A small RNA of *Mycoplasma capricolum* that resembles eukaryotic U6 small nuclear RNA

Chisato Ushida and Akira Muto*
Department of Biology, Faculty of Science, Hirosaki University, Bunkyo-cho 3, Hirosaki 036, Aomori, Japan

Received February 1, 1993; Revised and Accepted April 5, 1993

**ABSTRACT**

*Mycoplasma capricolum*, a parasitic prokaryote, contains several small stable RNAs, besides rRNAs and tRNAs. One of them, designated MCS4 RNA (125 nucleotides in length), has been isolated and sequenced. This RNA is abundant in the cell, and is encoded by two genes. Unexpectedly, MCS4 RNA has been found to reveal extensive sequence similarity to eukaryotic U6 snRNAs. This finding suggests that MCS4 and U6 snRNAs are derived from a common ancestral RNA that has existed before the divergence of prokaryotes and eukaryotes.

**INTRODUCTION**

In addition to rRNAs and tRNAs, diverse populations of RNA sequences have been found in bacterial cells (1). In *Escherichia coli*, there are several metabolically stable RNAs, such as M1 RNA (RNase P RNA) (2), 10Sa RNA (3), 6S RNA (4) and 4.5S RNA (5). M1 RNA is a catalytic part of a tRNA processing enzyme (6) and 4.5S RNA is suggested to be involved in translational process (7) or in membrane transport of proteins (8), while the functions of other RNAs are unknown. Several other low molecular weight RNAs, whose functions are unknown, have been also detected in *E. coli* (1, 9).

In this and previous studies, we have analyzed small stable RNAs of *Mycoplasma capricolum* (10). Mycoplasmas are small eubacteria, parasitic in eukaryotic tissues and organs. The genome size of mycoplasmas is 600 to 1,000 kilobase-pairs (kbp), the smallest of all known free-living organisms (for review, see refs. 11, 12). The *M. capricolum* genome contains only two sets of rRNA genes (13), 30 genes for 29 tRNA species (10, 14) and 400-500 genes for proteins (15). During our previous analysis of tRNAs and their genes of *M. capricolum*, we have found a new RNA species with 77 nucleotides in length (msRNA) (10). Samuelsson and Guindly (16) found a homologous sequence to msRNA in *M. mycoides*, a related strain of *M. capricolum*, and demonstrated that the RNA is a homolog of 4.5S RNA of *E. coli* (see also ref. 17). Simonneau and Hu (18) also found a homologous RNA in *M. pneumoniae*. Here, we report the structure of an another new RNA species, whose sequence resembles eukaryotic U6 small nuclear RNA (snRNA).

**MATERIALS AND METHODS**

**Isolation of small RNAs**

*M. capricolum* strain Kid (ATCC27343) was grown and harvested as previously described (14). Crude preparations of small RNAs were obtained by direct phenol extraction of the cells followed by selective NaCl and isopropanol precipitations (19). About 0.5 mg small RNAs were loaded on a 12% polyacrylamide—7 M urea gel (0.05×20×80 cm) made in TBE-buffer (89 mM Tris-borate, pH 8.3; 10 mM EDTA). Electrophoresis was performed in TBE buffer at 10 mA for 15 h. The gel was stained by ethidium-bromide (5 µg/ml), and RNAs were recovered from visible bands by elution and ethanol precipitation (16).

**RNA sequencing**

The 3'- or 5'-32P-labeling of RNAs was performed as described (14). The labeled RNAs were further purified by 8% polyacrylamide-7 M urea gel electrophoresis (0.05 ×20×40 cm) in TBE buffer. The nucleotide sequences were determined by the enzyme method of Donis-Keller (20), using RNA sequencing kit (Pharmacia-LKB).

**Cloning and DNA sequencing**

Southern hybridization was carried out as previously described (10). For cloning the RNA gene, the genomic DNA was digested with *BglII* and separated by 0.8% agarose gel electrophoresis. The DNA fragments with about 2 kbp in length were recovered from the gel, ligated to the *BglII* site of the vector plasmid, pUC119, and transformed to *E. coli* DH10B cells. Colony hybridization was performed to select the cells which carried the recombinant plasmid including the RNA gene, using 3'-32P-labeled RNA as a probe (10). Hybridization was carried out in 5 × SSC (1 × SSC: 0.15 M NaCl; 0.015 M sodium citrate, pH 7.2) containing 50% formamide and 0.1% SDS at 37°C for 20 h. The filter was washed in 5 × SSC at 37°C for 30 min, then in 2 × SSC at 50°C for 20 min, and autoradiographed. The DNA sequences were determined as previously described (10). The sequence data were analyzed by the computer program Hitachi DNASIS (Hitachi SK, Yokohama, Japan).

**Northern hybridization**

*M. capricolum* cells (0.5 g) were ground with quartz sands and suspended in 1 ml TMN-buffer (10 mM Tris—HCl, pH 7.6; 0.5
Nucleic Acids Research, 1993, Vol. 21, No. 11

mM MgCl2; 60 mM NH4Cl; 6 mM DTT) containing 10 μg/ml DNase I. The extract was centrifuged at 10,000 rpm for 15 min to remove sands and cell debris, and loaded on a 5—20% linear sucrose gradient (27 ml) made in TMN-buffer. Centrifugation was performed in SW25.1 rotor at 22,000 rpm for 20 h at 4°C in Beckman L8 ultracentrifuge. About 1.5 ml fractions were collected from the bottom of tube. After measuring the absorbance at 260 nm, RNA was extracted form each tube by phenol-method, precipitated with ethanol and dissolved in 25 μl distilled water. A 4 μl portion of the RNA solution was incubated at 65 °C in MOPS-buffer (20 mM N-morpholino-propane-sulphonic acid; 5 mM sodium acetate; 1 mM EDTA, pH 7.0) containing 50% formamide, separated by 1.75% agarose gel made in MOPS-buffer containing 18% formamide, and transferred to a Nylon membrane (Hybond-N+, Amersham). DNA fragment (30-mer) complementary to a part of MCS4 RNA intragenic sequence (from the nucleotide position number +48 to +77 of Figure 3b) was synthesized by Pharmacia-LKB DNA Synthesizer (Gene Assembler Plus). The 32P-5'-end labeled complementary DNA (cDNA) was used as a hybridization probe. Hybridization was performed as described above.

Figure 1. Polyacrylamide gel electrophoresis of small RNAs from Mycoplasma capricolum. About 0.5 mg total RNA was separated by denaturing 12% polyacrylamide gel electrophoresis, stained by ethidium-bromide, and photographed.

Figure 2. Southern hybridization of MCS4 RNA to the genomic DNA of M. capricolum. The DNA (4 μg) was digested with BgII (lane 1), EcoRI (lane 2), HindIII (lane 3) and XbaI (lane 4), respectively, and separated by 0.8% agarose gel electrophoresis. The gel was blotted to a Nylon membrane (Hybond-N+, Amersham), and hybridized with 32P-labeled MCS4 RNA. The 2.1 kbp BgII-fragment (shown by arrow) was cloned and sequenced (see Figure 3). The size markers are HindIII-digestion products of lambda phage DNA.

Figure 3. Nucleotide sequence of MCS4 RNA gene. a, Restriction map of the 2.1 kbp BgII-fragment containing the MCS4 RNA gene. The fragment was inserted at the BgII-site of pUC119 vector to generate PMC8015. The broad arrow below the map shows the position and direction of transcription of MCS4 RNA. P, T, and ORF represent promoter-, terminator-like structures and open reading frame, respectively. b, Nucleotide sequence encoding the gene for MCS4 RNA. The sequence of the non-coding strand is shown. The RNA coding sequence is underlined. Numbers indicate distance in nucleotides from the 5'-end of the RNA gene. Putative promoter -35 and -10 sequences are boxed. The initiation codon of ORF275 is shown by double-underline.

Materials
All restriction endonucleases, T4 RNA-ligase and T4-kinase were purchased from Takara-Shuzo Co. Ltd (Kyoto, Japan); Radioactive compounds (32P or 35S) were from Daiichi Pure Chemical Co. Ltd. (Tokyo, Japan). For sequencing DNA, 'Sequenase' kit of Toyobo Co, Ltd. (Osaka, Japan) was used.

RESULTS
Isolation of small RNAs
The total RNA was prepared from M. capricolum cells by direct phenol method (19) and fractionated by polyacrylamide gel electrophoresis. By staining the gel with ethidium-bromide, several bands could be observed in addition to 5S rRNA and bulk tRNA bands. Among them, a distinct band moving a little slower than 5S rRNA band (Figure 1) was cut out, and RNA was eluted from the gel. The RNA was labeled at the 3'- or 5'-end with 32P, and the sequences including the both terminals were determined by enzymatic method (20). The results revealed that the RNA comprised 125 nucleotides in length, and the sequence was homogeneous including both the 5'- and 3'-ends. Since the RNA showed no sequence identity to mycoplasmal rRNAs, tRNAs, or other known bacterial RNAs, we identified it as a new RNA species, and designated MCS4 RNA (We have identified six small RNA species including MCS4, and designated MCS1 to MCS6, respectively: our unpublished results). The RNA contains potential base-pairing sequences at the both 5'- and 3'-terminal regions (from position number +2 to +22 and...
Cloning and sequencing of MCS4 RNA gene

A Southern hybridization of genomic M.capricolum DNA with \(^{32}\)P-labeled MCS4 showed two major bands in the BglII-, EcoRI-, HindIII- and XbaI-digestion products, respectively (Figure 2), indicating that MCS4 RNA is encoded by two genes. The DNA fragment containing one of the genes, the 2.1 kbp BgII-fragment, was cloned in a plasmid vector pUC119, and the sequence of the fragment was determined. The location and organization of the genes in the plasmid (PMCB015) is shown in Figure 3a, and the sequence encompassing MCS4 RNA gene is in Figure 3b. The entire sequence of the DNA fragment will appear in the Nucleotide Sequence Databases under the accession number D13065. Upstream of the gene are two regions strikingly similar to the promoter consensus sequences, -35 and -10 boxes, found in E.coli and other eubacteria. Apart 39 bp from the 3'-end of MCS4 RNA gene, there exists an open reading frame capable of encoding a polypeptide with 275 amino acids (ORF275), followed by a possible termination signal, a long dyad-symmetrical sequence and T-stretches. No sequence homology of ORF275 was found with the known protein sequences.

Sedimentation behavior of MCS4 RNA

The M.capricolum cell extract was fractionated by sucrose gradient centrifugation under the low MgCl\(_2\) concentration (0.5 mM). Unexpectedly, the homology search with the known structural RNAs in the Databases has revealed that MCS4 exhibits an extensive sequence similarity to eukaryotic U6 snRNAs (see Discussion).

Cloning and sequencing of MCS4 RNA gene

A Southern hybridization of genomic M.capricolum DNA with \(^{32}\)P-labeled MCS4 showed two major bands in the BglII-, EcoRI-, HindIII- and XbaI-digestion products, respectively (Figure 2), indicating that MCS4 RNA is encoded by two genes. The DNA fragment containing one of the genes, the 2.1 kbp BgII-fragment, was cloned in a plasmid vector pUC119, and the sequence of the fragment was determined. The location and organization of the genes in the plasmid (PMCB015) is shown in Figure 3a, and the sequence encompassing MCS4 RNA gene is in Figure 3b. The entire sequence of the DNA fragment will appear in the Nucleotide Sequence Databases under the accession number D13065. Upstream of the gene are two regions strikingly similar to the promoter consensus sequences, -35 and -10 boxes, found in E.coli and other eubacteria. Apart 39 bp from the 3'-end of MCS4 RNA gene, there exists an open reading frame capable of encoding a polypeptide with 275 amino acids (ORF275), followed by a possible termination signal, a long dyad-symmetrical sequence and T-stretches. No sequence homology of ORF275 was found with the known protein sequences.

Sedimentation behavior of MCS4 RNA

The M.capricolum cell extract was fractionated by sucrose gradient centrifugation under the low MgCl\(_2\) concentration (0.5 mM). Unexpectedly, the homology search with the known structural RNAs in the Databases has revealed that MCS4 exhibits an extensive sequence similarity to eukaryotic U6 snRNAs (see Discussion).
mM), and distribution of the RNA was analyzed by hybridization to the complementary DNA (cDNA) sequence to a part of MCS4 RNA. As shown in Figure 4, only a single band of the expected size (about 125 nucleotides long) could be detected, showing that the cDNA probe specifically hybridized to MCS4 RNA. No other transcripts could be detected. The result showed a wide range distribution of MCS4 RNA sedimenting from about 15S to 50S. The RNA seems to exist in at least two populations of particles, the major peak of about 50S and the minor one of about 20S. It is not known whether the 50S component represents binding of MCS4 RNA to 50S ribosomal subunits, or an independent particle. However, since the RNA did not detected in 70S ribosome fraction in the cell extract prepared under the high MgCl₂ condition (10 mM) (data not shown), it is clear that MCS4 is not a structural component of 50S subunit.

**DISCUSSION**

In this work, we isolated and sequenced a new RNA species (125 nucleotides in length) from *M. capricolum*. The RNA, MCS4, is encoded by two genes, and is abundant in the cell (about one-fourth in amount of 5S rRNA (see Figure 1)).

Surprisingly, MCS4 RNA reveals extensive sequence similarity to eukaryotic U6 snRNAs. Figure 5 shows the sequence comparison between MCS4 RNA and ten representative U6 snRNAs (22–33). Since MCS4 RNA is longer than U6 RNAs (varying from 78 to 115 nucleotides in length), gaps are introduced mainly at the 5′-half region of U6 RNAs to improve the alignment. The MCS4 sequence shows 49 to 66% identity, excluding gaps, with either of the U6 sequences from variety of eukaryotes, vertebrates (22–24), invertebrates (25, 26), protozoans (27, 28), plants (29, 30), slime mold (31) and yeasts (32, 33). Taking into account of the phylogenetic distance between eukaryotes and mycoplasmas, the values are significantly high (e.g., the sequence identity between mammal and yeast U6 RNAs is about 60% (see also ref. 32)). Many consensus nucleotides among U6 RNAs are also conserved in MCS4 RNA. The middle part of the U6 sequence is more conserved than the 5′-terminal region (from the nucleotide position +2 to +22) of MCS4 RNA. These common features strongly suggest that MCS4 RNA is a structural homolog of U6 RNA. Then, this is the first finding of the prokaryotic RNA homologous to eukaryotic snRNA.

MCS4 RNA is one of the spliceosomal RNAs involved in pre-mRNA splicing (for reviews, see ref. 34–36). Five snRNA species, U1, U2, U4, U5, and U6, associate with proteins to form the small nuclear ribonucleoproteins (snRNPs). As shown in Figure 4, most MCS4 RNA sediments from about 15S to 50S regions, apparently heavier than free RNA, indicating that the RNA associates with proteins and/or RNAs to form multiple snRNPs. Does *M. capricolum* contain spliceosomas participating in the pre-mRNA splicing as eukaryotic cells? It is unlikely, however, because no nuclear mRNA-type intron has been found in eubacteria, and because the stretch from the nucleotide position +30 to +40, which is one of the most conserved regions in U6 RNAs for functionally important (32–35), is not retained in MCS4 RNA. Alternatively, MCS4 RNA may have some function(s) similar to but different from U6 RNA. Of five spliceosomal RNAs, U6 is the most highly conserved at the primary sequence level from lower to higher eukaryotes (33–36), suggesting that it plays a critical role in the splicing reaction. In fact, the evidence indicating that U6 functions as a catalytic element in splicing has been accumulating (37, 38).

ACKNOWLEDGEMENT

This research was supported by grants from the Ministry of Education, Science and Culture of Japan.

**REFERENCES**