JOURNAL OF BACTERIOLOGY, Oct. 1988, p. 4548–4554
0021-9193/88/104548-07$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 10

Nucleotide Sequence Analysis of a Gene Cloned from *Leptospira biflexa* Serovar *patoc* Which Complements an *argE* Defect in *Escherichia coli*

RICHARD L. ZUERNER[1,2] AND NYLES W. CHARON[1]*

Department of Microbiology and Immunology, Health Science Center, School of Medicine, West Virginia University, Morgantown, West Virginia 26506,[1] and National Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Ames, Iowa 50010[2]

Received 7 December 1987/Accepted 15 June 1988

The genus *Leptospira*, as a member of the order *Spirochaetales*, forms one of the most ancient evolutionary branches of the eubacteria. These spirochetes are morphologically and physiologically different from most eubacteria, and little is known about *Leptospira* genetics. In this communication, we report the first nucleotide sequence of a *Leptospira* gene. A gene which complements an *argE* mutation in *Escherichia coli* was isolated from a plasmid-based genomic library composed of *Leptospira biflexa* serovar *patoc* DNA. The functional region for the complementing activity was localized by transposon mutagenesis and restriction enzyme mapping and by subcloning. Nucleotide sequence analysis indicated a single open reading frame within the region containing *argE* complementing activity. The size of the predicted protein, 31,071 daltons, was in excellent agreement with data obtained from coupled transcription-translation reactions primed with cloned *L. biflexa* DNA. One surprising result was that the predicted amino acid sequence of this protein closely resembles portions of the β′ subunits of RNA polymerases from bacteria and chloroplasts.

Leptospira is a genetically diverse genus of long, slender, helical, motile bacteria (19). These spirochetes contain both pathogenic (*Leptospira interrogans*) and saphrophytic (*L. biflexa*) species (19). This genus contains a third species, *L. illini*, presently classified as species incertae sedis (19). Phylogenetic analysis involving oligonucleotide cataloging of bacterial 16S rRNAs suggests that the spirochetes, including *Leptospira* species, constitute one of the most ancient branches of the eubacteria (14, 44). The genes *Leptospira* forms one of the deepest evolutionary branches within the spirochete branch (31).

Our understanding of *Leptospira* genetics is limited to two previous reports involving cloning and analysis of a cluster of tryptophan (*trp*) biosynthetic genes from *L. biflexa* serovar *patoc* (45, 46). These studies suggest that the organization of the *trp* genes in *L. biflexa* is unique (46). In addition, transposon mutagenesis indicated that the *L. biflexa trp* genes were both transcribed and translated in vivo by *Escherichia coli* from sites initiating within the cloned DNA (45, 46).

We chose to expand our investigation of *Leptospira* genetics to include genes involved in arginine biosynthesis. The genetics of arginine biosynthesis in other bacteria, particularly *E. coli* and *Bacillus subtilis*, have been well characterized (11). Previous investigations of *Leptospira* physiology indicate that except for isoleucine, amino acids are synthesized by well-established biosynthetic pathways (9). We present here the analysis of an *L. biflexa* serovar *patoc* gene which complements an *E. coli* strain deficient in *argE* activity. In *E. coli*, the *argE* gene product converts *N*-acetylornithine to ornithine, an intermediate in arginine and polyamine biosynthesis (11).

(This work was presented in part at the 86th and 87th annual meetings of the American Society of Microbiology [R. L. Zuerner and N. W. Charon, Abstr. Annu. Meet. Am. Soc. Microbiol. 1986, H217, p. 163; R. L. Zuerner and N. W. Charon, Abstr. Annu. Meet. Am. Soc. Microbiol. 1987, H151, p. 164].)

MATERIALS AND METHODS

Materials. Restriction endonucleases, pBR322 DNA, lambda and M13 bacteriophage DNA, the nick-translation kit, ^{14}C-labeled protein molecular size markers, and deoxyribonucleotides were purchased from Bethesda Research Laboratories, Inc. Restriction endonucleases and T4 DNA ligase were purchased from New England BioLabs, Inc. Restriction endonucleases, molecular biology grade calf alkaline phosphatase, DNA polymerase I (Klenow fragment), and sequencing primer were purchased from Boehringer Mannheim Biochemicals. Mung bean nuclease and proteinase K were purchased from Pharmacia, Inc. Radiolabeled [^{35}S]methionine (1,475 Ci/mmol; 14.6 mCi/ml), [α-^{35}S]dATP (1,000 to 1,200 Ci/mmol; 8.1 mCi/ml), ^{14}C-labeled protein molecular size markers, and a coupled transcription-translation kit were purchased from Amersham Corp. Tran-[^{35}S]methionine (1,212 Ci/mmol; 6.9 mCi/ml) was purchased from ICN Radiochemicals Inc. GeneScreen and En^{3}Hance were purchased from Du Pont, NEN Research Products.

Bacterial strains and phage. The *E. coli* strains used in these studies were JA221 [*hsdR* (r$^-$ m$^+$) *recA* Δ*trpE5 leuB6*] (10) (from D. Yelton); AB1157 [*thr-1 leuB thi lacY galK ara-14 xyl-5 mtl-1* Δ(*gpt-proA*)*62 hisG4 argE3 supE44 rac rfbD1 mgl-51 rpsL31 tsx-33 kdgK51*] (35) (from K. Shimada), JM103 [Δ(*lac-pro*) *thi strA supE endA sbcB* F′ *traD proAB lacI*^{q} *lacZ* ΔM15] (27) (from D. Yelton); LE392 [*hsdR514* (r$^-$ m$^-$) *supE44 supF58 lacY1 galK2 galT22 metB1 trpR55*] (13) (from D. Yelton), and HB101 [*hsdR514* (r$^-$ m$^-$) *supE44 supF58 lacY1* or Δ(*lacIZY*)*6 galK2 galT22 metB1 trpR55*] (4) (from D. Yelton). The following *Leptospira* strains were used: *L. biflexa* serovar *patoc* strain Patoc I, *L. interrogans*

* Corresponding author.

serovar *bataviae* strain Van Tienen, and *L. illini* 3055 (obtained from the West Virginia University Laboratory) (8).

M13mp18 and M13mp19 phage vectors (27) were obtained as purified double-stranded replicative-form DNA from Bethesda Research Laboratories, Inc. Lambda 467 (*b221 rex*::Tn*5 c*I857 *O*am29 *P*am80) (13) was obtained from D. Yelton as a cell-free lysate.

Media and growth conditions. *E. coli* strains were routinely propagated in Luria broth (LB) medium (26) at 37°C; when used as a host for recombinant plasmids they were grown at 30°C. Complementation of auxotrophic mutations was performed by using minimal salts medium at 30°C (28). Transformed cells were incubated in Z broth at 37°C for phenotypic expression of antibiotic resistance genes (23). Either YT or DYT medium was used for all M13 phage infections, which were carried out at 37°C (27). The YM broth was used to grow *E. coli* recipients for lambda 467 infections (13). *Leptospira* strains were propagated at 30°C in the Ellinghausen and McCullough Tween 80-bovine serum albumin (Scientific Protein Laboratories) medium modified by Johnson and Harris (20).

DNA extraction. *Leptospira* DNA was extracted and purified as previously described (45). Plasmid DNA was purified from *E. coli* cells grown overnight in 100- to 250-ml cultures (12). Small-scale plasmid preparations were obtained by an alkaline lysis method from cells grown overnight in 1.5-ml cultures (18). Phage M13 replicative-form DNA was purified from 50 ml of late-logarithmic-phase M13-infected *E. coli* JM103 cultures (27) by using the protocol described for batch plasmid preparation.

Construction of genomic library. *L. biflexa* serovar *patoc* DNA was partially digested with *Bam*HI and ligated to *Bam*HI-digested, calf alkaline phosphatase-treated pBR322 DNA. Ligated DNA was used to transform *E. coli* JA221 to ampicillin resistance (Ap^r) (23). Transformants were pooled, and plasmid DNA was extracted. This pool of plasmid DNA was used in subsequent transformations of *E. coli* AB1157.

Gel electrophoresis and DNA blot analysis. Analytical and preparative gel electrophoresis was performed as described by Maniatis et al. (26). Specific DNA fragments were purified by electroelution from either agarose or polyacrylamide gels (26). Restriction fragments electrophoresed in agarose gels were blotted to GeneScreen with 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for Southern blot analysis (40). After hybridization, filters were washed with 0.1× SSC–0.5% sodium dodecyl sulfate (SDS) at 65°C, dried, sprayed with En^3Hance, and used to expose AR film (Eastman Kodak Co.) at −80°C. These conditions allowed an approximately 7% base mismatch as calculated by the method of Maniatis et al. (26).

Transposon mutagenesis. *E. coli* HB101 was transformed with pZC302 and transduced to kanamycin resistance (Km^r) with lambda 467 (13). Since *E. coli* AB1157 possesses an active K-12 restriction system, it was necessary to modify the pool of pZC302 insertion mutants obtained in HB101 (r^- m^-) prior to transformation of AB1157. This was achieved by first transforming *E. coli* JA221 (r^- m^+) to Km^r with the Tn*5*-modified pZC302 pool. Kanamycin-resistant transformants were pooled, plasmid DNA extracted, and used to transform *E. coli* AB1157 to Km^r. Phenotypes of the Tn*5*-modified plasmids in *E. coli* AB1157 were determined by replica plating to minimal medium lacking arginine. The Tn*5* insertion sites were located by restriction enzyme mapping of plasmid DNA.

In vitro-coupled transcription-translation assays. Polypeptides encoded by cloned *L. biflexa* DNA were identified through the use of an *E. coli*-coupled transcription-translation system. These cell extracts were primed with 2.5 to 5 μg of plasmid DNA, and the proteins synthesized were labeled with 45 μCi of [^{35}S]methionine as specified by the manufacturer. Proteins were denatured and electrophoresed in SDS-polyacrylamide gels (24). Radiolabeled proteins were identified by fluorography after impregnation of the gel with En^3Hance.

Nucleotide sequence analysis. *L. biflexa* DNA was subcloned into M13 phage vectors (27). Sequential overlapping deletions of the fragments subcloned into M13 phage vectors were constructed by use of exonuclease III and mung bean nuclease (16) or by digestion with appropriate restriction endonucleases. Each strand of *L. biflexa* DNA subcloned in M13 phage vectors was sequenced at least twice by using dideoxyribonucleotide chain termination reactions (34).

Nucleotide sequence data were analyzed for the presence of open reading frames (ORFs) by using a general sequence analysis program (25). Nucleotide and amino acid sequences were compared and aligned with the BIONET nucleotide sequence analysis package (39) (IntelliGenetics, Inc.) maintained by the U.S. Department of Agriculture-Agricultural Research Service Systems Research Resource. Potential secondary structures in RNA were determined by using the program described by Zuker and Stiegler (47).

RESULTS

Cloning an *L. biflexa* gene encoding *argE*-complementing activity. A genomic library of *L. biflexa* DNA containing approximately 2,750 individual Ap^r transformants was constructed in pBR322. Approximately 93% (93 of 100) of the Ap^r transformants contained insertions of *L. biflexa* DNA in the *Bam*HI site of pBR322 as determined by tetracycline sensitivity. This genomic library was used to transform *E. coli* AB1157 (Arg^-) to arginine phototrophy (Arg^+) and Ap^r. Two unique plasmid species were isolated from different Arg^+ Ap^r colonies: pZC300 (10.5 kilobase pairs [kb]) and pZC301 (6.6 kb). To eliminate the possibility that the Arg^+ phenotype in these transformants was a result of reversion at the chromosomal *argE* locus, we used purified pZC300 and pZC301 DNA to transform *E. coli* AB1157 to Ap^r and then screened for *argE* complementation by replica plating to minimal salts agar. All independent Ap^r colonies of *E. coli* AB1157 transformed with pZC300 (40 of 40) or pZC301 (40 of 40) were Arg^+. Conversely, all Ap^r colonies of *E. coli* AB1157 transformed with pBR322 (20 of 20) were Arg^-. Therefore, we concluded that cloned *L. biflexa* DNA was responsible for *argE* complementation and was not due to reversion of the chromosomal *argE* locus. Furthermore, the complementation observed with these plasmids was specific for *argE*. Attempts to complement other auxotrophic mutations contained in *E. coli* AB1157 (*leuB*, *proA*, and *hisG*) with argE-complementing plasmids were unsuccessful.

Restriction enzyme and DNA blot analysis. Restriction enzyme analysis was performed on pZC300 and pZC301. The results suggested that pZC300 and pZC301 had a common 2.2-kb *Bam*HI fragment (Fig. 1). Southern blot analysis was used to confirm that pZC300 and pZC301 contained *L. biflexa* serovar *patoc* DNA and to determine whether the 2.2-kb *Bam*HI fragments present in pZC300 and pZC301 were homologous. Genomic DNA from *E. coli*, *L. illini*, *L. interrogans*, and *L. biflexa* and purified pZC300 and pZC301 DNA were digested with *Bam*HI and blotted to nylon membranes (40). The immobilized DNA was used to hybridize the 2.2-kb *Bam*HI fragment of pZC301 which had been

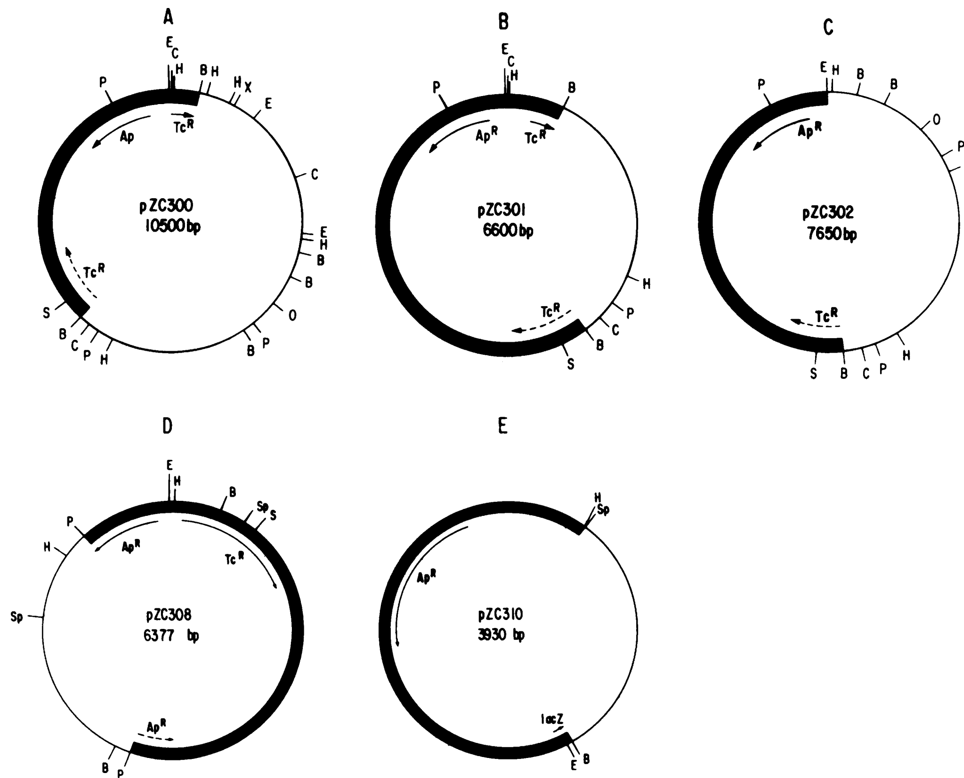

FIG. 1. Restriction maps of *argE*-complementing plasmids. Plasmids pZC300 (A) and pZC301 (B) were isolated from the genomic library. Plasmid pZC302 (C) was constructed from pZC300 by deletion of two *Eco*RI fragments. Plasmid pZC308 (D) was constructed by inserting the 2,061-bp *Pst*I fragment of pZC302 into pBR322. Plasmid pZC310 (E) was constructed by inserting the 1,244-bp *Bam*HI-*Sph*I fragment of pZC302 into the *lacZ* gene of pUC18. Abbreviations for restriction enzymes: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; K. *Kpn*I; O, *Xho*I; P, *Pst*I; S, *Sal*I; Sp, *Sph*I; and X, *Xba*I. ▬, Vector pBR322 DNA; ——, cloned *L. biflexa* DNA; →, approximate positions and orientations of the vector-encoded β-galactosidase gene (*lacZ*) and genes encoding resistance to ampicillin (Ap^r) and tetracycline (Tc^r); – – – , 3′ portion of insertionally inactivated antibiotic resistance genes.

labeled with [α-^{35}S]dATP by nick translation (specific activity, 1.1×10^6 dpm/μg). Under conditions of high stringency, hybridization between this probe and *L. biflexa*, pZC300, and pZC301 DNA, but not between the probe and DNA from *E. coli* or the other two *Leptospira* species, was detected (Fig. 2). These results confirm that pZC300 and pZC301 contain *L. biflexa* DNA and have a common 2.2-kb *Bam*HI fragment.

Construction and transposon mutagenesis of pZC302. Neither pZC300 or pZC301 was suitable for detailed analysis. Plasmid pZC300 contained more DNA than required to complement the *E. coli argE* gene, whereas pZC301 was unstable in several *E. coli* host strains. A new, stable *argE*-complementing plasmid, pZC302, was isolated by deleting the two small *Eco*RI fragments of pZC300 (Fig. 1).

The *argE*-complementing activity contained in pZC302 was localized by insertion mutagenesis with the Km^r transposon Tn*5*. Approximately 150 Km^r transformants of AB1157 were characterized by replica plating and restriction enzyme digestion. A schematic summary of the data is shown in Fig. 3A. On the basis of these data, a minimum of approximately 500 base pairs (bp) of the cloned *L. biflexa* DNA in pZC302 is required for *argE* complementation. Transposon Tn*5* insertions within *L. biflexa* DNA which flanks this region did not prevent *argE* complementation.

Coupled transcription-translation of *argE*-complementing plasmids. In vitro protein synthesis experiments were conducted with cell extracts of *E. coli* primed with pBR322,

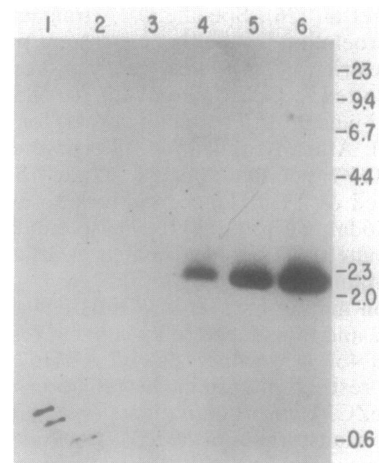

FIG. 2. Southern blot analysis of pZC300 and pZC301. DNA (5 μg) from *E. coli* (lane 1), *L. interrogans* (lane 2), *L. illini* (lane 3), *L. biflexa* (lane 4), and DNA (0.1 μg) from pZC301 (lane 5) and pZC300 (lane 6) were digested with *Bam*HI, electrophoresed, and blotted to nylon membranes. This nylon membrane was hybridized with approximately 2.7×10^5 dpm of labeled DNA (the 2.2-kb *Bam*HI fragment of pZC301) and washed, sprayed with En^{3}Hance, and used to expose AR film at −80°C. The resulting fluorograph is shown. Positions of *Hin*dIII-digested lambda DNA fragments are indicated, and the sizes are given in kilobases.

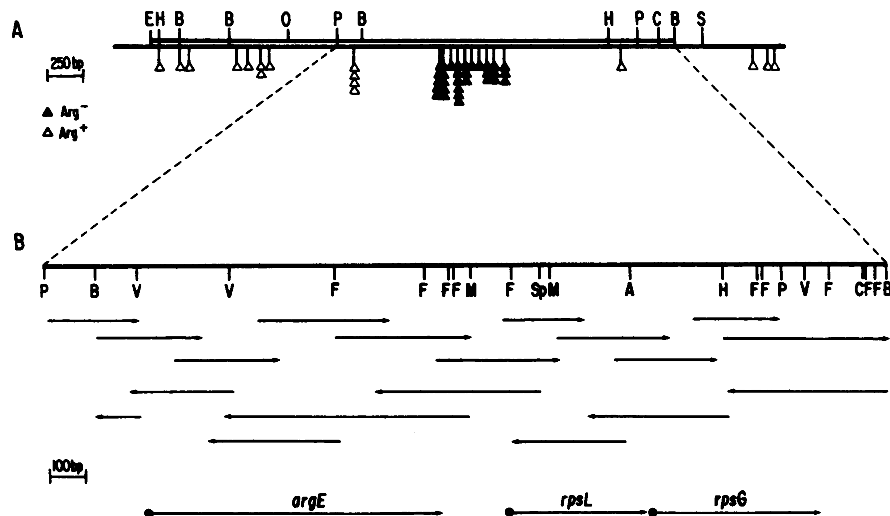

FIG. 3. Tn*5* mutagenesis and sequencing strategy of cloned *L. biflexa* DNA. (A) Restriction map of pZC302 and location of Tn*5* insertion sites. Symbols: ▲, Tn*5* insertions which prevented *argE* complementation; △, Tn*5* insertions which did not affect *argE* complementation. (B) Sequencing strategy of *argE*-complementing region of cloned *L. biflexa* DNA. The three ORFs detected within this sequence are identified. Abbreviations for restriction sites not given in Fig. 1: A, *Acc*I; F, *Hin*fI; M, *Mae*I; and V, *Eco*RV.

pZC300, pZC301, and pZC302 DNA. The radiolabeled proteins were separated by SDS-polyacrylamide gel electrophoresis and visualized by fluorography (Fig. 4A). A 31,500-dalton protein was synthesized in reactions primed with pBR322 DNA and corresponds to the β-lactamase (*bla*) gene product encoding ampicillin resistance. In addition to synthesis of β-lactamase, reactions primed with pZC300, pZC301, and pZC302 also encode proteins of 31,000, 17,500, and 15,000 daltons. Plasmid pZC300 also directed the synthesis of two unique proteins of 38,000 and 14,250 daltons.

Since the 31,000-dalton protein often comigrated with the vector-encoded β-lactamase and was difficult to resolve, a

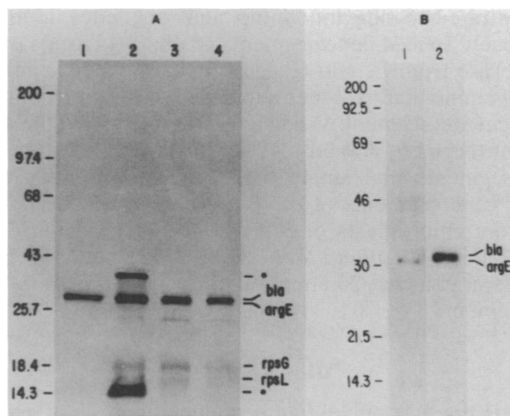

FIG. 4. Coupled transcription-translation of plasmids harboring the *L. biflexa argE* gene. (A) Radiolabeled proteins synthesized by cell extracts of *E. coli* primed with either pBR322 (lane 1), pZC300 (lane 2), pZC302 (lane 3), or pZC301 (lane 4) were separated on an SDS–5 to 15% polyacrylamide gel and visualized by fluorography. (B) Radiolabeled proteins synthesized by cell extracts primed with either pZC308 (lane 1) or pBR322 (lane 2) were separated on an SDS–12.5% polyacrylamide gel and visualized by fluorography. Positions and sizes of protein standards in kilodaltons are shown. Proteins encoded by the cloned 2.2-kb *Bam*HI fragment are identified with lines and labeled. Asterisks indicate proteins unique to pZC300.

new *argE*-complementing plasmid was constructed which disrupted β-lactamase synthesis. The 2.1-kb *Pst*I fragment of pZC302 was subcloned into the *Pst*I site of pBR322 located within the *bla* gene. The resulting plasmid, pZC308 (Fig. 1) was ampicillin sensitive and tetracycline resistant and complemented the *argE* mutation in *E. coli* AB1157. Coupled transcription-translation analysis of pZC308 detected synthesis of a 31,000-dalton protein (Fig. 4B), thus confirming its presence in *argE*-complementing plasmids.

Nucleotide sequencing. *L. biflexa* DNA extending from the *Pst*I site to the *Bam*HI site in pZC300 (as shown in Fig. 3B) was subcloned into M13mp18 and M13mp19 phage vectors. This DNA was sequenced by using dideoxyribonucleotide chain termination reactions (34) and the sequencing strategy shown in Fig. 3B. Three potential protein coding regions were detected within this sequence. Each ORF is on the same strand of DNA. The first ORF corresponds to the *argE* complementing region as defined by Tn*5* mapping (Fig. 3A). The nucleotide sequence and corresponding amino acid sequence for this region are shown in Fig. 5. A total of 279 bp was sequenced prior to the first methionine (Met) codon of this ORF (142 bp is shown), and 195 bp separated this ORF and the first Met codon of the next ORF. The predicted size of this protein is 31,971 daltons and is in excellent agreement with in vitro protein synthesis experiments primed with *argE*-complementing plasmids (Fig. 4).

Further subcloning of *argE*-complementing activity. Confirmation that the first ORF in the cloned *L. biflexa* DNA was responsible for *argE*-complementing activity was achieved by subcloning the 1,244-bp *Bam*HI-*Sph*I fragment of pZC302 (Fig. 3) into pUC18. The resulting plasmid, pZC310 (Fig. 1), contains 142 bp preceding and 269 bp following the putative *L. biflexa argE* gene. Plasmid pZC310 contains all of the putative *L. biflexa argE* gene and the first 27 codons of the second ORF (*rpsL*; see below). *E coli* AB1157 transformed with pZC310 was Arg^+, thus confirming that this ORF encodes the *L. biflexa argE*-complementing activity.

Analysis of sequenced DNA. The nucleotide sequence of the 2.2-kb *Bam*HI fragment was examined by using a number of DNA analysis programs. The second and third ORFs identified within this sequence share significant sequence

```
GGATCCGATGAATACTCTGTTCCAGTGGACGCTCTTCTCCAATTCCAAAATGGAGACAAG
        10        20        30        40        50        60

GTCAGGGAAGGGGATGTGATCTTCAAAATCCCGTCTGTGGCTGAAAAAACGCGAGATATC
        70        80        90       100       110       120

                            MetAsnPheSerLysLeuValValArgLysMetProAla
ACCGGTGGTCTCCCAAGGGTAGATGAACTTTTCGAAGCTCGTCGTCCGAAAGATGCCTGC
       130       140       150       160       170       180

 HisLeuGlnLysLeuThrValLysSerLysThrLysValLysSerLeuLysLysAsnGlu
ACACTTGCAGAAATTGACGGTAAAATCGAAGACAAAGGTGAAATCGTTAAAGAAAAACGA
       190       200       210       220       230       240

 TyrTyrThrIleIleProGluThrAlaGluGlnGluLysValLysValAlaIleProIle
ATATTATACGATCATTCCTGAAACAGCGGAACAAGAAAAAGTAAAAGTAGCAATTCCAAT
       250       260       270       280       290       300

 GlyLysGlnIleArgValArgGlnGlyAspPheValLysArgGlyAspGlnLeuAspGlu
CGGAAAACAAATCCGTGTTCGCCAGGGTGACTTTGTCAAACGCGGAGACCAGTTGGATGA
       310       320       330       340       350       360

 GlyAsnPheAspProHisAspIleLeuAlaIleLysGlyProAsnAlaLeuHisGluTyr
AGGAAATTTTGACCCGCATGATATCCTTGCGATCAAAGGACCAAATGCTCTTCACGAATA
       370       380       390       400       410       420

 LeuValSerGluValGlnGluValTyrArgLeuGlnGlyValHisIleAsnAspLysHis
CTTAGTTTCGGAAGTTCAGGAAGTTTACCGCTTACAAGGGGTTCATATCAATGATAAACA
       430       440       450       460       470       480

 IleGluValValValArgSerMetLeuArgLysValIleIleThrAspSerGlyAspThr
CATCGAAGTTGTGGTTCGTTCCATGCTGCGTAAGGTGATTATCACAGATAGTGGGGACAC
       490       500       510       520       530       540

 SerPheValAsnGlnGlnGlnValAspLysPheLeuPheAspGluGluAsnAspArgVal
ATCCTTTGTGAACCAACAACAAGTGGATAAATTCCTCTTTGATGAAGAAAACGACCGAGT
       550       560       570       580       590       600

 GluLysGluGlyGlySerProAlaGlnGlyThrProValLeuLeuGlyLeuThrLysAla
GGAAAAAGAAGGGGGATCTCCGGCACAAGGAACTCCTGTCCTTCTGGGATTAACAAAAGC
       610       620       630       640       650       660

 SerLeuAsnThrGluSerTyrPheSerAlaAlaSerPheGlnGluThrThrLysValLeu
ATCCCTCAACACTGAGTCTTATTTCTCGGCTGCATCATTCCAAGAAACAACAAAGGTTCT
       670       680       690       700       710       720

 ThrAspAlaAlaIleLysGlyLysThrAspAsnLeuMetGlyLeuLysGluAsnValIle
AACGGATGCGGCGATCAAAGGAAAAACAGACAACCTCATGGGTCTGAAAGAAAACGTAAT
       730       740       750       760       770       780

 IleGlyHisMetIleProAlaGlyThrGlyMetLysLysTyrArgAspIleGluValPhe
CATTGGTCACATGATCCCTGCGGGAACAGGCATGAAAAAATACCGTGACATTGAAGTTTT
       790       800       810       820       830       840

 LysAspLeuProGlyAspLeuAspTrpAspLeuAlaIleGlyGlyArgGlyArgArgSer
CAAAGACCTTCCTGGGGATTTGGATTGGGATCTGGCAATCGGAGGAAGAGGAAGAAGAAG
       850       860       870       880       890       900

 PheArgThrPheArgValGlySerCysPheHisCysHisThrLeuSerThrCysCysArg
TTTCCGAACTTTCCGAGTCGGCTCCTGTTTCCACTGCCACACTCTCTCGACTTGTTGCCG
       910       920       930       940       950       960

 ArgGlyArgGlyTER
AAGAGGACGAGGATGAAGATGAGTTGGAAGAAGAGTCTGATGACTCGGACGATGAAGACG
       970       980       990      1000      1010      1020

ACGACGATTAAACCAAACTTTGTCCCTAGTAATTCGTAAGAGTTTCTGGGGACAAAATCA
      1030      1040      1050      1060      1070      1080

TGTTTTCGTTTACAAAATGTCGCTATCTGGAATTTTGGACGAAAACGTCATTATTTTTTT
      1090      1100      1110      1120      1130      1140

                                Met
AAAGACAGAGAAGTAGAAGAAGGAATCGAATG
      1150      1160      1170
```

FIG. 5. Nucleotide sequence of *L. biflexa argE*-complementing gene and surrounding DNA. Nucleotide sequence and predicted amino acid sequence gene are shown. Potential −35 and −10 promoter sequences, the conserved 34-bp sequence, and potential ribosome-binding sequences are underlined.

```
                                        Segment A

L. biflexa    FDPHDILAIKGPNALHEYLVSEVQEVYRLQGVHINDKHIEVVVRSMLRKVIITDSGDTSF    76-135
E. coli       EA.....RLR.VH.VTR.I.N...D.......K........I..Q....AT.VNA.SSD.  1215-1274
M. polymorpha LWGFFLSTKISMEQGQIN..DQI.K..QS...Q.SN....II..Q.TS...TLED.M.NV  1176-1235
N. tabacum    LGIPWGFL.GAELTIRIS..NKI.Q...S...Q.HNR.L.II..QITS..LVSED.MSNV   654-716
                             AQS

                                        Segment B

L. biflexa    PVLLGLTKASLNTESYFSAASFQETTKVLTDAAIKGKTDNLMGLKENVIIGHMIPAGTGM   165-224
E. coli       RD...I.....A...FI.........R...E..VA..R.E.R.......V.RL......Y  1304-1363
M. polymorpha .I...I.......Q.FI.E.......R..AK..L..RI.W.K.......L.GLV.....S  1262-1321
N. tabacum    V....I.R.....Q.FI.E......AR..AK..LR.RI.W.K......VL.GV..V...F   743-802
```

FIG. 6. Amino acid sequence comparison of the *argE*-complementing gene of *L. biflexa* to the β′ subunits of RNA polymerases. Two gene segments (A and B) coding for the *L. biflexa argE*-complementing protein were compared with gene segments of the β′ subunits of RNA polymerases from *E. coli* and chloroplasts from *M. polymorpha* and *N. tabacum*. Dots signify identity.

similarity with ribosomal proteins S12 (*rpsL*) and S7 (*rpsG*), respectively (R. Zuerner, Ph.D. thesis, West Virginia University, Morgantown, 1986; Zuerner and Charon, Abstr. Annu. Meet. Am. Soc. Microbiol. 1987; R. Zuerner and N. Charon, manuscript in preparation). These two ORFs have the potential to encode proteins of 13,932 and 18,288 daltons for *rpsL* and *rpsG*, respectively. These results are consistent with the two smaller proteins synthesized in vitro from *argE*-complementing plasmids pZC300, pZC301, and pZC302 (Fig. 4). The location of these proteins in the cloned fragment is shown in Fig. 3.

Sequences located between nucleotides 51 and 56 and between nucleotides 75 and 81 resemble −35 and −10 regions, respectively, characteristic of *E. coli* promoters (33). Preceding *argE* (between nucleotides 84 and 117) and *rpsL* (between nucleotides 1073 and 1106) is the sequence CAAAATCNPyGTPyTNPyGNPyTPuNAAAANNNCGNNAT. This sequence precedes the first Met codon in *argE* by 26 bp. The purine-rich sequence AAGG precedes the first Met codon of each of these ORFs by 4 to 7 nucleotides. A potentially stable stem-loop structure ($\Delta G^\circ = -41.2$ kcal [172.4 kJ]) follows the *argE* ORF (between nucleotides 981 and 1128).

The *L. biflexa argE* gene was used to search GenBank (3) and EMBL (7) nucleotide sequence data bases for genes with similar sequences. Only the nucleotide sequence encoding the β′ subunit of RNA polymerase of *E. coli* (30) shared significant sequence similarity to the *L. biflexa argE* gene. The Protein Information Resource protein data base (38) was searched for amino acid sequence similarities to the predicted *L. biflexa argE* gene product. Two adjacent 60-amino-acid gene segments (residues 76 to 135 and 165 to 224) demonstrated significant amino acid sequence identity to two closely spaced gene segments of the β′ subunits of RNA polymerase from *E. coli* (Fig. 6). In both *L. biflexa* and *E. coli*, 29 amino acid residues separate the two segments. The first segment (segment A) had a 60% identity, and the second segment (segment B) had a 72% identity. In addition, these same segments had amino acid sequence identity to two similar gene segments of the β′ subunits of RNA polymerases from chloroplasts of *Marchantia polymorpha* (29) (35 and 68%) and *Nicotiana tabacum* (37) (30 and 62%) (Fig. 6). In these organisms, 26 amino acid residues separated these two segments.

DISCUSSION

This is the first nucleotide sequence analysis of any gene isolated from the spirochete *Leptospira* sp. The results reported here reveal some unique aspects related to *L. biflexa* gene organization and function.

The functional region for the *argE*-complementing activity was localized to a 2.2-kb *Bam*HI fragment during the initial cloning process. Further localization of *argE*-complementing activity was permitted through transposon mutagenesis of pZC302 with Tn*5* and by subcloning. A 1.2-kb fragment containing only one complete ORF encoding a 31,000-dalton protein complemented the *argE* defect in *E. coli*. These results indicate that this ORF, and not the *rpsL*

or *rpsG* gene, encodes *argE* complementing activity. *Leptospira* genetics are poorly characterized, and the function of this gene in *L. biflexa* cannot be determined. Furthermore, since this is the first published characterization of a gene which complements *argE* defects in *E. coli* from any source, comparisons with homologous genes from other bacteria are not possible. However, no homology was detected between the sequence reported here and the first 104 amino acid residues of the *E. coli argE* sequence made available to us (A. Boyen, J. Piette, R. Cunin, and N. Glansdorff, personal communication).

Nucleotide sequence analysis of the 2.2-kb *Bam*HI fragment of pZC300 identified genes encoding ribosomal proteins S12 (*rpsL*) and S7 (*rpsG*) (Zuerner and Charon, Abstr. Annu. Meet. Am. Soc. Microbiol. 1987; Zuerner and Charon, in preparation) located downstream from *argE*. The placement of a gene encoding *argE*-complementing activity adjacent to *rpsL* and *rpsG* is unusual among bacteria for which related chromosomal regions have been examined (2, 32).

The results presented here suggest that *E. coli* utilized *L. biflexa* sequences for transcription and translation initiation. Removal of the vector-encoded *tet* promoter (5, 43) during construction of pZC302 did not prevent *argE* complementation by this plasmid. Additionally, transposon Tn*5* insertions at either end of the cloned *L. biflexa* DNA did not prevent pZC302 complementation of the *argE* mutation in *E. coli* AB1157. Since Tn*5* introduces polar mutations (13), *E. coli* must initiate transcription and translation from within cloned *L. biflexa* DNA. Similar conclusions have been drawn on the basis of results from transposon mutagenesis of cloned *L. biflexa trp* genes (46). It cannot be determined from the data presented here or in previous reports (45, 46) whether *E. coli* initiates transcription of these genes from native *L. biflexa* promoters or sequences which fortuitously resemble *E. coli* promoters. A sequence resembling an *E. coli* promoter was located between nucleotides 51 and 81 and may be utilized by *E. coli* during transcription of the *L. biflexa argE* gene. The conserved 34-bp sequence which precedes both *argE* and *rpsL* may function in promoter recognition by *L. biflexa*. Additionally, transcription termination may occur within cloned *L. biflexa* DNA, since a potential stem-loop structure resembling transcription terminators in other eubacteria (17, 33) was identified following the *argE* ORF.

The predicted size of the *L. biflexa* gene product, 31,071 daltons, is in excellent agreement with results of protein synthesis studies using *argE* complementing plasmids. These data suggest that translation of the *argE* mRNA initiates at the first Met codon in the ORF. However, the 10th codon in this ORF also codes for methionine, and translation initiation at this codon would direct the synthesis of a 28,869-dalton protein. Although the data are consistent with translation initiation at either Met codon, we predict that translation initiation occurs at the first Met codon for the following reasons. In most procaryotes, translation initiation is directed by interactions between the 3′ end of the 16S rRNA and the 5′ end of the mRNA (reviewed in reference 21). The portion of the 16S rRNA sequence which binds to mRNA is often pyrimidine rich and is complementary to a purine-rich sequence on the mRNA (36, 41). The purine-rich sequence AAGG precedes each of the known *L. biflexa* ORFs: *argE* (this work) and *rpsL* and *rpsG* (Zuerner and Charon, Abstr. Annu. Meet. Am. Soc. Microbiol. 1987; Zuerner and Charon, in preparation). These sequences are optimally spaced for ribosome-binding sequences (15, 21, 42). Conversely, the sequence preceding the second Met codon in this ORF is pyrimidine rich. These results, as well as the similarity of this AAGG sequence to *E. coli* ribosome-binding sites (15, 21), suggest that *E. coli* may utilize this sequence during translation initiation of *L. biflexa* genes.

Perhaps the most intriguing finding of this work was the least expected: that the *L. biflexa argE* gene product would share amino acid sequence similarity with RNA polymerase β′ subunits. The function of the β′ subunit has not been clearly established. These subunits are generally considered to function in DNA binding (22), and putative DNA-binding domains within these protein subunits have been identified (1). The region of similarity between the RNA polymerase β′ subunits and the *L. biflexa argE* gene product is one of the most highly conserved regions among RNA polymerases (region VI [1, 6]) but is located outside of this putative DNA-binding domain (1). Helix-turn-helix structures typical of DNA-binding domains are not apparent in secondary structure predictions for the *argE*-complementing protein (data not shown). Therefore, it is unlikely that the region of similarity between this protein and the β′ subunit represents DNA-binding activity by the *argE*-complementing protein. Instead, sequence similarities between the *L. biflexa argE*-complementing protein and the RNA polymerase β′ subunits imply functional similarities. One possibility is that both the β′ subunit of RNA polymerases and *L. biflexa argE*-complementing activity possess *N*-acetylase of *N*-acetylornithinase activity; we note that the *E. coli argE* gene product demonstrates wide substrate specificity in deacetylating several compounds (11). Future biochemical analysis should clarify the nature of the cloned *L. biflexa argE*-complementing activity and establish whether this activity is shared by RNA polymerase β′ subunits.

ACKNOWLEDGMENTS

We appreciate the helpful discussions and suggestions of R. Stenberg, J. Strobl, and D. Yelton. We thank R. Hendrix, P. Kapke, and R. Stenberg for assistance in computer analysis of nucleotide sequence data and R. Cunin for communicating unpublished results. We especially appreciate the assistance of D. Yelton in critically reading the manuscript and of R. Hendrix in emphasizing the importance of the amino acid sequence identity between the *argE*-complementing gene and the gene coding for β′ subunit of *E. coli* RNA polymerase.

This research was supported by Public Health Service grant DE04645 from the National Institute of Dental Research and by National Institutes of Health Biomedical Research Support Grant S07-RR05433.

LITERATURE CITED

1. **Allison, L. A., M. Moyle, M. Shales, and C. J. Ingles.** 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell **42:**599–610.
2. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47:**180–230.
3. **Bilofsky, H. S., and C. Burks.** 1988. The GenBank genetic sequence data bank. Nucleic Acids Res. **16:**1861–1863.
4. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:**459–472.
5. **Brosius, J., R. L. Cate, and A. P. Perlmutter.** 1982. Precise location of two promoters for the β-lactamase gene of pBR322. S1 mapping of ribonucleic acid isolated from *Escherichia coli* or synthesized in vitro. J. Biol. Chem. **257:**9205–9210.
6. **Broyles, S. S., and B. Moss.** 1986. Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits. Proc. Natl. Acad. Sci. USA **83:**3141–3145.
7. **Cameron, G. N.** 1988. The EMBL data library. Nucleic Acids

Res. **16**:1865–1867.

8. **Carleton, O., N. W. Charon, P. Allender, and S. O'Brien.** 1979. Helix handedness of *Leptospira interrogans* as determined by scanning electron microscopy. J. Bacteriol. **137**:1413–1416.
9. **Charon, N. W., R. C. Johnson, and D. Peterson.** 1974. Amino acid biosynthesis in the spirochete *Leptospira*: evidence for a novel pathway of isoleucine biosynthesis. J. Bacteriol. **117**:203–211.
10. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
11. **Cunin, R., N. Glansdorff, A. Pierard, and V. Stalon.** 1986. Biosynthesis and metabolism of arginine in bacteria. Microbiol. Rev. **50**:314–352.
12. **Davis, R. W., D. Botstein, and J. R. Roth (ed.).** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **deBruijn, F. J., and J. R. Lupski.** 1984. The use of transposon Tn*5* mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids—a review. Gene **27**:131–149.
14. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, 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. Leuhrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaroytes. Science **209**:457–463.
15. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translation initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
17. **Holmes, W. M., T. Platt, and M. Rosenberg.** 1983. Termination of transcription in *E. coli*. Cell **32**:1029–1032.
18. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989–2998.
19. **Johnson, R. C., and S. Faine.** 1984. Genus I. *Leptospira* Noguchi 1917, 755^{AL}. p. 62–67. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins, Co., Baltimore.
20. **Johnson, R. C., and V. G. Harris.** 1967. Differentiation of pathogenic and saprophytic leptospires. I. Growth at low temperatures, J. Bacteriol. **94**:27–31.
21. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
22. **Kumar, S. A.** 1981. The structure and mechanism of action of bacterial DNA-dependent RNA polymerase. Prog. Biophys. Mol. Biol. **38**:165–210.
23. **Kushner, S. R.** 1978. An improved method for transformation of *Escherichia coli* with ColE1 derived plasmids, p. 17–23. *In* H. W. Boyer and S. Nicosia (ed.), Genetic engineering. Elsevier/North-Holland Biomedical Press, Amsterdam.
24. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
25. **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.
26. **Maniatias, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
28. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.
29. **Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki.** 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature (London) **322**:572–574.
30. **Ovchinnikov, Y. A., G. S. Monastyrskaya, V. V. Gubanov, S. O. Guryev, I. S. Salomatina, T. M. Shuvaeva, V. M. Lipkin, and E. D. Sverdlow.** 1982. The primary structure of *E. coli* RNA polymerase. Nucleotide sequence of the *rpoC* gene and amino acid sequence of the β′-subunit. Nucleic Acids Res. **10**:4035–4044.
31. **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.
32. **Piggot, P. J., and J. A. Hoch.** 1985. Revised genetic linkage map of *Bacillus subtilis*. Microbiol. Rev. **49**:158–179.
33. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Shimada, K., R. A. Weisberg, and M. E. Gottesman.** 1972. Prophage lambda at unusual chromosomal locations. I. Location of secondary attachment sites and properties of lysogens. J. Mol. Biol. **63**:483–503.
36. **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.
37. **Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B. Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, and M. Sugiura.** 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**: 2043–2049.
38. **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.
39. **Smith, D. H., D. Brutlag, P. Friedland, and L. H. Kedes.** 1986. BIONET: national computer resource for molecular biology. Nucleic Acids Res. **14**:17–20.
40. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
41. **Steitz, J. A., and K. Jakes.** 1975. How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **72**:4734–4738.
42. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.
43. **Stuber, D., and H. Bujard.** 1981. Organization of transcriptional signals in plasmids pBR322 and pACYC184. Proc. Natl. Acad. Sci. USA **78**:167–171.
44. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51**:221–271.
45. **Yelton, D. B., and N. W. Charon.** 1984. Cloning of a gene required for tryptophan biosynthesis from *L. biflexa* serovar *patoc* into *Escherichia coli*. Gene **28**:147–152.
46. **Yelton, D. B., and R. A. Cohen.** 1986. Analysis of cloned DNA from *L. biflexa* serovar *patoc* which complements a deletion of the *Escherichia coli trpE* gene. J. Bacteriol. **165**:41–46.
47. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.