Identification of a second protein encoded by influenza C virus RNA segment 6

Seiji Hongo, Kanetsu Sugawara, Hidekazu Nishimura, Yasushi Muraki, Fumio Kitame and Kiyoto Nakamura*

Department of Bacteriology, Yamagata University School of Medicine, Iida-Nishi, Yamagata 990-23, Japan

Influenza C virus matrix protein (M1) is encoded by a spliced mRNA derived from RNA segment 6. Unspliced mRNA from this RNA segment, which has not been previously identified, can potentially encode a polypeptide that contains an additional 132 amino acids on the carboxy terminus of the M1 protein. Here the nucleotide sequences of RNA segment 6 of four influenza C strains, isolated in Japan between 1964 and 1988, were compared with the previously determined sequence of C/Ann Arbor/1/50. The results indicated that the deduced amino acid sequence of the carboxy-terminal 132 amino acid domain is conserved fairly well although it is more divergent than the M1 protein sequence. Examination of RNA segment 6-specific mRNAs also showed that unspliced mRNA is present, although in small quantities (~13% of spliced mRNA), in influenza C virus-infected cells. To search for a polypeptide encoded by the unspliced mRNA, the extra carboxy-terminal domain was expressed in Escherichia coli as the glutathione S-transferase fusion protein, and rabbit immune serum was raised against the purified fusion protein. Immunoprecipitation experiments with this antiserum revealed that a previously unrecognized protein of apparent Mr ~ 18,000, designated CM2, is synthesized in influenza C virus-infected cells.

Introduction

Unspliced and spliced mRNAs are synthesized from RNA segment 7 of influenza A virus and are translated to yield the matrix proteins M1 and M2, respectively (Lamb & Lai, 1981). The matrix protein M1 underlies the lipid bilayer and adds rigidity to the virion envelope. The protein is also thought to play a key role in the assembly and budding of the virus by mediating interactions between the viral glycoproteins and the ribonucleoprotein complex (Lamb, 1989). M2 is an integral membrane protein and is present as a tetramer in large quantities on the surface of virus-infected cells (Lamb et al., 1985). This protein acts as a proton channel to allow acidification of the interior of the virion during virus uncoating and to regulate the pH of vesicular components of the trans-Golgi network (Sugrue & Hay, 1991; Pinto et al., 1992; Ciampor et al., 1992). RNA 7 of influenza B virus also codes for two proteins, M1 and BM2. These proteins are both synthesized from a bicistronic mRNA that is collinear with the vRNA (Horvath et al., 1990). The role of BM2 in virus replication is not known. It is generally believed that in influenza B virus-infected cells the NB protein encoded by RNA 6 performs a function similar to that of influenza A virus M2 (Sugrue & Hay, 1991).

The influenza C virus genome contains seven single-stranded RNA segments of negative polarity. RNA segment 6 of C/Ann Arbor/1/50 virus was cloned and sequenced by Yamashita et al. (1988), and was shown to code for the matrix protein (M1) of the virus. The RNA segment was 1180 nucleotides in length and contained a single open reading frame that could encode a polypeptide of 374 amino acids with a predicted Mr of about 42K, a value approximately 15K larger than the size of M1 estimated by PAGE (Yokota et al., 1983). Examination of RNA 6-specific mRNAs revealed that the predominant mRNA lacks a region from nucleotides 754 to 981 and encodes a 242 amino acid M1 protein (Mr 27K) (Yamashita et al., 1988). The collinear transcript of RNA 6 has not yet been identified in infected cells. This mRNA species, if it is synthesized, may be translated into a polypeptide that contains an additional 132 amino acids at the C terminus of the M1 protein (see Fig. 2).

Here, we provide data indicating that the deduced amino acid sequence of the extra carboxy-terminal domain is conserved fairly well among the five different influenza C strains although it is more divergent than the M1 protein sequence. Evidence is also presented that a
Fig. 1. Nucleotide sequences of the M genes of C/Yamagata/64 (YA/64), C/Aomori/74, C/Aichi/1/81, and C/Yamagata/1/88 (YA/1/88). The C/Ann Arbor/1/50 (AA/50) virus M gene (Yamashita et al., 1988) is used as the reference sequence and only nucleotides which differ from that sequence are shown. The sequences are given in the mRNA sense. The 5' twelve nucleotides indicated by asterisks were derived from the synthetic primer sequence. Nucleotides at the 3' end that were not contained in the cDNA clones are indicated by asterisks. The ATG codon responsible for the initiation of the M1 protein is boxed with solid lines, and the three ATG
polypeptide with an $M_r$ of 18K that contains at least a part of the 132 amino acid domain is synthesized in influenza C virus-infected cells.

**Methods**

**Viruses and cells.** Four influenza C strains (C/Yamagata/64, C/Aomori/74, C/Aichi/1/81, and C/Yamagata/1/88) isolated in Japan between 1964 and 1988 were each cloned twice by the limiting dilution method either in the amniotic cavity of 9-day-old embryonated hen's eggs (C/Yamagata/64, C/Aomori/74, C/Aichi/1/81) or in the HMV-II line of human malignant melanoma cells (C/Yamagata/1/88) and then propagated in the amniotic cavity of eggs. Virus was purified and virion RNA was extracted as described previously (Palese & Schulman, 1976). HMV-II cells were grown in RPMI 1640 medium containing 10% bovine serum.

**Cloning and sequencing of RNA segment 6.** Double-stranded cDNAs were synthesized from viral RNA by using the TimeSaver cDNA Synthesis Kit (Pharmacia) and a synthetic oligonucleotide primer, 5' d(AGCGAAGCAGG) 3', complementary to positions 1 to 12 at the 3' end of all influenza C virus RNAs. RNA 6-specific cDNAs were isolated, inserted into the EcoRI site of plasmid Bluescript II SK (+) (Stratagene) and then used to transform *Escherichia coli* DH5 cells (Toyobo) (Hongo et al., 1992). Nucleotide sequences of the cloned cDNA were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) using the Sequenase version 2.0 (United States Biochemical Corp.) according to the specification of the manufacturer.

**Isolation of poly(A)+ RNA from infected cells.** Monolayers of HMV-II cells were infected with C/Yamagata/1/88 virus at an m.o.i. of about 10 p.f.u./cell. After incubation at 34 °C for 12 h, total cellular RNA was isolated by a modification of the guanidium thiocyanate procedure (Chirgwin et al., 1979; Okayama et al., 1987). Poly(A)+ RNA was purified from total RNA by two cycles of adsorption onto and desorption from oligo(dT)-coated latex beads (Urabe et al., 1987). Poly(A)+ RNA was eluted from the beads by heating at 65 °C for 5 min in 400 μl of diethyl pyrocarbonate-treated distilled water.

**Northern blot analysis.** Poly(A)+ RNA (1 μg) isolated from virus strain C/Yamagata/1/88-infected HMV-II cells was separated on a 1.5% agarose gel containing 2.2 M-formaldehyde as described by Sambrook et al. (1989). The RNA was transferred to a Gene Screen membrane (DuPont-NEN) by the capillary transfer method and then immobilized by baking. The membrane was hybridized with 32P-labelled RNA 6-specific RNA probe (vRNA sense) that had been synthesized by *in vitro* transcription of pCM5 DNA using T3 RNA polymerase. Plasmid pCM5, which contains nucleotides 3 to 1168 of C/Yamagata/1/88 virus M gene, was linearized by digestion with *Hind*III prior to *in vitro* RNA transcription.

**PCR-amplification of poly(A)+ RNA and Southern blot analysis.** Poly(A)+ RNA from infected cells was first transcribed into cDNA with avian myeloblastosis virus reverse transcriptase XL (Life Sciences) and oligo(dT)$_{12-18}$ (Pharmacia) as a primer. The cDNA was then PCR-amplified (25 cycles) in the reaction mixture (100 μl) containing 10 mM-Tris-HCl, pH 8.3, 50 mM-KCl, 1.5 mM-MgCl$_2$, 0.001% (w/v) gelatin, 200 μM-dNTPs, 1 μM-oligonucleotide primers (positions 677 to 696 in the mRNA sense and positions 1128 to 1109 in the vRNA sense), and 2.5 U of AmpliTag polymerase (Perkin-Elmer Cetus). Each cycle consisted of 1 min of denaturation at 94 °C, 2 min of annealing at 55 °C and 3 min of extension at 72 °C, with a final extension of 7 min at 72 °C. The amplified dsDNAs were separated by electrophoresis on a 3% agarose gel, transferred to a Gene Screen membrane and then hybridized with either the 32P-labelled RNA 6-specific RNA probe described above or the synthetic oligonucleotide 5'dAGATTGCAA-GCAGAACAGG 3' (corresponding to positions 811 to 792 in the vRNA sense) labelled at the 5' end with [γ-32P]ATP (NEN) and T4 polynucleotide kinase (New England Biolabs).

**Ribonuclease protection assay.** 32P-labelled RNA probe (7 × 105 c.p.m.), synthesized by *in vitro* transcription of pCM5 DNA using the T3 promoter, was hybridized with total RNA (30 μg) or poly(A)+ RNA (10 μg) from infected cells at 42 °C overnight in 40 μl-PIPES buffer containing 0.4 M-sodium acetate, 1 mM-EDTA and 80% formamide. Hybrids were digested with RNase A (0.08 U) and RNase T1 (3 U) at 37 °C for 30 min using the Ribonuclease Protection Assay kit RPAII (Ambion) and then analysed on a 4% polyacrylamide gel containing 4 M-urea.

**Production of antiserum to a glutathione S-transferase (GST) fusion protein.** The 399 bp DNA fragment that corresponds to positions 783 to 1181 of RNA segment 6 (a region encoding amino acid residues 253 to 374 of Fig. 2) was obtained by PCR using plasmid pCM5-3'F (see later) as a template and one primer (5'dGGATCCACATGGCTACTCTGGTC 3') containing a BamHI restriction site (underlined) followed by the sequence corresponding to positions 783 to 802 (mRNA sense) and the second primer (5'dGAATTCGACATTACGGGCTCCTCTG 3') containing an EcoRI restriction site (underlined) followed by the sequence corresponding to positions 1115 to 1159 (vRNA sense). This DNA fragment was digested with BamHI and EcoRI and then cloned into the BamHI and EcoRI sites of pGEX-2T (Pharmacia) (pGEX/CMD) to express a region consisting of residues 253 to 374 as a GST fusion protein (GST/CMD). The nucleotide sequence across the junction of the gene fusion was verified by dideoxynucleotide sequencing. Plasmid pCM5-3'F, which contains nucleotides 3 to 1168 of the C/Yamagata/1/88 M gene and nucleotides 1169 to 1181 of the C/Ann Arbor/1/50 M gene, was constructed as follows: the 505 bp DNA fragment corresponding to positions 677 to 1181 was isolated by PCR utilizing pCM5 that had been subcloned into the EcoRI site of Bluescript II KS (+) (pCM5/KS +) as a template and two primers described above or the synthetic oligonucleotide 5'dAGATTGCAA-GCAGAACAGG 3' (corresponding to positions 811 to 792 in the vRNA sense) labelled at the 5' end with [γ-32P]ATP (NEN) and T4 polynucleotide kinase (New England Biolabs).

**PCR-amplification of poly(A)+ RNA and Southern blot analysis.** Poly(A)+ RNA from infected cells was first transcribed into cDNA with avian myeloblastosis virus reverse transcriptase XL (Life Sciences) and oligo(dT)$_{12-18}$ (Pharmacia) as a primer. The cDNA was then PCR-amplified (25 cycles) in the reaction mixture (100 μl) containing 10 mM-Tris-HCl, pH 8.3, 50 mM-KCl, 1.5 mM-MgCl$_2$, 0.001% (w/v) gelatin, 200 μM-dNTPs, 1 μM-oligonucleotide primers (positions 677 to 696 in the mRNA sense and positions 1128 to 1109 in the vRNA sense), and 2.5 U of AmpliTag polymerase (Perkin-Elmer Cetus). Each cycle consisted of 1 min of denaturation at 94 °C, 2 min of annealing at 55 °C and 3 min of extension at 72 °C, with a final extension of 7 min at 72 °C. The amplified dsDNAs were separated by electrophoresis on a 3% agarose gel, transferred to a Gene Screen membrane and then hybridized with either the 32P-labelled RNA 6-specific RNA probe described above or the synthetic oligonucleotide 5'dAGATTGCAA-GCAGAACAGG 3' (corresponding to positions 811 to 792 in the vRNA sense) labelled at the 5' end with [γ-32P]ATP (NEN) and T4 polynucleotide kinase (New England Biolabs).
3506  S. Hongo and others

**Results**

**Sequence analysis of the M genes of four different influenza C isolates**

The M gene sequences of four influenza C strains isolated in Japan over a period of 24 years were determined. The results are shown in Fig. 1, along with the previously reported sequence of C/Ann Arbor/1/50 M gene (Yamashita et al., 1988). None of the cDNA clones analysed here were full-length copies of RNA segment 6, each lacking 13 to 15 nucleotides at the 3' terminus (mRNA sense). Compared with the M gene of C/Ann Arbor/1/50, an insertion of one extra A residue occurred in the 5' non-coding region of the C/Yamagata/1/88 M gene (position 20) whereas no deletions or insertions were detected in the M genes of the other strains. Nucleotide substitutions in the M genes were observed at 52 positions (4.4 %), including an insertion at position 20, and were distributed uniformly throughout the molecule. The nucleotide sequences around the 5' and 3' end junctions of the intervening region of the M1 mRNA are conserved almost completely among the M gene sequences of the five influenza C viruses. The carboxy-terminal amino acid of the M protein is indicated by an arrowhead. The hydrophobic domains that can interact with a lipid bilayer are underlined. The possible N-glycosylation sites that lie in the region of residues 243 to 374 are boxed.

**Radioimmunoprecipitation (RIP).** The HMV-II cells infected with C/Yamagata/1/88 at an m.o.i. of 10 p.f.u./cell were labelled for 1 h at 42 °C post-infection with 15 μCi/ml of [35S]methionine in RPMI 1640 medium that lacked methionine. Cells were then disrupted in 0.01 M-Tris–HCl, pH 7.4 containing 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS and 0.15 M-NaCl and immunoprecipitated with either antisemur against the GST/CM2 fusion protein or rabbit antisemur against the C/Ann Arbor/1/50 virions according to procedures previously described (Sugawara et al., 1986). The immunoprecipitates obtained were analysed on 17.5% polyacrylamide gels containing 4 M-urea and processed for analysis by fluorography (Yokota et al., 1983).
Influenza C virus M gene products

Analysis of RNA segment 6 transcripts

To investigate the possibility that a collinear transcript of RNA segment 6 is synthesized in influenza C virus-infected cells, a Northern blot analysis of poly(A)+ RNA isolated from C/Yamagata/1/88 strain-infected HMV-II cells was made. As seen in Fig. 3(a), however, only a single intense band of ~1100 nucleotides that presumably represents a spliced M1 mRNA was detected. To further examine this possibility, poly(A)+ RNA from infected cells was reverse-transcribed into cDNA using oligo(dT)12-18 as a primer, and PCR-amplified DNA was prepared using the cDNA as a template and a pair of oligonucleotide primers that flank the intervening region (positions 755 to 982) of M1 mRNA. The PCR products were separated by gel electrophoresis and then analysed by Southern blotting with a pCM5-derived RNA probe (Fig. 3b). In addition to an intense band that corresponds in size to the amplified DNA of a spliced M1 mRNA, a faint band of ~430 nucleotides was detected. The latter band may represent the DNA product (expected size 452 nucleotides) derived from an unspliced RNA 6 mRNA. To confirm that the PCR-amplified DNA of ~430 nucleotides contains the sequence corresponding to the intervening region of M1 mRNA, a Southern blot analysis was done with a synthetic oligonucleotide (corresponding to positions 811 to 792 in the vRNA sense) end-labelled with [γ-32P]ATP as a probe. Fig. 3(c) indicates that this probe hybridized with amplified DNA of ~430 nucleotides, without hybridizing with that of ~200 nucleotides. Additionally, sequencing of the larger DNA fragment that had been cloned into the EcoRV site of pT7 Blue T vector (Novagen) confirmed again that it contained the sequence corresponding to nucleotides 755 to 982 of RNA 6 (data not shown).

The results described above suggest strongly that in addition to a spliced M1 mRNA, small quantities of an
Identification of the second protein encoded by RNA segment 6

To search for a putative polypeptide synthesized from an unspliced RNA 6 mRNA, extracts of the C/Yamagata/1/88 virus-infected HMV-II cells labelled with [35S]methionine for 1 h at 42 h post-infection were subjected to immunoprecipitation with rabbit immune serum raised against the GST/CM2 fusion protein. As clearly seen in Fig. 5, this antiserum immunoprecipitated a polypeptide of apparent Mr ≈ 18K, designated CM2, from infected cells (Fig. 5 lane 5) but not from uninfected cells (lane 4). It was also interesting that the GST/CM2 fusion protein antiserum precipitated, in addition to CM2, a material heterogeneous in electrophoretic mobility (Mr ≈ 22K to 30K).

Discussion

Comparison of the nucleotide sequence of RNA segment 6 of five different influenza C isolates (including C/Ann Arbor/1/50) revealed that the deduced amino acid sequence of the M1 protein is conserved almost completely over a period of 38 years. This was not unexpected since we observed previously that peptide maps of the M1s from seven different strains were indistinguishable from each other (Sugawara et al., 1983) and that there were no differences in the reactivity with monoclonal antibodies among the M1s of 23 strains isolated over a 41 year period (Sugawara et al., 1991). The extensive M1 protein sequence conservation is presumably due to functional constraints that could arise from possible interactions with other viral components such as the lipid envelope, the HE glycoprotein and the ribonucleoprotein complex. It was also evident that the unique 132 amino acid domain of the protein that could be generated from a collinear transcript of RNA 6 is conserved although it is more divergent than the M1 protein, and this is consistent with the hypothesis that the unspliced mRNA may be translated into a functional protein that contains this extra carboxy-terminal domain. It is interesting to note here that the M2 proteins of human influenza A viruses have been shown to be much more divergent than the M1 proteins (Buckler-White et al., 1986; Ito et al., 1991).

Previously, Yamashita et al. (1988) analysed RNA 6 mRNAs from C/Ann Arbor/1/50 virus-infected MDCK cells using several different techniques but could not obtain evidence for the presence of a full-length unspliced mRNA. By contrast, it was demonstrated here by Southern blot analysis of the PCR-amplified DNA derived from RNA 6 mRNAs that there was a small amount of unspliced mRNA, in addition to a large amount of spliced M1 mRNA, in C/Yamagata/1/88.
virus-infected HMV-II cells. When quantitated by ribonuclease protection assays, the unspliced mRNA was found to be about 13% as abundant as the spliced one. It is unlikely that all of the collinear transcripts detected represent full-length template RNA that may have contaminated the poly(A)\(^+\) preparation; the amount of full-length template RNA synthesized in influenza C virus-infected cells should be very low compared with the amount of unspliced mRNA, since ribonuclease protection assays with total RNA and poly(A)\(^+\)RNA yielded results virtually identical to each other (Fig. 4).

It is unlikely that a 42K protein corresponding to the entire M1 coding sequence is produced from unspliced RNA 6 mRNA since no polypeptide other than M1 could be immunoprecipitated from the lysates of influenza C virus-infected cells with any of the nine anti-M1 monoclonal antibodies tested (Sugawara et al., 1991). It was demonstrated however that a polyclonal rabbit antiserum raised against the purified GST/CM2 fusion protein immunoprecipitated a polypeptide with an apparent \(M_r\) \(\approx\) 18K (CM2) from C/Yamagata/1/88 virus-infected HMV-II cells, without precipitating a 42K protein. In addition to CM2, material that migrated heterogeneously on polyacrylamide gels (\(M_r\) \(\approx\) 22K to 30K) was also precipitated with this immune serum, an observation that suggests that the CM2 protein may be modified, e.g. by the processing of carbohydrate chains, to a heterogeneous form of higher \(M_r\). The mechanism by which the CM2 protein is generated from the unspliced mRNA remains to be understood. Just upstream from the 5’ donor splice site, however, three initiation codons lie in the same reading frame as that used for M1, at nucleotide positions 732 to 734, 741 to 743, and 747 to 749 (Fig. 1). One of these AUGs (positions 732 to 734) is in the context considered most favourable for ribosome initiation (A/GNNAUGG) (Kozak, 1986). Recently, evidence has been presented which suggests that Sendai virus X protein is initiated by ribosomes \(\sim\) 1500 nucleotides from the cap group, in a cap-dependent fashion (Curran & Kolakofsky, 1988) and that translation initiation by an internal ribosome-binding mechanism is used by the mRNA encoding immunoglobulin heavy chain binding protein (Macejak & Sarnow, 1991). It is not impossible, therefore, that the collinear mRNA might be translated from the AUG initiation codon at nucleotides 732 to 734 to the termination codon at nucleotides 1149 to 1151, yielding a protein composed of 139 amino acids (residues 236 to 374 of Fig. 2).

Examination of the predicted sequence of the putative 139-amino acid protein suggests that it shares some structural features with influenza A virus M2 protein. For example, the region spanning residues 287 to 318 (numbering corresponds to that shown in Fig. 2) contains only two charged amino acids (K306, E310) and has a hydrophathy index of \(> 2\) (data not shown), a value normally found for regions that interact with membranes. The sequence of this region, like that of the transmembrane domain of influenza A virus M2 (Buckler-White et al., 1986; Ito et al., 1991), is strikingly conserved among the influenza C isolates analysed, with only one substitution (I to T) at position 308 of C/Yamagata/64. Inspection of the predicted sequence of the region consisting of residues 253 to 286 indicates the presence of three conserved cysteines (positions 260, 265, and 279), in addition to the presence of one or two potential sites for N-glycosylation (residues 270 and 281). This region is similar in these respects to the extracellular domain of influenza A virus M2 protein (Lamb, 1989). The cytoplasmic domain of M2, composed of 54 carboxy-terminal amino acids, is strikingly hydrophilic, and is modified by the addition of palmitate to a cysteine residue in most of influenza A virus strains (Sugrue et al., 1990; Veit et al., 1991). The carboxy-terminal region of the putative 139 amino acid protein is also hydrophilic, 21 out of the 56 amino acids being charged ones. The region also contains a cysteine residue at position 324 that is conserved among all the influenza C strains analysed. The amino-terminal region consisting of residues 236 to 252 has no counterpart in the M2 protein of influenza A virus. This region contains a strongly hydrophobic domain (residues 241 to 252), and might represent the leader sequence of the putative protein.

We thank T. Itoh (Yamagata University Laboratory Animal Center) for excellent technical assistance, and Drs K. Nakajima (Nagoya City University), T. Odagiri (Jichi Medical School) and M. Nakamura (Tohoku University) for helpful discussions. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education Science and Culture, Japan, and the Naito Foundation.

References


(Received 19 April 1994; Accepted 13 July 1994)