Strain-variable editing during transcription of the P gene of mumps virus may lead to the generation of non-structural proteins NS1 (V) and NS2

G. D. Elliott, R. P. Yeo, M. A. Afzal, E. J. B. Simpson, J. A. Curran and B. K. Rima*

School of Biology and Biochemistry, Medical Biology Centre, The Queen's University of Belfast, Northern Ireland BT9 7BL, U.K.

The sequence of the P (phosphoprotein) gene of mumps virus has been determined. It has two open reading frames, the first of which probably encodes the NS1 (or V) protein of mumps virus. Expression of the P protein requires the insertion of two non-templated residues to link the two ORFs in a process analogous to that observed in the P/V gene of simian virus type 5 to which mumps virus is closely related. Strain differences in the accuracy of insertion of non-templated G residues in the P/V gene transcripts have been described.

Introduction

Non-structural proteins have been described in para-
myxoviruses such as Sendai virus (Lamb & Choppin, 1978), Newcastle disease virus (NDV) (Collins et al., 1982; Chambers & Samson, 1982), simian virus type 5 (SV5) (Peluso et al., 1977) and mumps virus (Rima et al., 1980) as well as in the morbilliviruses (Rima & Martin, 1979; Vainionpaa, 1979; Campbell et al., 1980) and in the pneumovirus respiratory syncytial virus (Huang & Wertz, 1982; Venkatesan et al., 1983). The function of these proteins remains unclear, but the manner in which they are generated has been established in most members of the paramyxoviridae. In the pneumoviruses, separate genes appear to exist that code for some of the non-
structural proteins (Collins et al., 1984). In viruses belonging to the Sendai virus subgroup of the paramyxoviruses and in the morbilliviruses, non-structural 'C' proteins are generated from translation of the P (phosphoprotein) messenger RNA in different frames (Giorgi et al., 1983; Bellini et al., 1985; Barrett et al., 1985). In the case of Sendai virus multiple C proteins with different starting codons have been observed (Curran & Kolakofsky, 1988; Vidal et al., 1990 and references therein). These non-structural proteins have no amino acid sequence similarity to the P protein. However, in NDV (Collins et al., 1982) and mumps virus (Herrler & Compans 1982; Simpson et al., 1984) the non-structural proteins appear to share peptides with the P protein.

Recently, the non-structural V protein of SV5 has been shown to be translated from the unaltered transcript of the P gene, whereas the generation of the P protein requires the insertion of non-templated residues into the P gene transcripts (Thomas et al., 1988). A similar insertional mechanism has been shown to generate a 'V'-like protein from the P gene of morbilliviruses (Cattaneo et al., 1989).

Although mumps virus appears to be very closely related to SV5 (Rima, 1989) two recent papers which report the sequence of the P gene of mumps virus (Takeuchi et al., 1988; Elango et al., 1989) suggest that the generation of the non-structural proteins of mumps virus involves internal translational start sites. We provide here nucleotide sequence evidence that suggests that the generation of these proteins in mumps virus is similar to the process demonstrated for SV5. This provides an explanation for the generation of the non-structural proteins of mumps virus.

Methods

Cells and viruses. Mumps virus was grown either in 8 day old embryonated eggs for the SBL1/E (egg-adapted) strain or in Vero cells for the SBL1/V Vero cell-adapted strain, the Enders strain and the BF strain. The SBL1 strains were a generous gift from Dr C. Örvell (Statens Bakteriologiska Laboratoriet, Stockholm, Sweden). The Enders and BF strains have been described before (Rima et al., 1980).

Cloning, sequencing and polymerase chain reaction. Cloning and sequencing of cDNA was carried out as described before (Elliott et al., 1989). For the polymerase chain reaction (PCR) poly(A)+ RNA was extracted from infected cells as described before (Elliott et al., 1989). Genomic RNA was obtained by extraction of RNA from purified nucleocapsids obtained from infected Vero cells. Nucleocapsids were purified according to the method of Kolakofsky (1976) as modified by Takeuchi et al. (1988). The extracted RNA species were reverse-transcribed with specific primers for the synthesis of the complemen-
tary strand in 25 μl reaction mixtures. To make the complementary positive strand from genomic RNA we utilized a primer with the sequence 5' GGAAAAGAGAGAATGATTA 3' (nucleotides 461 to 479). To prime from mRNA we used 5' CAAGATGTTGCAGGC- GAGC 3' (complementary to nucleotides 774 to 756). After first-strand synthesis the resulting RNA/DNA hybrid was subjected to PCR. The reverse transcription reaction mixture was diluted with PCR buffer, Taq DNA polymerase (0.5 units), the second oligonucleotide and deoxynucleotides to a final volume of 100 μl and PCR was carried out according to the manufacturer's instructions (Perkin-Elmer Cetus). The amplified products were cut with the appropriate restriction enzymes (ClaI and Sau3A) and ligated into BamHI- and Accl-cut M13 mp9 and M13 mp8 for sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977).

Results

cDNA cloning of the P gene of mumps virus

The non-structural proteins NS1 (23K) and NS2 (18K) of mumps virus have been shown to share tryptic peptides when analysed by HPLC (Herrler & Compans, 1982). Two-dimensional fingerprinting (Simpson et al., 1984) has shown that the [35S]methionine-labelled tryptic peptides in NS2, NS1 and the P protein form a nested set.

In order to elucidate how the non-structural proteins were generated we decided to study the P gene of mumps virus further by cDNA cloning and nucleotide sequence determination. cDNA clones prepared from the purified genome of SBL1/E have been generated and sequenced as described before (Elliott et al., 1989). Fig. 1 describes the subcloning and sequencing strategy employed for this gene. cDNA clones generated from Enders virus-infected cell mRNA were also available in this study (Elliott et al., 1989 and further clones generated from such material for this study). Two mumps virus-specific cDNA clones for almost all parts of the P gene were generated and these have been sequenced in both directions. Fig. 2 shows the sequence determined from the genomic cDNA clones of the P gene of SBL1/E.

This sequence shows the presence of two open reading frames (ORFs) covering the 5' end to the 3' end of the gene and varies from those published earlier for SBL1/V (Elango et al., 1989) and the Enders and Miyahara strains (Takeuchi et al., 1988) particularly between residues 521 and 560 as detailed in Fig. 3. It was clear that nucleotides 531 to 536 of the sequence that we have found to be present in genomic clones of SBL1/E were the same as those published for SBL1/V (Elango et al., 1989) but that nucleotides 543 to 550 were the same as those presented for the Enders and Miyahara strains reported by Takeuchi et al. (1988). One of the clones of an mRNA extracted from Enders virus-infected cells had nine G residues in the mRNA sequence after position 530 instead of the eight proposed by Takeuchi et al. (1988) and the six G residues in the genomic cDNA clones. Thus we decided to determine the nucleotide sequence of a larger number of genome- and mRNA-derived cDNA clones around this site.

Sequence variation around the insertion site

We synthesized oligonucleotide primers at either side of the region of interest and carried out PCR using cDNA reverse-transcribed from infected cell RNA oligo(dT)-selected twice, or with genomic RNA extracted from purified nucleocapsid RNA so that the generation of genomic or mRNA cDNA clones was assured. The amplified gene products were cut with ClaI and Sau3A restriction enzymes, force-cloned into M13 mp8 and mp9 and sequenced using the dideoxynucleotide method. A total of 40 clones of genomic cDNA from the Enders, SBL1 and BF strains of mumps virus were analysed and
Fig. 2. Nucleotide sequence of the P gene of mumps virus. The sequence represented is the positive antigenome. Possible translation products are indicated.
Fig. 3. Sequence variation in the published sequences around the insertion site in the mumps virus P gene. MIY and END1, Miyahara and Enders strain sequences determined by Takeuchi et al. (1988). These probably represent edited transcripts and not antigenomic sequences; SBL1, sequence of the Vero cell-grown SBL1 strain determined by Elango et al. (1989); SBL2, sequence of the egg-grown SBL1 strain (this work); END2, Enders strain sequence (this work); BF, sequence of the BF strain of mumps virus (this work). The sources of RNA used are indicated as either M (mRNA) or G (genomic RNA) or both. The amino acid sequence of the Miyahara strain and expressed mutations is indicated.

Fig. 4. Sequence variability at the insertion site. Cloned amplified product from the mRNA extracted from Enders virus-infected Vero cells was sequenced. The G reaction products of 14 subclones are shown. Clones with insertions containing one, two, three or four extra G residues are presented. Clones with five or more inserted G residues have not been observed.

Table 1. Strain-dependent sequence variation at the editing site of mumps virus

<table>
<thead>
<tr>
<th>Number of G residues inserted</th>
<th>Protein encoded</th>
<th>Enders strain</th>
<th>SBL1 strain</th>
<th>BF strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NS1</td>
<td>12*</td>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>1</td>
<td>NS2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>P</td>
<td>12</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>NS1-like</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>NS2-like</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

* Number of cDNA clones with the indicated number of G residues.

Table 1 shows that cDNA clones with more than four G residues inserted were not observed among the 153 clones analysed.

It was apparent that the SBL1 strain was more accurate in inserting only the correct two G residues than the other two strains. This insertion process has been referred to as editing of the transcripts (Thomas et al., 1988; Cattaneo et al., 1989; Vidal et al., 1990). It was clear that the SBL strain was accurate in this process and very few misedited transcripts were observed. The Enders strain appeared to be much more prone to misediting. These results were observed in a number of experiments with the different strains and with RNA samples extracted at different times post-infection. These, therefore, appeared to represent real strain-specific differences in the editing phenomenon. They also implied that the editing function is probably virus- and not host-encoded. Very few edited P-encoding transcripts were observed when the BF strain was analysed. These data are discussed below.
**Discussion**

The sequence of the P gene and its transcripts have allowed us to suggest how the generation of the NS1 protein takes place in mumps virus. We have found that at position 531 in the sequence given in Fig. 2 there are a variable number of extra G residues incorporated in the mRNA transcripts. These will give rise to mRNA molecules that encode the P and NS2 proteins, whilst direct unedited transcripts from the P/NS1 gene encode the NS1 (V) protein. It has previously been shown that the NS proteins contain a nested set of tryptic peptides also present in the P protein (Herrler & Compans, 1982; Simpson et al., 1984). These NS proteins were shown to have no precursor–product relationship and were primary translation products from mRNA molecules similar in size to that of the P protein (Simpson et al., 1984). The identification of a cysteine (C)-rich domain in the NS1 protein, the absence of C in the NS2 protein and the very low C content of the P protein (Simpson et al., 1984) are inconsistent with the suggestion of Takeuchi et al. (1988) and Elango et al. (1989) that the NS proteins are generated from the P mRNA by internal starts of translation. We conclude therefore that the mumps virus NS proteins and P proteins are generated from the mRNA by the same process as has been shown to be responsible for the synthesis of the V and P proteins of SV5, but with the difference that the inaccuracy of the process or misediting in some strains also appears to be able to lead to the generation of another protein, NS2. In about 40% of the transcripts two extra G residues have been inserted to generate transcripts that contain an ORF of sufficient size to encode the P protein.

The derived amino acid sequence of the mumps virus P protein indicates very high levels of homology (37%) between mumps virus and SV5, 19% homology between mumps virus and NDV and insignificant levels of homology with the other paramyxoviruses. This is further evidence for the proposed close relationship between SV5, NDV and mumps virus (Rima, 1989).

Hightower et al. (1984) reported that the non-structural proteins of NDV shared amino-terminal peptides with the NDV P protein. Recently McGinnies et al. (1988) suggested that the non-structural proteins of NDV were generated from the same ORF as the P protein. The mechanism responsible for the generation of NS proteins in NDV, the third member of this subgroup of the paramyxoviruses (Rima, 1989) is not yet clear. It will be interesting to establish the precise sequence of the P gene and its transcripts for this and other paramyxoviruses. The alignment of protein sequences of the paramyxoviruses reveals no conservation of amino-acid residues in the P proteins (Rima, 1989) or in the C proteins, whereas in the V/NS1-like proteins there are a substantial number of conserved residues (Thomas et al., 1988). Thus, it may well be that the shorter V-like proteins expressed from this gene are functionally important, as indicated by the level of transcripts encoding them.

The incorporation of non-templated residues appears to be due to the presence of sequences, related to the polyadenylation signal, containing all the bases normally found in front of a set of U residues at which the polyadenylase, by slippage, generates the poly(A) tails at the end of mRNA, except that these are replaced by C residues in the insertion site. Sequences similar to polyadenylation sites have also been identified in the other paramyxoviruses in which the phenomenon of the incorporation of non-templated residues has been documented (Thomas et al., 1988; Cattaneo et al., 1989) or postulated. Thomas and coworkers (1988) suggested that the secondary structure of the RNA template around the insertion site was important. However, we were not able to fold the template around the switch site into a thermstable structure similar to the one proposed for SV5.

The data presented here indicate strain variability in the generation of the NS1 and NS2 proteins. In the SBL strain only transcripts encoding the NS1 and P proteins were detectable in roughly a 2:1 ratio and this correlated with the synthesis of these proteins in vivo. In the Enders strain it was noted that the process of insertion of non-template residues was less accurate and transcripts with between one and four inserted G residues were observed. The ones with one or four residues were supposed to give rise to the NS2 protein and this was demonstrated in Enders virus-infected cells. Earlier we reported that the Belfast (BF) strain of mumps virus induced synthesis of greater amounts of NS proteins in infected cells than the Enders strain. Scanning of 14C-labelled protein profiles in SDS–PAGE showed that in the BF strain the relative molar proportions of P, NS1 and NS2 protein were 1-0, 1-8 and 0-5 respectively and in the Enders strain the proportions were 1-0, 1-0 and 0-7 (Simpson et al., 1984).

The values for the Enders strain correlate well with the relative numbers of transcripts for each of the proteins and indicate strain variation in the ability to generate various non-structural proteins. However, the data for the BF strain were surprising in the sense that the number of transcripts that would putatively encode the P protein was rather smaller than expected. The data given are the results of repeated experiments and extractions of RNA. We have shown that the mRNA fraction that was analysed is capable of generating the P protein in in vitro translation and thus we cannot yet conclude that insertion of G residues into transcripts is the only mechanism for the generation of the P and NS2 proteins. Further experiments with specific antisera are under way to elucidate this further. We do not consider that PCR is
the cause of the aberrant results (low levels of P transcripts) seen for the BF strain, because we have not observed any problems in the other strains and the genomic RNA sequence of the BF strain. Direct cloning of the region without PCR has also demonstrated that insertion occurs at the editing site. In many ways the difficulty with the results for the BF strain are opposite to those experienced with Sendai virus where 31% of the transcripts seem to encode a V-like protein instead of the P protein, but no such protein has been found in vivo (Vidal et al., 1990). In the measles virus editing system we have recently been able to identify the X protein observed earlier (Rima et al., 1981) as the V-like cysteine-rich translation product in infected Vero cells.

In conclusion, by establishing the sequence of the genome of the SBL strain of mumps virus we have obtained indications that the organization of the ORFs and expression of the P/V gene of mumps virus is similar to that of SV5, except that additional proteins are generated in mumps virus-infected cells by the reduced accuracy of insertion of non-templated residues in different strains. The proposed mechanism not only explains the data in our earlier paper (Simpson et al., 1984), which we found difficult to explain using the models of gene expression then available, but also the distribution of cysteine over the P and NS proteins, whereas the suggestion in earlier reports that the latter were generated from internal starts of translation does not. The observed variability in the transcripts from a single genomic sequence showing different numbers of G residues inserted after position 531 supports the interpretation that mumps virus uses a similar mechanism to SV5 and measles virus for the expression of non-structural V-like proteins.

We thank the Medical Research Council (U.K.) for grant support and the Government of Pakistan for studentship support to M.A.A.

References


(Received 22 December 1989; Accepted 6 March 1990)