Complete Sequence and Transcriptional Analysis of the *spo0F* Region of the *Bacillus subtilis* Chromosome†

KATHLEEN TRACH,¹ JOHN W. CHAPMAN,² PATRICK PIGGOT,³ DOMINIQUE LECOQ,[‡]
AND JAMES A. HOCH^{1*}

Division of Cellular Biology, Department of Basic and Clinical Research, Research Institute of Scripps Clinic, La Jolla, California 92037¹; Microbiology Division, National Institute for Medical Research, Mill Hill, London, England²; and Department of Microbiology & Immunology, Temple University School of Medicine, Philadelphia, Pennsylvania 19140³

Received 2 May 1988/Accepted 17 June 1988

The total sequence of a 6,314-base-pair *Bgl*III fragment of the *Bacillus subtilis* chromosome containing the *spo0F* locus has been accomplished. Several genes of interest have been identified on this DNA fragment. The *ctrA* locus was recognized as coding for CTP synthetase by comparison of its deduced sequence with that of *Escherichia coli* CTP synthetase. A total of 53% of the residues are identical between the enzymes from these organisms. The *spo0F* locus was followed immediately by a locus, *tsr*, required for RNA synthesis in this organism. Temperature-sensitive mutations within the *tsr* locus have been identified, but strains with deletions of the locus are nonviable. It was concluded that *tsr* codes for an unknown essential component of the RNA synthesis machinery. The *tsr* gene was followed by another open reading frame which could code for a protein of 19,975 *M_r*. This gene was translated in vivo, but deletion-insertion mutations within the gene had no phenotype. The gene was cotranscribed with the *tsr* gene, although about 50% of the transcripts terminated between the two genes. The *rev-4* mutation which reverts the sporulation-defective phenotype of erythromycin-resistant mutants was located to a partial open reading frame at the end of the fragment. Disruption of this open reading frame by deletion-insertion mutation did not result in a discernible phenotype. S1 protection experiments located the start sites of transcription for several of the genes on this fragment. The *spo0F* gene was found to be monocistronic. Regulation of the identified genes was investigated by using β -galactosidase gene fusions.

Stage 0 of the sporulation process is that time during which the bacterium decides to initiate those events required for the onset of sporulation. These initial steps in the sporulation process are under genetic control. Mutants have been obtained which are blocked at stage 0 of sporulation and are, presumably, deficient in regulation of the onset of the developmental sequence. These *spo0* mutations map in eight loci (14). Mutations in five of these loci, *spo0A*, *spo0B*, *spo0E*, *spo0F*, and *spo0H*, result not only in inhibition of sporulation, but also in the blockage of expression of a wide variety of genes (5). One approach to understanding the functions of the products of the *spo0* genes has been the identification of the gene products by cloning and sequencing. The *spo0A* (12, 18), *spo0B* (4, 10), *spo0F* (34, 39), and *spo0H* (8) genes have been characterized by this method.

The *spo0F* gene was the first *spo0* gene cloned, although the early studies misidentified the *spo0F* open reading frame (31). The *spo0F* gene was found subsequently to code for a protein of *M_r* 14,286 with considerable amino acid sequence similarity to the amino-terminal portion of the *spo0A* product (34). This intriguing homology was discovered to be characteristic of a large variety of regulatory proteins that respond to signals from the environment (reviewed in reference 27). Overproduction of the *spo0F* gene product by placing the gene on a multicopy vector led to the paradoxical situation of sporulation inhibition by a protein required for sporulation (25, 40). Little else was known about the *spo0F* gene

product, particularly whether it had any relationship with the genes closely linked to it on the chromosome.

In this report, we extend the sequencing of the *spo0F* locus to include those genes surrounding it. We show that the *spo0F* locus is followed immediately by a locus, *tsr*, required for RNA synthesis in this organism. A pleiotropic suppressor mutation, *rev-4*, of ribosomal mutations has been located in the cluster of genes surrounding the *spo0F* locus. The sequence of the linked *ctrA* locus allowed its identity as the gene for CTP synthetase. None of these loci appears to be directly related to *spo0F* expression. Studies of the mRNA produced by these genes has allowed the description of the transcription start sites and the promoters for these genes. Regulation of the genes was studied by *lacZ* fusions.

MATERIALS AND METHODS

Bacterial strains, media, and transformation. *Escherichia coli* DH5 α (F⁻ *endA1 hsdR17 supE44 thi-1 λ^- recA1 gyrA96 relA1 ϕ 80 dlacZ Δ M15*) was used for the propagation of all plasmids in this study. Competent cells were purchased from Bethesda Research Laboratories and transformed according to their procedure. *E. coli* JM109 [*endA1 hsdR17 supE44 thi-1 λ^- recA1 gyrA96 relA1 Δ (lac-proAB)* (F' *traD36 proAB lacI^q lacZ Δ M15*)] was used as the host for M13 phage derivatives according to a protocol provided by Pharmacia. Selection was on L-broth agar, and, when appropriate, ampicillin, chloramphenicol, and X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) were added to 100, 20, and 40 μ g ml⁻¹, respectively. *E. coli* AR1062 (F' *thr-1 ara-13 leu-6 azi-8 tonA2 lacY T6^s minA1 gal6 λ^- minB2 Str^r malA1 xyl-7 mtl-2 thi-1 sup*) was used in minicell expression experiments.

* Corresponding author.

† Publication no. BCR-5108, Department of Basic and Clinical Research, Research Institute of Scripps Clinic.

‡ Present address: Laboratoire de Genetique, Centre de Biologie Agro-Industrielle, 78850 Thiverval-Grignon, France.

Bacillus subtilis strains were maintained on Schaeffer sporulation medium and transformed in a procedure described by Anagnostopoulos and Spizizen (1, 29). *B. subtilis* W168 (wild type) was used for gene inactivation and *lacZ* fusion studies. Selection following transformation with plasmids carrying a chloramphenicol resistance marker was on Schaeffer medium supplemented with 5 μ g of chloramphenicol ml^{-1} . Auxotrophic requirements were detected by plating on minimal medium containing 0.5% glucose. Expression of β -galactosidase fusion proteins was observed with the addition of 40 μ g of X-Gal ml^{-1} to Schaeffer-plus-chloramphenicol plates. mRNAs were isolated from *B. subtilis* JH642 (*trpC2 phe-1*) and UOT0550 (*trpC2 metB51 leuA8 nonB1 Δ spo0F sof-1*) after culture in Schaeffer medium with 0.5% glucose.

B. subtilis strains used in gene mapping were CU479 (*trpC2 ctrA1*), OSB450 (*trpC2 tsr-450*) OSB459 (*trpC2 tsr-459*) (L. Brown, Oregon State University), and 168 (*ery-1 rev-4*) (30).

Enzymes and reagents. Restriction enzymes were from New England BioLabs or Bethesda Research Laboratories and were used according to their recommended assay procedures. Bacteriophage T4 polynucleotide kinase was from United States Biochemicals Corp. Phage T4 DNA ligase, *E. coli* DNA polymerase-large fragment, ultrapure agarose, X-Gal, and isopropyl- β -D-thiogalactoside were obtained from Bethesda Research Laboratories. Calf intestinal alkaline phosphatase was purchased from Boehringer Mannheim Biochemicals. FPLC-pure mung bean nuclease was obtained from Pharmacia. DNA polymerase I and λ DNA were from New England BioLabs. Acrylagel and Bisacrylagel were from National Diagnostics. Sequencing kits for Maxam and Gilbert or M13 dideoxy methods were purchased from Dupont, NEN Research Products and Pharmacia, respectively. [γ - 32 P]ATP (crude fraction, 7,000 Ci/mmol) was obtained from ICN Pharmaceuticals. α - 32 P-labeled deoxynucleoside triphosphates (3,000 Ci/mmol) and [α - 35 S]dATP (>1,000 Ci/mmol) were from Amersham Corp. B38-4000 glass bubbles were from 3M. RNase Block was obtained from Stratagene.

Nucleic acid isolation. Plasmid DNA was isolated by the boiling method of Holmes and Quigley (16) with an added phenol extraction step or in a modification of the alkaline extraction procedure of Birnboim and Doly (3).

Gel-purified fragments for nick translation, nuclease S1 analysis, DNA sequencing, or routine subcloning were isolated from 4% (19:1 cross-link) polyacrylamide gels (electrode buffer, 1 $\times$ TBE, is 100 mM Tris, 100 mM boric acid, 1 mM EDTA, pH 8.2) by electroelution in dialysis bags containing TE buffer (10 mM Tris hydrochloride, pH 8.0, 1 mM EDTA).

B. subtilis chromosomal DNA was prepared by a procedure described by Marmer (23) with some modifications. *B. subtilis* mRNA was isolated by the following procedure. A 100-ml portion of Schaeffer medium plus glucose was inoculated with a loopful of cells from an overnight plate. Bacterial growth was monitored at a wavelength of 525 nm, and cells were harvested at either transition (T_0) or 2 h posttransition (T_2). The culture was chilled in ice water for 10 min and centrifuged at $8,000 \times g$ for 10 min at 4°C. The cell pellet was suspended in 4 ml of ice-cold LETS buffer and transferred to a 10-ml disposable polypropylene tube containing 2 ml of baked glass beads and 4 ml of cold phenol saturated with TE buffer. LETS buffer is 100 mM lithium chloride–10 mM EDTA–1% NaDodSO₄. LETS buffer was autoclaved in the presence of 0.1% diethylpyrocarbonate.

Tris hydrochloride, pH 7.5, was added subsequently to the buffer to a concentration of 10 mM, and the pH was corrected to 7.5 by the addition of lithium hydroxide. The cell suspension was sonicated on ice until lysis was evident, usually three to five bursts for 20 s each separated by 60-s intervals. A 2-ml amount of chloroform-isoamyl alcohol (24:1) was added to the lysate, and the combination was mixed by vigorous shaking. The phases were separated by centrifugation in an HB4 rotor for 10 min at 8,000 rpm. The aqueous phase was removed to a fresh tube and sequentially extracted with phenol-chloroform (1:1, then 1:2). Finally, the aqueous phase was made 0.2 M in lithium chloride, and RNase Block (RNase inhibitor) was added to 10 U ml^{-1} . Three volumes of ethanol were added next, and the RNA solution was stored at -20°C at least overnight or longer without detectable degradation.

Nucleic acid transfers. A 20- μ g portion of chromosomal DNA from *B. subtilis* JH642 and UOT0550 was digested with the restriction endonuclease *EcoRI* and *PstI* and electrophoresed in a 1% agarose gel containing 1 μ g of ethidium bromide (electrode buffer, 1 $\times$ TBE) ml^{-1} . *EcoRI*-*HindIII*-digested lambda DNA was the size marker. The DNA was transferred to nitrocellulose by the Southern procedure (32) and probed with the 0.6-kilobase (kb) *PstI*-*PvuII* restriction fragment containing the *spo0F* and *tsr* loci. The probe was labeled by nick translation (26) with [α - 32 P]dATP and *E. coli* DNA polymerase I. Hybridization was overnight at 37°C in a mixture of 50% formamide, 5 $\times$ SSC (0.75 M sodium chloride, 0.075 M sodium citrate), 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.02% bovine serum albumin, 0.02% NaDodSO₄, 100 μ g of sonicated calf thymus DNA ml^{-1} , and probe following a 2-h prehybridization with the above solution minus probe. After sequential washes at 37°C in 50% formamide–5 $\times$ SSC–0.02% NaDodSO₄, then 5 $\times$ SSC–0.02% NaDodSO₄, and finally, 2 $\times$ SSC, the filter was air dried and exposed to Kodak XRP film overnight at room temperature.

For the RNA blot, 1 ml of JH642 or UOT0550 RNA in ethanol was placed in a Microfuge for 10 min at 4°C, air dried, and suspended in 100 μ l of diethyl pyrocarbonate (depc)-treated water containing 25 U of RNase Block ml^{-1} . A 2- μ l portion (approximately 5 to 10 μ g of RNA) was electrophoresed in a 1% agarose gel containing 2.2 M formaldehyde as described by Lehrach et al. (20). The RNA was blotted onto nitrocellulose by capillary transfer in 20 $\times$ SSC (3 M sodium chloride, 0.3 M sodium citrate). The probe and hybridization and wash conditions were the same as those outlined for the Southern blot. Background hybridization to the abundant rRNA species served as molecular weight markers.

S1 nuclease analysis. Nuclease S1 protection assays were performed for the *orfS*, *spo0F*, *tsr*, and *orfU* loci by the method of Berk and Sharp (2). End-labeled DNA probes complementary to the above mRNAs were prepared as follows. Plasmid pJH4183 was digested with *SacI*, and the 3' extension was removed with *E. coli* DNA polymerase I-large fragment in the presence of the four deoxynucleotide triphosphates. Following dephosphorylation with calf intestinal alkaline phosphatase, the 5' ends were 32 P labeled, using [γ - 32 P]ATP and T4 polynucleotide kinase in a reaction described previously (15). The kinased plasmid DNA was divided into two aliquots and digested with either *HpaI* or *PstI* to give a 0.45-kb *SacI*-*HpaI* probe or a 0.53-kb *SacI*-*PstI* probe for the *spo0F* and *orfS* transcription starts, respectively. The probes for S1 nuclease protection of *tsr* and *orfU* mRNAs were obtained by 5'-end labeling the 2.2- and 0.4-kb *EcoRI* fragments of ppp41.

For S1 mapping, approximately 250 µg of JH642 or UOTO550 RNA was ethanol precipitated with 50,000 cpm of end-labeled DNA probe, air dried, and suspended in 20 µl of S1 hybridization buffer {80% deionized formamide, 40 mM PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)], pH 6.4, 1 mM EDTA, pH 8.0, 0.4 M sodium chloride} containing 25 U of RNase Block ml⁻¹. The hybridization mixture was incubated at 70°C for 10 min and then rapidly transferred to and submerged in a 49°C water bath for 3 h. The hybridization reaction was terminated with the addition of 250 µl of S1 dilution buffer (0.28 M sodium chloride, 0.05 M sodium acetate, pH 4.6, 4.5 mM zinc sulfate) containing 100, 250, 500, or 1,000 U of S1 nuclease ml⁻¹. A probe-minus RNA control was included. S1 digestion was at 37°C for 1 h. Sonicated calf thymus DNA was then added to 0.5 µg ml⁻¹ and ammonium acetate was added to 0.5 M, and the digestion was extracted once with phenol-chloroform-isoamylalcohol (25:24:1). The protected hybrids were collected by precipitation with 2 volumes of isopropanol, dried briefly under vacuum, and suspended in a 99% formamide loading dye. After heating at 90°C for 3 min, the S1 samples for *spo0F* and *tsr* were loaded on a 4% (19:1 cross-link) polyacrylamide gel containing 7.4 M urea (electrode buffer, 1× TBE) along with labeled size markers (kinased Bethesda Research Laboratories kilobase ladder). The dried gel was autoradiographed overnight on Kodak XRP film with one screen at -70°C. The S1-protected fragments for the *orfS* and *tsr* probes were electrophoresed in a 6% (19:1 cross-link) polyacrylamide gel containing 7.4 M urea (electrode buffer, 1× TBE) adjacent to a Maxam and Gilbert sequence ladder of the probe fragments.

DNA sequence analysis. Extensive DNA sequencing on both strands was performed by both the Maxam-Gilbert chemical cleavage method (24) and the dideoxy-chain termination method of Sanger et al. (28). ³²P labeling of 5' and 3' termini and the casting and processing of sequencing gels have been described by us (10, 15). Bacteriophage M13mp18 and M13mp19 (38) were used for the cloning and dideoxy sequencing of various fragments. Supercoil plasmid sequencing was performed by the method of Chen and Seeburg (7).

The region between the *spo0F* and *tsr* loci was reisolated on a 1.1-kb *Pst*I-*Eco*RI fragment from the chromosome of wild-type *B. subtilis* 168 and sequenced by using both methods to rule out a cloned mutation on the original plasmid. A *Pst*I-*Eco*RI digest of W168 chromosomal DNA was fractionated on a 0.8% agarose gel in 1× TBE, and the region around 1.1 kb was excised and electroeluted into a dialysis bag containing TE buffer. The presence of the desired fragment in this fraction of chromosomal DNA was confirmed by Southern analysis (32), using the analogous fragment from the original plasmid that had been ³²P labeled by nick translation (26). The chromosomal DNA fraction was phenol and chloroform extracted and ethanol precipitated before ligation to *Pst*I-*Eco*RI-restricted pJF751 (9). Plasmid DNA from ampicillin-sensitive, chloramphenicol-resistant *E. coli* transformants was prepared and restriction mapped until one identical to the previously cloned fragment was found. DNA sequence ladders for both the newly isolated fragment and the original cloned fragment were run side by side on sequencing gels.

Plasmid constructions. The cloning of the 6.3-kb *Bgl*II fragment containing the *B. subtilis* *spo0F* locus on plasmid pPP41 has been described by us (34). Plasmids pJH4133, pJH4122, and pJH4164 contain each of the three *Pst*I fragments found within the insert of pPP41 cloned in the *Pst*I site

of the integrative vector pJH101 (11) (see Fig. 3). Also, derivatives of pJH101 are the pJH419X series of plasmids containing each of the *Hind*III fragments from the insert of pPP41. Plasmids pJH 4182 and 4183 contain the larger 3.2-kb *Cl*aI fragment of pPP41 in both possible orientations in the *Cl*aI site of pJH101. All plasmid constructions were verified by restriction analysis.

Translation fusions to the eighth codon of *lacZ* were constructed in the integrative vector pJF751 (9) (see also Fig. 3). For pJH4534 (*orfS-lacZ*), the 0.3-kb *Pst*I-*Hpa*I fragment was ligated to a *Pst*I-*Eco*RI-restricted vector in which the *Eco*RI site had been blunted in a fill-in reaction with dATP, TTP, and *E. coli* DNA polymerase I-large fragment (Klenow enzyme). Plasmid pJH4530 (*spo0F-lacZ*) contains a 1.2-kb *Eco*RI-*Pst*I insert in the *Eco*RI-*Bam*HI sites of pJH751. The *Pst*I site of the inserted fragment was blunted with the Klenow enzyme before ligation to the *Bam*HI site of the vector, which was made blunt by mung bean nuclease digestion. The 1.44-kb insert of plasmid pJH4531 (*orfX-lacZ*) is the blunt *Nde*I (by fill-in)-*Eco*RI fragment in the blunt (mung bean nuclease-treated) *Bam*HI-*Eco*RI sites of the vector. Plasmid pJH4532 (*orfY-lacZ*) contains a 1.47-kb blunt *Hind*III (by fill-in)-*Eco*RI insert in *Eco*RI-*Sma*I. Plasmid pJH4533 (*tsr-lacZ*) and pJH4535 (*orfU-lacZ*) contain 2.1- and 2.5-kb *Eco*RI fragments, respectively, in the *Eco*RI site of pJH751. The 2.5-kb fragment was produced by partial *Eco*RI digestion of pPP41.

Plasmid DNA from chloramphenicol-resistant *E. coli* transformants was prepared, and the correct orientation of the inserts relative to the *lacZ* was confirmed by restriction analysis. An in-frame fusion between the open reading frame of interest and *lacZ* was suggested by the production of a blue colony when grown in the presence of X-Gal.

The chloramphenicol resistance insertion-deletion plasmids indicated in Fig. 4 were constructed in vector pUC18 (38). Plasmid pJM105 (M. Perego and J. A. Hoch, unpublished data) contains the *Staphylococcus aureus* chloramphenicol acetyltransferase gene (*cat*) from pC194 (17) on a 1.0-kb *Hpa*II-*Sau*3AI fragment in the *Eco*RV site of the polylinker in the Bluescript (Stratagene) vector. Plasmid pJM105, which therefore contains the chloramphenicol resistance marker in cassette form, was used as the source for this gene in the constructions. All of the restriction fragment termini were unmodified except for the *Pst*I site in plasmid pJH4527, which was blunted to accept the *Hinc*II end of the *cat* fragment. Amp^r colonies resulting from the transformation of ligation mixtures were replica plated to chloramphenicol. Plasmid DNA from Amp^r Cm^r transformants was prepared and verified by restriction analysis for orientation of the *cat* gene with respect to the direction of transcription/translation of the interrupted gene. For plasmid pJH4522, the lack of an asymmetric restriction site in the 0.3-kb *Hind*III fragment required supercoil plasmid sequencing to confirm the correct orientation of this fragment.

Minicell analysis. Minicells from *E. coli* AR1052 were prepared by using a 10 to 30% glycerol gradient as described by Stone (33). Minicell aliquots (100 µl) containing pJH101, pJH4182, or pJH4183 were incubated with [³⁵S]methionine essentially as described by Frazer and Curtiss (13). Labeled proteins were separated on a 15% polyacrylamide-sodium dodecyl sulfate gel system (19) and visualized by autoradiography of the fixed and dried gel. The molecular weights of the major labeled proteins were estimated by comparison with a set of low-molecular-weight protein markers (Sigma Chemical Co.).

Expression of *lacZ*-fusion proteins in *B. subtilis*. Super-

FIG. 1. Restriction map and sequencing strategy for the *Bgl*II fragment containing the *spo0F* locus. The potential open reading frames and the direction of their transcription are indicated. The arrows show the direction and extent of each sequencing reaction. Arrows with closed circles were 5'-end labeled and those with crosses were 3'-end labeled in the Maxam and Gilbert technique. The rest of the reactions were carried out by dideoxy sequencing reactions. The arrows within the braces show the extent of the reactions carried out on the independently isolated clone (see text). Restriction fragments used for the S1 experiments are shown below the open reading frames. The end labeled is indicated by the asterisk.

coiled plasmid DNA was transformed into W168, and transformants were selected on Schaeffer medium plus chloramphenicol and X-Gal. In the absence of a detectably blue colony for pJH4531 and pJH4532, the *cat*-containing integrated DNA was amplified by plating transformants on increasing concentrations of chloramphenicol until a stable sporulation-negative phenotype was observed at 100 μ g of chloramphenicol ml^{-1} . The colony color on X-Gal was noted. For the β -galactosidase assays, strains carrying pJH4530, pJH4533, pJH4534, or pJH4535 were grown in Schaeffer sporulation medium containing 5 μ g of chloramphenicol ml^{-1} . At 30-min intervals, 6-ml samples were removed and the optical density at 525 nm and β -galactosidase activity were measured as described previously (9).

Gene inactivation in *B. subtilis*. *cat* insertion-deletion plasmids were transformed into W168 in supercoiled and linearized form. Transformants on Schaeffer medium plus chloramphenicol were scored for colonies per microgram of plasmid DNA and for a sporulation defect. Auxotrophic requirements were tested for by plating on a minimal medium with glucose and comparing growth rate and colony morphology with the wild type. In the case of W168::pJH4521 (*ctrA-cat*), minimal-plus-glucose plates were supplemented with adenine, guanine, cytosine, thymine, uracil, and cytidine at 20 μ g ml^{-1} .

RESULTS

Sequence of the *spo0F* locus and surrounding region. The original *Bgl*II fragment containing the *spo0F* gene was identified by transformation of a *spo0F* mutation (25). The sequence of part of this fragment has been reported earlier (34). The entire sequence of the fragment was undertaken to clarify the relationships of the *spo0F* gene to surrounding genes. Sequence analysis was accomplished through a combination of Maxam and Gilbert (24) and dideoxy sequencing (28) of the approximately 6.3-kb *Bgl*II fragment (Fig. 1). The

entire fragment was sequenced on both strands except for about 100 base pairs (bp) at the extreme left end of the fragment. The second strand was not done because the sequence was identical to that found by Lampe et al. (M. Lampe, C. Binnie, R. Schmidt, and R. Losick, Gene, in press). Several open reading frames were identified in addition to the previously described *spo0F* locus. The complete sequence of this fragment is shown in Fig. 2. Major open reading frames preceded by feasible ribosome-binding sites have been identified. The location of terminators was identified through the use of a computer terminator program (6). The terminator-associated inverted repeat structures have also been identified (Fig. 2).

***ctrA* locus.** The *ctrA* locus was identified on subclones of the original *Bgl*II fragment by subcloning fragments into the integrative vector pJH101. These plasmids were then used to transform the *ctrA1* mutation to prototrophy. The results of

TABLE 1. Transformation of mutant alleles by insertional vectors

Plasmid ^a	Transformation for:		
	<i>ctrA</i>	<i>tsr-450</i>	<i>rev-4</i>
pPP41	+	+	+
pJH4133	+	—	—
pJH4122	—	+	—
pJH4164	NT ^b	—	+
pJH4182	NT	+	—
pJH4190	NT	—	—
pJH4192	NT	—	+
pJH4199	NT	—	NT
pJH4193	NT	+	NT
pJH4194	NT	—	NT

^a The inserts in these vectors are shown in Fig. 3 except for pPP41, which contains the entire *Bgl*II fragment.

^b NT, Not tested.

FIG. 2. Sequence of the *spo0F* locus and surrounding DNA. Potential ribosome-binding sites preceding deduced start codons of open reading frames are underlined. The arrows indicate the locations of potential stem-loop structures involved in terminators identified by computer analysis (6). Major (P1) and minor (P2) promoters are indicated.

B

```

      1710              1740              1770              1800
AAGTGTATCAGGATGAAGTGGTGTACGAGCGCCACC GCCCACCGCTATGAGTTTAACAATGAATTCAAGCAGCAAATGGAAGAGCAAGGCTTCGTATTCTCAGGCACAAGCCCTGATGGAC
E V Y Q D E V V Y E R H R H R Y E F N N E F R Q Q M E E Q G F V F S G T S P D G

      1830              1860              1890              1920
GCCTTGTTGAAATCATCGAGCTGAAAGACCACCTTGGTTCGTGGCTTCTCAGTTCACCCAGAGTTCAAGTCAAGACCGACAAGACCTCAGCCTCTATTCAAAGGCTTTATCGGAGCGT
R L V E I I E L K D H P W F V A S Q F H P E F K S R P T R P Q P L F K G F I G A

      1950              1980              2010              2040
CTGTGGAAGCTGCAAAATCAGAAGTAATAAAAAAGACTTGCCGTTTGTCAAAAAGCGGGCAAGTCTTTTTTATTGTTTCCGGCAAGTCCAGCATGTACAATAAGATATTATGCATACTCT
>>>>>>>>>>>>>>>><<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
S V E A A N Q K *                                * END orfs

      2070              2100              2130              2160
TGCAGAATTTCTTTTAGTGTAACGATAAAGCGTGATAGACATACAGTGCTGTGCATTAAAGAGGGAAACCCGTAACGGGTGCCGCTTCATCGGGAACAAGCGGCATTTTAAATTTAGAA
<<<< <<< <<<<

      2190              2220              2250              2280
TCAATATGCACATCACAACTACTGTTCTATTCCCGCACTGTCTTTGGCTGCGGGCGATACGCATACGACTCCCGCACCTTTTTCATAAAGCTCTTTTGGGAGCTCCTGTTCTTCGGCTGTG

      2310              2340              2370              2400
CCCGTCCCTGAGCAAAACATCAGCACCTGTGCTTTCTGTCACTTCCTCAGCGTTTCTGGAAAGGCTTTGACAGATGGGAAGGGTCTTTGCTTTGGGTGGCCTCATAAAGACGCCCC

      2430              2460              2490              2520
TGAAGCTCATTGCTCCGTATAAATAAATGGAATGCCCGCTGATCACCGCTTGAGCAAGCAGCCGCGCCCCATCTTCAATCGCGTCAGATTCTTATCCTGAATGCGGGAAAAAATACCT

      2550              +1 <---- P1              2610              +1 <---- P2              2640
GTAACTGCGTCGTGAATATTTTCAACATAAGTGAATCCTCCTTTATAACGTACAATATCAGTATACAACAAAAGAGAAAATGCTCAGAAAAATGCTGTAAGTAGACTATTATAATTTAA
(M <----- orfs)                                P2 <----->

      2670              2700              2730              2760
GGAAATAGGAAAAATCAAACAGAATACATACAATACTGCTTACTTTTACGAAAAATCATAATATTGGGGTGTAATAATGATGAATGAAAAAATTTTAACTGTTGATGATCAATACGGCATTTC
+1                P1 ---->          +1                -----   M M N E K I L I V D D Q Y G I
                                                (spoOF ---->)

      2790              2820              2850              2880
GTATTTTGCTAAATGAAGTGTTCAATAAAGAAGGCTACCAGACGTTTCAGGCTGCGAAGCGCCTGCAGGCGCTTGACATTGTGACAAAAAGACGGCCCGACCTTGTGCTGTTGGACATGA
R I L L N E V F N K E G Y Q T F Q A A N G L Q A L D I V T K E R P D L V L L D M

      2910              2940              2970              3000
AAATCCCAGCATGGACGGAATCGAAATCTTAAACGGATGAAGGTCAATTGACGAAAACATCCGGGTCAATTATCATGACGGCATACGGAGAGCTCGACATGATCCAGGAATCGAAGGAAT
K I P G M D G I E I L K R M K V I D E N I R V I I M T A Y G E L D M I Q E S K E

      3030              3060              3090              3120
TGGGCGCTCTGACGCACTTTGCCAAGCGTTTGACATCGACGAAATCAGAGACGCGCTCAAAAAATATCTGCCCTGAAGTCTAACTGACAAAAAGAAGAAACAAATGAATCATGTCATT
L G A L T H F A K P F D I D E I R D A V K K Y L P L K S *           (orfX -->) M N H V I
                                                                >>>>    <<<<<

      P1 --> +1                3180     NdeI              3210             HindIII              3240
ATGTTGCCGATTTGTGAAAAAGTTGGTATCCTAGTTATGGAGAAAAAGCGATCTGAGTATTACATATGACAGCAATATATGGGTGCTAGGGTGGAAAGCTTTTTCGCTAGAAGAC
M L P I C R K V G I L V M E K K R S E Y L H M T A I Y G S C *
                                                                (orfY-->)       M L G W K A F F A R R

      3270              3300              3330              3360
AATCAGGCTACATAAGGAGGACATTGACATGCCTTTAGTTTCTATGACGGAATGTTGAATACAGCTAAAGAAAAAAGGGTACGCTGTAGGACAATTTAACTTAACAACCTTGAGTTT
Q S G Y I R R T F D M P L V S M T E M L N T A K E K R V R C R T I *

```

C

3390 3420 3450 3480
 ACTCAAGCGATTTTACAAGCAGCTGAAGAAGAAAAATCTCCAGTGATCCTTGGTGTTCGGAAGGTGCGGGACGCTACATGGGCGGCTTCAAAACAGTCGTTGCTATGGTAAAAGCACTT

3510 3540 3570 3600
 ATGGAAGAGTACAAAGTGACAGTTCTGTGCGATTCACTTAGACCACGGTTCAAGCTTTGAATCTTGTGCGAAAAGCGATTGATGCAAGGTTTACATCTGTAATGATCGACGCTTCTCAC
 (tsr --->) M T V P V A I H L D H G S S F E S C A K A I H A G F T S V M I D A S H

3630 3660 3690 3720
 CATCCATTTGAAGAAAACGTAGCAACAACGTAAAGTTGTTGAGCTTGCTCATTCCACGGAGTATCTGTAGAAGCTGAGCTTGAACAGTAGGCGGACAAGAAGATGACGTTATCGCT
 H P F E E N V A T T A K V V E L A H F H G V S V E A E L G T V G G Q E D D V I A

3750 3780 3810 3840
 GAAGGCGTAATCTACGCTGATCCTAAAGAGTGCCCAAGAGCTTGTGAGCGCACTGGCATCGACTGCCTTGACCTGCATTAGGTTCTGTTTACGGGCCCTTACAAAGGTGAGCCAAACCTT
 E G V I Y A D P K E C Q E L V E R T G I D C L A P A L G S V H G P Y K G E P N L

3870 3900 3930 3960
 GGCTTCAAGAAAATGGAAGAGATCGGAAAATCAACTGGCCTTCCGCTAGTTCTTCACGGCGGTACTGGAATTCGACTGCTGACATCAAAAAATCAATTTCACTTGGTACAGCTAAATC
 G F K E M E E I G K S T G L P L V L H G G T G I P T A D I K K S I S L G T A K I

3990 4020 4050 4080
 AACGTGAATACTGAAAACCAATTTCTTCTGCAAAAGCTGTTGCGGAAACATTGGCAGCTAAGCCTGACGAGTACGATCCACGTAAATACCTTGGACCAGCTCGTGAAGCGATCAAAGAA
 N V N T E N Q I S S A K A V R E T L A A K P D E Y D P R K Y L G P A R E A I K E

4110 4140 4170 4200
 ACAGTTATCGGTAAATGCGTGAATTTGGTCTTCAAACCAAGCTTAATTCAATTGGAACTTTTATAGCCGCTGACAGCTTGACAGGCGGTTTTCCGTCTATCTTTAGAAAGCGCAAT
 T V I G K M R E F G S S N Q A * >>>>>> >> << <<<<<<<

4230 4260 4290 4320
 CATATTCGCAGTTCAGAAAATTTCTTAATCAGGGAGGCTTTTTGAAAATGTTATTCTTTGTTGATACAGCCAATATCGATGAAATTAGAGAAGCGAATGAATTAGGAATTCGCGCGGTG
 ----- M L F F V D T A N I D E I R E A N E L G I L A G
 (orfU ---->)

4350 4380 4410 4440
 TAACGACGAATCCTAGTTTAGTAGCAAAAGGAAGCTAACGTATCATTCCACGACGCTCTTCGCGAGATCACAGAGCTCGTGAAAGGGTCTGTAAGCGCAGAGGTTATTTCTTTGAAAGCTG
 V T T N P S L V A K E A N V S F H D R L R E I T D V V K G S V S A E V I S L K A

4470 4500 4530 4560
 AGGAAATGATCGAGGAAGGAAAAGAACTGGCGAAGATCGCTCCGAACATTACGGTGAAAATCCCAATGACGTCTGACGGTTTAAAGCGGTAAGAGCACTTACTGGCTTAGGCATCAAAA
 E E M I E E G K E L A K I A P N I T V K I P M T S D G L K A V R A L T G L G I K

4590 4620 4650 4680
 CAAACGTTACATTGATCTTCAATGCCAACAGGCGCTTCTTGCTGCCAGAGCGGGGGCAACATATGTATCTCCATTCTGGGACGTTTAGATGACATCGGCCACAACGGGCTTGACCTGA
 T N V T L I F N A N Q A L L A A R A G A T Y V S P F L G R L D D I G H N G L D L

4710 4740 4770 4800
 TTTCAGAAGTTAAACAGATTTTGTGACATTCACGGCCTTGACACGCAATCATTGCAGCGTCAATCCGCCATCCGACGACGTGACAGAAGCTGCTCTTAGAGGGGCTCATACGGCACAAT
 I S E V K Q I F D I H G L D T Q I I A A S I R H P Q H V T E A A L R G A H T A Q

4830 4860 4890 4920
 GCCGCTGAAAGTCATTGATGCGCTCACTAAACACCCGTTAACAGACAAAGGAATCGAACAATTCCTGGCAGACTGGAACAAATAATGAAAGGGGCGGCAACAGCTTTATGCCTGTTTGC
 C R * >>>> >>>>>>>>> > < <<<<<<<

4950 4980 5010 5040
 CGGGCCTTTGTATTTACCGACCTGATACAGCGGATCTTCAGATGCTGACGCCGCTTTACAGAGAAGGAATCATTGGGCTGCAGAAGCTGATATCGGAATGAAGGGTATGCGTACTGCTT
 << <<<<

D

5070 5100 5130 5160
 TTGCCGGATGACGACGATGTCTCTCGAGCCGGGGTCTGTGCATCTTTTGAACCTGTACCCCCAAAGCTTCGCTATTTCTTCTGATATTTACAATGTGTGATTCTTCTC

5190 5220 5250 5280
 TTCGAATTGGCTAGATTTTGGACATGAGCTTTATTTGCAGAGACAGAATAAAATGACATTTATTTAGCTCCGCAGCTATCTTCTTATGGAAGGAGACTATCATGGAAAAGTTGAATATT
 >>>> <<<< ----- M E K L N I
 (orfR ---->)

5310 5340 5370 5400
 GCCGGCGGTGACTCGTTAAACGGTACAGTACATATCAGCGGCGCTAAAAACAGCGCTGTTGCGTTAATACCTGCAACCATTTTGGCAAATCCGAGGTGACAATTGAAGGGCTTCCAGAG
 A G G D S L N G T V H I S G A K N S A V A L I P A T I L A N S E V T I E G L P E

5430 5460 5490 5520
 ATTTGAGATATTGAAACGCTGCGTGACCTGTTAAAGGAAATCGGCGGCAACGTGCATTTTGAAGTGGGAAATGGTTGTTGACCCTACGTCGATGATCAGCATGCCGCTTCTAACGGG
 I S D I E T L R D L L K E I G G N V H F E N G E M V V D P T S M I S M P L P N G

5550 5580 5610 5640
 AAAGTAAAAAGCTTCGCGCGTCATATTATTTAATGGGGCGATGCTCGGCGGCTTCAAGCAGGCGGTTCATTGGATTGCCTGGCGGCTGTCACTTAGGGCCCCGTCGATTGATCAGCAT
 K V K K L R A S Y Y L M G A M L G R F K Q A V I G L P G G C H L G P R P I D Q H

5670 5700 5730 5760
 ATCAAAGGCTTTGAAGCACTCGGAGCTGAAGTAACCAATGAACAAGGCCCACTTTATTTGCGAGCTGAAAGGCTGAGAGGCGCACGGATTTATTTAGATGTCGTAAGCGTTGGGGCAACG
 I K G F E A L G A E V T N E Q G A I Y L R A E R L R G A R I Y L D V V S V G A T

5790 5820 5850 5880
 ATTAACATTATGCTCGCGCTGTTTTGGCAGAAGGGAAAACGATCATCGAAAACGCTGCCAAGGAGCCTGAGATCATTGACGTCGCGACATTGCTTACCAGCATGGGCGCCAAATCAAA
 I N I M L A A V L A E G K T I I E N A A K E P E I I D V A T L L T S M G A K I K

5910 5940 5970 6000
 GGTGCGGGCACCAATGTGATTGCAATCGACGGCGTGAAGGAAGTGCACGGCTGCAAGCATACGATCATTCCGGACAGAATTGAAGCCGGACATTTATGATTGCAGGGGCTGCAATGGG
 G A G T N V I R I D G V K E L N G C K H T I I P D R I E A G T F M I A G A A M G

6030 6060 6090 6120
 AAGGAAGTCATTATCGATAACGTCATCCCTACTCATCTTGAGTCGTTAACGGCAAAGCTGAGAGAAATGGGCTATCATATCGAAACAAGCGACGACAGCTCCTCATTGTGCGCGGGCAG
 K E V I I D N V I P T H L E S L T A K L R E M G Y H I E T S D D Q L L I V G G Q

6150 6180 6210 6240
 AAGAACTTAAAGCCGGTTGACGTCAAAACCTCGTATACCGGGGTTTCCGACTGATTACAGCAGCCGATGACGGCGCTCCTGACAAGGGCGAAAGGGACGAGTGTCGTACAGACACC
 K N L K P V D V K T L V Y P G F P T D L Q Q P M T A L L T R A K G T S V V T D T

6270 6300
 ATCTACTCGGCAAGATTCAAGCACATTGATGAGCTGAGACGAATGGGTGCCAATATGAAAGTAGAAGGCAGATC
 I Y S A R F K H I D E L R R M G A N M K V E G R

these experiments are shown in Table 1, using some of the plasmids shown in Fig. 3. Both the original *Bg/II* fragment and the fragment contained within plasmid pJH4133 were able to transform the *ctrA1* allele to prototrophy. We had shown in previous studies that insertion of a fragment of λ into the *orfS* open reading frame did not give rise to a *ctr* phenotype (34). Therefore, the large open reading frame preceding *orfS* was identified, by elimination, as the *ctrA* gene. This putative *ctrA* open reading frame codes for 535 amino acids, yielding a protein of 59,717 M_r . An insertion-deletion mutant was obtained in this open reading frame by inserting a chloramphenicol resistance determinant in the *HindIII* and *EcoRI* sites (Fig. 4). When this plasmid,

pJH4521, was linearized and transformed into a wild-type strain for chloramphenicol resistance, the resultant strain acquired the *CtrA* phenotype. This strain differed from the original strain bearing the *ctrA1* mutation, which could grow on minimal medium in the presence of ammonium ion but required cytidine in its absence, by growing only very slowly on minimal medium without the addition of cytidine in the presence of ammonium ion. This locus was followed closely by two potential terminators. The strongest of these, as defined by the terminator program, begins at base 2023. This is within the coding region of the *orfS* open reading frame running in the opposite direction. This putative terminator is preceded by another putative terminator of somewhat less

FIG. 5. S1 nuclease experiments on transcription of *orfS*. mRNAs from JH642 (lane 1) or UOT0550 (lane 2) were subjected to S1 analysis as described in Materials and Methods. The probe used in these experiments was labeled at the *SacI* site within *orfS* (Fig. 1). The sequencing reactions of the complementary strands were carried out on DNA labeled at the *SacI* site.

To determine the transcription initiation site of the *spo0F* gene, an *HpaI-SacI* fragment was labeled at the *SacI* site and used in S1 experiments with mRNA prepared from a wild-type sporulating strain. The results of this experiment are shown in Fig. 6. Increasing levels of S1 nuclease show that a major protected species of approximately 300 bp was observed. This corresponds very closely to the start site observed by Smith and colleagues (21) and confirms that the promoter they identified is indeed the major promoter for this gene. Although not shown, if the autoradiograph was overexposed considerably, a minor band appeared at 330 bp. This band corresponds to the weak promoter previously observed (21). Since these authors have already run the S1-protected species against a sequencing ladder, it was deemed unnecessary to further characterize the origin of these transcripts.

***orfX* and *orfY*.** Immediately following the *spo0F* locus were two potential open reading frames that could code for small peptides. These open reading frames were preceded by reasonable ribosome-binding sites. Since both of these open reading frames would have been deleted by the integration of pJH4522, it was important to verify whether these open reading frames were capable of coding for proteins. To test this possibility, we constructed two *lacZ* translation fusions, one to each open reading frame. The *lac* fusion pJH4531 was constructed at the *NdeI* site (nucleotide 3184) in *orfX*, and plasmid pJH4532 was constructed in *orfY* at a *HindIII* site (nucleotide 3220; Fig. 3). Both of these fusions were designed to be in frame, and *E. coli* colonies bearing them were blue on X-Gal plates. When these plasmids were transferred into a *B. subtilis* wild-type strain by transformation leading to a Campbell integration, both gave a white, i.e., Lac⁻, phenotype. Thus, it seemed that neither *orfX* nor *orfY* was

significantly translated in *B. subtilis*. To verify that this was indeed the case, the plasmids were amplified by plating these chloramphenicol-resistant strains on 50 and 100 μ g of chloramphenicol ml⁻¹. At 100 μ g of chloramphenicol ml⁻¹, the colonies that survived were 100% spore defective, which is a consequence of the amplification of the *spo0F* gene (25, 40). All of these strains containing an amplified *orfX* or *orfY* were still Lac⁻. Thus, it was concluded that these open reading frames are likely not to code for any protein in vivo. This result suggests that the anomalous behavior of pJH4522 which deletes *orfX* and the start of *orfY* was probably not due to deletion of these open reading frames since they are not detectively translated in vivo. It seems possible that the deletion in pJH4522 which starts at the *HindIII* site may in fact delete the promoter for the gene following *spo0F*, *tsr*.

***tsr* locus.** The *tsr* locus was defined by two mutations, *tsr-450* and *tsr-459*, which were originally isolated as temperature-sensitive RNA synthesis mutants (L. Brown, manuscript in preparation). These mutations were found to lie downstream from the *spo0F* locus by three-factor transformation crosses with *spo0F* and *ctrA* mutations (L. Brown, personal communication). We attempted to determine the location of these alleles by using transformation with the integrative plasmids shown in Fig. 3. The results of this analysis (Table 1) show that the *tsr-450* and *tsr-459* mutations could be transformed to prototrophy by plasmids pJH4122 and pJH4182, which would place these mutations in *spo0F*, *orfX*, *orfY*, or the open reading frame labeled *tsr* (Fig. 3). In addition, plasmid pJH4193, which is wholly contained

FIG. 6. S1 experiments on transcription of *spo0F* and *tsr*. The *spo0F* transcript was assayed by using an *HpaI-SacI* restriction fragment labeled at the *SacI* site (Fig. 1). Lane B, Probe alone; lane C, probe and S1 (100 μ g ml⁻¹) without RNA; lane D, probe and S1 and RNA. The *tsr* transcript was assayed with an *EcoRI* fragment labeled at the *EcoRI* site within *tsr* (Fig. 1). Lane E, Probe alone; lane F, probe and S1 (100 μ g ml⁻¹) without RNA; lane G, probe and S1 and RNA. Lanes A and H are molecular weight standards of (top to bottom) 0.51, 0.4, 0.34, 0.3, and 0.22 kilobase pairs.

within the putative *tsr* open reading frame, was capable of transforming these mutations to prototrophy although at a very low frequency, suggesting that most of the transformants were nonviable. In this case the few chloramphenicol-resistant transformants obtained were prototrophic for the *tsr* locus but the majority of the cells within the colony were sensitive to the antibiotic. Thus, the transformants were not ordinary Campbell integrants, but their structure was not investigated further. To verify the location of the *tsr* alleles, we constructed a deletion-insertion chloramphenicol vector, pJH4523, with the chloramphenicol resistance gene inserted between the *Pvu*II and the *Eco*RI sites (Fig. 4). This vector will not transform *B. subtilis* for chloramphenicol resistance as either an intact plasmid or a linearized plasmid, suggesting that the *Pvu*II to *Eco*RI fragment is wholly contained within the transcription unit for *tsr*. This further implies that the promoter for *tsr* is upstream, i.e., to the left in Fig. 4, of the *Pvu*II site, which makes more credible the proposition that plasmid pJH4522 may indeed delete the promoter for this gene.

The start site for transcription of the *tsr* gene was first located by S1 protection experiments, using the *Eco*RI fragment shown in Fig. 1. The *Eco*RI site of the fragment within the *tsr* gene was labeled, and this fragment of DNA was used as a probe. The results of these experiments are also shown in Fig. 6. A 770-bp protected fragment was found in these studies. This would place the promoter for the *tsr* gene very close to the end of the *spo0F* gene. To further pinpoint the location of this transcriptional start site, an *Eco*RI-*Pvu*II fragment (Fig. 1) was labeled at the *Pvu*II site and similar experiments were undertaken. In this case the protected fragment was electrophoresed beside a sequencing ladder to find the actual transcription initiation site. The result of this experiment is shown in Fig. 7, in which the start site for the *tsr* gene was located at nucleotide 3160. Examination of the region just upstream of this promoter suggests that it may, in fact, correspond to a sigma-43 promoter.

From the sequence (Fig. 2), it was difficult to readily assign a translation start site for the *tsr* gene since none of the potential start codons was preceded by an easily discernible ribosome-binding site. We thought that there might be a sequence mistake from a cloning artifact in this region which could be responsible for this problem. To test this possibility, we recloned this region from wild-type *B. subtilis* DNA and resequenced the gap region between *tsr* and *spo0F* (Fig. 1). The sequence was identical to the one we had already cloned.

Minicell experiments were undertaken, using plasmid pJH4182 to identify the protein products coded by this region (Fig. 8). This analysis included a plasmid identical to pJH4182, pJH4183, with the exception that the insert was oriented in the opposite direction in the vector. Minicell analysis of both of these plasmids revealed three major protein bands directed by the insert (Fig. 8). Band 3 (Fig. 8) is close to the correct size for the *spo0F* gene product, and band 2 is the correct size for the *orfS* gene product determined from the nucleotide sequence. Band 1, which is approximately 21,000 M_r , was found in both plasmids at the same position, suggesting that it is the product of the *tsr* locus and not a fusion protein generated by the insert and the vector. Using this information, we could show a tentative start site for translation in Fig. 2 that results in a protein of molecular weight 22,800.

Characterization of the RNA from the *tsr* locus. Northern (RNA) blots were undertaken to size the mRNAs from the *spo0F*-*tsr* region to investigate the possibility of a multi-

FIG. 7. S1 nuclease analysis of the *tsr* transcript. The *Eco*RI-*Pvu*II fragment (Fig. 1) labeled at the *Pvu*II site was used as a probe in these experiments. The protected fragment after S1 nuclease treatment was compared with a Maxam-Gilbert sequencing ladder of the same fragment. The sequencing reactions (left to right) are as follows: G, G + A, T + C, and C.

cistronic mRNA. In these experiments the 0.6-kb *Pst*-*Pvu*II probe fragment covering the carboxy terminus of the *spo0F* gene and the region between the *spo0F* gene and the *tsr* gene was used as a probe. Nick-translated probe was used to hybridize to RNA separated on a formaldehyde gel. The results of these experiments are shown in Fig. 9. The RNA from the wild-type strain, JH642, harvested at T_0 , or at 2 h after the beginning of sporulation, is shown (Fig. 9) as well as RNA isolated from strain UOT0550, which contains a deletion of the *Sac*I to *Sac*I region containing the promoters for the *spo0F* and *orfS* genes. It can be seen from the results that there is no difference in the amount or size of RNA detected in these strains. We were never able to detect on a Northern blot RNA corresponding to the *spo0F* gene itself. We believe that even at T_2 the level of *spo0F* RNA is below the limit of detection by this method. There are, however, two bands of nominal sizes 2.0 and 1.15 kb. These bands must correlate to the mRNA for the *tsr* gene since they are present in undiminished amounts in strain UOT0550, which should not produce any *spo0F* mRNA. The sizes of these fragments are consistent with the proposition that mRNA for the *tsr* gene can terminate directly after the *tsr* gene at a putative terminator located at approximately nucleotide 4160 or can read through into the *orfU* strain with termination at nucleotide 4930. The calculated sizes for such mRNAs derived from the sequence of the gene would be 1.0 and 1.8 kb. The level of both mRNAs seemed to decrease at T_2 , as fivefold more RNA was needed to generate a signal equivalent to log-phase or T_0 RNA.

orfU. Following the *tsr* gene was the open reading frame *orfU* of unknown function. *orfU* was found to code for 186 amino acids, giving a gene product of 19,975 M_r . This open reading frame was discerned to be transcribed and translated through the use of a *lac* fusion, pJH4535, to the *Eco*RI site

FIG. 8. Minicell analysis of the product of the *tsr* gene. Autoradiograph of the protein products produced in minicells from the vector with no insert (lane A), plasmid pJH4182 (lane B), and the same plasmid with the insert in the opposite orientation, pJH4183 (lane C). The molecular weights of bands 1, 2, and 3 were calculated to be 21,000, 19,000, and 12,750, respectively, from molecular weight standards (not shown).

within *orfU* (Fig. 3). This fusion was found to produce blue colonies, i.e., *lac*⁺, in *E. coli* strains bearing it and when it was integrated into *B. subtilis*. Generation of an insertionally activated *orfU* was accomplished through the insertion of the chloramphenicol resistance determinant in the *Clal*-*EcoRI* region generating pJH4524 (Fig. 4). This chloramphenicol insertion when transformed into wild-type *B. subtilis* after linearization gave chloramphenicol-resistant transformants which were perfectly capable of growing on minimal media and made spores at the wild-type level. Thus, even though *orfU* appears to be transcribed and translated in vivo, inactivation of the gene has no discernible effect on growth or sporulation. Labeling of a fragment at the *EcoRI* site in *orfU* (Fig. 1) followed by S1 analysis gave no evidence of a transcription start site between *tsr* and *orfU*, confirming that *orfU* is cotranscribed with *tsr* (data not shown).

orfR. *orfR* is the final gene in the fragment sequenced, and only a partial sequence has been obtained because it extends beyond the end of the fragment. *orfR* is most probably the location of the mutation *rev-4*, which is a pleiotropic suppressor of the sporulation defect of certain ribosomal mutations induced by erythromycin resistance (30). The location of the *rev-4* allele was determined by using the plasmids outlined in Fig. 3, and the results are presented in Table 1. Both plasmids pJH4164 and pJH4192 were capable of transforming the *rev-4* mutation to prototrophy. Insertion of the wholly contained vector pJH4190 into the chromosome or insertion of the chloramphenicol vector pJH4525 (Fig. 4) had no effect on the ability of wild-type strains to grow on minimal media, and such strains were not sporulation defective. Thus, it seems most likely that the product encoded by *orfR* is not necessary for growth or sporulation in this organism.

Regulation of genes in the *spo0F* region. To study the regulation of genes in the *spo0F* region, we constructed a series of *lac* fusions in the translational fusion vector pJH751. This is an integrative vector and therefore yields a single copy of the *lac* fusion upon integration into the chromosome. Plasmids were constructed with the fragments shown in Fig. 3. These plasmids were then integrated into strain 168, and the production of β -galactosidase was monitored during growth and sporulation. Studies of a similar nature have been carried out with the *spo0F-lac* fusions (36). The results of our experiments (Fig. 10) confirm that the *spo0F* locus is translated at a relatively constant rate during the log phase of growth, but at T_0 of sporulation there is a significant, fourfold increase in the rate of expression of this gene. *orfS*, however, seems to be regulated as a low-level constitutive during log phase and decreases in expression as sporulation proceeds. The *lac* fusion to the *tsr* gene and to *orfU* followed a similar pattern which might be expected if they both come from the same transcript as predicted from the Northern blot experiments. Both of these genes were expressed at high levels, with accumulation of the proteins occurring continuously during the exponential phase of growth. However, at T_0 , there was a precipitous fall in the accumulation of both *tsr* and *orfU*, although the synthesis of *tsr* appears to be maintained at a higher rate. This would be consistent with our observations on the relative amounts of the message populations for *tsr* and *tsr* plus *orfU*.

DISCUSSION

The original intent of this research was to determine whether the *spo0F* gene was monocistronic or part of a polycistronic operon. Several distinctive loci had been

FIG. 9. Northern blot analysis of transcripts from the *tsr* gene. RNA preparations from strain UOT0550 (lane A) or JH642 (lane B) harvested at T_0 and JH642 harvested at T_2 (lane C) were subjected to electrophoresis, blotting, and hybridization as described in Materials and Methods. The probe used was the *Pst*I-*Pvu*II restriction fragment encompassing the 3' end of the *spo0F* gene and the 5' end of the *tsr* mRNA. Five times more RNA was used in lane C than in lanes A and B. The locations of 16S and 23S RNAs are indicated. The nominal sizes of the hybridizing RNAs were calculated as 2.0 and 1.15 kilobase pairs.

FIG. 10. Experiments with *lac* fusion plasmids. The *lac* fusion plasmids shown in Fig. 3 were integrated into the chromosome of strain JH642, and β -galactosidase was assayed as a function of growth and sporulation. Results from β -galactosidase assays are indicated: *lac* fusion to *orfS* (pJH4534), ■; *lac* fusions to *spo0F* (pJH4530), ▲; *lac* fusion to *tsr* (pJH4533), ●; *lac* fusions to *orfU* (pJH4535), ○.

mapped very close to *spo0F*, and we found that mutations in three of these, *ctrA*, *tsr*, and *rev*, could be corrected with the 6.3-kb *Bgl*II fragment we had isolated as containing *spo0F* (25). A rationale could be developed as to why any one of these might be cotranscribed with *spo0F*. Sequence analysis revealed that the *spo0F* locus was followed by two small open reading frames, *orfX* and *orfY*, that could code for proteins of 35 and 44 amino acids, respectively. Since small peptides such as the product of the *sacQ* locus, 46 amino acids, have a profound effect on the synthesis of a variety of proteins (37), it seemed important to determine whether these open reading frames following the *spo0F* gene were translated. Construction of *lacZ* fusion proteins to both of these open reading frames allowed us to test this possibility. Neither fusion was translationally active even when the copy number of the fusion in the chromosome was amplified by selecting for increased chloramphenicol resistance. There is a potential terminator just beyond the end of the *spo0F* gene. It seems likely that *spo0F* transcription terminates before *orfX* and *orfY*. Thus, the *spo0F* gene appears to be monocistronic, like the majority of the *spo0* genes.

The transcription start site for the *tsr* locus was located just downstream of the putative *spo0F* terminator at base 3160 (Fig. 2), which indicates that mRNA is produced for *orfY*. It seems possible that *orfY* has something to do with the regulation or activity of the *tsr* gene product. However, that we cannot demonstrate translation of this open reading

frame makes this interpretation unlikely. The leader sequence between the transcription start site and the putative start of *tsr* translation is 335 bp, which is abnormally long for such a sequence in bacteria. The role of this leader sequence and whether any other translation products emanate from it are unclear. Mutagenesis experiments in this region may help to resolve this dilemma.

The *tsr* locus is enigmatic. Mutations in the *tsr* locus clearly have a temperature-sensitive phenotype and stop RNA synthesis immediately upon transfer to the nonpermissive temperature (Brown, personal communication). The *tsr* gene is only capable of coding for a protein of between 20,000 and 25,000 *M_r*, and the minicell experiments suggest that the size of the product of this locus is closer to 21,000 *M_r*. This is smaller than any of the major subunits of RNA polymerase except for delta, which is approximately this size. The delta factor has recently been cloned, sequenced, and mapped, and its gene lies just upstream of the *ctrA* gene (Lampe et al., in press). Moreover, genes for the other subunits have already been identified and do not map in this region (22). Thus, the immediate cessation of RNA synthesis observed in *tsr* mutants upon temperature increase cannot be ascribed to temperature sensitivity mutations in the structural genes for the major subunits of RNA polymerase. We entertained the possibility that the product of the *tsr* gene might be involved in some modification reaction of RNA polymerase which, in the absence of the *tsr* gene product, results in a temperature-sensitive RNA polymerase. This proposition seems unlikely since we were unable to produce a deletion mutation within the *tsr* gene. It seems more reasonable to assume that the product of the *tsr* gene is, itself, temperature sensitive in strains carrying the *tsr-450* and *tsr-459* mutations. Whatever the function of the *tsr* gene product, it must be a crucial component or affect a crucial component of the RNA synthesis machinery.

Northern analysis as well as S1 experiments (unpublished) established that *orfU* was cotranscribed with *tsr*. Approximately half of the time the *tsr* transcript terminated after *tsr* and before *orfU*. These results suggested that the *orfU* gene product may bear some relationship to *tsr* function. Mutations in *orfU*, however, were without phenotype.

Both the *spo0F* gene and *orfS* are transcribed from two promoters. In the case of *spo0F*, the major promoter was found downstream of the minor promoter (21). Our studies confirmed this assignment. *orfS* is located upstream of the *spo0F* locus and transcribed from the opposite strand. The minor promoter for this gene is also located upstream of the major promoter and overlaps with the minor promoter of *spo0F*. The sequences of these minor promoters do not allow an unambiguous assignment of sigma factors responsible for their transcription; therefore, that they overlap is of uncertain significance.

Two of the open reading frames, *orfS* and *orfU*, are translated in vivo, yet chloramphenicol insertion-deletion mutations can be produced in *orfU* and λ fragments can be inserted in *orfS* (34) with the result that neither growth nor sporulation is detectably altered. Thus, the role of these genes in the cell is completely unknown. The *rev-4* mutation maps in *orfR*. The mutation results in the phenotype of reversal of the sporulation defect of certain erythromycin-resistant mutations (30). However, it is clear from both insertional inactivation studies with wholly contained insertional vectors and the use of a chloramphenicol insertion-deletion mutation that the function of the *orfR* gene product can be deleted without detectable effect or either growth or sporulation of the wild-type strain. Thus, this study has not

- nol. 69:1-84.
14. Hoch, J. A. 1976. Genetics of bacterial sporulation. *Adv. Genet.* **18**:69-99.
 15. Hoch, J. A., K. Trach, F. Kawamura, and H. Saito. 1985. Identification of the transcriptional suppressor *sof-1* as an alteration in the *spo0A* protein. *J. Bacteriol.* **161**:552-555.
 16. Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114**:193-197.
 17. Horinouchi, S., and B. Weisblum. 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* **150**:815-825.
 18. Kudoh, J., T. Ikeuchi, and K. Kurahashi. 1985. Nucleotide sequences of the sporulation gene *spo0A* and its mutant genes of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **82**:2665-2668.
 19. Laemmli, U. K. 1970. Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
 20. Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker. 1977. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical re-examination. *Biochemistry* **16**:4743-4751.
 21. Lewandoski, M., E. Dubnau, and I. Smith. 1986. Transcriptional regulation of the *spo0F* gene of *Bacillus subtilis*. *J. Bacteriol.* **168**:870-877.
 22. Losick, R., R. Youngman, and P. J. Piggot. 1986. Genetics of endospore formation in *Bacillus subtilis*. *Annu. Rev. Genet.* **20**:625-669.
 23. Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J. Mol. Biol.* **3**:208-218.
 24. Maxam, A. M., and G. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
 25. Piggot, P. J., J. W. Chapman, and C. A. M. Curtis. 1985. Analysis of the control of *spo* gene expression in *Bacillus subtilis*, p. 15-21. In J. A. Hoch and P. Setlow (ed.), *Molecular biology of microbial differentiation*. American Society for Microbiology, Washington, D.C.
 26. Rigby, P. W. J., M. Dieckman, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**:237-251.
 27. Ronson, C. W., B. T. Nixon, and F. M. Ausubel. 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. *Cell* **49**:579-581.
 28. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
 29. Schaeffer, P., J. Miller, and J. Aubert. 1965. Catabolic repression of bacterial sporulation. *Proc. Natl. Acad. Sci. USA* **54**:701-711.
 30. Sharrock, R. A., and T. Leighton. 1982. Suppression of defective-sporulation phenotypes by the *Bacillus subtilis* mutation *rev4*. *Mol. Gen. Genet.* **186**:432-438.
 31. Shimotsu, H., F. Kawamura, Y. Kobayashi, and H. Saito. 1983. Early sporulation gene *spo0F*: nucleotide sequence and analysis of gene product. *Proc. Natl. Acad. Sci. USA* **80**:658-662.
 32. Southern, E. M. 1975. The detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
 33. Stone, A. B. 1974. A simplified method for preparing sucrose gradients. *Biochem. J.* **137**:117-118.
 34. Trach, K., J. W. Chapman, P. J. Piggot, and J. A. Hoch. 1985. Deduced product of the stage 0 sporulation gene *spo0F* shares homology with the *spo0F*, *ompR* and *sfrA* proteins. *Proc. Natl. Acad. Sci. USA* **82**:7260-7264.
 35. Weng, M., C. A. Makaroff, and H. Zalkin. 1986. Nucleotide sequence of *Escherichia coli* *pyrG* encoding CTP synthetase. *J. Biol. Chem.* **261**:5568-5574.
 36. Yamashita, S., H. Yoshikawa, F. Kawamura, H. Takahashi, T. Yamamoto, Y. Kobayashi, and H. Saito. 1986. The effect of *spo0* mutations on the expression of *spo0A*- and *spo0F-lacZ* fusions. *Mol. Gen. Genet.* **205**:28-33.
 37. 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.
 38. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the m13mp18 and pUC10 vectors. *Gene* **33**:103-119.
 39. Yoshikawa, H., J. Kazami, S. Yamashita, T. Chibazakura, H. Sone, F. Kawamura, M. Oda, M. Isaka, Y. Kobayashi, and H. Saito. 1986. Revised assignment for the *Bacillus subtilis* *spo0F* gene and its homology with *spo0A* and with two *Escherichia coli* genes. *Nucleic Acids Res.* **14**:1063-1072.
 40. Yoshikawa, H., H. Sone, T. Seki, F. Kawamura, and H. Saito. 1988. Functions of *spo0F* upstream region for the sporulation of *Bacillus subtilis*, p. 29-34. In J. A. Hoch and P. Setlow (ed.), *Sporulation and germination genes*. American Society for Microbiology, Washington, D.C.