JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 678–686
0021-9193/91/020678-09$02.00/0

Vol. 173, No. 2

The *Bacillus subtilis sin* Gene, a Regulator of Alternate Developmental Processes, Codes for a DNA-Binding Protein

NAND KISHORE GAUR,[1] JOEL OPPENHEIM,[2] AND ISSAR SMITH[1,2]*

Department of Microbiology, The Public Health Research Institute,[1] and Department of Microbiology, New York University School of Medicine,[2] New York, New York 10016

Received 24 September 1990/Accepted 14 November 1990

The *sin* gene of *Bacillus subtilis* encodes a dual-function regulatory protein, Sin, which is a negative as well as a positive regulator of alternate developmental processes that are induced at the end of vegetative growth in response to nutrient depletion. Sin has been purified to homogeneity by using a simple two-step procedure. It was found to bind to the developmentally regulated *aprE* (alkaline protease) gene at two sites in vitro. The stronger Sin-binding site (SBS-1) is located more than 200 bp upstream from the transcription start site. It is required for Sin repression of *aprE* expression in vivo, as strains bearing SBS-1 deletions were not affected by the *sin* gene. The second, weaker Sin-binding site lies on a DNA fragment that contains the *aprE* promoter. Results of DNase I, exonuclease III, and dimethyl sulfate footprinting analysis of SBS-1 suggested that Sin binding involves two adjacent binding sites which appear to contain two different partial dyad symmetries. An analysis of the predicted amino acid sequence of Sin revealed a potential leucine zipper protein dimerization motif which is flanked by two helix-turn-helix motifs that could be involved in recognizing two different dyad symmetries.

Bacillus subtilis is a soil bacterium that has evolved several adaptive mechanisms to survive under adverse environmental conditions. Among these are production of extracellular enzymes and antibiotics, acquisition of competence, and development of motility and sporulation, all of which require several unlinked genes for their successful completion (13, 22, 26). These developmentally regulated events appear only at the end of exponential growth, and there appear to be specific nutritional signals that induce them. For example, the presence of excess glucose prevents sporulation, whereas glucose is required for acquisition of competence, a natural ability of cells to take up DNA (4). The mechanism of global regulation that controls these alternate developmental pathways is not known, although many genes are involved. This regulatory network must involve sensing of changing nutritional conditions, and this information then should lead to turning on and turning off of appropriate genes. Several *spo0* genes and other genes involved in alternate developmental pathways are believed to be involved in sensing environmental changes. Some of these genes encode proteins that have amino acid homologies to the effector or sensor class of proteins of the two-component system (26).

Expression of the *aprE* gene, which encodes a major extracellular protease, subtilisin (also called an alkaline protease or extracellular serine protease), is closely associated with the onset of sporulation. Like *spo0* genes that are also required for initiation of sporulation, *aprE* expression is controlled by several regulatory genes, such as *degU* (11), *degQ* (1, 31), *degR* (28, 32), *hpr* (12, 21), *sen* (30), *abrB* (6), *pai* (14), and *sin* (8). Henner et al. (11) have identified target sites of *hpr*, *degQ*, and *degU* regulatory genes in the *aprE* gene by using several *B. subtilis* strains that have a series of promoter upstream deletions in *aprE*. These results suggest that the target sites for these genes lie relatively far upstream from the *aprE* promoter. The direct DNA binding of AbrB to the *aprE* promoter has been demonstrated by gel retardation assay and footprinting analysis (27).

The *sin* gene, previously cloned and characterized in our laboratory, is involved in controlling many late growth developmental processes (8). An inactivation of the *sin* gene (loss of function) results in loss of competence and motility. The effect on competence has now been demonstrated at the level of late *com* gene expression (10). In addition, the loss of function results in filamentous growth and a deficiency in autolysin production (8, 24). Sekiguchi et al. (25) have shown *sin* to be identical to the *B. subtilis flaD* locus. Elevated levels of Sin (gain of function) repress sporulation and the production of extracellular proteases, including subtilisin. Recent work has shown that in vivo, Sin represses expression of three sporulation genes, *spoIIE*, *spoIIG*, and *spoIIA*, and binds to at least the *spoIIA* promoter (17a).

As an initial step toward understanding the molecular mechanism of Sin regulatory functions, we describe in this report the purification of the protein, determination of its in vivo target site in the *aprE* gene, its in vitro binding to the *aprE* gene by gel retardation assay and footprinting analysis, and identification of two possible helix-turn-helix (H-T-H) DNA-binding motifs and a putative leucine zipper motif in the Sin amino acid sequence.

MATERIALS AND METHODS

Plasmid and strain construction. Plasmid pAP was a generous gift of Ed Chang. This plasmid has a 1,200-bp fragment of the *aprE* gene including the promoter cloned into pUC19 at the *Hin*dIII site. pIS226 was constructed by cloning a *Hae*III fragment from pAP containing the promoter into pUC19 at the *Sma*I site. To construct pIS169, the *Mbo*I fragment from pIS74 containing the wild-type *sin* gene and its promoter (P_3) was cloned in pBD9 (9) at the *Bam*HI site. Construction of pIS21 and pIS119, which contain the intact *sin* gene, has been described previously (8).

BG strains of *B. subtilis*, carrying an *aprE-lacZ* translational fusion and containing deletions of the region upstream

* Corresponding author.

from the *aprE* promoter, were generously provided by Dennis Henner and are described by Henner et al. (11). To construct a strain having a deletion and insertion in the *sin* gene, plasmid pIS141 was used. pIS141 was constructed by cloning a *Hpa*II fragment from pBD64, which contained the chloramphenicol acetyltransferase (*cat*) gene, into pIS21, which had been restricted with *Bal*I and *Nru*I, removing the internal *Bal*I-*Nru*I fragment from the *sin* gene. Linearized pIS141, which had both Cm^r and Em^r determinants, the latter from pIS21, was transformed into IS75, selecting for Cm^r colonies. Em^s and Cm^r transformants had the mutated *sin* gene which had replaced the wild-type locus. One of these transformants (IS432) is phenotypically similar to IS354, which had the *sin* gene disrupted by a Campbell-type event (8). The *cat* gene in IS432 remains stably integrated in the chromosome even after several passages without chloramphenicol.

Isolation and transformation of plasmid DNA in *B. subtilis* was done as described by Gryczan et al. (9).

β-Galactosidase assay. The β-galactosidase activity was determined as described previously (7), and activity is expressed in Miller units (19).

Purification of the Sin protein. *B. subtilis* IS404, which contains the *spo0AΔ204*, *trpC2*, and *pheA1* mutations and harboring pIS119, was grown in VY broth containing chloramphenicol (5 μg/ml) in a Lab-Line Hi-Density fermentor at 37°C. All of the following purification steps were performed at 4°C. Cells were pelleted by centrifugation from a mid-log-phase culture and were washed once with buffer B (10 mM Tris chloride [pH 8.0], 10 mM $MgCl_2$, 1 mM EDTA, 0.3 mM dithiothreitol, 5% glycerol, 1 mM phenylmethylsulfonyl fluoride, 0.3 M NaCl). A cell pellet (28 g) obtained from 10 liters of culture was suspended in 60 ml of buffer B and lysed by two passages through a French press at a pressure of 13,000 lb/in^2. The lysate was immediately centrifuged in a Sorvall centrifuge, using an SS34 rotor at 20,000 rpm for 1 h at 4°C. The supernatant solution was collected and is referred to as the crude extract. The crude extract was directly applied to a 12-ml heparin-agarose column that was previously equilibrated with buffer B, and the column was washed with 30 ml of buffer B. Proteins were eluted with 40 ml of buffer B containing a linear gradient of 0.3 to 0.8 M NaCl. Fractions of 2 ml were collected at a flow rate of 15 ml/h. Elution of Sin was monitored by running a 5-μl sample from alternate fractions on 15% sodium dodecyl sulfate (SDS)-polyacrylamide gels stained with Coomassie blue and by Western immunoblotting, and the Sin-containing fractions corresponding to 0.5 to 0.62 M NaCl were pooled. The pooled fraction was treated with 4 volumes of cold acetone, leaving the sample overnight at −20°C, and it was centrifuged at 12,000 rpm in an SS34 rotor for 30 min at 4°C. The acetone pellet, containing Sin, was dried and dissolved in 0.4 ml of buffer B containing 0.5% SDS. The resuspended pellet was directly applied to a Sephadex G-75 column (15 ml) preequilibrated with buffer B. The column was washed with 20 ml of buffer B, and 1-ml fractions were collected, maintaining a 10-ml/h flow rate. Alternate fractions were analyzed for the Sin protein by electrophoresis on 15% SDS-polyacrylamide gels stained with Coomassie blue and by Western blotting. The Sin-containing fractions were pooled and stored at −70°C in several aliquots. The purified protein did not show any loss of DNA-binding activity for more than a year.

Protein concentrations were determined by using Bio-Rad reagent. Bovine serum albumin (BSA; Sigma) served as a protein standard.

Preparation of anti-Sin antibodies and Western blotting. A 14-mer peptide corresponding to the C-terminal end of the Sin protein (FLDYQKWRKSQKEE) was synthesized by Bioresearch. The peptide was coupled to a carrier, keyhole limpet hemocyanin, via a cysteine residue that was attached to the N-terminal end of the peptide. An emulsified preparation of conjugated peptide (1 mg) in complete Freund adjuvant was injected subcutaneously in two female white rabbits. The rabbits were given a booster dose by injecting 1 mg of conjugated peptide in incomplete Freund adjuvant. This was repeated every week for three weeks. Blood was drawn from the ear 6 weeks after the first injection, and serum was prepared from it.

For Western blotting, extracts (prepared as described above) were subjected to electrophoresis on a 15% SDS-polyacrylamide gel, and proteins were transferred to a nitrocellulose membrane (29). The membrane was incubated with serum that had been preincubated with IS432 crude extract (20 μl of serum and 100 μl of extract) for 2 h, and detection was performed by using alkaline phosphatase-conjugated goat antibodies (Calbiochem) (2).

Gel retardation assay. *Not*I-restricted pAP plasmid DNA was 5′ end labeled as described previously (7) and then secondarily cut with *Dde*I or with both *Dde*I and *Ssp*I. The binding reaction was done in a 20-μl volume with increasing amounts of purified Sin protein in binding buffer which contained 10 mM Tris (pH 8.0), 6 mM mercaptoethanol, 1 mM EDTA, 100 nM NaCl, 2 μg of BSA, and 200 ng of sonicated calf thymus DNA. The reaction was initiated by addition of labeled DNA, and it was incubated at room temperature for 15 min. After addition of 5 μl of 50% glycerol, the samples were run immediately on 5% polyacrylamide gels in standard Tris-borate-EDTA (TBE) buffer at 100 V. The gel was prerun at 100 V overnight before application of samples. Loading dye containing 0.1% xylene cyanol and 0.1% bromophenol blue was run in a separate lane.

DNase I, dimethyl sulfate (DMS), and exonuclease III footprinting. Plasmid pAP was used to make both top-strand and bottom-strand probes. For the top-strand footprinting, a probe 5′ end labeled at the *Not*I site and secondarily cut at the *Hpa*I site was used. The bottom-strand probe was labeled at the *Dde*I site and secondarily cut with *Eco*RI. The probe (2 to 5 ng) was incubated at room temperature with different amounts of purified Sin in 100 μl of binding buffer. After 15 min of incubation, 6 to 7 μl of Promega DNase I (1 U/μl) was added and incubated at room temperature. After 1 min of incubation, 10 μl of DNase stop solution (0.2 M EDTA, 100 μg of tRNA per ml), 10 μl of 3 M sodium acetate (pH 7.0), and 500 μl of cold ethanol were added. The mixture was immediately placed on dry ice for 10 min, and the DNA was then collected by centrifugation. After drying, the DNA was suspended in 5 μl of formamide dye (0.1% xylene cyanol, 0.1% bromophenol blue, 10 mM disodium EDTA, 95% deionized formamide). Samples were electrophoresed onto 6% (for the top-strand probe) and 10% or 8% (for the bottom-strand probe) sequencing gels.

DMS methylation protection (footprinting) was performed in 200 μl of buffer containing 50 mM sodium cacodylate (pH 8.0), 1 mM EDTA, and 2 μg of sonicated calf thymus DNA. Binding of the probe (2 to 5 ng) in the binding buffer with different amounts of purified Sin was done at room temperature for 15 min. The DNA was methylated by addition of 1 μl of DMS and incubation at room temperature. After 1 min of incubation, 50 μl of DMS stop solution (1.5 M sodium acetate [pH 7.0], 1.0 M mercaptoethanol, 100 μg of yeast tRNA per ml), and 750 μl of cold ethanol were added. After

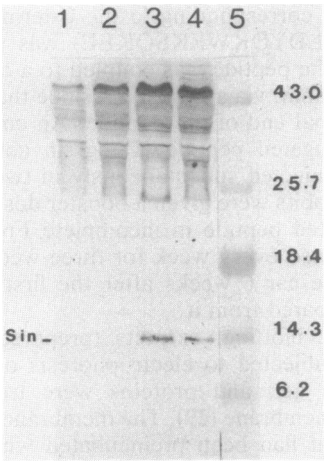

FIG. 1. Immunological detection of Sin protein. Western blot analysis of cell extracts with whole serum containing Sin antibodies was performed as described in Materials and Methods. Lanes: 1, crude extracts prepared from IS432, which is IS75 bearing a deletion in the *sin* gene (70 μg); 2, IS75, a wild-type strain carrying pC194 (106 μg); 3, IS75 harboring pIS119 (80 μg); 4, IS75 harboring pIS74 (99 μg); 5, prestained size markers (Bethesda Research Laboratories) (in descending order: ovalbumin, α-chymotrypsinogen, β-lactoglobulin, lysozyme, and bovine trypsin inhibitor). Sizes (in kilodaltons) are shown on the right.

the preparation was mixed and then kept on dry ice for 5 min, the DNA was pelleted by centrifugation for 5 min. The strand cleavage reaction with piperidine was done as described by Maxam and Gilbert (18). The samples were run on 6 or 10% sequencing gel as described for DNase footprinting.

Exonuclease III footprinting was done by forming the Sin-DNA probe complexes in 20 μl of binding buffer, as described above, in the gel retardation assay except that binding buffer also contained 10 mM $MgCl_2$. After incubation with 385 mM Sin protein for 15 min at room temperature, 875 U of exonuclease III was added and the reaction was continued at room temperature for 30 min. The reaction was terminated by addition of 20 μl of DNase stop solution, 5 μl of 3 M sodium acetate (pH 7.0), and 120 μl of cold ethanol, and the DNA was precipitated by centrifugation as described above. The samples were analyzed on 6% sequencing gel as described above. pIS226 plasmid DNA was used to make top- and bottom-strand probes. For the top-strand probe, DNA was 5′ end labeled at the *Bam*HI site and secondarily cut with *Eco*RI. The bottom-strand probe was made by 5′ end labeling at the *Dde*I site and secondarily cut with *Bam*HI.

RESULTS

Identification of the Sin protein. Sin, the protein product of the *sin* gene, was identified by using immunological methods. A hydropathy profile of the predicted amino acid sequence of the *sin* gene ORF2 revealed the C-terminal end to be the most hydrophilic. A 14-mer synthetic peptide from this region that was coupled to keyhole limpet hemocyanin was used to raise antibodies in rabbits. With use of whole serum, a specific protein band corresponding to the predicted molecular mass of Sin (13 kDa) was identified in a Western blot (Fig. 1). This band was observed in strains carrying multicopy plasmids that have the wild-type *sin* gene, pIS74 and pIS119. The intensity of the band is also related to the plasmid copy number. pIS119 is a higher-copy-number plasmid and has a more intense band than pIS74. The protein was not detected in a wild-type strain (IS75), a strain carrying a control plasmid (pC194), or a *sin* deletion strain (IS432) under these conditions. This result may be due to low levels of Sin in a wild-type strain that has one copy of the gene and is consistent with the very low β-galactosidase activity observed in a *sin-lacZ* translational fusion strain (7).

Purification of the Sin protein. We expected Sin to be a DNA-binding protein because the predicted amino acid sequence had a strong H-T-H motif in the N-terminal end (8). This motif is present in many regulatory proteins that bind to specific DNA sequences (20). Since many of these proteins have been purified by using heparin-based affinity columns, the ability of Sin to bind to heparin-agarose was tested. In a preliminary experiment, we observed that Sin binds strongly to heparin-agarose and could be eluted at about 0.5 M NaCl (results not shown). We also observed that Sin is extremely unstable when purified from a wild-type strain harboring pIS119, the high-copy-number plasmid with the wild-type *sin* gene. However, when the protein was purified from a *spo0A* strain, harboring pIS119, the protein was much more stable (unpublished results). We therefore purified Sin from a *spo0A* strain carrying pIS119. A crude extract prepared from this strain was passed through a heparin-agarose column, using a low salt concentration, and bound protein was eluted with a linear NaCl gradient as described in Materials and Methods. Sin was almost pure after the affinity column, as very few proteins eluted from heparin agarose in the part of the gradient displayed (Fig. 2A and B). A protein migrating at an apparent molecular mass of approximately 10 kDa was eluted at a higher salt concentration than Sin (Fig. 2A, fractions 32 to 34). It did not react with the Sin antibody, providing further proof for the specificity of the antibody for the Sin protein showing maximum elution in fractions 26 to 28 (Fig. 2B). Contaminating proteins were removed by passing the pooled heparin-agarose fractions containing the Sin protein through a Sephadex G-75 column. Figure 2C compares the heparin-agarose-purified Sin protein (lane 2) with Sephadex G-75-purified Sin (lane 3). The Sephadex G-75-purified protein was directly used for in vitro binding studies with the *aprE* promoter (see below). Sin exists as a tetramer in solution, as judged by native polyacrylamide gel electrophoresis and molecular sieve chromatography (results not shown).

In vivo determination of Sin target site in the *aprE* gene. It was previously observed that extracellular protease production at the end of vegetative growth was inhibited in a *B. subtilis* strain carrying a multicopy *sin* plasmid (8). The major extracellular protease induced at the end of vegetative growth is subtilisin, also known as serine protease or alkaline protease. However, these experiments could not distinguish between the inhibition of gene expression, protein processing, or secretion. We therefore used the *aprE* gene (the gene encoding alkaline protease) to study the regulatory function of Sin. Henner et al. have constructed a series of upstream promoter deletions in the *aprE* gene to determine the target sites of the *hpr*, *degU*, and *degQ* genes (11). In these strains, translational fusions between the *aprE* gene and the *Escherichia coli lacZ* gene at the eighth codon were integrated in the chromosome in single copy at the *amyE* locus. Since these fusions are in the *aprE* signal peptide, the fusion protein remains cytoplasmic. Initially, we used two strains, BG4057 and BG4160, to determine the effect of the multicopy *sin* plasmid (pIS21) on *aprE* expression. BG4057 has approximately 600-bp sequences upstream from the

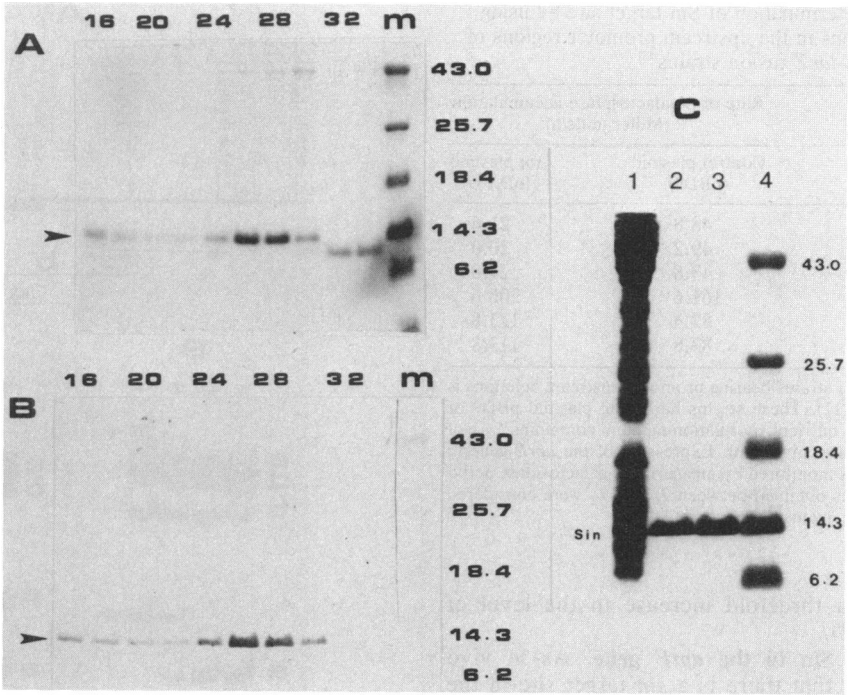

FIG. 2. Purification of Sin protein. A crude extract from *B. subtilis* IS404 carrying pIS119 was prepared by disruption in a French pressure cell. The extract, in a low-salt solution, was passed through a heparin-agarose column. Bound proteins were eluted with a linear 0.3 to 0.8 M NaCl gradient. (A) Alternate fractions of 5 μl were separated by electrophoresis on 15% polyacrylamide gels, and the proteins were visualized by staining with Coomassie blue. Fraction numbers are indicated above the lanes. Lane m shows the migration of prestained protein markers (Bethesda Research Laboratories). Size (in kilodaltons) are shown on the right. (B) As in panel A, but the separated proteins were electrophoretically transferred to a nitrocellulose filter, and Sin protein was visualized by Western immunoblotting. Fractions 24 to 29, eluting at NaCl concentrations of 0.5 to 0.6 M, were pooled, concentrated, and further purified on a Sephadex G-75 column. Peak fractions containing Sin were pooled and concentrated. (C) The crude extract and pooled fractions at different stages of purification were separated on 15% SDS-polyacrylamide gels and stained with Coomassie blue. Lanes: 1, crude extract (30 μg); 2, pooled heparin-agarose fraction (3 μg); 3, Sephadex G-75 fraction (2.4 μg); 4, prestained size markers (Bethesda Research Laboratories) (in descending order: ovalbumin, α-chymotrypsinogen, β-lactoglobulin, lysozyme, and bovine trypsin inhibitor). Sizes (in kilodaltons) are shown on the right.

aprE transcription start site, whereas BG4160 has a deletion up to the −52-bp region. The deletion itself had no effect on *aprE* expression (Fig. 3), as previously reported (11). However, the multicopy *sin* gene inhibited synthesis of β-galactosidase in BG4057, whereas synthesis in BG4160 remained relatively unaffected. This finding suggested that the target site for *sin* action, whether direct or indirect, is in the region upstream from the promoter. To further define the *sin* target site, we used other strains that have promoter upstream deletions terminating between the −600 and −52 regions. Strains carrying a multicopy *sin* plasmid (pIS169) showed less β-galactosidase synthesis than did strains carrying a control plasmid (pBD9) when the deletions terminated at or before the −340 region (Table 1). Strains having a deletion up to or downstream from the −244 region did not show inhibition of β-galactosidase synthesis by the presence of a multicopy *sin* plasmid. This result indicated that the sequence required for the *sin* effect should lie within the −340 and −244 region. It is interesting that accumulation of β-galactosidase in strains carrying the control plasmid (pBD9) with a single chromosomal copy of the *sin* gene was increased about threefold when the region between −340 to −244 was deleted. This result suggests that the *sin* gene exerts its repressive effect on *aprE* expression by directly or indirectly interacting with this region even when present in single copy. In agreement with this observation, we had previously noted that disruption of the chromosomal copy of

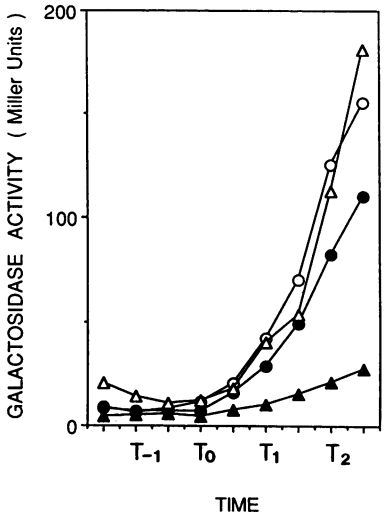

FIG. 3. Effect of multicopy *sin* gene on the expression of *aprE-lacZ* fusions. Expression by various strains of the *aprE* gene in single chromosomal copy was monitored by measuring β-galactosidase activity of *aprE-lacZ* translational fusions. Enzyme activity is expressed in Miller units. The growth time in hours is represented relative to the end of vegetative growth (T_0) in nutrient sporulation medium at 32°C. Symbols: Δ, BG4057 harboring pE194; ▲, BG4057 harboring pIS21; ○, BG4160 harboring pE194; ●, BG4160 harboring pIS21.

TABLE 1. In vivo determination of Sin target site by using strains that have deletions in the upstream promoter regions of *aprE-lacZ* fusion strains[a]

Strain	Deletion endpoint	Rate of β-galactosidase accumulation (Miller units/h)	
		Control plasmid (pBD9)	*sin* plasmid (pIS169)
BG4057	−600	48.8	21.6
BG4224	−412	49.2	20.0
BG4226	−340	43.8	21.0
BG4225	−244	161.6	206.6
BG4197	−200	82.4	123.8
BG4201	−164	83.8	113.8

[a] Construction of *B. subtilis* strains bearing promoter upstream deletions is described by Henner et al. (11). These strains harboring plasmid pBD9 or pIS169 were grown at 37°C in nutrient sporulation medium containing 5 μg of chloramphenicol and erythromycin per ml. Expression of the *aprE* gene in single chromosomal copy was monitored by measuring β-galactosidase activity at 30-min intervals. Values obtained between T_0 and T_2 were considered for calculating the rate of β-galactosidase accumulation.

the *sin* gene caused a threefold increase in the level of subtilisin production (8).

In vitro binding of Sin to the *aprE* gene. As in vivo experiments suggested that there is a *sin* target site in the *aprE* gene between the −340 and −244 region, we decided to examine the DNA-binding activity of purified Sin to the *aprE* gene. A plasmid (pAP) carrying a 1,200-bp fragment that contains the *aprE* promoter was used to determine the Sin-binding activity in a gel retardation assay (Fig. 4A). DNA fragment a, which contains the in vivo-determined *sin* target site, was specifically retarded in the presence of Sin (Fig. 4B). The extent of retardation was dependent on the concentration of added Sin. Another retarded band that showed very weak binding to Sin was also observed in the same gel. To identify the origin of this minor retarded band, the DNA was restricted with *Ssp*I, which cleaved only fragment b and gave rise to a smaller fragment, c (Fig. 4A). The minor retarded band was not observed when *Ssp*I-cut DNA was analyzed for Sin binding in a gel retardation assay (Fig. 4C). However, binding to fragment a remained unaffected under these conditions. These results imply that sin also binds to DNA fragment b somewhere between *Dde*I and *Ssp*I or that *Ssp*I is within the binding site. Sin binding to these two DNA fragments, a and b, was specific because the remaining two DNA fragments, d and e, were not retarded by added Sin under these conditions (Fig. 4B and C). Fragments d and e were not resolved from each other in these electrophoretic separations and correspond to junction fragments between the vector and the *aprE* insert. However, it must be emphasized that we were unable to characterize further the weak binding site on fragment b by any of the footprinting techniques described below, and it is not clear whether this weak interaction is specific. However, the gel retardation experiments were performed with vast excesses of heterologous DNA (up to 200-fold excess), which prevented the binding of Sin to the vector fragments (Fig. 4).

Other experiments indicate that Sin represses expression of *sin* gene and certain stage II sporulation genes (17a). The Sin protein binds to restricted promoter regions of these genes and not to other fragments present in the gel retardation assays, indicating the specific nature of the Sin-DNA interaction.

Footprint analysis of Sin binding to the *aprE* target site. To determine the exact location of the promoter-distal Sin-

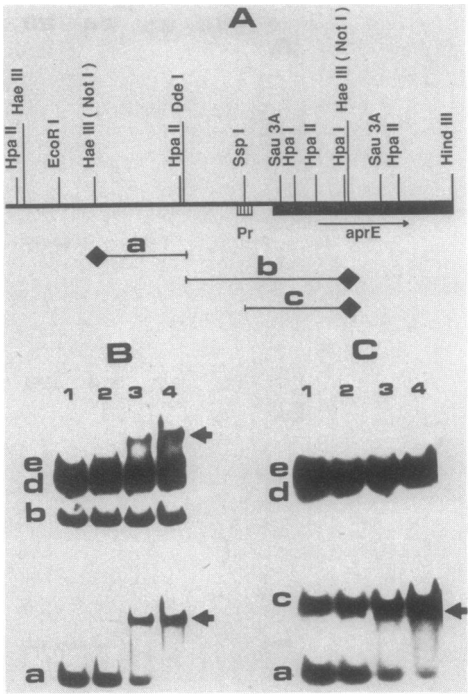

FIG. 4. Gel retardation assay. (A) Restriction map of a 1,200-bp fragment in plasmid pAP that contains the *aprE* promoter and part of the structural gene. DNA fragments a, b, and c are indicated below the map. DNA fragments d and e are junction fragments between the vector and the *aprE* insert. A diamond on a DNA fragment represents the labeled 5′ end. (B) Assay with *Dde*I-cut DNA. *Not*I-restricted pAP DNA was 5′ end labeled and then cut with *Dde*I. A constant amount of labeled DNA was allowed to bind with different amounts of Sin as described in Materials and Methods. Sin concentrations: lane 1, no protein; lane 2, 7.7 nM; lane 3, 30.8 nM; lane 4, 123.0 nM. An arrow indicates the retarded band. (C) Assay with *Dde*I-*Ssp*I-cut DNA. *Not*I-restricted pAP DNA was 5′ end labeled and then cut with *Dde*I and *Ssp*I. A constant amount of labeled DNA was allowed to bind with different amounts of Sin as described in Materials and Methods. Sin concentrations: lane 1, no protein; lane 2, 7.7 nM; lane 3, 30.8 nM; lane 4, 123.0 nM. An arrow indicates the retarded band.

binding site (SBS-1) in the *aprE* gene, we did various footprinting analyses. Sin protected a 46-bp region from −220 to −265 from DNase I cleavage when the labeled top-strand probe was used (Fig. 5A). Complete protection could be seen when Sin protein was present at a monomer concentration of 7.7 nM. This amount of protein is at least 10 times less than the amount required for complete binding in the gel retardation assay. Since binding conditions were identical in both assays, this DNA-protein complex is probably less stable when analyzed in the gel retardation assay. To further define the binding, DMS methylation protection of guanine residues was also analyzed after binding to Sin. DMS methylates guanine at the N-7 position in the major groove of the DNA, and a protein bound to it would prevent G methylation. The Sin protein protected some guanine residues in DMS footprinting of the labeled top-strand probe (Fig. 5B). These guanine residues correspond to positions −229, −240, −241, −250, and −262 and would lie on the same side of a B-DNA helix. Guanine residues that lie on the opposite side of the helix (at positions −233, −244, −253, and −254) were not protected from DMS methylation.

Similarly, with use of a bottom-strand probe, a 46-bp

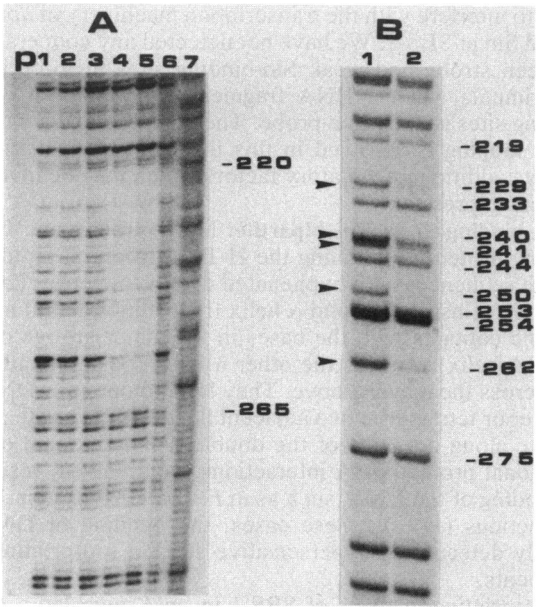

FIG. 5. DNase I and DMS footprint analysis for the top strand. (A) DNase I footprinting of the *Not*I 5′-end-labeled top-strand probe by using different amounts of Sin as described in Materials and Methods. Sin concentrations: lane 1, no protein, lane 2, 1.5 nM; lane 3, 3.0 nM; lane 4, 7.7 nM; lane 5, 30.8 nM; lane 6, no protein. Lane 7, G chemical cleavage ladder of the probe; lane P, DNase I-untreated probe. (B) DMS footprinting of the *Not*I 5′-end-labeled top-strand probe by using different amounts of Sin as described in Materials and Methods. Sin concentrations: lane 1, no protein; lane 2, 38.5 nM. An arrow indicates the partial or complete protection of a guanine residue at 38.5 nM Sin. Nucleotide positions relative to the transcriptional start site are shown on the right.

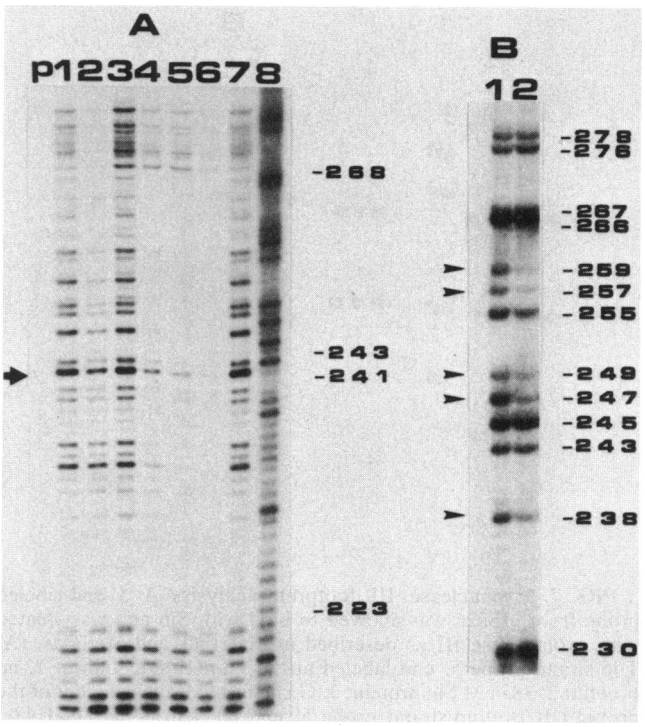

FIG. 6. DNase I and DMS footprint analysis for the bottom strand. (A) DNase I footprinting of the *Dde*I 5′-end-labeled bottom-strand probe by using different amounts of Sin as described in Materials and Methods. Sin concentrations: lane 1, no protein; lane 2, 1.5 nM; lane 3, 3.0 nM; lane 4, 7.7 nM; lane 5, 30.8 nM; lane 6, 61.6 nM; lane 7, no protein. Lane 8, G chemical cleavage ladder of the probe; lane P, DNase I-untreated probe. An arrow indicates the partially protected band. (B) DMS footprinting of the *Dde*I 5′-end-labeled bottom-strand probe by using different amounts of Sin as described in Materials and Methods. Sin concentrations: lane 1, no protein; lane 2, 38.5 nM. An arrow indicates the partial or complete protection of a guanine residue at 38.5 nM Sin. Nucleotide positions relative to the transcriptional start site are shown on the right.

region from −223 to −268 was protected by Sin from DNase I cleavage (Fig. 6A). A band at position −242 was partially protected from DNase I cleavage even at higher concentrations of Sin. The protection of some guanine residues from DMS methylation was also observed in the bottom strand after binding to the Sin protein (Fig. 6B). These protected guanine residues (−238, −247, −249, −257, and −259) lie on the same side of B-DNA helix, and unprotected guanine residues (−230, −243, −245, −255, −266, and −267) would lie on the other side of the DNA helix.

To further probe this unusually large Sin-binding site, exonuclease III footprinting was performed. This analysis allows the determination of multiple binding sites which have different relative affinities for a DNA-binding protein. Exonuclease III removes mononucleotides from the 3′ end of duplex DNA progressively, and this removal is impeded by a DNA-bound protein. Exonuclease III footprinting on the top-strand probe showed two major stop points corresponding to positions at −219 and −240 (Fig. 7A), whereas exonuclease III footprinting on the bottom-strand probe resulted in only one major stop at position −269 (Fig. 7B). These results imply that this *aprE* region has two adjacent Sin-binding sites, SBS-1A and SBS-1B (Fig. 8). Since two stop points were obtained only on the top-strand probe, these results also suggested that SBS-1B is weaker than SBS-1A. These positions of exonuclease III stop points are in good agreement with the boundaries defined by DNase I footprinting.

DISCUSSION

The *sin* gene encodes a DNA-binding protein that regulates expression of the *aprE* gene. We have identified a protein, Sin, encoded by a regulatory gene (*sin*) by using antibodies raised against a synthetic peptide. In confirmation of our earlier results obtained with gene fusion experiments (7), this protein appears to be in low abundance in wild-type *B. subtilis* strains. The *sin* gene has been implicated as a positive and negative regulator in the control of alternate developmental processes that are initiated at the end of exponential growth in response to nutrient depletion (26). Therefore, Sin may behave like a developmental switch protein that controls alternate developmental pathways in *B. subtilis*.

We have purified Sin protein to homogeneity, using a simple two-step procedure. The protein binds to the *aprE* gene at two sites. The expression of *aprE*, a gene whose expression is linked to the end of vegetative growth and the onset of sporulation, is negatively controlled by the *sin* gene. A strong Sin-binding site (SBS-1) is located on a DNA fragment that contains sequences from the −200 to −412 region relative to transcription start site. A promoter-containing DNA fragment binds weakly to Sin in a gel retardation assay. This fragment contains sequences from −200 to

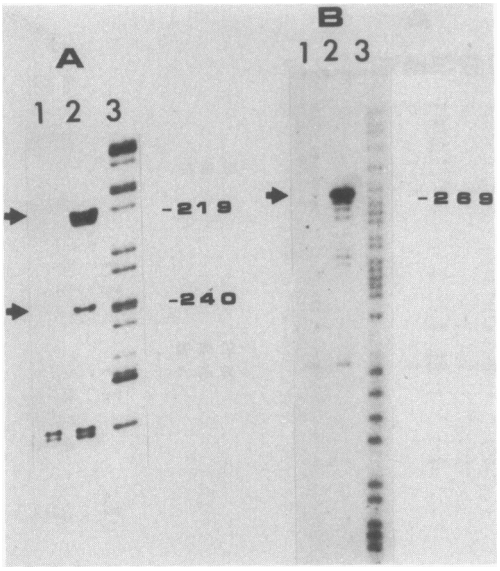

FIG. 7. Exonuclease III footprint analysis. A 5'-end-labeled probe from pIS226 was allowed to bind with Sin and was treated with exonuclease III as described in Materials and Methods. (A) Top-strand probe 5' end labeled at the *Bam*HI site. Lanes: 1, no protein; 2, 385 nM Sin protein; 3, G chemical cleavage ladder of the probe. (B) Bottom-strand probe 5' end labeled at the *Dde*I site. Lanes: 1, no protein; 2, 385 nM Sin protein; lane 3, A + G chemical cleavage ladder of the probe. The positions of the exonuclease III major stop points are indicated by arrows. Nucleotide positions relative to the transcriptional start site are shown on the right.

+281 relative to the transcription start site. However, we were unable to obtain any footprint on this binding site, and additional evidence for its existence is therefore needed.

The significance of a strong Sin-binding site (SBS-1) in regulation of *aprE* gene expression is supported by evidence obtained from in vivo experiments. A *B. subtilis* strain carrying a promoter upstream deletion that terminates at position −244 in the *aprE* gene is insensitive to Sin repression. These results imply that SBS-1 is essential in vivo for Sin repression of *aprE* expression, as the deletion would remove a part of SBS-1. It is too early to predict the mechanism based on these results that allows repression of *aprE* expression by Sin binding to a site about 200 bp upstream from the promoter. However, there must be a way

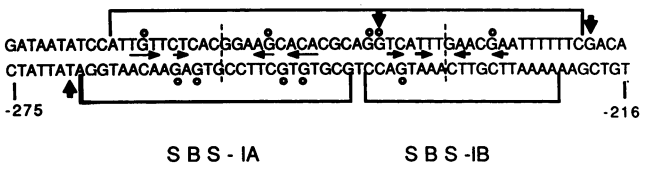

FIG. 8. Overview of Sin interactions with the *aprE* promoter distal binding site (SBS-1). The base pair numbers (−275 to −216) refer to positions relative to the transcriptional start site. Sequences protected from DNase I cleavage after binding to Sin are indicated by brackets. Guanine residues that are protected from DMS methylation after binding to Sin are indicated by circles. Thick arrows at positions −219 and −240 on the top strand and −269 on the bottom strand represent exonuclease III stop points after Sin binding. Arrows between the complementary DNA strands indicate the bases that make dyad symmetries, and the vertical broken line represents the center of dyad symmetry.

so as to interfere with the transcription machinery of *aprE* by bound Sin at SBS-1. We have not detected any cooperativity between strong and weak Sin-binding sites in footprinting experiments when a DNA fragment carrying both of the binding sites is used as a probe. Therefore, it appears that if DNA looping is involved in this mechanism, then it must involve additional regulatory factors which may be involved in *aprE* expression.

Sin binding to *aprE* is bipartite. Many procaryotic DNA-binding proteins containing the H-T-H structural motif recognize sequences with hyphenated dyad symmetries (20). In these proteins, the second α helix (recognition helix) makes specific contacts with the bases in the major groove of the B-DNA helix, whereas the other α helix (positioning helix) lies across the major groove. They bind to operator DNA as a dimer or tetramer in two adjacent helical turns of the major groove along one face of the double helix. In some cases, additional protein-DNA interactions are involved, resulting in bending of the DNA, such as in *E. coli* CAP protein-DNA interactions (17). In these cases, the bending of DNA is usually detected as hypersensitive sites in footprinting experiments.

Footprinting analysis of SBS-1 in *aprE* revealed a 46-bp region which was protected by Sin from DNase I cleavage in both top- and bottom-strand probes (Fig. 8). The protected region is unusually long compared with other proteins that bind to a single site on DNA. These sites usually protect from 20 to 25 bp (two helical turns) in a DNase I footprint. A 46-bp region would have four helical turns of the major groove of a B-DNA helix, indicating that this region might have two adjacent Sin-binding sites. A nucleotide (T) at position −242 may separate these two sites, as it is not completely protected in the bottom-strand DNase I footprint. Exonuclease III footprinting experiments demonstrated clearly that this region has two adjacent Sin-binding sites. The results obtained in DMS footprinting experiments suggested that guanine residues which are protected from methylation lie on the same face of the DNA. We never observed any hypersensitive site in DNase I or DMS footprinting experiments, suggesting that DNA is not grossly distorted upon Sin binding to this site. Guanine residues located on the other side of the DNA helix are not protected from methylation. This region appears to have two partial dyad symmetries (SBS-1A and SBS-1B) consisting of different DNA sequences. If these binding sites are different, as their DNA sequences suggest, then Sin binding to SBS-1A and SBS-1B may be mediated by two different DNA-binding domains or a single DNA-binding domain that is capable of recognizing different DNA sequences. Although Sin binding to *aprE* involves two adjacent binding sites, it remains to be seen whether it binds to other genes in a similar fashion.

Predicted structural motifs in Sin. The possibility of the existence of two DNA-binding motifs in Sin prompted us to search for a putative DNA-binding domain(s) in the predicted amino acid sequence. A potential H-T-H motif in this protein at the N-terminal end at amino acids 17 to 36 was previously reported (8). Computer analysis using Chou-Fasman (3) and hydrophobic moment (5) programs for secondary structures in Sin revealed that there could be another H-T-H motif at the C-terminal end at amino acid positions 81 to 100 (Fig. 9A). This motif has all the essential features of an ideal H-T-H structure except that it has phenylalanine in place of isoleucine or valine or leucine in the recognition helix. However, this phenylalanine should be capable of making the required hydrophobic interaction with the conserved alanine and could be a functionally equivalent

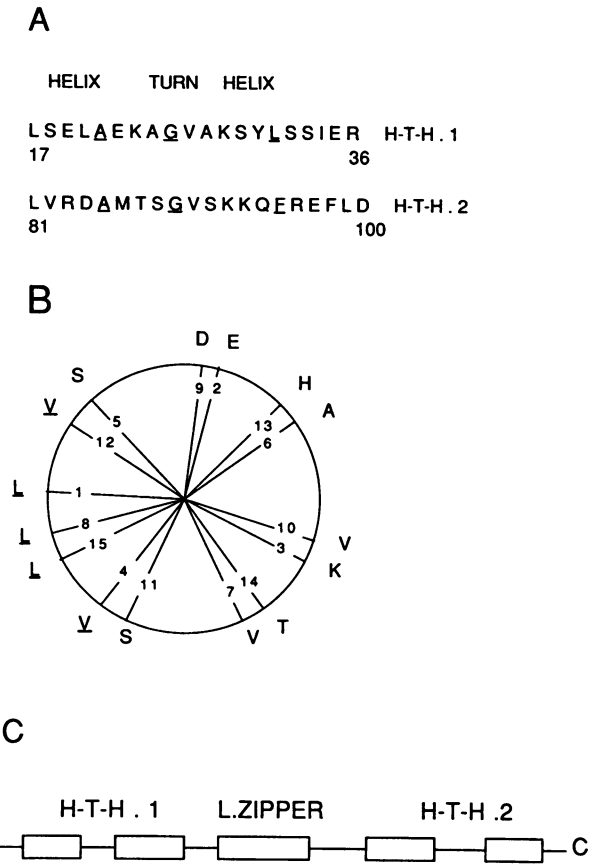

FIG. 9. Predicted structural motifs in Sin. (A) Amino acid sequence of two H-T-H motifs, H-T-H-1 and H-T-H-2, represented in single-letter code. Highly conserved amino acids are underlined; the positions of first and last amino acids are indicated below them. (B) Amino acid sequence from positions 47 to 61, displayed on a helical wheel (23). The hydrophobic residues that are located on one side of the helix are underlined. Note three leucine residues on this side of the helix that make heptad repeats. (C) Positions of structural motifs relative to each other.

replacement. It is conceivable that Sin binds to DNA by using these two H-T-H motifs.

Since Sin exists as a tetramer in solution, it must have at least two interfaces that allow it to have protein-protein interactions. An interface for protein dimerization, the leucine zipper, has been described in many eucaryotic proteins (16). It is an amphipathic α helix consisting of heptad repeats of leucine residues and mediates dimer formation between two parallel polypeptides, resulting in a coiled-coil structure. Sin has this domain, consisting of three leucine residues in heptad repeats in a region predicted to be an α helix. This helix has hydrophobic amino acid residues at one side and all hydrophilic residues on the opposite side when displayed on a helical wheel (Fig. 9B). This motif may provide one interface to make a Sin dimer. However, the nature of another interface that allows a Sin tetramer to be formed is not known. It is important to note that these predicted motifs are completely retained in the *Bacillus licheniformis sin* gene (25), suggesting an essential functional significance of these structural features.

Sin and other developmental regulatory proteins. As a soil bacterium, *B. subtilis* must be expected to have regulatory mechanisms that allow it to adjust rapidly to the changing environmental conditions by turning on or turning off appropriate developmental pathways. Sin, being a negative and positive regulator, may perform this function by switching its regulatory activities in response to environmental changes. The predicted structural motifs, two H-T-H and one leucine zipper, may play a role in the alternate functioning of Sin. AbrB, another *B. subtilis* regulatory protein, is a negative and a positive regulator of genes that are induced at the end of vegetative growth (27). The DNA-binding property of AbrB is intrinsically different from that of Sin. It involves protection of a relatively large region in the DNase I footprint (50 to 120 bp), and the protein appears to recognize some kind of structure in the DNA. A *B. subtilis* switch protein, 14 kDa, that alters the specificity of RNA polymerase containing σ^K is involved during the transition from morphological stages IV to V of sporulation (15). This protein enhances transcription of the *spoIVCB* gene and represses *cotD* transcription in vitro. Although it is clear that these proteins are dual-function regulators, the mechanism that allows them to switch regulatory activities is not known.

The confirmation of predicted Sin motifs (two H-T-H and one leucine zipper) must come from structural and mutational studies. The role of these motifs, if any, in switching the Sin regulatory properties from being a negative to a positive regulator at the end of vegetative growth should also provide an insight in the functioning of this protein. We will be addressing these questions in the future to define the role of Sin as a developmental switch.

ACKNOWLEDGMENTS

We thank Kettly Cabane for technical assistance and other members of the Smith group, Uma Bai, Jeannie Dubnau, Ines Mandic-Mulec, and Mima Predich, for critical and helpful discussions. We are also grateful to David Dubnau and members of his group for helpful discussions. Special thanks to Annabel Howard for expert secretarial assistance.

This work was supported by National Institutes of Health grant GM32651, awarded to I.S. Computer analysis was performed on a Vax 11/750 purchased with funds from National Science Foundation grant PCM-8313516, awarded to the Public Health Research Institute.

REFERENCES

1. **Amory, A. F., E. Kunst, A. Aubert, and G. Rapoport.** 1987. Characterization of the *sacQ* genes from *Bacillus licheniformis* and *Bacillus subtilis*. J. Bacteriol. **169**:324–333.
2. **Blake, M., K. Johnston, G. Russell-Jones, and E. Gotschlich.** 1984. A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on western blots. Anal. Biochem. **136**:175–179.
3. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:45–148.
4. **Dubnau, D.** 1989. The competence regulon of *Bacillus subtilis*, p. 147–166. *In* I. Smith, R. Slepecky, and P. Setlow (ed.) Regulation of procaryotic development. American Society for Microbiology, Washington, D.C.
5. **Eisenberg, D.** 1984. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. **53**:595–623.
6. **Ferrari, E., D. J. Henner, M. Perego, and J. A. Hoch.** 1988. Transcription of *Bacillus subtilis* subtilisin and expression of subtilisin in sporulation mutants. J. Bacteriol. **170**:289–295.
7. **Gaur, N. K., K. Cabane, and I. Smith.** 1988. Structure and expression of the *Bacillus subtilis sin* operon. J. Bacteriol. **170**:1046–1053.
8. **Gaur, N. K., E. Dubnau, and I. Smith.** 1986. Characterization of a cloned *Bacillus subtilis* gene that inhibits sporulation in multiple copies. J. Bacteriol. **168**:860–869.
9. **Gryczan, T. J., S. Contente, and D. Dubnau.** 1978. Characteri-

zation of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. J. Bacteriol. **134**:318–329.
10. **Guillen, N., Y. Weinrauch, and D. A. Dubnau.** 1989. Cloning and characterization of the regulatory *Bacillus subtilis* competence genes *comA* and *comB*. J. Bacteriol. **171**:5354–5361.
11. **Henner, D. J., E. Ferrari, M. Perego, and J. A. Hoch.** 1988. Location of the targets of the *hpr-97*, *sacU32*(Hy), and *sacQ36*(Hy) mutations in upstream regions of the subtilisin promoter. J. Bacteriol. **170**:296–300.
12. **Hiegerd, T. B., J. A. Hoch, and J. Spizizen.** 1972. Hyperprotease-producing mutants of *Bacillus subtilis*. J. Bacteriol. **112**: 1026–1028.
13. **Hoch, J. A.** 1976. Genetics of bacterial sporulation. Adv. Genet. **18**:69–99.
14. **Honjo, M., A. Nakayama, K. Fukazawa, K. Kawamura, K. Ando, M. Hori, and Y. Furutani.** 1990. A novel *Bacillus subtilis* gene involved in negative control of sporulation and degradative-enzyme production. J. Bacteriol. **172**:1783–1790.
15. **Kroos, L., B. Kunkel, and R. Losick.** 1989. Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. Science **243**:526–528.
16. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
17. **Liu-Johnson, H., M. R. Gartenberg, and D. M. Crothers.** 1986. The DNA binding domain and bending angle of E. coli CAP protein. Cell **47**:995–1005.
17a. **Mandic-Mulec, I., N. Gaur, and I. Smith.** Unpublished data.
18. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**: 499–560.
19. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
21. **Perego, M., and J. A. Hoch.** 1988. Sequence analysis and regulation of the *hpr* locus, a regulatory gene for protease production and sporulation in *Bacillus subtilis*. J. Bacteriol. **170**:2560–2567.
22. **Piggot, P. J., and J. G. Coote.** 1976. Genetic aspects of bacterial endospore formation. Bacteriol. Rev. **40**:908–962.
23. **Schiffer, M., and A. B. Edmundson.** 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. **7**:121–135.
24. **Sekiguchi, J., B. Ezaki, K. Kodama, and T. Akamatsu.** 1988. Molecular cloning of a gene affecting the autolysin level and flagellation in *Bacillus subtilis*. J. Gen. Microbiol. **134**:1611–1621.
25. **Sekiguchi, J., H. Ohsu, A. Kuroda, H. Moriyama, and T. Akamatsu.** 1990. Nucleotide sequences of the *Bacillus subtilis flaD* locus and a *B. licheniformis* homologue affecting the autolysin level and flagellation. J. Gen. Microbiol. **136**:1223–1230.
26. **Smith, I.** 1989. Initiation of sporulation, p. 185–210. *In* I. Smith, R. Slepecky, and P. Setlow (ed.), Regulation of procaryotic development. American Society for Microbiology, Washington, D.C.
27. **Strauch, M. A., G. B. Spiegelman, M. Perego, W. C. Johnson, D. Burbulys, and J. A. Hoch.** 1989. The transition state transcription regulator *abrB* of *Bacillus subtilis* is a DNA binding protein. EMBO J. **8**:1615–1621.
28. **Tanaka, T., M. Kawata, Y. Nagami, and H. Uchiyama.** 1987. *prtR* enhances the mRNA level of the *Bacillus subtilis* extracellular protease. J. Bacteriol. **169**:3044–3050.
29. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
30. **Wong, S. L., L. F. Wang, and R. H. Doi.** 1988. Cloning and nucleotide sequence of senN, a novel *Bacillus natto* (*B. subtilis*) gene that regulates expression of extracellular protein genes. J. Gen. Microbiol. **134**:3264–3276.
31. **Yang, M., E. Ferrari, E. Chen, and D. J. Henner.** 1986. Identification of the pleiotropic *sacQ* gene of *Bacillus subtilis*. J. Bacteriol. **166**:113–119.
32. **Yang, M., H. Shimotsu, E. Ferrari, and D. J. Henner.** 1987. Characterization and mapping of the *Bacillus subtilis prtR* gene. J. Bacteriol. **169**:434–437.