JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 541–548
0021-9193/91/020541-08$02.00/0

Vol. 173, No. 2

Nucleotide Sequence and Analysis of a Gene Encoding Anthranilate Synthase Component I in *Spirochaeta aurantia*

B. BRAHAMSHA,[1] CHUN-YA HAN,[2] IRVING P. CRAWFORD,[2]† AND E. P. GREENBERG[2]*

Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637,[1] and Department of Microbiology, University of Iowa, Iowa City, Iowa 52242[2]

Received 22 June 1990/Accepted 7 November 1990

A *Spirochaeta aurantia* DNA fragment containing the *trpE* gene and flanking chromosomal DNA was cloned, and the sequence of the *trpE* structural gene plus 870 bp upstream and 1,257 bp downstream of *trpE* was determined. The *S. aurantia trpE* gene codes for a polypeptide of 482 amino acid residues with a predicted molecular weight of 53,629 that showed sequence similarity to TrpE proteins from other organisms. The *S. aurantia* TrpE polypeptide is not more closely related to the other published spirochete TrpE sequence (that of *Leptospira biflexa*) than to TrpE polypeptides of other bacteria. Two additional complete open reading frames and one partial open reading frame were identified in the sequenced DNA. One of the complete open reading frames and the partial open reading frame are upstream of *trpE* and are encoded on the DNA strand opposite that containing *trpE*. The other open reading frame is downstream of *trpE* and on the same DNA strand as *trpE*. On the basis of the results of a protein sequence data base search, it appears that *trpE* is the only tryptophan biosynthesis gene in the sequenced DNA. This is in contrast to *L. biflexa*, in which *trpE* is separated from *trpG* by only 64 bp.

The spirochetes constitute one of the major lines of eubacterial descent (24). Although these organisms share a distinctive morphology and a unique type of motility, there exists among them a great deal of diversity with respect to ecology and metabolism (18), and 16S rRNA oligonucleotide analysis has revealed an early divergence of the two major clusters: the *Spirochaeta-Treponema-Borrelia* cluster and the *Leptospira* cluster (24). In part because there has not yet been a genetic transfer system described for any spirochete, little is known about gene organization in these bacteria. However, several spirochete genes have been cloned and sequenced (for example, see references 1, 3, 19, 23, and 29 to 31), and the chromosome of one of the spirochete species, *Borrelia burgdorferi*, is so far unique among procaryotes in that it is linear rather than circular (13).

Because the tryptophan biosynthetic pathway represents a useful paradigm for the comparative study of gene organization and expression (5–7, 28), spirochete *trp* genes are of particular interest. The *trpE* gene and the adjacent *trpG* gene, which encode components I and II, respectively, of anthranilate synthase (AS), the first enzyme in the tryptophan biosynthetic pathway, were cloned from *Leptospira biflexa* (29, 30) and have been sequenced (31). To begin our analysis of the molecular genetics of the *Spirochaeta-Treponema-Borrelia* cluster, we cloned *Spirochaeta aurantia* DNA that complemented an *Escherichia coli trpE* deletion (1). *S. aurantia* is a facultatively anaerobic bacterium isolated from freshwater sediments (4) and has received considerable attention as an easily cultivated model organism for studying various aspects of spirochete physiology and molecular biology (2, 3, 14–17).

In *E. coli*, the *S. aurantia trpE*-complementing DNA directed the synthesis of two polypeptides with apparent molecular weights of 59,000 and 47,000. The two polypeptides appear to be encoded by a single gene, with the larger polypeptide representing a readthrough product. A transposon insertion analysis indicated that the 47,000-dalton polypeptide was required for complementation (1). Furthermore, the results of in vitro enzyme assays and in vivo growth studies indicated that the cloned DNA encoded an ASI but not an ASII activity (1) and that the *S. aurantia* ASI was capable of interacting with *E. coli* ASII.

As a next step in the analysis of *S. aurantia trp* genes, we have sequenced and analyzed the cloned *trpE*-complementing DNA and the DNA which flanks it on the *S. aurantia* chromosome. We have also compared the deduced amino acid sequence of the encoded TrpE polypeptide with those of polypeptides encoded by *trpE* genes from other organisms, including the distantly related spirochete *L. biflexa*.

MATERIALS AND METHODS

Plasmids. The plasmids used in this study are described below. The construction of pBG100, which contains 2.7 kbp of *S. aurantia* DNA, has been described elsewhere (1). A plasmid containing approximately 8.6 kbp of *S. aurantia* DNA, pBG101, was isolated from an *S. aurantia* genomic library by complementation of the *trpE* deletion in *E. coli* JA221. The construction of the genomic library, which consisted of *Sau*3AI fragments of *S. aurantia* DNA ligated into the *Bam*HI site of pBR322, and the selection of complementing plasmids have been described previously (1).

To facilitate DNA sequencing, *S. aurantia* DNA fragments from pBG100 and pBG101 were subcloned. In pBG100 there are three *Sal*I sites and two *Kpn*I sites (Fig. 1). A deletion between the *Sal*I site in the pBR322 portion of pBG100 and the *Kpn*I site outside *trpE* was made, using the *Sal*I-*Kpn*I segment of the multiple-cloning site of pUC19 to connect the ends (p1422). Also useful for sequencing was p1420, which contained the 650-bp *Kpn*I-*Sal*I fragment from the middle of the *trpE* gene inserted (in reverse orientation) into the largest *Kpn*I-*Sal*I segment of pBG100. Two additional plasmids, p1446 and p1451, were used to sequence spirochete DNA from pBG100. To construct p1446, the ca.

* Corresponding author.
† Deceased 8 October 1989.

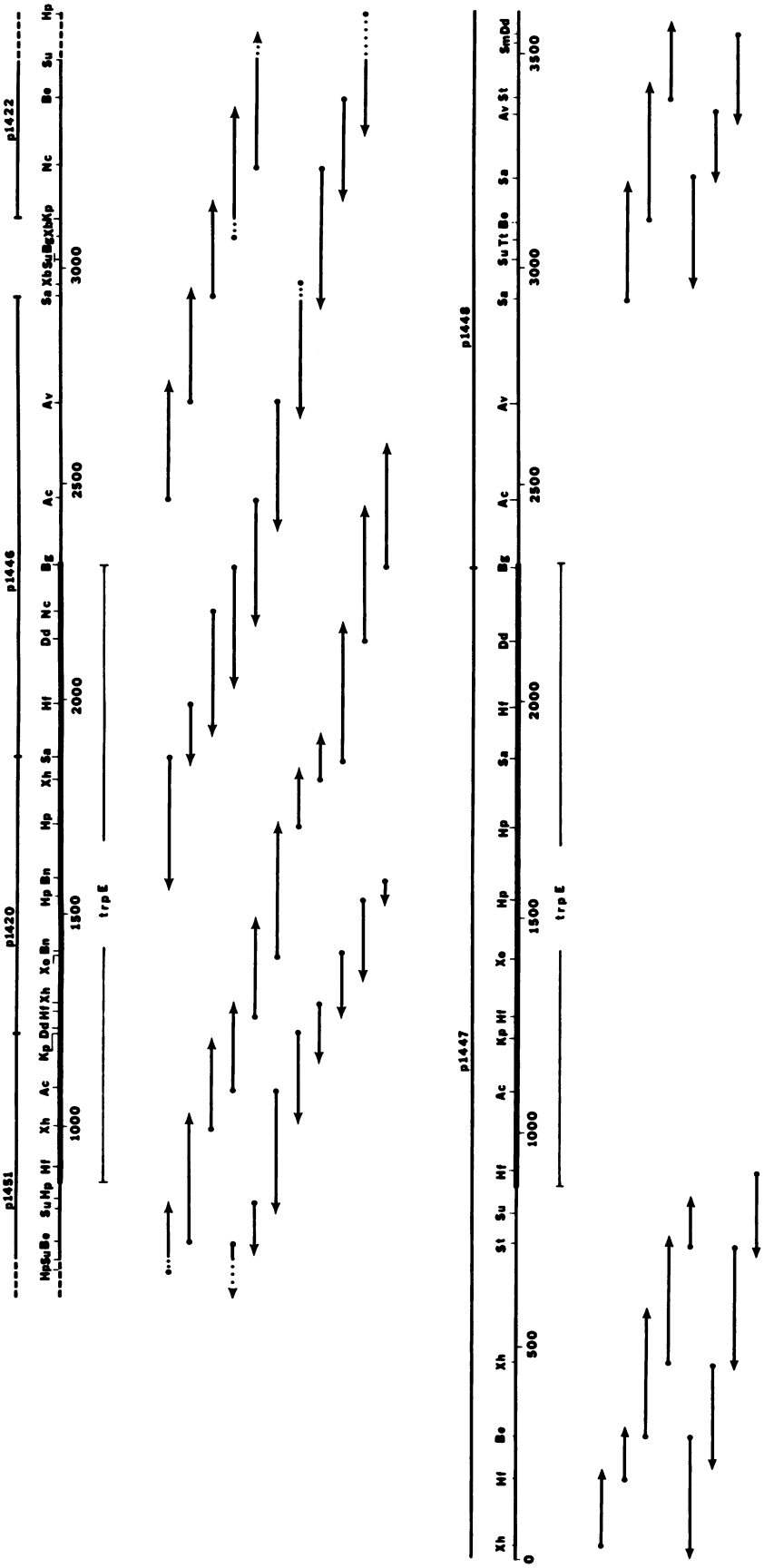

FIG. 1. Restriction maps of pBG100 (top) and pBG101 (bottom) and DNA-sequencing strategy. The *trpE*-containing segment common to pBG100 and pBG101 extends from the *Sau*3A1 site at position 814 to the *Sau*3A1 site at position 3021. Only the restriction enzyme sites used in subcloning or radioactively labeling fragments are shown. The bars above the plasmids represent the segments of *S. aurantia* DNA from pBG100 or pBG101 that were subcloned (p1451, 1420, 1446, 1447, and 1448) or that remained after deletion (p1422). These derivative plasmids, made to facilitate sequencing, are designated above the bars. Dashed lines at the ends of pBG100 represent the vector, pBR322. Dotted portions of the sequencing arrows similarly indicate nonspirochete DNA. The region from the filled circle to the arrowhead shows the length of clearly readable DNA sequence. Arrows pointing to the right show sequence on the *trpE* coding strand; those pointing to the left show sequence on the noncoding strand. The *trpE* coding region is indicated by the bold line for each plasmid, with translation proceeding from left to right. Abbreviations: Ac, *Acc*I; Av, *Ava*I; Be, *Bst*EII; Bg, *Bgl*II; Bn, *Bst*NI; Dd, *Dde*I; Hp, *Hin*pI; Hf, *Hin*fI; Kp, *Kpn*I; Nc, *Nco*I; Sa, *Sal*I; Sm, *Sma*I; St, *Sty*I; Su, *Sau*3AI; Tt, *Tth*I; Xb, *Xba*I; Xh, *Xho*I.

1-kbp *Sal*I-*Sal*I fragment containing the distal part of *trpE* was cloned into pUC8, and to construct p1451, the ca. 800-bp *Sal*I-*Kpn*I fragment that contained part of the pBR322 *tet* gene and the 5′ end of the *trpE* gene was cloned into pUC19.

Two subclones were made to facilitate sequencing of the *trpE*-flanking DNA cloned in pBG101. One, p1447, contains the ca. 4-kbp *Bgl*II-*Bgl*II fragment with most of *trpE* and its 5′ flanking region cloned into the *Bam*HI site of pUC19, and the other, p1448, contains the ca. 2-kbp *Bgl*II-*Eco*RI fragment that has the rest of *trpE* and its 3′ flanking region similarly cloned into pUC19.

DNA-DNA hybridizations. Restriction fragments to be used as probes were electroeluted from 5% polyacrylamide gels. After ethanol precipitation, the DNA fragments were ^{32}P labeled by nick translation (25). Southern blot analyses were carried out as described previously (1).

DNA sequencing. Fragments were labeled with ^{32}P by fill-in labeling using the large fragment of DNA polymerase I and the appropriate α-^{32}P-labeled radioactive deoxynucleotide (26). Labeled fragments electroeluted from polyacrylamide gels were sequenced by the procedure of Maxam and Gilbert (22); reaction mixtures were subjected to electrophoresis through 8% urea-polyacrylamide gels as described by Sanger and Coulson (27), except for the addition of 25% formamide to the gels to minimize compressions. Sequence data were analyzed with the aid of the PCS computer program (21).

Normalized alignment score. The alignment used was extracted from one published elsewhere (7) containing more sequences. To estimate similarity between homologous or paralogous sequences we employed the normalized alignment score (8), which considers only identical residues, with a double value given to Cys residues and a gap value of 2.5. Each score was normalized by dividing by the number of positions compared and multiplying by 1,000. All scores reported are significantly higher than would be obtained by chance after randomization of one of the sequences. For sequences greater than 200 residues long, similarity values greater than 160 indicate that common ancestry is probable, and values above 280 make it certain; for longer sequences, somewhat lower values are significant (9).

Nucleotide sequence accession number. The DNA sequence reported here has been deposited in GenBank (accession number M55917).

RESULTS

DNA-sequencing strategy. The strategy used to sequence the *S. aurantia trpE* gene and flanking regions is shown in Fig. 1. We chose initially to sequence the insert of chromosomal DNA in pBG100, the smaller of the two plasmids that complemented the *trpE* mutation in *E. coli* JA221. In the sequence we found an intact gene that encoded a protein resembling TrpE proteins of other bacteria (see below). To obtain additional *trpE*-flanking sequence we examined pBG101 (Fig. 1) and in the process found that the sequences of *S. aurantia* DNA in pBG100 and pBG101 differed starting at the *Sau*3A1 site downstream of *trpE* in pBG100 and at the rightmost *Sau*3A1 site of pBG101 (Fig. 1). The sequence obtained from pBG101 overlapped the sequence obtained from pBG100 in the region common to both plasmids by 47 bp on the left side and 171 bp on the right side of this region. The ends of the common regions of pBG100 and pBG101 consist of *Sau*3A1 sites, as expected from the manner in which these two plasmids were obtained (1).

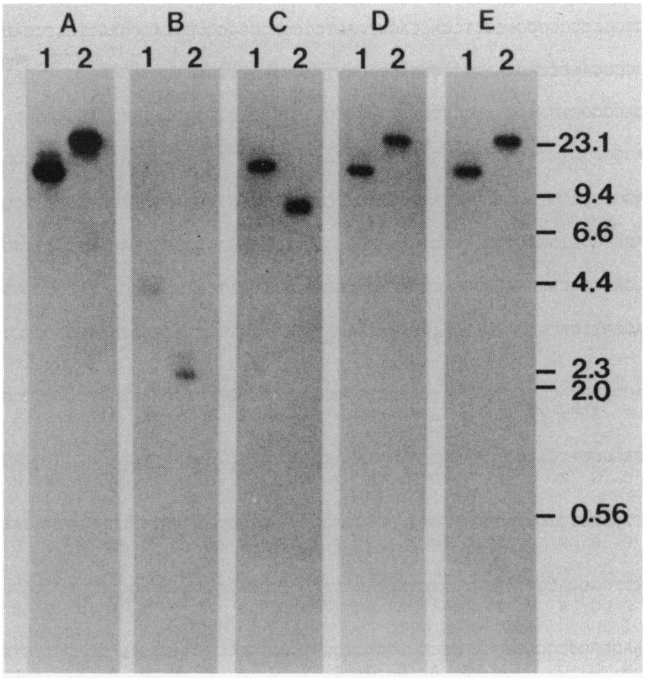

FIG. 2. Southern blot analysis of pBG100 and pBG101. The probes were (A) the *Sal*I-*Sal*I insert from p1446, which served as a probe for the *trpE* gene segment common to pBG100 and pBG101, (B) the 369-bp *Bgl*II-*Bst*EII fragment unique to the right end of pBG100, (C) the 120-bp *Sau*3A1-*Sau*3A1 fragment of p1451 unique to the left end of pBG100, (D) the 472-bp *Tth*I-*Sma*I fragment from p1448 unique to the right end of pBG101, and (E) the 460-bp *Bst*EII-*Sty*I fragment from p1447 unique to the left end of pBG101. Lanes 1, *Nru*I digests of *S. aurantia* DNA; lanes 2, *Bcl*I digests of *S. aurantia* DNA. The numbers to the right indicate the sizes (in kilobases) of standard fragments prepared by digestion of lambda DNA with *Hin*dIII.

DNA-DNA hybridizations. To determine which of the two *trpE*-complementing plasmids, pBG100 or pBG101, contained the flanking chromosomal DNA, a Southern blot analysis was carried out in which *Bcl*I and *Nru*I digests of *S. aurantia* chromosomal DNA were probed with DNA common to pBG100 and pBG101, DNA unique to the left end of pBG100, DNA unique to the right end of pBG100, DNA unique to the left end of pBG101, or DNA unique to the right end of pBG101. The probes unique to pBG101 and the probe to the common section hybridized with a 25-kbp *Bcl*I fragment and a 14-kbp *Nru*I fragment (Fig. 2). On the other hand, the probe unique to the right end of pBG100 recognized a 2.3-kbp *Bcl*I fragment and a 4-kbp *Nru*I fragment, and the one unique to the left end hybridized to an 8.6-kbp *Bcl*I fragment and a 17-kbp *Nru*I fragment (Fig. 2). The results indicate that pBG101 contains the chromosomal *trpE*-flanking sequences, while those in pBG100 arose from the ligation of *Sau*3A1 fragments from elsewhere in the chromosome.

DNA sequence analysis. We have obtained the sequence of 3,575 bp of contiguous *S. aurantia* chromosomal DNA (Fig. 3); this region contains an open reading frame (ORF) of 482 codons ending with a TAG stop codon (Fig. 3). The ORF encodes a polypeptide that shows considerable similarity to other TrpE polypeptides (Fig. 3 and see below). The start codon for this ORF is preceded by a region showing com-

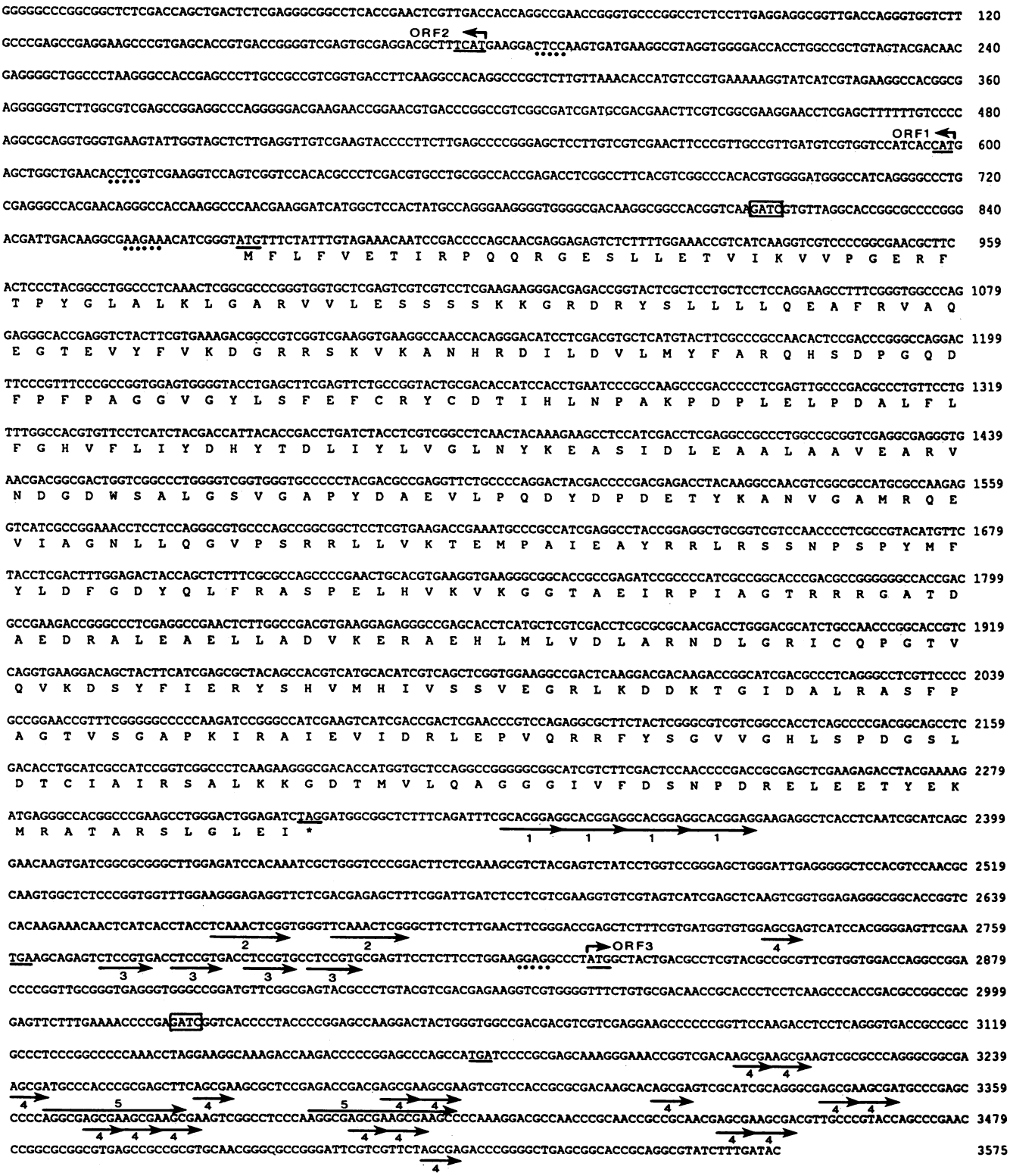

FIG. 3. DNA sequence and deduced amino acid sequence of *S. aurantia trpE* and flanking regions. Start and stop codons for *trpE* and the three ORFs are underlined, as is the downstream TGA stop codon in frame with *trpE*. A potential stem-and-loop structure is located between the *trpE* stop codon and the in-frame downstream TGA stop codon around bp 2405 to 2549. Presumed ribosome-binding sites for *trpE* and the three ORFs are marked by dotted lines. The two *Sau*3A1 sites (GATC) marking the boundaries of the segment common to pBG100 and pBG101 are boxed. Outside the common segment the pBG101 sequence is shown. Repeated sequences are indicated by the numbered arrows: 1, GCACGGAG; 2, TCAAACTCGG; 3, CTCCGTG; 4, AGCGA; 5, AGGCGAGCGAAGCGAAGC.

TABLE 1. Codon usage in *S. aurantia* genes

Codon	No. of times used in:				Codon	No. of times used in:			
	trpE	ORF1+ORF2+ORF3	*trpE* extension	*flaA*		*trpE*	ORF1+ORF2+ORF3	*trpE* extension	*flaA*
TTT	7	4	3	1	TAT	0	0	0	0
C	13	17	4	12	C	21	12	2	4
A	0	0	0	0	A	0	0	0	0
G	3	3	3	6	G	1	0	0	1
CTT	1	1	2	3	CAT	1	0	0	1
C	32	20	14	10	C	9	4	6	9
A	1	1	0	0	A	4	2	4	1
G	15	7	3	5	G	10	4	0	13
ATT	0	0	1	5	AAT	1	0	1	0
C	22	4	7	9	C	8	8	1	9
A	0	0	0	0	A	3	3	4	2
G	9	6	0	5	G	18	19	3	8
GTT	2	1	2	3	GAT	0	2	3	3
C	17	13	8	12	C	35	18	1	10
A	1	0	1	4	A	14	2	5	6
G	18	15	5	7	G	20	12	13	12
TCT	1	0	2	3	TGT	0	0	0	0
C	5	1	0	5	C	4	1	0	0
A	0	1	0	0	A	0	2	1	0
G	14	9	2	4	G	1	2	1	1
CCT	0	4	0	2	CGT	1	0	1	1
C	22	13	1	9	C	11	5	3	6
A	0	1	0	6	A	7	0	1	3
G	3	5	2	5	G	13	9	3	1
ACT	1	2	0	2	AGT	0	2	0	2
C	15	14	2	9	C	9	5	3	2
A	1	1	0	0	A	0	0	1	0
G	1	5	2	5	G	7	5	3	1
GCT	0	1	2	6	GGT	2	2	4	13
C	40	24	0	16	C	23	15	5	17
A	0	2	2	2	A	8	3	4	2
G	3	2	2	7	G	5	6	5	11

plementarity to the 3′ end of the *S. aurantia* 16S rRNA (5′-GAUCACCUCCUUU-3′ [27a]) located 9 bp ahead (Fig. 3). Further upstream from *trpE* there are several sequences that could serve as weak promoters in *E. coli*. However, to our knowledge there is no available information on the nature of spirochete promoters. The codon usage for the *S. aurantia trpE* gene (Tables 1 and 2) is characteristic of genes from bacteria with DNA that has a high G+C content (10). Furthermore, the codon usage is consistent with that for *flaA*, the only other *S. aurantia* gene we know to have been sequenced (3).

TABLE 2. G+C content in *S. aurantia* genes

Codon position	G+C content (%) in:			
	trpE	ORF1+ORF2+ ORF3	*trpE* extension	*flaA*
First base	65.8	60.7	68.9	68.2
Second base	40.8	45.1	35.1	46.7
Third base	88.8	89.4	70.2	76.5
Avg	65.0	63.2	58.1	63.8
Total no. of codons	483	318	148	302

We found no evidence of an ORF corresponding to *trpG*, which encodes ASII. This is in contrast to the situation for *L. biflexa*, in which the *trpG* structural gene follows *trpE* by 64 bp (31). Upstream of *trpE* we observed one complete and one partial ORF (Fig. 3) showing codon usage characteristic of genes from DNA with high G+C contents (Tables 1 and 2). The orientation of these putative genes is opposite that of *trpE*. We designated them ORF1 (complete, 140 codons plus TGA) and ORF2 (partial, 60 codons). The ORF1 stop codon overlaps the ORF2 start codon. There are 271 bp between the start codon of ORF1 and the start codon of *trpE*. Potential ribosome-binding sites are located 6 and 14 bp ahead of the suspected ATG start codons of the two putative genes (Fig. 3).

Downstream of *trpE* we observed an ORF (Fig. 3) showing codon usage characteristic of genes from DNA with high G+C contents (Tables 1 and 2). We designated this ORF3 (115 codons plus TGA). It is located several hundred base pairs downstream of the *trpE* translation stop codon, and transcription would be expected to proceed in the same direction as transcription of *trpE*. A potential ribosome-

binding site precedes this ORF by 9 bp (Fig. 3). Between *trpE* and ORF3 is an 8-bp sequence repeated four times. This is followed by a possible stem-and-loop type of secondary structure (bp 2405 to 2549; calculated free energy of formation, −51.3 kcal [ca. −215 KJ]/mol), after which there is a 10-bp repeat separated by 5 bp. There is another string of direct repeats just upstream of ORF3 (Fig. 3). Finally, downstream of ORF3 there is an 18-bp direct repeat, and the sequence AGCGA occurs 17 times in this region (Fig. 3).

The ORF corresponding to *trpE* encodes a polypeptide of 482 amino acid residues having a predicted molecular weight of 53,629. The ORF ends with an amber codon, TAG. The next stop codon in the same reading frame is 441 bp downstream, a TGA (Fig. 3). This can explain the two sizes of gene products seen when *S. aurantia trpE* DNA was expressed in *E. coli* maxicells (1), since the maxicell strain contained an amber suppressor. Codon usage in the region between the TAG and TGA stop codons is not highly biased towards G or C in the third position, as is the case for the three ORFs, *trpE*, and the previously sequenced *flaA* (Table 2).

Homology with other enzymes. Little difficulty was encountered in aligning the amino acid sequence deduced for the *S. aurantia* TrpE, ASI, with other published sequences, including the eucaryotic homolog *Saccharomyces cerevisiae* ASI and the paralogous *p*-aminobenzoate synthase from *E. coli* (Fig. 4 and reference 7). Table 3 presents normalized alignment scores calculated from Fig. 4. Several features noted previously in sequence comparisons from nonspirochete proteins are also observed when the two spirochete ASs are included. The N-terminal half of the protein is much less conserved than the C-terminal half (normalized alignment scores for the N-terminal half frequently fall below the level required to demonstrate homology). Moreover, the scores for alignment of nonenteric bacterial sequences to the paralogous subunit from *p*-aminobenzoate synthase from *E. coli* are often as high as those for alignment to the homologous AS subunit (7). The highest score for the polypeptides in Table 3 is for the comparison of *E. coli* TrpE with the TrpE of a gram-positive, coryneform species, *Brevibacterium lactofermentans*. The second highest is also for a gram-negative–gram-positive comparison, *Pseudomonas aeruginosa* TrpE with *Bacillus subtilis* TrpE. The *S. aurantia* polypeptide is approximately equidistant to all of the other proteins in the alignment; it actually scores a little higher with the *E. coli* TrpE and the *B. subtilis* TrpE than with the TrpE of the other spirochete, *L. biflexa*. The latter scores slightly higher with the *S. aurantia* TrpE than with anything else, but this score is not significantly higher than the score with the *P. aeruginosa* TrpE or the *B. subtilis* TrpE.

The amino acid sequences from ORFs 1, 2, and 3 were compared with sequences in a data base of known proteins with the FASTA program developed by Pearson and Lipman (25). The survey did not detect any identities that we judged to be significant.

DISCUSSION

We have sequenced the complete *S. aurantia trpE* gene and flanking regions. In all, 3,575 bp of contiguous *S. aurantia* chromosomal DNA was sequenced (Fig. 1 and 3). This includes 870 bp of DNA upstream of *trpE* and 1,257 bp of downstream DNA (Fig. 3). In addition to *trpE*, three ORFs that appear to be *S. aurantia* genes were identified in the *S. aurantia* DNA. The proteins encoded by these ORFs did not show significant similarity to TrpG proteins from other bacteria or to any other proteins in the protein sequence data base at the time of our search. In other bacteria, ASI is encoded by *trpE* and ASII is encoded by *trpG* or its equivalent (10–12, 20). The apparent lack of *trpG* in the cloned *trpE*-containing DNA supports our previous hypothesis that the *S. aurantia* ASI polypeptide can interact with the *E. coli* ASII polypeptide to form an AS active with either glutamine or NH_3 as the amino donor (1). The fact that the *S. aurantia trpE* gene is not closely linked to a *trpG* gene differs from the arrangement found in the *L. biflexa* chromosome, in which *trpE* and *trpG* are separated by 64 bp (31). This difference is not surprising considering the great evolutionary divergence between *S. aurantia* and *L. biflexa* (24). The *trp* genes are organized in many arrangements in other bacteria (5, 7, 10–12, 28).

It was reported previously that in *E. coli* maxicells, the *S. aurantia trpE*-containing DNA directed the synthesis of two polypeptides with apparent molecular weights of 59,000 and 47,000, with the larger polypeptide representing a readthrough product encoded by the same gene that encoded the smaller polypeptide (1). The ORF corresponding to *trpE* coded for a polypeptide consisting of 482 amino acid residues (Fig. 3) and having a predicted molecular weight of 53,629. This ORF stops with the amber codon, TAG (Fig. 3). The next stop codon in the same reading frame is 441 bp downstream of the amber codon; it is a TGA (Fig. 3). These facts can explain why two gene products are observed when *S. aurantia trpE* DNA is expressed in *E. coli*, since the maxicell strain used contains an amber suppressor (1). We conclude that the smaller of the two gene products is the full-length TrpE polypeptide for the following reasons. A transposon insertion analysis indicated that the full-length, large polypeptide was not required for *trpE* complementation in *E. coli* (1). Codon usage between the TAG and TGA stop codons is not highly biased towards G or C in the third position, as is the case for the three ORFs and *trpE* (up to the TAG stop codon [Table 2]). The sequence of the small, 53,629-dalton protein is similar to sequences of other TrpE proteins (Fig. 4), and the sequence of the TrpE extension does not show such a similarity (data not shown). It should also be pointed out that within the sequence between the TAG and TGA there are several direct repeats, and this region also contains a sequence (between bp 2405 and 2549) that could fold into a stem-and-loop structure.

Besides the repeated DNA sequences mentioned above, other repeated sequences were observed also. Just upstream and 185 bp downstream of ORF3, direct repeat sequences were observed, and, curiously, the sequence AGCGA occurs 17 times in a 247-bp region starting 31 bp downstream of the ORF3 stop codon (Fig. 3).

Our alignment of the *S. aurantia trpE* product with other TrpE polypeptides shows clear sequence similarities, particularly in the C-terminal portion of the polypeptides (Fig. 4). Normalized alignment scores indicate that the *S. aurantia* TrpE protein is no more related to the TrpE protein of the distantly related spirochete, *L. biflexa*, than it is to TrpE proteins of other nonspirochetes (Table 3). It must be pointed out, however, that in the aligned TrpE sequences (Fig. 4) there are 19 amino acid residues unique to the two spirochetes. An example is the aligned sequence LIYD beginning at residue 155 in the *S. aurantia* TrpE protein and at residue 138 in the *L. biflexa* protein (Fig. 4). The significance of these sequence identities is unknown; however, the analysis of spirochete 16S rRNA sequences has revealed a similar situation (24). The 16S rRNAs of members of the *Leptospira* cluster are no more related to those of the

```
S.a. trpE                      M FLFVETIRPQ QRGESLLETV IKVVPGERFT PYGLALKLGA RVVLESSSSK KGRDRYSLLL   61
L.b. trpE                     MS QTLPKIKIPK KPNYNSLALA EGIEFWELF- -RVIEAKYEN CFLLESAGDN QYDSRYSVIG
P.a. trpE                     MN REEFLRLAAD GYNRIPLSFE TLADFDTPLS IYLKLADAPN SYLLESVQGG EKWGRYSIIG
B.s. trpE                  MNFQS NISAFLEDSL SHHTIPIVET FTVDTLTPI- QMIEKLDREI TYLLESKDDT STWSRYSFIG
S.c. trpE  MTASIKIQPD IDSLKQLQQQ NDDSSINMYP VYAYLPSLDL TPHVAYLKL- AQLNNPDRKE SFLLES-AKT NNEDRYSFIG
E.c. pabB                             M KTLSPAVITL LWRQDAAEF- YFSRLSHLPW AMLLHSGYAD HPYSRFDIVV
E.c. trpE                          MQTQKP TLELLTCEGA YRDNPTALF- -HQLCGDRPA TLLLESADID SKDDLKSLLL
B.l. trpE                             MST NPHVFSLDVR YHEDASALF- AHLGGTTADD AALLESADIT TENGISSLAV
                                                               *              ***       ***

LQE------- ---------- ---------- ---------- AFRVAQEGTE VYFVKDGRRS KVKA--NHRD ILDVLMYFAR Q-HSDPGQ-D  110
FQP------- ---------- ---------- ---------- -SHLI----- ILEIDGKKYP VEN---PYFA LRELTDYNSL ---------S
LPC------- ---------- ---------- ---------R TVLRVYDHQV RISIDGVETE RFDCADPLAF VEEFKARYQV P-TVPGLPRF
LNP------- ---------- ---------- ---------- FLTIKEEQGR FSAADQDSKS LYTGNELKEV LNWMNTTYKI KTPELGIP-F
ISP------- ---------- ---------- ---------- RKTIKTGPTE GIETD----- ------PLEI LEKEMSTFKV AENVPGLPKL
AEP------- ---------- ---------- ---------- ICTLTTFGKE TVVSESEKRT TTTDD-PLQV LQQVLDRADI RPTHNEDLPF
VDSALRITAL GDTVTIQALS GNGEALLALL DNALPAGVES EQSPNCRVLR FPPVSPLLDE DARL--CSLS VFDAFRLLQN LLNVPKEERE
LKSSVRITCT GNTVVTQPLT DSGRAVVARL TQQL------ GQYNTAENTF SFPASDAVDE RERL--TAPS TIEVLRKLQF --ESGYSDAS

FPFPAGGVGY LSFEFCRYCD TIHLNPAK-- -PDPLELPDA LFLFGHVFLI YDHYTDLIYL VGLNY----- --KEASIDLE AALAAVEARV  190
ISYAGGFVGY LGYQSMQFFE PKLQLKPH-- -PDFPAMIFG LYLDG---LI YDKFTGELIY ---------- --------FD NGTNRIHEVN
DGGLVGYFGY DCVRYVEKRL ATCPNPDPLG NPDILLMVSD AVVVFDNLAG KIHAIVLADP SEENA----- ----YERGQA RLEELLERLR
VGGAVGYLSY DMIPLIEPSV PSHTKETDM- -EKCMLFVCR TLIAYDHETK NVHFIQYARL TGEETKNEKM DVFHQNHLEL QNLIEKMMDQ
SGGAIGYISY DCVRYFEPKT RRPLKDVLR- LPEAYLMLCD TIIAFDNVFQ RFQIIHNINT NETSL----- --EEGYQAAA QIITDIVSKL
QGGALGLFGY DLGRRFESLP EIAEQDIVL- -PDMAVGIYD WALIVDHQRH TVSLLSHNDV NARRA----- ---------- ----------
AMFFSGLFSY DLVAGFEDLP QLSAENNC-- -PDFCFYLAE TLMVIDHQKK STRIQASLFA PNEEE----- ---------K QRLTARLNEL
LPLLMGGFAF DFLETFETLP AVEESVNTY- -PDYQFVLAE IVLDINHQDQ TAKLTGVSNA PGELE----- --------AE LNKLSLLIDA
     *   *                        **

NDGDWSALGS VGAPYDAEVL PQDYDPDETY KANVGAMRQE VIAGNLLQGV PSRRLLVKTE MPAIE---AY RRLRSSNPSP YMFYLDFGD-  276
QILEQLKKEN SQKPKATVSL VKAGLSKEVH KQMVEEALEE VKAGNTFQCQ IGFEEIYQVD GNPLA---IY ETLREINPSP HMYYVNLEL-
QPITPRRGLD LEAAQGREPA FRASFTREDY ENAVGRIKDY ILAGDCMQVV PSQRMSIEFK AAPID---LY RALRCFNPTP YMYFFNFGD-
KNIKELFLSA DSYKTPSFET VSSNYEKSAF MADVEKIKSY IKAGDIFQGV LSQKFEVPIK ADAFE---LY RVLRIVNPSP YMYYMKLLD-
DRRFLANTIP EQPPIKPNQL LNRMWARKVT KITSPTLKKH IKKGDIIQGV PSQRVARPSR YILSIFTDIY RHLRTINPSP YLFYIDCLD-
--WLESQQFS PQEDFTLTSD WQSNMTREQY GEKFRQVQEY LHSGDCYQVN LAQRFHATYS GDEWQ---AF LQLNQANRAP FSAFLRLEQ-
RQQLTEAAPP LPVVSVPHMR CECNQSDEEF GGVVRLLQKA IRAGEIFQVV PSRRFSLPCP SPLA----AY YVLKKSNPSP YMFFMQDND-
ALPATEHAYQ TTPHDGDTLR VVADIPDAQF RTQINELKEN IYNGDIYQVV PARTFTAPCP DAFA----AY LQLRATNPSP YMFYIRGLNE
                        *     *        **   * * ** *               *   **  **** ** *    *

---YQLFRAS PELHVKVKG- -GTAEIRPIA GTRRRGA--- ---TDAEDRA LEAELLADVK ERAEHLMLVD LARNDLGRIC QPGTVQVKDS  355
---VTILGAS PSSLFRLRQ- -GEMESFPLA GTTKRGV--- ---DAKEDTL LARKLLTDPK EIAEHNMLID LHRNDVGRVA KFGTVKVRRR
---FHVVGSS PEVLVRVED- -GLVTVRPIA GTRPRGI--- ---NEEADLA LEQDLLSDAK EIAEHLMLID LGRNDVGRVS DIGAVKVTEK
---REIVGSS PERLIHVQD- -GHLEIHPIA GTRKRGA--- ---DKAEDER LKVELMKDEK EKAEHYMLVD LARNDIGRVA EYGSVSVPEF
---FQIIGAS PELLCKSDS- KNRVITHPIA GTVKRGA--- ---ATEEDDA GADQLRGSLK DRAEHVMLVD LARNDINRIC DPLTTSVDKL
---GAILSLS PERFILCDN- -SEIQTRPIK GTLPRLP--- ---DPQEDSK QAVKLANSAK DRAENLMIVD LMRNDIGRVA VAGSVKVPEL
---FTLFGAS PESSLKYDAT SRQIEIYPIA GTRPRGRRAD GSLDRDLDSR IELEMRTDHK ELSEHLMLVD LARNDLARIC TPGSRYVADL
GRSYELFGAS PESNLKFTAA NRELQLYPIA GTRPRGLNPD GSINDELDIR NELDMRTDAK EIADDTMLVD LARNDLARVS VPASRRVADL
        ** **               *** *** **          **       *  * * * *** **** ***** **     * * *

YFIERYSHVM HIVSSVEGRL KDDKTGIDAL RASFPAGTVS GAPKIRAI-E VIDRLEPVQR RFYSGVVGHL SPDG-SLDTC IAIRSALKKG  443
FDVKRFSHVQ HISSEVVGIL SSKEDMFSGL ASSFPRGTLS GAPKIESDSK IIERIEKSPR GPYGGAVGSF GLNG-DCTFA IPIRSFFVNG
MVIERYSNVM HIVSNVTGQL REGLSAMDAL RAILPAGTLS GAPKIRAM-E IIDELEPVKR GVYGGAVGYL AWNG-NMDTA IAIRTAVIKN
TKIVSFSHVM HIISVVTGRL KKGVHPVDAL MSAFPAGTLT GAPKIRAM-Q LLQELEPTPR ETYGGCIAYI GFDG-NIDSC ITIRTMSVKN
LTIQKFSHVQ HLVSQVSGVL RPEKTRFDAF RTNFPAGTVS GAPKVRAM-E LIAELEGERR GVYAGAVGHW SYDGKTMDNC IALRTMVYKD
FVVEPFPAVH HLVSTITAQL PEQLHASDLL RAAFPGGSIT GAPKVRAM-E IIDELEPQRR NAWCGSIGYL SFCG-NMDTS ITIRTLTAIN
TKVDRYSYVM HLVSRVVGEL RHDLDALHAY RACMNMGTLS GAPKVRAM-Q LIAEAEGRRR GSYGGAVGYF TAHG-DLDTC IVIRSALVEN
LQVDRYSRVM HLVSRVTATL DPELDALDAY RACMNMGTLT GAPKLRAM-E LLRGVEKRRR GSYGGAVGYL RGNG-DMDNC IVIRSAFVQD
    * * ** * ** * * *        *** ** ** ** * **** **  *  *  **   *   * * **      *   * * * **

DTMVLQAGGG IVFDSNPDRE LEETYEKMRA TARSLGLEI*                                                        482
KKGFVRASGG IVFdsEPEDE YQEIINKMAS VRKALDLHKG P*
GELHVQAGGG IVADSVPALE WEETINKRRA MFRAVALAEQ SVE*
GVASIQAGAG IVADSVPEAE YEESCNKAGA LLKTIHIAED MFHSKEDKAD EQISTIVR*
GILTLqaggg ivydsleyde MLETMNNDGQ SQYYCASRRI VGRYRRISLK RAFSVFFPLD DIFIVFE*
GQIFCSAGGG IVADSQEEAE YQETFDKVNR ILKQLEK*
GIATVQAGAG VVLDSVPQSE ADETRNKARA VLRAIATAHH AQETF*
GVAAVQAGAG VVRDSNPQSE ADETLHKAYA VLNAIALAAG STLEVIR*
     ***** ** ** * *   **  *  *
```

FIG. 4. Deduced amino acid sequences for six TrpE proteins and one PabB protein. Standard single-letter abbreviations are used. Hyphens indicate gaps introduced to increase similarity. Residues in lowercase letters (*L. biflexa* residues 544 to 545 and *S. cerevisiae* residues 535 to 550) differ from the published versions; these residues were in an alternate reading frame, indicative of possible sequence errors. Asterisks indicate the positions of *S. aurantia* residues that are identical to residues of at least four of the other seven aligned sequences.

TABLE 3. Normalized alignment score for various TrpE proteins.

Organism	Alignment score with:						
	L. biflexa TrpE	*P. aeruginosa* TrpE	*B. subtilis* TrpE	*S. cerevisiae* TrpE	*E. coli* PabB	*E. coli* TrpE	*B. lactofermentans* TrpE
S. aurantia TrpE	253	213	261	248	213	269	222
L. biflexa TrpE		248	233	180	188	187	177
P. aeruginosa TrpE			319	272	282	252	216
B. subtilis TrpE				255	270	240	205
S. cerevisiae TrpE					220	224	208
E. coli PabB						198	185
E. coli TrpE							426

Spirochaeta-Treponema-Borellia cluster than they are to other nonspirochetes, except for the presence of certain signature sequences found in all spirochete 16S RNAs but not in 16S RNAs of other bacteria (24).

ACKNOWLEDGMENTS

We thank Charles Yanofsky for critically reading the manuscript.
This work was supported by grants from the National Science Foundation (DMB 8606653 to I.P.C. and DCB 8913528 to E.P.G.) and by Public Health Service grant AI 23275 from the National Institutes of Health (to E.P.G.).

REFERENCES

1. **Brahamsha, B., and E. P. Greenberg.** 1987. Complementation of a *trpE* deletion in *Escherichia coli* by *Spirochaeta aurantia* DNA encoding anthranilate synthetase component I activity. J. Bacteriol. **169**:3764–3769.
2. **Brahamsha, B., and E. P. Greenberg.** 1988. A biochemical and cytological analysis of the complex periplasmic flagella from *Spirochaeta aurantia*. J. Bacteriol. **170**:4023–4032.
3. **Brahamsha, B., and E. P. Greenberg.** 1989. Cloning and sequence analysis of *flaA*, a gene encoding a *Spirochaeta aurantia* flagellar filament surface antigen. J. Bacteriol. **171**:1692–1697.
4. **Breznak, J. A., and E. Canale-Parola.** 1975. Morphology and physiology of *Spirochaeta aurantia* strains isolated from aquatic habitats. Arch. Microbiol. **105**:1–12.
5. **Crawford, I. P.** 1975. Gene rearrangements in the evolution of the tryptophan synthetic pathway. Bacteriol. Rev. **39**:87–120.
6. **Crawford, I. P.** 1980. Comparative studies on the regulation of tryptophan synthesis. Crit. Rev. Biochem. **6**:175–189.
7. **Crawford, I. P.** 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu. Rev. Microbiol. **43**:567–600.
8. **Doolittle, R. F.** 1986. Of urfs and orfs: a primer on how to analyze derived amino acid sequences, p. 12–15. University Science Books, Mill Valley, Calif.
9. **Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure.** 1986. Relationships of human protein sequences to those of other organisms. Cold Spring Harbor Symp. Quant. Biol. **51**:447–456.
10. **Essar, D. W., L. Eberly, and I. P. Crawford.** 1990. Evolutionary differences in chromosome locations of four early genes of the tryptophan pathway in fluorescent pseudomonads: DNA sequences and characterization of *Pseudomonas putida trpE* and *trpGDC*. J. Bacteriol. **172**:867–883.
11. **Essar, D. W., L. Eberly, A. Hadero, and I. P. Crawford.** 1990. Identification and characterization of genes for second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. J. Bacteriol. **172**:884–900.
12. **Essar, D. W., L. Eberly, C.-Y. Han, and I. P. Crawford.** 1990. DNA sequences and characterization of four early genes of the tryptophan pathway in *Pseudomonas aeruginosa*. J. Bacteriol. **172**:853–866.
13. **Ferdows, M. S., and A. G. Barbour.** 1989. Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. Proc. Natl. Acad. Sci. USA **86**:5969–5973.
14. **Fosnaugh, K., and E. P. Greenberg.** 1988. Motility and chemotaxis of *Spirochaeta aurantia*: computer-assisted motion analysis. J. Bacteriol. **170**:1768–1774.
15. **Fosnaugh, K., and E. P. Greenberg.** 1989. Chemotaxis mutants of *Spirochaeta aurantia*. J. Bacteriol. **171**:606–611.
16. **Greenberg, E. P., B. Brahamsha, and K. Fosnaugh.** 1985. The motile behavior of *Spirochaeta aurantia*: a twist in chemosensory transduction in bacteria, p. 107–117. *In* M. Eisenbach and M. Balban (ed.), Sensing and response in microorganisms. Elsevier Science Publishers, Amsterdam.
17. **Greenberg, E. P., and E. Canale-Parola.** 1977. Chemotaxis in *Spirochaeta aurantia*. J. Bacteriol. **130**:485–494.
18. **Harwood, C. S., and E. Canale-Parola.** 1984. Ecology of spirochetes. Annu. Rev. Microbiol. **38**:161–192.
19. **Howe, T. R., L. W. Mayer, and A. G. Barbour.** 1985. A single recombinant plasmid expressing two major outer surface proteins of the Lyme disease spirochete. Science **227**:645–646.
20. **Jackson, E. N., and C. Yanofsky.** 1974. Localization of two functions of the phosphoribosyl anthranilate transferase of *Escherichia coli* to distinct regions of the polypeptide chain. J. Bacteriol. **117**:502–508.
21. **Lagrimini, L. M., S. T. Brentano, and J. E. Donelson.** 1984. A DNA sequence analysis package for the IBM personal computer. Nucleic Acids Res. **12**:605–614.
22. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
23. **Pallesen, L., and P. Hindersson.** 1989. Cloning and sequencing of a *Treponema pallidum* gene encoding a 31.3-kilodalton endoflagellar subunit (FlaB2). Infect. Immun. **57**:2166–2172.
24. **Paster, B. J., E. Stackebrandt, R. B. Hespell, C. M. Hahn, and C. R. Woese.** 1984. The phylogeny of spirochetes. Syst. Appl. Microbiol. **5**:337–351.
25. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
26. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
27. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
27a. **Woese, C.** Personal communication.
28. **Yanofsky, C.** 1984. Comparison of regulatory and structural regions of genes of tryptophan metabolism. Mol. Biol. Evol. **1**:143–161.
29. **Yelton, D. B., and N. W. Charon.** 1984. Cloning of a gene required for tryptophan biosynthesis from *Leptospira biflexa* serovar *patoc* into *Escherichia coli*. Gene **28**:147–152.
30. **Yelton, D. B., and R. A. Cohen.** 1986. Analysis of cloned DNA from *Leptospira biflexa* serovar *patoc* which complements a deletion of the *Escherichia coli trpE* gene. J. Bacteriol. **165**:41–46.
31. **Yelton, D. B., and S. L. Peng.** 1989. Identification and nucleotide sequence of the *Leptospira biflexa* serovar *patoc trpE* and *trpG* genes. J. Bacteriol. **171**:2083–2089.