Molecular Cloning and Characterization of Six Genes, Determination of Gene Order and Intergenic Sequences and Leader Sequence of Mumps Virus

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

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

(Accepted 21 July 1988)

SUMMARY

mRNA isolated from mumps virus-infected Vero cells was converted into cDNA and cloned into the PstI site of the plasmid pBR322. After screening with 32P-labelled cDNA synthesized from poly(A)+ RNA of uninfected or mumps virus-infected Vero cells, five different groups of virus-specific clones were obtained. The virus specificity of the clones was confirmed by Northern blot analysis, in which the cDNA inserts from the five different groups hybridized to mRNAs of about 2100, 1500, 1450, 2000 and 2200 nucleotides. By the use of oligonucleotides synthesized on the basis of sequences obtained from the five cDNA clones and mRNAs, the sequence of the intergenic and surrounding areas was determined. During genome sequencing, a separate gene was identified between the fusion protein (F) gene and the haemagglutinin–neuraminidase protein (HN) gene. Using oligonucleotides synthesized on the basis of the new gene sequence, cDNA clones with poly(A) were isolated from the cDNA library. The gene order was determined to be 3' NC-P-M-F-SH-HN-L 5' (where NC, P, M, SH, and L represent the genes for the nucleocapsid, phosphoprotein or polymerase-associated, matrix or membrane, small hydrophobic and large proteins respectively). There is one nucleotide between the P and M (A), M and F (A), and HN and L genes (G), two between the NC and P (AA) and SH and HN (3'-CG) genes, and seven between the F and SH genes (3' GAUUUUA) as intergenic sequence. The leader sequence at the 3' end of the genome has been determined by sequencing the dicistronic leader–NC mRNA using oligonucleotide primers. The sequence from the 3' terminus to the NC gene start of the mumps virus genome is similar in length (55 nucleotides) to that present in Sendai virus, Newcastle disease virus and parainfluenza virus type 3, and the first five nucleotides are conserved in all negative-stranded RNA virus genomes sequenced to date.

INTRODUCTION

Mumps virus, an enveloped RNA virus, is a member of the genus Paramyxovirus, within the family Paramyxoviridae. The genome is an unsegmented negative- and single-stranded RNA of about 15000 nucleotides (East & Kingbury, 1971). Like other paramyxoviruses, mumps virus contains a nucleocapsid (NC) protein (M, 72K), a phosphoprotein or polymerase-associated (P) protein (M, 45K), a membrane or matrix (M) protein (M, 40K), a fusion (F) protein consisting of two subunits, F1 (M, 61K) and F2 (M, 16K) joined together by disulphide bonding, a haemagglutinin–neuraminidase (HN) (M, 79K) and a large (L) protein (M, 200K) (Örvell, 1978; McCarthy & Johnson, 1980; Rima et al., 1980; Herrler & Compans, 1982). The virus encodes two non-structural proteins in infected cells, which are related to the P protein (Rima et al., 1980; Herrler & Compans, 1982).
Despite the medical significance of mumps virus, comparatively limited data are available on its molecular biology. Although there was a report on cDNA cloning of the NC and P protein genes (Curran et al., 1985), no sequences were published, and only very recently was the sequence of the F protein gene of a non-fusion strain published (Waxham et al., 1987).

In order to understand the genome organization and, as a first step, to study transcription and replication in this negative-stranded RNA virus, we have constructed a cDNA library using mRNA isolated from mumps virus-infected Vero cells. Here we describe the construction and characterization of cDNA clones for six mRNAs, the determination of the gene order and intergenic sequences, the analysis of the gene start and gene end sequences and the determination of the leader sequence.

**METHODS**

**Cells and virus.** Vero cells were grown in Eagle's MEM supplemented with 5% heat-inactivated foetal calf serum. The SBL-1 strain of mumps virus (Orvell, 1978) was plaque-purified three times and grown in Vero cell monolayers.

**Isolation of mRNA.** Vero cells grown in roller bottles were infected at a multiplicity of 1 p.f.u./cell. At 10 h post-infection (p.i.), actinomycin D (5 μg/ml) and [3H]uridine (10 μCi/ml) were added. The cells were harvested at 20 h p.i., washed with phosphate-buffered saline and then suspended in RS buffer (0.01 M-Tris-HCl pH 8.5, 0.01 M-NaCl and 0.0015 M-MgCl₂) diluted with two volumes of water. A 10% solution of Nonidet P40 was added to a final concentration of 0.5% and the mixture was incubated on ice for 10 min. The nuclei were then removed by centrifugation at 1875 g for 10 min. A 10% solution of SDS was added to the cytoplasmic extract to a final concentration of 1% and RNA was isolated by extraction with phenol followed by phenol–chloroform (Wertz, 1978) and precipitated with ethanol. Poly(A)+ RNA was purified by oligo(dT)-cellulose chromatography (Maniatis et al., 1982).

**Isolation of genomic RNA.** Vero cells grown in roller bottles were infected at a multiplicity of 1 p.f.u./cell. At 48 h p.i., the medium was clarified at 8000 r.p.m. (Sorvall GSA rotor) for 10 min and the virus was pelleted at 13000 r.p.m. (Sorvall GSA rotor) for 90 min. The virus was then purified first by sedimentation through 30% sucrose onto a cushion of 65% sucrose and then by banding in a 10 to 60% discontinuous sucrose gradient at 35000 r.p.m. (Beckman SW40 rotor) for 90 min. The banded virus was diluted with TNE buffer (10 mM-Tris-HCl pH 7.4, 100 mM-NaCl, 1 mM-EDTA) and pelleted by centrifugation at 35000 r.p.m. (Beckman SW40 rotor) for 60 min. RNA was isolated from the purified virions by first incubating with 500 μg/ml proteinase K, 0.3% SDS at 56 °C for 20 min and then extracting with phenol–chloroform, followed by ethanol precipitation.

**Construction of cDNA library.** Poly(A)+ RNA (10 μg) was used to synthesize cDNA by a modification of the method of Gubler & Hoffman as described by D'Alessio et al. (1987), but using avian myeloblastosis virus reverse transcriptase. The double-stranded cDNA was dC-tailed and annealed to PstI-cut dG-tailed pBR322 and used to transform competent *Escherichia coli* HB101 cells. Tetracycline-resistant transformants (about 4000) were grown and stored in 96-well tissue culture plates.

**Northern blot analysis.** Poly(A)+ RNA from virus-infected or uninfected Vero cells was electrophoresed on formaldehyde-agarose gels, transferred to a nitrocellulose membrane and hybridized to 32P-labelled cDNA inserts or oligonucleotides as described (Maniatis et al., 1982; Woods, 1984).

**Genome sequencing.** mRNA-sense oligonucleotides [NC, 5'-d(AAATCCACTACACTCCA)-3'; P, 5'-d(AACTGAGAGCCACAATG)-3'; M, 5'-d(GCAAATTTGACCGGTGG)-3'; F, 5'-d(GTGATTCGTCC-TGCAAT)-3'; SH, 5'-d(ATATTGACCATTAACCA)-3'; HN, 5'-d(TGGGCGTGTCTCATTTC)-3'], about 25 nucleotides upstream of each gene end, were synthesized using a Pharmacia Gene Assembler. Methods for labelling of oligonucleotides and RNA sequencing will be described elsewhere.

**RESULTS**

**Identification of virus-specific cDNA clones**

About 500 cDNA clones were screened for the presence of cDNA inserts by isolating plasmid DNA from overnight cultures and digesting the plasmid DNA with PstI. Clones with inserts of more than 800 bp were grouped together by cross-hybridization. After dot blot hybridization with 32P-labelled cDNA synthesized from poly(A)+ RNA of virus-infected or uninfected cells, five groups of clones that hybridized only to the cDNA synthesized from virus mRNA were selected for analysis.

The viral specificity of these five groups of clones was confirmed by Northern blot analysis. One large cDNA insert from each group was nick-translated and hybridized to a nitrocellulose blot of poly(A)+ RNA from uninfected or infected cells; the results are shown in Fig. 1. All the
Genome organization of mumps virus

Fig. 1. Northern blot analysis of mumps virus mRNA. Five µg of poly(A)-containing RNA from uninfected Vero cells (lanes 1) and mumps virus-infected Vero cells (lanes 2) were electrophoresed on formaldehyde-1% agarose gels, transferred to nitrocellulose paper and hybridized to 32P-labelled cDNA inserts of (a) pMNI, (b) pMP1, (c) pMM1, (d) pMF1 and (e) pMH1; 5' end-labelled oligonucleotide was used to identify the L mRNA (f). The positions of mumps virus mRNAs and genome (G) are marked. The additional bands seen in the case of M and F are readthrough mRNAs. NC, P and M were from the same gel whereas F, HN and L were from three different gels.

cDNA inserts hybridized to the mRNA from virus-infected Vero cells but not to the cellular mRNA, confirming their viral origin. The cDNA inserts from pMNI (1700 bp), pMP1 (1350 bp), pMM1 (1350 bp) and pMH1 (1000 bp) hybridized mainly to species of about 2100, 1500, 1450 and 2200 nucleotides, respectively, whereas the insert from pMF1 (900 bp) hybridized to two mRNAs of 2000 and 2400 nucleotides. All the inserts also hybridized with an RNA of about 15000 nucleotides, i.e. the positive-sense genome that was present in the RNA preparation.

Determination of gene order and intergenic sequences

The 3' end of mRNA of each gene was determined by sequencing cDNA clones with poly(A), and the 5' end by sequencing cDNA clones and mRNA.

Oligonucleotides complementary to the genomic RNA at the 5' end of each gene and about 25 nucleotides away from the intergenic area [i.e. away from the start of the poly(A), in the mRNA sense] were synthesized from the nucleotide sequences of the NC, P, M, F and HN mRNAs. After 5' end labelling, these oligonucleotides were annealed to the genomic RNA (5 to 10 µg) and dideoxynucleotide sequencing reactions were carried out. The sequence after the F gene did not have the 5' end sequence of any mRNA we have sequenced. Based on this new sequence another oligonucleotide, 5'-d(ATATTGACCATTACCCA), was synthesized and used for genome sequencing. The sequence obtained now contained the sequence of the 5' end of the HN mRNA.

Identification of a new small hydrophobic protein gene

If a separate gene was present between the F and HN genes, there had to be an mRNA species synthesized from it. In order to verify this, two oligonucleotides, 5'-d(GGTCACGAGACGTTACG) and 5'-d(ATATTGACCATTACCCA) were synthesized on the basis of the sequence present between the F and HN genes, and used as probes to screen 1000 colonies. Out of six colonies that hybridized to these probes, two had cDNA inserts of about 400 bp, and these were selected for sequencing. The sequence obtained from these two clones (unpublished data) confirmed the existence of an independent gene for a small hydrophobic (SH) protein. This gene, by analogy with simian virus 5 (SV5) (Hiebert et al., 1985), is referred to as the SH gene.
Fig. 2. Comparison of the intergenic and flanking gene sequences and determination of gene order. Genomic RNA sequences (upper sequences) were obtained by sequencing the genomic RNA with oligonucleotide primers (see Methods), and mRNA sequences (lower sequences) were obtained by sequencing cDNA inserts (unpublished data). The intergenic sequences are underlined.

Mapping of the L gene

Using the HN mRNA-sense oligonucleotide, 5'-d(TGGGCGTGTCTCATTTC), we were able to sequence part of the gene next to the HN gene in the genome. To find out whether the L gene follows the HN gene in the gene order, a genome-sense oligonucleotide, 5'-d(GCTTATATCTACGATG), synthesized on the basis of the sequence next to the HN gene, was hybridized to a nitrocellulose blot of mRNA from uninfected and mumps virus-infected Vero cells. As shown in Fig. 1(f), the oligonucleotide hybridized specifically to RNAs of about 7000 and 15000 nucleotides. By analogy to the L mRNA of other paramyxoviruses, this 7000 nucleotide RNA should be the L mRNA and the 15000 nucleotide RNA should be the positive-sense genome.

To sequence the 5' end of the L mRNA, the oligonucleotide that was used to identify this mRNA was used. Except for the first nucleotide, which could not be determined by dideoxynucleotide sequencing, the sequence 5'-NGGCCAGAAUGGCGG... (mRNA-sense) was shown to contain the L gene start sequence.

Gene order and intergenic sequence

Fig. 2 shows the intergenic and flanking sequences obtained by genome sequencing, and the sequences at the 5' and 3' ends of mRNAs determined by cDNA and mRNA sequencing.
Comparison of the genome sequence with the 5' and 3' mRNA sequences gives the gene order and the intergenic sequences. The gene order thus obtained for mumps virus is: 3' NC-P-M-F-SH-HN-L 5'. There is one nucleotide between the P and M (A), M and F (A) and HN and L (G) genes, two between the NC and P (AA) and SH and HN (3'-CG) genes, and seven between the F and SH (3'-GAUUUUA) genes as intergenic sequences.

Analysis of gene start and gene end sequences

Mumps virus genes start with 3'-UUC, with the following exceptions: the P and L genes start with 3'-UCC, the NC, F and HN genes with 3'-UUCGG, the P and L genes with 3'-UCCGG, the M gene with 3'-UUCGU, and the SH gene with 3'-UUCUU. Mumps virus genes have been shown to terminate at a C residue (excluding the L gene which is not determined), except the P gene which terminates at an A residue. The consensus gene start sequence, therefore, is 3'-UUCGU and the consensus gene end sequence is 3'-AUAAAUUUA.

Determination of the leader sequence

When the oligonucleotide 5'-d(AGCCCGGCCTCGAAATC) was used to determine the exact 5' terminus of the NC mRNA, additional sequence bands were seen beyond the NC 5' terminus. These represented sequences from a minor dicistronic RNA derived from the readthrough transcription of the leader region plus the NC gene. Based on this sequence an oligonucleotide, 5'-d(TGTTTCTTACACTA), was synthesized and used in mRNA sequencing to obtain the 3' end sequence of the genome. In order to obtain the sequence between the leader and the NC gene an oligonucleotide, 5'-d(AGGGGAGAAAGAAGATG), complementary to the genome at its 3' end was used to sequence the genomic RNA. The complete 3' end sequence of the mumps genome (excluding the first nucleotide), present before the NC gene start, is shown in Fig. 3.

**DISCUSSION**

In this paper we describe the construction of a cDNA library and the identification of clones specific for six genes of mumps virus. Five different groups of cDNA clones were obtained, and one clone with a large insert was selected from each group for analysis. In Northern blot analysis the cDNA inserts from pMN1 (1700 bp), pMP1 (1350 bp), pMM1 (1350 bp), pMF1 (900 bp) and pMH1 (1000 bp), and an oligonucleotide synthesized on the basis of the L mRNA sequence, hybridized to viral mRNAs of about 2100, 1500, 1450, 2000, 2200 and 7000 nucleotides, respectively. The cDNA insert from pMF1 also hybridized to an mRNA of 2400 nucleotides. The sequencing of the cDNA insert from pMF1 and the mRNA showed that the insert contained only the F mRNA sequence (data not shown), which indicated that the 2000 nucleotide RNA is the F mRNA and the 2400 nucleotide RNA is a readthrough mRNA of the F gene plus the next gene. During the course of the genome sequencing, an SH protein gene was identified and the corresponding cDNA clones were isolated from the cDNA library. The exact length (in nucleotides) of the mRNAs, excluding poly(A), determined by sequencing of the cDNA inserts and the mRNAs are: NC, 1844; P, 1311; M, 1248; F, 1721; SH, 310; HN, 1887 (unpublished data). Sequencing of the L gene has not yet been completed.
Table 1. The gene order in the genomes of six paramyxoviruses, determined by sequence analysis

<table>
<thead>
<tr>
<th>Virus</th>
<th>Gene order 3'-5'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mumps virus</td>
<td>NC P M F SH HN L</td>
</tr>
<tr>
<td>SV5</td>
<td>[NC]† [P]† M F SH HN [L]†</td>
</tr>
<tr>
<td>NDV</td>
<td>NC P M F SH HN L</td>
</tr>
<tr>
<td>Sendai virus</td>
<td>NC P +C M F HN L</td>
</tr>
<tr>
<td>PIV3</td>
<td>NC P +C M F HN L</td>
</tr>
<tr>
<td>Measles virus</td>
<td>NC P +C M F HN L</td>
</tr>
</tbody>
</table>

* Mumps virus (from Fig. 2); SV5 (Paterson et al., 1984; Hiebert et al., 1985); NDV (Ishida et al., 1986; Sato et al., 1987a, b); Sendai virus (Shioda et al., 1983; Gupta & Kingsbury, 1984); PIV3 (Spriggs & Collins, 1986); measles virus (Cattaneo et al., 1987).
† The order of the genes shown in brackets was not determined by sequencing and is tentative.

So far only the NC, P, M, F, HN and L proteins of mumps virus have been described in the literature. Here we show the presence of a separate gene in the genome and its counterpart in the cDNA library. Recently a similar gene has been identified in SV5 (Hiebert et al., 1985). It is present at the same locus of the genome in both viruses. The gene product of the mumps virus SH gene has not yet been identified, and whether it is a structural or non-structural protein remains to be determined.

Comparison of the 5' and 3' end sequences of mRNA with the genome sequence identified the gene order as 3' NC-P-M-F-SH-HN-L 5'. There is no consensus within the intergenic sequences either in sequence or in length, as there is one nucleotide between the P and M (A), M and F (A) and HN and L (G) genes, two between the NC and P (AA) and SH and HN (3'-CG) genes, and seven between the F and SH genes (3'-GAUUUUA) as intergenic sequences. This is in contrast to the situation in Sendai virus [3'-GAA or GGG (Gupta & Kingsbury, 1984)], vesicular stomatitis virus (VSV) [3'-CA or -GA (Rose, 1980)], measles virus [3'-GAA or -GCA (Richardson et al., 1985; Cattaneo et al., 1987)] or parainfluenza virus type 3 (PIV3) [3'-GAA (Spriggs & Collins, 1986)]. Newcastle disease virus (NDV) was shown to have A (P-M), G (M-F), 3'-CA (NC-P), 31 nucleotides (F-HN) and 48 nucleotides (HN-L) as intergenic sequences (Ishida et al., 1986; Sato et al., 1987a, b).

Table 1 shows the gene order of different paramyxoviruses. Although the complete gene order of SV5 has not been determined, the arrangement of genes surrounding the SH gene has. The existence of an extra gene of similar size in mumps virus and SV5 and the higher degree of relationship at the amino acid level (unpublished data) indicate that mumps virus, SV5 and NDV may have a common ancestor.

Mumps virus genes do not have identical gene start as well as gene end sequences, but the consensus derived for gene start is 3'-UUCUCCNUU... and gene end is 3'...AUACU (Rose, 1980). Sendai virus also has a common gene start, 3'-UCCANUUUC... and gene end, 3'...UNAUUCU (Blumberg et al., 1985; Giorgi et al., 1983; Gupta & Kingsbury, 1982, 1984; Shioda et al., 1983). Similar gene start (3'-UGCNAUUC...) and gene end (3'...AUACU) sequences have been described for NDV (Ishida et al., 1986; Sato et al., 1987a, b; Yusoff et al., 1987). PIV3 has a semi-conserved gene start (3'-UCCUNNUUUC...) and gene end (3'...UUUAAUU...)(Spriggs & Collins, 1986).

The gene end sequences of all negative-stranded RNA viruses contain oligo(U). The number of U residues ranges from four, e.g. respiratory syncytial virus (Collins et al., 1986) to seven, e.g. VSV (Rose, 1980), and the number of U residues is either the same (five in Sendai virus and seven in VSV) or variable (six or seven in mumps virus, four to seven in respiratory syncytial virus, five or six in PIV3, six or seven in NDV) within the virus. These oligo(U) residues are thought to be copied repeatedly by the polymerase to give rise to the mRNA poly(A) tail (Schubert et al., 1980).
What regulates transcription termination in negative-stranded RNA viruses? It has been shown in Sendai virus that an absence of two U residues (out of five) plus the intergenic trinucleotide and the downstream common gene start sequence causes exclusive readthrough mRNA transcription (Hsu et al., 1985). However, it is not known whether the absence of one or more U residues, the intergenic sequence, the downstream common gene start sequence, or a combination of more than one of the above will cause the readthrough transcription.

It is known that the gene end sequences are somewhat semi-conserved or conserved, and the number of U residues after the gene end varies within the same virus. It is also known that the intergenic sequence varies within each virus and there is no gene start sequence found downstream of the L gene (Schubert & Lazzarini, 1981; Re et al., 1983; Yusoff et al., 1987; Crowley et al., 1988). Absence of a gene start sequence downstream of the L gene might suggest that downstream gene start sequences do not participate in transcription termination. One might therefore hypothesize that either a conserved or semi-conserved gene end sequence with a tract of U residues plus any sequence will act as a transcription termination signal and cause polyadenylation.

The leader sequences of all paramyxoviruses except measles virus examined to date (Sendai virus, NDV and PIV3) are identical in length (55 nucleotides) (the measles virus leader sequence is 56 nucleotides) (Shioda et al., 1983; Gupta & Kingsbury, 1984; Kurilla et al., 1985; Ishida et al., 1986; Dimock et al., 1986; Sakai et al., 1987). The consensus 3' end sequence of the genome may indicate an important role in transcription and/or replication.

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

REFERENCES


(Received 2 May 1988)