mRNA Sequence and Deduced Amino Acid Sequence of the Mumps Virus Small Hydrophobic Protein Gene

NARAYANASAMY ELANGO,† JAN KÖVAMEES, TAMAS M. VARSANYI, AND ERLING NORRBY

Department of Virology, School of Medicine, Karolinska Institute, Stockholm S-105 21, Sweden

Received 18 July 1988/Accepted 7 November 1988

The mRNA of a putative small hydrophobic protein (SH) of mumps virus was identified in mumps virus-infected Vero cells, and its complete nucleotide sequence was determined by sequencing the genomic RNA and cDNA clones and partial sequencing of mRNA. The SH mRNA is 310 nucleotides long excluding the poly(A) and contains a single open reading frame encoding a protein of 57 amino acids with a calculated molecular weight of 6,719. The predicted protein is highly hydrophobic and contains a stretch of 25 hydrophobic amino acids near the amino terminus which could act as a membrane anchor region. There is no homology between the putative SH protein of mumps virus and the SH protein of simian virus 5, even though the SH genes are located in the same locus in the corresponding genome. One interesting observation is that the hydrophobic domain of simian virus 5 SH protein is at the carboxyl terminus, whereas that of mumps virus putative SH protein is near the amino terminus.

Mumps virus, an enveloped RNA virus, is a member of the Paramyxoviridae family. The virus envelope contains a nonglycosylated matrix protein (M) lining the inside of the envelope and two glycoproteins, a hemagglutinin-neuraminidase protein (HN) and a fusion protein (F), which form spikelike projections on the outer surface of the virus particle. Inside the envelope, there is a ribonucleoprotein complex which contains the genome of about 15,000-nucleotide-long negative-sense RNA, a nucleocapsid protein (NP), a polymerase-associated protein (P), and a large protein (L). Two nonstructural proteins related to the P protein were seen in mumps virus-infected cells (9).

We have recently described the construction and characterization of cDNA clones for mumps virus mRNAs which led to the identification of seven viral genes including a small hydrophobic protein (SH) gene, located between the F and HN genes, in the genome (1). Here we present evidence for the presence of SH mRNA in mumps virus-infected cells, the complete nucleotide sequence of the SH mRNA, and the deduced amino acid sequence and compare the SH proteins of mumps virus and simian virus 5 (SV5).

Isolation of SH-specific cDNA clones. When an mRNA sense oligonucleotide synthesized on the basis of the 3' end of the mumps virus fusion protein mRNA was applied to sequence the mumps virus genome, the deduced nucleotide sequence (about 200 nucleotides) did not contain the 5'-end sequence of any of the known mumps virus mRNAs (1). An oligonucleotide, 5'-ATATTGACCATTAACA, was then synthesized on the basis of the new sequence and applied to sequence the genome further. The sequence obtained now included the 5'-end sequence of the HN mRNA. The intervening sequence, between the F gene terminus to HN gene start, was 325 nucleotides long. If the region between the F and HN genes was transcribed to mRNA, then there should be corresponding cDNA clones in the cDNA library constructed with the mRNA isolated from mumps virus-infected Vero cells. When two oligonucleotides synthesized on the basis of the intervening sequence were used as probes to screen 1,000 cDNA clones, two cDNA clones with 400-base-pair inserts hybridized to the oligonucleotide probes. They were designated pMS1 and pMS2 and were selected for sequencing.

Identification of SH mRNA. To obtain evidence that SH mRNA is transcribed from the mumps virus genome and to determine the size of the SH mRNA, we performed Northern blot (RNA blot) analysis (5) with mRNA from mumps virus-infected Vero cells. The 32P-labeled cDNA insert of pMS1 hybridized to three RNA species of about 550, 2,500, and 15,000 nucleotides, respectively (Fig. 1). When a 32P-labeled cDNA insert of an F cDNA clone was used as the hybridization probe, it hybridized to a monocistronic F mRNA of 2,000 nucleotides, to a mRNA of 2,500 nucleo-

![FIG. 1. Identification of the SH mRNA. Poly(A)-containing RNA (5 μg) from uninfected Vero cells (lane C) and mumps virus-infected Vero cells (lane V) was electrophoresed on a formaldehyde-agarose (1.5%) gel, transferred to nitrocellulose paper, and hybridized to 32P-labeled cDNA insert of pMS1. The positions of SH mRNA (SH), F-plus-SH dicistronic mRNA (F+SH), and genome RNA (G) are marked. Lane M, HaeIII digest of φX174 replicative-form DNA.](http://jvi.asm.org/Downloaded_from)
nucleotides, and to the 15,000-nucleotide-long genomic RNA, whereas a 32P-labeled cDNA insert of an HN cDNA clone hybridized to 2,000-nucleotide-long HN mRNA and 15,000-nucleotide-long genomic RNA, (1). The 550-nucleotide-long mRNA which hybridized to the cDNA insert of pMS1 was, therefore, the SH mRNA, and the 2,500-nucleotide-long mRNA should represent F-plus-SH dicistronic mRNA.

Nucleotide sequence of SH mRNA. The cDNA inserts from both pMS1 and pMS2 were subcloned into the M13 bacteriophage vector M13mp18 and sequenced by the dideoxynucleotide sequencing method (6) with an mRNA sense oligonucleotide (positions 135 to 151, Fig. 2) and a genome sense oligonucleotide (positions 151 to 135, Fig. 2) synthesized on the basis of the genomic sequence. Both cDNA inserts had poly(A), which indicated that the 3' end of the mRNA has been cloned. pMS2 was missing 39 nucleotides of pMS1. An oligonucleotide complementary to nucleotides 37 to 21 (Fig. 2) was used to map the 5' end of the SH mRNA as described elsewhere (1a). Except for the terminal nucleotide, which could not be determined by dideoxynucleotide sequencing of the RNA, the rest of the mRNA sequence was complementary to the sequence obtained by genome sequencing. The

![Hydropathy plots of the mumps virus SH protein and SV5 SH protein. The relative hydrophobicity and hydrophilicity of the proteins were calculated with a segment length of nine amino acids and plotted from the amino terminus to the carboxyl terminus of the proteins. The dotted midline represents the overall average of hydrophathy of amino acid compositions found in most sequenced proteins. The plots were plotted by the Kyte and Doolittle (3) program as part of the Pustell sequence analysis programs of International Biotechnologies, Inc. (New Haven, Conn.).](image-url)
cDNA insert from pMS1 was missing 19 nucleotides, whereas the pMS2 insert was missing 58 nucleotides of the mRNA.

The complete nucleotide sequence of the SH mRNA, excluding the poly(A), is shown in Fig. 2. The SH mRNA is 310 nucleotides long, excluding the poly(A). The sequence contains a single open reading frame which starts at AUG at nucleotides 51 to 53 and ends at UAG at nucleotides 222 to 224. The sequence around the initiation codon, ACUAUGC, corresponds only partially to the consensus sequence 5XXAUGG present in most eucaryotic mRNAs (2). The putative protein is 57 amino acids long with a calculated molecular weight of 6,719. The protein is rich in hydrophobic amino acids, and there is one acidic amino acid (Asp) and four basic amino acids (three Arg and one Lys). Near the amino terminus, there is a stretch of 25 hydrophobic amino acids which could anchor the protein in the cell membrane or virus envelope. The presence of a hydrophobic domain near the amino terminus is similar to the SH protein of paramyxoviruses and the respiratory syncytial virus glycoprotein G (7, 8). The SH protein does not have any potential N-glycosylation sites (Asn-X-Ser or Asn-X-Thr), but after the hydrophobic domain, there are three serine and one threonine residues which are the potential O-glycosylation sites.

Comparison of the mumps virus putative SH protein with the SV5 SH protein (1b) did not show any homology, even though the SH genes are located in the same locus in the corresponding genome. The hydropathy plots of the SH proteins of mumps virus and SV5 are shown in Fig. 3. The hydrophobic domain of mumps virus SH protein is near the amino terminus, whereas that of SV5 is at the carboxyl terminus.

At present, we are unable to show the presence of the SH protein in infected cells because we see a protein of 6.7 kilodaltons in control cell extract as well as in mumps virus-infected cell extract. We consistently failed to translate the hybrid-selected SH mRNA and SH mRNA present in total mRNA with rabbit reticulocyte lysate as well as wheat germ extract. Monoclonal or polyclonal antipeptide antibodies against the SH protein might identify the protein and provide an answer to the question of whether this protein is a membrane protein and if so whether it is present in the virus envelope or in the membrane of infected cells like the nonglycosylated M2 protein of influenza A viruses (4).

This work was supported by project B88-16X-00116-24B of the Medical Research Council, Sweden, and by a grant from the Wallenberg Foundation, Stockholm, Sweden.

LITERATURE CITED