JOURNAL OF BACTERIOLOGY, Nov. 1987, p. 4950–4961
0021-9193/87/114950-12$02.00/0

Vol. 169, No. 11

Spiroplasma Virus 4: Nucleotide Sequence of the Viral DNA, Regulatory Signals, and Proposed Genome Organization

JOËL RENAUDIN, MARIE-CLAIRE PASCAREL, AND JOSEPH-MARIE BOVÉ*

Laboratoire de Biologie Cellulaire et Moléculaire, I.N.R.A. et Université de Bordeaux II, Domaine de la Grande Ferrade, 33140 Pont de la Maye, France

Received 10 February 1987/Accepted 23 July 1987

The replicative form (RF) of spiroplasma virus 4 (SpV4) has been cloned in *Escherichia coli*, and the cloned RF has been shown to be infectious by transfection (M. C. Pascarel-Devilder, J. Renaudin, and J.-M. Bové, Virology 151:390–393, 1986). The cloned SpV4 RF was randomly subcloned and was fully sequenced by the dideoxy chain termination technique, using the M13 cloning and sequencing system. The nucleotide sequence of the SpV4 genome contains 4,421 nucleotides with a G+C content of 32 mol%. The triplet TGA is not a termination codon but, as in *Mycoplasma capricolum* (F. Yamao, A. Muto, Y. Kawauchi, M. Iwami, S. Iwagani, Y. Azumi, and S. Osawa, Proc. Natl. Acad. Sci. USA 82:2306–2309, 1985), probably codes for tryptophan. With these assumptions, nine open reading frames (ORFs) were identified. All nine are characterized by an ATG or GTG initiation codon, one or several termination codons, and a Shine-Dalgarno sequence upstream of the initiation codon. The nine ORFs are distributed in all three reading frames. One of the ORFs (ORF1) corresponds to the 60,000-dalton capsid protein gene. Analysis of codon usage showed that T- and A-terminated codons are preferably used, reflecting the low G+C content (32 mol%) of the SpV4 genome. The viral DNA contains two G+C-rich inverted repeat sequences. One could be involved in transcription termination and the other in initiation of cDNA strand synthesis. The SpV4 genome was found to contain at least three promoterlike sequences quasi-identical to those of eubacteria. These results fully support the bacterial origin of spiroplasmas.

Spiroplasma virus 4 (SpV4) is an isometric virus with single-stranded circular DNA that produces a lytic infection of the helical mollicute *Spiroplasma melliferum* (22). The 4.4-kilobase viral DNA is one of the smallest genomes of procaryotic DNA viruses. Possible bacterial equivalents of SpV4 are the coliphages G4 and ϕX174. The genomes of these phages are only slightly larger than that of SpV4 and code for at least 10 proteins. The SpV4 DNA might also code for a relatively large number of proteins despite its small size. Therefore, SpV4 DNA seemed to be a good candidate for nucleotide sequencing to obtain information on gene structure and regulatory signals in the spiroplasmas. Such data are still very scarce in the mollicutes in general and the spiroplasmas in particular. We have recently cloned the double-stranded replicative form (RF) of SpV4. The cloned RF was proved to be infectious by transfection, indicating that no sequences were lost during cloning (20). We report here the full sequence of the cloned SpV4 DNA. The viral genome has nine open reading frames (ORFs) provided that TGA is not taken as a termination codon. The regulatory signals are very similar to those of eubacterial sequences, in agreement with recent views on the phylogeny of the mollicutes, regarded as a coherent phylogenetic group deriving by regressive evolution from low-G+C gram-positive bacteria (26).

MATERIALS AND METHODS

Bacteria and bacteriophage. *Escherichia coli* HB101 was used for propagating recombinant plasmids containing SpV4 RF DNA, and *E. coli* TG1 was used as the host for bacteriophage M13. (*E. coli* HB101 and TG1 as well as phage M13mp8 were kindly supplied by S. Wain Hobson [Institut Pasteur, Paris].)

Enzymes and chemicals. Restriction endonucleases, DNA polymerase I (Klenow fragment), calf intestine phosphatase, T4 DNA ligase, isopropyl-β-D-thiogalactopyranoside, and 5-bromo-4-chloro-3-indoyl-β-galactoside (X-gal) were purchased from Boehringer GMb H (Mannheim, Federal Republic of Germany). A nick translation kit, an M13 sequencing kit, and the labeled nucleotides [α-^{32}P]dCTP (110 TBq/mmol) and [α-^{35}S]dATPαS (22 TBq/mmol) were purchased from Amersham Corp. (Arlington Heights, Ill.). *N,N'*-Methylene-bisacrylamide and acrylamide were obtained from Bio-Rad Laboratories (Richmond, Calif.). Urea was from E. Merck AG (Darmstadt, Federal Republic of Germany). Agarose and low-melting-point agarose were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.).

SpV4 RF DNA. Culture of *S. melliferum* G1 and propagation of SpV4 in this spiroplasma have been described previously (22). Purification of the SpV4 RF DNA and its cloning in *E. coli* have been described recently (20).

Shotgun cloning and dideoxy sequencing of SpV4 RF DNA with bacteriophage M13. SpV4 RF DNA fragments were randomly generated by sonication (4), cloned in *E. coli* TG1 after insertion into the RF of phage M13mp8 used as a vector (17), and sequenced by the dideoxy chain termination method (30) as follows. Recombinant plasmid pESV4-13 containing the full-size SpV4 RF inserted at the *Cla*I site of the *E. coli* plasmid vector pBR328 (20) was sonicated at 10 W for 45 s. The 400- to 800-base-pair fragments were purified by electrophoresis on a 1% low-melting-point agarose gel and

* Corresponding author.

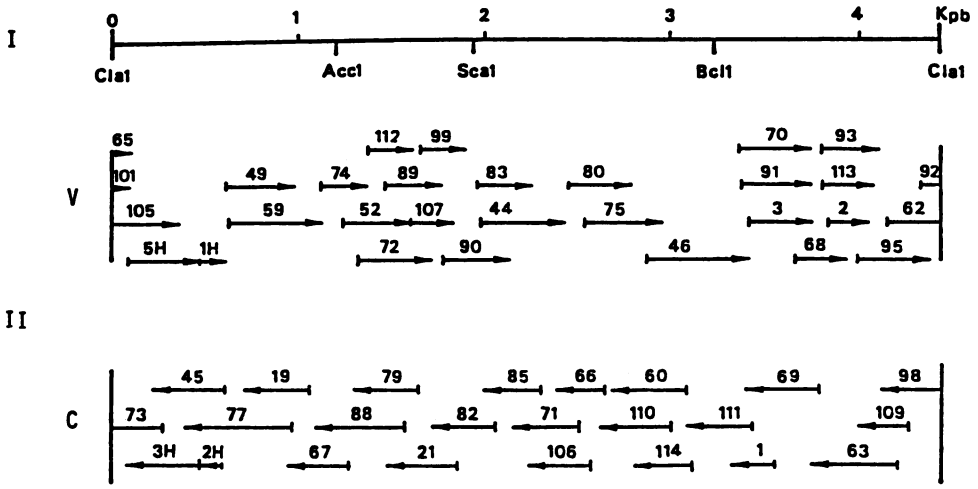

FIG. 1. Sequencing strategy. I, *Cla*I-linearized restriction map of SpV4 RF DNA with unique restriction sites. II, Directions, alignments, and numbers of the sequenced DNA fragments. V, Viral DNA strand; C, complementary strand. kbp, Kilobase pairs.

blunt ended by a fill-in reaction with DNA polymerase I (Klenow fragment) and all four deoxyribonucleotide 5'-triphosphates (1). The blunt-ended fragments were ligated to the dephosphorylated *Sma*I-linearized M13mp8 RF vector. The ligation mixture was used to transform *E. coli* TG1 cells by the method of Hanahan (10). Among the recombinant phage giving colorless plaques, those containing SpV4 DNA were further selected by in situ hybridization (15) with an SpV4-specific probe made by nick translation (25) of the SpV4 RF. A total of 114 hybridization-positive subclones were obtained.

In addition, two *Hin*cII restriction fragments (139 and 390 base pairs) of SpV4 RF were separately cloned in both orientations, using the same dephosphorylated *Sma*I-linearized M13mp8 RF vector.

Preparation of single-stranded DNA templates from the recombinant phages, annealing the forward 17-mer universal primer to templates, and sequencing reactions were performed following the *M13 Cloning and Sequencing Handbook* (1), except that for the sequencing reaction, concentrations of ddATP and ddTTP working solutions were lowered to 0.015 and 0.05 mM, respectively. [α-^{35}S]dATPαS (22 TBq/mmol) was used as the labeled nucleotide.

Sequencing gel electrophoresis. Sequencing reaction mixtures were loaded onto a 0.4-mm-thick, 50-cm-long polyacrylamide gel containing 7 M urea and 6.5% acrylamide in Tris-borate-EDTA buffer (pH 8.3). Electrophoresis was performed at 36 W constant power for 4 h (short run) or 8 h (long run). Gels were fixed in a mixture of 10% acetic acid and 10% methanol for 20 min before being dried under vacuum. They were autoradiographed overnight at room temperature with Du Pont Cronex 4 X-ray films.

Sequence analysis. Computer analysis of the nucleotide sequence was performed by using the alignment program NUCALN of Wilbur and Lipman (33) and the translational program NUMSEQ of Fristensky et al. (5). Hydropathy profiles of putative polypeptides were displayed by the method of Kyte and Doolittle (14).

Determination of NH_2-terminal amino acid sequence of SpV4 capsid protein. Proteins were purified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and electroblotted onto glass fiber sheets coated with Polybrene (32). The immobilized proteins were subjected to automatic gas-phase sequence analysis essentially as described by Hewick et al. (11).

RESULTS

Nucleotide sequence of SpV4 DNA. The double-stranded RF DNA was found to contain 4,421 base pairs. The DNA sequencing strategy is outlined in Fig. 1. Sequence data for both strands were obtained for all 4,421 base pairs with overlaps between the junctions. On the basis of hybridization experiments, it was shown that the II-V sequences of Fig. 1 correspond to the viral DNA strand of the RF and that the II-C sequences correspond to the complementary strand.

The nucleotide sequence of the single-stranded circular viral DNA (V strand) is indicated in Fig. 2. Since, for cloning purposes, the circular RF was initially linearized with restriction endonuclease *Cla*I (20) and since *Cla*I cuts the sequence 5'-ATCGAT-3' between T and C, C was chosen arbitrarily as nucleotide number 1 and T as nucleotide number 4,421.

The base composition of SpV4 DNA is 34% A, 33.9% T, 11.8% C, and 20.2% G. The G+C content is 32 mol%, slightly higher than that of the spiroplasma host DNA (26 mol%) (3).

Distribution of ORFs on SpV4 genome. The SpV4 DNA sequence was analyzed by the NUMSEQ translational program. Figure 3 summarizes the positions of termination codons on the viral DNA strand (V) and the complementary strand (C) in two cases (II and III): when all three termination codons are used (III) and when TAA and TAG but not TGA are represented (II). Only when termination codon TGA was omitted (II) could an ORF large enough to fit the 60,000-dalton capsid protein be identified on the viral DNA strand (II-V) in ORF2. The reason for not considering TGA as a termination codon derives from the finding of Yamao et al. (36) that in *Mycoplasma capricolum* TGA is not a termination codon, but codes for tryptophan (see Discussion). In the following results, the assumption will be made that TGA is not a termination codon, and only the ORFs corresponding to part II of Fig. 3 will be considered.

The positions of methionine codon ATG on the viral DNA strand (V) and the complementary strand (C) are represented as vertical bars in Fig. 4. The ORFs of Fig. 3, panel II, that

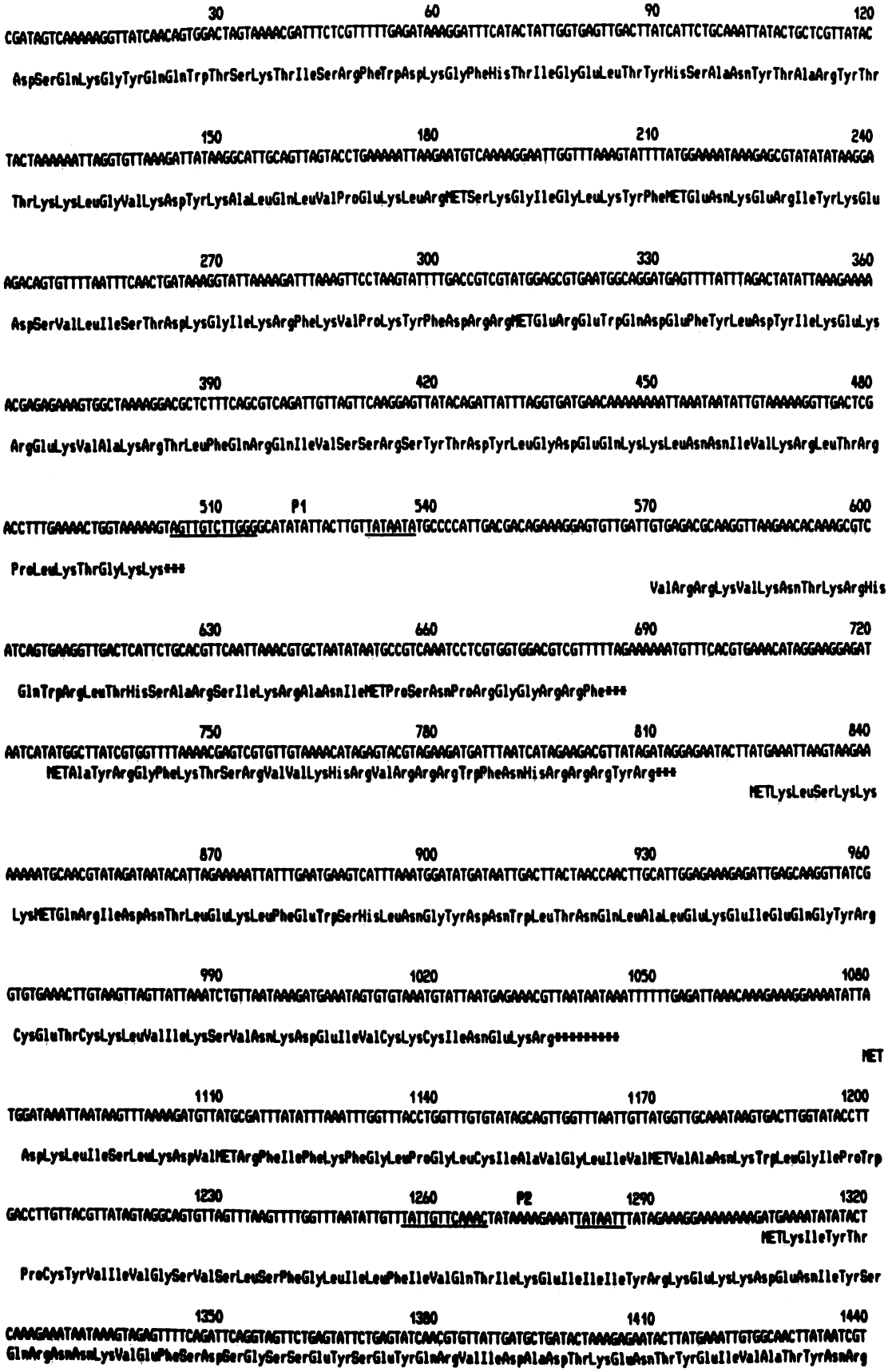

FIG. 2. Nucleotide sequence of SpV4 DNA and derived amino acid sequences of the putative polypeptides. The −10 and −35 regions of promoterlike sequences P1, P2, and P3 are underlined.

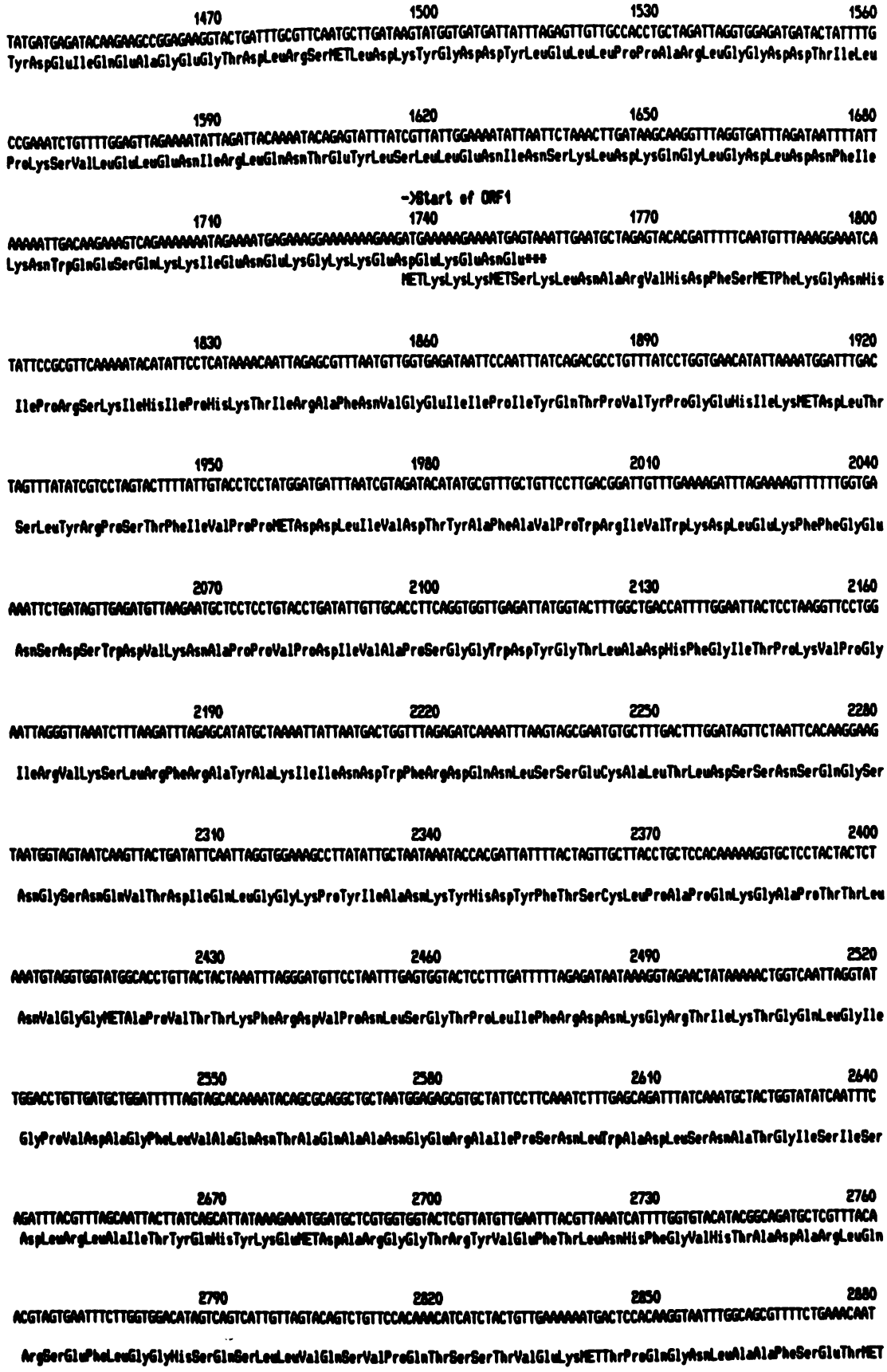

FIG. 2—(*Continued*)

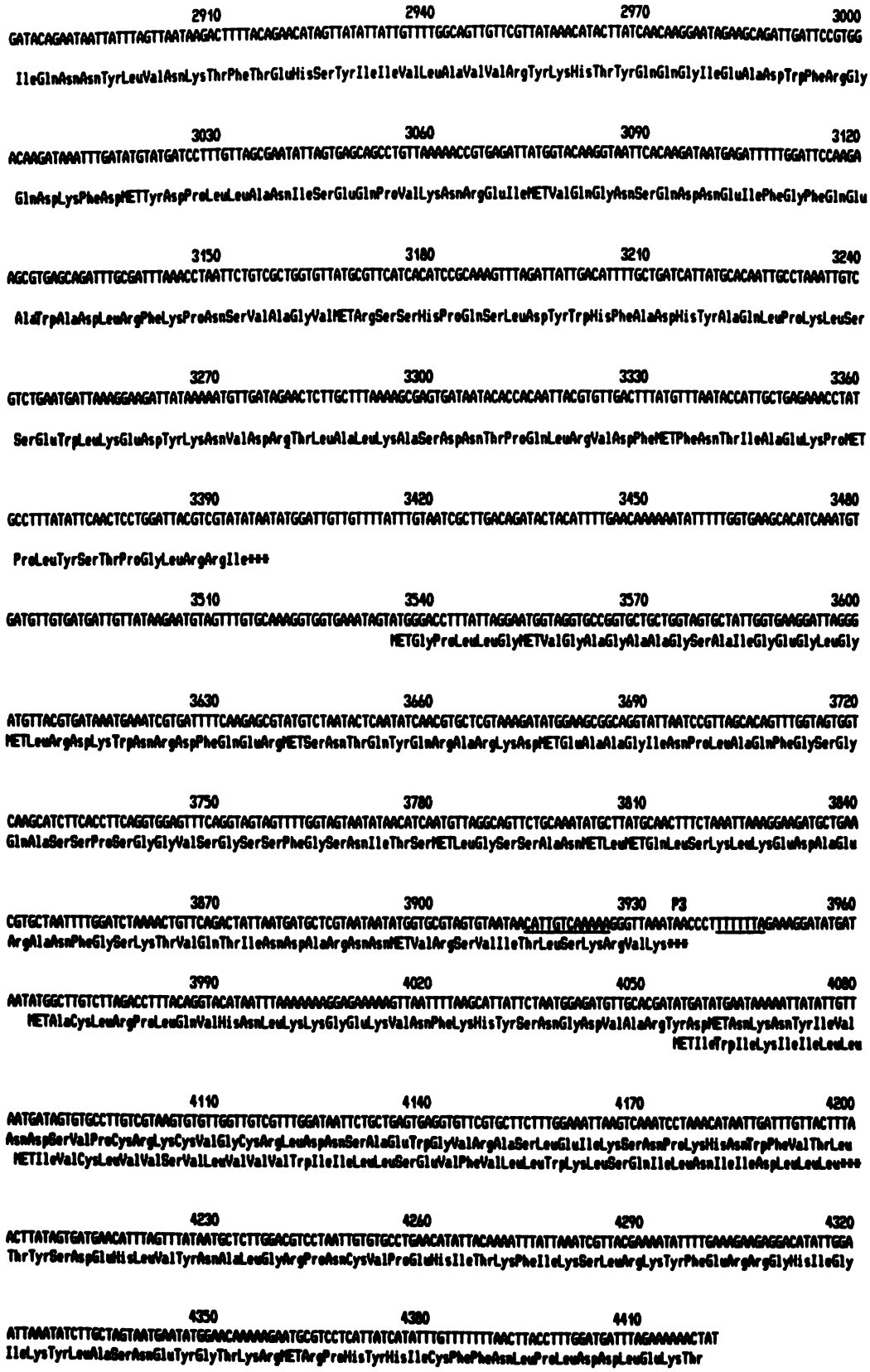

FIG. 2—(*Continued*)

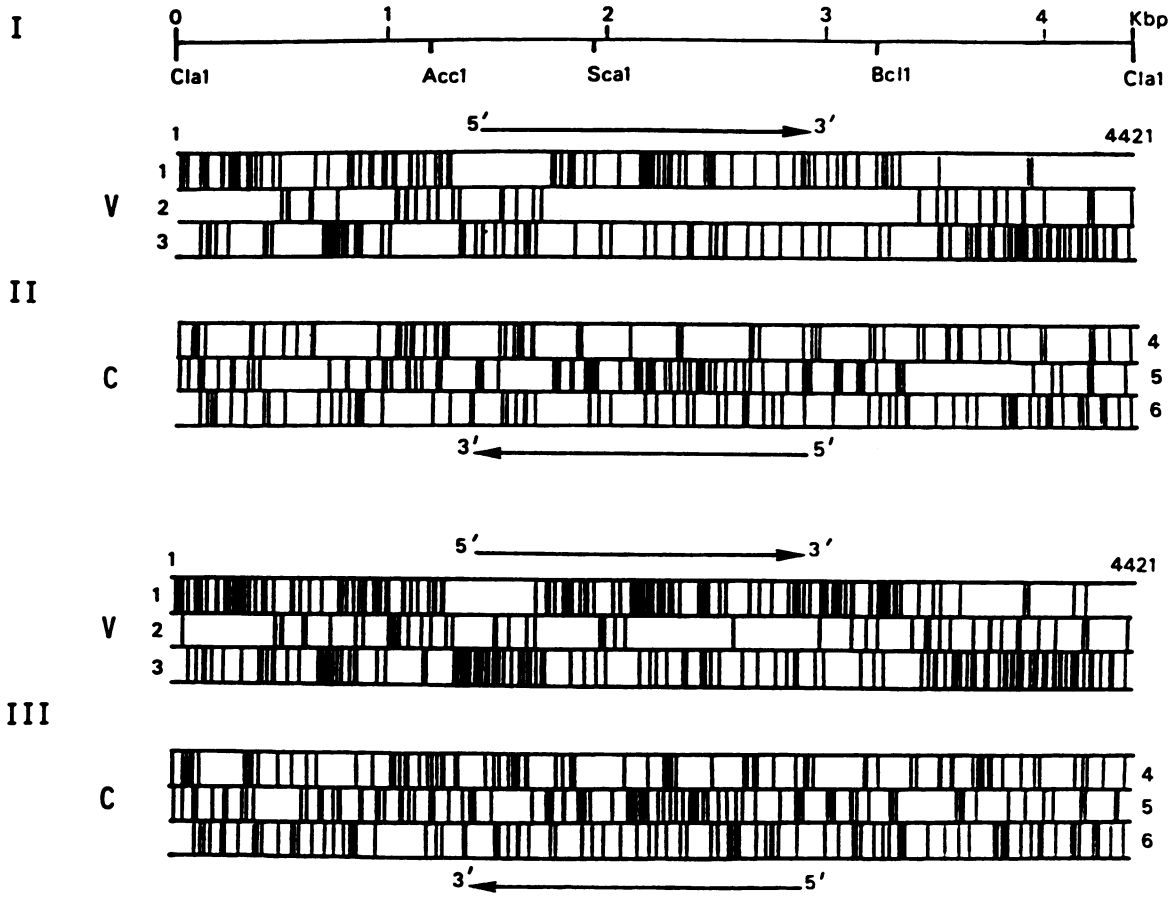

FIG. 3. Potential protein-coding regions. I, Same as in Fig. 1. II, Positions of termination codons TAA and TAG are shown as vertical bars in all three reading frames of the viral DNA strand (V: 1, 2, 3) and of the complementary strand (C: 4, 5, 6). III, Same as II, but in addition to TAA and TAG, termination codon TGA is also positioned. The reading frames are defined as follows: the first nucleotide of the first codon of frame 1 is nucleotide number 1; in frame 2 the first nucleotide is nucleotide number 2; in frame 3 it is nucleotide number 3. kbp, Kilobase pairs.

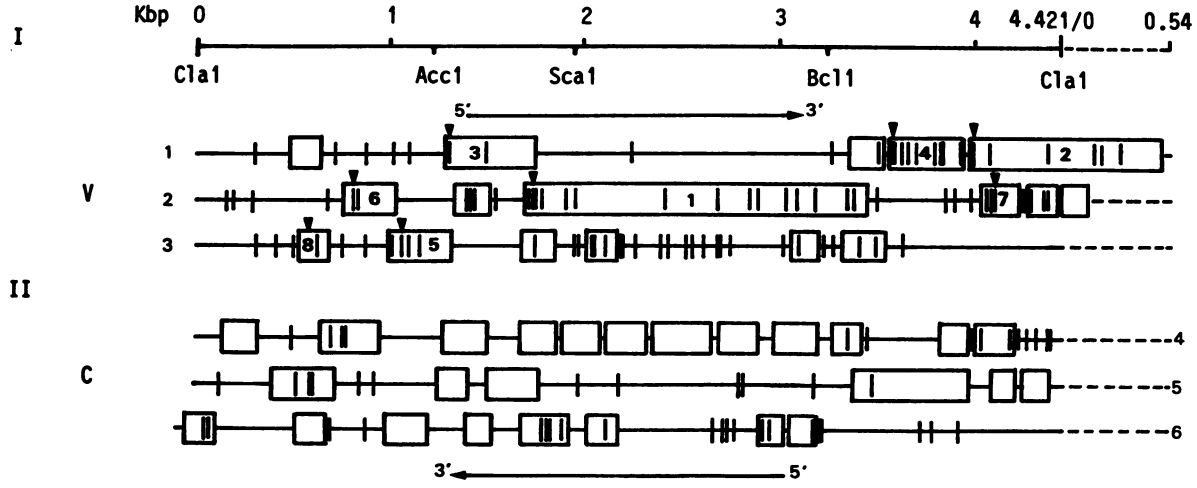

FIG. 4. Summary of methionine codons in reading frames 1, 2, and 3 of the viral DNA strand (V) and frames 4, 5, and 6 of the complementary strand (C). I, Same as in Fig. 1. Positions of the methionine codon are indicated by vertical bars. ORFs larger than 120 nucleotides are represented as boxes. The numbered boxes are ORFs with a ribosome-binding site upstream of the initiation codon, shown by an arrowhead. kbp, Kilobase pairs.

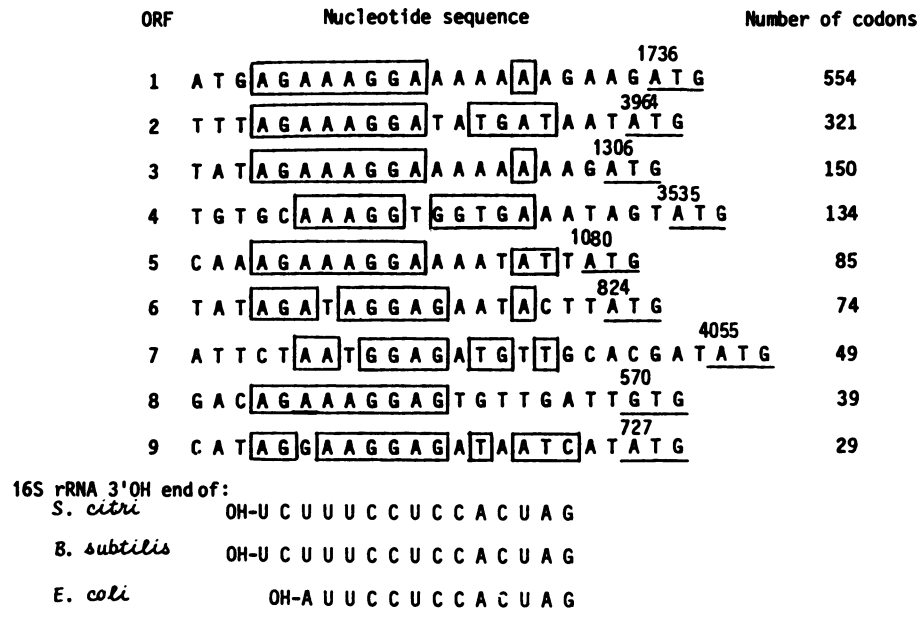

FIG. 5. Shine-Dalgarno sequences associated with ORFs of SpV4 DNA. Nine ORFs have, upstream of the initiation codon, sequences that are complementary to the 3'-OH end of 16S rRNA of *S. citri* (in boxes), *B. subtilis*, and *E. coli*. The initiation codon is underlined. The number of codons includes the termination codon(s).

are larger than 120 nucleotides are indicated as boxes in Fig. 4. Only some of these ORFs have an ATG initiation codon at the 5' end (arrowheads in Fig. 4), and almost all of these are located on the viral DNA strand (V).

A bacterial coding ORF possesses, at 5 to 10 nucleotides upstream of the initiation codon, a Shine-Dalgarno sequence complementary to the 3'-OH end of the 16S rRNA (31). This site, approximately six nucleotides long, often contains the sequence AGGA and is involved in ribosome binding. The nucleotide sequence at the 3'-OH end of *S. melliferum* 16S rRNA is not known. A search for Shine-Dalgarno sequences on the SpV4 DNA was done by comparing the sequences upstream to the initiation codons of the ORFs with the 3'-OH end of three others 16S rRNAs: that of *E. coli*, that of *Bacillus subtilis*, and that of *Spiroplasma citri* (Fig. 5). *S. citri* is serologically related to *S. melliferum*, and these two spiroplasmas have 65% DNA homology (2). The sequence of the 15 terminal nucleotides of *S. citri* 16S rRNA is the same as that of the following mollicutes and gram-positive bacteria: *M. capricolum*, *Mycoplasma* sp. strain PG50, *Acholeplasma laidlawii*, *Clostridium innocuum*, and *B. subtilis* (6, 12, 18, 34). Undoubtedly, this 3'-OH terminal sequence is highly conserved and differs from that of *E. coli* at the very 3'-OH end, 3'-UCU being replaced by 3'-A in *E. coli*. It is therefore very likely that the 3'-OH end of 16S rRNA of *S. melliferum* is identical to that of *S. citri*. Figure 5 shows the nine ORFs which have a Shine-Dalgarno sequence complementary to *S. citri* 16S rRNA. These nine ORFs have been numbered 1 to 9 from the largest to the smallest; they are indicated by their number in Fig. 4, except ORF9, which is shorter than 120 nucleotides. ORF1, the largest, has the size expected for the 60,000-dalton capsid protein, the only viral protein identified so far. The Shine-Dalgarno sequences are up to three nucleotides larger when evaluated against the 3'-UCU-terminated 16S rRNA of *S. citri* or *B. subtilis* than against the 3'-A-terminated 16S rRNA of *E. coli* (Fig. 5).

In summary, nine putative coding ORFs have been identified on the SpV4 viral DNA. Each ORF is bordered by an initiation codon and at least one termination codon, and possesses, upstream of the initiation codon, a Shine-Dalgarno sequence. ORF1 has the expected size for the 60,000-dalton capsid protein gene. The initiation codon of ORFs 1, 2, 3, 4, 5, 6, 7, and 9 is ATG, and that of ORF8 is GTG. The GTG codon of ORF8 is part of the sequence TTGTG (Fig. 5). The codon formed by the first three nucleotides (TTG) has been described as an initiation codon in gram-positive bacteria and phages (16). Hence, TTG could also be the start of an ORF, however, one that is shorter than ORF8 starting with GTG. The first termination codon is TAG for ORFs 2, 8, and 9 and TAA for the other six ORFs. ORFs 5 and 6 are terminated by two and three adjacent TAA

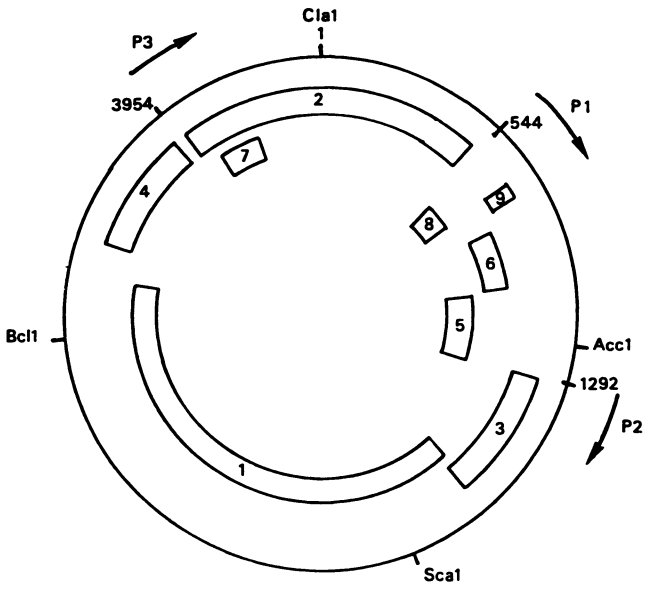

FIG. 6. Proposed SpV4 genome organization.

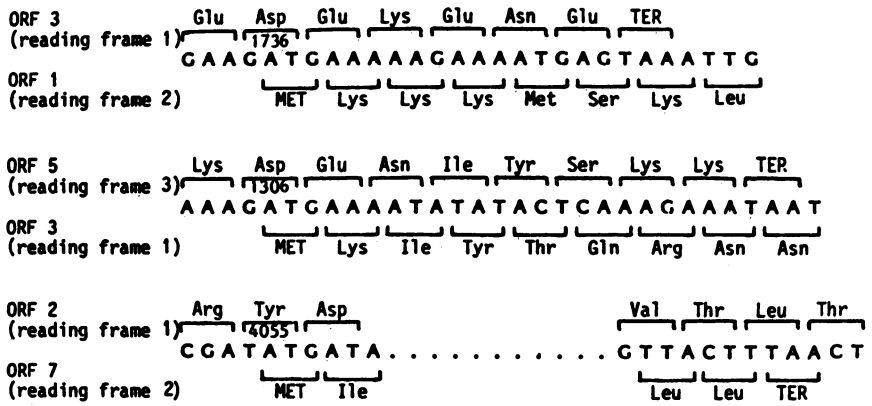

FIG. 7. Nucleotide and putative amino acid sequences of ORF overlapping regions.

codons, respectively. The nine putative coding ORFs involve all three reading frames (Fig. 6). There are three overlapping regions, between ORFs 5 and 3, 3 and 1, and 2 and 7. ORF2 fully overlaps ORF7. The nucleotide sequences of overlapping regions are shown in Fig. 7.

ORF of capsid protein. The size of the capsid protein of SpV4 was found to be 60,000 daltons by polyacrylamide gel electrophoresis (22). ORF1, which extends from nucleotide 1736 to nucleotide 3397, has a coding capacity of 63,900 daltons of protein and is the only ORF large enough to accommodate the capsid protein. The N-terminal amino acid sequence of the capsid protein, determined as described in Materials and Methods, was found to be Met-Lys-Lys-Lys-Met-Ser. . . . This is precisely the amino acid sequence predicted from the nucleotide sequence of ORF1 (Fig. 2). It is therefore very likely that ORF1 is indeed the gene for the SpV4 capsid protein.

Codon usage. The codon usage for the capsid protein has been determined from the nucleotide sequence of its gene (ORF1). For codons specifying the same amino acid and differing by only the nucleotide at position 3, those with A or T in the third position are used much more frequently than those terminated by C or G; those terminated by T are more frequently used than those ending with A (Table 1). Within the six codons for arginine, AGA and AGG starting with A are more frequently used than, respectively, CGA and CGG starting with C. A similar situation occurs with leucine, for which the two codons starting with T are preferably used over those starting with C. The preferred usage of codons having A or T in the first or third position as described above for ORF1 is also true for the other eight ORFs.

Hydropathy profiles. From their nucleotide sequences, we determined the hydropathy (14) curves of ORF1 (capsid protein) and the other eight ORFs. The hydropathy profiles for the putative proteins of ORFs 2, 3, 4, 6, and 8 show alternating hydrophilic and hydrophobic regions. The capsid protein profile is similar but shows a pronounced hydrophobic peak in the region corresponding to nucleotide 2940. The putative polypeptides corresponding to ORFs 7 and 9 are peculiar in that ORF7 has only hydrophobic regions, while ORF9 is fully hydrophilic. For ORF5, the profile shows two large hydrophobic regions and a small hydrophilic region at the C-terminal end.

Regulatory signals. Three promoterlike sequences (P1, P2, P3) were identified (Fig. 8). Their positions on the SpV4 genome are indicated in Fig. 6. Sequence P1 is close to the consensus sequence of bacterial promoters, recognized by *E. coli* RNA polymerase carrying sigma factor σ70 or the *B. subtilis* holoenzyme with sigma factor σ43 (24). It has a perfect Pribnow (TATPuATPu) box (21) located at −10 of a CAT box. The A residue of CAT could be the start (+1) of mRNA transcription. Also, the sequence of P1 at the −35 region is only two nucleotides short of the most conserved sequence in that region, namely, GTTGACA. Finally, at −43, there is an A+T-rich region, characteristic of a generalized promoter. Promoterlike sequence P2 has regions similar to the −10, −35, and −43 regions of the consensus bacterial promoter, but the initiation point of mRNA transcription does not involve a CAT box and remains ambiguous. The start of mRNA transcription from promoterlike sequence P3 is also ambiguous; the region at −35 has a good consensus sequence, but in the region at −10 the A's of the TATPuATPu box are replaced by T's. Preliminary Northern blot (RNA blot) analyses of viral mRNAs (data not shown)

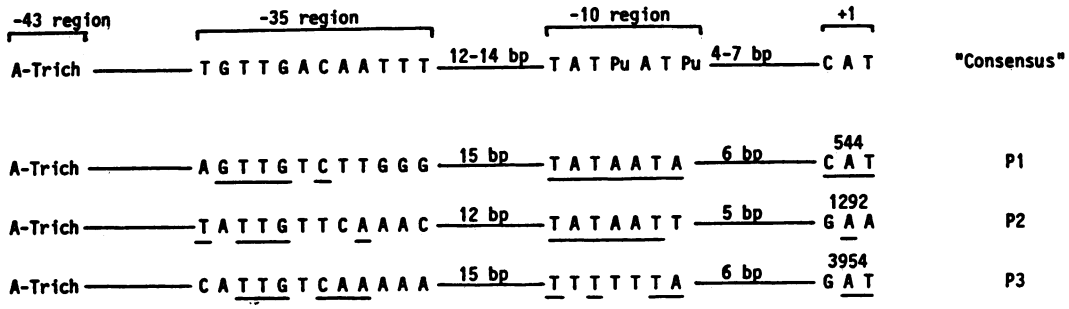

FIG. 8. Comparison of the three promoterlike sequences of SpV4 DNA (P1, P2, and P3) with a consensus promoter sequence (28). Underlined residues are those that agree with the consensus sequence. bp, Base pairs.

TABLE 1. Codon usage for SpV4 capsid protein

Codon	Usage (%)	Codon	Usage (%)
TTT-Phe	25 (4.5)	TAT-Tyr	20 (3.6)
TTC-Phe	2 (0.4)	TAC-Tyr	1 (0.2)
TTA-Leu	23 (4.2)	TAA-TER	1 (0.2)
TTG-Leu	14 (2.5)	TAG-TER	0 (0.0)
CTT-Leu	3 (0.5)	CAT-His	14 (2.5)
CTC-Leu	0 (0.0)	CAC-His	2 (0.4)
CTA-Leu	1 (0.2)	CAA-Gln	19 (3.4)
CTG-Leu	0 (0.0)	CAG-Gln	7 (1.3)
ATT-Ile	26 (4.7)	AAT-Asn	30 (5.4)
ATC-Ile	1 (0.2)	AAC-Asn	1 (0.2)
ATA-Ile	7 (1.3)	AAA-Lys	25 (4.5)
ATG-Met	14 (2.5)	AAG-Lys	7 (1.3)
GTT-Val	23 (4.2)	GAT-Asp	32 (5.8)
GTC-Val	1 (0.2)	GAC-Asp	3 (0.5)
GTA-Val	9 (1.6)	GAA-Glu	14 (2.5)
GTG-Val	0 (0.0)	GAG-Glu	6 (1.1)
TCT-Ser	8 (1.4)	TGT-Cys	1 (0.2)
TCC-Ser	0 (0.0)	TGC-Cys	1 (0.2)
TCA-Ser	14 (2.5)	TGA-Trp	9 (1.6)
TCG-Ser	1 (0.2)	TGG-Trp	1 (0.2)
CCT-Pro	28 (5.1)	CGT-Arg	15 (2.7)
CCC-Pro	0 (0.0)	CGC-Arg	0 (0.0)
CCA-Pro	5 (0.9)	CGA-Arg	1 (0.2)
CCG-Pro	2 (0.4)	CGG-Arg	1 (0.2)
ACT-Thr	23 (4.2)	AGT-Ser	16 (2.9)
ACC-Thr	1 (0.2)	AGC-Ser	1 (0.2)
ACA-Thr	7 (1.3)	AGA-Arg	8 (1.4)
ACG-Thr	3 (0.5)	AGG-Arg	2 (0.4)
GCT-Ala	20 (3.6)	GGT-Gly	23 (4.2)
GCC-Ala	0 (0.0)	GGC-Gly	0 (0.0)
GCA-Ala	12 (2.2)	GGA-Gly	13 (2.3)
GCG-Ala	7 (1.3)	GGG-Gly	0 (0.0)

seem to indicate that P1, P2, and P3 are functional but that P1 is more efficient. Two additional promoterlike sequences were localized around nucleotides 905 and 2117, respectively. No evidence for their involvement in transcription was obtained.

Two inverted repeat sequences were located on the SpV4 DNA, one around nucleotide 3932 and the other around nucleotide 528. The sequence at nucleotide 3932 is shown in Fig. 10 as double-stranded DNA; the inverted repeats, 11 nucleotides long, are underlined. The putative RNA transcript from this double-stranded DNA has the ability to form a hairpin structure (Fig. 10). This structure has a UUUUUUUA-3'-OH sequence, typical of a transcription terminator independent of factor rho, and could be involved in transcription termination. It should be noticed that, as shown in Fig. 10, the sequence around nucleotide 3932 is part of the promoterlike region P3 (Fig. 8). A similar situation has also been described for phage ϕX174, in which the promoter of gene A overlaps the main transcription terminator (9).

The inverted repeat sequence at nucleotide 528 is located in the untranslated region between ORF2 and ORF8 (Fig. 6) and can form a secondary structure as shown in Fig. 9. With seven G · C base pairs of which five are in a row, this hairpin structure should be quite stable. The hairpin structure of SpV4 is reminiscent of similar structures on the single-stranded DNAs of phages M13, G4, and ϕX174 (9). For M13 and G4, there is a unique origin of cDNA strand synthesis. The hairpin structure of M13 DNA, located in the untranslated region between genes II and IV, and that of G4 DNA, located in the intergenic space between genes F and G, are involved in the transcription of the RNA primer required for complementary strand synthesis. For ϕX174, the origins of complementary strand synthesis are multiple and almost randomly located. The hairpin structure between genes F and G is involved in the formation of the preprimosome (13).

DISCUSSION

With 4,421 nucleotide residues, the single-stranded circular DNA of *S. melliferum* SpV4 is one of the smallest procaryotic viral DNAs known. Despite its small size it seems to possess nine ORFs. The identification of these ORFs is based on the assumption that TGA is not a termination codon. No ORFs could be detected when all three

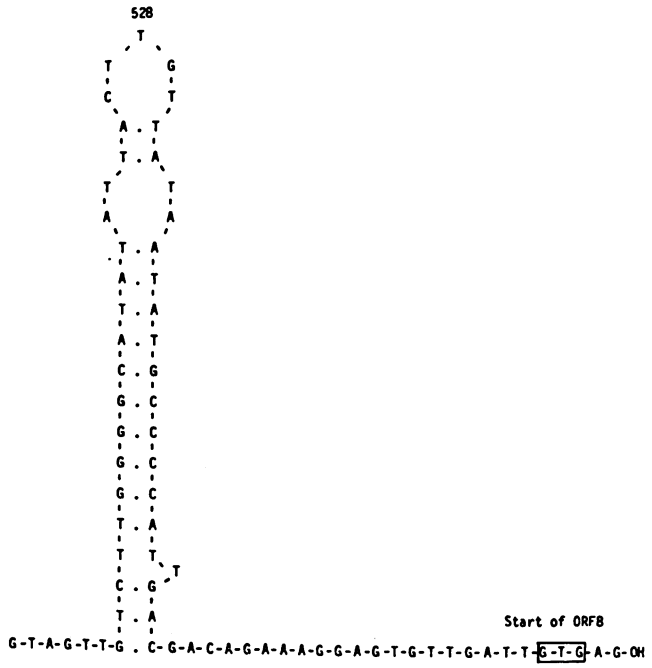

FIG. 9. Secondary structure of untranslated region between ORF2 and ORF8.

termination codons were considered as such. The finding that in the mollicute *S. melliferum* TGA does not seem to function as a termination codon is similar to that in the mollicute *M. capricolum* for which Yamao et al. (36) have shown that TGA codes for tryptophan. It is likely that in spiroplasmas, TGA also codes for tryptophan, as we have recently proposed (23) for the following reason. The gene for spiralin, the major membrane protein of *S. citri*, has been cloned in *E. coli* HB101, and the recombinant bacterial clone expresses spiralin (19). The replicative form of SpV4 has also been cloned in *E. coli* HB101 and found to be infectious in spiroplasmas by transfection (20), indicating that the cloned RF contained in particular a biologically active capsid protein gene. However, in the recombinant *E. coli* clones, expression of the capsid protein did not occur, whatever the cloning strategy used. These results are easily explained if it is assumed that TGA codes for tryptophan. Indeed, spiralin contains no tryptophan (35), hence no TGA codon, and can thus be expressed in the bacterium. On the contrary, the gene for the capsid protein (ORF1) has nine TGA codons (Fig. 2; Table 1) and cannot be fully expressed in *E. coli*, since the ribosomes will stop at the first UGA codon encountered on the mRNA. The absence of TGA codons from the spiralin gene has recently been confirmed by the sequence determination of this gene (C. Chevalier, C. Saillard, and J.-M. Bové, unpublished data).

The universal codon for tryptophan is TGG. The gene for the capsid protein contains one TGG codon besides the nine TGA codons. Under the assumption that TGA codes for tryptophan in spiroplasmas, the fact that there are nine TGA codons for one TGG codon is only a particular case of a more general phenomenon, namely, the preferred usage of A- (or T-) terminated codons in SpV4 ORFs (Table 1). The predominance of A- or T-terminated codons probably reflects the high A+T content of SpV4 DNA: 68 mol%, compared with 55.2% for ϕX174 DNA and 54.3% for G4 DNA. Even though SpV4 DNA contains as much T (33.9 mol%) as A (34 mol%), the codons terminated by T are more frequent than those ended by A. The high frequency of T in the third position was also noted for ϕX174 codons and to a lesser extent for G4 codons, but it is more pronounced for SpV4 DNA. The more frequent use of T rather than A as the last codon base allows more options for anticodons read by using wobble.

The promoterlike regions of the SpV4 DNA that were identified had sequences very similar to those of the bacterial promoters used by *E. coli* or *B. subtilis* RNA polymerases when σ70 or σ43 are the respective sigma factors. Also, each one of the nine viral ORFs has a bacterial Shine-Dalgarno sequence similar to that found in gram-positive bacteria (16).

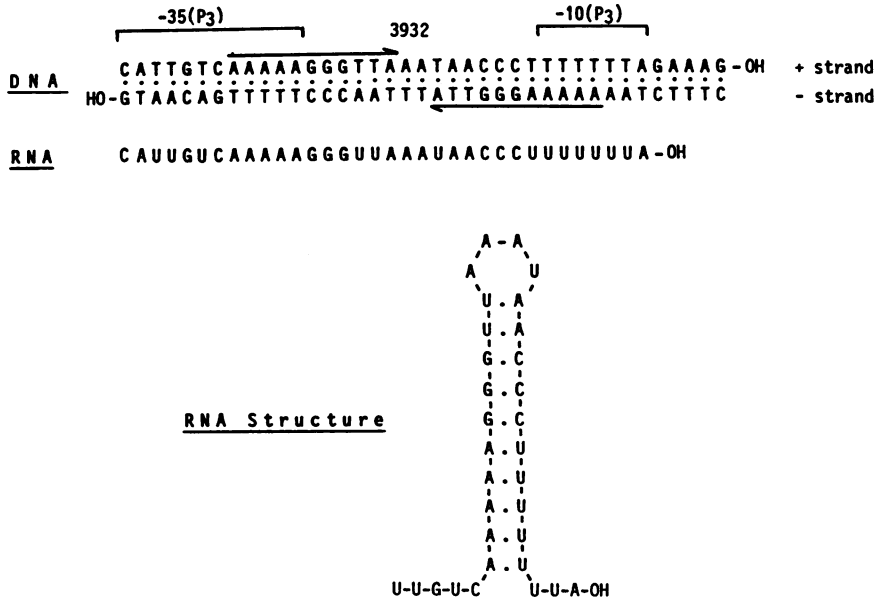

FIG. 10. DNA sequence and RNA structure of SpV4 putative transcription terminator.

It is quite probable that the host spiroplasma genome has the same bacteriumlike promoters and Shine-Dalgarno sequences as the SpV4 genome. This seems to be so since spiroplasma genes such as the spiralin gene can be expressed in *E. coli* from their own promoters (19). This implies that the same spiroplasma promoters are recognized by two RNA polymerases: the one from spiroplasma and the one from *E. coli*. This is not surprising, since we have recently shown that the RNA polymerase from *S. melliferum* has a subunit structure of the type $\beta\beta'\alpha_2$ (7), very similar to that of the eubacterial enzyme. In addition, Rogers et al. (27) have obtained from sequencing data direct evidence for the occurrence in *S. melliferum* of at least one bacteriumlike promoter upstream to a tRNA gene cluster.

The spiroplasmal virus SpV4 and the bacterial viruses G4 and ϕX174 have several similarities. The three viruses are isometric, lytic, and contain single-stranded circular DNA, of 4,421 nucleotides for SpV4 and 5,386 and 5,577 nucleotides, respectively, for ϕX174 and G4 (8, 29). Despite their small sizes, they code for approximately the same number of proteins, 11 for G4 and ϕX174 and 9 for SpV4. However, SpV4 has only one capsid protein, while G4 and ϕX174 have four. The genes or ORFs are located in all three reading frames with several overlapping regions.

The only identified SpV4 protein is the capsid protein. ORF1 has been identified as the gene for this protein. With 1,662 nucleotide residues, this gene represents more than one-third of the genome. The genes of the four capsid proteins of G4 and ϕX174 occupy more than one-half of the genome.

ACKNOWLEDGMENTS

We thank J. Vandekerckhove, M. Puype, and J. Van Damme for sequencing the NH_2-terminal end of SpV4 capsid protein.

LITERATURE CITED

1. **Amersham International plc.** 1984. M13 cloning and sequencing handbook. Amersham International plc., London.
2. **Bové, J.-M., C. Saillard, P. Junca, J. R. Degorce-Dumas, B. Ricard, A. Nhami, R. F. Whitcomb, D. Williamson, and J. G. Tully.** 1982. Guanine-plus-cytosine content, hybridization percentages and EcorI restriction enzyme profiles of spiroplasmal DNA. Rev. Infect. Dis. **4**:5129–5135.
3. **Clark, T. B., R. F. Whitcomb, J. G. Tully, C. Mouchès, C. Saillard, J.-M. Bové, H. Wroblewski, P. Carle, D. L. Rose, R. B. Henegar, and D. Williamson.** 1985. *Spiroplasma melliferum*, a new species from the honeybee (*Apis mellifera*). Int. J. Syst. Bacteriol. **35**:296–308.
4. **Deininger, P. L.** 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. **129**:219–223.
5. **Fristensky, B., J. Lis, and R. Wu.** 1982. Portable micro computer software for nucleotide sequence analysis. Nucleic Acids Res. **10**:6451–6463.
6. **Frydenderg, J., and C. Christiansen.** 1985. The sequence of 16S rRNA from mycoplasma strain PG50. DNA **4**:127–137.
7. **Gadeau, A. P., C. Mouchès, and J.-M. Bové.** 1986. Probable insensitivity of mollicutes to rifampin and characterization of spiroplasmal DNA-dependent RNA polymerase. J. Bacteriol. **166**:824–828.
8. **Godson, G. N., B. G. Barrell, R. Staden, and J. C. Fiddes.** 1978. Nucleotide sequence of bacteriophage G4 DNA. Nature (London) **276**:236–247.
9. **Godson, G. N., J. C. Fiddes, B. G. Barrell, and F. Sanger.** 1978. Comparative DNA sequence analysis of the G4 and ϕX174 genomes, p. 51–83. *In* D. T. Denharat, D. Dressler, and D. S. Ray (ed.), The single stranded DNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
11. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256**:7990–7997.
12. **Iwami, M., A. Muto, F. Yamao, and S. Osawa.** 1984. Nucleotide sequence of rrnB 16S ribosomal RNA gene from *Mycoplasma capricolum*. Mol. Gen. Genet. **196**:317–322.
13. **Kornberg, A. (ed.).** 1982. Supplement to DNA replication. W. H. Freeman & Co., San Francisco.
14. **Kyte, J., and R. F. Doolitte.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **McLaughlin, J. R., C. L. Murray, and D. L. Thomas.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
17. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
18. **Moran, C. P., L. Naomi, S. F. J. Legrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**:339–346.
19. **Mouchès, 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.
20. **Pascarel-Devilder, M. C., J. Renaudin, and J.-M. Bové.** 1986. The spiroplasma virus 4 replicative form cloned in *Escherichia coli* transfects spiroplasmas. Virology **151**:390–393.
21. **Pribnow, D.** 1975. Nucleotide sequence of an RNA polymerase binding site at an early T4 promoter. Proc. Natl. Acad. Sci. USA **72**:343–361.
22. **Renaudin, J., M. C. Pascarel, M. Garnier, P. Carle-Junca, and J.-M. Bové.** 1984. SpV4, a new spiroplasma virus with circular, single stranded DNA. Ann. Virol. **135E**:343–361.
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. **Reznikoff, W. S., D. A. Siegele, D. W. Cowing, and C. A. Gross.** 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. **19**:355–387.
25. **Rigby, P. W. J., M. Dieckmann, L. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237.
26. **Rogers, M. J., J. Simmons, R. T. Walker, W. G. Weisburg, C. R. Woese, R. J. Tanner, I. M. Robinson, D. A. Stahl, G. Olsen, R. H. Leach, and J. Maniloff.** 1985. Construction of the mycoplasma evolutionary tree from 5s rRNA sequence data. Proc. Natl. Acad. Sci. USA **82**:1160–1164.
27. **Rogers, M. J., A. A. Steinmetz, and R. T. Walker.** 1986. The nucleotide sequence of a tRNA gene cluster from *Spiroplasma melliferum*. Nucleic Acids Res. **14**:3145.
28. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
29. **Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchinson III, P. M. Slocombe, and M. Smith.** 1977. Nucleotide sequence of bacteriophage ϕX174 DNA. Nature (London) **265**:68.
30. **Sanger, F., S. Nicklenm, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **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**:1343–1346.

32. **Vandekerckhove, J., G. Bauw, M. Puype, J. Van Damme, and M. Van Montagu.** 1985. Protein-blotting on Polybrene-coated glass-fiber sheets. Eur. J. Biochem. **152**:9–19.
33. **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.
34. **Woese, C. R., J. Maniloff, and L. B. Zablen.** 1980. Phylogenetic analysis of the mycoplasmas. Proc. Natl. Acad. Sci. USA **77**:494–498.
35. **Wroblewski, H., D. Robic, D. Thomas, and A. Blanchard.** 1984. Comparison of the amino acid compositions and antigenic properties of spiralins purified from the plasma membranes of different spiroplasmas. Ann. Microbiol. (Paris) **135A**:73–82.
36. **Yamao, F., A. Muto, Y. Kawauchi, M. Iwami, S. Iwagani, Y. Azumi, and S. Osawa.** 1985. UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. USA **82**: 2306–2309.