JOURNAL OF BACTERIOLOGY, Jan. 1995, p. 66–74 Vol. 177, No. 1
0021-9193/95/$04.00+0

A Cluster of Four Genes Encoding Enzymes for Five Steps in the Folate Biosynthetic Pathway of *Streptococcus pneumoniae*

SANFORD A. LACKS,[1]* BILL GREENBERG,[1] AND PALOMA LOPEZ[2]

Biology Department, Brookhaven National Laboratory, Upton, New York 11973,[1] and Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain[2]

Received 29 August 1994/Accepted 1 November 1994

Two genes, *sulB* and *sulC*, in a folate biosynthetic gene cluster of *Streptococcus pneumoniae* were identified after determination of the DNA sequence between two previously reported genes, *sulA* and *sulD*, in a cloned segment of chromosomal DNA containing a mutation to sulfonamide resistance. The gene products, SulB and SulC, correspond to polypeptides of 49 and 21 kDa, respectively. SulC has GTP cyclohydrolase activity and catalyzes the first step in the folate biosynthetic pathway. SulB apparently has dihydrofolate synthetase activity in that it complements a *folC* mutant of *Escherichia coli* and thus catalyzes the last step in the pathway. Prior work showed that SulA, a dihydropteroate synthase, and SulD, a bifunctional enzyme, catalyze three intervening steps. Mapping of the mRNA transcribed from the operon was consistent with its beginning at a promoter with a −35 site (gTGtCc) and an extended −10 site (T-TG-TAaAAT) and its termination at the end of a hairpin structure, which would give a transcript 3,745 nucleotides in length. SulC showed a considerable conservation of sequence by comparison with proven or putative GTP cyclohydrolases from four unrelated species, with 38 to 53% of the residues being identical. A similar comparison of SulB with dihydrofolate synthetases showed an identity of only 26 to 37%. Overall, comparisons of the five folate biosynthetic enzymes in different species suggest that *S. pneumoniae* is related more closely to other gram-positive bacteria, less closely to eucaryotes, and least closely to the gram-negative *E. coli*. The varied arrangements of folate biosynthetic genes in different species imply an early evolutionary period of fluidity in genomic rearrangement.

Folate derivatives are essential cofactors in the biosynthesis of purines, pyrimidines, and amino acids in living cells. Whereas animals, including human beings, require folate in their diet, most bacterial species synthesize their own and are thus subject to inhibition by compounds, such as sulfonamides, that block folate biosynthesis. Sulfonamides and other folate antagonists are important chemotherapeutic agents used effectively against bacterial infections, human cancer, and more recently against *Pneumocystis* pneumonia resulting from AIDS (45).

The folate biosynthetic pathway depicted in Fig. 1 was elucidated by studies of cell extracts and purified enzymes from several bacterial species (32, 36), including *Streptococcus pneumoniae* (31). It begins with GTP, which is converted by GTP cyclohydrolase (GTPCH) to dihydroneopterin triphosphate. [Pterin is the trivial name for 2-amino-4-hydroxypteridine, and neopterin is the trivial name for 6-(D-*erythro*-1′,2′,3′-trihydroxypropyl)pterin.] The enzyme first breaks the five-member ring of guanine to release formate and an intermediate that recyclizes to give a pteridine ring. Removal of the phosphate residues from dihydroneopterin triphosphate must occur, presumably by the action of a phosphatase, but the nature of this enzyme has not been demonstrated in *S. pneumoniae* or any other system. Dihydroneopterin aldolase (DHNA) then acts on the product to give 6-hydroxymethyl-7,8-dihydropterin. The latter is converted to 6-hydroxymethyl-7,8-dihydropterin pyrophosphate by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (PPPK) and ATP. Dihydropteroate synthase (DHPS) catalyzes the linkage to *p*-aminobenzoate to give 7,8-dihydropteroate. Finally, the addition of glutamate to dihydropteroate to produce dihydrofolate is catalyzed by dihydrofolate synthetase (DHFS).

Genetic and molecular analysis of this pathway in *S. pneumoniae* began with the cloning of a 10-kb chromosomal fragment containing a marker for sulfonamide resistance (*sul*) in the vector plasmid pLS1 (18), with *S. pneumoniae* as the host species (38). Part of this fragment was transferred to an *Escherichia coli* expression system and shown to produce four polypeptides, SulA, SulB, SulC, and SulD (20, 22). The sulfonamide resistance mutation, *sul*-d, was located within *sulA*. The 34-kDa product of *sulA*, SulA, was shown to have DHPS activity (20). The *sul*-d mutation was sufficient to confer resistance to sulfonamides, but it did so only when present, *cis*, in the full gene cluster (20). SulD, a 31-kDa polypeptide coded for by *sulD*, was shown to be a bifunctional protein carrying both PPPK and DHNA activities (22, 23). On the basis of these findings, putative genes encoding DHPS, PPPK, and DHNA were identified in *Bacillus subtilis* (37) and *Pneumocystis carinii* (43, 44). Later, genes for PPPK and DHPS were independently identified in *E. coli* (7, 8, 40).

An *E. coli* gene, *folC*, encoding DHFS was cloned earlier (3, 4), and human and bacterial homologs were found (9, 26, 29, 41). Rat cDNA encoding GTPCH was cloned and sequenced (12), and a homolog in *B. subtilis* was identified (1). In the present work, we report the sequences of *sulB* and *sulC*, show that these genes encode the pneumococcal DHFS and GTPCH, compare the various homologs of these enzymes, and map the mRNA produced by the folate biosynthetic operon in *S. pneumoniae*. We also compare the arrangements of these genes in various species.

MATERIALS AND METHODS

Bacterial strains and plasmids. Strains and plasmids used are listed in Table 1.

Growth and transformation of bacteria. Cultures of *S. pneumoniae* were grown and transformed as previously described (17). Transformants were se-

* Corresponding author. Phone: (516) 282-3369. Fax: (516) 282-3407.

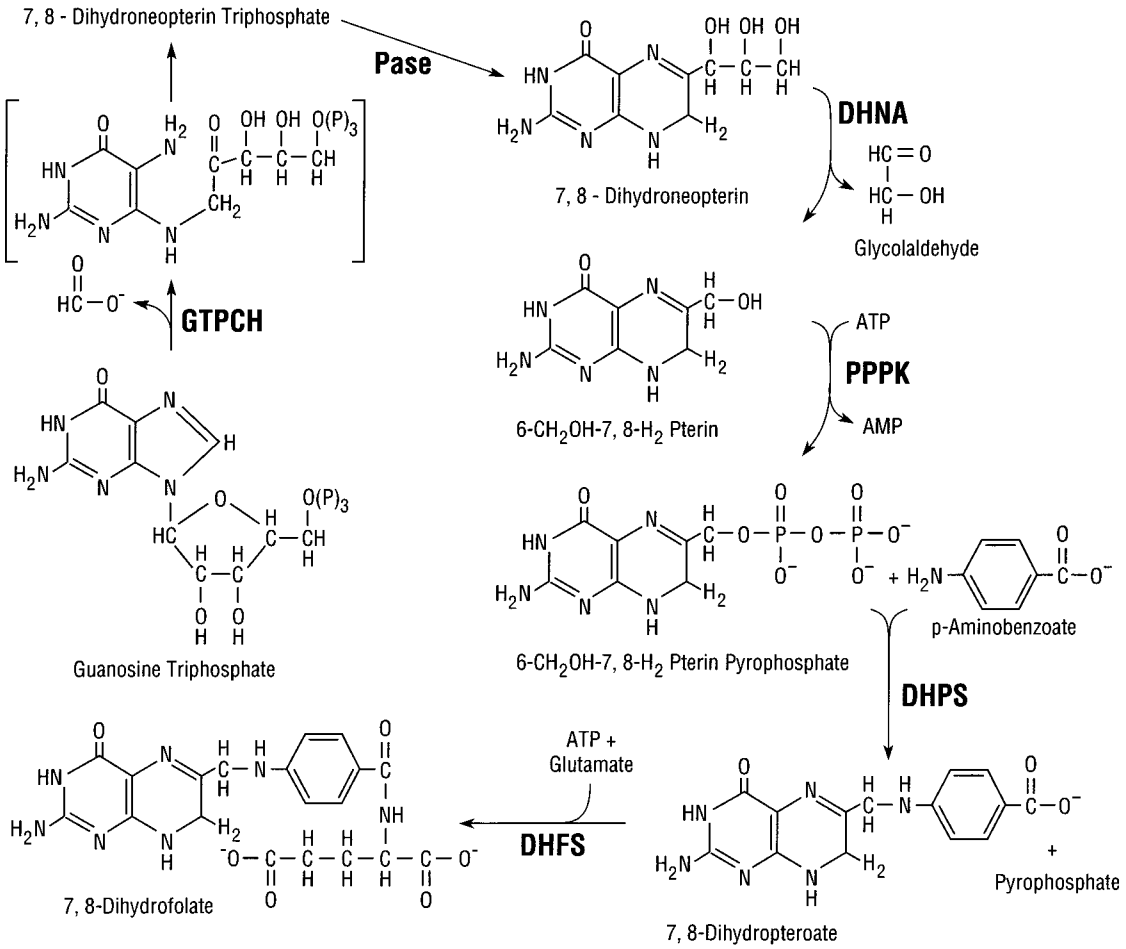

FIG. 1. Pathway for folate biosynthesis, beginning with guanosine triphosphate and proceeding clockwise. Identified enzymes include GTPCH, DHNA, PPPK, DHPS, and DHFS. The phosphatase(s) that presumably removes the phosphate residues of 7,8-dihydroneopterin triphosphate (Pase) has not been identified. Brackets indicate a presumed reaction intermediate. Triphosphate residues are indicated as $(P)_3$.

lected in agar medium containing tetracycline at 1.0 μg/ml. In general, *E. coli* was grown in L broth and transformed according to the method of Kushner (16). Transformants were selected with ampicillin at 50 μg/ml. To test the abilities of various plasmids to complement the *folC* defect in strain SF4, cultures were grown in L broth to an optical density at 600 nm of 0.4 and then diluted and plated in E medium (42) with and without methionine at 50 μg/ml. All transformations were carried out with plasmid DNA at 1.0 μg/ml.

DNA sequence determination. Purified pLS830 was prepared from cells of *S. pneumoniae* 708 by a modification of the procedure of Currier and Nester (6). Trace amounts of RNA were removed from the plasmid samples by treatment of 0.1 mg of DNA dissolved in 0.3 ml of buffer (10 mM Tris-HCl [pH 7.6]–1 mM EDTA) with 4.5 μl of pancreatic RNase (4 mg/ml; heated for 15 min at 90°C) for 30 min at 37°C. After shaking of the solution with an equal volume of buffer-equilibrated phenol, the DNA in the aqueous phase was precipitated with ethanol and dissolved in 0.2 ml of buffer. This material was then fractionated by gel filtration on a Bio-Gel A-1.5m column (0.9 by 20 cm). Fractions (1.0 ml) containing plasmid DNA were pooled. After plasmid cleavage with restriction enzymes, the DNA fragments were treated with alkaline phosphatase and labeled at their 5′ ends with [γ-^{32}P]ATP and T4 polynucleotide kinase (27). Nucleotide sequences on both strands of DNA were determined by the chemical method of Maxam and Gilbert (27).

TABLE 1. Bacterial strains and plasmids used

Strain or plasmid	Relevant feature(s)	Reference
S. pneumoniae 708	Derivative of R6, wild-type *sul* (*sul*-s)	21
E. coli BL21(DE3)	Inducible T7 RNA polymerase	39
E. coli SF4	F$^-$ *folC strA recA* Tn*10*:*srlC*	4
Plasmids		
pLS830	Contains cloned *sul* segment (*sul*-d)	21
pET-5	Contains T7 RNA polymerase promoter	33
pLS23	Intact pLS830 *sul* fragment in pET-5	22
pLS24	Deletion of pLS23 in *sulA*	22
pLS27	Deletion of pLS23 spanning *sulBC*	22
pLS29	Deletion in pLS23 of 4 bp in *sulB*	22
pLS30	Deletion in pLS23 of 4 bp in *sulC*	22
pLS31	Deletion in pLS23 of 4 bp in *sulD*	22

Preparation of cell extracts. Cultures of *E. coli* BL21(DE3) containing various plasmids were grown with shaking at 37°C in M9 medium with ampicillin at 0.2 mg/ml until an optical density at 600 nm of 0.5 was reached, at which time IPTG (isopropyl-β-D-thiogalactopyranoside) was added to 0.5 mM, and the culture was incubated for 90 min more. For protein purification, cells from 1 liter of culture were centrifuged, washed, and suspended in 10 ml of buffer containing 20 mM Tris-HCl (pH 7.6), 50 mM NaCl, 1 mM EDTA, and 3 mM β-mercaptoethanol. For enzyme assays, cells from 25 ml of an induced culture were suspended in 5 ml of buffer containing 0.1 M Tris-HCl (pH 8.5), 0.5 M NaCl, and 1 mM dithiothreitol. The cells were disrupted by passage through a French pressure cell at 20,000 lb/in^2, and the crude extract was clarified by centrifugation for 20 min at 20,000 × *g*. The protein was determined by the method of Lowry et al. (25).

Determination of amino-terminal polypeptide sequence. SulC protein purified by gel filtration as described (22) was subjected to preparative electrophoresis in the presence of sodium dodecyl sulfate on a 10 to 20% polyacrylamide gradient gel prepared from "ultrapure" acrylamide (International Biotechnologies). The polypeptide band was visualized by placing the gel in 4 M sodium acetate at 4°C for 30 min (13), and the band of interest was excised. The polypeptide was

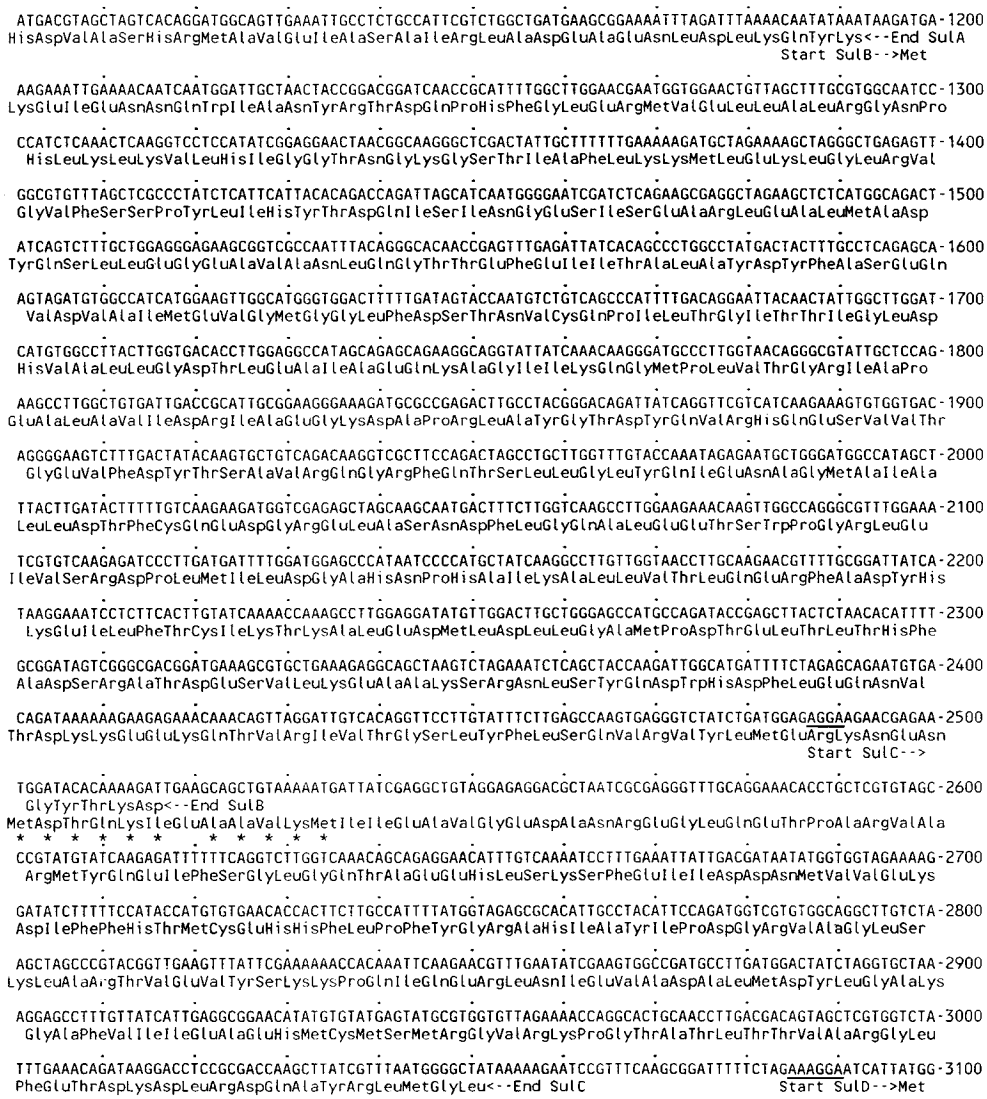

FIG. 2. Nucleotide sequence of a 2-kb portion of the *sul* segment of the *S. pneumoniae* chromosome containing *sulB* and *sulC*. The sequence of one DNA strand is shown with its 5′ end at 1101 and its 3′ end at 3100, with numbering extending from the reported *sulA* sequence (20). The predicted amino acid sequences of the SulB and SulC polypeptides are shown together with putative mRNA ribosome binding sites for SulC and SulD (underlined). Determination of the amino-terminal sequence of the purified SulC polypeptide revealed the residues marked with asterisks.

electroeluted with an ISCO sample concentrator with a chamber solution of 50 mM NH_4HCO_3 and 0.1% sodium dodecyl sulfate. The solution was lyophilized, and the residue was dissolved in 0.4 ml of water. This was repeated four times to remove NH_4HCO_3. Approximately 1 nmol of polypeptide was subjected to amino-terminal sequence determination in a gas-phase sequenator (Applied Biosystems).

Enzyme assay. GTPCH was assayed as described by Babitzke et al. (1). The mixtures contained, in a volume of 0.5 ml, 0.1 M Tris-HCl (pH 8.5), 25 mM EDTA, 1 mM dithiothreitol, 1 mM GTP, and 0.1 ml of the crude extract. After incubation at 42°C for 30 min, 0.5 ml of an aqueous suspension containing 20 mg of activated charcoal was added and the mixture was filtered through a type HA Millipore filter (pore size, 0.45 μm). After it was washed with 5 ml of water and 5 ml of 5% ethanol, the neopterin triphosphate product was eluted from the charcoal by 3.1% NaOH in 50% ethanol. To increase the fluorescence yield, samples (1.0 ml) of the NaOH-containing extracts were neutralized by the addition of 0.1 ml of 1 M Tris-HCl (pH 7.4) and 0.14 ml of 4 N HCl. With excitation at 265 nm, the fluorescence emission at 450 nm was measured in a SimoAminco 500 spectrofluorometer and compared with a neopterin standard under the same conditions.

Mapping of mRNA endpoints. RNA was prepared as previously described (24) from *S. pneumoniae* 708 and 708(pLS830). The 5′ end of *sul* mRNA was determined by primer extension with 5′-GGCTTTACTTGACATAT-3′, which is complementary to nucleotides 258 to 242 in the *sul* DNA (20). One picomole of this

TABLE 2. GTPCH activity in extracts of *E. coli* expressing various *sul* genes[a]

Plasmid	*sul* genotype[b]	GTPCH activity (nmol/h/mg of protein)
pET-5	None	2
pLS23	*ABCD*	32
pLS24	*BCD*	40
pLS27	*AD*	3
pLS29	*ACD*	37
pLS30	*ABD*	1
pLS31	*ABC*	20

[a] Cultures of BL21(DE3) containing the indicated plasmids were induced, extracts were prepared, and samples containing approximately 50 μg of protein were assayed for GTPCH activity as indicated in Materials and Methods.

[b] Intact pneumococcal *sul* genes carried by the plasmids are indicated.

TABLE 3. Complementation of an *E. coli* mutant deficient in DHFS by various mutants of the *sul* gene cluster[a]

Plasmid	*sul* genotype[b]	Colonies (no.) on synthetic medium	
		Without Met	With Met
None	None	0	45
pLS23	*ABCD*	0	309
pLS24	*BCD*	86	141
pLS27	*AD*	0	159
pLS29	*ACD*	0	91
pLS30	*ABD*	85	184
pLS31	*ABC*	66	80

[a] Cultures of strain SF4 containing various plasmids were grown in L broth and diluted and plated in E medium ± methionine at 50 μg/ml.
[b] Intact pneumococcal *sul* genes carried by the plasmids are indicated.

oligonucleotide was annealed with 15 μg of total RNA in 50 mM Tris-HCl (pH 8.3), 40 mM KCl, 1 mM $MgCl_2$, 1 mM dithiothreitol, and 33 U of RNase inhibitor (Promega) in 14 μl for 5 min at 65°C. After addition of dATP, dGTP, and dTTP, each to 100 μM, and dCTP to 10 μM, 15 μCi of [α-^{32}P]dCTP and 20 U of avian myeloblastosis virus reverse transcriptase were added (volume now 20 μl) and the mixture was incubated at 42°C for 30 min. The products were examined on a sequencing gel and compared with a dideoxynucleotide-terminated DNA sequencing ladder obtained with the same oligonucleotide.

To map the 3′ end of the mRNA, pLS23 was first cut with *Ava*I and *Sac*I. The 604-bp fragment was isolated and cut with *Hin*PI; the two resulting fragments were labeled at their 3′ ends with [α-^{32}P]dCTP by the Klenow fragment. Further digestion of the *Hin*PI-*Ava*I fragments by *Taq*I gave a 199-bp *Hin*PI-*Taq*I fragment labeled at the 3′-*Hin*PI end and including the 3′ end of *sulD*. The labeled strand was used as a probe for nuclease S1 mapping as previously described (24) but with the following modifications. This probe (containing 18,000 cpm) was hybridized by heating for 10 min at 85°C and then for 3 h at 46°C with 150 μg of RNA from strain 708(pLS830) in 50 μl of 50% formamide containing 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.3), 0.4 M NaCl, and 1 mM EDTA. The hybridized sample was treated with 100 U of nuclease S1 for 20 min at 30°C and processed as described previously (24). The length of the protected mRNA was determined by comparison with a sequencing ladder from an unrelated DNA fragment.

Nucleotide sequence accession number. The DNA sequence data described in the paper have been deposited in GenBank with the accession number U16156.

RESULTS

DNA sequence of the *sulB* and *sulC* genes. The DNA sequence of a 2-kb segment of the cloned *sul* chromosomal region of *S. pneumoniae* is shown in Fig. 2. Beginning at bp 1101 of the previously reported *sulA* sequence (20), it links that gene to *sulD*, with bp 3100 corresponding to bp 933 of the previously reported *sulD* sequence (22) (add 2,167 to the numbering in reference 22 to match the present base pair designation). Analysis of the newly reported segment revealed two open reading frames in the same orientation between *sulA* and *sulD*. Called *sulB* and *sulC*, they are capable of encoding polypeptides of 49 kDa and 21 kDa, respectively (Fig. 2). High expression of the *sul* gene segment in *E. coli* previously showed

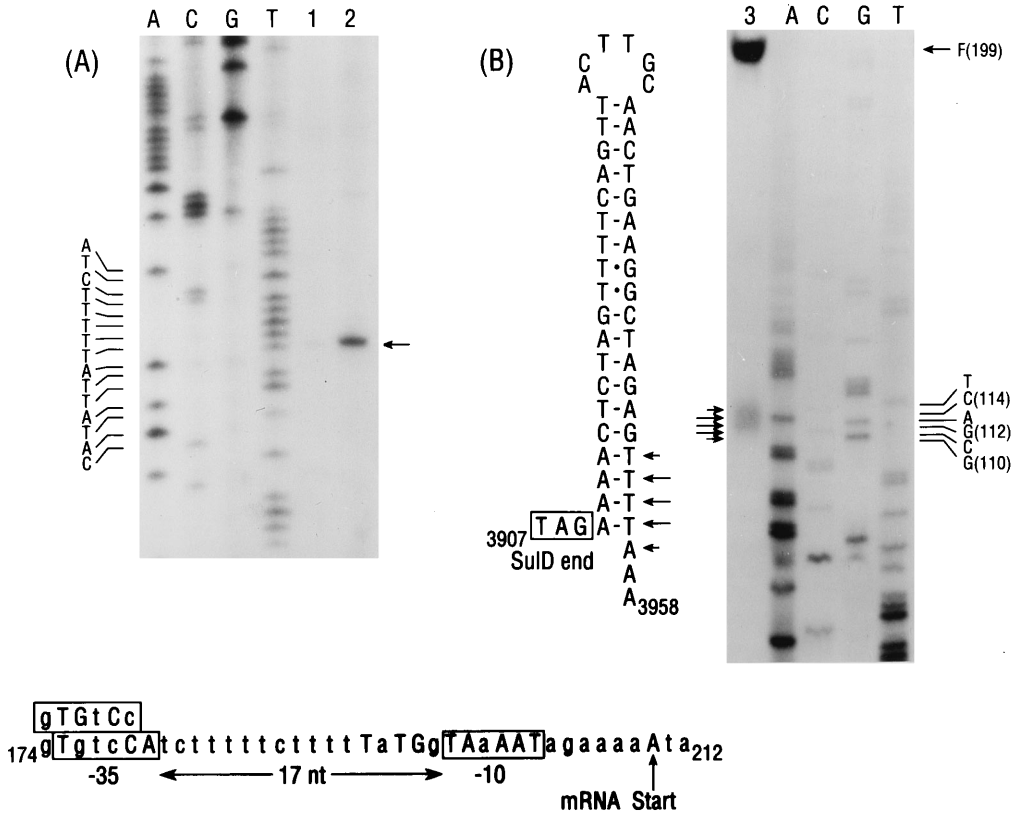

FIG. 3. Mapping of endpoints of the folate pathway mRNA. (A) Mapping of the 5′ endpoint by primer extension. Lanes A, C, G, and T show DNA sequencing ladders primed by the oligonucleotide used for primer extension on RNA from *S. pneumoniae* 708 (lane 1) and 708(pLS830) (lane 2). The arrow marks the primer extension product with the 5′-labeled oligonucleotide. The sequence at the bottom indicates the start site and the promoter, with the −10 site and two possible −35 sites boxed. Uppercase letters indicate agreement with the consensus sequence (and also the mRNA start). (B) Mapping of the 3′ endpoint by S1 nuclease digestion. Lane 3, RNA from *S. pneumoniae* 708(pLS830) hybridized to the 3′-labeled *Hin*PI-*Taq*I restriction fragment and treated with S1 nuclease. Lanes A, C, G, and T, sequencing ladders for unrelated DNA (lengths from the beginning of the primer are shown on the right). F, full-length restriction fragment. Unlabeled arrows indicate termination points, with half-arrows corresponding to minor bands. At the left is depicted a segment of DNA following the *sulD* gene showing the ability of the mRNA to fold into a hairpin.

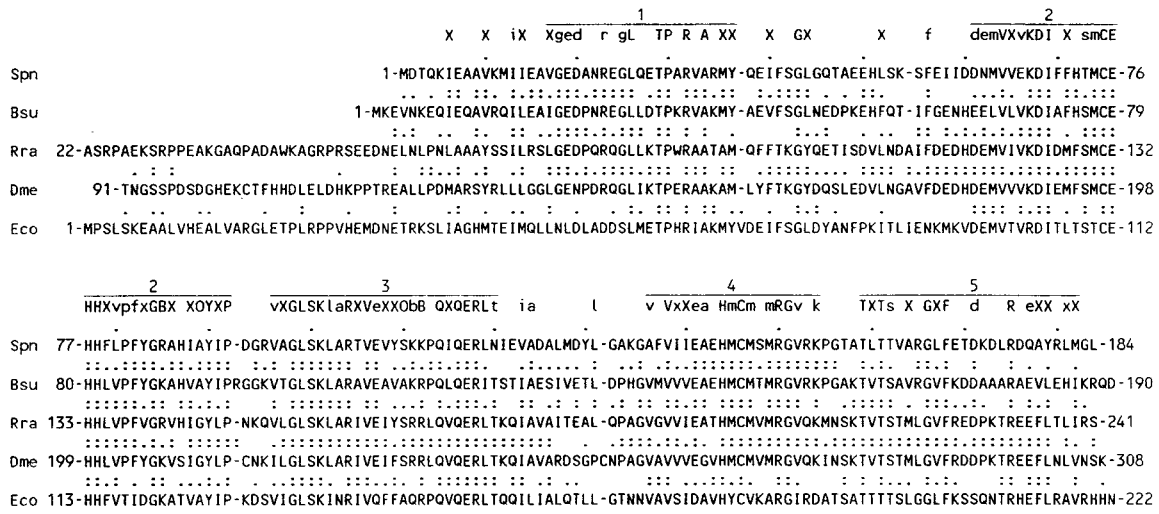

FIG. 4. Sequence comparison of SulC and homologous proteins. Numbers indicate positions in the protein sequence, with 1 corresponding to the N terminus. Dashes within the sequences indicate gaps giving optimal alignment. Spn, *Streptococcus pneumoniae*; Bsu, *Bacillus subtilis* (11); Rra, *Rattus rattus* (12); Dme, *Drosophila melanogaster* (28); Eco, *Escherichia coli* (14); :, identical residues; ., similar residues (based on the groups A, G, and P; D, E, N, and Q; F, I, L, M, V, W, and Y; H, K, and R; and S and T). Above the Spn sequence are indicated residues invariant among all five species (uppercase) or any four (lowercase). X and x represent hydrophobic residues A, F, I, L, M, V, W, and Y; O represents a residue (A, G, or P) with a small side chain. Boxes of conserved residues are indicated by numbered overlines.

the production of polypeptides of these sizes, which we shall call SulB and SulC, in addition to polypeptides SulA and SulD. The SulC polypeptide was purified, and its amino-terminal sequence was determined. The results show that SulC begins at the predicted site in that 11 of the 12 predicted amino acid residues were observed (Fig. 2). No product corresponding to the Glu residue was observed in the seventh step. Insufficient amounts of SulB were obtained for amino-terminal analysis. A Shine-Dalgarno sequence for ribosome binding is present before the start of *sulC* (Fig. 2); no such sequence is associated with *sulB*. At the position complementary to the Shine-Dalgarno sequence, the nucleotide sequence of *S. pneumoniae* 16S rRNA is identical to that of *E. coli* (2).

Identification of SulC function. When the *sulC* sequence was determined and the protein it encoded was compared with those in the GenBank database, a high degree of similarity was found with the GTPCH protein of the rat (12) and the protein produced by the *mtrA* gene of *B. subtilis* (11), with 43 and 53% identities of amino acid residues, respectively (see Fig. 4). The *mtrA* gene product was subsequently shown to have GTPCH activity (1). We have determined that the *sulC* product has GTPCH activity by expressing the protein from plasmids carrying the *sul* segment in *E. coli*. Only plasmids carrying an intact *sulC* gene expressed this enzyme activity, which was measured in extracts of *E. coli* containing various deleted derivatives of pLS23 (Table 2). These deletions removed either part of *sulA* (pLS24), parts of both *sulB* and *sulC* (pLS27), or 4-bp segments within each of the genes *sulB* (pLS29), *sulC* (pLS30), and *sulD* (pLS31). Only pLS27 and pLS30 failed to express the pneumococcal GTPCH.

Identification of SulB function. The *sulB* sequence encoded a protein that showed similarity (28% identity) to the previously determined *E. coli* DHFS coded for by the *folC* gene (3). We have not been successful in demonstrating DHFS activity in extracts of cells carrying this gene, but we were able to show that plasmids containing an intact *sulB* gene could complement a mutant *folC* gene in *E. coli*. The *E. coli* mutant strain SF4, as a result of a defect in DHFS, requires methionine for growth (4). We transferred plasmids containing the entire *sul* segment or various deleted derivatives into strain SF4 and tested colony-forming ability in the presence and absence of methionine (Table 3). Plasmids pLS24, pLS30, and pLS31, which are defective in genes other than *sulB*, complemented SF4 in the absence of methionine; plasmids pLS27 and pLS29, which are defective in *sulB*, did not. We cannot explain the behavior of pLS23, which contains an intact *sul* segment but did not complement the SF4 mutation. Otherwise, all the data support a DHFS function for the *sulB* product.

Mapping of the *sul* transcription product. From the DNA sequence, a consensus procaryotic promoter was observed upstream from *sulA* (20) and a palindrome corresponding to a potential terminator was evident downstream from *sulD* (22). The palindrome was positioned between *sulD* and a convergent open reading frame, so it could serve as a terminator for mRNAs transcribed in the two directions. The 5′ end of the *sul* mRNA was mapped by primer extension and shown to start at A-210 in the reported DNA sequence (20) for mRNA transcribed both from the chromosome and from plasmid pLS830 (Fig. 3A). This start site corresponds to a promoter with a −10 site close to consensus (34) and with all three bases corresponding to extended −10 sites (15), as shown in Fig. 3A. Two possible −35 sites, at a distance of 17 or 18 nucleotides from the −10 TAaAAT box, both show consensus bases in three of the six positions. The 3′ end of the *sul* mRNA was determined by S1 nuclease digestion of RNA from *S. pneumoniae* containing pLS830 hybridized with a 3′-end-labeled DNA segment extending from positions 3843 to 4041, as numbered in the present work. Several closely spaced bands were obtained, with the three heaviest indicating termination of the mRNA at the end of the hairpin structure (bp 3955) deduced from the DNA sequence (Fig. 3B). These results are consistent with transcription of the *sul* genes as a polycistronic mRNA of 3,745 ± 2 nucleotides in length.

Comparison of the SulC and SulB proteins with their homologs. The protein sequence of SulC is compared in Fig. 4 with those of its rat homolog and the MtrB protein of *B. subtilis*. All three proteins have been shown to function as GTPCH. In addition, two apparently homologous proteins

FIG. 5. Sequence comparison of SulB and homologous proteins. Numbers indicate positions in the protein sequence, with 1 corresponding to the N terminus. Dashes within the sequences indicate gaps giving optimal alignment. Spn, *Streptococcus pneumoniae*; Bsu, *Bacillus subtilis* (26, 29); Lca, *Lactobacillus casei* (41); Hsa, *Homo sapiens* (9); Eco, *Escherichia coli* (3); Sce, *Saccharomyces cerevisiae* (GenBank accession no. Z28131); :, identical residues; ., similar residues (based on the groups A, G, and P; D, E, N, and Q; F, I, L, M, V, W, and Y; H, K, and R; and S and T). Above the Spn sequence are indicated residues conserved among all six examples (uppercase) or any five (lowercase). X and x represent hydrophobic residues A, F, I, L, M, V, W, and Y; O and o represent residues (A, G, and P) with a small side chain. Boxes of conserved residues are indicated by numbered overlines.

identified by nucleic acid sequences in *E. coli* and *Drosophila melanogaster* are included in the comparison. Homologous proteins identified by nucleic acid sequencing of human and mouse nucleic acids are very similar to the rat protein and will not be considered separately. It is evident from Fig. 4 that a very high degree of sequence conservation exists for this protein. The Spn and Bsu proteins have 53% of their amino acid residues in common. Among the five homologs, 37 amino acid residues, or 20%, are invariant. An additional 20% have only hydrophobic side chains. These figures are considerably higher than those that would be found among exemplars of other proteins in these not closely related species, as witnessed for SulB (see below). The basis for such conservation may be the necessity for a protein of relatively small size to carry out multiple reactions. As shown in Fig. 1, the enzyme recognizes GTP, breaks a bond in a ring, removes a formate group, and

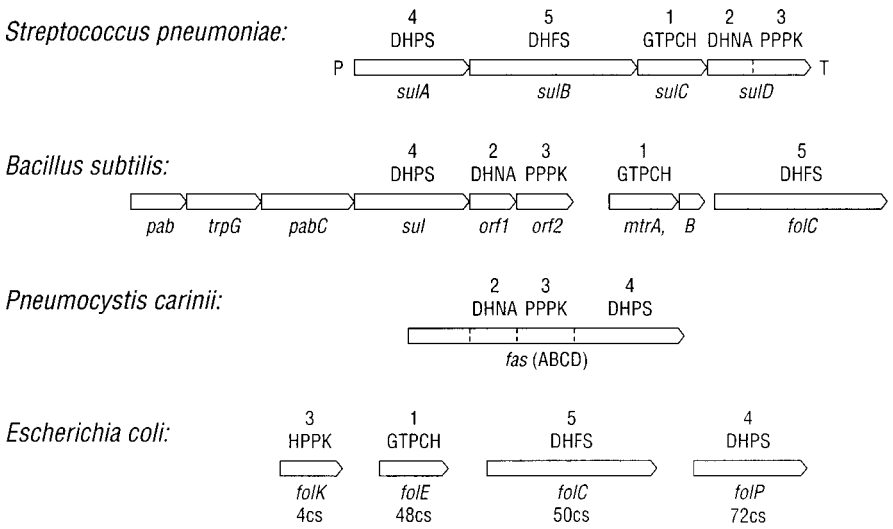

FIG. 6. Folate biosynthetic gene arrangements in different species. Genes or gene segments corresponding to enzyme domains are drawn to scale as boxlike arrows, with the orientation indicated by the arrowhead. Genes not shown adjacent are far apart in the genome. Gene designations and enzyme designations are given below and above the arrows, respectively. The numbers refer to the order of the steps in the folate pathway. The chromosomal locations of the *E. coli* genes were determined, and they are indicated in centisomes (cs). P and T refer to experimentally verified transcription promoters and terminators.

reforms a ring. Presumably, there is not much variability that can be tolerated in the structural basis for this performance. In keeping with multiple functions, there are five boxes of especially highly conserved regions, as indicated in Fig. 4. The first of these was hypothesized to correspond to a pterin-binding site (12), but possible functions of the other boxes are unknown.

In Fig. 5, the SulB sequence is compared with those of five homologous proteins, of which three (Eco, Hsa, and Lca) have been shown to exhibit DHFS activity and folylpoly(γ-glutamate) synthetase activity as well (9, 14, 41). The Bsu and Sce homologs were identified by nucleic acid sequencing. This group of proteins shows significant homology over their entire lengths (except for an insert of approximately 100 amino acid residues in Hsa). Six boxes of especially high conservation are evident (Fig. 5).

DISCUSSION

Because the 5′ and 3′ endpoints of *sul* mRNA correspond to the only promoter and transcription terminator evident in the nucleotide sequence, we presume that the operon is expressed as a single transcript. In matching the consensus promoter sequence, the *sul* promoter is weak at its −35 site (3 of 6 nucleotides match) and strong at its −10 site (5 of 6 nucleotides match), and it shows an extended −10 sequence (15). Good matches to the Shine-Dalgarno sequence (35) are found upstream of all the genes except *sulB*. Not surprisingly, the *sulB* gene product was found in the least amount when the operon was expressed in minicells of *B. subtilis* (unpublished results) or on an *E. coli* plasmid (22). Whether its expression is closer to that of the other genes under natural conditions in the chromosome of *S. pneumoniae* is not known. The proximity of the SulB start site to the termination of SulA might allow enhancement of SulB production by translational coupling (30). The homologous *E. coli* gene, *folC*, is also expressed relatively poorly in maxicells (3), and the corresponding enzyme is present at low levels in all cells studied (41).

Our findings indicate that all five enzymatic activities that have been demonstrated to function in folate biosynthesis are coded for by a single gene cluster in *S. pneumoniae*. Inasmuch as one of the protein products, SulD, carries two of these activities, it is conceivable that one of the other proteins may harbor another activity, such as the phosphatase needed to convert the product of GTPCH action to dihydroneopterin. The syncretion of two activities in SulD that are coded for as separate polypeptides in *B. subtilis* (37) and *E. coli* (40) is carried even further in the case of *P. carinii*, in which the polypeptide product of a single gene, *fas*, also contains a DHPS region and an additional segment of unknown function in addition to the two activities of SulD (44). It was shown that the *fas* product can express its multiple activities, like SulD, without prior cleavage (43).

The clustering of the genes for folate biosynthesis, presumably to help regulate their expression, is more marked for *S. pneumoniae* than for the other species in which these genes have been identified. Figure 6 depicts the arrangement of the genes in the different species. Although the three genes encoding DHPS, DHNA, and PPPK are found together in *B. subtilis*, the intervening genes in *S. pneumoniae* that encode DHFS and GTPCH are found elsewhere and separately. In addition, the three former genes are part of a larger cluster that includes two genes for *p*-aminobenzoate synthesis and the *trpG* gene, which has a role in aromatic synthesis (37). Interestingly, GTPCH is coded for by *mtrA*, which forms an operon with *mtrB*, the product of which acts as repressor of the *trp* operon and possibly as a translational repressor of *trpG* expression (1). In *P. carinii*, as mentioned, three activities are encoded together, but the genes for the other activities have not been located. In the case of *E. coli*, which encodes PPPK on a separate polypeptide, none of the four genes identified so far are linked. In particular, the DNA immediately upstream of the PPPK homolog does not encode DHNA, as in the other three species, but corresponds to the *pcnB* gene (19), which encodes a poly(A) polymerase (5).

The five genes or gene segments encoding the different activities show various degrees of conservation, as indicated in Table 4, which lists the percentages of identical amino acids between *S. pneumoniae* and the other species. For example, with *B. subtilis*, which showed the greatest similarity to pneumococcal proteins, the degrees of identity ranged from 28%

TABLE 4. Extent of homology between folate biosynthetic proteins of *S. pneumoniae* and those of other species

Species	Percent residues identical in *S. pneumoniae*				
	GTPCH	DHNA[a]	PPPK[b]	DHPS	DHFS
Bacillus subtilis	53	28	38	41	37
Lactobacillus casei					34
Pneumocystis carinii[c]		20	36	41	
Homo sapiens	42				31
Drosophila melanogaster	40				
Escherichia coli	38		32	35	28
Saccharomyces cerevisiae					26

[a] 119 N-terminal amino acid residues of SulD.
[b] 151 C-terminal amino acid residues of SulD.
[c] Corresponding segments of Fas protein (44).

for DHNA to 53% for GTPCH. Another gram-positive bacterium, *Lactobacillus casei*, gave a 34% identity with *S. pneumoniae*, compared with 37% for *B. subtilis*, for the single protein examined (DHFS). For each protein, *E. coli* usually showed the least identity, averaging nine percentage points below the identity values for *B. subtilis*. The identity values for eucaryotic proteins were in between, averaging seven percentage points below those for *B. subtilis*. Although these data are admittedly very limited, they suggest that the gram-positive bacteria diverged from the gram-negative bacteria earlier than they did from the progenitor of eucaryotes. At the time when the bacterial and eucaryotic progenitors separated, presumably more than a billion years ago, a prevalence of introns in their genomes (as hypothesized in reference 10) may have fostered a fluidity in gene rearrangement that accounts for the variety of folate biosynthetic gene organization that we see today.

ACKNOWLEDGMENTS

We thank A. L. Bognar for providing us with *E. coli* SF4. Research at Brookhaven National Laboratory was carried out under the auspices of the Office of Health and Environmental Research, U.S. Department of Energy; research at the Centro de Investigaciones Biológicas was carried out under the auspices of the Consejo Superior de Investigaciones Científicas.

This work was supported in part by Public Health Service grants GM29721 and AI14885 to S.A.L., by NATO grant 0119-88 to P.L. and S.A.L., and by grant Bio94-1029 from the Comision Interministerial de Ciencia y Tecnología.

REFERENCES

1. **Babitzke, P., P. Gollnick, and C. Yanofsky.** 1992. The *mtrAB* operon of *Bacillus subtilis* encodes GTP cyclohydrolase I (MtrA), an enzyme involved in folic acid biosynthesis, and MtrB, a regulator of tryptophan biosynthesis. J. Bacteriol. **174:**2059–2064.
2. **Bacot, C. M., and R. H. Reeves.** 1991. Novel tRNA gene organization in the 16S-23S intergenic spacer of the *Streptococcus pneumoniae* rRNA gene cluster. J. Bacteriol. **173:**4234–4236.
3. **Bognar, A. L., C. Osborne, and B. Shane.** 1987. Primary structure of the *Escherichia coli folC* gene and its folylpolyglutamate synthetase-dihydrofolate synthetase product and regulation of expression by an upstream gene. J. Biol. Chem. **262:**12337–12342.
4. **Bognar, A. L., C. Osborne, B. Shane, S. C. Singer, and R. Ferone.** 1985. Folylpoly-γ-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the *Escherichia coli folC* gene and purification and properties of the gene product. J. Biol. Chem. **260:**5625–5630.
5. **Cao, G. J., and N. Sarkar.** 1992. Identification of the gene for an *Escherichia coli* poly(A) polymerase. Proc. Natl. Acad. Sci. USA **89:**10380–10384.
6. **Currier, T. C., and E. W. Nester.** 1976. Isolation of covalently closed circular DNA of high molecular weight from bacteria. Anal. Biochem. **76:**431–441.
7. **Dallas, W. S., I. K. Dev, and P. H. Ray.** 1993. The dihydropteroate synthase gene, *folP*, is near the leucine tRNA gene, *leuU*, on the *Escherichia coli* chromosome. J. Bacteriol. **175:**7743–7744.
8. **Dallas, W. S., J. E. Gowen, P. H. Ray, M. J. Cox, and I. K. Dev.** 1992. Cloning, sequencing, and enhanced expression of the dihydropteroate synthase gene of *Escherichia coli* MC4100. J. Bacteriol. **174:**5961–5970.
9. **Garrow, T. A., A. Admon, and B. Shane.** 1992. Expression cloning of a human cDNA encoding folylpoly(γ-glutamate) synthetase and determination of its primary structure. Proc. Natl. Acad. Sci. USA **89:**9151–9155.
10. **Gilbert, W.** 1985. Genes-in-pieces revisited. Science **228:**823–824.
11. **Gollnick, P., S. Ishino, M. I. Kuroda, D. J. Henner, and C. Yanofsky.** 1990. The *mtr* locus is a two-gene operon required for transcription attenuation in the *trp* operon of *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **87:**8726–8730.
12. **Hatakeyama, K., Y. Inoue, T. Harada, and H. Kagamiyama.** 1991. Cloning and sequencing of cDNA encoding rat GTP cyclohydrolase I. J. Biol. Chem. **266:**765–769.
13. **Higgins, R. C., and M. E. Dahmus.** 1979. Rapid visualization of protein bands in preparative SDS-polyacrylamide gels. Anal. Biochem. **93:**257–260.
14. **Katzenmeier, G., C. Schmid, J. Kellermann, F. Lottspeich, and A. Bacher.** 1991. Biosynthesis of tetrahydrofolate. Sequence of GTP cyclohydrolase I from *Escherichia coli*. Biol. Chem. Hoppe-Seyler **372:**991–997.
15. **Keilty, S., and M. Rosenberg.** 1987. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. J. Biol. Chem. **262:**6389–6395.
16. **Kushner, S. R.** 1978. An improved method for transformation of *Escherichia coli* with ColE1 derived plasmids, p. 17–23. *In* H. W. Boyer and S. Nicosia (ed.), Genetic engineering. Elsevier/North-Holland, Amsterdam.
17. **Lacks, S.** 1966. Integration efficiency and genetic recombination in pneumococcal transformation. Genetics **53:**207–235.
18. **Lacks, S. A., P. Lopez, B. Greenberg, and M. Espinosa.** 1986. Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. J. Mol. Biol. **192:**753–765.
19. **Liu, J., and J. S. Parkinson.** 1989. Genetics and sequence analysis of the *pcnB* locus, an *Escherichia coli* gene involved in plasmid copy number control. J. Bacteriol. **171:**1254–1261.
20. **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.
21. **Lopez, P., M. Espinosa, and S. A. Lacks.** 1984. Physical structure and genetic expression of the sulfonamide-resistance plasmid pLS80 and its derivatives in *Streptococcus pneumoniae* and *Bacillus subtilis*. Mol. Gen. Genet. **195:**402–410.
22. **Lopez, P., B. Greenberg, and S. A. Lacks.** 1990. DNA sequence of folate biosynthesis gene *sulD*, encoding hydroxymethyldihydropterin pyrophosphokinase in *Streptococcus pneumoniae*, and characterization of the enzyme. J. Bacteriol. **172:**4766–4774.
23. **Lopez, P., and S. A. Lacks.** 1993. A bifunctional protein in the folate biosynthetic pathway of *Streptococcus pneumoniae* with dihydroneopterin aldolase and hydroxymethyldihydropterin pyrophosphokinase activities. J. Bacteriol. **175:**2214–2220.
24. **Lopez, P., S. Martinez, A. Diaz, M. Espinosa, and S. A. Lacks.** 1989. Characterization of the *polA* gene of *Streptococcus pneumoniae* and comparison of the DNA polymerase I it encodes to homologous enzymes from *Escherichia coli* and phage T7. J. Biol. Chem. **264:**4255–4263.
25. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurements with the Folin phenol reagent. J. Biol. Chem. **193:**265–275.
26. **Margolis, P. S., A. Driks, and R. Losick.** 1993. Sporulation gene *spoIIB* from *Bacillus subtilis*. J. Bacteriol. **175:**528–540.
27. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
28. **McLean, J. R., S. Krishnakumar, and J. M. O'Donnell.** 1993. Multiple mRNAs from the Punch locus of *Drosophila melanogaster* encode isoforms of GTP cyclohydrolase I with distinct N-terminal domains. J. Biol. Chem. **268:**27191–27197.
29. **Mohan, S., J. Aghion, N. Guillen, and D. A. Dubnau.** 1989. Molecular cloning and characterization of *comC*, a late competence gene of *Bacillus subtilis*. J. Bacteriol. **171:**6043–6051.
30. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling in expression of the tryptophan operon of *E. coli*. Genetics **95:**785–795.
31. **Ortiz, P. J.** 1970. Dihydrofolate and dihydropteroate synthesis by partially purified enzymes from wild-type and sulfonamide-resistant pneumococcus. Biochemistry **9:**355–361.
32. **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.
33. **Rosenberg, A. H., B. N. Lade, D. Chui, S. Lin, J. J. Dunn, and W. F. Studier.** 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene **56:**125–135.
34. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13:**319–353.
35. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254:**34–38.
36. **Shiota, T., C. M. Baugh, R. Jackson, and R. Dillard.** 1969. The enzymatic synthesis of hydroxymethyldihydropteridine pyrophosphate and dihydrofolate. Biochemistry **8:**5022–5028.

37. **Slock, J., D. P. Stahly, C.-Y. Han, E. W. Six, and I. P. Crawford.** 1990. 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. J. Bacteriol. **172:**7211–7226.
38. **Stassi, D. L., P. Lopez, M. Espinosa, and S. A. Lacks.** 1981. Cloning of chromosomal genes in *Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA **78:**7028–7032.
39. **Studier, F. W., and B. A. Moffatt.** 1986. Use of T7 RNA polymerase to direct selective, high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
40. **Talarico, T. L., P. H. Ray, I. K. Dev, B. M. Merrill, and W. S. Dallas.** 1992. Cloning, sequence analysis, and overexpression of *Escherichia coli folK*, the gene coding for 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase. J. Bacteriol. **174:**5971–5977.
41. **Toy, J., and A. L. Bognar.** 1990. Cloning and expression of the gene encoding *Lactobacillus casei* folylpoly-γ-glutamate synthetase in *Escherichia coli* and determination of its primary structure. J. Biol. Chem. **265:**2492–2499.
42. **Vogel, H. J., and D. M. Bonner.** 1956. Acetylornithinase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem. **218:**97–106.
43. **Volpe, F., S. P. Ballantine, and C. J. Delves.** 1993. The multifunctional folic acid synthesis *fas* gene of *Pneumocystis carinii* encodes dihydroneopterin aldolase, hydroxymethyldihydropterin pyrophosphokinase and dihydropteroate synthase. Eur. J. Biochem. **216:**449–458.
44. **Volpe, F., M. Dyer, J. G. Scaife, G. Darby, D. K. Stammers, and C. J. Delves.** 1992. The multifunctional folic acid synthesis *fas* gene of *Pneumocystis carinii* appears to encode dihydropteroate synthase and hydroxymethyldihydropterin pyrophosphokinase. Gene **112:**213–218.
45. **Winston, D. J., W. K. Lau, R. P. Gale, and L. S. Young.** 1980. Trimethoprim-sulfamethoxazole for the treatment of *Pneumocystis carinii* pneumonia. Ann. Intern. Med. **92:**762–769.