JOURNAL OF BACTERIOLOGY, Apr. 1989, p. 2083–2089
0021-9193/89/042083-07$02.00/0

Vol. 171, No. 4

Identification and Nucleotide Sequence of the *Leptospira biflexa* Serovar *patoc trpE* and *trpG* Genes

DAVID B. YELTON* AND STANFORD L. PENG

Department of Microbiology and Immunology, 2095 Health Science Center North, West Virginia University, Morgantown, West Virginia 26506

Received 22 August 1988/Accepted 13 January 1989

***Leptospira biflexa* is a representative of an evolutionarily distinct group of eubacteria. In order to better understand the genetic organization and gene regulatory mechanisms of this species, we have chosen to study the genes required for tryptophan biosynthesis in this bacterium. The nucleotide sequence of the region of the *L. biflexa* serovar *patoc* chromosome encoding the *trpE* and *trpG* genes has been determined. Four open reading frames (ORFs) were identified in this region, but only three ORFs were translated into proteins when the cloned genes were introduced into *Escherichia coli*. Analysis of the predicted amino acid sequences of the proteins encoded by the ORFs allowed us to identify the *trpE* and *trpG* genes of *L. biflexa*. Enzyme assays confirmed the identity of these two ORFs. Anthranilate synthase from *L. biflexa* was found to be subject to feedback inhibition by tryptophan. Codon usage analysis showed that there was a bias in *L. biflexa* towards the use of codons rich in A and T, as would be expected from its G+C content of 37%. Comparison of the amino acid sequences of the *trpE* gene product and the *trpG* gene product with corresponding gene products from other bacteria showed regions of highly conserved sequence.**

Tryptophan biosynthesis has been examined in a variety of microorganisms (4, 33). The organization and control of the genes encoding the enzymes required for tryptophan biosynthesis have also been examined in several bacterial genera (4, 33). The tryptophan operon of *Escherichia coli* is often used as a model for gene regulation involving repression and attenuation (33). Because of the genetic diversity of bacterial species and their apparently rapid rates of evolution, it is of interest to examine genes for tryptophan biosynthesis in a variety of bacterial genera to uncover those rules which may govern genetic organization and control.

Oligonucleotide analysis of 16S rRNA has revealed that the spirochetes represent a distinct eubacterial group which diverged from other eubacteria at an early time (6, 32). *Leptospira* is one genus within the spirochete group (22). Its oligonucleotide pattern is distinct and has an S_{AB} value of <0.2 (22). *Leptospira biflexa* has a unique metabolism which uses fatty acids or alcohols as its sole source of carbon and energy (12). Its distinct mode of motility allows it to both swim freely in liquid media and move in a corkscrew-like manner through semisolid medium (2, 7).

Our laboratory has been interested in understanding the genetics of these microorganisms. Because classical genetic exchange systems presently are not available for *L. biflexa*, we have used the techniques of gene cloning to begin a study into the genetics of this bacterium. Previously we have cloned genes from *L. biflexa* which complement deletions in the *trpE* and *trp(G)D* genes of *E. coli* and have demonstrated that these genes reside on a 4-kilobase fragment of leptospiral DNA (35, 36). Three proteins (22, 53.6, and 23.5 kilodaltons [kDa]) are produced from this DNA fragment when it is introduced into *E. coli* (36).

Transposon mutagenesis of this cloned DNA has shown that in *E. coli* a single transcriptional unit encodes the 22- and 53.6-kDa proteins, whereas a second unit encodes the 23.5-kDa protein (36). Insertion of a transposon into the gene encoding any one of these proteins prevents complementation of a *trpEG* deletion in *E. coli* (36). The present study was undertaken to determine the nucleotide sequence of the gene encoding each of these three proteins, to identify the *trpE* and *trpG* genes of *L. biflexa*, and to compare the deduced amino acid sequences of the leptospiral *trpE* and *trpG* gene products with similar proteins from other bacteria.

MATERIALS AND METHODS

Bacterial strains, bacteriophage, and plasmids. *E. coli* JA221 Δ*trpE5 leuB6 hsdR* (*hsdR hsdM*$^+$) was used for preparation of plasmid DNA (3). *E. coli* JM103 [Δ(*lac-pro*) *thi strA supE endA sbcB* (F′ *traD proAB lacI*^{q} *lacZ* ΔM15)] was used to propagate bacteriophage M13 and to produce replicative-form DNA from M13 (20). Bacteriophage M13 and clonal deletions derived from it were used as sources of DNA for nucleotide sequencing reactions (20). Plasmid pYC6 (36) and derivatives of it produced in the pGEM plasmid (Promega) were also used for nucleotide sequencing reactions. Clonal derivatives of M13 and pGEM were produced by cloning restriction fragments of pYC6 into the appropriate sites of the polylinkers present on these DNA molecules. Additional subclones were produced by creating nested deletions with exonuclease III and mung bean nuclease (11).

Media. *E. coli* strains were grown in L broth or on L agar (1.5%) at 37°C (21). M13 bacteriophage infections were done in YT medium at 37°C (17). Medium B was prepared as described previously (38).

Enzyme assays. Anthranilate synthase (AS) was assayed with a spectrofluorimeter as described previously (14, 24, 28). The reaction was carried out at 37°C in either phosphate buffer (pH 7.5) with glutamine or in Tris buffer (pH 8.7) with NH_4Cl.

DNA preparation. Plasmid DNA was extracted from 100-ml samples of overnight cultures of *E. coli* (17). Replicative-form DNA from bacteriophage M13 was prepared from 100-ml samples of late-logarithmic-phase cultures of *E. coli* JM103 infected with the bacteriophage by the same protocol as was used for plasmid isolation. All supercoiled DNA

* Corresponding author.

preparations were purified through two cycles of equilibrium centrifugation in CsCl-ethidium bromide. Single-stranded M13 DNA was extracted from purified bacteriophage particles as described previously (20).

Materials. Restriction endonucleases and M13 replicative-form DNA were obtained from Bethesda Research Laboratories, Inc. T4 DNA ligase was obtained from New England BioLabs, Inc. Calf intestinal alkaline phosphatase, exonuclease III, and M13 sequencing primer were purchased from Boehringer Mannheim Biochemicals. Additional sequencing primers were produced in the West Virginia University Recombinant DNA Facility with an Applied Biosystems model 380B DNA synthesizer. Mung bean nuclease and proteinase K were obtained from Pharmacia. DNA sequencing kits were purchased from Promega (GemSeq K/RT) and U.S. Biochemical Corp. (Sequenase). Radiolabeled [^{35}S]dATP (1,000 Ci/mmol) was obtained from Amersham Corp.

Nucleotide sequencing. All sequencing reactions were performed by using the dideoxy-chain termination method of Sanger et al. (23). Single-stranded M13 clonal DNA was sequenced with a modified T7 DNA polymerase (Sequenase) and dITP as recommended by the manufacturer. Double-stranded plasmid DNAs were sequenced by using Klenow fragment polymerase and dGTP. The products of the sequencing reactions were resolved on a 6 or 8% polyacrylamide gel (0.4 mm by 20 cm by 40 cm). Electrophoresis was carried out at 45 mA constant current. Gels were double or triple loaded. After electrophoresis, the gel was fixed with 5% methanol–5% acetic acid, dried under vacuum at 80°C, and exposed to X-ray film for 40 to 72 h.

Sequence analysis. Analysis of sequence data was performed by using the DNASTAR system maintained in the West Virginia University Recombinant DNA Facility.

RESULTS

Assay of AS from *L. biflexa*. AS consists of two subunits, ASI and ASII. ASI catalyzes the production of anthranilic acid from chorismic acid; ASII acts as an amidotransferase. Mutants of *E. coli* defective in ASII can grow in medium B containing $(NH_4)_2SO_4$ if the pH of the medium is adjusted to 8 (38). At this pH, the free ammonia which is available allows ASI to convert chorismic acid to anthranilic acid without the amidotransferase activity of ASII (38). At pH 6, the amidotransferase activity of ASII is required, as free ammonia is not available (38). We used medium B at both pHs to identify the genes encoding the ASI and ASII subunits of *L. biflexa*. We also used a spectrofluorimetric assay to directly measure enzyme activity. The results of these experiments are shown in Table 1. These results demonstrate that ASI of *L. biflexa* is able to utilize ammonia to convert chorismic acid to anthranilic acid. They also show that the mutant identified as Tn*177* has an intact ASI subunit but a defective ASII subunit and that mutant Tn*158* has a defective ASI subunit. We have reported that Tn*177* is located in the 23-kDa protein and that Tn*158* is located in the 53.5-kDa protein (36).

AS from other bacteria is sensitive to feedback inhibition by tryptophan (37). Feedback inhibition has been reported for the isoleucine, leucine, and valine pathways of *L. biflexa* (31). To determine whether AS from *L. biflexa* is subject to feedback inhibition, we measured the activity of the cloned leptospiral enzyme in the presence of various concentrations of tryptophan. At 5×10^{-6} M tryptophan, 50% of the enzyme activity was inhibited; 90% inhibition occurred at 3×10^{-5} M. AS from *E. coli* was similarly sensitive to inhibition by tryptophan.

TABLE 1. AS activities in strains of *E. coli* containing a *trpED* deletion and carrying cloned *Leptospira* genes

E. coli strain (plasmid)	Specific activity[a]		Growth[b] on minimal medium at pH:	
	Gln	NH_3	6	8
W3110 Δ*trpED23*	0.00	0.00	0.0	0.0
W3110 Δ*trpED23*(pYC6)	0.95	25.70	0.9	0.7
W3110 Δ*trpED23*(pYC6-Tn*158*)	0.00	0.00	0.0	0.0
W3110 Δ*trpED23*(pYC6-Tn*177*)	0.00	13.30	0.0	0.4
Wild type	0.83	0.32	0.8	1.2

[a] Specific activity is measured here in nanomoles of anthranilate synthesized per minute per milligram of protein with either glutamine (Gln) or ammonia (NH_3) as the amino donor. The enzyme preparations from the deletion strains were partially purified before the assays were performed.

[b] Growth was monitored by measuring the optical densities of the cultures at 600 nm at 24 h postinoculation. The cultures were incubated on a shaker at 37°C. Initial optical density readings were 0.0 in all instances.

Sequencing strategy. A partial restriction endonuclease cleavage map and the sequencing strategy employed for the cloned *trpEG* region (3,480 base pairs) of *L. biflexa* is shown in Fig. 1. Both strands of the DNA molecule were sequenced essentially in their entirety. Four major open reading frames (ORFs) were identified and are located in Fig. 1. Three ORFs correspond to proteins of 22, 53.5, and 23.6 kDa that are produced in *E. coli* maxicells (36). No evidence was found to indicate that ORF-4 was translated into protein in *E. coli* maxicells. In addition, 242 base pairs of 5′-flanking DNA and 60 base pairs of 3′-flanking DNA were sequenced.

Identification of encoded proteins. The nucleotide sequence of the region is shown in Fig. 2. The deduced amino acid sequences of the three ORFs translated in *E. coli* are also shown. ORF-1 (5′-proximal end of the cloned DNA fragment) encodes a protein of 21,584 Da. This protein is highly basic and shares weak amino acid homologies with many DNA-binding proteins when compared with all the sequences present in the protein identification resource data bank (27). ORF-2 encodes a protein of 51,775 Da. This

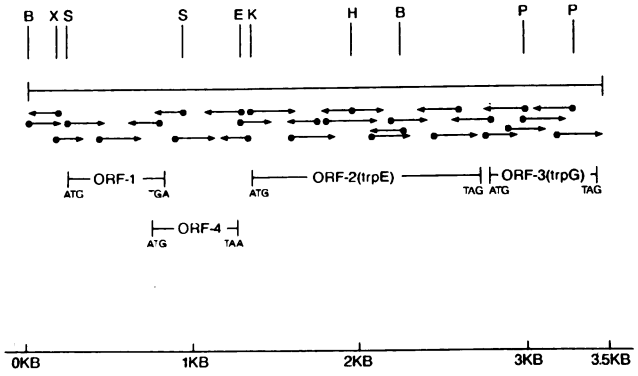

FIG. 1. Sequencing strategy. The region of cloned *L. biflexa* DNA which was sequenced is shown. Restriction sites are indicated. The sequencing strategy is indicated by the arrows. The location of the four ORFs is diagrammed. ORF-1 corresponds to the 22-kDa protein, ORF-2 (*trpE*) corresponds to the 53.5-kDa protein, and ORF-3 (*trpG*) corresponds to the 23.6-kDa protein (27). ORF-4 apparently is not translated in *E. coli*. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; S, *Sst*I; X, *Xba*I; KB, kilobase.

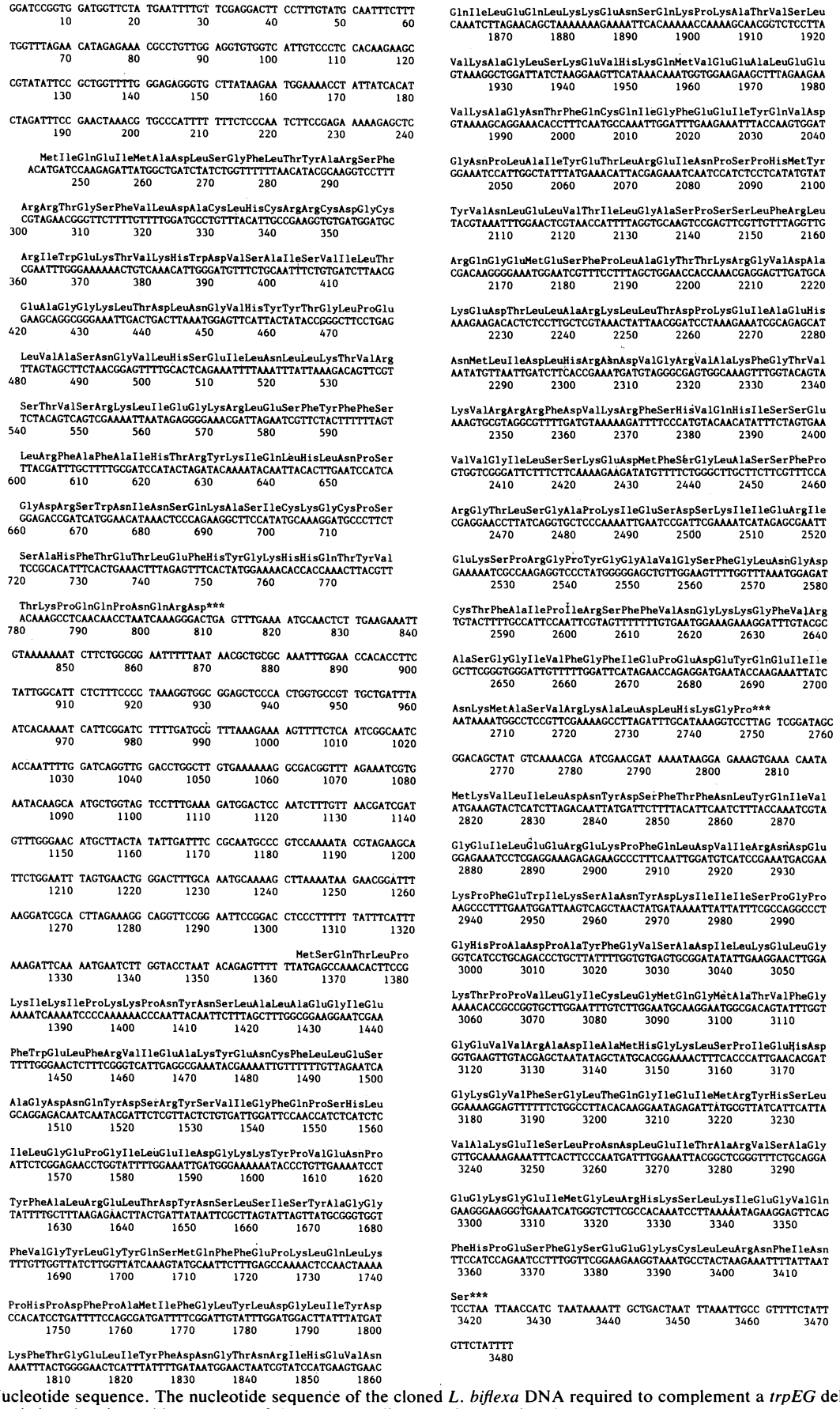

FIG. 2. Nucleotide sequence. The nucleotide sequence of the cloned *L. biflexa* DNA required to complement a *trpEG* deletion in *E. coli* is shown. The deduced amino acid sequences of the corresponding proteins are also shown.

```
E. coli     1  MQTQKPTLELLTCEGAYRDNPTALF  25
L. biflexa  1  MS-Q--TLPKIKIPKKPNYNSLALA  22

 26  HQLCGDRPATL----------LLESADIDSKD  47
 23  EGIEFWELFRVIEAKYENCFLLESAGDNQYD  53

 48  DLKSLLLVDSALRITALGDTVTIQALSGNGE  78
 54  S----------------------------  54

 79  ALLALLDNALPAGVESEQSPNCRVLRFPPVS  109
 55  ----------------RYSVIGFQPSHLILGE  70

110  PLLDEDARLCSLSVFDAFRLLQNLLNVPKEE  140
 71  PGILEIDGKKYPVENPYFALRELTDYNSLSI  101

141  REAMFFSGLFSYDLVAGFEDLPQLSAENNCP  171
102  SYAGGFVGYLGYQSMQFFEPKLQLKPH---P  129

172  DF--CFYLAETLMVIDHQKKSTRIQASLFA  199
130  DFPAMIFGLYLDGLIYD--KFTGEL--IYFD  156

200  PNEEEKQRLTARLNELRQQLTEAAPPLPVVS  230
157  NGTNRIHEVNQILEQLKKE--NSQKPKATVS  185

231  VPHMRCECNQSDEEFGGVVRLLQKAIRAGEI  261
186  L--VKAG--LSKEVHKQMVEEALEEVKAGNT  212

262  FQ-VVPSRRFSLPCPSPLAAYYVLKKSNPSP  291
213  FQCQIGFEEIYQVDGNPLAIYETLREINPSP  243

292  YMFFMQDNDFTLFGASPESSLKYDATSRQIE  322
244  HMYYVNLELVTILGASP-SSL-FRLRQGEME  272

323  IYPIAGTRPRGRRADGSLDRDLDSRIELEMR  353
273  SFPLAGT---TKRGVDAKEDTLLAR---KLL  297

354  TDHKELSEHLMLVDLARNDLARICTPGSRYV  384
298  TDPKEIAEHNMLIDLHRNDVGRVAKFGTVKV  328

385  ADLTKVDRYSYVMHLVSRVVGELRHDLDALH  415
329  RRRFDVKRFSHVQHISSEVVGILSSKEDMFS  359

416  AYRACMNMGTLSGAPKV-RAMQLIAEAEGRR  445
360  GLASSFPRGTLSGAPKIESDSKIIERIEKSP  390

446  RGSYGGAVGYFTAHGDLDTCIVIRSALVENG  476
391  RGPYGGAVGSFGLNGDCTFAIPIRSFFVNGK  421

477  IATVQAGAGVVLDSV-PQSEADETRNKARAV  506
422  KGFVRASGGIVFGFIEPEDEYQEIINKMASV  452

507  LRAIATAHHAQETF*  520
453  RKALDLHKGP*  462
```

FIG. 3. Alignment of the amino acid sequences of ASI from *E. coli* and *L. biflexa*. The amino acid sequences of these two ASI proteins were aligned by using the DNASTAR system. To optimize the alignment, gaps were introduced where necessary. Conservation of amino acid sequence is indicated by overlining.

protein has extensive amino acid similarity (26 to 30% identity and about 25% substitution of related amino acids) with ASI from several bacteria (27) and has the corresponding enzymatic activity. Therefore we designate this ORF as the *trpE* gene of *L. biflexa*. ORF-3 encodes a protein of 22,132 Da. This protein has extensive amino acid similarity (36 to 49% identity and approximately 26% substitution of related amino acids) with ASII and with *p*-aminobenzoate synthase component II of several bacteria (27). It acts as a glutamine amidotransferase. Because this ORF is immediately downstream from *trpE* and has the appropriate amino acid sequence and enzymatic activity, we designate it as the *trpG* gene of *L. biflexa*. The molecular masses predicted for these three proteins on the basis of the nucleotide sequence data are similar to those which we determined previously by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (36).

Amino acid similarities of ASI and ASII from *L. biflexa* and *E. coli*. An alignment of the amino acid sequences of ASI from *L. biflexa* and *E. coli* is shown in Fig. 3. Where necessary to maximize alignments, gaps have been introduced in the sequences. Of the amino acids in the two ASI proteins, 29% are identical and 19% are substitutions of similar amino acids. Most of the conserved sequence is found in the COOH end of the molecules.

```
E. coli     1  MADILLLDNIDSFTYNL---ADQL  21
L. biflexa  1  MK-VLILDNYDSFTFNLYQIVGEI  23

 22  RSNGHNV----VIYRNHIPAQTLIERLATMS  48
 24  LEEREKPFQLDVI-RND---EKPFEWIKSAN  50

 49  NPVLMLSPGPGVPSEAGC---MPELLTRLRG  76
 51  YDKIIISPGPGHPADPAYFGVSADILKELGK  81

 77  KLPIIGICLGHQAIVEAYGGYVGQAGEILHG  107
 82  TPPVLGICLGMQGMATVFGGEVVRANIAMHG  112

108  KASSIEHDGQAMFAGLTNPLPVARYHSLVGS  138
113  KLSPIEHDGKGVFSGLTQGIEIMRYHSLVAK  143

139  NI--PAGLTINAHF-----NGMVMAVRHDAD  162
144  EISLPNDLEITARVSAGEGKGEIMGLRHKSL  174

163  RVCGFQFHPESILTTQGARLLEQTLAWAQHK  193
175  KIEGVQFHPESFGSEEGKCLLRNFINS*  201

194  LEPANTLQP...  202
```

FIG. 4. Alignment of the amino acid sequences of ASII from *E. coli* and *L. biflexa*. The amino acid sequences of these two ASII proteins were aligned by using the DNASTAR system. To optimize the alignment, gaps were introduced where necessary. Conservation of amino acid sequence is indicated by overlining.

An alignment of the amino acid sequences of the ASII molecules is shown in Fig. 4. Gaps were introduced to maximize alignments. Conservation of amino acid sequence is seen throughout the molecules. Of the amino acids in the two proteins, 36% are identical and 28% are substitutions of related amino acids.

Translation initiation regions. The nucleotide sequences of the regions surrounding the presumed sites for translation initiation are shown in Fig. 5. At 5 nucleotides preceding the initiation codon for ORF-1 is the potential Shine-Dalgarno sequence, AAAGAG (26). At 6 nucleotides preceding the initiation codon of the *trpE* gene is a potential Shine-Dalgarno sequence, AGAG, while at 7 bases before the initiation codon for *trpG* is the potential Shine-Dalgarno sequence, TAAGGAGAAAGTG. The genes for *trpE* and *trpG* are each preceded by the sequence TCAAAAYGA ATC. This sequence does not precede either ORF-1 or ORF-4. Between the termination codon for *trpE* and the initiation codon for *trpG* are 64 apparently untranslated nucleotides (Fig. 2). This differs from many other bacteria in which the coding regions for these proteins overlap (5). There are no methionine codons in the 60 nucleotides after the *trpG* termination codon, which again suggests an untranslated region preceding the next ORF.

Codon usage. The codons used in *L. biflexa* for *trpE* and *trpG* peptides are shown in Table 2 and compared with those used in *E. coli* and *Bacillus subtilis* for *trp* genes. Sometimes codons similar to those used in *B. subtilis* are used in *L. biflexa*, as in the case of histidine and phenylalanine. *E. coli* codon patterns for lysine are used in *L. biflexa*. There are also examples of codon usage patterns which are distinct from either *E. coli* or *B. subtilis*, such as with glutamine and proline. Analysis of the nucleotide occupying position 3 of a

```
                                                MetIleGlnGluIle
ORF-1:  ACTAAACGTGCCATTTTTTTTCTCCCAATCTTCCGAGAAAAAGAGCTCACATGATCCAAGAGATT

                                                MetSerGlnThrLeu
trpE:   TTCATTTAAAGATTCAAAATGAATCTTGGTACCTAATACAGAGTTTTTTATGAGCCAAACACTT

                                                MetGluSerThrHis
trpG:   CTATGTCAAAACGAATCGAACGATAAAATAAGGAGAAAGTGAAACAATAATGGAAAGTACTCAT
```

FIG. 5. Translation initiation regions of *L. biflexa*. Fifty nucleotides upstream and twelve downstream from the presumed translation initiation codons of the cloned genes are shown. Potential Shine-Dalgarno sequences are underlined. A potential regulatory sequence is overlined.

TABLE 2. Codons used for *trp* genes of *L. biflexa* serovar *patoc* compared with the *trp* genes of *B. subtilis* and *E. coli*

Amino acid	Codon	No. (%) of codons		
		L. biflexa trpE and *trpG*	*B. subtilis* total *trp* genes	*E. coli* total *trp* genes
Ala	GCA	10 (29)	54 (33)	44 (17)
Ala	GCG	6 (17)	39 (24)	91 (35)
Ala	GCC	3 (9)	27 (17)	86 (33)
Ala	GCT	16 (45)	42 (26)	37 (14)
Arg	AGA	5 (18)	18 (25)	3 (3)
Arg	AGG	2 (7)	5 (7)	1 (1)
Arg	CGA	10 (36)	10 (14)	5 (4)
Arg	CGG	2 (7)	15 (21)	3 (3)
Arg	CGC	2 (7)	13 (16)	67 (57)
Arg	CGT	7 (25)	13 (18)	39 (33)
Asn	AAC	3 (11)	26 (39)	48 (60)
Asn	AAT	24 (89)	41 (61)	32 (40)
Asp	GAC	5 (17)	38 (36)	43 (37)
Asp	GAT	24 (83)	67 (64)	73 (63)
Cys	TGC	2 (40)	4 (22)	18 (56)
Cys	TGT	3 (60)	14 (78)	14 (44)
Gln	CAA	19 (90)	35 (49)	38 (37)
Gln	CAG	2 (10)	36 (51)	65 (63)
Glu	GAA	49 (86)	107 (61)	96 (72)
Glu	GAG	8 (14)	67 (39)	38 (28)
Gly	GGA	31 (50)	40 (28)	15 (9)
Gly	GGG	9 (15)	26 (18)	22 (13)
Gly	GGC	2 (3)	39 (27)	68 (41)
Gly	GGT	20 (32)	39 (27)	62 (37)
His	CAC	4 (25)	13 (28)	31 (49)
His	CAT	12 (75)	33 (72)	32 (51)
Ile	ATA	6 (12)	26 (19)	2 (2)
Ile	ATC	12 (24)	40 (30)	43 (38)
Ile	ATT	32 (64)	69 (51)	67 (60)
Leu	CTA	5 (8)	11 (6)	11 (4)
Leu	CTG	0 (0)	36 (20)	133 (54)
Leu	CTC	9 (13)	21 (12)	24 (10)
Leu	CTT	17 (28)	48 (27)	20 (8)
Leu	TTA	18 (29)	44 (24)	30 (12)
Leu	TTG	14 (22)	26 (14)	30 (12)
Lys	AAA	39 (81)	96 (71)	64 (82)
Lys	AAG	9 (19)	40 (29)	14 (18)
Met	ATG	15 (100)	56 (100)	50 (100)
Phe	TTC	6 (16)	20 (25)	37 (49)
Phe	TTT	32 (84)	58 (75)	39 (51)
Pro	CCA	13 (36)	7 (9)	18 (17)
Pro	CCG	4 (11)	41 (52)	63 (58)
Pro	CCC	8 (22)	8 (10)	15 (14)
Pro	CCT	11 (31)	23 (29)	12 (11)
Ser	AGC	1 (2)	37 (26)	33 (27)
Ser	AGT	11 (22)	7 (5)	17 (14)
Ser	TCA	8 (16)	32 (23)	12 (10)
Ser	TCG	9 (18)	15 (11)	21 (17)
Ser	TCC	7 (14)	13 (9)	21 (17)
Ser	TCT	13 (28)	37 (26)	17 (14)
Thr	ACA	7 (35)	49 (45)	13 (13)
Thr	ACG	3 (15)	23 (21)	22 (23)
Thr	ACC	5 (25)	15 (13)	50 (52)
Thr	ACT	5 (25)	23 (21)	12 (12)
Trp	TGG	2 (100)	7 (100)	5 (100)
Tyr	TAC	10 (40)	23 (38)	23 (37)
Tyr	TAT	15 (60)	38 (62)	39 (63)
Val	GTA	13 (33)	26 (19)	18 (13)
Val	GTG	10 (25)	36 (26)	59 (43)
Val	GTC	4 (11)	28 (20)	28 (20)
Val	GTT	12 (31)	46 (34)	33 (24)

codon revealed a preference for A or T (Table 3). Since *L. biflexa* is only 37% G+C, this preference for an A or T in position 3 is not surprising and is consistent with what is seen with other bacteria in which the nucleotide in position 3 of a codon reflects the overall base composition of the organism (5, 13, 15, 19).

DISCUSSION

The gene order in *L. biflexa*, *trpE* preceding *trpG*, is similar to what is found in many other bacteria. Unlike some other systems, there is no overlap of these two genes (4, 5). Sequence analysis of the *trpE* and *trpG* gene products of *L. biflexa* revealed that they share amino acid homologies with similar proteins from other bacteria. The regions of conserved amino acid sequences are found in the same places in which other bacteria show conservation of amino acid sequences (Fig. 3 and 4). These results suggest that these regions are essential for the proper functioning of the enzymes.

TABLE 3. Nucleotides at position 3 in codons of *L. biflexa trp* genes

Nucleotide in position 3	% Use of nucleotide			
	ORF-1	*trpE*	*trpG*	Average
A	34	35	34	35
T	33	38	37	37
G	15	15	16	15
C	18	12	13	13
A or T	67	73	71	71
G or C	33	27	29	29

The *trpG* proteins of other bacteria have a cysteine residue at position 85 and either a lysine or an arginine residue at position 109; these two amino acids are part of the active site of the enzyme (5). A similar arrangement is found in *Saccharomyces cerevisiae* (39). *L. biflexa* has a cysteine residue at position 89 and a lysine residue at position 113 (Fig. 4). The spacing of these amino acids is such that they could function as part of the active site for the leptospiral enzyme. ASII from *L. biflexa* also shows conservation of the amino acid sequences -DXXDSF- (positions 7 to 12), -PGP- (58 to 60), -HG- (111 and 112), -RYHSL- (136 to 140), and -FHPES- (181 to 185). Conservation of these amino acid sequences is found in all other ASIIs which have been examined (15).

ASI from *L. biflexa* also shows conservation of amino acid sequence when compared with ASIs from other organisms. Most of this conservation occurs in the COOH end of the molecule, as has been noted previously with other organisms (8, 9). An extensive gap was introduced into the amino-terminal end of the leptospiral protein to maximize its alignment with the corresponding *E. coli* protein. A deletion presumably occurred in the corresponding region of the leptospiral gene during its evolution. A histidine residue (His-342) may be equivalent to the conserved histidine essential for enzymatic activity (9, 30). Similar to other organisms, several conserved arginine residues are also found in the COOH end of the molecule. In *L. biflexa*, these arginine residues occur at positions 282, 314, 319, 391, and 414. Strict analogy with other organisms would dictate the presence of another conserved arginine (9). This arginine would be found at position 377, but in *L. biflexa*, it is replaced with a glutamate. Presumably one of the five conserved arginines is the essential arginine required for enzymatic activity (9, 30). The essential cysteine (30) is missing and is replaced by alanine. A similar replacement of cysteine with alanine occurs in *B. subtilis* (9). Near the NH_2 terminus of the molecule, the sequence -LLES- (43 to 46) is found. This sequence has been shown to be important for feedback inhibition by tryptophan in *Brevibacterium lactofermentum* (18); it is also found in ASIs from other organisms which are feedback inhibited by tryptophan (9).

Analysis of the codons used by *L. biflexa* revealed a preference for codons rich in A or T which can be correlated with the 37% G+C content of the organism. For instance, in *L. biflexa*, TTA, CTT, or TTG is used for leucine, rather than the CTG codon preferred by other bacteria (10). In addition, TTT is used 84% of the time as a phenylalanine codon in *L. biflexa*, whereas most other bacteria use TTC as often as TTT. In *B. subtilis*, which is 43% G+C, AAT is used for asparagine as in *L. biflexa*; however, in *B. subtilis*, CAG is used for glutamine, while CAA is used in *L. biflexa*. The most striking example of the preference for A or T in codons of *L. biflexa*, however, is found in proline codons. All other bacteria which have been examined use CCG overwhelmingly as their proline codon; *L. biflexa* uses either CCA or CCT. Yeasts, on the other hand, use TTA or TTG for leucine, CAA for glutamine, and CCA or CCT for proline, which are the same as those used by *L. biflexa* (10).

The genes encoding *p*-aminobenzoate synthase and AS are thought to have evolved from common precursors (8, 9). The *trpE* and *trpG* genes of *E. coli* contain no codons for tryptophan, while the *pabB* and *pabA* genes of *E. coli* contain seven and three tryptophan codons, respectively (8). *L. biflexa* contains a single Trp codon in both its *trpE* and *trpG* genes. Thus, it appears as though the genes encoding AS in *L. biflexa* are intermediate in the evolutionary process.

ORF-1 is expressed in *E. coli* and encodes a protein with DNA-binding capabilities. The positioning of this gene 5′ to *trpE* is the same as the positioning of the gene encoding topoisomerase A in *E. coli*, although the *topA* gene is about 6 kilobases upstream from *trpE* (1). The *topA* protein of *E. coli* is about 100,000 Da (29), while ORF-1 encodes a protein of only 21,500 Da. Comparison of the amino acid sequences of these proteins revealed little similarity. Thus, it seems unlikely that ORF-1 encodes a topoisomerase activity. Whether the protein encoded by ORF-1 plays a regulatory role in *trp* gene expression in *L. biflexa* awaits further study.

Only a few regions within the sequenced *L. biflexa* DNA have the potential to form stem-loop structures. Potential stem-loop structures are centered at nucleotides 577.5, 1412, 1759.5, 2202.5, 2829, 3116, and 3422. Most of these are not located at positions where they might serve as regulators of the transcription process. However, one (centered on nucleotide 3422) is located immediately downstream of the *trpG* gene where it might serve as a transcription terminator. Whether it plays this role in vivo awaits analysis of the mRNAs produced in *L. biflexa*.

The lack of potential tertiary structure of the nucleotide sequence preceding the *trpE* gene of *L. biflexa* is unique. There is no evidence of a classical attenuator region preceding the *trpE* gene. Such regions are found in *E. coli* and related enteric bacteria (34) and in *Brevibacterium lactofermentum* (19). *B. subtilis* utilizes a different attenuation mechanism in which a regulatory protein in the presence of tryptophan binds to the nascent transcript and stops further transcription of the structural genes (16, 25). This nascent transcript is characterized by several inverted repeats that are necessary for the regulatory process (25). Again there is no evidence of such a regulatory region immediately preceding either the *trpE* or *trpG* genes of *L. biflexa*.

A 12-nucleotide repeat that might be involved in regulation does precede both the *trpE* and *trpG* genes of *L. biflexa*. The nucleotide sequence TCAAAAYGAATC precedes the initiation codon of *trpE* by 24 bases and the initiation codon of *trpG* by 32 bases. This sequence is not found before ORF-1 or ORF-4. Whether this sequence is involved in the regulation of expression of the *trp* genes of *L. biflexa* awaits further study. A similar sequence, TCAAAAT, precedes the first Met codon of the *argE* gene of *L. biflexa* by 52 nucleotides, where it is thought to form part of the promoter (40).

ORF-4 apparently is not expressed in *E. coli* (36). If expressed, it would encode a protein of 18,067 Da. Comparison of the amino acid sequence of this protein to those of proteins in the protein data bank revealed a strong similarity (41% identical amino acids and 21% related amino acids) to superoxide dismutase from *B. subtilis* spp. (27).

ACKNOWLEDGMENTS

We thank Irving Crawford, Nyles Charon, Richard Zuerner, and Richard Stenberg for their helpful comments and criticisms during the course of this work and preparation of the manuscript.

This work was supported by grants from the West Virginia University Dental Corporation, the West Virginia University Medical Corporation, BRSG 5 SO7-RRO5433, and Public Health Service grant 2 R01 DEO4645 from the National Institutes of Health. S. L. Peng was the recipient of an NIH Minority High School Student Research Apprentice Program award.

LITERATURE CITED

1. **Bachmann, B. J.** 1987. Linkage map of *Escherichia coli* K-12, edition 7, p. 807–876. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2.
2. **Berg, H. C., D. B. Bromley, and N. C. Charon.** 1978. Leptospiral motility. Symp. Soc. Gen. Microbiol. **28**:285–294.
3. **Chinault, A. C., and J. Carbon.** 1979. Overlap hybridization screening: isolation and characterization of overlapping DNA fragments surrounding the *leu2* gene on the yeast chromosome, III. Gene **5**:111–126.
4. **Crawford, I. P.** 1980. Comparative studies on the regulation of tryptophan synthesis. Crit. Rev. Biochem. **6**:175–189.
5. **Crawford, I. P., and L. Eberly.** 1986. Structure and regulation of the anthranilate synthase genes in *Pseudomonas aeruginosa*. I. Sequence of *trpG* encoding the glutamine amidotransferase subunit. Mol. Biol. Evol. **3**:436–448.
6. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum,**

L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. M. Chen, and C. R. Woese. 1980. The phylogeny of procaryotes. Science 209:457–463.
7. Goldstein, S. F., and N. C. Charon. 1988. The motility of the spirochete *Leptospira*. Cell Motil. Cytoskeleton 9:101–111.
8. Goncharoff, P., and B. P. Nichols. 1984. Nucleotide sequence of *Escherichia coli pabB* indicates a common evolutionary origin of *p*-aminobenzoate synthetase and anthranilate synthetase. J. Bacteriol. 159:57–62.
9. Goncharoff, P., and B. P. Nichols. 1988. Evolution of aminobenzoate synthases: nucleotide sequences of *Salmonella typhimurium* and *Klebsiella aerogenes pabB*. Mol. Biol. Evol. 5:531–548.
10. 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.
11. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359.
12. Henneberry, R. C., and C. D. Cox. 1970. β-Oxidation of fatty acids by *Leptospira*. Can. J. Microbiol. 16:41–45.
13. Henner, D. J., L. Band, and H. Shimotsu. 1984. Nucleotide sequence of the *Bacillus subtilis* tryptophan operon. Gene 34:169–177.
14. Ito, J., E. C. Cox, and C. Yanofsky. 1969. Anthranilate synthetase, an enzyme specified by the tryptophan operon of *Escherichia coli*: purification and characterization of component I. J. Bacteriol. 97:725–733.
15. Kaplan, J. B., P. Goncharoff, A. M. Seibold, and B. P. Nichols. 1984. Nucleotide sequence of the *Acinetobacter calcoaceticus trpGDC* gene cluster. Mol. Biol. Evol. 1:456–472.
16. Kuroda, M. I., D. Henner, and C. Yanofsky. 1988. *cis*-Acting sites in the transcript of the *Bacillus subtilis trp* operon regulate expression of the operon. J. Bacteriol. 170:3080–3088.
17. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. Matsui, K., K. Miwa, and K. Sano. 1987. Two single-base-pair substitutions causing desensitization to tryptophan feedback inhibition of anthranilate synthase and enhanced expression of tryptophan genes of *Brevibacterium lactofermentum*. J. Bacteriol. 169:5330–5332.
19. Matsui, K., K. Sano, and E. Ohtsubo. 1987. Sequence analysis of the *Brevibacterium lactofermentum trp* operon. Mol. Gen. Genet. 209:299–305.
20. Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101:20–78.
21. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. Paster, B. J., E. Stackebrandt, R. B. Hespell, C. M. Hahn, and C. R. Woese. 1984. The phylogeny of the spirochetes. Syst. Appl. Microbiol. 5:337–351.
23. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
24. Sawula, R. V., and I. P. Crawford. 1973. Anthranilate synthetase of *Acinetobacter calcoaceticus*: separation and partial purification of subunits. J. Biol. Chem. 248:3573–3581.
25. Shimotsu, H., M. I. Kuroda, C. Yanofsky, and D. J. Henner. 1986. Novel form of transcription attenuation regulates expression of the *Bacillus subtilis* tryptophan operon. J. Bacteriol. 166:461–471.
26. Shine, J., and L. Dalgarno. 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71:1342–1346.
27. Sidman, K. E., D. G. George, W. C. Barker, and L. T. Hunt. 1988. The protein identification resource (PIR). Nucleic Acids Res. 16:1869–1871.
28. Tamir, H., and P. R. Srinivasan. 1969. Purification and properties of anthranilate synthase from *Salmonella typhimurium*. J. Biol. Chem. 244:6507–6513.
29. Tse-Dinh, Y.-C., and J. C. Wang. 1986. Complete nucleotide sequence of the *topA* gene encoding *Escherichia coli* DNA topoisomerase I. J. Mol. Biol. 191:321–331.
30. Tso, J. Y., and H. Zalkin. 1981. Chemical modifications of *Serratia marcescens* anthranilate synthase component I. J. Biol. Chem. 256:9901–9908.
31. Westfall, H., N. W. Charon, and D. E. Peterson. 1983. Multiple pathways for isoleucine biosynthesis in the spirochete *Leptospira*. J. Bacteriol. 154:846–853.
32. Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51: 221–271.
33. Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. Nature (London) 289:751–758.
34. Yanofsky, C. 1984. Comparison of the regulatory and structural genes of tryptophan metabolism. Mol. Biol. Evol. 1:143–161.
35. Yelton, D. B., and N. C. Charon. 1984. Cloning of a gene required for tryptophan biosynthesis from *Leptospira biflexa* serovar *patoc* into *Escherichia coli*. Gene 28:147–152.
36. 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.
37. Zalkin, H. 1973. Anthranilate synthetase. Adv. Enzymol. 38: 1–39.
38. Zalkin, H., and T. Murphy. 1975. Utilization of ammonia for tryptophan synthesis. Biochem. Biophys. Res. Commun. 67: 1370–1377.
39. Zalkin, H., J. L. Paluh, M. van Cleemput, W. S. Moye, and C. Yanofsky. 1984. Nucleotide sequence of *Saccharomyces cerevisiae* genes *TRP2* and *TRP3* encoding bifunctional anthranilate synthase:indole-3-glycerol phosphate synthase. J. Biol. Chem. 259:3985–3992.
40. Zurner, R. L., and N. W. Charon. 1988. Nucleotide sequence analysis of a gene cloned from *Leptospira biflexa* serovar *patoc* which complements an *argE* defect in *Escherichia coli*. J. Bacteriol. 170:4548–4554.