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, DONALD P. STAHLRY, CHUN-YA HAN, ERICH W. SIX, and IRVING P. CRAWFORD†

Department of Microbiology, University of Iowa, Iowa City, Iowa 52242, and Department of Biology, King’s College, Wilkes-Barre, Pennsylvania 18711.

Received 7 May 1990/Accepted 12 September 1990

McDonald and Burke (J. Bacteriol. 149:391–394, 1982) previously cloned a sulfinamide-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 glutamate 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 glutamate (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 sulfinamide 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

<table>
<thead>
<tr>
<th>Bacterial strain or plasmid</th>
<th>Description or genotype*</th>
<th>Reference or sourceb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. subtilis</strong> 1A2</td>
<td>Prototrophic</td>
<td>BGSC</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A491</td>
<td>trpG7 dfpA24 trk-24</td>
<td>BGSC</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> ASB342(KpK0101)</td>
<td>trpC2 (Km’ pab” trpG” pabC” Sul’ ORFI” )</td>
<td>W. F. Burke; 29</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A491(pK0101)</td>
<td>Same as 1A491 but (Km’ pab” trpG” pabC” Sul’ ORFI” )</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A491(pJS3)</td>
<td>Same as 1A491 but (Km’ Em’ trpG” )</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A491(pJS4)</td>
<td>Same as 1A491 but (Km’ Em’ pab” trpG” )</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A21IpJS6</td>
<td>Same as 1A2 but Cm’ pabG”</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A21pCH1</td>
<td>Same as 1A2 but Cm’ pabG”</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A2O2pCH2</td>
<td>Same as 1A2 but Cm’ pabG”</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A2pCH3</td>
<td>Same as 1A2 but Cm’ pabC”</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1A2O3pCH4</td>
<td>Same as 1A2 but Cm’ pabC”</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. subtilis</strong> 1E27(pBD9)</td>
<td>thr-5 trpC2 (Em’ Km’ )</td>
<td>BGSC</td>
</tr>
<tr>
<td><strong>B. larvae</strong> NRRL B-3555(pC194)</td>
<td>recA1 endA1 gyrA96 thi hsdR17 supE44 relA1 Δ(lac-proAB)(F’ trd36 proAB lacP2 ΔM15)</td>
<td>43</td>
</tr>
<tr>
<td>E. coli J1M09</td>
<td>Same as J1M09 but (Ap”)</td>
<td>This study</td>
</tr>
<tr>
<td>E. coli J1M09(pJS1)</td>
<td>Same as J1M09 but (Ap’ trpG’ pabC’ Sul’ ORFI’ )</td>
<td>This study</td>
</tr>
<tr>
<td>E. coli J1M09(pJS3)</td>
<td>Same as J1M09 but (Ap’ trpG’ pabC’ )</td>
<td>This study</td>
</tr>
<tr>
<td>E. coli J1M09(pJS5)</td>
<td>Same as J1M09 but (Ap’ pabC’ )</td>
<td>This study</td>
</tr>
<tr>
<td>E. coli J1M09(pJS6)</td>
<td>Same as J1M09 but (Ap’ Cm’ )</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmid</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pK0101</td>
<td>Km’ pab” trpG” pabC” Sul’ ORFI”</td>
<td>29</td>
</tr>
<tr>
<td>pJS1</td>
<td>Ap’</td>
<td>This study</td>
</tr>
<tr>
<td>pJS2</td>
<td>Ap’ trpG” pabC” Sul’ ORFI”</td>
<td>This study</td>
</tr>
<tr>
<td>pJS3</td>
<td>Km’ Em’ trpG”</td>
<td>This study</td>
</tr>
<tr>
<td>pJS4</td>
<td>Km’ Em’ pab” trpG”</td>
<td>This study</td>
</tr>
<tr>
<td>pJS5</td>
<td>Ap’ trpG”</td>
<td>This study</td>
</tr>
<tr>
<td>pJS6</td>
<td>Ap’, Cm’ cassette in trpG of pJS5</td>
<td>This study</td>
</tr>
<tr>
<td>pCH1</td>
<td>Ap’, Cm’ cassette (orientation same as operon) in pab of pJS1</td>
<td>This study</td>
</tr>
<tr>
<td>pCH2</td>
<td>Ap’, Cm’ cassette (orientation opposite of operon) in pab of pJS1</td>
<td>This study</td>
</tr>
<tr>
<td>pCH3</td>
<td>Ap’, Cm’ cassette (orientation same as operon) in pab of pJS2</td>
<td>This study</td>
</tr>
<tr>
<td>pCH4</td>
<td>Ap’, Cm’ cassette (orientation opposite of operon) in pab of pJS2</td>
<td>This study</td>
</tr>
<tr>
<td>pBD9</td>
<td>Km’ Em’</td>
<td>11</td>
</tr>
<tr>
<td>pCI194</td>
<td>Cm’</td>
<td>15, 16</td>
</tr>
<tr>
<td>pUC18</td>
<td>Ap’</td>
<td>BRL</td>
</tr>
</tbody>
</table>

*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-CH₂-OH-7,8-dihydropterin pyrophosphate to form 7,8-dihydropteroate, the immediate precursor of folic acid. This, 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 MgSO₄, 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% K₂HPO₄, 0.76% KH₂PO₄, 0.1% trisodium citrate, and 0.6% (NH₄)₂SO₄ (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 MgSO₄, 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 MgSO₄; 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-isooamyl alcohol (50:48:2, vol/vol/vol) followed by chloroform-isooamyl 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 EcoRI 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.

Plasmid pK0101, isolated from *B. subtilis* ASB342 (pK0101), was cut with EcoRI. The 1.3- and 3.6-kb fragments derived from the insert were separately cloned into the *EcoRI* site of pUC18 (43) for ease of sequencing. Plasmids derived from transformants of *E. coli* JM109 containing the 1.3- and 3.6-kb *EcoRI* 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 *trpG* 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 *EcoRI* and *ClaI*, and both the 0.5-kb *EcoRI-ClaI* fragment and the 0.8-kb *ClaI-ClaI* fragment were isolated following electrophoresis in low-gelling-temperature agarose. Vector pBD9 (11) was cut into two fragments with *EcoRI* and *ClaI*. The largest fragment (5.5 kb) contains the origin of replication of
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 pKO101. The ligation mixture was transformed into *B. subtilis* 1A491 and Km<sup>+</sup> transformants were isolated. One of the transformants contained a plasmid, referred to as pJS3, that contained the two pKO101-derived fragments in the same order and orientations present in pKO101. This plasmid

![Nucleic acid sequence](image-url)

**Fig. 2.** Nucleic acid sequence of the 4,880 bp of *B. subtilis* chromosomal DNA in pKO101. 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).
contains the last 384 bp of \textit{pab} and the complete reading frame for the presumed \textit{trpG}; i.e., the only complete gene in the insert DNA is the presumed \textit{trpG}. \textit{pJS3} in \textit{B. subtilis 1A491} complemented the \textit{trpG} mutation, eliminating its partial \textit{p}-aminobenzoate growth requirement just as \textit{pKO101} did.

\textbf{Sequence determination.} The entire sequence of both strands of the \textit{pJS1} and \textit{pJS2} inserts was determined, with
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-
sequence 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. calcoaceticus TrpG, 167; B. lactofermentans TrpG, 145; and E. coli TrpG, 50.

Fig. 2—Continued.
**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.

![Alignment of the B. subtilis Pab protein with PabB of E. coli and TrpE of selected microorganisms.](https://jb.asm.org/content/36/7/7217/F3.large.jpg)

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.

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 of these two proteins was 7218.

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 of these two proteins was 7218, 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.

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 of these two proteins was 7218, 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.

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 of these two proteins was 7218, 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.
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

**FIG. 3—Continued.**
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.I., 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, TrpGpd. The stop sites for protein synthesis are marked by asterisks (*). (•) indicates residue identities in at least four of the five proteins.

sequested 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 segment in pK0101 to the right of the EagI target site. The EagI 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') gene of pMII101 (34; P. Youngman, personal communication) which resides on a 1.5-kb Sau3A fragment, was inserted by blunt-end ligation into the single NruI site of pJS1 that lies within the pab gene. Two plasmids were obtained by transformation of E. coli JM109: pCH1 had the Cm' fragment inserted in the pab gene in an orientation such that transcription of the Cm' gene would occur in the same direction as that of pab (Fig. 6A); pCH2 had the Cm' fragment inserted in the opposite orientation.

For insertional inactivation of trpG, we inserted the chloramphenicol resistance gene of pC194. The Cm' gene of pC194 resides on a 1.6-kb Clal fragment. Although trpG contains a single Clal site, there are two Clal 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 EagI and XbaI, isolating the 6.3-kb fragment, filling in the staggered ends by incubation with the Klone fragment of DNA polymerase, and performing blunt-end ligation. The ligation mixture was transformed into B. subtilis 1A491, and selection was made for Kmr'. The blunt-end ligation destroyed both the EagI and XbaI sites. This plasmid was named pJS4. The trpG gene in pJS4 was excised by cleavage with EcoRI and Aval and was inserted into pUC18 appropriately prepared by removal of the small Aval-EcoRI fragment from the polylinker region. The resulting plasmid, pJS5, was recovered following transformation of E. coli JM109. The Clal-Clal fragment of pC194 containing the Cm' gene was inserted into the Clal site of pJS5, generating pJS6 containing the insertionally inactivated trpG gene. The pJS6-containing isolate actually had two Cm' cassettes inserted into trpG (Fig. 6B).
The Cm' gene of pMI1101 was inserted into pabC. The 1.5-kb Smal fragment of pMI1101, which contains the Cm' gene, was inserted by blunt-end ligation into the single Bsh11 site of pJS2 that lies within pabC. Before ligation, the Bsh11-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' fragment inserted in pabC in an orientation such that transcription of the Cm' gene would occur in the same direction as that of pabC (Fig. 6C); pCH4 had the Cm' fragment inserted in the opposite orientation.

Integration of the insertional 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 insertional inactivated genes incorporated into the chromosome. Prior to transformation of B. subtilis 1A2, pCH1 and pCH2 were linearized with Smal. The transformants with the insertional inactivated pab gene are referred to as B. subtilis 1A2pCH1 and B. subtilis 1A2pCH2. pJS6 was linearized with BamHI. The strain with the integrated, insertional inactivated trpG is referred to as B. subtilis 1A2pJS6. pCH3 and pCH4 were linearized with Smal. The transformants with the insertional inactivated pabC are referred to as B. subtilis 1A2pCH3 and B. subtilis 1A2pCH4.

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 1A2pCH1, the strain with the insertional 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.
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 *Cmr* cassette from pMI1101 inserted so that the *Cmr* 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 1A2OpCH1 pab; (C) B. subtilis 1A2Op156 trpG; (D) B. subtilis 1A2OpCH3 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 MgSO4; 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% (NH4)2SO4]; 1 ml of 50% glucose; 0.1 ml of 1 M MgSO4. 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) (Δ), 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 amphoteric, 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 (Mr, 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-CH2OH-7,8-

H2-perin 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 H2-peridine-CH2OH pyrophosphokinase of E. coli. (Neither the E. coli enzyme nor its gene has been sequenced.) This enzyme catalyzes the reaction of ATP and H2-peridine-CH2OH to form AMP and 6-CH2OH-7,8-H2-perin 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' 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 1A21pCH2, 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' 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' 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 nonterminating 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 amphiolic, 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 8006653 (I.P.C.) from the National Science Foundation.

ADDITIONAL PROOF

After this paper was accepted, we became aware of the sequence of the H2-p-pteridine-CH2OH 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 naphthylamino group ORP being 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

zoate synthases: nucleotide sequences of Salmonella typhimu-
rium and Klebsiella aerogenes pabB. Mol. Biol. Evol. 3:191-
204.
45.
27. Matsui, K., K. Sano, and E. Ohtsubo. 1986. Complete nucleotide and deduced amino acid sequences of the Brevibacterium lac-

FOVIC ACID OPERON OF B. SUBTILIS 7225

VOL. 172, 1990

Downloaded from http://jb.asm.org/ on August 14, 2017 by guest


