The *Bacillus subtilis* *spo0B* Stage 0 Sporulation Operon Encodes an Essential GTP-Binding Protein†

KATHLEEN TRACH AND JAMES A. HOCH*

Division of Cellular Biology, Department of Basic and Clinical Research, Research Institute of Scripps Clinic,
10666 North Torrey Pines Road, La Jolla, California 92037

Received 29 September 1988/Accepted 6 December 1988

Transcription of the *spo0B* gene and genes downstream of it was investigated by S1 nuclease protection experiments. The *spo0B* gene was transcribed from a single promoter, and this transcript extended through a gene, *obg*, coding for a 47,668 M_r protein. Transcription of this operon ended in a stem-loop structure. The sequence of the deduced *obg* protein contained a region with homology to known GTP-binding proteins in the nucleotide-binding regions. The amino-terminal portion of this protein showed homology to mammalian collagen, suggesting a structural role. The purified *obg* protein was shown to bind [α - 32 P]GTP in vitro. Several attempts to inactivate the *obg* gene were unsuccessful, indicating that the *obg* gene product was essential for growth. The possible function of this protein and its relationship to RAS-like proteins and sporulation was discussed. Immediately downstream of the *obg* gene were two genes involved in phenylalanine biosynthesis, *pheB* and *pheA*. The *pheA* gene coded for monofunctional prephenate dehydratase, on the basis of the high homology of the deduced amino acid sequence to prephenate dehydratases of bacterial origin. The sequence of the *pheB* gene product was not homologous to chorismate mutase, and its function remains unknown. Transcription of the *phe* genes was shown to begin at the stem-loop structure between *obg* and *pheB*. The possibility was entertained that *phe* gene transcription arises from processing or antitermination of the *spo0B* transcript.

The onset of sporulation in *Bacillus subtilis* requires the products of several stage 0 sporulation genes (20). Among these is the *spo0B* gene, which maps near the *pheA* locus on the *B. subtilis* chromosome. The *spo0B* gene has been cloned and characterized in several laboratories. A specialized transducing bacteriophage carrying the *spo0B* gene was isolated by Hirochika et al. (19). This phage complemented several *spo0B* alleles. Bonamy and Szulmajster (6) isolated the *spo0B* locus on a plasmid vector. Their analyses suggested that the *spo0B* gene coded for a product of M_r 24,000. The *spo0B* gene and the chromosomal region surrounding this gene was cloned in the λ vector Charon 4A by Ferrari et al. (12). Using the *trpE26* system of merodiploid formation, these investigators were able to show that all of the known alleles of the *spo0B* locus make up a single complementation group. Furthermore, the locations of several of the surrounding loci were mapped on the cloned fragments.

The sequence of the *spo0B* gene was reported by Bouvier et al. (7) and by Ferrari et al. (14). The deduced product of the *spo0B* gene was a protein of M_r 22,542. S1 nuclease protection analysis allowed the identification of the transcription start site for *spo0B* (7). Similar S1 experiments with different probes showed that the mRNA for the *spo0B* gene extended beyond this gene into a partially sequenced open reading frame (14). The results were consistent with the conclusion that the *spo0B* gene and this open reading frame are part of an operon of undetermined length (14).

In the work described in this communication, we extended our sequencing and S1 nuclease protection analyses of the *spo0B* operon. The second gene in this operon was shown to code for an essential GTP-binding protein of unknown

function. Immediately following this gene and separated from it by a terminator were two open reading frames coding for components of the phenylalanine biosynthetic pathway. The results suggested, but did not prove, that these *phe* genes were part of a transcript originating at the *spo0B* promoter.

MATERIALS AND METHODS

Bacterial strains, media, and transformation. *Escherichia coli* DH5 α (F^- *endA1 hsdR17 supE44 thi-1 λ^- relA1 gyrA96 recA1 ϕ 80dlacZ Δ M15*) competent cells were purchased from Bethesda Research Laboratories, Inc., and transformed as specified in the accompanying procedure. *E. coli* JM109 [*endA1 hsdR17 supE44 thi-1 λ^- relA1 gyrA96 relA1 Δ (lac-proAB)* (F' *traD36 proAB lacI^q, lacZ Δ M15*)] were made competent and transformed by the method of Dagert and Ehrlich (9). *E. coli* strains were maintained on L broth, and, when appropriate, ampicillin, chloramphenicol, and X-Gal (5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside; Bethesda Research Laboratories) were added to 100, 10, and 40 μ g ml $^{-1}$, respectively.

B. subtilis strains were maintained on Schaeffer sporulation medium and transformed in a procedure described by Anagnostopoulos and Spizizen (2). *B. subtilis* JH642 (*trpC2 phe-1*) was used for the purification of mRNA. Strain W168 (wild type) was used for gene inactivation studies, and selection was done on Schaeffer agar plates containing chloramphenicol at 5 μ g ml $^{-1}$.

Enzymes and reagents. Restriction enzymes were obtained from New England BioLabs, Inc., or Bethesda Research Laboratories and were used as recommended by the supplier. Bacteriophage T4 polynucleotide kinase was from United States Biochemical Corp. Bacteriophage T4 DNA ligase, *E. coli* DNA polymerase I large fragment, prestained protein molecular weight standards (high range), and the

* Corresponding author.

† Publication no. 5533-BCR from the Department of Basic and Clinical Research, Research Institute of Scripps Clinic.

FIG. 1. Restriction map, S1 probes, and sequencing strategy for the *spo0B* locus. The potential open reading frames and the direction of their transcription are indicated. The arrows above the restriction map show the direction and extent of each Maxam-Gilbert sequencing reaction: $\bullet$, 5' end labeled; $\circ$, 3' end labeled. The *BstNI-HindIII* insert in plasmid pJH4604 is shown. Restriction fragments used in the S1 nuclease protection assays are shown below the open reading frames. The end-labeled site is indicated by the asterisk.

1-kilobase (kb) DNA ladder were obtained from Bethesda Research Laboratories. Calf intestinal alkaline phosphatase was purchased from Boehringer Mannheim Biochemicals. Acrylagel and Bisacrylagel were obtained from National Diagnostics Inc. Maxam-Gilbert DNA sequencing reactions were performed with reagents supplied by Du Pont, NEN Research Products. $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (crude fraction, 7,000 Ci mmol^{-1}) was obtained from ICN Pharmaceuticals Inc. $[\alpha\text{-}^{32}\text{P}]\text{dGTP}$ (3,000 Ci mmol^{-1}) and $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ (3,000 Ci mmol^{-1}) were purchased from Amersham Corp. Lambda DNA was obtained from New England BioLabs.

Restriction mapping and DNA sequence analysis. A restriction map of the DNA discussed below was generated by digestion with the restriction enzymes indicated either alone or in various combinations. The DNA fragments were analyzed in 0.8% agarose gels, using $1\times$ TBE electrode buffer (100 mM Tris, 100 mM boric acid, 1 mM EDTA [pH 8.2]), and their sizes were estimated from a comparison with a lambda DNA standard cut with *EcoRI* and *HindIII*. The DNA sequence of both strands was determined by using the Maxam and Gilbert chemical cleavage method (26). We have already described the ^{32}P labeling of 5' and 3' termini and the processing of sequencing gels (14). The restriction map and sequencing strategy are shown in Fig. 1.

Nucleic acid isolation. Plasmid DNA was isolated by the boiling method of Holmes and Quigley (21), with an added phenol-chloroform (1:1) extraction step or in a modification of the alkaline extraction procedure of Birnboim and Doly (5). Gel-purified fragments for subcloning, nuclease S1 analysis, or DNA sequencing were isolated from 4% (19:1 cross-link) polyacrylamide gels (electrode buffer- $1\times$ TBE) by electroelution in dialysis bags containing TE buffer (10 mM Tris hydrochloride [pH 8.0], 1 mM EDTA). We have already described the purification of *B. subtilis* JH642 mRNA from transition stage cells (29).

S1 nuclease analysis. Nuclease S1 protection assays with mRNA prepared from strain JH642 were performed by the method of Berk and Sharp (4). The end-labeled DNA probes complementary to *spo0B*, *OrfA*, and *pheB* mRNAs shown in Fig. 1 were prepared by the following method. Plasmid pJH4222 contains the 5.6-kilobase-pair (kbp) *EcoRI-HindIII* fragment shown in Fig. 1 in the *EcoRI* and *HindIII* sites of the integrative vector pJH101 (13). In separate reactions, pJH4222 was digested with *HindIII*, *PstI*, *BglI*, or *NcoI*. The 3' extensions on the *PstI* and *BglI* ends were removed with *E. coli* DNA polymerase I large fragment in the presence of the four deoxynucleoside triphosphates. After dephosphorylation with calf intestinal alkaline phosphatase, the 5' ends were ^{32}P labeled with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and T4 polynucleotide kinase. The plasmid DNAs labeled at *HindIII*, *BglI*, and *NcoI* were next digested with *BglIII* to give the 0.72-, 1.3-, and 2.7-kbp single end-labeled probes complementary to *spo0B*, *OrfA*, and *pheB* mRNAs, respectively (Fig. 1). A second 5'-end-labeled S1 probe for the *spo0B* transcript was the 0.85-kbp *PstI-EcoRI* fragment labeled at the *PstI* site. A 3'-end-labeled probe used to determine the termination of transcription of *spo0B* mRNA was the 3.4-kbp *HindIII-PstI* fragment in which the *HindIII* site in *spo0B* was labeled in a fill-in reaction with $[\alpha\text{-}^{32}\text{P}]\text{dATP}$, $[\alpha\text{-}^{32}\text{P}]\text{dGTP}$, and *E. coli* DNA polymerase I large fragment.

The conditions for probe-mRNA hybridization and S1 nuclease digestion were identical to those described by Trach et al. (29). The protected hybrids were denatured and electrophoresed on 4% (19:1 cross-link) polyacrylamide gels containing 7.4 M urea (electrode buffer- $1\times$ TBE) with the 1-kb ladder size standard which had been ^{32}P labeled with polynucleotide kinase. The gels were dried and autoradiographed for 1 to 2 days on Kodak XRP film with one intensifying screen at -70°C .

Construction of CAT insertion plasmids. The chloramphen-

+1

ATAATACACTTATACAGACTCCTAAATAAGAAATTAAATGATTGGGAGTGCGAAAAATGAAGGATGTTCACAAAATCAAGAAGAAAATATAAGCGCACGCGCATTAACAACGAAGCTG
spoOB --> M K D V S K N Q E E N I S D T A L T N E L

160 PstI PvulI HindIII 200

ATT CAT CT G CTT GGCCATT CCCGGCAT GATT GGAT GAATA AGCTGCAGCTGATTAAGGAAACTTAAGCTTACAGAAGTATGCCGTGTCTTTGAAATGATTGAAGAAATGGTTATAGAC
I H L L G H S R H D W M N K L Q L I K G N L S L Q K Y D R V F E M I E E M V I D

280 320 360

GCAAAGCACGAATCAAAGCTCTCAAACCTGAAACACCGGCATTTGGCGTTTGATTTTTCTTACGTTAATTGGAACCATTATATGACGCTTGAATATGAAGTTCTCGGAGAAATTAAG
A K H E S K L S N L K T P H L A F D F L T F N W K T H Y M T L E Y E V L G E I K

400 440 480

GATTTGTGGCTTATGATCAAAGCTGGCGAAACTGATGAGAAAGCTGTTTCATCTGTTTGATCAAGCAGTCAGCAGAGAGAGTGAAATCATTTAACGGTTTCGCTTCAAACGGATCAT
D L S A Y D Q K L A K L M R K L F H L F D Q A V S R E S E N H L T V S L Q T D H

PvuII 520 EcoRV 600

CCTGACAGACAGCTGATTCTGTACCTTGATTTTCACGGCGCCTTTGCCGATCCATCTGCTTTTGATGATATTCCGCAGAATGGATATGAGGACGTTGATATCATGCGTTTTGAAATCAGC
P D R Q L I L Y L D F H G A F A D P S A F D D I R Q N G Y E D V D I M R F E I T

640 680 720

AGTCACGAATGTCTGATTGAAATTGGGTTGGACTAGCGGAGTTTTAACGGTTTAGAACGGAGGACATTATGTTTGATAGTCAGGTCAAAGTATATGTAAAGGCGGCGACGGCGGCAAC
S H E C L I E I G L D * orfA --> M F V D Q V K V Y V K G G D G G N

BglI

>>>> >>>>> >>>> <<<< <<<<<<<<<<<<<<<<<<<

800 840

GGTATGGTTGCGTTTTGCGCGTGAAAAATATGTGCCGAAAGGCGGCCCTGCCGCGGTGACGGAGGAAAGGGAGGAGACGTCGTTTTGAGTAGATGAAGGTTCTCCGCAACCTGATGGAT
G M V A F R R E K Y V P K G G P A G G D G G K G G D V V F E V D E G L R T L M D

DraI 880 920 960

TTTAGATACAAAAACACTTTAAAGCGATTGCGGGCGAGCATGGCATGTCCA AAA ACCAGCACGGGCGGAAATGCTGATGATATGGTCATTAAAGTTCCGCCGGGCACCGTTGTGACAGAC
F R Y K K H F K A I R G E H G M S K N Q H G R N A D D M V I K V P P G T V V T D
#

1000 1040 1080

GATGATACAAAACAGGTCATCGCTGATTTAACAGAGCACGGACAGCGGGCTGTTCATTGCAAGAGGCGGAAGAGGCGGAAGAGGAATAGCCGCTTTGCTACACCGGCTAATCCCAGCGCT
D D T K Q V I A D L T E H G Q R A V I A R G G R G G R G N S R F A T P A N P A P
#

1120 1160 1200

CAGCTTTGAGAAACGCGGAGCGCGGAAAGAACGCTACATTGTTCTTGAATTTAAAGTGCTTGAGATGTCGGACTTGTCGGGTTCCCGAGTGTCGGAAATCTACTTTGCTGTCTGTT
Q L S E N G E P G K E R Y I V L E L K V L A D V G L V G F P S V G K S T L L S V

1240 1280 1320

GTCTCATCTGCAAAACCGAAAATTGCGGACTATCACTTTACAACGCTTGTCGCCGAATCTCGGCATGGTTGAACGGATGACGGACGCAGCTTTGTGTCGGCTGATTGCGACGGGCTGATT
V S S A K P K I A D Y H F T T L V P N L G M V E T D D G R S F V M A D L P G L I

1360 1400 1440

GAAGGCGCACACCAAGGCGTCGGACTGGGCCACAGTTTTTGC GCC ATATTGAACGGACGAGGGTTATTGTTTCATGTCATTGACATGTCAGGCTTAGAAGGCGGTGATCCATATGATGAT
E G A H Q G V G L G H Q F L R H I E R T R V I V H V I D M S G L E G R D P Y D D

1480 1520 DraI

TATCTTACGATTAACCAGGAGCTGAGCGAGTACAATCTCGCTCTCACTGAGCGTCCGCAAATTATCGTTGCAAAATAAATGGACATGCCGGAGGCTGCGGAAAATCTCGAAGCCTTTAA
Y L T I N Q E L S E Y N L R L T E R P Q I I V A N K M D M P E A A E N L E A F K

1600 1640 EcoRI 1680

GAAAAGCTGACGGATGATTATCCGGTATTTCCCGATCAGTCGGTGACAAGAGAAGGTCTGCGTGAGCTTCTGTTTGAGGTTGCCAATCAGCTTGAGAACACACCGGAATTTCCCGTGTAT
E K L T D D Y P V F P I S A V T R E G L R E L L F E V A N Q L E N T P E F P L Y

FIG. 2. Sequence of the *spo0B* locus. Amino acid residues indicated by # in OrfA are homologous to mammalian collagen (15), whereas residues indicated by * form the GTP-binding site of the protein.

icol resistance insertion-deletion plasmids shown in Fig. 2 were constructed in the vector pUC18 (30). Plasmids pJH4503, pJH4504, pJH4506, and pJH4507 contain the chloramphenicol acetyltransferase gene (CAT) from pC194 (22) which had been subcloned in cassette form in the multiple cloning site of Bluescript (Stratagene Inc.) as plasmid pJM105 (M. Perego and J. A. Hoch, unpublished results). The CAT gene insertion in the above plasmids includes the CAT transcriptional terminator.

Plasmid pJH4670 contains a CAT insertion in *spo0B* where the transcriptional terminator for CAT has been deleted. A total of 115 base pairs between the *StuI* and *MboI* sites in the CAT gene were deleted, and the truncated gene was reintroduced into Bluescript (plasmid pJM102b; Perego and Hoch, unpublished). This deletion includes six carboxy-terminal amino acids and the translational and transcriptional terminators; however, the ability to confer chloramphenicol resistance is not noticeably affected.

Plasmid DNA from Amp^r Cm^r colonies resulting from transformation of ligation mixtures in *E. coli* was prepared, and the constructions were verified by restriction analysis.

Gene inactivation in *B. subtilis*. *B. subtilis* W168 (wild type) was transformed with the CAT insertion-deletion plasmids in supercoiled and linearized form. Transformants were selected on Schaeffer agar plates containing 5 µg of chloramphenicol ml⁻¹. Plates were scored for colonies per microgram of transforming DNA and for a sporulation defect. Auxotrophic requirements were detected by plating original transformants on minimal media. For strains W168::pJH4508 and W168::pJH4506, minimal-medium plates were supplemented with phenylalanine at 20 µg ml⁻¹.

GTP photoaffinity cross-linking. Expression of the *B. subtilis* OrfA protein in *E. coli* was placed under the control of the hybrid *tac* promoter in the vector pKK223-3 (Amersham Corp.). Plasmid pJH4604 was constructed containing the 3.8-kbp partial *Bst*NI-*Hind*III fragment from pJH4222 in the *Sma*I-*Hind*III sites of pKK223-3 (Fig. 1). The *Bst*NI end of the inserted fragment was blunted in a fill-in reaction with *E. coli* DNA polymerase I large fragment and deoxynucleoside triphosphates prior to ligation to the *Sma*I site of the vector.

Plasmid pJH4604 was maintained in the *lacI*^r host *E. coli* JM109. For preparation of a protein extract, bacterial cells containing pJH4604 or the pKK223-3 vector without an insert (control strain) were grown at 37°C in 100 ml of LB containing 100 µg of ampicillin ml⁻¹. When the cultures reached an optical density at 550 nm of 1.0, the inducer IPTG (isopropyl-β-D-thiogalactopyranoside; Bethesda Research Laboratories) was added to a final concentration of 2 mM, and the induction was continued at 37°C for 2.5 h. The cells were collected by centrifugation at 8,000 × *g* for 10 min, washed once with one-quarter volume of cold 1× sonication buffer (SB), and finally resuspended in 2 ml of cold 1× SB (50 mM Tris hydrochloride [pH 7.4], 100 mM NaCl, and 5 mM MgCl₂). The bacterial cells were disrupted by sonication on ice. The crude extracts were centrifuged at 15,000 × *g* for 5 min at 4°C to remove cell debris.

Control or OrfA-induced crude extracts (100 µl each) were diluted with 100 µl of distilled H₂O and applied to separate columns containing Affi-Gel Blue (bed volume, 0.5 ml; Bio-Rad Laboratories) previously equilibrated with 1/2× SB. After extensive rinsing with 1/2× SB, the columns were eluted with 0.3 M KCl in 1/2× SB in a total volume of 800 µl. The eluants were dialyzed against 1 liter of 1/4× SB at 4°C with one change and then concentrated to a volume of 200 µl by lyophilization.

The partially purified OrfA and the control protein sam-

ples were assayed for GTP-photoaffinity cross-linking in a modification of the procedures described previously (11, 27). Each reaction mixture contained 24 µl of partially purified protein in 1× SB and 1.0 µM [α -³²P]GTP (>400 Ci mmol⁻¹; Amersham Corp.). The reactions were incubated in individual wells of a microdilution plate on ice for 30 min in the dark and then exposed for 60 min to UV light (two 15-W General Electric G15T8 germicidal UV lamps) at a distance of 3 cm. A UV-minus control sample for the partially purified OrfA protein was included and was incubated on ice in the dark for 90 min.

Following irradiation, the proteins in each reaction mixture were precipitated with 5 volumes of cold 5% trichloroacetic acid. After 60 min on ice, the solutions were centrifuged at 15,000 × *g* for 10 min at 4°C, and the pellets were rinsed once with cold 5% trichloroacetic acid. The pellets were dissolved in 50 µl of sample buffer and boiled for 5 min, and 25-µl portions were analyzed on a sodium dodecyl sulfate-13.5% polyacrylamide gel (24). The gel was fixed, dried, and autoradiographed with Kodak XRP-1 film for 5 h at -70°C with one intensifying screen.

RESULTS

Sequence of the *spo0B* operon. The sequence of the region downstream of the *spo0B* gene was obtained by the sequencing strategy shown in Fig. 1. The complete sequence is depicted in Fig. 2. Immediately downstream of the *spo0B* gene, an open reading frame (OrfA) was found that encodes a protein of 428 amino acid residues with a calculated *M_r* of 47,688. This open reading frame was separated from two additional open reading frames by a stem-loop structure with strong similarity to procaryotic terminators as determined by the computer program of Brendel and Trifonov (8). A similar structure was identified after these open reading frames by the same program. Results discussed below indicate that these open reading frames code for proteins of *M_r* 16,659 and 31,901 that are involved in phenylalanine biosynthesis, and therefore they were labeled *pheB* and *pheA*, respectively.

The terminator computer program did not find any terminator structures between *spo0B* and OrfA. A large, strong stem-loop structure was found beginning at codon 19 (Fig. 2, base 721) of OrfA. The terminator computer program did not recognize any terminatorlike structures in this stem-loop structure. This feature of the primary structure of the OrfA gene may have to do with the secondary structure of its mRNA.

Gene inactivation studies. The phenotypes of mutants with each open reading frame inactivated were determined by insertion of a CAT cassette. The CAT cassette was inserted in the *spo0B* gene in both orientations, resulting in plasmids pJH4503 and pJH4507 (Fig. 3). Linearization of these plasmids and transformation into several recipients did not result in Cm^r colonies, suggesting that these configurations were lethal. We suspected that this result might have been due to the terminator carried by the CAT cassette. To test this possibility, we constructed a CAT cassette without the terminator (M. Perego and J. A. Hoch, manuscript in preparation), and a plasmid constructed with it, pJH4670, transformed sporulating strains to Cm^r and sporulation deficiency. These deletion-insertion strains had no additional properties not previously attributed to *spo0B* mutant strains.

The lethality of a terminator-bearing CAT cassette in the *spo0B* gene was most probably due to termination of the transcript for the OrfA. We have previously shown that

FIG. 3. Chromosomal insertion-deletion plasmids. The chloramphenicol resistance determinant was inserted in the restriction sites shown, with deletion of intervening material. The direction of transcription of the inserted CAT gene is indicated by the arrows.

inactivation of OrfA by an integrative vector did not give rise to transformed colonies, suggesting that OrfA coded for an essential gene (14). The present CAT cassette results lend support to our contention that *spo0B* and OrfA are part of the same operon (14). A CAT cassette construction within OrfA, pJH4504, was unable to transform recipients to Cm^r , further underscoring the possibility that the OrfA gene product is essential for growth.

CAT insertion-deletion constructions within the downstream open reading frames, pJH4508 and pJH4506, resulted in a simple phenylalanine requirement when Cm^r transformants were obtained. This tentatively identifies these genes as part of the phenylalanine biosynthetic pathway. It also eliminates a trivial possibility that CAT cassette insertion in OrfA was lethal only because the terminator prevented expression of genes downstream of it.

S1 protection experiments. Bouvier et al. (7) mapped the 5' end of the *spo0B* transcript to a position approximately 60 bp upstream of the translation initiation site for *spo0B* (Fig. 2). In our S1 protection experiments to define the site of transcription initiation, we used two probes from within the *spo0B* gene labeled at the *Hind*III or *Pst*I restriction sites (Fig. 1). S1-protected fragments from experiments with these probes were 160 and 180 bp from the *Hind*III and *Pst*I sites, respectively (Fig. 4). Furthermore, the protected fragments were run against a sequencing ladder, with the result that the same start site of transcription as reported by Bouvier et al. (7) was found (data not shown). No evidence was obtained for multiple promoters of the *spo0B* operon.

In our previous studies we showed through S1 protection experiments that there was no evidence for transcript initiation between the *spo0B* gene and the OrfA gene (14). We confirmed these results by S1 protection experiments with a fragment labeled at the *Bgl*II site within OrfA (Fig. 1). The S1-protected fragment in this experiment was 765 bases, which corresponds within 5 bases to the start site of *spo0B* transcription mapped above (Fig. 5).

In an experiment to map the 3' end of the *spo0B* mRNA, a 3.4-kbp *Hind*III-*Pst*I fragment was labeled at the *Hind*III site (Fig. 2, nucleotide 188) within the *spo0B* gene. The protected fragment in this case was 1.8 kbp, which corre-

FIG. 4. S1 mapping of the start of transcription of *spo0B*. Lane A shows the molecular weight standards labeled with ^{32}P as described in Materials and Methods. The sizes of the fragments are given in kilobase pairs. S1 protection experiments: lane B, *Hind*III-*Bgl*II probe with S1 nuclease; lane C, *Hind*III-*Bgl*II probe with S1 nuclease and mRNA; lane D, *Pst*I-*Eco*RI probe with S1 nuclease; lane E, *Pst*I-*Eco*RI probe with S1 nuclease and mRNA.

FIG. 5. S1 mapping of the start of transcription of OrfA. S1 protection experiments: lane A, *Bgl*I-*Bgl*II probe alone; lane B, *Bgl*II-*Bgl*II probe with S1 nuclease; lane C, *Bgl*II-*Bgl*II probe with S1 nuclease and mRNA. Lane D contains the molecular weight standards of Fig. 4, lane A, and two additional labeled fragments of 0.85 and 0.72 kbp to increase the accuracy in this region.

sponds to termination at the stem-loop structure within the intergenic region between OrfA and *pheB* (data not shown). Thus, all the available evidence suggests that *spo0B* and OrfA constitute an operon.

The position of the start site for transcription of the *pheB* and *pheA* genes was determined by S1 protection experiments. A *Bgl*I-*Nco*I fragment (Fig. 1) was labeled at the *Nco*I site and used in these experiments. The protected fragment was about 225 bp, which would locate the transcription start site within the stem-loop structure preceding the *pheB* gene (Fig. 6). There is no homology upstream of this region to known promoters. This suggests that the *phe* genes may be transcribed from the *spo0B* promoter and that the putative start site observed may be a site of RNA processing.

Identification of the *pheA* gene product. The *pheA* gene was shown to code for prephenate dehydratase by homology to the sequences of known prephenate dehydratase enzymes (Fig. 7). The *B. subtilis* enzyme seems to be like the *Corynebacterium glutamicum* prephenate dehydratase (16) in that it is not fused to chorismate mutase as in *E. coli* (23). Of 285 residues of the *B. subtilis* enzyme, 51 (18%) are identical in both *C. glutamicum* and *E. coli* and 39 additional residues are conserved substitutions. Many more homologies could be seen if the enzymes were compared in a pairwise fashion.

The presumed product of the *pheB* gene is chorismate mutase. The deduced product of this gene, however, has no

FIG. 6. S1 mapping of the start of transcription of *pheB*. S1 protection experiments: lane A, *Nco*I-*Bgl*II probe alone; lane B, *Nco*I-*Bgl*II probe with S1 nuclease; lane C, *Nco*I-*Bgl*II probe with S1 nuclease and mRNA. Lane D contains the molecular weight standards of Fig. 4, lane A.

significant homology to the chorismate mutase portion of the bifunctional chorismate mutase-prephenate dehydratase of *E. coli* (23). In addition, it shows no significant homology to the tyrosine-specific chorismate mutase-prephenate dehydratase of *E. coli* (23) or of *B. subtilis* (18). The chorismate mutase portions of these last two enzymes, however, show very little homology. The only evidence that *pheB* is an enzyme involved in phenylalanine biosynthesis is the Phe requirement of CAT insertions in it. These experiments were carried out with the terminator-bearing CAT cassette, and so it is still possible that the observed Phe requirement is a termination effect on *pheA* transcription.

Location of a consensus GTP-binding domain in OrfA. Examination of the amino acid sequence of the deduced OrfA gene product revealed the presence of a consensus GTP-binding sequence. The sequence found is shown in comparison with other known GTP-binding proteins in Table 1. The OrfA sequence is similar to the other GTP-binding sites in sequence and spacing (10). It resembles rat G_i protein or bovine α -transducin in having two possible region 2 phosphoryl-binding sequences, DXGX. The region 1 phosphoryl-binding sequence is very similar to the RAS-like *E. coli* protein Era. Only two residues in the sequence GXPX-VGK differ, and the four following residues, STLL, are identical in the two proteins. Furthermore, the P in position 3 of both proteins is the only known substitution for G in RAS proteins that does not lead to oncogenic activity. Thus, the putative GTP-binding sequences of Era, the OrfA gene product, and the RAS proteins show many similarities.

GTP binding to the OrfA protein. The suspected GTP-

C. glut	M S D A P T V V A Y L G P A G T G T E E A L Y K F A D A G V
E. coli	I N P H S A R I A F L G P K G S Y S H L A A R Q Y A A R H -
B. sub	M K V G Y L G P A A T F T H L A V - - - - S S C
C. glut	F G D G E I E Q L P A - K S P Q E - A V D A V R H G T A Q F
E. coli	F - E Q F I E S G C A - K F A - D - I F N Q V E T G Q A D Y
B. sub	F Q N G - A E H V - A Y R T I P E C I D A A V - A G E V D F
C. glut	A V V A I E N F V D G P V T P T F D A L D Q G S N V Q I I A
E. coli	A V V P I E N T S S G A I N D V Y D - L L Q H T S L S I V G
B. sub	A F V P L E N A L E G S V N L T I D Y L I H E Q P L P I V G
C. glut	E E E L D I A F S I M V R P G T S - - - L A D V K T L A T H
E. coli	E M T L T I D H C L L V S G T T D - - - L S T I N T V Y S H
B. sub	E M T L P I H Q H L L V H P S R E N A W K E L D K - I Y S H
C. glut	P V G Y Q Q V - K N W M A - T T I P D A M Y L S A S S N G A
E. coli	P Q P F Q Q C S K F L N R - Y P H W K I E Y - - T E S T S A
B. sub	S H A I A Q C H K F L H R H F P S V P Y E Y - - A N S T G A
C. glut	G A Q M V A E G T A D A A A A - - P S T A A E L F G L E R L
E. coli	A M E K V A Q A K S P H V A A L G S E A G G T L Y G L Q V L
B. sub	A A K F V S D H P E L N I G V I A N D M A A S T Y E L K I V
C. glut	V D D V A D V R G A R T R F V A V Q - - - - A Q A - A V S
E. coli	E R I E A N Q R Q N F T R F V V L A - - - - R K A I N V S
B. sub	K R D I Q D Y R D N H T R F V I L S P D E N I S F E V N - S
C. glut	E P T G H D R T S V I F S L P N V - P - G S L V R A L N E F
E. coli	D Q F - P A K T T L L M A T G Q Q - A - G A L V E A L - - L
B. sub	K L S S R P K T T L M V M L P Q D D Q S G A L H R V L S A F
C. glut	A I R G V D L - - T R I E S R P T R K V F G T Y R F H L D I
E. coli	V L R N H N L I M T R L E S R P I N G N P W E E M F Y L D I
B. sub	S W R N L N L - - S K I E S R P T K T G L G H Y F F I I D I
C. glut	S G H I R D I P V A E A L R A L H L Q A E E L V F V G S W P
E. coli	Q A N L E S A E M Q K A L K E L G E I T R S M K V L G C Y P
B. sub	E K A F D D V L I P G A N Q E L E E - A - - - - L G C K V
C. glut	S N R A E D S T P Q T D Q L A K L H K A D
E. coli	S E N V V P V D P T
B. sub	- - R L L G A Y - Q S Y Q L

FIG. 7. Homology between prephenate dehydratases. The amino acid sequence of prephenate dehydratases of *C. glutamicum* (16), *E. coli* (23) and the *B. subtilis* *pheA* gene product are aligned for maximum homology. Vertical columns with identical residues in all three proteins are indicated by two asterisks. One asterisk indicates a very conservative substitution in one or more proteins. The *E. coli* sequence shown begins with residue 99.

binding activity of the OrfA gene product led us to determine whether this protein was bound by Cibacron Blue F3GA, a compound with binding affinity for enzymes containing a dinucleotide fold (28). Crude extracts prepared from strains overproducing the OrfA protein were passed over an immobilized Cibacron Blue F3GA column (Affi-Gel Blue), and the bound protein was examined by gel electrophoresis (Fig. 8). The Affi-Gel Blue column quantitatively removed the OrfA protein from the crude extract, giving a preparation of OrfA protein estimated to be over 90% pure. Extracts from cells

carrying the vector with no insert did not produce or lead to purification of a protein similar to OrfA (Fig. 8).

The Affi-Gel Blue-purified OrfA protein was examined for its ability to bind GTP by incubation with [α - 32 P]GTP and by cross-linking under UV light. Cross-linking of GTP to the OrfA protein was detected by autoradiography of OrfA protein separated by gel electrophoresis. The results of this analysis showed that GTP could be cross-linked to the OrfA protein (Fig. 9) and, hence, that the OrfA protein should be classified as a GTP-binding protein.

DISCUSSION

Results from S1 protection experiments indicated that transcription from the *spo0B* promoter extends through OrfA immediately downstream of *spo0B*. Many of these transcripts appear to terminate at a stem-loop structure located between OrfA and two genes involved in phenylalanine biosynthesis, *pheB* and *pheA*. Transcripts for these two genes appear to originate at this same stem-loop structure. Several possible explanations of these results were entertained. The mRNA for the *phe* genes may originate from an independent transcript whose promoter is located within the region of the stem-loop structure. However, we could find no homology to any known promoters in this region. The *phe* gene transcript may be processed from a longer transcript, or the stem-loop structure may be a terminator for a transcript that could include all of these genes in the presence of an antiterminator activity. Thus, these *phe* genes may not be controlled at the level of transcription initiation but rather by antitermination. This antitermination activity might be the role of the *pheB* gene product whose function is unknown and whose sequence has no homology to chorismate mutase, the missing enzyme in this pathway. It is interesting that the carboxy-terminal region of OrfA has a repeating glutamic acid-phenylalanine (EF) motif. Perhaps these phenylalanine codons constitute ribosome pause sites when levels of charged phenylalanine tRNAs are low and therefore could be involved in an antitermination mode of regulation of expression of the *pheB* and *pheA* genes. Certainly, more sophisticated experiments are required before the observed results are fully understood.

The *pheA* gene product is clearly prephenate dehydratase, as determined by its homology to authentic prephenate dehydratase of *C. glutamicum* and *E. coli*. It resembles the *C. glutamicum* enzyme in that it is not fused to chorismate mutase as in *E. coli*, but it is homologous to both enzymes.

The product of OrfA is most probably a GTP-binding protein, and therefore this gene has been designated *obg* or *spo0B*-associated GTP-binding protein. GTP binding to the purified *obg* protein was shown in vitro by UV cross-linking experiments. This result, in conjunction with the strong consensus GTP-binding site observed within the deduced amino sequence of the gene, provides evidence that the binding of GTP must play a role in whatever function this protein plays in the cell. This function is clearly essential to the viability of the cell, since we have been unable to insert ionally inactivate this gene.

The GTP-binding site of the *obg* protein is very similar to the same site in Era, an *E. coli* RAS-like protein which is also essential for growth (1, 25). Era possesses GTPase activity like that of RAS proteins. Both GTP-binding and GTPase activities are functions of the amino-terminal domains of Era and RAS proteins, and in this structural feature they differ from the OrfA protein. The GTP-binding site of

TABLE 1. Comparison of GTP-binding domain sequences

Protein and source	Consensus sequences				Guanine-binding site (NKXD)
	Phosphoryl-binding sequences				
	GXXXXGK		DXXG		
	Sequence	Residue separation (40 to 80 or 130 to 170) ^a	Sequence	Residue separation (40 to 80) ^a	
EF-Tu (<i>E. coli</i>) ^b	GHVDHGK	56	DCPG	51	NKCD
EF-G (<i>E. coli</i>)	AHIDAGK	65	DTPG	50	NKMD
LepA (<i>E. coli</i>)	AHIDHGK	60	DTPG	50	NKID
IF-2 (<i>E. coli</i>)	GHVDHGK	40	DTPG	50	NKID
Era (<i>E. coli</i>)	GRPNVGK	40	DTPG	50	NKVD
H- <i>ras</i> , N- <i>ras</i> (human)	GAGGVGK	41	DTAG	55	NKCD
G _s protein (rat)	GAGESGK	170	DVGG	65	NKQD
G _i protein (rat)	GAGESGK	77	DLSG	143	NKKD
	GAGESGK	155	DVGG	65	
α-Transducin (cattle)	GAGESGK	104	DSAG	115	NKKD
	GAGESGK	154	DVGG	65	
OrfA (<i>B. subtilis</i>)	GFPSVGK	41	DLPG	67	NKMD
	GFPSVGK	74	DMSG	34	

^a Number of amino acid residues separating portions of the consensus sequences.

^b The GTP-binding sequences of these proteins were taken from references 1 and 10.

the *obg* protein is located in the middle of the protein (Fig. 2). The *obg* protein can, for convenience, be divided into three equal domains (Fig. 2). The amino-terminal domain is particularly glycine rich and contains a region from residue 69 to residue 148 with high homology to the alpha 1 chain of mammalian collagen (15). This similarity to collagen may indicate that the *obg* protein interacts with structural proteins or at least favors their environment. Thus, the collagenlike amino domain of this protein may serve to anchor the *obg* protein at a particular location in the cell. The carboxy-terminal domain of the *obg* protein is not particularly hydrophobic and therefore may not be membrane associated like the RAS proteins. Furthermore, it lacks the obligatory cysteine residue acylated by palmitic acid (3). We could find

no significant homology between the amino acid sequence of the carboxy-terminal domain and any other known protein. Its role remains elusive. It will be of interest to study the location of this protein within the cell and to see whether it possesses GTPase activity like the RAS proteins and Era.

Our vision of the role of the *spo0* gene products in the cell is that they are part of a multicomponent complex which senses the metabolic state of the cell, processes this information, and drives the cell toward either vegetative growth or the initiation of sporulation. In this role, the complex may interact with many essential proteins with functions in septation and cell division as well as, perhaps, DNA synthesis. The finding of a gene for an essential protein in an operon structure with the *spo0B* gene was not totally unexpected and is consistent with this viewpoint. However, there is no evidence to link the function of the *spo0B* gene to that of the *obg* protein. Furthermore, being in the same operon does not constitute prima facie evidence for unity in function of the

FIG. 8. Binding of the OrfA gene product to Cibacron Blue. Polyacrylamide gel-separated fractions from Affi-Gel Blue experiments described in Materials and Methods are shown. The positions of molecular mass standards are indicated in kilodaltons (Kd). Lanes: A, IPTG-induced crude extract from *E. coli* JM109 bearing pJH4604; B, flowthrough fraction of pJH4604 crude extract applied to Affi-Gel Blue column; C, elution with 0.3 M KCl of proteins bound to Affi-Gel Blue in the pJH4604 crude extract; D, IPTG-induced crude extract from JM109 bearing the pKK223-3 plasmid without an insert.

FIG. 9. Photoaffinity cross-linking of GTP to the OrfA protein. Lanes: A, [α -³²P]GTP-UV-cross-linked proteins in Affi-Gel Blue-purified fraction from pKK223-3; B, [α -³²P]GTP-UV-cross-linked proteins in Affi-Gel Blue-purified fraction from pJH4604; C, protein fraction from lane B incubated with [α -³²P]GTP in the absence of UV light. Molecular mass standards are as in Fig. 8.

components. The tantalizing feature of this particular association of a gene for a GTP-binding protein with a *spo0* gene is that there is a large body of data suggesting that fluctuations in the GTP content of the cell are the primary cause of the initiation of sporulation in bacteria and yeasts (17). If this long-standing hypothesis is true, the *obg* protein may form the bridge between GTP levels and sporulation.

ACKNOWLEDGMENTS

This research was supported, in part, by Public Health Service grant GM19416 from the National Institute of General Medical Sciences.

We thank a reviewer of the manuscript for pointing out the possible ribosome pause sites in *obg*.

LITERATURE CITED

- Ahnn, J., P. E. March, H. E. Takiff, and M. Inouye. 1986. A GTP-binding protein of *Escherichia coli* has homology to yeast RAS proteins. *Proc. Natl. Acad. Sci. USA* **83**:8849–8853.
- Anagnostopoulos, C., and J. Spizizen. 1961. Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.* **81**:741–746.
- Barbacid, M. 1987. *ras* genes. *Annu. Rev. Biochem.* **56**:779–827.
- Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1-endonuclease digested hybrids. *Cell* **12**:721–732.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**:1513–1523.
- Bonamy, C., and J. Szulmajster. 1982. Cloning and expression of *Bacillus subtilis* spore genes. *Mol. Gen. Genet.* **188**:202–210.
- Bouvier, J., P. Stragier, C. Bonamy, and J. Szulmajster. 1984. Nucleotide sequence of the *spo0B* gene of *Bacillus subtilis* and regulation of its expression. *Proc. Natl. Acad. Sci. USA* **81**:7012–7016.
- Brendel, V., and E. N. Trifonov. 1984. A computer algorithm for testing potential prokaryotic terminators. *Nucleic Acids Res.* **12**:4411–4428.
- Dagert, M., and S. D. Ehrlich. 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* **6**:23–28.
- Dever, T. E., M. J. Glynias, and W. C. Merrick. 1987. GTP-binding domain: three consensus elements with distinct spacing. *Proc. Natl. Acad. Sci. USA* **84**:1814–1818.
- Eriksson, S., I. W. Caras, and D. W. Martin, Jr. 1982. Direct photoaffinity labeling of an allosteric site on subunit protein M1 of mouse ribonucleotide reductase by dTTP. *Proc. Natl. Acad. Sci. USA* **79**:81–85.
- Ferrari, F. A., D. Lang, E. Ferrari, and J. A. Hoch. 1982. Molecular cloning of the *spo0B* sporulation locus in bacteriophage lambda. *J. Bacteriol.* **152**:809–814.
- Ferrari, F. A., A. Nguyen, D. Lang, and J. A. Hoch. 1983. Construction and properties of an integrable plasmid for *Bacillus subtilis*. *J. Bacteriol.* **154**:1513–1515.
- Ferrari, F. A., K. Trach, and J. A. Hoch. 1985. Sequence analysis of the *spo0B* locus reveals a polycistronic transcription unit. *J. Bacteriol.* **161**:556–562.
- Fietzek, P. P., P. Wendt, I. Kell, and K. Kuhn. 1972. The covalent structure of collagen: amino acid sequence of alpha 1-CB3 from calf skin collagen. *FEBS Lett.* **26**:74–76.
- Follettie, M. T., and A. J. Sinskey. 1986. Molecular cloning and nucleotide sequence of the *Corynebacterium glutamicum pheA* gene. *J. Bacteriol.* **167**:695–702.
- Freese, E., E. B. Freese, E. R. Allen, Z. Olenpska-Beer, C. Orrego, A. Varma, and H. Wabiko. 1985. Metabolic initiation of spore development, p. 194–202. In J. A. Hoch and P. Setlow (ed.), *Molecular biology of microbial differentiation*. American Society for Microbiology, Washington, D.C.
- Henner, D. J., L. Band, G. Flaggs, and E. Chen. 1986. The organization and nucleotide sequence of the *Bacillus subtilis* *hisH*, *tyrA*, and *aroE* genes. *Gene* **49**:147–152.
- Hirochika, H., Y. Kobayashi, F. Kawamura, and H. Saito. 1981. Cloning of sporulation gene *spo0B* of *Bacillus subtilis* and its genetic and biochemical analysis. *J. Bacteriol.* **146**:494–505.
- Hoch, J. A. 1976. Genetics of bacterial sporulation. *Adv. Genet.* **18**:69–99.
- Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114**:193–197.
- 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.
- Hudson, G. S., and B. E. Davidson. 1984. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K12. *J. Mol. Biol.* **180**:1023–1051.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680–685.
- March, P. E., C. G. Lerner, J. Ahnn, X. Cui, and M. Inouye. 1988. The *Escherichia coli* Ras-like protein (Era) has GTPase activity and is essential for cell growth. *Oncogene* **2**:539–544.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499–560.
- Stein, R. B., P. S. Robinson, and E. M. Scolnick. 1984. Photoaffinity labeling with GTP of viral p21 ras protein expressed in *Escherichia coli*. *J. Virol.* **50**:343–351.
- Thompson, S. T., K. H. Cass, and E. Stellwagen. 1975. Blue dextran-sepharose: an affinity column for the dinucleotide fold in proteins. *Proc. Natl. Acad. Sci. USA* **72**:669–672.
- Trach, K., J. W. Chapman, P. Piggot, D. LeCoq, and J. A. Hoch. 1988. Complete sequence and transcriptional analysis of the *spo0F* region of the *Bacillus subtilis* chromosome. *J. Bacteriol.* **170**:4194–4208.
- 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.