hadden, C. and E.W. Nester. 1968. Purification of competent cells in Bacillus subtilis transformation system. J. Bacteriol. 95: 876–885.

Hahn, J., M. Albano, and D. Dubnau. 1987. Isolation and characterization of competence mutants in Bacillus subtilis. J. Bacteriol. 169: 3104–3109.

Haseltine Cahn, F. and M.S. Fox. 1968. Fractionation of transformable bacteria from competent cultures of Bacillus subtilis on renograffin gradients. J. Bacteriol. 95: 867–875.

Henner, D.J., M. Yang, and E. Ferrari. 1988. Localization of Bacillus subtilis sacU(Hy) mutations to two linked genes with similarities to the conserved proacyrloic family of two-component signalling systems. J. Bacteriol. 170: 5102–5109.

Hoch, J.A., K. Trach, F. Kawamura, and H. Saito. 1985. Identification of the transcriptional suppressor sof-1 as an alteration in the spoDA protein. J. Bacteriol. 161: 552–555.

Klein, P., M. Kanehisa, and C. DeLisi. 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta 815: 468–476.

Kunst, F., M. Debarbouille, T. Msadek, M. Young, C. Mauel, D. Karamata, A. Klier, G. Rapoport, and R. Dedonder. 1988. Deduced polypeptide encoded by the Bacillus subtilis sacU locus share homology with two-component sensor-regulator systems. J. Bacteriol. 170: 5093–5101.

Kyte, J. and R.A. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105–132.

Lipman, D.J. and W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science 227: 1435–1441.

Mehler, L.S., T.J.G. Regensburg-Tuink, R.B. Bours, N.J.A. Sedee, R.A. Schilperoot, and P.J.J. Hooykaas. 1989. Membrane topology and functional analysis of the sensory protein VirA of Agrobacterium tumefaciens. EMBO J. 8: 1919–1925.

Miller, J.H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Monod, M., C. Denoya, and D. Dubnau. 1986. Sequence and properties of pM13, a macrolide-lincosamide-streptogramin B resistance plasmid from Bacillus subtilis. J. Bacteriol. 167: 138–147.

Msadek, T., F. Kunst, D. Henner, G. Rapoport, A. Klier, and R. Dedonder. 1990. The signal transduction pathway controlling the synthesis of a class of degradative enzymes in Bacillus subtilis: Expression of the regulatory genes and analysis of mutations in degS and degU. J. Bacteriol. 172: 824–834.

Nakano, M.M. and P. Zubo. 1989. Cloning and characterization of srfB: A regulatory gene involved in suractin production and competence in Bacillus subtilis. J. Bacteriol. 171: 5347–5353.

Nohto, T., S. Noji, S. Taniguchi, and T. Saito. 1989. The narX and narL genes encoding the nitrate-sensing regulators of
Escherichia coli are homologous to a family of prokaryotic two-component regulatory genes. *Nucleic Acids Res.* 17: 2947–2957.

Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* 85: 2444–2448.

Perego, M., S.P. Cole, D. Burbulys, K. Trach, and J.A. Hoch. 1989. Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins SpoOA and SpoOF of *Bacillus subtilis*. *J. Bacteriol.* 171: 6187–6196.

Projan, S.J., M. Monod, C.S. Narayanan, and D. Dubnau. 1987. Replication properties of pIM13, a naturally occurring plasmid found in *Bacillus subtilis* and its close relative pE5, a plasmid native to *Staphylococcus aureus*. *J. Bacteriol.* 169: 5131–5139.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74: 5463–5467.

Schaeffer, P., J. Millet, and J. Aubert. 1965. Catabolic repression of bacterial sporulation. *Proc. Natl. Acad. Sci.* 54: 704–711.

Shine, J. and L. Dalgarno. 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci.* 71: 1342–1346.

Slauch, J.M., S. Garrett, D.E. Jackson, and T.J. Silhavy. 1988. *envZ* functions through *ompR* to control porin gene expression in *Escherichia coli* K-12. *J. Bacteriol.* 170: 439–441.

Stock, J.B., A.J. Ninfa, and A.M. Stock. 1989. Protein phosphorylation and the regulation of adaptive responses in bacteria. *Microbiol. Rev.* 53: 450–490.

Tanaka, T. and M. Kawata. 1988. Cloning and characterization of *Bacillus subtilis* iep, which has positive and negative effects on production of extracellular proteases. *J. Bacteriol.* 170: 3593–3600.

Ulmanen, L., K. Lundstrom, P. Lehtovaara, M. Sarvas, M. Ruohonen, and I. Palva. 1985. Transcription and translation of foreign genes in *Bacillus subtilis* by the aid of a secretion vector. *J. Bacteriol.* 162: 176–182.

von Heijne, G. 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* 5: 3021–3027.

———. 1989. Control of topology and mode of assembly of a polytopic protein by positively charged residues. *Nature* 341: 456–458.

von Heijne, G. and Y. Gavel. 1988. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* 174: 671–678.

Weber, R.F. and P.M. Silverman. 1988. The Cpx proteins of *Escherichia coli* K-12. Structure of the CpxA polypeptide as an inner membrane component. *J. Mol. Biol.* 203: 467–478.

Weinrauch, Y., N. Guillen, and D. Dubnau. 1989. Sequence and transcription mapping of *comB* and of *comA*, a *Bacillus subtilis* competence gene that exhibits similarity to members of the two-component prokaryotic regulatory system. *J. Bacteriol.* 171: 5362–5375.

Weston, L.A. and R.J. Kadner. 1987. Identification of UhP polypeptides and evidence for their role in exogenous induction of the sugar phosphate transport system of *Escherichia coli* K-12. *J. Bacteriol.* 169: 3546–3555.
A Bacillus subtilis regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems.

Y Weinrauch, R Penchev, E Dubnau, et al.

Genes Dev. 1990, 4:
Access the most recent version at doi:10.1101/gad.4.5.860