Fusion Glycoprotein of Measles Virus: Nucleotide Sequence of the Gene and Comparison with Other Paramyxoviruses

By ROBIN BUCKLAND, CHRISTOPHE GERALD, RICHARD BARKER and T. FABIAN WILD

SUMMARY

The sequence of the fusion (F) glycoprotein mRNA of the Hallé strain of measles virus was determined from a cDNA clone representing the entire length of the mRNA. It contained 2384 nucleotides, excluding poly(A), with a 5' consensus sequence typical of paramyxoviruses and a 3' terminus found in measles virus mRNAs. The coding sequence was preceded by an unusually long (580 nucleotide) 5' non-translated region, which contained 44% cytosine. The longest open reading frame coded for a polypeptide of 553 amino acids with a predicted molecular weight of 59.84k. Comparison of the sequence with that of the Edmonston strain of measles virus showed that the gene is highly conserved. No amino acid differences were observed between the two strains.

The F polypeptide had three regions of high hydrophobicity: an N-terminal signal peptide, the N-terminus of F1 and a C-terminal membrane-spanning region. The four potential asparagine-linked glycosylation sites (one in the signal peptide) were all in the F2 subunit. Comparison of the measles virus F amino acid sequence with other paramyxoviruses revealed homologies with these viruses. Certain regions such as the N terminus of F1 and ten cysteine residues which probably impose structural restraints were highly conserved.

INTRODUCTION

The paramyxoviruses have been extensively studied in recent years as they provoke serious disease in children and may initiate persistent infections. In order to characterize these viruses, a number of laboratories have studied the organization and expression of the virus genes. The paramyxoviruses belong to the negative strand RNA virus group, and recently the prototype of the group, Sendai virus, has been completely sequenced (Blumberg et al., 1984, 1985a, b; Dowling et al., 1983; Shioda et al., 1986). The gene order was determined as 3'-nucleoprotein (NP)-phosphoprotein (P)-matrix (M)-fusion (F)-haemagglutinin (HA/HN)-large protein (L)-5'.

Similar studies on measles virus, which is one of the most important members of this group, are incomplete. Cloning and sequence analysis of the virus genome has confirmed the same order as Sendai virus for the first three genes (Rozenblatt et al., 1985; Bellini et al., 1985, 1986). Indirect evidence from transcriptional mapping has confirmed the position of the remaining genes (Dowling et al., 1986; Rima et al., 1986). Two of the remaining genes, HA and F, are particularly interesting as they code for the two virus glycoproteins. The HA is the virus protein responsible for cell attachment (Walsh et al., 1984) and is the dominant antigen to which neutralizing antibodies are directed (Giraudon & Wild, 1985; Norrby, 1985). The F protein is responsible for cell penetration (Merz et al., 1980). It is synthesized as a precursor polypeptide...
F0 and is cleaved by cellular proteases to give the biologically active polypeptide F1,2. In order to study the biological function of the measles virus glycoproteins and their relative importance in immunization we have cloned their corresponding mRNAs. In the present paper we have determined the nucleotide sequence and the predicted amino acid sequence for the F protein. The protein is shown to have a high homology with other paramyxoviruses. Its mRNA has an unique and unusual 5' non-coding region which may have a role in translation control.

After the completion of this work, Richardson et al. (1986) published the sequence of the F gene for the Edmonston strain of measles virus. We have compared this sequence with that of the Hallé strain.

**METHODS**

**Cells and virus.** Vero cells were cultivated in Eagle's MEM containing 5% foetal calf serum plus antibiotics (100 units/ml penicillin; 100 μg/ml streptomycin). The Hallé strain of measles virus (Horta-Barbosa et al., 1971) was plaque-purified twice in Vero cells; stocks were made from this virus and stored at -70 °C.

**Preparation of RNA and construction of cDNA library.** The extraction of the RNA from measles virus-infected cells and the preparation of the cDNA from mRNA by a modified Okayama & Berg (1982) method was performed as previously described (Gerald et al., 1986).

**Identification of measles virus F clones.** Escherichia coli transformants were screened for the presence of measles virus F cDNA by in situ colony hybridization (Grunstein & Hogness, 1975) with a 32P-labelled nick-translated 650 bp cDNA fragment of the F gene. This cDNA was kindly provided by Drs B. M. Blumberg and P. C. Dowling (Medical School, Newark, N.J., U.S.A.) and corresponds to a fragment near the 3' end of the F mRNA. Colonies that gave positive hybridization signals were grown in L broth containing 50 μg/ml ampicillin and their plasmid DNAs were isolated by standard procedures.

**Nucleic acid sequencing.** Nucleotide sequences were determined by the dideoxy chain-terminating method (Sanger et al., 1977). Because of the high GC content of the 5' non-coding region, this part of the sequence was confirmed using the method of Maxam & Gilbert (1980).

**RESULTS AND DISCUSSION**

**Construction and identification of cDNA clones**

mRNA from measles virus-infected Vero cells was annealed to the oligo(dT)-tailed primer of vector pCD-1. cDNA was synthesized and the construction completed as described in Methods. The cDNA plasmids were transfected into E. coli MC 1061 and bacterial transformants were screened with an F-specific cDNA. A number of positive clones were selected and the sizes of their inserts were determined on agarose gels after digestion with BamHI. Digestion with BamHI leaves 30 nucleotides at the 5' end and 100 nucleotides at the 3' end of plasmid DNA attached to the cDNA. The clone with the largest insert, pCD-F.A, after digestion with BamHI gave two fragments corresponding to 2200 and 350 bp. Northern blot analysis of measles virus RNA showed that these cDNAs hybridized with an mRNA of approximately 2400 nucleotides which has been postulated to be the F mRNA (Dowling et al., 1986; Rima et al., 1986). This pCD-F.A clone was selected for nucleotide sequence analysis.

**Analysis of the nucleotide sequence**

The nucleotide sequence of the pCD-F.A clone is shown in Fig. 1. The cDNA contained 2384 nucleotides excluding poly(A). The first three nucleotides (TCG) at the 5' end of the F cDNA are followed by the consensus sequence AGGGCCAAGG, which is similar to that found at the beginning of the other measles virus mRNAs (Bellini et al., 1985, 1986; Gerald et al., 1986; Rozenblatt et al., 1985) and those of Sendai virus (Gupta & Kingsbury, 1984). Thus, it is highly probable that this clone represents a full length copy of the F mRNA. The 3' polyadenylated region is preceded by the sequence TTAATT which is similar to that in the measles virus NP, P and HA mRNAs.

**Coding region**

The first and longest open reading frame starts with the ATG at nucleotides 581 to 583 and terminates with the TGA at nucleotides 2240 to 2242. This could code for a protein of 553 amino acids. However, the context around this ATG codon, CAAATGT, is less favourable for protein
Measles virus fusion protein sequence

The deduced amino acid sequence is shown. The presumptive signal sequence and membrane anchorage region are underlined. Potential glycosylation sites are boxed. The point of cleavage of F0 to F1,2 is shown by an arrow.
Fig. 2. Hydropathy plot of the measles virus fusion protein. The relative hydrophobicities and hydrophilicities were calculated as described by Kyte & Doolittle (1982) using a segment length of nine amino acids. The cleavage point of F0 is shown by the arrow.

synthesis than that of a second ATG codon, which follows in the same reading frame three codons later, ATCATGG (Kozak, 1983). Several viruses have two possible initiating codons in the same reading frame (see Kozak, 1986). Both methionines have been shown to be used when the first is not optimal for initiation. The N terminus of the predicted measles virus polypeptide has the properties ascribed to a signal peptide (von Heijne, 1985). The predicted cleavage site would be the glutamine at either position 22 or 27. This would produce an F protein containing either 531 (57.47K) or 526 (56.96K) amino acids respectively. Alternatively, if the signal peptide is not cleaved, as in certain cellular glycoproteins such as rhodopsin (Schechter et al., 1979), the F protein would have 553 amino acids corresponding to a molecular weight of 59.84K.

The glycosylated form of the uncleaved F0 or F1,2 measles virus protein has been estimated by SDS–PAGE to be 60K (Hardwick & Bussell, 1978; Rima & Martin, 1979; Stallcup et al., 1979). Our sequence determination predicts four glycosylation sites, at the asparagines at positions 9, 32, 64 and 70, of which the first is in the proposed signal peptide and would be lost if cleavage takes place and the last is probably not used as it is followed by a cysteine which is involved in a disulphide linkage. Thus, taking into account the glycosylation, the predicted size of the F protein is slightly greater than that determined by PAGE.

The hydropathy profile (Fig. 2) reveals a hydrophobic region consisting of 29 amino acids (492 to 520) close to the C terminus which is probably the transmembrane region. This would leave 33 amino acids on the cytoplasmic side. Three serines and a threonine are situated towards the outside and three cysteines towards the inside of the transmembrane region. These may serve as receptors for acylation, as radiolabelling measles virus-infected cells with [³H]palmitic acid specifically labels the F protein (T. F. Wild, unpublished results). Studies with Sindbis and Semliki Forest viruses have shown that the serine and threonine residues in the transmembrane region are acylated (Schmidt, 1982; Rice & Strauss, 1981), whereas Chambers et al. (1986) have correlated acylation in Newcastle disease virus with the presence of cysteine in this region.

The cleavage of the F0 protein of Sendai virus to F1,2 by cellular proteases gives rise to the biologically active form (Scheid & Choppin, 1977). The involvement of the N terminus of the F1 of measles virus in the fusion activity has been inferred from results of experiments with short peptides (Norby, 1971) and its amino acid sequence has been determined for the LEC strain of measles virus (Varsanyi et al., 1985). This sequence corresponds with a single exception (residue 154, Thr to Arg) to our sequence determined for the Hallé strain. This puts the cleavage site at
Measles virus fusion protein sequence

Arg–Phe at positions 115 to 116. The cleavage would give an F1 polypeptide of 47.03K which is higher than the 40K calculated from PAGE (Mountcastle & Choppin, 1977; Graves et al., 1978; Hardwick & Bussell, 1978; Tyrrell & Norrby, 1978) probably because of its hydrophobic nature (Fig. 2). The F2 polypeptide produced would be either 10.54K or 10.06K (excluding glycosylation) depending upon the cleavage site of the signal peptide. It would contain the two usable glycosylation sites and confirms the previous observations that only the F2 is glycosylated (Hardwick & Bussell, 1978; Tyrrell & Norrby, 1978). The efficiency of the cleavage depends on the sequence preceding this site. The F proteins of viruses such as simian virus 5 (SV5) which contain five arginine residues (Paterson et al., 1984) are more readily cleaved than Sendai virus F, which has a single arginine residue (Blumberg et al., 1985b). The amino acid sequence preceding the measles virus F cleavage site, Arg-Arg-His-Lys-Arg, predicts that the F0 polypeptide would be readily cleaved.

Non-coding regions

The F protein coding region is preceded by 580 nucleotides at the 5' end and has 142 nucleotides at the 3' end [excluding the poly(A)]. The 5' region is unusual in that it contains a high level of cytosine (44%). Secondary structure in this sequence could play a role in controlling translation (Kozak, 1986). However, despite the G+C content of 64%, there is no strong secondary structure prediction for the measles virus mRNA. Although such a sequence is not found amongst the other paramyxovirus F mRNAs, certain cellular mRNAs, such as those encoding human thymidylate synthase (Takeishi et al., 1985), avian muscle creatine kinase (Ordahl et al., 1984) and c-erb-A (Sap et al., 1986) have a high cytosine content in the 5' non-coding region. Thus it may be possible that during evolution, measles virus has recovered a cellular mRNA sequence by mechanisms similar to that proposed for influenza virus transcription (Krug et al., 1979). Alternatively, it may have integrated sequences from CpG-rich regions (Bird, 1986). The role of the 5' non-coding region may be a regulatory one, as removal of this sequence from c-erb-A leads to an increase in translation (Sap et al., 1986).

Comparison with the Edmonston strain

The recently published sequence of the F gene of the Edmonston strain of measles virus (Richardson et al., 1986) permits a comparison with our sequence for the Halle strain, which was isolated from a patient suffering from subacute sclerosing panencephalitis. In the coding region, there were only four nucleotide changes, all of which are silent and do not affect the amino acid coding. Comparing the non-coding regions, and bearing in mind that the 5' terminus is missing from the Edmonston virus sequence, there were 11 changes at the 5' end and two deletions at the 3' end. Thus the F protein is highly conserved.

Comparison with other paramyxoviruses

The size of the F protein of measles virus from our nucleotide sequence analysis and from PAGE is in the same order of magnitude as that reported for other paramyxoviruses, i.e. 529 to 565 amino acids (Blumberg et al., 1985b; Paterson et al., 1984; McGinnies & Morrison, 1986; Spriggs et al., 1986). Thus, it was surprising that the measles virus mRNA was approximately one-third larger. With the exception of parainfluenza 3 virus, which has a 193 nucleotide 5' non-coding sequence (Spriggs et al., 1986), the others contain 29 to 52 additional nucleotides. In measles virus this region may regulate the translation of F, as too great synthesis of the F protein would lead to fusion and cell death.

Alignment of the amino acid sequences of measles virus F with those of other paramyxoviruses to give maximum homology (Table 1) revealed a number of features. The overall homology was 25-3%, with SV5, 26-2% with Sendai virus, 24-2% with parainfluenza 3 virus and 24-0% with Newcastle disease virus. The greatest homology (64 to 78%) was found in the hydrophobic region (112 to 140) responsible for cell fusion activity. Other regions of the polypeptide displayed varying relationships with those of individual viruses (Table 2). Thus, between residues 361 and 400 there is a high homology (46-5%) with SV5, whereas amino acids 268 to 293 share greatest homology with Sendai virus (40%). The evolutionary changes observed
Table 1. Comparison of paramyxovirus F protein amino acid sequences*

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>SEN</th>
<th>PF3</th>
<th>SV5</th>
<th>NDV</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>EIA</td>
<td>LMK</td>
<td>KTV</td>
<td>NVS</td>
<td>IFS</td>
</tr>
<tr>
<td>150</td>
<td>OVN</td>
<td>EQR</td>
<td>185</td>
<td>276</td>
<td>196</td>
</tr>
<tr>
<td>200</td>
<td>LGQ</td>
<td>LNL</td>
<td>LST</td>
<td>IGG</td>
<td>LAY</td>
</tr>
<tr>
<td>250</td>
<td>IGG</td>
<td>LNS</td>
<td>LVG</td>
<td>SMQ</td>
<td>LGG</td>
</tr>
<tr>
<td>300</td>
<td>1QD</td>
<td>LOK</td>
<td>202</td>
<td>310</td>
<td>211</td>
</tr>
<tr>
<td>350</td>
<td>155</td>
<td>213</td>
<td>105</td>
<td>227</td>
<td>116</td>
</tr>
<tr>
<td>400</td>
<td>1QD</td>
<td>213</td>
<td>105</td>
<td>227</td>
<td>116</td>
</tr>
</tbody>
</table>

* The measles virus (MV) F sequence is compared to those of Sendai virus (SEN; Blumberg et al., 1985b), parainfluenza 3 virus (PF3; Spriggs et al., 1986), SV5 (Paterson et al., 1984) and Newcastle disease virus (NDV; McGinnnes & Morrison, 1986). The comparison starts at amino acid 4 of the MV sequence. The solid arrows mark the conserved cysteines, the dotted arrow the point of cleavage of F0 to F1,2. Identical amino acids are indicated by asterisks.

Table 2. Comparison of the homologies of the F protein of measles virus with other paramyxoviruses*

<table>
<thead>
<tr>
<th></th>
<th>Overall homology (%)</th>
<th>Residue no.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>112–140</td>
<td>159–221</td>
</tr>
<tr>
<td>SV5</td>
<td>25:3</td>
<td>76</td>
</tr>
<tr>
<td>SEN</td>
<td>26:2</td>
<td>78</td>
</tr>
<tr>
<td>PF3</td>
<td>24:2</td>
<td>64</td>
</tr>
<tr>
<td>NDV</td>
<td>24:0</td>
<td>64</td>
</tr>
</tbody>
</table>

* The alignment of the amino acid sequences and abbreviations are as in Table 1. Homologies are expressed as percentages.

within the paramyxoviruses reflect the different selection pressures. Until the structure and function of these regions can be defined, it is difficult to attach importance to such homologies.

Comparison of the different F proteins has shown that a number of constraints are imposed on the protein structure. This is provided by eight well conserved cysteines between positions 337 and 423, and a cysteine at position 71 in the F2 polypeptide which may possibly be bonded to a cysteine at position 198 in the F1 chain. The grouped cysteine residues would give a bunched structure. Studies on the cleavage of F0 to F1,2 in Sendai virus have shown that there is a change in the protein leading to an increase in the $\alpha$-helical configuration (Hsu et al., 1981). The proposed cleavage site in the measles virus F protein would also release an $\alpha$-helical region on the F1 polypeptide. After cleavage, the polypeptide binds more non-ionic detergent and so it has
Measles virus fusion protein sequence

been proposed that cleavage revolvs the bunched region which could act as a hinge. This would expose the hydrophobic region for possible interaction with the cell membrane. Thus, any variation which occurs in the F protein must respect the restraints imposed by the cysteine residues.

The least homology between the F proteins of the paramyxoviruses is at the two ends. Thus, the signal peptide, the transmembrane region and the cytoplasm side residues show the greatest variation.

General considerations

The F protein is highly conserved in the paramyxovirus family. Although it plays an important role in immunization, it is not the major antigen that induces neutralizing antibodies (Norrby, 1985). Studies have shown that the cell fusion activity is located at the N-terminal region of the F1 polypeptide and it is probably by this mechanism that the virus penetrates the cell. However, the manner in which antibodies can inhibit virus penetration and play a role in immunity is unknown. Thus, it will be interesting to determine whether the fusion-inhibiting antibodies react directly with the hydrophobic region or by interacting with other regions and inhibit by steric hindrance. The possession of a long 5' non-coding sequence with a high cytosine content in measles virus F mRNA is unique in the paramyxoviruses. As mentioned above, similar sequences on cellular mRNAs can regulate translation. Studies are underway to investigate whether this region has a regulatory role in translation of the measles virus F protein.

We would like to thank Drs P. C. Dowling and B. M. Blumberg for supplying the measles virus F probe, David Meyer, Steven Fuller and Patrick Argos for most helpful discussions and Roy Omond for patient advice with the computing. C.G. is a recipient of a fellowship from the Cancer Research Institute, New York.

REFERENCES


Measles virus fusion protein sequence


(Received 5 January 1987)