Journal of Bacteriology, July 1984, p. 57–62
0021-9193/84/070057-06$02.00/0

Vol. 159, No. 1

Nucleotide Sequence of *Escherichia coli pabB* Indicates a Common Evolutionary Origin of *p*-Aminobenzoate Synthetase and Anthranilate Synthetase

PAUL GONCHAROFF and BRIAN P. NICHOLS*

Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60680

Received 16 January 1984/Accepted 13 April 1984

Biochemical and immunological experiments have suggested that the *Escherichia coli* enzyme *p*-aminobenzoate synthetase and anthranilate synthetase are structurally related. Both enzymes are composed of two nonidentical subunits. Anthranilate synthetase is composed of proteins encoded by the genes *trp(G)D* and *trpE*, whereas *p*-aminobenzoate synthetase is composed of proteins encoded by *pabA* and *pabB*. These two enzymes catalyze similar reactions and produce similar products. The nucleotide sequences of *pabA* and *trp(G)D* have been determined and indicate a common evolutionary origin of these two genes. Here we present the nucleotide sequence of *pabB* and compare it with that of *trpE*. Similarities are 26% at the amino acid level and 40% at the nucleotide level. We propose that *pabB* and *trpE* arose from a common ancestor and hence that there is a common ancestry of genes encoding *p*-aminobenzoate synthetase and anthranilate synthetase.

Chorismate is a common precursor in the biosynthesis of *p*-aminobenzoate (PABA) and anthranilate. PABA is a component of the vitamin folic acid, whereas anthranilate is an intermediate in the tryptophan biosynthetic pathway. These two specific reactions involving chorismate are similar and yield similar products. Both of these reactions utilize chorismate and glutamine to produce pyruvate, glutamate, and either PABA or anthranilate. The structure of PABA and anthranilate differ only by the position of the amino group on the benzene ring. The enzymes that catalyze these two reactions, *p*-aminobenzoate synthetase (PABS) and anthranilate synthetase (AS), both consist of nonidentical subunits, component I (CoI) and component II (CoII). In PABA synthesis, PABS CoI cannot function with AS CoII. Similarly, in anthranilate synthesis, AS CoI cannot function with PABS CoII. This is inferred by the inability of wild-type *trp* genes to complement *pabA* and *pabB* mutants and the inability of *pab* genes to complement *trpE* and *trp(G)D* mutants (27). Nevertheless, it has been suggested that *Escherichia coli* PABS and AS are structurally related since antibodies raised against AS cross-react with fractionated extracts containing PABS (19).

In each synthetase complex, CoI alone can bind chorismate and ammonia to form the aromatic product (8; S. Doktor and B. P. Nichols, unpublished results). CoII contains a glutamine amidotransferase activity whose function is to transfer the amide group from glutamine to CoI. *E. coli* PABS CoI and CoII are encoded by two unlinked genes, *pabB* and *pabA*, respectively (5, 6). AS CoI and CoII of *E. coli* are encoded by two linked genes, *trpE* and *trpD*, respectively (26). *trpD* encodes a bifunctional protein that contains the glutamine amidotransferase and the anthranilate phosphoribosyl transferase of the tryptophan pathway (26). In *Serratia marcescens*, these two enzymes are encoded by separate genes, *trpG* and *trpD*, respectively (15). To avoid confusion, we will use *trp(G)* to refer to that portion of the *trpD* gene that encodes the glutamine amidotransferase activity of *E. coli* AS CoII (2).

Recently, common evolutionary origins of several gene pairs have been reported (3, 11, 25). Gene duplications increase the genetic potential of an organism (9) since after a gene duplication event, one of the genes is free to respond to selective pressures which may result in an altered function. It has been hypothesized that *E. coli pabA* and *trp(G)* arose from a common ancestor via a gene duplication event, and nucleotide sequence comparisons support this view (10). This work investigates the evolutionary relationship of the CoI subunits of PABS and AS encoded by *E. coli pabB* and *trpE*.

MATERIALS AND METHODS

Bacterial strains. *E. coli* AB3303 (*pabB3 thi-1 his-4 argE3 lacY1 galK2 xyl-5 mtl-1 rpsL-704 tsx-29* or *tsx-358 supE44*) (6) was provided by B. Bachmann. *E. coli* JM103 (13) and phages M13mp8 and M13mp9 (14) were obtained from New England Biolabs.

Enzymes and DNA manipulations. Restriction endonucleases were purchased from New England Biolabs, Boehringer Mannheim, or P-L Biochemicals, Inc., or were prepared in this laboratory by published procedures. *E. coli* DNA polymerase I (Klenow fragment) was purchased from Boehringer Mannheim. Plasmid DNA was prepared by the Birnboim and Doly method (1). Transformations of bacterial cells with plasmid DNA were performed as described by Mandel and Higa (12).

DNA sequence determination and analysis. DNA sequence analysis was performed by the method of Sanger et al. (22). M13 manipulations were performed as described by Messing et al. (13). The polyacrylamide-urea gel electrophoresis system described by Sanger and Coulson (21) was used. Sequence analysis was performed in part by a computer with the program of J. B. Kaplan (unpublished).

RESULTS

Plasmid construction and sequence determination of *E. coli pabB*. Two independent cloning experiments yielded plasmids carrying *E. coli pabB*. *E. coli* genomic DNA was partially digested with either *Eco*RI or *Hin*dIII. Fragments resulting from these two digests were ligated into either *Eco*RI-digested or *Hin*dIII-digested plasmid pBR322. The products of these constructions were used to transform *E. coli* AB3303 (*pabB3*) to PABA independence and ampicillin

* Corresponding author.

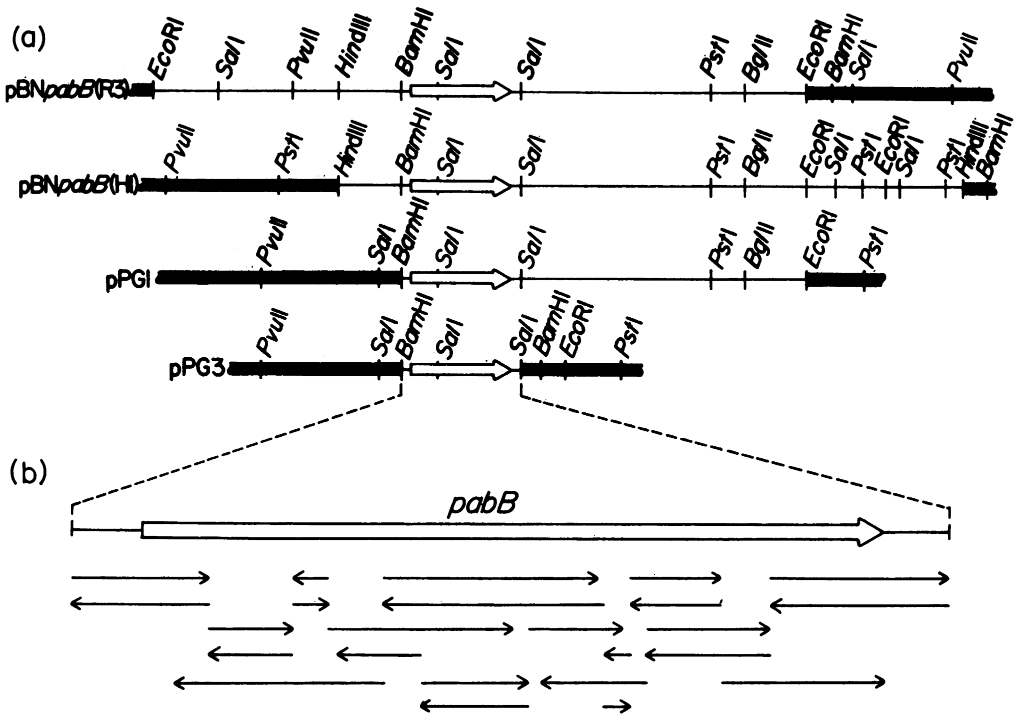

FIG. 1. (a) Linear restriction maps of *pabB*-containing plasmids. The heavy dark lines represent pBR322 vector DNA, and the light lines represent inserted DNA. The open arrow within the inserted DNA corresponds to the coding region of the *pabB* and indicates direction of transcription. (b) DNA sequencing strategy. Arrows represent DNA sequence determined from specific *Sau*3A, *Taq*I, or *Hae*III fragments cloned into M13 bacteriophage vectors.

resistance. Plasmids were isolated from the resulting colonies. Two of these plasmids were named pBN*pabB*(R3) and pBN*pabB*(H1), corresponding to the *Eco*RI and *Hin*dIII constructions, respectively. Besides the 4.3 kilobases (kb) of pBR322 vector DNA, pBN*pabB*(R3) contained 9 kb of *E. coli* DNA; pBN*pabB*(H1) contained 8.7 kb. The restriction maps of these two plasmids (Fig. 1a) were found to be identical within the 6.5-kb *Hin*dIII-*Eco*RI fragment. This indicated that the *pabB* gene was contained within this fragment. Further subclones were constructed from pBN*pabB*(R3).

The subcloning strategy of pBN*pabB*(R3) is shown in Fig. 1a. Plasmids pBN*pabB*(R3) and pBR322 were digested with *Bam*HI and *Eco*RI. This mixture of fragments was ligated and used to transform *E. coli* AB3303 to PABA independence and ampicillin resistance. One of the plasmids isolated in this construction, pPG1, was a pBR322 derivative that contained a 5.6-kb *Bam*HI-*Eco*RI fragment from pBN*pabB*(R3). A *Sal*I reduction of pPG1 did not yield PABA-independent colonies, suggesting that at least one of the *Sal*I sites lay within *pabB*. A 1.9-kb fragment was isolated from pPG1 after a partial *Sal*I digest. The fragment contained the 275-base pair (bp) *Sal*I-*Bam*HI portion of pBR322 and a 1,622-bp portion of *E. coli* DNA containing *pabB*. pPG3 was constructed by ligation of the 1.9-kb *Sal*I fragment into the *Sal*I site of pBR322 and transformation of *E. coli* AB3303 to ampicillin resistance and PABA independence. pPG3 contains the 1,622-bp *Bam*HI-*Sal*I fragment flanked by 275-bp direct repeats.

The 1,622-bp *Bam*HI-*Sal*I fragment of plasmid pPG3 was the source of smaller DNA fragments used to determine the nucleotide sequence of *pabB*. *Sau*3A, *Taq*I, and *Hae*III restriction fragments were isolated from 5% polyacrylamide gels and were ligated into M13mp8 or M13mp9 bacteriophage vectors restricted with *Bam*HI, *Acc*I, and *Hin*cII, respectively. Single-stranded DNA was prepared from *E. coli* JM103 that had been transfected with the recombinant phages. The nucleotide sequence of the 1,622-bp *Bam*HI-*Sal*I fragment was determined completely on both strands of the DNA (Fig. 2).

Translational reading frame of *pabB*. Examination of all possible translational frames of the 1,622-bp *Bam*HI-*Sal*I fragment identified one 1,359-bp continuous reading frame. No other reading frame exceeded a total length of 300 bp. This 1,359-bp open reading frame has the potential for encoding a protein containing 453 amino acid residues with a calculated molecular weight of 50,958. This figure is in close agreement with two independently determined molecular weights for *E. coli* PABS CoI. Gel permeation chromatography yielded a molecular weight of 46,000 (5), and sodium dodecyl sulfate-polyacrylamide gel electrophoresis yielded a molecular weight of 53,000 (Seibold and Nichols, unpublished results). These data and the fact that plasmid pPG3 can complement the *pabB* mutation of strain AB3303 provide strong evidence that the nucleotide and amino acid sequences shown in Fig. 2 are those of *E. coli pabB*.

Amino acid and nucleotide similarities of *pabB* and *trpE*. The amino acid sequences of PABS CoI and AS CoI are aligned and presented in Fig. 3. Six gaps have been inserted between the two sequences at positions 124 to 125, 130 to 131, 160 to 161, 228, 271 to 272, and 281 to 282 to align areas of similarity. Other alignments are possible if more gaps are inserted, but care was taken to use a minimal number of gaps in aligning the two sequences. Some gaps have been inserted arbitrarily within specific regions of the two sequences. For example, the gap present between positions 160 and 161 could have been placed between positions 148 and 149 without affecting the overall amount of similarity. No gaps

```
                                                                                                         -130      -120
                                                                                                        GGATCCGCTCGACAG

      -110      -100       -90       -80       -70       -60       -50       -40       -30       -20       -10
CTTTTTGCTGCTGTTCATGGGTGAGTGAAGGTAAACCTGCAAACATTGTTAACTCCTGCTAAATTGTTGGCGCTAATTATTTCATGCTACCCGGCACATAGCCAGTAGAGTCAGGACTG

           10           20           30            40           50           60            70           80           90
Met Lys Thr Leu Ser Pro Ala Val Ile Thr Leu Leu Trp Arg Gln Asp Ala Ala Glu Phe Tyr Phe Ser Arg Leu Ser His Leu Pro Trp
ATG AAG ACG TTA TCT CCC GCT GTG ATT ACT TTA CTC TGG CGT CAG GAC GCC GCT GAA TTT TAT TTC TCC CGC TTA AGC CAC CTG CCG TGG

          100          110          120           130          140          150           160          170          180
Ala Met Leu Leu His Ser Gly Tyr Ala Asp His Pro Tyr Ser Arg Phe Asp Ile Val Val Ala Glu Pro Ile Cys Thr Leu Thr Thr Phe
GCG ATG CTT TTA CAC TCC GGC TAT GCC GAT CAT CCG TAT AGC CGC TTT GAT ATT GTG GTC GCC GAG CCG ATT TGC ACT TTA ACC ACT TTC

          190          200          210           220          230          240           250          260          270
Gly Lys Glu Thr Val Val Ser Glu Ser Glu Lys Arg Thr Thr Thr Thr Asp Asp Pro Leu Gln Val Leu Gln Gln Val Leu Asp Arg Ala
GGT AAA GAA ACC GTT GTT AGT GAA AGC GAA AAA CGC ACA ACG ACC ACT GAT GAC CCG CTA CAG GTG CTC CAG CAG GTG CTG GAT CGC GCA

          280          290          300           310          320          330           340          350          360
Asp Ile Arg Pro Thr His Asn Glu Asp Leu Pro Phe Gln Gly Gly Ala Leu Gly Leu Phe Gly Tyr Asp Leu Gly Arg Arg Phe Glu Ser
GAC ATT CGC CCA ACG CAT AAC GAA GAT TTG CCA TTT CAG GGC GGC GCA CTG GGG TTG TTT GGC TAC GAT CTG GGC CGC CGT TTT GAG TCA

          370          380          390           400          410          420           430          440          450
Leu Pro Glu Ile Ala Glu Gln Asp Ile Val Leu Pro Asp Met Ala Val Gly Ile Tyr Asp Trp Ala Leu Ile Val Asp His Gln Arg His
CTG CCA GAA ATT GCG GAA CAA GAT ATC GTT CTG CCG GAT ATG GCA GTG GGT ATC TAC GAT TGG GCG CTC ATT GTC GAC CAC CAG CGT CAT

          460          470          480           490          500          510           520          530          540
Thr Val Ser Leu Leu Ser His Asn Asp Val Asn Ala Arg Arg Ala Trp Leu Glu Ser Gln Gln Phe Ser Pro Gln Glu Asp Phe Thr Leu
ACA GTT TCT TTG CTG AGT CAT AAT GAT GTC AAT GCC CGT CGG GCC TGG CTG GAA AGC CAG CAA TTC TCG CCG CAG GAA GAT TTC ACG CTC

          550          560          570           580          590          600           610          620          630
Thr Ser Asp Trp Gln Ser Asn Met Thr Arg Glu Gln Tyr Gly Glu Lys Phe Arg Gln Val Gln Glu Tyr Leu His Ser Gly Asp Cys Tyr
ACT TCC GAC TGG CAA TCC AAT ATG ACC CGC GAG CAG TAC GGC GAA AAA TTT CGC CAG GTA CAG GAA TAT CTG CAC AGC GGT GAT TGC TAT

          640          650          660           670          680          690           700          710          720
Gln Val Asn Leu Ala Gln Arg Phe His Ala Thr Tyr Ser Gly Asp Glu Trp Gln Ala Phe Leu Gln Leu Asn Gln Ala Asn Arg Ala Pro
CAG GTG AAT CTC GCC CAA CGT TTT CAT GCG ACC TAT TCT GGC GAT GAA TGG CAG GCA TTC CTT CAG CTT AAT CAG GCC AAC CGC GCG CCA

          730          740          750           760          770          780           790          800          810
Phe Ser Ala Phe Leu Arg Leu Glu Gln Gly Ala Ile Leu Ser Leu Ser Pro Glu Arg Phe Ile Leu Cys Asp Asn Ser Glu Ile Gln Thr
TTT AGC GCT TTT TTA CGT CTT GAA CAG GGT GCA ATT TTA AGC CTT TCG CCA GAG CGG TTT ATT CTT TGT GAT AAT AGT GAA ATC CAG ACC

          820          830          840           850          860          870           880          890          900
Arg Pro Ile Lys Gly Thr Leu Pro Arg Leu Pro Asp Pro Gln Glu Asp Ser Lys Gln Ala Val Lys Leu Ala Asn Ser Ala Lys Asp Arg
CGC CCG ATT AAA GGC ACG CTA CCA CGC CTG CCC GAT CCT CAG GAA GAT AGC AAA CAA GCA GTA AAA CTG GCG AAC TCA GCG AAA GAT CGT

          910          920          930           940          950          960           970          980          990
Ala Glu Asn Leu Met Ile Val Asp Leu Met Arg Asn Asp Ile Gly Arg Val Ala Val Ala Gly Ser Val Lys Val Pro Glu Leu Phe Val
GCC GAA AAT CTG ATG ATT GTC GAT TTA ATG CGT AAT GAT ATC GGT CGT GTT GCC GTA GCA GGT TCG GTA AAA GTA CCA GAG CTG TTC GTG

         1000         1010         1020          1030         1040         1050          1060         1070         1080
Val Glu Pro Phe Pro Ala Val His His Leu Val Ser Thr Ile Thr Ala Gln Leu Pro Glu Gln Leu His Ala Ser Asp Leu Leu Arg Ala
GTG GAA CCC TTC CCT GCC GTG CAT CAT CTG GTC AGC ACC ATA ACG GCG CAA CTA CCA GAA CAG TTA CAC GCC AGC GAT CTG CTG CGC GCA

         1090         1100         1110          1120         1130         1140          1150         1160         1170
Ala Phe Pro Gly Gly Ser Ile Thr Gly Ala Pro Lys Val Arg Ala Met Glu Ile Ile Asp Glu Leu Glu Pro Gln Arg Arg Asn Ala Trp
GCT TTT CCT GGT GGC TCA ATA ACC GGG GCT CCG AAA GTA CGG GCT ATG GAA ATT ATC GAC GAA CTG GAA CCG CAG CGA CGC AAT GCC TGG

         1180         1190         1200          1210         1220         1230          1240         1250         1260
Cys Gly Ser Ile Gly Tyr Leu Ser Phe Cys Gly Asn Met Asp Thr Ser Ile Thr Ile Arg Thr Leu Thr Ala Ile Asn Gly Gln Ile Phe
TGC GGC AGC ATT GGC TAT TTG AGC TTT TGC GGC AAC ATG GAT ACC AGT ATT ACT ATC CGC ACG CTG ACT GCC ATT AAC GGA CAA ATT TTC

         1270         1280         1290          1300         1310         1320          1330         1340         1350
Cys Ser Ala Gly Gly Gly Ile Val Ala Asp Ser Gln Glu Glu Ala Glu Tyr Gln Glu Thr Phe Asp Lys Val Asn Arg Ile Leu Lys Gln
TGC TCT GCG GGC GGT GGA ATT GTC GCC GAT AGC CAG GAA GAA GCG GAA TAT CAG GAA ACT TTT GAT AAA GTT AAT CGT ATC CTG AAG CAA

         1360       1370      1380      1390      1400      1410      1420      1430      1440      1450      1460
Leu Glu Lys End
CTG GAG AAG TAA GACGTGGAATACCGTAGCCTGACGCTTGATGATTTTTTATCGCGCTTTCAACTTTTGCGCCCGCAAATTAACCGGGAAACCCTAAATCATCGTCAGGCTGCTG

 1470      1480
TGTTAATCCCCATCGTCCGTCGAC
```

FIG. 2. Nucleotide and amino acid sequence of *E. coli pabB*. The complete nucleotide sequence of the 1,622-bp *Bam*HI-*Sal*I fragment is shown along with the predicted amino acid sequence. Underlined portions of the sequence correspond to possible regulatory regions discussed in the text.

had to be inserted from position 282 to the end of the comparison. Out of 450 amino acids compared, 26% similarity exists between PABS CoI and AS CoI. Using this same alignment, a similarity of 40% exists at the nucleotide level. The extent of similarity at the nucleotide level throughout the alignment of *pabB* and *trpE* is shown in Fig. 4.

Codon usage in *pabB*. Codon usage is believed to be primarily dependent on the relative amounts of isoaccepting tRNA molecules present in the cell (7). The codon usage of *pabB* is similar to that of *trpE* (Table 1) and is consistent with the codon usage of *E. coli* genes which are not highly expressed (4). These data, including the amino acid alignment of *pabB* with *trpE*, further support the proposed translational frame of *E. coli pabB* (Fig. 2).

PABS CoI contains seven tryptophan residues, whereas AS CoI contains none. Furthermore, it has been reported that PABS CoII contains three tryptophan residues, whereas its counterpart, AS CoII, contains none (10, 16). It has been proposed that AS CoI and CoII, which together make up AS, are tryptophan free so that chorismate can be channeled into the tryptophan biosynthetic pathway under severe tryptophan starvation conditions (17). Thus, the presence of tryp-

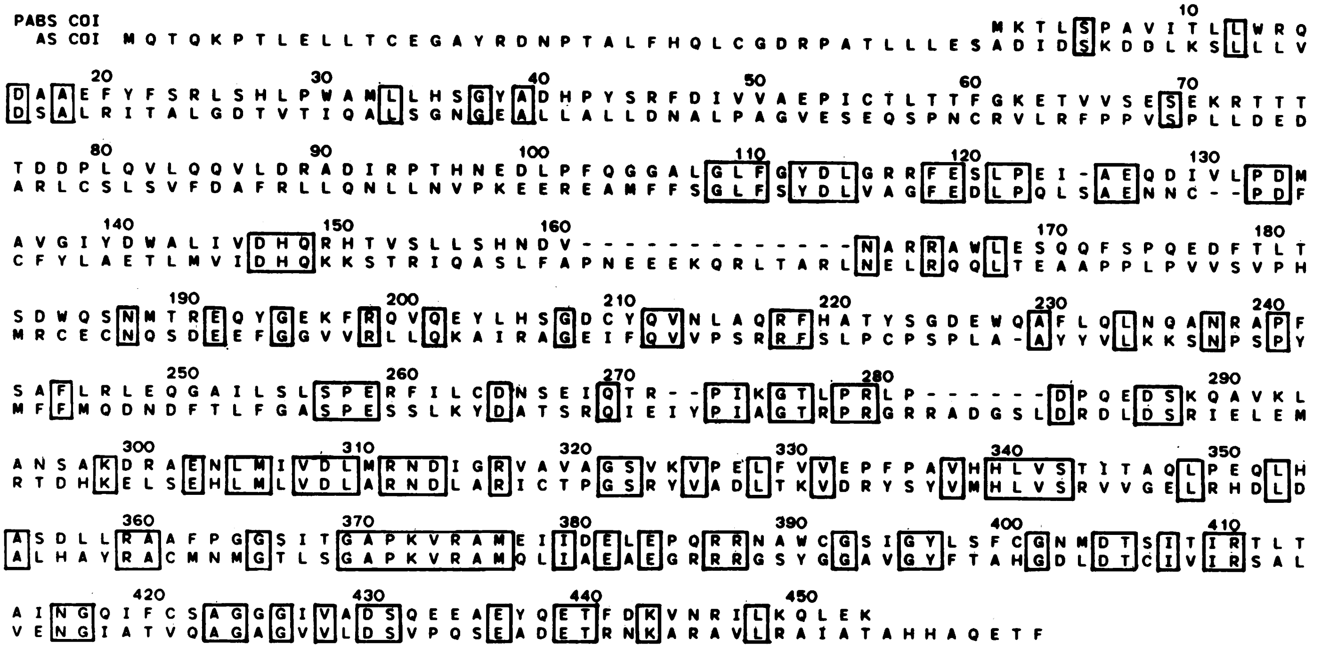
PABS COI
AS COI
10
MKTLSPAVITLLWRQ
MQTQKPTLELLTCEGAYRDNPTALFHQLCGDRPATLLLESADIDSKDDLKSLLLV
20 30 40 50 60 70
DAAEFYFSRLSHLPWAMLLHSGYADHPYSRFDIVVAEPICTLTTFGKETVVSESEKRTTT
DSALRITALGDTVTIQALSGNGEALLALLDNALPAGVESEQSPNCRVLRFPPVSPLLDED
80 90 100 110 120 130
TDDPLQVLQQVLDRADIRPTHNEDLPFQGGALGLFGYDLGRRFESLPEI-AEQDIVLPDM
ARLCSLSVFDAFRLLQNLLNVPKEEREAMFFSGLFSYDLVAGFEDLPQLSAENNC--PDF
140 150 160 170 180
AVGIYDWALIVDHQRHTVSLLSHNDV-------------NARRAWLESQQFSPQEDFTLT
CFYLAETLMVIDHQKKSTRIQASLFAPNEEEKQRLTARLNELRQQLTEAAPPLPVVSVPH
190 200 210 220 230 240
SDWQSNMTREQYGEKFRQVQEYLHSGDCYQVNLAQRFHATYSGDEWQAFLQLNQANRAPF
MRCECNQSDEEFGGVVRLLQKAIRAGEIFQVVPSRRFSLPCPSPLA-AYYVLKKSNPSPY
250 260 270 280 290
SAFLRLEQGAILSLSPERFILCDNSEIQTR--PIKGTLPRLP------DPQEDSKQAVKL
MFFMQDNDFTLFGASPESSLKYDATSRQIEIYPIAGTRPRGRRADGSLDRDLDSRIELEM
300 310 320 330 340 350
ANSAKDRAENLMIVDLMRNDIGRVAVAGSVKVPELFVVEPFPAVHHLVSTITAQLPEQLH
RTDHKELSEHLMLVDLARNDLARICTPGSRYVADLTKVDRYSYVMHLVSRVVGELRHDLD
360 370 380 390 400 410
ASDLLRAAFPGGSITGAPKVRAMEIIDELEPQRRNAWCGSIGYLSFCGNMDTSITIRTLT
ALHAYRACMNMGTLSGAPKVRAMQLIAEAEGRRRGSYGGAVGYFTAHGDLDTCIVIRSAL
420 430 440 450
AINGQIFCSAGGGIVADSQEEAEYQETFDKVNRILKQLEK
VENGIATVQAGAGVVLDSVPQSEADETRNKARAVLRAIATAHHAQETF

FIG. 3. Amino acid alignment of PABS CoI and AS CoI. Reference numbers above the two sequences correspond to the PABS CoI sequence.

tophan residues in PABS CoI and CoII would aid in this process. Presumably, other proteins which use chorismate as a substrate, like those encoded by *pheA* and *tyrA*, will be found to contain tryptophan residues as well.

Putative regulatory sequences in *pabB*. Unlike *trpE*, *pabB* contains no attenuator-like sequences in its 5′ flanking region. In addition, little similarity exists between the 5′ flanking regions of these two genes. Nevertheless, the 5′ region of *pabB* contains a promoter-like sequence extending from nucleotide positions -81 to -35. The region from positions -73 to -67 has a possible RNA polymerase binding site (5′-TGTTAAC-3′ [20]), and 21 bp downstream from this area there is a Pribnow-like box (5′-TAAT-3′ [18]) at positions −46 to −43. The positions of these putative regulatory sequences suggest that transcription initiation occurs at position −36 before the initiation codon (23). Within the putative mRNA leader sequence, there is a Shine-Delgarno ribosome binding site at positions −9 to −4 (5′-TCAGGA-3′ [24]; these putative regulatory sequences are underlined in Fig. 2). At present, the translational start point of *pabB* is unknown. We believe, however, that the translation of *pabB* initiates at the methionine codon shown in Fig. 2 because the putative regulatory regions lie directly upstream from it.

Examination of the 3′ region of *pabB* revealed no canonical rho-independent termination sequence. Currently, experiments are in progress to determine whether the abovementioned proposed regulatory regions affect *pabB* expression.

DISCUSSION

We have established the complete nucleotide sequence of *E. coli pabB*, including the 5′ and 3′ flanking regions. The coding sequence of *pabB* can be aligned with that of *E. coli trpE*. The two genes have similarities of 40% at the nucleotide level and 26% at the amino acid level. These data suggest that *pabB* and *trpE* arose from a common ancestor since similarities exist throughout most of the alignment of

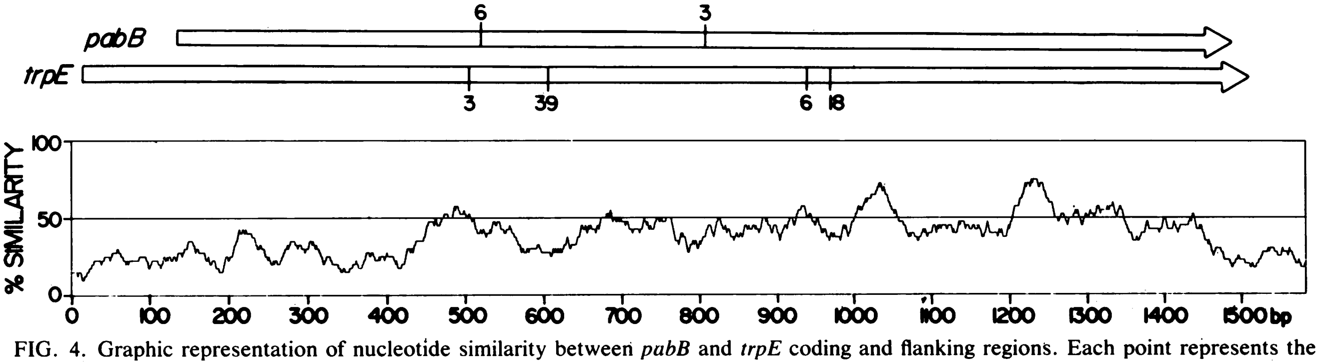

FIG. 4. Graphic representation of nucleotide similarity between *pabB* and *trpE* coding and flanking regions. Each point represents the amount of similarity present within a stretch of 40 nucleotides and is placed at the beginning of the corresponding nucleotide region. Above this graph are open arrows corresponding to the coding regions of *pabB* and *trpE*. Relative positions of the gaps introduced in the amino acid alignment of these two gene products are indicated by vertical lines drawn through the open arrows. Reference numbers corresponding to these lines represent the numbers of nucleotides deleted.

TABLE 1. Codon usage for *E. coli pabB* and *trpE*

Amino acid	Codon	Codons in gene[a]:		Amino acid	Codon	Codons in gene[a]:	
		pabB	*trpE*			*pabB*	*trpE*
Phe	TTT	13 (62)	8 (40)		TAG	0 (0)	0 (0)
	TTC	8 (38)	12 (60)				
				His	CAT	7 (58)	7 (64)
Leu	TTA	9 (19)	7 (11)		CAC	5 (42)	4 (36)
	TTG	4 (9)	5 (8)				
	CTT	6 (13)	6 (9)	Gln	CAA	8 (27)	8 (38)
	CTC	5 (11)	10 (15)		CAG	22 (73)	13 (62)
	CTA	3 (6)	4 (6)				
	CTG	20 (43)	35 (52)	Asn	AAT	10 (67)	10 (59)
					AAC	5 (33)	7 (41)
Ile	ATT	16 (64)	12 (71)				
	ATC	7 (28)	5 (29)	Lys	AAA	10 (77)	12 (80)
	ATA	2 (8)	0 (0)		AAG	3 (23)	3 (20)
Met	ATG	8 (100)	12 (100)	Asp	GAT	23 (79)	21 (60)
					GAC	6 (21)	14 (40)
Val	GTT	6 (22)	5 (15)				
	GTC	6 (22)	6 (18)	Glu	GAA	25 (81)	30 (86)
	GTA	6 (22)	8 (24)		GAG	6 (19)	5 (14)
	GTG	9 (33)	14 (42)				
				Cys	TGT	1 (17)	5 (42)
Ser	TCT	4 (13)	6 (16)		TGC	5 (83)	7 (58)
	TCC	4 (13)	5 (14)				
	TCA	3 (10)	4 (11)	End	TGA	0 (0)	1 (100)
	TCG	3 (10)	5 (14)				
				Trp	TGG	7 (100)	0 (0)
Pro	CCT	3 (13)	3 (11)				
	CCC	3 (13)	5 (18)	Arg	CGT	10 (36)	17 (43)
	CCA	8 (35)	5 (18)		CGC	14 (50)	21 (53)
	CCG	9 (39)	15 (54)		CGA	1 (4)	1 (3)
					CGG	3 (11)	0 (0)
Thr	ACT	8 (31)	4 (16)				
	ACC	9 (35)	14 (56)	Ser	AGT	4 (13)	4 (11)
	ACA	2 (8)	4 (16)		AGC	13 (42)	13 (35)
	ACG	7 (27)	3 (12)				
				Arg	AGA	0 (0)	1 (3)
Ala	GCT	6 (16)	12 (23)		AGG	0 (0)	0 (0)
	GCC	14 (37)	19 (36)				
	GCA	8 (21)	6 (11)	Gly	GGT	8 (32)	11 (39)
	GCG	10 (26)	16 (30)		GGC	13 (52)	12 (43)
					GGA	2 (8)	3 (11)
Tyr	TAT	8 (73)	9 (64)		GGG	2 (8)	2 (7)
	TAC	3 (27)	5 (36)				
End	TAA	1 (100)	0 (0)				

[a] Numbers in parentheses show the percentage of residues of the indicated amino acid coded by the indicated codon.

the two genes (Fig. 3). If these genes arose through convergent evolution, we would expect to see little similarity between them. The data show that this is not the case. Stretches of no similarity between the two sequences do, however, exist. This is an indication of the great amount of divergence which has occurred between the two genes.

In addition to our results indicating that PABS CoI and AS CoI have a common evolutionary origin, it has been proposed that PABS CoII and AS CoII have also evolved from a common ancestor (10). Both sets of data are in agreement with immunological studies suggesting that *E. coli* PABS and AS share common antigenic determinants (19). The amount of similarity between PABS CoII and AS CoII, 44% at the amino acid level, is greater than the 26% amino acid similarity present between PABS CoI and AS CoI. This is to be expected since PABS CoII and AS CoII have identical roles of transferring the amino group from glutamine to the CoI subunit of each respective enzyme complex.

On the other hand, the lower amount of similarity present between PABS CoI and AS CoI can be ascribed to several factors. (i) AS CoI responds to feedback inhibition by tryptophan, whereas PABS CoI does not. Therefore, tryptophan binding areas of AS CoI will represent an area (or areas) with no similarity to PABS CoI. This region could possibly exist in the amino-terminal end of AS CoI since this area shows the least amount of similarity to PABS CoI. (ii) PABS CoI and AS CoI have slightly different catalytic functions. (iii) Subunit interaction areas are different in PABS CoI and AS CoI.

Since the complete nucleotide sequences of all the genes coding for PABS and AS are known, we can now consider the evolutionary origins of these two enzymes. It has been proposed that the development of new enzyme functions is most easily achieved by recruiting proteins which already exist and catalyze similar reactions (9). We believe that the existence of PABS and AS is an actual case of acquisitive evolution which occurred after duplication of the genes encoding the initial synthetase complex. The fact that this duplication event actually occurred is supported by data presented here and elsewhere (10).

ACKNOWLEDGMENTS

We thank Jeffrey B. Kaplan for advice and criticism and for help with the figures. We also thank Charles Yanofsky, in whose laboratory pBN*pabB*(R3) and pBN*pabB*(H1) were constructed.

This work was supported by a grant from the University of Illinois at Chicago Circle Research Board and by Public Health Service grant AI 18639 from the National Institutes of Health.

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
2. **Crawford, I. P.** 1975. Gene rearrangements in the evolution of the tryptophan synthetic pathway. Bacteriol. Rev. **39:**87–120.
3. **Gough, J. A., and N. E. Murray.** 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166:**1–19.
4. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9:**r43–r74.
5. **Huang, M., and F. Gibson.** 1970. Biosynthesis of 4-aminobenzoate in *Escherichia coli*. J. Bacteriol. **102:**767–773.
6. **Huang, M., and J. Pittard.** 1967. Genetic analysis of mutant strains of *Escherichia coli* requiring *p*-aminobenzoic acid for growth. J. Bacteriol. **93:**1938–1942.
7. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151:**389–409.
8. **Ito, J., E. C. Cox, and C. Yanofsky.** 1968. Anthranilate synthetase, an enzyme specified by the tryptophan operon of *Escherichia coli*: purification and characterization of component I. J. Bacteriol. **97:**725–733.
9. **Jensen, R. A.** 1976. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. **30:**409–425.
10. **Kaplan, J. B., and B. P. Nichols.** 1983. Nucleotide sequence of *Escherichia coli pabA* and its evolutionary relationship to *trp(G)D*. J. Mol. Biol. **168:**451–468.
11. **Krikos, A., N. Mutoh, A. Boyd, and M. I. Simon.** 1983. Sensory transducers of *E. coli* are composed of discrete structural and functional domains. Cell **33:**615–622.
12. **Mandel, M., and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53:**159–162.
13. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9:**309–321.
14. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19:**269–276.
15. **Miozzari, G. F., and C. Yanofsky.** 1979. The regulatory region of the *trp* operon of *Serratia marcescens*. Nature (London) **276:**684–689.
16. **Nichols, B. P., G. F. Miozzari, M. vanCleemput, G. N. Bennet, and C. Yanofsky.** 1980. Nucleotide sequences of the *trpG* regions of *Escherichia coli*, *Shigella dysenteriae*, *Salmonella typhimurium* and *Serratia marcescens*. J. Mol. Biol. **142:**503–517.
17. **Nichols, B. P., M. vanCleemput, and C. Yanofsky.** 1981. Nucleotide sequence of *Escherichia coli trpE*. Anthranilate synthetase component I contains no tryptophan residues. J. Mol. Biol. **146:**45–54.
18. **Pribnow, D.** 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J. Mol. Biol. **99:**419–443.
19. **Reiners, J. J., L. J. Messenger, and H. Zalkin.** 1978. Immunological cross-reactivity of *Escherichia coli* anthranilate synthetase, glutamate synthase and other proteins. J. Biol. Chem. **253:**1226–1233.
20. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13:**319–353.
21. **Sanger, F., and A. R. Coulson.** 1978. Use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87:**107–110.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74:**5463–5467.
23. **Scherer, G. E. F., M. D. Walkinshaw, and S. Arnott.** 1978. A computer aided oligonucleotide analysis provides a model sequence for RNA polymerase-promoter recognition in *E. coli*. Nucleic Acids Res. **5:**3759–3773.
24. **Shine, J., and L. Delgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71:**1342–1346.
25. **Squires, C. H., M. DeFelice, J. Devereux, and J. M. Calvo.** 1983. Molecular structure of *ilvIH* and its evolutionary relationship to *ilvG* in *Escherichia coli* K12. Nucleic Acids Res. **11:**5299–5313.
26. **Yanofsky, C., T. Platt, I. P. Crawford, B. P. Nichols, G. E. Christie, H. Horowitz, M. vanCleemput, and A. M. Wu.** 1981. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. Nucleic Acids Res. **9:**6647–6668.
27. **Zalkin, H., and T. Murphy.** 1975. Utilization of ammonia for tryptophan synthesis. Biochem. Biophys. Res. Commun. **67:**1370–1377.