JOURNAL OF BACTERIOLOGY, Dec. 1990, p. 7211–7226
0021-9193/90/127211-16$02.00/0

Vol. 172, No. 12

An Apparent *Bacillus subtilis* Folic Acid Biosynthetic Operon Containing *pab*, an Amphibolic *trpG* Gene, a Third Gene Required for Synthesis of *para*-Aminobenzoic Acid, and the Dihydropteroate Synthase Gene

JAMES SLOCK,[1] DONALD P. STAHLY,[2]* CHUN-YA HAN,[2] ERICH W. SIX,[2] AND IRVING P. CRAWFORD[2]†

Department of Microbiology, University of Iowa, Iowa City, Iowa 52242,[2] and Department of Biology, King's College, Wilkes-Barre, Pennsylvania 18711[1]

Received 7 May 1990/Accepted 12 September 1990

McDonald and Burke (J. Bacteriol. 149:391–394, 1982) previously cloned a sulfanilamide-resistance gene, *sul*, residing on a 4.9-kb segment of *Bacillus subtilis* chromosomal DNA, into plasmid pUB110. In this study we determined the nucleotide sequence of the entire 4.9-kb fragment. Genes identified on the fragment include *pab*, *trpG*, *pabC*, *sul*, one complete unidentified open reading frame, and one incomplete unidentified open reading frame. The first three of these genes, *pab*, *trpG*, and *pabC*, are required for synthesis of *p*-aminobenzoic acid. The *trpG* gene encodes an amphibolic glutamine amidotransferase required for synthesis of both *p*-aminobenzoate and anthranilate, the latter an intermediate in the tryptophan biosynthetic pathway. The *pabC* gene may encode a *B. subtilis* analog of enzyme X, an enzyme needed for *p*-aminobenzoate synthesis in *Escherichia coli*. The *sul* gene probably encodes dihydropteroate synthase, the enzyme responsible for formation of 7,8-dihydropteroate, the immediate precursor of folic acid. All six of the cloned genes are arranged in a single operon. Since all four of the identified genes are needed for folate biosynthesis, we refer to this operon as a folic acid operon. Expression of the *trpG* gene is known to be negatively controlled by tryptophan. We propose that this regulation is at the level of translation. This hypothesis is supported by the finding of an apparent Mtr-binding site which overlaps with the *trpG* ribosome-binding site.

Six of the seven tryptophan pathway genes in *Bacillus subtilis* are clustered in the *trpEDCFBA* operon (1, 5) found at 205 degrees on the *B. subtilis* map (36). These genes are coordinately regulated (13) by a recently revealed, unusual mechanism (22, 41), and the complete nucleotide sequence of the operon has been determined (12). The seventh *trp* gene is not linked to this operon; it participates in both folate and tryptophan syntheses and is located near the *pab* gene at 10 degrees on the map (36). This gene was originally designated *trpX* (18), but here it will be called *trpG* to conform to the terminology used for similar amphibolic genes in *Acinetobacter calcoaceticus* (40) and *Pseudomonas acidovorans* (4).

TrpG participates in the first step in the synthesis of tryptophan, the conversion of chorismate and glutamine to anthranilate, glutamate, and pyruvate. A reaction using the same substrates provides the *p*-aminobenzoate moiety for folate synthesis. Recently, it was discovered that in *Escherichia coli* three genes are involved in *p*-aminobenzoate synthesis instead of two as in anthranilate synthesis (31). Two of the three show sequence similarities to the large and small subunits of anthranilate synthase (10, 20); the third, which apparently performs the final aromatization reaction, has no counterpart in anthranilate synthase. The large and small subunits of anthranilate and *p*-aminobenzoate synthases are related in function as well as sequence, the former binding chorismate and the latter binding glutamine (10, 20). In each enzyme complex the large subunit can function independently, given high concentrations of ammonia and a high pH.

In the three organisms mentioned above (*A. calcoaceticus*, *P. acidovorans*, and *B. subtilis*), a single gene encodes the small glutamine amidotransferase subunit of both anthranilate synthase and *p*-aminobenzoate synthase. In *B. subtilis* the large subunits for these enzymes are specified by the *trpE* and *pab* genes, respectively. In this report we show that in *B. subtilis* the gene for the common small subunit, *trpG*, lies immediately downstream from, and is probably cotranscribed with, the *pab* gene. A gene immediately downstream from *trpG*, referred to as *pabC*, may encode the third polypeptide required for the synthesis of *p*-aminobenzoate.

Kane (17) reported that the synthesis of TrpG in *B. subtilis* is negatively regulated by tryptophan. In *A. calcoaceticus* and *P. acidovorans*, the amphibolic *trpG* genes are located in an operon with two other *trp* genes (4, 40). However, in *B. subtilis trpG* is located in an operon with *pab*, making its regulation paradoxical. The DNA sequence of the region between *pab* and *trpG* of *B. subtilis*, reported in this paper, suggests a possible explanation of this paradox.

McDonald and Burke (29) had previously cloned a sulfanilamide resistance gene, *sul*, residing on a 4.9-kb segment of *B. subtilis* chromosomal DNA, into plasmid pUB110. After determining that McDonald and Burke's plasmid, pK0101, would complement Kane's *trpG* mutant, we determined the nucleotide sequence of the entire fragment. In addition to containing the *pab*, *trpG*, and *pabC* genes referred to above, the DNA fragment also has a gene immediately downstream from *pabC* that is probably the *sul* gene. The fragment also contains an unidentified gene, *ORF1*, and part of another unidentified gene, *ORF2*. Evidence is presented indicating that *sul* probably encodes dihydropteroate synthase, the enzyme responsible for condensation of *p*-aminobenzoate

---

* Corresponding author.

† Deceased 8 October 1989.

TABLE 1. Bacterial strains and plasmids

Bacterial strain or plasmid	Description or genotype[a]	Reference or source[b]
Bacterial strain		
B. subtilis 1A2	Prototrophic	BGSC
B. subtilis 1A491	*trpG7 dfrA24 trk-24*	BGSC
B. subtilis ASB342(pK0101)	*trpC2* (Km^r *pab*$^+$ *trpG*$^+$ *pabC*$^+$ Sul^r *ORF1*$^+$)	W. F. Burke; 29
B. subtilis 1A491(pK0101)	Same as 1A491 but (Km^r *pab*$^+$ *trpG*$^+$ *pabC*$^+$ Sul^r *ORF1*$^+$)	This study
B. subtilis 1A491(pJS3)	Same as 1A491 but (Km^r Em^r *trpG*$^+$)	This study
B. subtilis 1A491(pJS4)	Same as 1A491 but (Km^r Em^r *pab*$^+$ *trpG*$^+$)	This study
B. subtilis 1A2ΩpJS6	Same as 1A2 but Cm^r *trpG*	This study
B. subtilis 1A2ΩpCH1	Same as 1A2 but Cm^r *pab*	This study
B. subtilis 1A2ΩpCH2	Same as 1A2 but Cm^r *pab*	This study
B. subtilis 1A2ΩpCH3	Same as 1A2 but Cm^r *pabC*	This study
B. subtilis 1A2ΩpCH4	Same as 1A2 but Cm^r *pabC*	This study
B. subtilis 1E27(pBD9)	*thr-5 trpC2* (Em^r Km^r)	BGSC
B. larvae NRRL B-3555(pC194)	(Cm^r)	2
E. coli JM109	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* λ^- Δ(*lac-proAB*)(F′ *traD36 proAB lacI*^{q}*Z* Δ*M15*)	43
E. coli JM109(pJS1)	Same as JM109 but (Ap^r)	This study
E. coli JM109(pJS2)	Same as JM109 but (Ap^r *trpG*$^+$ *pabC*$^+$ Sul^r *ORF1*$^+$)	This study
E. coli JM109(pJS5)	Same as JM109 but (Ap^r *trpG*$^+$)	This study
E. coli JM109(pJS6)	Same as JM109 but (Ap^r Cm^r)	This study
Plasmid		
pK0101	Km^r *pab*$^+$ *trpG*$^+$ *pabC* Sul^r *ORF1*$^+$	29
pJS1	Ap^r	This study
pJS2	Ap^r *trpG*$^+$ *pabC*$^+$ Sul^r *ORF1*$^+$	This study
pJS3	Km^r Em^r *trpG*$^+$	This study
pJS4	Km^r Em^r *pab*$^+$ *trpG*$^+$	This study
pJS5	Ap^r *trpG*$^+$	This study
pJS6	Ap^r, Cm^r cassette in *trpG* of pJS5	This study
pCH1	Ap^r, Cm^r cassette (orientation same as operon) in *pab* of pJS1	This study
pCH2	Ap^r, Cm^r cassette (orientation opposite of operon) in *pab* of pJS1	This study
pCH3	Ap^r, Cm^r cassette (orientation same as operon) in *pabC* of pJS2	This study
pCH4	Ap^r, Cm^r cassette (orientation opposite of operon) in *pabC* of pJS2	This study
pBD9	Km^r Em^r	11
pC194	Cm^r	15, 16
pUC18	Ap^r	BRL

[a] Genotype in parentheses is due to genes carried by the resident plasmid.

[b] BGSC, Bacillus Genetic Stock Center; BRL, Bethesda Research Laboratories.

with 6-CH$_2$-OH-7,8-dihydropterin pyrophosphate to form 7,8-dihydropteroate, the immediate precursor of folic acid. Thus, at least four genes concerned with folic acid biosynthesis are clustered on the *B. subtilis* chromosome.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are shown in Table 1.

Media and growth conditions. In general, *B. subtilis* and *E. coli* strains were grown on L agar or L broth (30). For strains possessing plasmids, the following amounts of antibiotics were added unless otherwise specified: chloramphenicol (Cm), 25 μg/ml; kanamycin (Km), 25 μg/ml; ampicillin (Ap), 100 μg/ml; erythromycin (Em), 10 μg/ml; and sulfanilamide, 1,500 μg/ml. Medium containing 50 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) per ml was used to detect transformants containing plasmids with an insert in the *lacZ* gene. GM1 and GM2 media (35) were used for transformation of *B. subtilis* cells.

TrpG$^+$ transformants of *B. subtilis* 1A491 (a strain that contains a leaky *trpG7* mutation) were recognized by growth on a medium lacking *p*-aminobenzoate. The selective medium (SM) contained the following substances in the proportions indicated: Bott and Wilson (BW) salts (3), 100 ml; 50% glucose, 1 ml; 1 M MgSO$_4$, 0.1 ml; tryptophan, 2.0 mg; phenylalanine, 2.0 mg; tyrosine, 1.0 mg; and Difco purified agar, 1.5 g. Complete medium was SM supplemented with 1.0 mg of *p*-aminobenzoate per 100 ml. BW salts consisted of 1.24% K$_2$HPO$_4$, 0.76% KH$_2$PO$_4$, 0.1% trisodium citrate, and 0.6% (NH$_4$)$_2$SO$_4$ (adjusted to pH 6.7). After transformation and before plating on SM or complete medium, cells were grown overnight in an amino acid-supplemented medium at 37°C with shaking. This medium contained 100 ml of BW salts, 1.0 ml of 50% glucose, 0.1 ml of 1 M MgSO$_4$, and 5.0 ml of BW amino acids without tryptophan. BW amino acids contained valine, lysine, threonine, glycine, aspartic acid, methionine, histidine, tryptophan, and arginine, each at 500 μg/ml.

Media used for testing growth requirements of the prototrophic *B. subtilis* 1A2 and strains containing insertionally inactivated genes are specified in the legend to Fig. 7. Inocula for these experiments were prepared by growth in a medium containing the following substances in the proportions indicated: 100 ml of BW salts; 1 ml of 50% glucose; 0.1 ml of 1 M MgSO$_4$; 5 ml of BW amino acids; and 1 ml of 10% yeast extract. In testing for growth requirements, it was essential before plating or inoculating broth media to wash the cells thoroughly to remove residual tryptophan and/or *p*-aminobenzoic acid. Cell pellets were routinely washed by repeated (10 times) suspension in the appropriate basal medium and centrifugation.

Transformation. *E. coli* was transformed by the method of

Cohen et al. (6). *B. subtilis* was transformed as described by Piggot et al. (35).

Plasmid preparation. Plasmid DNA was prepared from 1.5-ml cultures of *B. subtilis* and *E. coli* as described by Rodriguez and Tait (38). When the DNA was to be digested with restriction enzymes, it was treated with phenol-chloroform-isoamyl alcohol (50:48:2, vol/vol/vol) followed by chloroform-isoamyl alcohol (24:1, vol/vol) (26) after the sodium acetate neutralization step.

Restriction enzyme analyses and agarose gel electrophoresis. Restriction endonuclease digestion was carried out under conditions recommended by the suppliers. The enzymes were obtained from New England BioLabs or Bethesda Research Laboratories. Analytical agarose gel electrophoresis was performed at room temperature, using agarose (EM Laboratories, Inc.) concentrations of between 0.8 and 1.0%. Preparative agarose gel electrophoresis was performed at 4°C, using 0.7% low-gelling-temperature agarose (SeaPlaque Agarose; FMC BioProducts). DNA fragment size was estimated by comparison with a "1-kb ladder" supplied by Bethesda Research Laboratories.

Cloning of DNA fragments into plasmid vectors. DNA fragments and linearized plasmid vectors were purified by agarose gel electrophoresis in low-gelling-temperature agarose before ligation. Standard procedures (26) were used for ligation.

DNA sequencing and sequence analysis. DNA fragments were labeled with ^{32}P by one of two methods: 3′ fill-in labeling with the large fragment of DNA polymerase I and the appropriate radioactive deoxynucleotide triphosphate (44); or T4 polynucleotide kinase and [γ-^{32}P]ATP (28). Sequencing reactions were carried out by the procedure of Maxam and Gilbert (28) and developed by 8% urea–polyacrylamide gel electrophoresis by the method of Sanger and Coulson (39). Figure 2 shows the sequencing strategy that was used. Sequence data were analyzed with the aid of the PCS computer program (24). Amino acid sequences of the cloned genes were compared with sequences of other related proteins after appropriate alignment. Initial alignments were made by use of the FASTA program (33). To estimate similarity between two sequences, we used the normalized alignment score of Doolittle (8). With this method, the number of identical residues is multiplied by 10 (20 for cysteines) and the number of gaps is multiplied by −25. The score is normalized by dividing by the average length of the two sequences and multiplying by 100. Doolittle (8) presented graphic relationships between normalized alignment score and number of residues in compared sequences that permit determination of the relative certainty of relatedness.

RESULTS

Physical map of pK0101 and subcloning the two *Eco*RI fragments of the pK0101 insert. Figure 1 shows a physical map of pK0101, the plasmid constructed by McDonald and Burke (29) consisting of pUB110 with a 4.9-kb insert of chromosomal DNA derived from *B. subtilis* VB157, a sulfanilamide-resistant strain. pK0101 contains the *sul* gene, since transformation of the plasmid into various wild-type strains of *B. subtilis* rendered them resistant to sulfanilamide (29). Figure 1 shows restriction enzyme sites that were used in subcloning. Also shown are the locations of *pab*, *trpG*, *pabC*, *sul*, one complete open reading frame, and one partial open reading frame, as determined in this study.

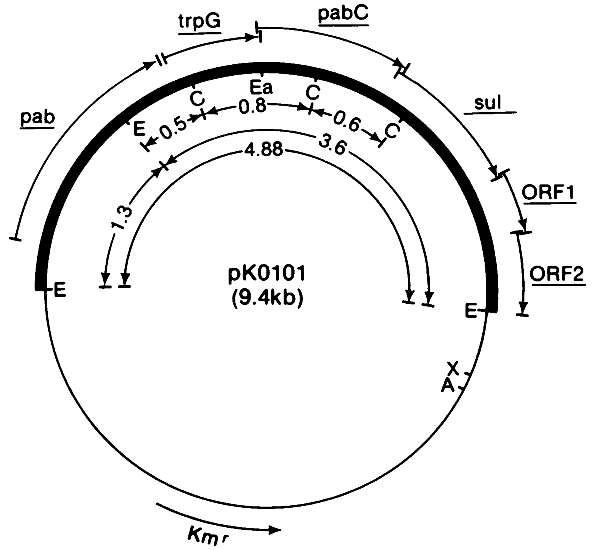

FIG. 1. Physical map of pK0101. The plasmid consists of a 4.9-kb fragment of chromosomal DNA from *B. subtilis* VB157 (thick line) inserted into the *Eco*RI site of pUB110 (thin line) (29). Genes identified in this study are shown. Note the overlap of all genes except *pab* and *trpG*. A, *Ava*I; C, *Cla*I; Ea, *Eag*I; E, *Eco*RI; X, *Xba*I.

Plasmid pK0101, isolated from *B. subtilis* ASB342 (pK0101), was cut with *Eco*RI. The 1.3- and 3.6-kb fragments derived from the insert were separately cloned into the *Eco*RI site of pUC18 (43) for ease of sequencing. Plasmids derived from transformants of *E. coli* JM109 containing the 1.3- and 3.6-kb *Eco*RI fragments were named pJS1 and pJS2, respectively.

Complementation of the *trpG* mutation in *B. subtilis* 1A491 by the presumed *trpG* gene in pK0101. The first question addressed was whether pK0101 would complement the *trpG* mutation in *B. subtilis*. *B. subtilis* 1A491 is a glutamine amidotransferase-negative mutant containing a mutation designated *trpX7* by Kane et al. (18). Hereafter we refer to this mutation as *trpG7*. This strain is a leaky mutant in which we were able to demonstrate only a partial growth requirement for *p*-aminobenzoate in the presence of tryptophan. When washed cells were plated on SM, no colonies were seen after 24 h at 37°C, but minute colonies appeared after 36 h of incubation. When the mutant was plated on SM supplemented with *p*-aminobenzoate, small colonies appeared after 24 h of incubation. Prototrophic *B. subtilis* 1A2 gave small colonies after 24 h of incubation when plated on either of the above media.

Plasmid pK0101 was transformed into competent cells of *B. subtilis* 1A491, the *trpG* mutant. The resultant *B. subtilis* 1A491(pK0101) grew on plates in a manner identical to that of the prototrophic strain 1A2 described above. Thus, the cloned fragment did complement the *trpG7* mutation in *B. subtilis* 1A491.

The presumed *trpG* gene in pK0101 was subcloned into plasmid pBD9 to demonstrate more clearly that the gene identified as *trpG* by sequence comparisons (see below) was, in fact, *trpG*. Plasmid pK0101 was digested with *Eco*RI and *Cla*I, and both the 0.5-kb *Eco*RI-*Cla*I fragment and the 0.8-kb *Cla*I-*Cla*I fragment were isolated following electrophoresis in low-gelling-temperature agarose. Vector pBD9 (11) was cut into two fragments with *Eco*RI and *Cla*I. The largest fragment (5.5 kb) contains the origin of replication of

```
                            30                            60                            90
GAATTCTCGGAGGCATCTCATCAGGTGCGGCAATTTATGCTGCTCTTCAAGTAGCAAAAAAACTCGGAAAAGGCAAAAAAGTATTGGCCA

                           120                           150                           180
TCATTCCAAGTAACGGCGAACGTTATCTAAGCACGCCGCTGTACCAATTCGATTAAAAAAAGCCAAAACTCCCGGTTCGCCGGGAGTTTT

                           210                           240                           270
TTTATATTTCGTGCATCAAATATGCCATGTCTGCGACGAATGTTATTCTTGGAAGAGAAATTCACTTTTTTCACTAACAACATTGCTTTA

                           300          -----→ pab   330                           360
CAATTAAAAACAAGTAAAAAAGAAAATAGAGGAATGATACAAATGGCACAACGCAGACCGGCAGGCAAAAAAATACCTTTTCAAAAAGAC
                              rbs          M  A  Q  R  R  P  A  G  K  K  I  P  F  Q  K  D

                           390                           420                           450
TCATTCTTACAACAATTTGAGAAACTTGCGCAATCCCGGAAACATCATGTACTTCTCGAAAGTGCAAGAGGCGGCAGATATAGTATAGCC
 S  F  L  Q  Q  F  E  K  L  A  Q  S  R  K  H  H  V  L  L  E  S  A  R  G  G  R  Y  S  I  A

                           480                           510                           540
GGTCTTGATCCAATTGCGACTGTGAAAGGAAAAGACGGAATAACTACAATTAAGCATGGTGATGAGATGCTGTTTAAAGAAGGTGATCCA
 G  L  D  P  I  A  T  V  K  G  K  D  G  I  T  T  I  K  H  G  D  E  M  L  F  K  E  G  D  P

                           570                           600                           630
TTACGGGCCTTCCACAGCTGGTTTAAAACACTGGAAACAGAAACGAATCATGAGTTCCCTGACTTTCAAGGCGGGGCAATCGGGTTTCTC
 L  R  A  F  H  S  W  F  K  T  L  E  T  E  T  N  H  E  F  P  D  F  Q  G  G  A  I  G  F  L

                           660                           690                           720
AGCTATGATTACGCACGGTACATTGAAAATTTTAAAATGCTCTCATTAGATGATTTAGAAACACCAGATATTTATTTTCTTGTTTTTGAT
 S  Y  D  Y  A  R  Y  I  E  N  F  K  M  L  S  L  D  D  L  E  T  P  D  I  Y  F  L  V  F  D

                           750                           780                           810
GATATAGCAGTTTATGACCATCAAGAAGAGTCTCTATGGCTGATTACTCATGTTAATGGTTCTGATCAGGAAACAGCGGATGTGAAGCTA
 D  I  A  V  Y  D  H  Q  E  E  S  L  W  L  I  T  H  V  N  G  S  D  Q  E  T  A  D  V  K  L

                           840                           870                           900
TCTGAGTTAGAGCAGATGTGGTTGACTGAGCTTCCCGCTGTCACTTCGCGAGAGATGAAGCCTGAAACAGCTGGTTCTTTCGCGGCGCCA
 S  E  L  E  Q  M  W  L  T  E  L  P  A  V  T  S  R  E  M  K  P  E  T  A  G  S  F  A  A  P

                           930                           960                           990
TTTACCGAGGATGGGTTCTCACAAGCTGTAGAGAAAATCAAACAATACATTGCCAGCGGAGATGTGTTTCAAGTCAATCTATCAATAAGG
 F  T  E  D  G  F  S  Q  A  V  E  K  I  K  Q  Y  I  A  S  G  D  V  F  Q  V  N  L  S  I  R

                          1020                          1050                          1080
CAGTCACAGTCACTGTCTGTCCACCCATATCAAATTTACAAAACCTTGAGAGAAGTAAATCCTTCTCCTTATATGGCGTATTTAGAAACA
 Q  S  Q  S  L  S  V  H  P  Y  Q  I  Y  K  T  L  R  E  V  N  P  S  P  Y  M  A  Y  L  E  T

                          1110                          1140                          1170
CCTGATTTCCAAATCATTTGCGGATCGCCTGAACTGCTTGTCAGCAAAAAGGGCAAGCTATTAGAGACGAGACCGATTGCGGGCACCCGT
 P  D  F  Q  I  I  C  G  S  P  E  L  L  V  S  K  K  G  K  L  L  E  T  R  P  I  A  G  T  R

                          1200                          1230                          1260
TCCAGAGGGAAAACAAATGAAGAAGACGAGGCGCTTGCAAACGAATTGATACACAATGAAAAAGAACGCGCGGAACATGTCATGCTGGTT
 S  R  G  K  T  N  E  E  D  E  A  L  A  N  E  L  I  H  N  E  K  E  R  A  E  H  V  M  L  V

                          1290                          1320                          1350
GATCTTGAGCGAAATGATCTGGGAAGAGTATCACGTTACGGGTCTGTGCGCGTAAATGAATTCATGGCAATTGAAAAATACTCGCATGTG
 D  L  E  R  N  D  L  G  R  V  S  R  Y  G  S  V  R  V  N  E  F  M  A  I  E  K  Y  S  H  V
```

FIG. 2. Nucleic acid sequence of the 4,880 bp of *B. subtilis* chromosomal DNA in pK0101. The deduced amino acid sequences of Pab, TrpG, PabC, Sul, ORF1, and part of ORF2 are also shown. The proposed ribosome-binding sites (rbs) are underlined. The proposed Mtr-binding site (mtr) is identified by a line above the sequence. The stop sites for protein synthesis are marked by asterisks (*). These sequences have been submitted to GenBank (accession no. M34053).

pBD9 and a kanamycin resistance gene. This fragment was isolated following agarose gel electrophoresis and used as a vector to simultaneously clone the two fragments isolated from pK0101. The ligation mixture was transformed into *B. subtilis* 1A491 and Km^r transformants were isolated. One of the transformants contained a plasmid, referred to as pJS3, that contained the two pK0101-derived fragments in the same order and orientations present in pK0101. This plasmid

```
                          1380                          1410                          1440
ATGCACATTGTGTCTAATGTCCAAGGTGAACTGCAGGATGGGTATGATGCTGTAGATATTATTCATGCTGTGTTTCCCGGAGGAACCATT
 M  H  I  V  S  N  V  Q  G  E  L  Q  D  G  Y  D  A  V  D  I  I  H  A  V  F  P  G  G  T  I 

                          1470                          1500                          1530
ACTGGTGCACCGAAAGTAAGAACGATGGAAATTATAGAAGAACTTGAGCCGACACGCCGAGGGCTTTATACTGGATCTATAGGATGGTTT
 T  G  A  P  K  V  R  T  M  E  I  I  E  E  L  E  P  T  R  R  G  L  Y  T  G  S  I  G  W  F 

                          1560                          1590                          1620
GGATATAATCACGATCTGCAGTTTAATATCGTCATTCGAACCATTTATGCAACCGGAGGGCAGGCATTTATGCAGTCCGGTGCAGGAGTT
 G  Y  N  H  D  L  Q  F  N  I  V  I  R  T  I  Y  A  T  G  G  Q  A  F  M  Q  S  G  A  G  V 

                          1650                          1680                          1710
GTGATTGATTCTGTTCCGAAGCACGAATACAAGGAATCATTCAAAAAAGCTTTTGCGATGCAAAGAGCATTAGAGCTGAGCGAAGAAGAG
 V  I  D  S  V  P  K  H  E  Y  K  E  S  F  K  K  A  F  A  M  Q  R  A  L  E  L  S  E  E  E 

                          -----→ trpG
            Mtr           1740                          1770                          1800
ACAAAAATTAGATGAGGTGAGCGGAGAAATGATTTTAATGATTGATAACTACGATTCATTCACGTACAACTTGGTACAGTATTTGGGCGA
 T  K  I  R  * rbs           M  I  L  M  I  D  N  Y  D  S  F  T  Y  N  L  V  Q  Y  L  G  E

                          1830                          1860                          1890
GCTTGGGGAAGAGCTGGTTGTGAAACGCAATGACAGCATCACAATCGATGAAATTGAAGAACTGTCTCCGGACTTTCTGATGATATCTCC
  L  G  E  E  L  V  V  K  R  N  D  S  I  T  I  D  E  I  E  E  L  S  P  D  F  L  M  I  S  P

                          1920                          1950                          1980
CGGACCGTGCAGCCCTGATGAGGCGGGAATCAGCCTCGAAGCAATTAAACATTTCGCAGGGAAAATTCCTATTTTCGGTGTATGTCTCGG
  G  P  C  S  P  D  E  A  G  I  S  L  E  A  I  K  H  F  A  G  K  I  P  I  F  G  V  C  L  G

                          2010                          2040                          2070
ACATCAGTCCATCGCACAAGTGTTCGGTGGTGATGTTGTTAGGGCAGAACGGCTTATGCACGGGAAAACCTCGGATATCGAGCATGACGG
  H  Q  S  I  A  Q  V  F  G  G  D  V  V  R  A  E  R  L  M  H  G  K  T  S  D  I  E  H  D  G

                          2100                          2130                          2160
CAAAACCATTTTTGAAGGGTTGAAAAATCCCCTTGTTGCGACGCGATACCACTCGCTGATCGTAAAACCTGAGACGCTGCCAAGCTGTTT
  K  T  I  F  E  G  L  K  N  P  L  V  A  T  R  Y  H  S  L  I  V  K  P  E  T  L  P  S  C  F

                          2190                          2220                          2250
TACAGTAACAGCACAAACGAAAGAAGGAGAAATCATGGCTATTCGCCACAATGACCTCCCGATAGAGGGTGTGCAATTTCACCCAGAGTC
  T  V  T  A  Q  T  K  E  G  E  I  M  A  I  R  H  N  D  L  P  I  E  G  V  Q  F  H  P  E  S

                          2280                          2310             -----→ pabC  2340
TATTATGACCTCCTTTGGGAAAGAAATGCTCAGAAATTTTATTGAGACATATCGCAAGGAAGTTATTGCGTGATGATATATGTGAACGGC
  I  M  T  S  F  G  K  E  M  L  R  N  F  I  E  T  Y  R  K  E  V  I  A  * M  I  Y  V  N  G
                                                         rbs

                          2370                          2400                          2430
CGGTATATGGAGGAGAAAGATGCAGTTCTTTCTCCTTTTGACCATGGGTTTTTATATGGGATCGGTGTTTTCGAAACCTTCAGGCTCTAC
 R  Y  M  E  E  K  D  A  V  L  S  P  F  D  H  G  F  L  Y  G  I  G  V  F  E  T  F  R  L  Y 

                          2460                          2490                          2520
GAAGGCTGCCCGTTTTTGCTTGATTGGCATATAGAGAGGCTTGAACGCGCGCTTAAGGATCTGCAAATCGAATATACTGTTTCAAAGCAT
 E  G  C  P  F  L  L  D  W  H  I  E  R  L  E  R  A  L  K  D  L  Q  I  E  Y  T  V  S  K  H 

                          2550                          2580                          2610
GAGATCCTTGAGATGCTGGACAAGCTGCTTAAGCTAAATGATATCAAGGATGGCAATGCCCGCGTCAGACTGAACATATCAGCGGGAATC
 E  I  L  E  M  L  D  K  L  L  K  L  N  D  I  K  D  G  N  A  R  V  R  L  N  I  S  A  G  I 
```

FIG. 2—*Continued.*

contains the last 384 bp of *pab* and the complete reading frame for the presumed *trpG*; i.e., the only complete gene in the insert DNA is the presumed *trpG*. pJS3 in *B. subtilis* 1A491 complemented the *trpG7* mutation, eliminating its partial *p*-aminobenzoate growth requirement just as pK0101 did.

Sequence determination. The entire sequence of both strands of the pJS1 and pJS2 inserts was determined, with

```
                          2640                          2670                          2700
AGCGATAAAGGCTTTGTCGCCCAGACGTACGATAAGCCCACCGTTTTATGCTTTGTGAATCAGCTGAAACCAGAAAGTCTTCCTTTGCAA
 S  D  K  G  F  V  A  Q  T  Y  D  K  P  T  V  L  C  F  V  N  Q  L  K  P  E  S  L  P  L  Q

                          2730                          2760                          2790
AAAGAAGGAAAGGTTTTATCGATACGGAGAAACACTCCAGAGGGATCGTTCCGGTTAAAATCTCATCATTATCTAAATAATATGTATGCG
 K  E  G  K  V  L  S  I  R  R  N  T  P  E  G  S  F  R  L  K  S  H  H  Y  L  N  N  M  Y  A

                          2820                          2850                          2880
AAACGGGAAATTGGAAATGATCCGCGTGTCGAGGGAATTTTCTTAACGGAGGATGGAGCTGTTGCAGAGGGTATCATATCTAATGTCTTC
 K  R  E  I  G  N  D  P  R  V  E  G  I  F  L  T  E  D  G  A  V  A  E  G  I  I  S  N  V  F

                          2910                          2940                          2970
TGGAGAAAAGGTCGCTGTATCTATACGCCATCGCTTGATACCGGTATTCTGGACGGCGTCACGAGACGTTTTATCATAGAAAATGCAAAG
 W  R  K  G  R  C  I  Y  T  P  S  L  D  T  G  I  L  D  G  V  T  R  R  F  I  I  E  N  A  K

                          3000                          3030                          3060
GATATCGGCCTGGAGCTCAAAACGGGCCGATATGAGCTAGAAGCTCTCTTGACAGCTGACGAAGCGTGGATGACGAACTCGGTTCTTGAA
 D  I  G  L  E  L  K  T  G  R  Y  E  L  E  A  L  L  T  A  D  E  A  W  M  T  N  S  V  L  E

                          3090                          3120                          3150
ATCATCCCGTTTACTAAAATAGAAGAAGTAAACTATGGAAGCCAAAGCGGAGAAGCAACCTCCGCTCTTCAATTGTTATACAAAAAAGAA
 I  I  P  F  T  K  I  E  E  V  N  Y  G  S  Q  S  G  E  A  T  S  A  L  Q  L  L  Y  K  K  E

                          3180     -----→ sul           3210                          3240
ATAAAGAACATGATTCATGAAAAGGGAGGAAGAGCATGGCGCAGCACACAATAGATCAAACACAAGTAATCCACACTAAGCCCAGCGCTT
 I  K  N  M  I  H  E  K  G  G  R  A  W  R  S  T  Q  *
                          rbs       M  A  Q  H  T  I  D  Q  T  Q  V  I  H  T  K  P  S  A

                          3270                          3300                          3330
TATCATATAAAGAGAAGACGCTGGTGATGGGAATTTTAAACGTAACGCCTGACTCTTTCTCGGACGGCGGAAAATATGACAGCTTGGACA
L  S  Y  K  E  K  T  L  V  M  G  I  L  N  V  T  P  D  S  F  S  D  G  G  K  Y  D  S  L  D

                          3360                          3390                          3420
AGGCGCTGCTGCACGCGAAAGAGATGATCGATGATGGTGCCCATATCATTGATATTGGAGGGGAATCGACAAGGCCTGGCGCTGAGTGCG
K  A  L  L  H  A  K  E  M  I  D  D  G  A  H  I  I  D  I  G  G  E  S  T  R  P  G  A  E  C

                          3450                          3480                          3510
TATCTGAGGATGAGGAGATGTCCAGAGTCATTCCGGTGATTGAGCGGATTACGAAAGAGCTTGGTGTTCCTATTTCTGTAGACACGTACA
V  S  E  D  E  E  M  S  R  V  I  P  V  I  E  R  I  T  K  E  L  G  V  P  I  S  V  D  T  Y

                          3540                          3570                          3600
AGGCTTCTGTCGCAGATGAAGCAGTGAAAGCCGGTGCATCCATTATCAATGATATTTGGGGAGCCAAACATGATCCGAAGATGGCTTCCG
K  A  S  V  A  D  E  A  V  K  A  G  A  S  I  I  N  D  I  W  G  A  K  H  D  P  K  M  A  S

                          3630                          3660                          3690
TTGCAGCTGAACATAATGTTCCAATTGTACTCATGCATAACCGCCCTGAAAGAAACTACAATGACTTATTGCCGGATATGCTGTCGGACT
V  A  A  E  H  N  V  P  I  V  L  M  H  N  R  P  E  R  N  Y  N  D  L  L  P  D  M  L  S  D

                          3720                          3750                          3780
TAATGGAGAGTGTAAAAATTGCTGTTGAGGCCGGAGTAGACGAGAAGAACATTATTCTTGATCCTGGTATCGGTTTCCCGAAAACCTATC
L  M  E  S  V  K  I  A  V  E  A  G  V  D  E  K  N  I  I  L  D  P  G  I  G  F  P  K  T  Y

                          3810                          3840                          3870
ACGATAACTTGGCAGTGATGAACAAACTAGAGATTTTCAGCGGATTGGGATATCCGGTTCTTCTGGCAACCTCCCGAAAAAGATTCATCG
H  D  N  L  A  V  M  N  K  L  E  I  F  S  G  L  G  Y  P  V  L  L  A  T  S  R  K  R  F  I
```

FIG. 2—*Continued.*

overlaps of all restriction sites used for labeling. Figure 2 shows the sequence of the 4,880 bp plus the derived amino acid sequence of the six proteins. Analysis of the pJS1 insert sequence revealed a long open reading frame (337 codons) preceded by a consensus ribosome-binding site. The amino acid sequence deduced from this open reading frame showed considerable similarity to the first 75% of the sequence of the large subunits of *p*-aminobenzoate synthase from *E. coli* and other enteric bacteria (10). Subsequent sequencing of 1.4 kb at one end of pJS2 (Fig. 2) disclosed the remainder of the *pab* gene, a short intercistronic region, and a second open reading frame encoding 194 residues with pronounced se-

```
                          3900                          3930                          3960
GACGTGTTCTGGATCTTCCGCCTGAGGAGCGGGCTGAGGGCACAGGCGCGACTGTGTGTCTCGGCATTCAAAAAGGCTGTGACATTGTCA
G  R  V  L  D  L  P  P  E  E  R  A  E  G  T  G  A  T  V  C  L  G  I  Q  K  G  C  D  I  V

                          3990                          4020               ---→ ORF1 4050
GGGTCCATGATGTAAAGCAAATTGCCAGAATGGCGAAAATGATGGACGCGATGCTGAATAAGGGAGGGGTGCACCATGGATAAAGTTTAT
R  V  H  D  V  K  Q  I  A  R  M  A  K  M  M  D  A  M  L  N  K  G  G  V  H  H  G  *
                                                               rbs          M  D  K  V  Y

                          4080                          4110                          4140
GTAGAAGGTATGGAGTTTTACGGATATCACGGTGTGTTCACAGAAGAAAACAAACTTGGCCAGCGGTTTAAAGTCGATTTAACCGCTGAG
 V  E  G  M  E  F  Y  G  Y  H  G  V  F  T  E  E  N  K  L  G  Q  R  F  K  V  D  L  T  A  E

                          4170                          4200                          4230
CTGGATTTAAGCAAAGCTGGACAGACAGACGACCTTGAGCAAACGATCAACTATGCTGAGCTCTATCACGTATGTAAAGATATCGTGGAA
 L  D  L  S  K  A  G  Q  T  D  D  L  E  Q  T  I  N  Y  A  E  L  Y  H  V  C  K  D  I  V  E

                          4260                          4290                          4320
GGGGAGCCTGTGAAATTGGTGGAAACGCTGGCGGAACGTATTGCTGGCACTGTTCTCGGAAAATTTCAGCCTGTTCAGCAATGTACGGTG
 G  E  P  V  K  L  V  E  T  L  A  E  R  I  A  G  T  V  L  G  K  F  Q  P  V  Q  Q  C  T  V

                          4350                          4380               ---→ ORF2 4410
AAAGTGATTAAGCCAGACCCGCCAATTCCCGGACACTATAAATCAGTAGCAATTGAAATTACGAGAAAAAAGTCATGAACAACATAGCTT
 K  V  I  K  P  D  P  P  I  P  G  H  Y  K  S  V  A  I  E  I  T  R  K  K  S  *
                                                             rbs           M  N  N  I  A

                          4440                          4470                          4500
ATATTGCACTTGGATCTAATATTGGAGATAGAGAAACGTATTTAAGGCAAGCAGTGGCTTTACTGCATCAGCATGCTGCGGTGACAGTCA
Y  I  A  L  G  S  N  I  G  D  R  E  T  Y  L  R  Q  A  V  A  L  L  H  Q  H  A  A  V  T  V

                          4530                          4560                          4590
CTAAAGTGTCGTCTATTTACGAAACTGACCCGGTCGGATACGAAGATCAAGCTCAATTTTTGAATATGGCTGTTGAAATCAAGACATCAT
T  K  V  S  S  I  Y  E  T  D  P  V  G  Y  E  D  Q  A  Q  F  L  N  M  A  V  E  I  K  T  S

                          4620                          4650                          4680
TGAACCCTTTTGAACTCCTTGAACTGACGCAGCAGATAGAAAATGAATTAGGCAGAACAAGGGAAGTAAGATGGGGACCGCGGACGGCAG
L  N  P  F  E  L  L  E  L  T  Q  Q  I  E  N  E  L  G  R  T  R  E  V  R  W  G  P  R  T  A

                          4710                          4740                          4770
ACCTTGACATTTTGTTATTTAATCGTGAAAATATTGAAACAGAGCAACTAATTGTTCCGCATCCGAGAATGTATGAGCGTTTGTTTGTCC
D  L  D  I  L  L  F  N  R  E  N  I  E  T  E  Q  L  I  V  P  H  P  R  M  Y  E  R  L  F  V

                          4800                          4830                          4860
TTGCGCCGCTTGCGGAAATTTGCCAGCAGGTTGAAAAAGAGGCTACAAGCGCCGAAACAGACCAAGAAGGTGTAAGAGTATGGAAGCAGA
L  A  P  L  A  E  I  C  Q  Q  V  E  K  E  A  T  S  A  E  T  D  Q  E  G  V  R  V  W  K  Q

                4880
AATCTGGGGTAGACGAATTC
K  S  G  V  D  E  F
```

FIG. 2—*Continued.*

quence similarity to the small subunits of *p*-aminobenzoate synthase from enteric bacteria and the *trpG* products of a variety of organisms (20). As stated above, three complete open reading frames and one partial open reading frame were located downstream from *trpG*.

The amino acid sequence of the *B. subtilis pab* gene product can be aligned with *E. coli* PabB, *B. subtilis* TrpE, *Brevibacterium lactofermentum* TrpE (27), and *P. aeruginosa* TrpE (9) (Fig. 3). In the C-terminal half of the molecule, there are about 35% identities in at least four of the five proteins over about 250 residues. Although alignment of the proximal half of these molecules is more problematical, with <8% residue identities in at least four of the five proteins in the best alignment found, a similar result is seen with all evolutionarily diverse large *p*-aminobenzoate synthase subunit sequences studied (7, 19) and is not peculiar to the *B. subtilis* Pab protein.

Alignment of the *B. subtilis* TrpG protein with *E. coli* PabA, *Brevibacterium lactofermentans* TrpG, *E. coli* TrpG, and the only other amphibolic glutamine amidotransferase subunit that has been sequenced, *A. calcoaceticus* TrpG, is shown in Fig. 4. Both in the number of residue identities and the position of gaps introduced to maximize these identities, the *B. subtilis* protein resembles *A. calcoaceticus* TrpG and *E. coli* PabA more than *Brevibacterium lactofermentans* TrpG or *E. coli* TrpG. The number of amino acids in the other proteins that were identical to those in *B. subtilis* TrpG (194 amino acids total) were as follows: *E. coli* PabA, 110; *A.*

```
                                               10                  20
B.l.TrpE                            M S T N P H V F S L D V R Y H E D A S A L F
P.a.TrpE          M N R E E F L R L A A D G Y N R I P L S F E T L A D F D T P L
B.s.TrpE    M N F Q S N I S A F L E D S L S H H T I P I V E T F T V D T L T P I
E.c.PabB                                M K T L S P A V I T L L W R Q D A A E F
B.s.Pab                               M A Q R R P A G K K I P F Q K D S F L Q Q
                                                                    ^

       30                  40                  50                  60
- A H L G G T T A D D A A L L E S A D I T T R N G I S S L A V L K S S V R I T C
S I Y L K L A D A P N S Y L L E S V Q G G E K W G R Y S I I G L P C R - - - - -
- Q M I E K L D R E I T Y L L E S K D D T S T W S R Y S F I G L N P - - - - - -
- Y F S R L S H L P W A M L L H S G Y A D H P Y S R F D I V V A E P - - - - - -
- F E K L A Q S R K H H V L L E S A R G G - - - - R Y S I A G L D P - - - - - -
                          ^ ^ ^ ^                 ^   ^       ^

       70                  80                  90                 100
T G N T V V T Q P L T D S G R A V V A R L T Q Q L G Q Y N T A - E N - - - - - -
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - T V L R V Y D H
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - F L T I K E E Q
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - I C T L T T F G
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - I A T V K G K D

      110                 120                 130                 140
T F S F P A S D A V D E R E R L - - T A P S T I E V L R K L Q F - - E S G Y S D
Q V R I S I D G V E T E R F D C A D P L A F V E E F K A R Y Q V P - T V P G L P
G R F S A A D Q D S K S L Y T G N E L K E V L N W M N T T Y K I K T P E L G I P
K E T V V S E S E K R T T T T D D - P L Q V L Q Q V L D R A D I R P T H N E D L
G I T T I K H G D E M L F K E G D - P L R A F H S W F K T L E T E - T N H E F P

      150                 160                 170                 180
A S L P L L M G G F A F D F L E T F E T L P A V E E S V N T Y - - P D Y Q F V L
R F D G G L V G Y F G Y D C V R Y V E K R L A T C P N P D P L G N P D I L L M V
- F V G G A V G Y L S Y D M I P L I E P S V P S H T K E T D M - - E K C M L F V
P F Q G G A L G L F G Y D L G R R F E S L P E I A E Q D I V L - - P D M A V G I
D F Q G G A I G F L S Y D Y A R Y I E N F K M L S L D D L E T - - P D I Y F L V
  ^   ^ ^     ^       ^ ^           ^                             ^ ^

      190                 200                 210                 220
A E I V L D I N H Q D Q T A K L T G V S N A P G E L E - - - - - - - - - - - - -
S D A V V V F D N L A G K I H A I V L A D P S E E N A - - - - - - - - - Y E R G
C R T L I A Y D H E T K N V H F I Q Y A R L T G E E T K N E K M D V F H Q N H L
Y D W A L I V D H Q R H T V S L L S H N D V N A R R A - - - - - - - - - - - - -
F D D I A V Y D H Q E E S L W L I T H V N G S D Q E T - - - - - - - - - - - - -
              ^ ^

      230                 240                 250                 260
A E L N K L S L L I D A A L P A T E H A Y Q T T P H D G D T L R V V A D I P D A
Q A R L E E L L E R L R Q P I T P R R G L D L E A A Q G R E P A F R A S F T R E
E L Q N L I E K M M D Q K N I K E L F L S A D S Y K T P S F E T V S S N Y E K S
- - - - - - - - - - - - - - W L E S Q Q F S P Q E D F T L T S D W Q S N M T R E
A D V K L S E L E Q M W L - T E L P A V T S R E M K P E T A G S F A A P F T E D
```

FIG. 3. Alignment of the *B. subtilis* Pab protein with PabB of *E. coli* and TrpE of selected microorganisms. Hyphens indicate gaps introduced to increase similarity. B.l., *Brevibacterium lactofermentans*; P.a., *P. aeruginosa*; B.s., *B. subtilis*; E.c., *E. coli*. The stop sites for protein synthesis are marked by asterisks (*). (^) indicates residue identities in at least four of the five proteins.

calcoaceticus TrpG, 108; *E. coli* Trp(G), 83; and *Brevibacterium lactofermentans* TrpG, 56. There are about 31% identities in at least four of the five proteins.

A search of the translated version of the Genetics Sequence Data Bank (GenBank) for other proteins possessing significant amino acid sequence similarity to PabC was conducted. Only IlvE from *E. coli* (21), the branched-chain amino acid transaminase B that functions in biosynthesis of isoleucine and valine, had a significantly similar sequence. The normalized alignment score (8) for the two proteins was 219, a value indicating that the relatedness of the two proteins is probably significant. A role of PabC in synthesis of *p*-aminobenzoic acid synthesis is established below.

The protein in GenBank that had an amino acid sequence most similar to Sul was the dihydropteroate synthase of *Streptococcus pneumoniae* (25). This enzyme catalyzes the condensation of *p*-aminobenzoate with 6-CH_2OH-7,8-H_2-pterin pyrophosphate to form 7,8-dihydropteroate, the immediate precursor of folic acid. Figure 5 shows the alignment of these two Sul proteins. The normalized alignment score

```
      270                 280                 290                 300
Q F R T Q I N E L K E N I Y N G D I Y Q V V P A R T F T A P C P D A F A - - - -
D Y E N A V G R I K D Y I L A G D C M Q V V P S Q R M S I E F K A A P I D - - -
A F M A D V E K I K S Y I K A G D I F Q G V L S Q K F E V P I K A D A F E - - -
Q Y G E K F R Q V Q E Y L H S G D C Y Q V N L A Q R F H A T Y S G D E W Q - - -
G F S Q A V E K I K Q Y I A S G D V F Q V N L S I R Q S Q S L S V H P Y Q - - -
                  ^   ^ ^     ^ ^     ^ ^

      310                 320                 330                 340
A Y L Q L R A T N P S P Y M F Y I R G L N E G R S Y E L F G A S P E S N L K F T
L Y R A L R C F N P T P Y M Y F F N F G D - - - - F H V V G S S P E V L V R V E
L Y R V L R I V N P S P Y M Y Y M K L L D - - - - R E I V G S S P E R L I H V Q
A F L Q L N Q A N R A P F S A F L R L E Q - - - - G A I L S L S P E R F I L C D
I Y K T L R E V N P S P Y M A Y L E T P D - - - - F Q I I C G S P E L L V S E K
  ^     ^ ^     ^ ^   ^ ^ ^   ^                               ^ ^ ^

      350                 360                 370                 380
A A N R E L Q L Y P I A G T R P R G L N P D G S I N D E L D I R N E L D M R T D
D - - G L V T V R P I A G T R P R G I - - - - - - N E E A D L A L E Q D L L S D
D - - G H L E I H P I A G T R K R G A - - - - - - D K A E D E R L K V E L M K D
N - - S E I Q T R P I K G T L P R L P - - - - - - D P Q E D S K Q A V K L A N S
G - - K L L E T R P I A G T R S R G K - - - - - - T N E E D E A L A N E L I H N
                  ^ ^ ^ ^ ^ ^   ^ ^                       ^             ^

      390                 400                 410                 420
A K E I A D D T M L V D L A R N D L A R V S V P A S R R V A D L L Q V D R Y S R
A K E I A E H L M L I D L G R N D V G R V S D I G A V K V T E K M V I E R Y S N
E K E K A E H Y M L V D L A R N D I G R V A E Y G S V S V P E F T K I V S F S H
A K D R A E N L M I V D L M R N D I G R V A V A G S V K V P E L F V V E P F P A
E K E R A E H V M L V D L E R N D L G R V S R Y G S V R V N E F M A I E K Y S H
  ^ ^   ^ ^     ^   ^ ^ ^   ^ ^ ^   ^ ^ ^       ^ ^ ^   ^   ^               ^

      430                 440                 450                 460
V M H L V S R V T A T L D P E L D A L D A Y R A C M N M G T L T G A P K L R A M
V M H I V S N V T G Q L R E G L S A M D A L R A I L P A G T L S G A P K I R A M
V M H I I S V V T G R L K K G V H P V D A L M S A F P A G T L T G A P K I R A M
V H H L V S T I T A Q L P E Q L H A S D L L R A A F P G G S I T G A P K V R A M
V M H I V S N V Q G E L Q D G Y D A V D I I H A V F P G G T I T G A P K V R T M
^ ^ ^   ^ ^   ^ ^     ^           ^   ^       ^     ^   ^ ^   ^ ^ ^ ^ ^   ^ ^ ^

      470                 480                 490                 500
- E L L R G V E K R R R G S Y G G A V G Y L R G N G - D M D N C I V I R S A F V
- E I I D E L E P V K R G V Y G G A V G Y L A W N G - N M D T A I A I R T A V I
- Q L L Q E L E P T P R E T Y G G C I A Y I G F D G - N I D S C I T I R T M S V
- E I I D E L E P Q R R N A W C G S I G Y L S F C G - N M D T S I T I R T L T A
- E I I E E L E P T R R G L Y T G S I G W F G Y N H - D L Q F N I V I R T I Y A
      ^   ^ ^ ^ ^     ^     ^   ^     ^ ^         ^       ^     ^   ^ ^ ^

      510                 520                 530                 540
Q D G V A A V Q A G A G V V R D S N P Q S E A D E T L H K A Y A V L N A I A L A
K N G E L H V Q A G G G I V A D S V P A L E W E E T I N K R R A M F R A V A L A
K N G V A S I Q A G A G I V A D S V P E A E Y E E S C N K A G A L L K T I H I A
I N G Q I F C S A G G G I V A D S Q E E A E Y Q E T F D K V N R I L K Q L E K *
T G G Q A F M Q S G A G V V I D S V P K H E Y K E S F K K A F A M Q R A L E L S
    ^         ^ ^ ^   ^   ^   ^ ^   ^     ^     ^             ^

      550                 560
A G S T L E V I R *
E Q S V E *
E D M F H S K E D K A D E Q I S T I V R *

E E E T K I R *
```

FIG. 3—*Continued.*

for the two proteins was 246, a value indicating almost certain relatedness. In the Discussion, evidence is summarized in support of the hypothesis that the gene referred to as *sul* is the dihydropteroate synthase gene.

No significant match was detected between the deduced amino acid composition of ORF1 and proteins in GenBank. Also, no significant match was apparent with the portion of ORF2 analyzed. Only the 5′ end of *ORF2* was present in the

```
                               10                  20                  30
E. c. PabA        M I L L I D N Y D S F T W N L Y Q Y F C E L G A D V L V K R N D A L T L A
B. s. TrpG        M I L M I D N Y D S F T Y N L V Q Y L G E L G E E L V V K R N D S I T I D
A. c. TrpG        M L L M I D N Y D S F T Y N I V Q Y F G E L N Q D V K V V R N D Q V T L E
E. c. Trp(G)  M A D I L L L D N I D S F T Y N L A D Q L R S N G H N V V I Y R N H I P A Q T
B. l. TrpG    M T H V V L I D N H D S F V Y N L V D A F A V A G Y K C T V F R N T V P V E T
                      ^   ^ ^ ^   ^ ^ ^ ^ ^ ^ ^               ^         ^   ^ ^

                   50                  60                  70                  80
D I D A L K P Q K - - - I V I S P G P C T P D E A G I S L D V I R H Y A G R L P I L G V C L
E I E E L S P D F - - - L M I S P G P C S P D E A G I S L E A I K H F A G K I P I F G V C L
D I E R W Q P K Y - - - L V V G P G P C S P T E A G I S I P A I H H F A G R I P L L G V C L
L I E R L A T M S N P V L M L S P G P G V P S E A G C M P E L L T R L R G K L P I I G I C L
I L A A N P D L - - - - I C L S P G P G Y P A D A G N M M A L I E R T L G Q I P L L G I C L
  ^                           ^ ^ ^ ^     ^   ^ ^ ^           ^         ^     ^     ^   ^ ^

       90                  100                 110                 120                 130
G H Q A M A Q A F G G K V V R A A K V M H G K T S P I T H N G E G V F R G L P N P - - - - -
G H Q S I A Q V F G G D V V R A E R L M H G K T S D I E H D G K T I F E G L K N P - - - - -
G H Q A I G Q A F G G N I I R A K T V M H G R L S D M Y H T D K G I F S N L P S P - - - - -
G H Q A I V E A Y G G Y V G Q A G E I L H G K A S S I E H D G Q A M F A G L T N P - - - - -
G Y Q A L I E Y H G G K V E P C G - P V H G T T D N M I L T D A G V Q S P V F A G L A T D V
^ ^ ^ ^           ^ ^   ^     ^         ^ ^     ^       ^           ^     ^     ^

               140                 150                 160                 170
- - - - - - - - - - - L T V T R Y H S L V V E P D S L P A C F D V T A W S E T - - - - - R E
- - - - - - - - - - - L V A T R Y H S L I V K P E T L P S C F T V T A Q T K E - - - - - G E
- - - - - - - - - - - F S A T R Y H S L V I E Q E S L P E C L E V T C W T N Q N D G S I E E
- - - - - - - - - - - L P V A R Y H S L V G S N - - I P A G L T I N A H F N G - - - - - - M
E P D H P E V P G R K V P I G R Y H S L G C V V - - A P D G I E S L G T C S S - - E I G D V
                              ^ ^ ^ ^ ^               ^

   180                 190                 200                 210                 220
I M G I R H R Q W D L E G V Q F H P E S I L S E Q G H Q L L A N F L H R *
I M A I R H N D L P I E G V Q F H P E S I M T S F G K E M L R N F I E T Y R K E V I A *
I M G V K H K T L P V E G V Q F H P E S I L S Q H G H Q I F K N F L E I Y A *
V M A V R H D A D R V C G F Q F H P E S I L T T Q G A R L L E Q T L A W A Q H K L E P A . .
I M A A R T T D G K A I G L Q F H P E S V L S P T G P I I L S R C V E Q L L A N *
^ ^     ^ ^             ^   ^ ^ ^ ^ ^ ^ ^ ^       ^       ^
```

FIG. 4. Alignment of TrpG proteins and one PabA protein from selected microorganisms. Hyphens indicate gaps introduced to increase similarity. E.c., *E. coli*; B.s., *B. subtilis*; A.c., *A. calcoaceticus*; B.l., *Brevibacterium lactofermentans*. The "G" in *E. coli* Trp(G) is in parentheses to indicate that the sequence shown represents the TrpG portion of the fusion protein, TrpGD. The stop sites for protein synthesis are marked by asterisks (*). (^) indicates residue identities in at least four of the five proteins.

sequenced DNA fragment (sequence coding for 162 amino acids).

Location of the *sul* gene. Kane (17) reported that the order of the genes he studied was *pab-trpG-sul-lysS*. Thus, the *sul* gene should be to the right of *trpG* on pK0101 (Fig. 1). Plasmid pJS4 was derived from pK0101 by deletion of the entire cloned fragment in pK0101 to the right of the *Eag*I target site. The *Eag*I site is only 14 bp to the right of the termination codon of *trpG*. When *B. subtilis* 1A491(pJS4) was plated on SM containing sulfanilamide, no growth occurred. This confirmed the general location of the *sul* gene. Additional experimentation will be needed to locate it precisely. Based on sequence analysis (see above), the open reading frame immediately downstream from *pabC* is almost certain to be the *sul* gene.

Insertional inactivation of *pab*, *trpG*, and *pabC*. One way to gain an understanding of the function of a gene is to inactivate it in vitro, insert it into the chromosome of a prototrophic strain by gene replacement, and determine whether growth requirements have been created. This section describes construction of plasmids containing insertionally inactivated *pab*, *trpG*, and *pabC* genes.

The chloramphenicol resistance (Cm^r) gene of pMI1101 (34; P. Youngman, personal communication) which resides on a 1.5-kb *Sma*I fragment, was inserted by blunt-end ligation into the single *Nru*I site of pJS1 that lies within the *pab* gene. Two plasmids were obtained by transformation of *E. coli* JM109: pCH1 had the Cm^r fragment inserted in the *pab* gene in an orientation such that transcription of the Cm^r gene would occur in the same direction as that of *pab* (Fig. 6A); pCH2 had the Cm^r fragment inserted in the opposite orientation.

For insertional inactivation of *trpG*, we inserted the chloramphenicol resistance gene of pC194. The Cm^r gene of pC194 resides on a 1.6-kb *Cla*I fragment. Although *trpG* contains a single *Cla*I site, there are two *Cla*I sites in the cloned *B. subtilis* chromosomal DNA in pK0101 to the right of *trpG* (Fig. 1). We deleted the region to the right of *trpG* in pK0101 by digesting with *Eag*I and *Xba*I, isolating the 6.3-kb fragment, filling in the staggered ends by incubation with the Klenow fragment of DNA polymerase, and performing blunt-end ligation. The ligation mixture was transformed into *B. subtilis* 1A491, and selection was made for Km^r. The blunt-end ligation destroyed both the *Eag*I and *Xba*I sites. This plasmid was named pJS4. The *trpG* gene in pJS4 was excised by cleavage with *Eco*RI and *Ava*I and was inserted into pUC18 appropriately prepared by removal of the small *Ava*I-*Eco*RI fragment from the polylinker region. The resulting plasmid, pJS5, was recovered following transformation of *E. coli* JM109. The *Cla*I-*Cla*I fragment of pC194 containing the Cm^r gene was inserted into the *Cla*I site of pJS5, generating pJS6 containing the insertionally inactivated *trpG* gene. The pJS6-containing isolate actually had two Cm^r cassettes inserted into *trpG* (Fig. 6B).

```
                  10        20        30        40        50        60
B.s.Sul   MAQHTIDQTQVIHTKPSALSYKEKTLVMGILNVTPDSFSDGGKYDSLDKALLHAKEMIDD
                          |      ||   || |||||||||||    |  ||  |   |
S.p.Sul                  MSSKANHAKTVICGIINVTPDSFSDGGQFFALEQALQQARKLIAE
                                 10        20        30        40

                  70        80        90       100       110
B.s.Sul   GAHIIDIGGESTRPGAECV-SE-DEEMSRVIPVIERITKELGVPISVDTYKASVADEAVK
          ||   ||||||||||   |  |  ||  || |||  | ||  | || || |  ||  |
S.p.Sul   GASMLDIGGESTRPGSSYVEIEIEEEIQRVVPVIKAIRKESDVLISIDTWKSQVAEAALA
             50        60        70        80        90       100

          120       130       140       150       160       170
B.s.Sul   AGASIINDIWGAKHDPKMASVAAEHNVPIVLMHNRPERNYNDLLPDMLSDLMESVKIAVE
          |||   ||| |   | ||  | ||     | | |                         |
S.p.Sul   AGADLVNDITGLMGDEKMPHVVAEARAQVVIMFNPVMARPQHPSSLIFPHFGFGQAFTEE
            110       120       130       140       150       160

          180       190       200       210       220       230
B.s.Sul   AGVDEKNIILDPGIGFPKTYHDNLAVMNKLEIFSGLGYPVLLATSRKRFIGRVLDLPPEE
             |                    |          |  |       |      |
S.p.Sul   ELADFETLPIEELMEAFFERALARAAEAGIAPENILLDPGIGFGLTKKENLLLLRDLDKL
            170       180       190       200       210       220

          240       250       260       270       280
B.s.Sul   RAEGTGATVCLGIQKGCDIVRVHDVKQIARMAKMMDAMLNKGGVHHG
             |                                        |
S.p.Sul   HQKGYPIFLGVSRKRFVINILEENGFEVNPETELGFRNRDTASAHVTSIAARQGVEVVRV
            230       240       250       260       270       280

S.p.Sul   HDVASHRMAVEIASAIRLADEAENLDLKQYK
            290       300       310
```

FIG. 5. Alignment of the Sul proteins of *B. subtilis* (B.s.) and *S. pneumoniae* (S.p.). Hyphens indicate gaps introduced to increase similarity. Vertical lines indicate residue identities.

The Cm^r gene of pMI1101 was inserted into *pabC*. The 1.5-kb *Sma*I fragment of pMI1101, which contains the Cm^r gene, was inserted by blunt-end ligation into the single *Bss*HII site of pJS2 that lies within *pabC*. Before ligation, the *Bss*HII-generated cohesive ends were filled in by use of DNA polymerase (Klenow fragment). Two plasmids were obtained from transformants of *E. coli* JM109: pCH3 had the Cm^r fragment inserted in *pabC* in an orientation such that transcription of the Cm^r gene would occur in the same direction as that of *pabC* (Fig. 6C); pCH4 had the Cm^r fragment inserted in the opposite orientation.

Integration of the insertionally inactivated genes into the chromosome of a prototrophic *B. subtilis* strain. All of the newly constructed plasmids described in the preceding section are capable of autonomous replication in *E. coli* but not in *B. subtilis*. Each plasmid was linearized and transformed into prototrophic *B. subtilis* 1A2. Chloramphenicol-resistant transformants should have the insertionally inactivated genes incorporated into the chromosome. Prior to transformation of *B. subtilis* 1A2, pCH1 and pCH2 were linearized with *Sma*I. The transformants with the insertionally inactivated *pab* gene are referred to as *B. subtilis* 1A2ΩpCH1 and *B. subtilis* 1A2ΩpCH2. pJS6 was linearized with *Bam*HI. The strain with the integrated, insertionally inactivated *trpG* is referred to as *B. subtilis* 1A2ΩpJS6. pCH3 and pCH4 were linearized with *Sma*I. The transformants with the insertionally inactivated *pabC* are referred to as *B. subtilis* 1A2ΩpCH3 and *B. subtilis* 1A2ΩpCH4.

Growth requirements of the *B. subtilis* strains containing insertionally inactivated *trpG*, *pab*, and *pabC* genes. The inoculum for growth experiments was obtained by overnight growth in a medium containing yeast extract and tryptophan (see Methods and Materials). The cells were washed 10 times in the appropriate basal medium lacking tryptophan and *p*-aminobenzoic acid before use as inoculum. The compositions of the basal media used for growing each of the strains of *B. subtilis* are included in the legend to Fig. 7.

Figure 7A shows the growth of the prototrophic strain, *B. subtilis* 1A2, in (i) basal medium and in basal medium supplemented with (ii) tryptophan, (iii) *p*-aminobenzoic acid, and (iv) tryptophan plus *p*-aminobenzoic acid. No growth requirement for tryptophan or *p*-aminobenzoate was observed; the generation time was the same (0.67 h) in all four media. The basal medium for the above experiment contained 45 mM ammonium sulfate. No requirement for tryptophan or *p*-aminobenzoic acid was evident when the basal medium contained 0.45 mM ammonium sulfate (data not shown).

Figure 7B shows the growth of *B. subtilis* 1A2ΩpCH1, the strain with the insertionally inactivated *pab* gene, in the four media. A definite requirement for *p*-aminobenzoic acid and not tryptophan was evident. The generation times in basal medium plus *p*-aminobenzoic acid and in basal medium plus *p*-aminobenzoic acid plus tryptophan were about the same (1.3 h). There was also very little difference in the generation times observed for cells growing in basal medium and basal medium plus tryptophan (1.7 and 2.1 h, respectively). Essentially the same results were obtained with *B. subtilis*

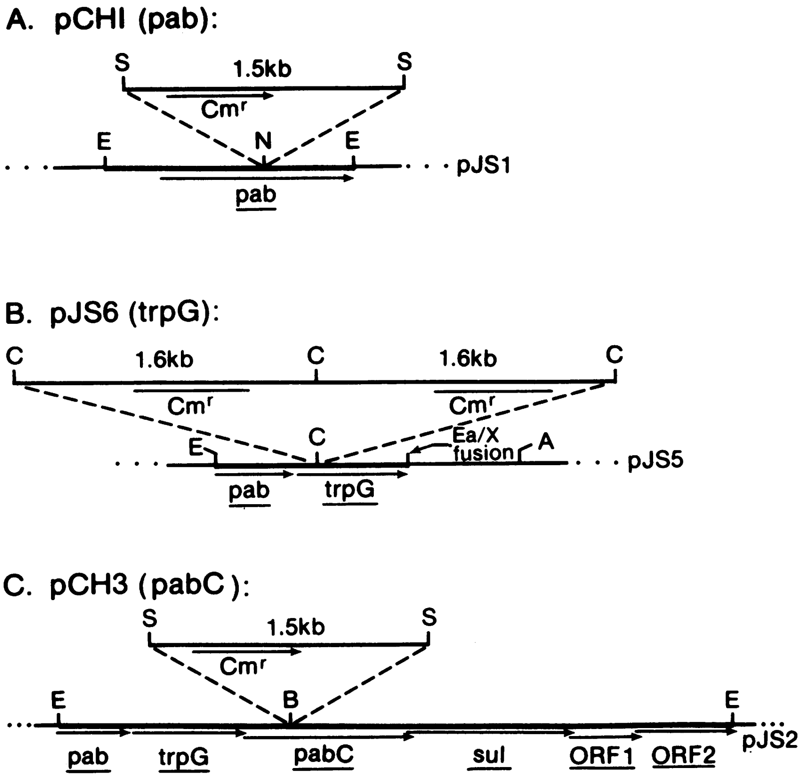

FIG. 6. Schematic diagrams of plasmids containing insertionally inactivated *pab*, *trpG*, and *pabC* genes. (A) pCH1 contains the Cm^r gene from pMI1101 inserted into the *pab* gene of pJS1. (B) pJS6 contains two Cm^r gene-containing segments from pC194 inserted into the *trpG* gene of pJS5. The orientation of the two Cm^r cassettes was not determined. (C) pCH3 contains the Cm^r gene from pMI1101 inserted into the *pabC* gene of pJS2. A, *Ava*I; B, *Bss*HII; C, *Cla*I; E, *Eco*RI; N, *Nru*I.

1A2ΩpCH2 (data not shown). The results were the same with both strains whether high- or low-ammonia-containing medium was used. Speculation concerning the absence of an absolute requirement for *p*-aminobenzoic acid in this strain is included in the Discussion.

Figure 7C shows the growth of *B. subtilis* 1A2ΩpJS6, the strain with the insertionally inactivated *trpG* gene, in the four media. An absolute requirement for tryptophan and a partial requirement for *p*-aminobenzoic acid were evident. Essentially no growth occurred in basal medium or in basal medium plus *p*-aminobenzoic acid. Growth did occur in basal medium plus tryptophan, but the generation time was 2.2 times longer than in basal medium plus tryptophan plus *p*-aminobenzoic acid (4.7 h in basal medium plus tryptophan and 2.1 h in basal medium plus tryptophan plus *p*-aminobenzoic acid). The basal medium for the above experiment contained 0.9 mM ammonium sulfate. When a similar experiment was conducted with a basal medium containing 45 mM ammonium sulfate, the growth was reasonably good even in basal medium (generation time, 3.0 h; data not shown). This was expected since ammonia, when present at a high concentration, can be used rather than glutamine for both the anthranilate synthase and *p*-aminobenzoic acid synthase reactions; i.e., glutamine amidotransferase (TrpG) is not absolutely required. The explanation for the partial (rather than complete) requirement for *p*-aminobenzoic acid, when 0.9 mM ammonium sulfate was present in the growth medium (Fig. 1), is less certain (see Discussion).

Figure 7D shows the growth of *B. subtilis* 1A2ΩpCH3, the strain with the insertionally inactivated *pabC* gene, in the four media. A definite requirement for *p*-aminobenzoic acid only was exhibited. The generation time in the absence of *p*-aminobenzoic acid (i.e., in basal medium and in basal medium plus tryptophan) was about 7.4 h. The generation time in the presence of *p*-aminobenzoic acid (i.e., in basal medium plus *p*-aminobenzoic acid and in basal medium plus tryptophan plus *p*-aminobenzoic acid) was 1.8 h.

B. subtilis 1A2ΩpCH3 has the Cm^r cassette from pMI1101 inserted so that the Cm^r gene is in the same orientation as *pabC*. The strain with the cassette inserted in the opposite orientation showed a definite growth defect in all media, including basal plus tryptophan plus *p*-aminobenzoic acid (data not shown). Perhaps the insertion affected expression of genes in the operon downstream from *pabC*. It is interesting that no such polarity effect was noted with the similar insertion of pCH2 into *pab* (data not shown).

DISCUSSION

The 4.9-kb chromosomal DNA fragment from *B. subtilis* that was analyzed apparently has at least six genes that are probably part of a single operon. No typical *B. subtilis* promoter was observed preceding *pab*, the first gene of the cluster. Thus, it is possible that one or more additional genes are upstream of *pab* on the chromosome. Also, only a part of *ORF2* is included in the cloned fragment. Thus, there may be other genes in the operon located downstream from *ORF2*. The distance between *pab* and *trpG* is too small (13 bp) for a promoter to be present. The other genes overlap. The apparent initiation codon of *pabC* precedes the termination codon of *trpG* by 1 bp. Similar overlaps occur between *pabC* and *sul* (19 bp), *sul* and *ORF1* (8 bp), and *ORF1* and *ORF2*

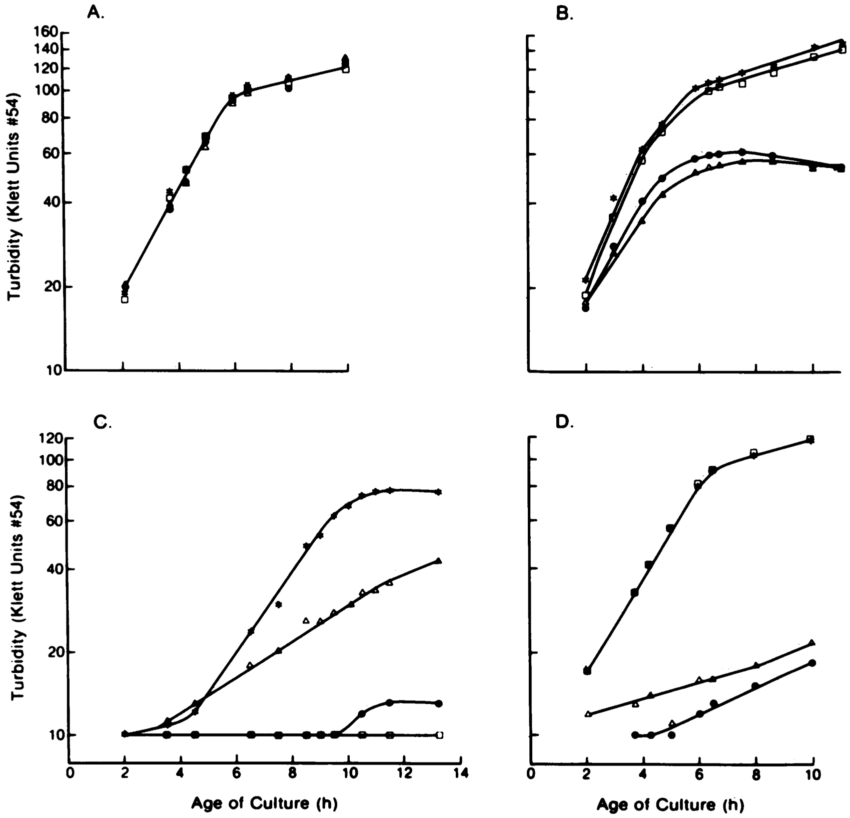

FIG. 7. Growth requirements of prototrophic *B. subtilis* and the strains containing the insertionally inactivated *trpG*, *pab*, and *pabC* genes. (A) *B. subtilis* 1A2, the prototrophic strain; (B) *B. subtilis* 1A2ΩpCH1 *pab*; (C) *B. subtilis* 1A2ΩpJS6 *trpG*; (D) *B. subtilis* 1A2ΩpCH3 *pabC*. The basal medium for growth of all of the strains except the one containing the insertionally inactivated *trpG* gene contained the following components in the proportions indicated: 100 ml of BW salts; 1 ml of 50% glucose; 0.1 ml of 1 M $MgSO_4$; and 5.0 ml of BW amino acids without tryptophan, glycine, and methionine. The basal medium for growth of the strain containing the insertionally inactivated *trpG* gene contained the following components in the proportions indicated: 100 ml of low-ammonia BW salts [containing 0.012% $(NH_4)_2SO_4$]; 1 ml of 50% glucose; 0.1 ml of 1 M $MgSO_4$. The ammonium sulfate concentrations in the BW salts used in the two media were 45 and 0.09 mM, respectively. Turbidity measurements were made during growth in basal medium (●) and in basal medium supplemented with tryptophan (20 μg/ml) (△), *p*-aminobenzoic acid (10 μg/ml) (□), and both substances (*).

(4 bp). This gene arrangement strongly suggests that all of the cloned genes belong to a single operon, an operon that we refer to as a folic acid biosynthetic operon because of the nature of four of the six genes. Overlapping genes are also apparent in other operons in *B. subtilis* and *E. coli*. For example, 75% of the genes of the *pur* and *trp* operons in *B. subtilis* are overlapping, a situation suggesting the occurrence of translational coupling (45).

To gain more information on the biological functions of Pab, TrpG, and PabC, the genes were insertionally inactivated and then introduced by replacement into the chromosome of prototrophic *B. subtilis* 1A2. Insertional inactivation of *pab* resulted in a partial requirement for *p*-aminobenzoic acid. It is uncertain why this requirement is not complete. One interesting possibility is that the anthranilate synthase large subunit (the *trpE* gene product) substitutes to a limited extent for Pab.

Insertional inactivation of *trpG* resulted in an absolute requirement for tryptophan and a partial requirement for *p*-aminobenzoic acid. An absolute requirement for both metabolites might have been expected, since TrpG (glutamine amidotransferase) of *B. subtilis* (like that of *A. calcoaceticus* [40] and *P. acidovorans* [4]) is amphibolic, functioning in synthesis of both anthranilate and *p*-aminobenzoate. The partial rather than complete requirement for *p*-aminobenzoate may be due to partial substitution of ammonia for glutamine. Ammonia at high concentrations is known to substitute for glutamine in both enzymatic reactions.

Insertional inactivation of *pabC* resulted in a complete requirement for *p*-aminobenzoate. Nichols et al. (31) recently showed that *p*-aminobenzoate synthesis in *E. coli* requires three polypeptides: a PabB subunit analogous to the *B. subtilis* Pab; a PabA subunit analogous to TrpG of *B. subtilis*; and a newly discovered enzyme X. In vitro synthesis of *p*-aminobenzoate from chorismate and glutamine required all three proteins. Nichols et al. (31) presented evidence that PabA and PabB react with chorismic acid and glutamine to form a diffusible intermediate that is converted by enzyme X to *p*-aminobenzoate. The nature of the diffusible intermediate is unknown, but it was speculated that it is 4-amino-4-deoxychorismate, a substance known to be convertible to *p*-aminobenzoate by crude preparations of *p*-aminobenzoate synthase from *E. coli* (42). One exciting possibility is that PabC in *B. subtilis* serves the same function as enzyme X in *E. coli*. The molecular weight of enzyme X of *E. coli* is 49,000 (31) and the enzyme consists of two identical subunits (M_r, 25,000) (B. P. Nichols, personal communication). The subunit's molecular weight indicates that it must consist of about 227 amino acids. PabC of *B. subtilis* has 293 amino acids. Neither enzyme X nor its gene in *E. coli* has yet been sequenced. A search of GenBank for other proteins possessing significant amino acid sequence similarity to PabC was conducted. Only IlvE from *E. coli* (21), the amino acid aminotransferase functioning in biosynthesis of isoleucine and valine, had a significantly similar sequence. The normalized alignment score (8) for the two proteins was 219, a value indicating that the relatedness of the two proteins is probably significant. More information concerning the role of PabC in *p*-aminobenzoate synthesis is required before we can speculate meaningfully about the similarity between the two proteins.

Data were presented indicating that the *sul* gene was to the right of *trpG* on the cloned fragment in pK0101 (Fig. 1). It is probable that the identification of the *sul* gene is correct, based on the following evidence. The amino acid sequence had significant base sequence similarity with the dihydropteroate synthase of *S. pneumoniae* (25). The normalized alignment score for the two proteins was 246, a value indicating almost certain relatedness between the two proteins. The size of the two genes is roughly similar; 942 bp for the dihydropteroate synthase gene of *S. pneumoniae* and 855 bp for *sul* of *B. subtilis*. Dihydropteroate synthase catalyzes the condensation of *p*-aminobenzoate with 6-CH_2OH-7,8-H_2-pterin pyrophosphate to form 7,8-dihydropteroate, the immediate precursor of folic acid. Thus, its presence in an operon with *p*-aminobenzoate synthase genes is logical. The *sul-d* mutation to sulfanilamide resistance in *S. pneumoniae* involves a 6-bp insertion in the dihydropteroate synthase gene (25) that causes a change in the kinetic properties of the enzyme. The mutant enzyme exhibits an increased K_m for *p*-aminobenzoate and an increased K_i for sulfanilamide (32). Therefore, it is very likely that the gene called *sul* in the *B. subtilis* chromosome is the dihydropteroate synthase gene. Presumably, if *sul* was insertionally inactivated and placed within the *B. subtilis* chromosome, the organism would be auxotrophic for folic acid. The impermeability of *B. subtilis* for folic acid (I. P. Crawford, unpublished observation) prevents direct testing of this hypothesis.

No sequence matches were observed between either the deduced amino acid sequences of ORF1 or the portion of ORF2 cloned and proteins in GenBank. Whether they are involved in folic acid biosynthesis remains to be determined. The size of ORF1 is similar to the H_2-pteridine-CH_2OH pyrophosphokinase of *E. coli*. (Neither the *E. coli* enzyme nor its gene has been sequenced.) This enzyme catalyzes the reaction of ATP and H_2-pterin-CH_2OH to form AMP and 6-CH_2OH-7,8-H_2-pterin pyrophosphate, the substance that subsequently reacts with *p*-aminobenzoic acid. The molecular weight of the *E. coli* enzyme is 15,000 (37), which should correspond to about 136 amino acids. ORF1 of *B. subtilis* has 120 amino acids. Although this correspondence is close, further investigation is required to identify ORF1.

There is a logical difficulty in interpreting the results of the gene inactivation experiments discussed above. These experiments involved insertion of Cm^r cassettes in the first three genes of the operon. It is theoretically possible that the phenotypes observed could be due to polarity effects on one or more downstream genes. In fact, the chloramphenicol resistance gene used does have a typical rho-independent transcriptional terminator (15). As indicated in Results, *B. subtilis* 1A2ΩpCH2, the strain with the insertion in *pabC* in an orientation opposite to that of *pabC*, did exhibit a growth defect that was not corrected by addition of *p*-aminobenzoate. An effect on expression of *sul*, the gene immediately downstream from *pabC*, for example, would be expected to result in a deficiency for folic acid not corrected for by addition of *p*-aminobenzoate. The strain with the Cm^r cassette inserted into *pabC* in the same orientation as *pabC* exhibited a growth defect correctable by *p*-aminobenzoate. Therefore, in this case there was no apparent polarity effect. No apparent polarity effect was observed with the strain containing the insertionally inactivated *trpG* gene. The requirements for tryptophan and *p*-aminobenzoate were observed in a low-ammonia-containing medium only, as would be expected of a *trpG* mutant. The phenotype resulting from inactivation of *pab* was a growth requirement for *p*-aminobenzoate only, in both high- and low-ammonia-containing media. If the phenotype was due to a polar effect on the first downstream gene, *trpG*, then both tryptophan and *p*-aminobenzoate would have been required in a low-ammonia-containing medium.

In summary, the phenotypes observed in the gene inactivation experiments were the phenotypes expected, based on the functions we proposed for these genes. Admittedly, however, the possibility of polar effects cannot be completely ruled out. More conclusive results might be obtained by complementation studies in which the strains containing inactivated chromosomal genes are transformed with plasmids containing individual good genes corresponding to the inactivated genes.

A logical question is why polar effects of gene inactivation were not more obvious. No definite answer can be given without additional experimentation. Perhaps the transcriptional terminator at the end of the Cm^r gene is not efficient and some readthrough occurs. There may be weak promoters within the operon. Another possibility is that the operon is subject to some type of antipolarity control.

As stated previously, synthesis of TrpG in *B. subtilis* is inhibited by tryptophan (17). The location of the *B. subtilis trpG* in an apparent folate operon makes this regulation paradoxical. It does not seem reasonable that transcriptional control of the entire operon by tryptophan would occur, since this would probably cause a deficiency in *p*-aminobenzoate. Translational control could, presumably, be exerted on *trpG* without affecting expression of the other genes of the operon. A clue that translational control is likely is provided by the finding of an apparent Mtr-binding site (22, 23) between −19 and −10, with respect to the initiation codon of *trpG*: the sequence AGATGAGGT. Control of the *trp* operon of *B. subtilis*, the operon containing all of the tryptophan biosynthetic genes except *trpG*, is mediated by a regulatory protein referred to as Mtr (methyltryptophan resistance) (14). Mtr activated by tryptophan appears to bind to the leader transcript and causes transcription termination (22). Binding of Mtr also is believed to inhibit translation of nonterminated *trpE* mRNA indirectly by favoring a transcript secondary structure that sequesters the *trpE* ribosome-binding site (22). The proposed Mtr-binding site overlaps with the probable *trpG* ribosome-binding site, GAGGTGA (Fig. 2). Therefore, we propose that Mtr, activated by tryptophan, regulates translation of *trpG* mRNA by preventing binding of ribosomes to the *trpG* ribosome-binding site. (The possibility was mentioned above that internal promoters exist within the operon. If so, regulation by tryptophan of transcription of the operon from one promoter might not totally shut down synthesis of *p*-aminobenzoate.)

In summary, a 4.8-kb fragment of *B. subtilis* chromosomal DNA contains five complete genes and one partial gene, all of which seem to be part of a single operon. Three of these genes, *pab*, *trpG*, and *pabC*, are required for synthesis of *p*-aminobenzoate. The *trpG* gene is an interesting gene because its gene product is amphibolic, acting as the glutamine amidotransferase for both *p*-aminobenzoate synthase and anthranilate synthase. The *pabC* gene may encode a third polypeptide required for synthesis of *p*-aminobenzoate, perhaps the *B. subtilis* analog of enzyme X discovered in *E. coli* (31). The *sul* gene probably encodes dihydropteroate synthase. Since at least four of the six genes of the operon are involved in folic acid biosynthesis, we feel justified in referring to the operon as a folic acid operon.

ACKNOWLEDGMENTS

We thank William Burke for the *B. subtilis* strain containing pK0101, the plasmid with the cloned segment of chromosomal DNA sequenced in this study. We are grateful to Charles Yanofsky, Brian Nichols, and Michael Feiss for encouragement and suggestions. We also thank Linda Johnson for secretarial assistance.

This work was supported by grant DMB 8606653 (I.P.C.) from the National Science Foundation.

ADDENDUM IN PROOF

After this paper was accepted, we became aware of the sequence of the H_2-pteridine-CH_2OH pyrophosphokinase (PPPK) gene, *sulD*, of *Streptococcus pneumoniae* (P. Lopez, B. Greenberg, and S. A. Lacks, J. Bacteriol. **172:**4766–4774, 1990). Hence we used the "Bestfit" program of the GCG package (J. Devereux, P. Haeberli, and O. Smithies, Nucleic Acids Res. **12:**387–395, 1984) to compare the deduced protein sequence (270 amino acid residues) of the *S. pneumoniae* enzyme with those for ORF1 and ORF2 of *B. subtilis*, as given in our paper. ORF1 (120 residues) aligned optimally with the N-terminal segment of the *S. pneumoniae* sequence, with a level of similarity indicative of probable relatedness. ORF2 (162 residues; incomplete) aligned optimally with the *S. pneumoniae* sequence immediately downstream from that showing best fit with ORF1, with a similarity level indicating certain relatedness between ORF2 and the *S. pneumoniae* protein. These findings suggest the following interpretations. Either ORF1 or (the complete) ORF2 would be a candidate for the gene encoding a *B. subtilis* PPPK, of a size similar to that of the *E. coli* enzyme (see Discussion), with the other ORF encoding an enzyme of unknown function. The larger *S. pneumoniae sulD* gene could conceivably encode a multifunctional enzyme combining moieties analogous to that of ORF1 and ORF2 of *B. subtilis*. Such fusion proteins are well known for the *trp* biosynthetic pathway of different bacterial species (7).

LITERATURE CITED

1. **Anagnostopoulos, C., and I. P. Crawford.** 1967. Le groupe des genes régissant la biosynthèse au tryptophan chez *Bacillus subtilis*. C. R. Acad. Sci. **265:**93–96.
2. **Bakhiet, N., and D. P. Stahly.** 1985. Studies on transfection and transformation of protoplasts of *Bacillus larvae*, *Bacillus subtilis*, and *Bacillus popilliae*. Appl. Environ. Microbiol. **49:**577–581.
3. **Bott, K. F., and G. A. Wilson.** 1968. Metabolic and nutritional factors influencing the development of competence for transfection of *Bacillus subtilis*. Bacteriol. Rev. **32:**370–378.
4. **Buvinger, W. E., L. C. Stone, and H. E. Heath.** 1981. Biochemical genetics of tryptophan synthesis in *Pseudomonas acidovorans*. J. Bacteriol. **147:**62–68.
5. **Carlton, B. C., and D. D. Whitt.** 1969. The isolation and genetic characterization of mutants of the tryptophan system of *Bacillus subtilis*. Genetics **62:**445–460.
6. **Cohen, S. N., A. C. Y. Chang, and L. Hsu.** 1973. Construction of biologically functional bacterial plasmids *in vitro*. Proc. Natl. Acad. Sci. USA **70:**3240–3244.
7. **Crawford, I. P.** 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu. Rev. Microbiol. **43:**567–600.
8. **Doolittle, R. F.** 1986. Of urfs and orfs: a primer on how to analyze derived amino acid sequences. University Science Books, Mill Valley, Calif.
9. **Essar, D. W., L. Eberly, C.-Y. Han, and I. P. Crawford.** 1990. DNA sequences and characterization of four early genes of the tryptophan pathway in *Pseudomonas aeruginosa*. J. Bacteriol. **172:**853–866.
10. **Goncharoff, P., and B. P. Nichols.** 1988. Evolution of aminobenzoate synthases: nucleotide sequences of *Salmonella typhimurium* and *Klebsiella aerogenes pabB*. Mol. Biol. Evol. **3:**191–204.
11. **Gryczan, T. J., and D. Dubnau.** 1978. Construction and properties of chimeric plasmids in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **75:**1428–1432.
12. **Henner, D. J., L. Band, and H. Shimotsu.** 1985. Nucleotide sequence of the *Bacillus subtilis* tryptophan operon. Gene **34:**169–177.
13. **Hoch, S. O., C. Anagnostopoulos, and I. P. Crawford.** 1969. Enzymes of the tryptophan operon of *Bacillus subtilis*. Biochem. Biophys. Res. Commun. **35:**838–844.
14. **Hoch, S. O., C. W. Roth, I. P. Crawford, and E. W. Nester.** 1971. Control of tryptophan biosynthesis by the methyltryptophan resistance gene in *Bacillus subtilis*. J. Bacteriol. **105:**38–45.
15. **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.
16. **Iordanescu, S., M. Surdeanu, P. Della Latta, and R. Novick.** 1978. Incompatibility and molecular relationships between small staphylococcal plasmids carrying the same resistance marker. Plasmid **1:**468–479.
17. **Kane, J. F.** 1977. Regulation of a common amidotransferase subunit. J. Bacteriol. **132:**419–425.
18. **Kane, J. F., W. M. Holmes, and R. A. Jensen.** 1972. Metabolic interlock: the dual function of a folate pathway gene as an extraoperonic gene of tryptophan biosynthesis. J. Biol. Chem. **247:**1587–1596.
19. **Kaplan, J. B., P. Goncharoff, A. M. Seibold, and B. P. Nichols.** 1984. Nucleotide sequence of the *Acinetobacter calcoaceticus trpGDC* gene cluster. Mol. Biol. Evol. **1:**456–472.
20. **Kaplan, J. B., W. K. Merkel, and B. P. Nichols.** 1985. Evolution of glutamine aminotransferase genes: nucleotide sequences of the *pabA* genes from *Salmonella typhimurium*, *Klebsiella aerogenes* and *Serratia marcescens*. J. Mol. Biol. **183:**327–340.
21. **Kuramitsu, S., T. Ogawa, H. Ogawa, and H. Kagamiyama.** 1985. Branched-chain amino acid aminotransferase of *Escherichia coli*: nucleotide sequence of the *ilvE* gene and its deduced amino acid sequence. J. Biochem. **97:**993–999.
22. **Kuroda, M. I., D. Henner, and C. Yanofsky.** 1988. *cis*-Acting sites in the transcript of the *Bacillus subtilis trp* operon regulate expression of the operon. J. Bacteriol. **170:**3080–3088.
23. **Kuroda, M. I., H. Shimotsu, D. J. Henner, and C. Yanofsky.** 1986. Regulatory elements common to the *Bacillus pumilus* and *Bacillus subtilis trp* operons. J. Bacteriol. **167:**792–798.
24. **Lagrimini, L. M., S. T. Brentano, and J. E. Donelson.** 1984. A DNA sequence analysis package for the IBM personal computer. Nucleic Acids Res. **12:**605–614.
25. **Lopez, P., M. Espinosa, B. Greenberg, and S. A. Lacks.** 1987. Sulfonamide resistance in *Streptococcus pneumoniae*: DNA sequence of the gene encoding dihydropteroate synthase and characterization of the enzyme. J. Bacteriol. **169:**4320–4326.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Matsui, K., K. Sano, and E. Ohtsubo.** 1986. Complete nucleotide and deduced amino acid sequences of the *Brevibacterium lac-*

tofermentum tryptophan operon. Nucleic Acids Res. **14**:10113–10114.

28. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
29. **McDonald, K. O., and W. F. Burke, Jr.** 1982. Cloning of the *Bacillus subtilis* sulfanilamide resistance gene in *Bacillus subtilis*. J. Bacteriol. **149**:391–394.
30. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Nichols, B. P., A. M. Seibold, and S. Z. Doktor.** 1989. *para*-Aminobenzoate synthesis from chorismate occurs in two steps. J. Biol. Chem. **264**:8587–8601.
32. **Ortiz, P. J., and R. D. Hotchkiss.** 1966. The enzymatic synthesis of dihydrofolate and dihydropteroate in cell-free preparations from wild-type and sulfonamide-resistant pneumococcus. Biochemistry **5**:67–74.
33. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
34. **Perkins, J. P., and P. J. Youngman.** 1984. A physical and functional analysis of Tn*917*, a *Streptococcus* transposon in the Tn*3* family that functions in *Bacillus*. Plasmid **12**:119–138.
35. **Piggot, P. J., C. A. M. Curtis, and H. DeLencastre.** 1984. Use of integrational plasmid vectors to demonstrate the polycistronic nature of a transcriptional unit (*spoIIA*) required for sporulation of *Bacillus subtilis*. J. Gen. Microbiol. **130**:2123–2136.
36. **Piggot, P. J., and J. A. Hoch.** 1985. Revised genetic linkage map of *Bacillus subtilis*. Microbiol. Rev. **49**:158–179.
37. **Richey, D. P., and G. M. Brown.** 1969. The biosynthesis of folic acid. IX. Purification and properties of the enzymes required for the formation of dihydropteroic acid. J. Biol. Chem. **244**:1582–1592.
38. **Rodriguez, R. L., and R. C. Tait.** 1983. Recombinant DNA techniques: an introduction, p. 164–165. Addison-Wesley Publication Co., London.
39. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
40. **Sawula, R. V., and I. P. Crawford.** 1972. Mapping of the tryptophan genes of *Acinetobacter calcoaceticus* by transformation. J. Bacteriol. **112**:797–805.
41. **Shimotsu, H., M. I. Kuroda, C. Yanofsky, and D. J. Henner.** 1986. Novel form of transcription attenuation regulates expression of the *Bacillus subtilis* tryptophan operon. J. Bacteriol. **166**:461–471.
42. **Teng, C.-Y. P., B. Ganem, S. Z. Doktor, B. P. Nichols, R. K. Bhatnagar, and L. C. Vining.** 1985. Total synthesis of (±)-4-amino-4-deoxychorismic acid: a key intermediate in the biosynthesis of p-aminobenzoic acid and L-(p-aminophenyl)alanine. J. Am. Chem. Soc. **107**:5008–5009.
43. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC vectors. Gene **33**:103–119.
44. **Yousaf, S. I., A. R. Carroll, and B. E. Clarke.** 1984. A new and improved method for 3′-end labeling DNA using [α-^{32}P]ddATP. Gene **27**:309–313.
45. **Zalkin, H., and D. J. Ebbole.** 1988. Organization and regulation of genes encoding biosynthetic enzymes in *Bacillus subtilis*. J. Biol. Chem. **263**:1595–1598.