JOURNAL OF BACTERIOLOGY, May 1990, p. 2693–2703
0021-9193/90/052693-11$02.00/0

Vol. 172, No. 5

Organization and Nucleotide Sequences of the *Spiroplasma citri* Genes for Ribosomal Protein S2, Elongation Factor Ts, Spiralin, Phosphofructokinase, Pyruvate Kinase, and an Unidentified Protein

C. CHEVALIER, C. SAILLARD, AND J. M. BOVÉ*

Laboratorie de Biologie Cellulaire et Moléculaire, Institut National de la Recherche Agronomique, and Université de Bordeaux II, Domaine de la Grande Ferrade, 33140 Pont de la Maye, France

Received 4 October 1989/Accepted 23 January 1990

The gene for spiralin, the major membrane protein of the helical mollicute *Spiroplasma citri*, was cloned in *Escherichia coli* as a 5-kilobase-pair (kbp) DNA fragment. The complete nucleotide sequence of the 5.0-kbp spiroplasmal DNA fragment was determined (GenBank accession no. M31161). The spiralin gene was identified by the size and amino acid composition of its translational product. Besides the spiralin gene, the spiroplasmal DNA fragment was found to contain five additional open reading frames (ORFs). The translational products of four of these ORFs were identified by their amino acid sequence homologies with known proteins: ribosomal protein S2, elongation factor Ts, phosphofructokinase, and pyruvate kinase, respectively encoded by the genes *rpsB*, *tsf*, *pfk*, and *pyk*. The product of the fifth ORF remains to be identified and was named protein X (*X* gene). The order of the above genes was *tsf*—*X*—spiralin gene—*pfk*—*pyk*. These genes were transcribed in one direction, while the gene for ribosomal protein S2 (*rpsB*) was transcribed in the opposite direction.

Spiroplasmas are mollicutes (mycoplasmas) (class *Mollicutes*) with motility and helical morphology (33). Analysis of the *Spiroplasma* genome by classical genetic techniques has been difficult. The dependence of these organisms on complex growth media and their poorly defined metabolic pathways explain the paucity of auxotrophic mutants and hence of genetic markers in these procaryotes. In the absence of such markers, characterization of cloned DNA fragments provides the means for analyzing the structural organization of *Spiroplasma* genes.

The gene for spiralin seemed to be of particular interest. Spiralin (35) is the major membrane protein of *Spiroplasma citri*, an important plant pathogen (28). It is apparently a transmembrane amphiphilic protein (34), and it is acylated (37). Acylation seems to be a characteristic of several membrane proteins of mollicutes. The role of spiralin is not known. It must, however, be a key protein of the spiroplasmal membrane, since spiralinlike proteins occur in *Spiroplasma* species other than *S. citri* (37) and probably in all spiroplasmas. Elucidation of the amino acid sequence of spiralin as deduced from the nucleotide sequence of its gene would undoubtly contribute to a better understanding of its structure and function, as well as the organization of its gene. Sequence determination of the spiralin gene was made possible by the fact that we have previously cloned the gene. A library of cloned genomic sequences of *S. citri* R8A2 (28) was constructed by incorporation of *Hin*dIII restriction fragments into plasmid pBR328 and cloning in *Escherichia coli* (21). The bacterial clone harboring recombinant plasmid pES1 was selected by its ability to express spiralin. Spiralin was indeed expressed in *E. coli* and was the first mollicute gene product to be fully expressed in a bacterium (21).

Recombinant plasmid pES1 was subcloned into plasmid pES3′. The 5-kilobase-pair (kbp) spiroplasmal insert of pES3′ has been entirely sequenced. Sequence analysis of the 5-kbp spiroplasmal insert of pES3′ made it possible to identify the spiralin gene and five additional open reading frames (ORFs). Four of these ORFs have been tentatively identified as genes for ribosomal protein S2, elongation factor Ts, pyruvate kinase (PYK), and 6-phosphofructokinase (PFK) by protein bank sequence comparisons. The organization of these genes is presented.

We further show that the gene for spiralin contains no UGA codons, accounting for the full expression of the spiralin gene in *E. coli*. We also report that the spiralin gene can be inactivated by mutational integration of the insertion element IS*1*.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* TG1 was used as the host for bacteriophage M13. All recombinant plasmids were propagated in *E. coli* HB101. The spiroplasmal DNA insert in plasmid pES1 (21), derived from *S. citri* R8A2 (28), was used as the source of the spiralin gene. The promoter selection vector pKK232.8 was obtained from Pharmacia Biochemicals (Uppsala, Sweden).

Chemical reagents and enzymes. Isopropyl-β-D-thiogalactoside, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, restriction enzymes, calf intestine phosphatase, and T4 DNA ligase were obtained from Boehringer (Mannheim, Federal Republic of Germany [FRG]). The Klenow fragment of DNA polymerase I was obtained from Pharmacia Biochemicals. The random primed DNA labeling kit, the M13 sequencing kit, and the labeled nucleotides [α-^{32}P]dCTP (110 TBq/mmol) and [α-^{35}S]dATP (22 TBq/mmol) were purchased from Amersham Corp. (Arlington Heights, Ill.). *N*,*N*′-Methylene-bisacrylamide and acrylamide were from Bio-Rad Laboratories (Richmond, Calif.). Urea was obtained from Merck (Darmstadt, FRG). Agarose and low-melting-point agarose were from Bethesda Research Laboratories (Gaithersburg, Md.).

Isolation of plasmid DNA. Analysis of recombinant plasmids was performed by the miniscreen method (5). Large-scale purification of plasmid DNA was done by the method of Rodriguez and Tait (26).

Analysis of spiralin expression in *E. coli*. The expression of

* Corresponding author.

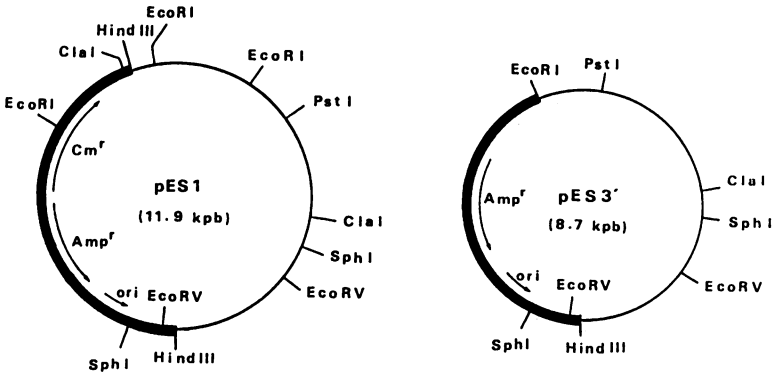

FIG. 1. Restriction maps of recombinant plasmids pES1 (21) and pES3′ expressing spiralin in *E. coli*. The heavy lines represent the pBR328 vector (3.7 kbp). ori, Origin of replication; Amp^r, gene for ampicillin resistance; Cm^r, gene for chloramphenicol resistance.

spiralin in *E. coli* was determined by enzyme-linked immunosorbent assay (ELISA) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by immunoblotting, as described previously (21).

Subcloning of pES1. Subcloning of pES1 was carried out as described previously (21). The recombinant DNA of this clone was labeled pES3′.

Shotgun cloning and dideoxy sequencing of the spiroplasmal DNA insert of pES3′. Recombinant plasmid pES3′ was used for shotgun cloning and sequencing of the *Eco*RI-*Hin*dIII restriction fragment of the spiroplasmal DNA insert. pES3′ was sonicated at 10 W for 45 s. The resulting randomly generated fragments were separated by gel electrophoresis with 1% low-melting-point agarose. Fragments of 400 to 800 bp were excised from the gel and purified by a procedure described by Guo and Wu (13), taking care that the melting temperature did not go beyond 65°C. The purified fragments were made blunt ended by a fill-in reaction with the Klenow enzyme. The blunt-ended fragments were ligated to the dephosphorylated *Sma*I-linearized M13mp8 RF vector and used to transform *E. coli* TG1 cells (14). In situ hybridization (20) was used to select the recombinant phages harboring spiroplasmal DNA by using the *Eco*RI-*Hin*dIII *S. citri* DNA insert purified from pES3′ as a radioactive probe. A total of 230 hybridization-positive subclones were obtained.

Preparation of single-stranded templates, annealing reactions, and sequencing reactions were carried out according to the "M13 Cloning and Sequencing Handbook" (Amersham Corp., London, England) except that lowered concentrations of ddATP (0.015 mM) and ddTTP (0.05 mM) were used as described previously (24). [α-^{35}S]dATP was used as the labeled nucleotide.

Sequence analysis. Gel reading was performed with a translucent digitizing pad for direct input of the sequence data into a computer file. The NUCALN alignment program of Wilbur and Lipman (32) was used for nucleotide sequence analysis. Predicted amino acid sequences were obtained with the NUMSEQ translational program of Fritensky et al. (10). Identification of proteins coded by the ORFs was done with the FASTP amino acid sequence alignment program of Lipman and Pearson (17), using the protein data bank of the National Biomedical Research Foundation (Georgetown University Medical Center, Washington, D.C.).

Characterization of the spiralin gene promoter. The *Eco*RI-*Hin*dIII spiroplasmal DNA insert of pES3′ was digested by the restriction enzyme *Rsa*I. The resulting restriction fragments were separated by electrophoresis in a 1% low-melting-point agarose gel. A 292-bp *Rsa*I-*Rsa*I fragment containing the spiralin gene promoter sequence was purified from the agarose and made blunt ended with the Klenow enzyme. It was then ligated into the promoter selection vector pKK232.8, which had been linearized with *Sma*I and dephosphorylated. The ligation mixture was used to transform *E. coli* HB101 cells. Recombinant clones were selected on LB medium containing ampicillin (30 μg/ml) (Sigma Chemical Co., St. Louis, Mo.). Ampicillin-resistant transformants were transferred to LB medium containing chloramphenicol (50 μg/ml) (Sigma Chemical Co.).

RESULTS

Subcloning the spiralin gene of plasmid pES1 into plasmid pES3′. Plasmid pES1 contains a 6.5-kbp fragment of *S. citri* DNA (21). The recombinant plasmid pES3′ is composed of the *Eco*RI-*Hin*dIII fragment of pBR328 (3.7 kbp) and the *Eco*RI-*Hin*dIII fragment (5.0 kbp) of the *S. citri* DNA insert carried by pES1 (Fig. 1). Plasmid pES3′ was in all respects identical to plasmid pES3 obtained previously (21). The bacterial transformant carrying pES3′ expressed spiralin as measured by ELISA (data not shown) and Western immunoblot (Fig. 2). The difference in migration between the *S. citri* spiralin (track 1) and the spiralin synthesized in both *E. coli* clones (tracks 2 and 3) has been discussed previously (21). It should be noted that the 5.0-kbp spiroplasmal DNA insert is much larger than the DNA required to encode spiralin (about 0.8 kbp).

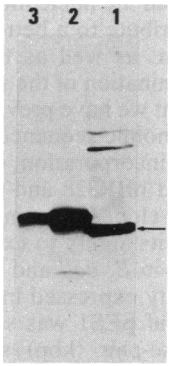

FIG. 2. Western blot of *S. citri* R8A2 lysate (track 1) and of two *E. coli* lysates harboring recombinant plasmid pES1 (track 2) or pES3′ (track 3). Polyclonal antibodies against whole *S. citri* cells detected spiralin as the major antigen in spiroplasmal lysates (arrow, track 1). Spiralin from bacterial lysates (track 2 and 3) had a slightly lower electrophoretic mobility (21).

Nucleotide sequence of the spiroplasmal DNA insert of plasmid pES3′. The complete nucleotide sequence of the pES3′ insert was determined by the dideoxynucleotide chain termination method (29) and was found to comprise 5,025 bp. Sequence data were obtained for both strands. The nucleotide sequence of the 5′→3′ *Eco*RI-*Hin*dIII strand, referred to as the plus strand, is shown in Fig. 3. The complementary sequence (minus strand) of the first 700 nucleotides is also shown in Fig. 3.

The base composition (molar percentage) was as follows: 37.3% A, 33.0% T, 17.0% G, and 12.7% C. The G+C content was 29.7%, a value slightly higher than that of *S. citri* genomic DNA (26.2%) (6).

Distribution of ORFs on the spiroplasmal DNA insert of pES3′. Six ORFs were found to be located on the *S. citri* DNA insert of pES3′. Their locations are shown in Fig. 4. Four complete ORFs (II to V) occurred on the DNA strand with the *Eco*RI site as the 5′ end (plus strand). They extended, left to right, from nucleotides 739 to 4401. A fifth ORF (VI) began at nucleotide 4455, but was interrupted at nucleotide 5025, the *Hin*dIII boundary of the spiroplasmal DNA fragment. No coding region was found on the plus strand from nucleotides 1 to 739. However, in this region, the start of a sixth ORF (ORF I) was present on the complementary strand (minus strand), beginning at nucleotide 549 and reading from right to left. The *Eco*RI boundary of the spiroplasmal DNA fragment interrupted ORF I. ORF I and ORF II are thus read in opposite directions.

An ATG served as the initiation codon of each ORF except for ORF III, where it was GTG. Each complete ORF (II to V) was terminated by TAA. Table 1 summarizes the positions of the six ORFs on the respective DNA strands of the *Eco*RI-*Hin*dIII spiroplasmal DNA insert of pES3′. The sizes of the putative proteins coded by each ORF are also given in Table 1. ORFs I and VI, located at the two ends of the spiroplasmal DNA insert, were incomplete, since ORF I and ORF VI were interrupted by the *Eco*RI and *Hin*dIII cloning sites, respectively. Hence, as indicated in Table 1, the complete translational products of ORF I and ORF VI have more amino acids than their partial ORFs situated at the ends of the spiroplasmal insert.

A bacterial ORF possesses a Shine-Dalgarno (SD) sequence 5 to 10 nucleotides upstream of the initiation codon. This sequence is complementary to the 3′ end of the 16S rRNA (30). Each of the six ORFs of the spiroplasmal DNA insert had an SD sequence. Figure 5 lists these sequences in comparison with the sequence of the 3′ ends of the 16S rRNAs of *S. citri*, *Bacillus subtilis*, and *E. coli*.

Identification of proteins encoded by different ORFs. Since the bacterial clone harboring pES3′ synthesized the complete spiralin protein, one of the four complete ORFs of the insert must be the spiralin gene.

The following results indicate that ORF IV is the spiralin gene. (i) The protein encoded by ORF IV had a mass of 25,282 Da, a value very close to that of purified spiralin (35). (ii) The amino acid composition of the ORF IV protein, as deduced from the sequencing data, was almost identical to that obtained by Wroblewski et al. (37) for purified spiralin (Table 2). Spiralin is composed of 241 amino acids. Four amino acids, alanine, valine, lysine, and threonine, account for more than 47.8% of all amino acids in the protein. Spiralin contains a low number of tyrosine, phenylalanine, and cysteine residues and lacks arginine, histidine, tryptophan, and internal methionine. (iii) The localization of ORF IV overlapping the *Cla*I restriction site and ending 151 nucleotides upstream of the *Sph*I site agrees well with data from subcloning experiments (6).

Identification of the proteins encoded by the other ORFs is summarized in Table 1. The protein product encoded by the interrupted ORF I shows 38% homology with the NH_2-terminal amino acid sequence of ribosomal protein S2 of *E. coli* (1) (Fig. 6A), indicating that ORF I is probably the *S. citri rpsB* gene. The amino acid sequence of the ORF II translational product had 46.0% homology with *E. coli* elongation factor EF-Ts (1) (Fig. 6B), suggesting that ORF II represents the *S. citri tsf* gene. The *S. citri* EF-Ts protein (297 amino acids) was 14 amino acids longer than that of *E. coli* (283 amino acids). No significant homology was observed between the putative protein of ORF III and any protein identified through the bank. The ORF III protein, which remains unidentified, has been named protein X. The protein encoded by ORF V had 45.8% homology with the sequence of PFK from *Bacillus stearothermophilus* (Bs-PFK) (16) (Fig. 6C). ORF IV may thus be the *pfk* gene of *S. citri*. The *S. citri* enzyme (327 amino acids) was seven amino acids larger than Bs-PFK (320 amino acids). Among the proteins listed in the bank, two showed homology with the truncated sequence of the ORF VI translational product: PYK from *Saccharomyces cerevisiae* (35.0% homology) and a similar enzyme from chicken muscle (40.0% homology) (7, 18) (Fig. 6D). ORF VI may thus correspond to the *pyk* gene encoding the *S. citri* PYK. The locations of the identified genes on the spiroplasmal DNA insert are indicated in Fig. 4.

Codon usage. The codon usage for spiralin was determined from the nucleotide sequence of its gene (Table 3). A preferential use of A- and T-rich codons was observed, especially when A or T occurred at the 3′ end of codons specifying the same amino acid; less than 8.0% of the codons (18 of 241) had a C or a G at the third position. For instance, among the 32 alanine codons, only 2 were terminated with G and none ended with C. Similar results applied to valine. All codons for phenylalanine, leucine, isoleucine, proline, threonine, glutamine, lysine, cysteine, and glycine were terminated by A or T. Among the six codons for leucine, those starting with T were preferentially used over those having a C at position 1 (11 TTC versus 2 CTT). Spiralin contains no tryptophan, and its gene lacks TGA as well as TGG triplets.

The preferred usage of codons with A or T in the first or third position as described above for the spiralin gene was also operative for the five other genes *rpsB*, *tsf*, *X*, *pfk*, and *pyk*. The genes for EF-Ts and protein X did not contain tryptophan codons and thus should be expressed entirely in *E. coli*. However the 5′ portion of the gene for ribosomal protein S2 had seven TGA codons (and one TGG codon), the *pfk* gene had two TGA codons, and the 5′ portion of the *pyk* gene had four TGA codons.

Regulatory signals. Four putative promoterlike sequences were identified (Fig. 7). Their positions on the spiroplasmal DNA insert are shown on Fig. 4.

Upstream of the spiralin gene, the sequence TGTAAT (Fig. 3, nucleotides 2477 to 2482) was only one nucleotide short of the consensus sequence of the −10 region (TATAAT) characteristic of promoters recognized by *E. coli* RNA polymerase carrying sigma factor σ^{70}, or the *B. subtilis* enzyme functioning with σ^{43} (8, 25). At the corresponding −35 region, the sequence TGTTATTT (Fig. 3, nucleotides 2452 to 2459) had the first four nucleotides of the consensus sequence TGTTGACA. An A+T-rich region occurred at −45. The transcription starting point (+1) could be the adenine residue (nucleotide 2493) of the sequence CAT, also

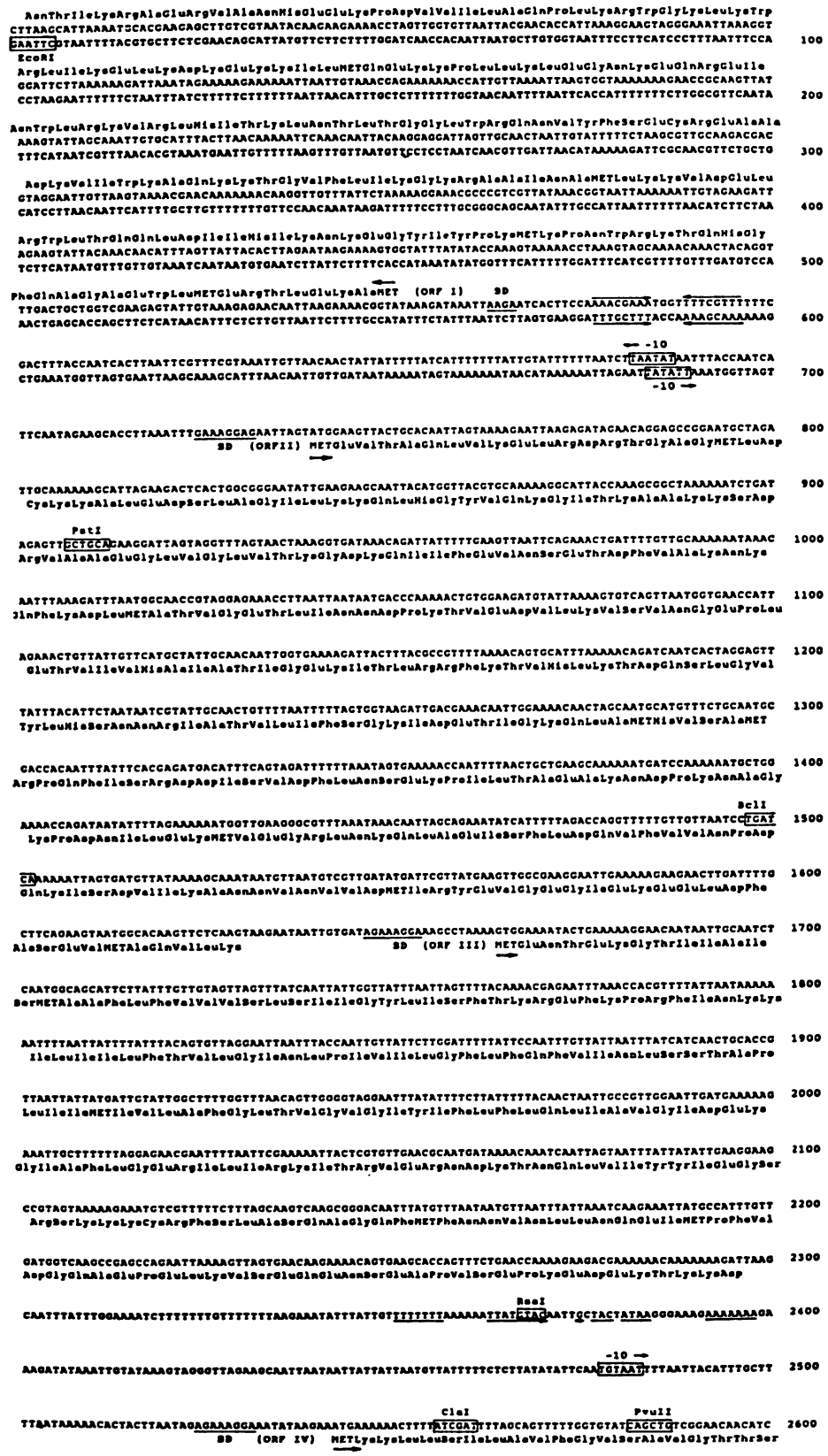

FIG. 3. Nucleotide sequence of the spiroplasmal DNA insert of plasmid pES3′ and amino acid sequence of corresponding ORFs. The sequence of the first 700 nucleotides is given for both DNA strands. Restriction sites are boxed. For each ORF, the direction of transcription is indicated by small arrows. The −10 regions of promoters are boxed. The ribosome-binding sites (SD) are underlined. Regions of inverted repeat are indicated by underlining arrows.

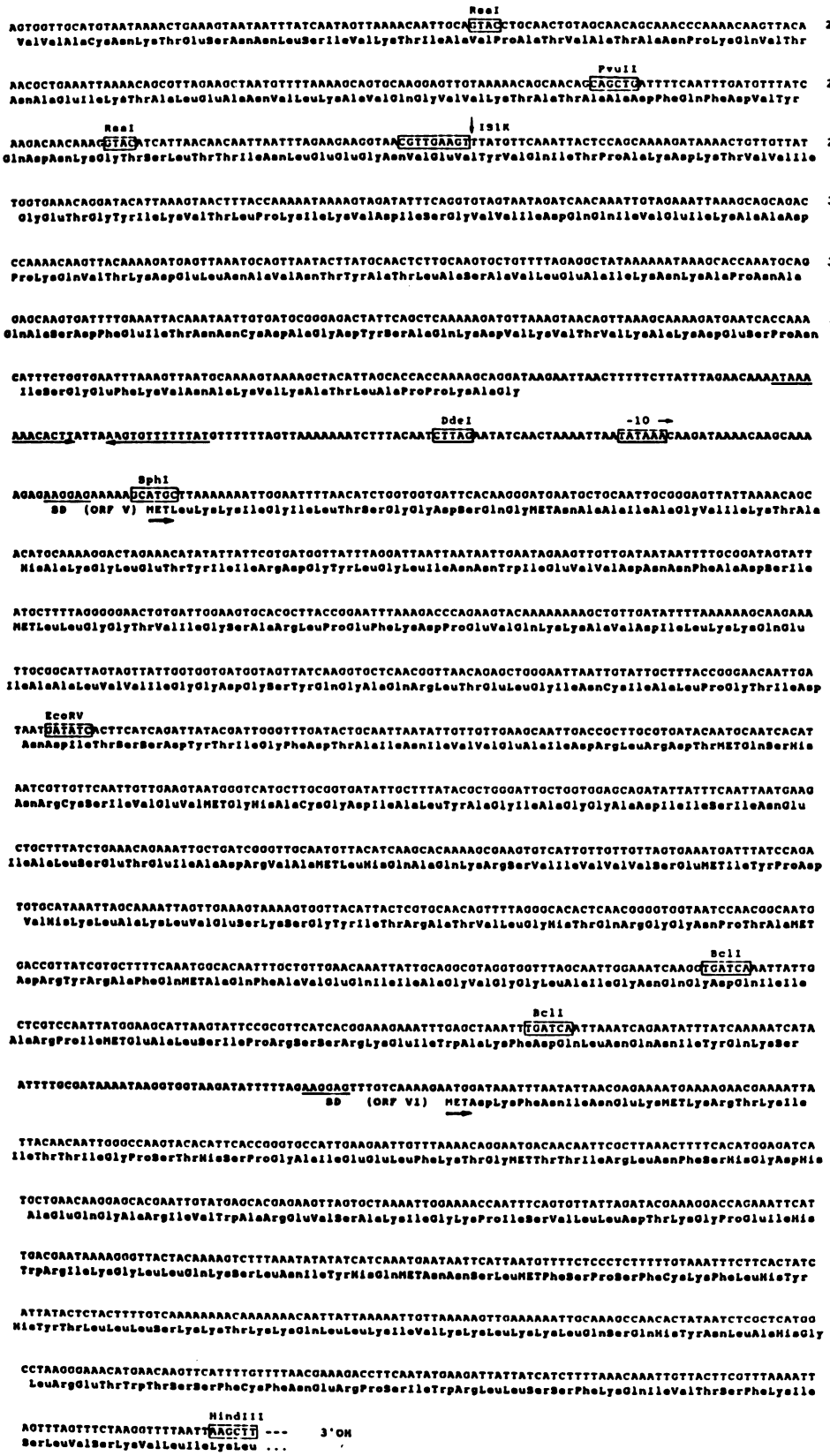

FIG. 3—*Continued.*

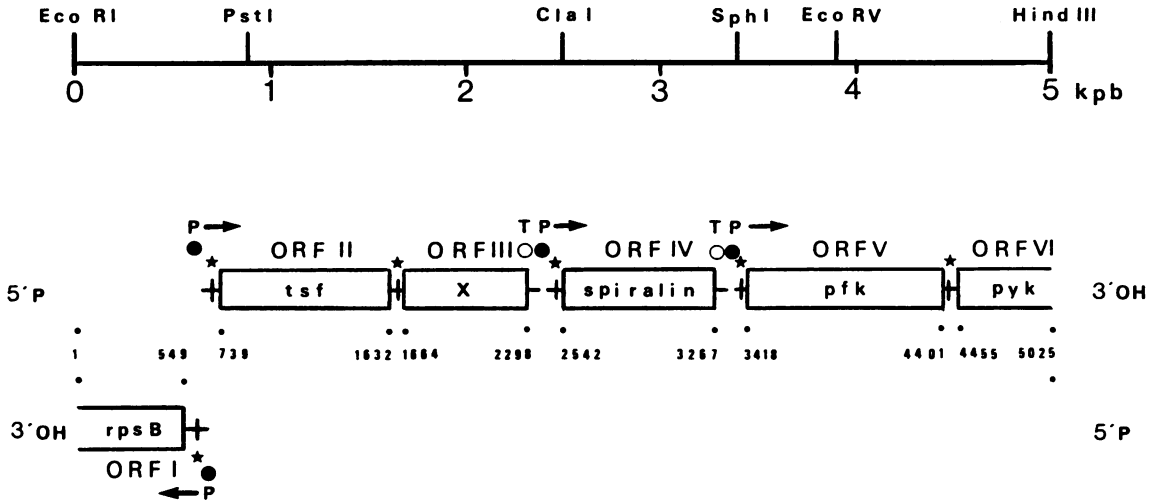

FIG. 4. Localization of ORFs and organization of corresponding genes on the *S. citri* DNA insert of plasmid pES3'. For each ORF, the direction of transcription is indicated by an arrow and the numbers of the first and last nucleotides are given. Promoters (P) and terminators (T) are represented by solid and open circles, respectively. Ribosome-binding sites are indicated by stars.

a consensus sequence, but less conserved than the −10 and −35 regions.

The promoter function of sequences composing the −10 and −35 regions of the putative spiralin promoter was examined by cloning the *Rsa*I-*Rsa*I restriction fragment (Fig. 3, nucleotides 2368 to 2660) into the *Sma*I site of the promoter selection vector pKK232.8. Insertion of a functional promoter sequence into *Sma*I, one of the multiple cloning sites of pKK232.8, resulted in expression of the chloramphenicol acetyltransferase gene and consequent chloramphenicol resistance of the relevant bacterial transformant.

Cloning the 292-bp-long *Rsa*I fragment into the *Sma*I site yielded 15 chloramphenicol-resistant transformants, all carrying the *Rsa*I insert in the same orientation (data not shown). The linearized vector, when treated in the same way but in the absence of foreign DNA, did not give rise to chloramphenicol-resistant transformants. Since the only promoterlike sequence present on the *Rsa*I restriction fragment is that extending from nucleotides 2452 to 2482, these results indicate that it is probably this sequence which is recognized as a promoter, at least in *E. coli*.

Upstream of the coding sequences of the spiroplasmal *rpsB*, *tsf*, and *pfk* genes, three respective promoterlike sequences were present (Fig. 4). The −10 sequences of these putative promoters were close to the bacterial consensus sequence (Fig. 7). The −35 regions were more poorly conserved with respect to the −35 consensus sequence TTGACA, but there was in all three an A+T-rich region at −43.

Two DNA sequences that could lead to transcription terminators on the mRNA were found downstream of the spiralin gene (Fig. 3, inverted repeat around nucleotide 3310) and the gene for protein X (Fig. 3, inverted repeat around nucleotide 2372). The putative terminator of the spiralin gene was followed by a row of uridine residues and could thus be independent of a termination factor such as rho (27). The stem of the protein X gene terminatorlike structure had an internal loop and ended with a row of seven U-A pairs, without a series of uridine residues at the 3' end.

The intergenic region between the termination codon of the *tsf* gene and the initiation codon of the adjacent protein X gene had only 31 nucleotide residues. The SD sequence for the protein X gene was the only remarkable sequence within this region. The *tsf* gene and the protein X gene are thus combined in a single transcription unit; transcription probably begins at the promoter located upstream of the *tsf* gene and ends at the terminator downstream of the protein X gene. Similarly, the *pfk* and *pyk* genes were also linked together in a single transcription unit. Whether this transcription unit contains only *pfk* and *pyk* or additional genes remains to be determined. However, the transcription prod-

TABLE 1. ORFs and putative proteins of *Spiroplasma citri* DNA insert of plasmid pES3' and homologous genes and proteins from eubacteria and eucaryotes

ORF	Location on spiroplasmal DNA		Size of putative spiroplasmal protein (no. of amino acids)	Homologous protein (origin)	Size (no. of amino acids)	% Homology	Corresponding gene
	First nucleotide	Last nucleotide					
I	549	1[a]	>183	Ribosomal protein S2 (*E. coli*)	240	38	*rpsB*
II	739	1632	297	Elongation factor Ts (*E. coli*)	283	46	*tsf*
III	1664	2298	211				
IV	2542	3267	241				
V	3418	4401	327	PFK (*B. stearothermophilus*)	320	45	*pfk*
VI	4455	5025[a]	>192	PYK (*S. cerevisiae*)	499	35	*pyk*
				PYK (chicken muscle)	530	40	

[a] The genes are truncated. ORF I is read on one DNA strand (minus strand), while ORFs II, III, IV, V, and VI are read on the other strand (plus strand).

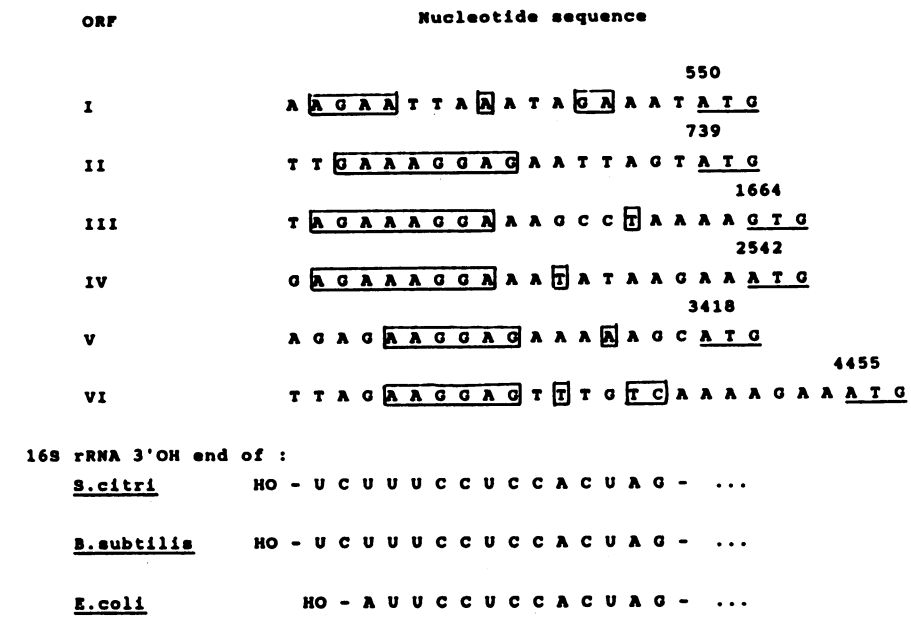

FIG. 5. SD sequences associated with the six ORFs on the spiroplasmal DNA insert of plasmid pES3′. Upstream of the initiation codon, the six ORFs have sequences that are complementary to the 3′ end of 16S rRNAs of *S. citri*, *B. subtilis*, and *E. coli*. These sequences are indicated within boxes. The initiation codons are underlined.

ucts of these two genes probably belong to the 4-kb mRNA that was revealed by Northern (RNA) blot analysis previously described (6).

A third inverted repeat (around nucleotide 585) was located between the promoter and the initiation codon of the *rpsB* gene, 86 bp downstream of the promoter or 28 bp upstream of the initiation codon ATG.

Insertional mutation of the spiralin gene. After transformation of *E. coli* by plasmid pES3′, a clone was obtained that lacked spiralin expression. The DNA insert of the recombinant plasmid harbored by this clone was found to be about 1,000 nucleotides larger than the insert present in pES3′. Sequencing showed that the spiralin gene of this plasmid was interrupted by the insertion element variant IS*1*K (15), and thus this clone represents an insertional mutation of the spiralin gene. As expected, this insertion resulted in a duplication of the spiroplasmal sequence at the target site. The 768-bp sequence of IS*1*K was inserted in the mutated spiralin gene after the final T (Fig. 3, nucleotide 2858) of CGTTGAAGT; this sequence was repeated at the end of IS*1*K.

TABLE 2. Amino acid composition of spiralin

Amino acid	No./molecule[a]	% of total (mol%)	
		Sequence[a]	Protein[b]
Ala	32	13.28	12.84
Val	32	13.28	11.81
Lys	27	11.20	10.84
Thr	23	9.54	9.64
Asp	13	5.39	
Asn	17	7.05	13.62
Glu	13	5.39	
Gln	9	3.73	10.02
Ile	16	6.64	6.35
Leu	13	5.39	4.79
Ser	12	4.98	6.68
Gly	12	4.98	5.00
Pro	9	3.73	3.50
Tyr	5	2.07	2.45
Phe	5	2.07	1.83
Cys	2	0.83	0.55
Met	1	0.41	0
Arg	0	0	0
His	0	0	0
Trp	0	0	0

[a] Determined from nucleotide sequence.
[b] Determined from purified protein (37).

DISCUSSION

The full 5.0-kbp spiroplasmal insert of pES3′ containing the spiralin gene was sequenced. Six ORFs were found. The translational product of one of these (ORF IV) had a size and an amino acid composition very similar to those of spiralin purified from *S. citri* (37). The only discrepancy found was in the N-terminal methionine, since no methionine was detected in purified spiralin (37). It is known that methionine is cleaved from the N-terminal end of native peptides when it is followed by Ala, Ser, Gly, Pro, Thr, or Val as the second amino acid; but with Arg, Asn, Asp, Gln, Glu, Ile, Leu, Lys, or Met as the second amino acid, methionine is not cleaved (31). In the spiralin gene, the initiator methionine codon is followed by AAA, a codon for Lys, and according to the above rule, methionine should remain present. As spiralin does not have an N-terminal methionine, the rule apparently does not apply to *S. citri*.

In *Mycoplasma* spp. (38) and in *Spiroplasma* spp. (23, 24), TGA does not serve as a termination codon, but as a codon for tryptophan, in addition to the "universal" tryptophan codon TGG. The coding portion of the spiralin gene contains neither TGA nor TGG codons, in agreement with the absence of tryptophan in spiralin. The absence of TGA codons (read as tryptophan in *S. citri* but as "stop" in *E. coli*) accounts for the full expression of spiralin in the bacterial transformant.

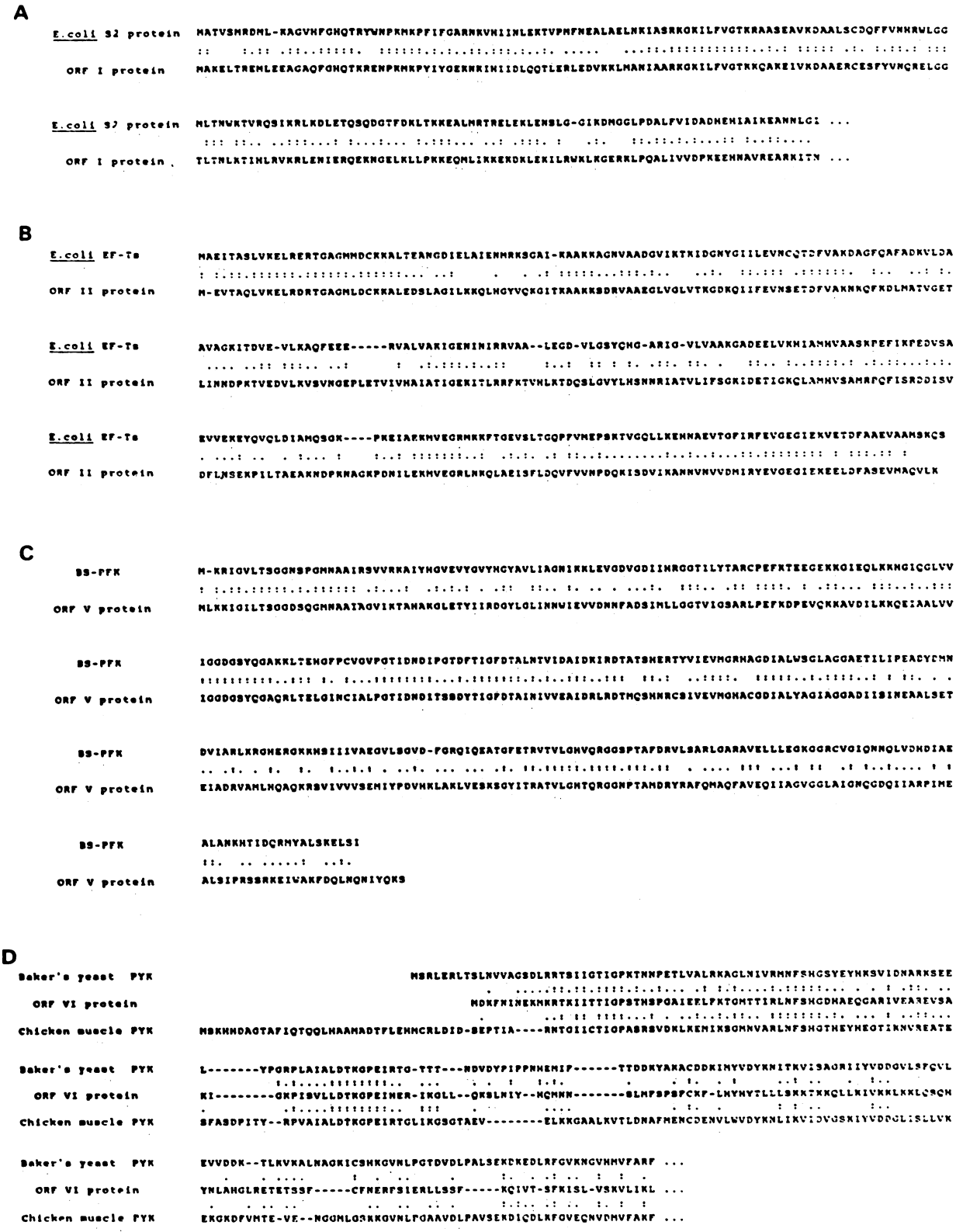

FIG. 6. Sequence homologies between products of the different spiroplasmal ORFs and known proteins from the data bank. Product of interrupted ORF I with *E. coli* ribosomal protein S2 (1) (A); product of ORF II with *E. coli* elongation factor Ts (1) (B); product of ORF V and *B. stearothermophilus* PFK (16) (C); product of ORF VI and PYK from bakers' yeast (7) and chicken muscle (18) (D).

TABLE 3. Codon usage for spiralin

1st base	2nd base								3rd base
	U	No.	C	No.	A	No.	G	No.	
U	Phe	5	Ser	1	Tyr	4	Cys	2	U
	Phe	0	Ser	0	Tyr	1	Cys	0	C
	Leu	11	Ser	7	Ter	1	Trp	0	A
	Leu	0	Ser	1	Ter	0	Trp	0	G
C	Leu	2	Pro	1	His	0	Arg	0	U
	Leu	0	Pro	0	His	0	Arg	0	C
	Leu	0	Pro	8	Gln	9	Arg	0	A
	Leu	0	Pro	0	Gln	0	Arg	0	G
A	Ile	12	Thr	7	Asn	12	Ser	3	U
	Ile	0	Thr	0	Asn	5	Ser	0	C
	Ile	4	Thr	16	Lys	27	Arg	0	A
	Met	1	Thr	0	Lys	0	Arg	0	G
G	Val	18	Ala	8	Asp	10	Gly	6	U
	Val	1	Ala	0	Asp	3	Gly	0	C
	Val	11	Ala	22	Glu	11	Gly	6	A
	Val	2	Ala	2	Glu	2	Gly	0	G

The codon usage for spiralin is very similar to that described for the capsid protein of SpV4 (24): a preferred usage of A (or T)-terminated codons is observed, reflecting the high A+T content (74%) of *S. citri* DNA.

In addition to ORF IV (the gene for spiralin), five other ORFs have been located on the spiroplasmal DNA insert of pES3′ (Fig. 4). Each of these six ORFs is characterized by an initiation codon, a termination codon, and an SD sequence, complementary to the 3′ end of 16S rRNA. In *S. citri*, the 3′ end of 16S rRNA is identical to that of *B. subtilis* (OH-UCUUUCCUCCACUAG) and is 2 bases longer than in *E. coli* (where the terminal OH-A replaces the OH-UCU of *S. citri*). These observations agree with the finding that gram-positive bacteria generally have SD sequences longer than those found in gram-negative bacteria (20), and this rule appears to apply to the SD sequences of ORFs II, III, and IV, where the base-pairing between the SD sequence and the 16S rRNA can occur over 8 bases in a row (Fig. 5). For ORFs V and VI, 6 bases are involved, but only 4 bases for ORF I. In all cases except ORF I, the characteristic sequence AAGGA occurs.

Sequence analysis software programs have suggested that the products of ORFs I, II, V, and VI correspond respectively to ribosomal protein S2, elongation factor Ts, PFK, and PYK. Identification of these components seems unambiguous, since only one candidate protein was given by the bank for each one of the ORFs. Homology between the candidate protein and the ORF translational products was moderately high (between 35 and 46%) (Table 1), even when the sequence comparisons involved eucaryotic proteins such as PYK from *S. cerevisiae* or chicken muscle. On the basis of these results, it is likely that ORFs I, II, V, and VI correspond respectively to *S. citri* genes *rpsB*, *tsf*, *pfk*, and *pyk*.

Four promoterlike sequences on the spiroplasmal DNA insert (Fig. 4) were identified upstream of ORFs I, II, IV, and V. The promoter of the spiralin gene was shown to be functional in an *E. coli* promoter selection vector. The spiroplasmal core RNA polymerase has the eubacterial subunit structure ($\alpha_2\beta\beta'$) (11), which recognizes spiroplasmal promoters with the consensus sequence TATAAT at −10 and TTGACA at −35 (6).

Two regions with inverted repeats, leading to possible transcription terminators, were found downstream of ORFs III and IV. Assuming that the four promoterlike sequences and the two terminatorlike sequences described above are functional, it follows (Fig. 4) that the spiralin gene (ORF IV) is transcribed from a monocistronic transcription unit but that the genes for PFK (ORF V) and PYK (ORF VI) are linked together as a single transcription unit. Similarly, *tsf* (ORF II) and the gene for the unidentified protein X (ORF III) represent a single transcription unit. Nothing can be said about the transcription unit of *rpsB* (ORF I). It is noteworthy that in *S. citri* the genes for two key enzymes of glycolysis, PFK and PYK, are part of the same transcription unit, which suggests that these enzymes must be produced at the same time and in similar amounts. In *E. coli* the situation is different. The *pfkA* gene is 48 min away from the *pyk* gene, with the *rpsB-tsf* transcription unit in between (2, 9). Recently, the presence of PFK and PYK in *S. citri* has been demonstrated (22). The PFK in *S. citri* is of the ATP-dependent form; no pyrophosphate-dependent PFK activity

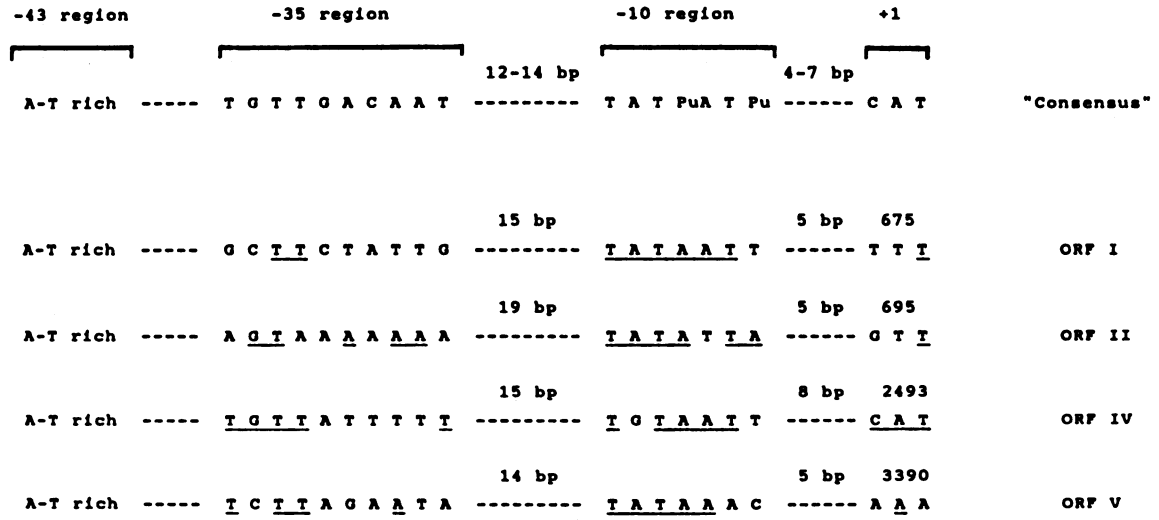

FIG. 7. Comparison of the four promoterlike sequences identified on the spiroplasmal DNA insert of plasmid pES3′ with a consensus promoter sequence (27). Underlined residues are those that agree with the consensus sequence. Pu, Purine.

was found. The PFK gene that we have identified in *S. citri* probably codes for the ATP-dependent PFK.

As shown in Fig. 4, *rpsB* (ORF I) and *tsf* (ORF II) are divergent genes with overlapping promoters (3). The results of this study provide the first description of divergent genes in mollicutes. The region between the promoter of *rpsB* (ORF I) and the ATG initiation codon contains an inverted repeat, 28 bp upstream of the ATG codon. This sequence is such that a protein dimer with dyad symmetry would be able to bind to the same side of the DNA helix, one monomer in contact with the direct sequence, the other with the inverted sequence. Hence the inverted repeat would represent a region where a polypeptide with regulatory functions could bind to control transcription in both directions. It would be interesting to investigate whether protein X may function as a regulatory protein that acts at the inverted repeat (within the divergent transcription unit) to control transcription of the nonregulatory S2 protein.

In *E. coli*, the genes for S2 (*rpsB*) and EF-Ts (*tsf*) are not divergent as they are in *S. citri*. They belong to a single transcription unit, with *rpsB* being promoter proximal (1, 3), and are transcribed in the same direction (12).

Currently, efforts are being directed towards understanding the organization of the spiralin molecule in the *Spiroplasma* membrane. The results of these studies will form the basis of a subsequent report.

ACKNOWLEDGMENTS

We express special thanks to F. Laigret for help in monitoring computer analysis. Thanks are due to J. Gandar for technical assistance in protein gel electrophoresis. We thank J. G. Tully and D. L. Williamson for critical review of the manuscript.

Comparisons with protein data banks were performed at the computer facilities at CITI2 in Paris on a DPS8 BULL computer with the help of the French Ministère de l'Education Nationale (programme mobilisateur "essor des biotechnologies").

LITERATURE CITED

1. **An, G., D. S. Bendiak, L. A. Mamelak, and J. D. Friesen.** 1981. Organization and nucleotide sequence of a new ribosomal operon in *E. coli* containing the genes for ribosomal protein S2 and elongation factor Ts. Nucleic Acids Res. **9**:4163–4172.
2. **Bachmann, B.** 1987. Linkage map of *Escherichia coli* K-12, p. 807–876. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
3. **Beck, C. F., and R. A. J. Warren.** 1988. Divergent promoters, a common form of gene organization. Microbiol. Rev. **52**:318–326.
4. **Bendiak, D. S., and J. D. Friesen.** 1981. Organization of genes in the four minute region of the *E. coli* chromosome: evidence that *rpsB* and *tsf* are co-transcribed. Mol. Gen. Genet. **181**:356–362.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
6. **Bové, J. M., P. Carle, M. Garnier, F. Laigret, J. Renaudin, and C. Saillard.** 1989. Molecular and cellular biology of spiroplasmas, p. 243–364. *In* R. F. Whitcomb and J. G. Tully (ed.), The mycoplasmas, vol. 5. Academic Press, Inc., New York.
7. **Burke, R. L., P. Tekamp-Olson, and R. Najarian.** 1983. The isolation, characterization, and sequence of the pyruvate kinase gene of *Saccharomyces cerevisiae*. J. Biol. Chem. **258**:2193–2201.
8. **Doi, R. H., and L. Wang.** 1986. Multiple procaryotic ribonucleic acid polymerase sigma factors. Microbiol. Rev. **50**:227–243.
9. **Fraenkel, D. G.** 1987. Glycolisis, pentose phosphate pathway and Entner-Dauderoff pathway, p. 142–150. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
10. **Fritensky, L., J. Lis, and R. Wu.** 1982. Portable microcomputer software for nucleotide sequence analysis. Nucleic Acids Res. **10**:6451–6463.
11. **Gadeau, A. P., C. Mouchés, and J. M. Bové.** 1986. Probable insensitivity of mollicutes to rifampin and characterization of spiroplasmal DNA-dependent RNA polymerase. J. Bacteriol. **166**:824–828.
12. **Grunberg-Manago, M.** 1987. Regulation of the expression of aminoacyl-tRNA synthetases and translation factors, p. 1386–1409. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
13. **Guo, L. H., and R. Wu.** 1983. Exonuclease III: use for DNA sequence analysis and in specific deletions of nucleotides. Methods Enzymol. **100**:60–96.
14. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
15. **Johnsrud, L.** 1979. DNA sequence of the transposable element IS*1*. Mol. Gen. Genet. **169**:213–218.
16. **Kolb, E., P. J. Hudson, and J. I. Harris.** 1980. Phosphofructokinase: complete aminoacid sequence of the enzyme from *Bacillus stearothermophilus*. Eur. J. Biochem. **108**:587–597.
17. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1440.
18. **Lonberg, N., and W. Gilbert.** 1983. Primary structure of chicken muscle pyruvate kinase mRNA. Proc. Natl. Acad. Sci. USA **80**:3661–3665.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
21. **Mouchés, C., T. Candresse, G. Barroso, C. Saillard, H. Wroblewski, and J. M. Bové.** 1985. Gene for spiralin, the major membrane protein of the helical mollicute *Spiroplasma citri*: cloning and expression in *Escherichia coli*. J. Bacteriol. **164**:1094–1099.
22. **Pollack, J. D., M. C. McElwain, D. DeSantis, J. T. Manolukas, J. G. Tully, C.-J. Chang, R. F. Whitcomb, K. J. Hackett, and M. V. Williams.** 1989. Metabolism of members of the *Spiroplasmataceae*. Int. J. Syst. Bacteriol. **39**:406–412.
23. **Renaudin, J., M.-C. Pascarel, C. Saillard, C. Chevalier, and J. M. Bové.** 1986. Chez les spiroplasmes le codon UGA n'est pas non-sens et semble coder pour le tryptophane. C.R. Acad. Sci. Paris Ser. III **303**:539–540.
24. **Renaudin, J., M.-C. Pascarel-Devilder, and J. M. Bové.** 1987. Spiroplasma virus 4 (SpV4): nucleotide sequence of the viral DNA, regulatory signals, and proposed genome organization. J. Bacteriol. **169**:4950–4961.
25. **Reznikoff, W. S., D. A. Siegele, D. W. Cowing, and C. A. Cross.** 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. **19**:355–387.
26. **Rodriguez, R. L., and R. C. Tait.** 1979. Recombinant DNA techniques: an introduction. Addison-Wesley Publishing Co., Boston.
27. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
28. **Saglio, P., M. Lhospital, D. Lafleche, G. Dupont, J. M. Bové, J. G. Tully, and E. A. Freundt.** 1973. *Spiroplasma citri* gen. and sp. nov.: a mycoplasma-like organism associated with "stubborn" disease of citrus. Int. J. Syst. Bacteriol. **23**:191–204.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **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.
31. **Tsunasawa, S., J. W. Stewart, and F. Sherman.** 1985. Amino-

terminal processing of mutant forms of yeast iso-1-cytochrome c. J. Biol. Chem. **260**:5382–5391.

32. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.
33. **Williamson, D. L., J. G. Tully, and R. F. Whitcomb.** 1989. The genus *Spiroplasma*, p. 71–107. *In* R. F. Whitcomb and J. G. Tully (ed.), The mycoplasmas, vol. 5. Academic Press, Inc., New York.
34. **Wroblewski, H.** 1978. Spiralin: its topomolecular anatomy and its possible function in the *Spiroplasma citri* cell membrane. Zentralbl. Bakteriol. **241**:179–180.
35. **Wroblewski, H., K. E. Johansson, and S. Hjerten.** 1977. Purification and characterization of spiralin, the main protein of the *Spiroplasma citri* membrane. Biochim. Biophys. Acta **465**: 275–289.
36. **Wroblewski, H., S. Nyström, A. Blanchard, and A. Wieslander.** 1989. Topology and acylation of spiralin. J. Bacteriol. **171**: 5039–5047.
37. **Wroblewski, H., D. Robic, D. Thomas, and A. Blanchard.** 1984. Comparison of the aminoacid compositions and antigenic properties of spiralins purified from the plasma membranes of different spiroplasmas. Ann. Microbiol. (Inst. Pasteur) **135A**: 73–82.
38. **Yamao, F., A. Muto, Y. Kawauchi, M. Iwami, S. Iwagami, Y. Azumi, and S. Osawa.** 1985. UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. USA **82**: 2306–2309.