Genome organisation and genetic relationships among the morbilliviruses

T. BARRETT and B.W.J. MAHY*

Summary: Measles virus (MV) and canine distemper virus (CDV) cDNA clones have been used to study the morbillivirus group of the Paramyxoviridae. Strong nucleotide homology has been shown to exist between the N genes of MV and CDV and between the P genes of all members of the group. Each member of the group induces similar mRNA species on infection in Vero cells and their order on the genome can be deduced by analysis of the bicistronic mRNAs produced by readthrough of the mRNAs. Recently, the coding function of the mRNAs has been demonstrated and the gene order shown to be 3' N-P/C-M-F-H-L-5', similar to that found in other paramyxoviruses.

KEYWORDS: Antigen structure - Distemper virus - Genes - Measles virus - Morbillivirus - Nucleotides - Paramyxoviridae.

INTRODUCTION

Measles virus (MV), canine distemper virus (CDV), rinderpest virus (RV) and peste des petits ruminants virus (PPRV) are an antigenically closely related group within the family Paramyxoviridae (16, 17, 22). They are known as morbillviruses after the characteristics spots seen in measles infections (morbillus = pustule). They are important and dangerous pathogens in their host species. In addition to the initial acute infection, these viruses are associated with chronic neurological diseases in rare cases. A chronic degenerative disease of the brain (subacute sclerosing panencephalitis or SSPE) is associated with persistent infection with MV in humans. A similar disease is associated with CDV in dogs. Although cross-species infection has been demonstrated, no serious disease occurs in the non-natural host (11). Epidemiological and molecular evidence has been used to suggest an aetiological role for CDV in some chronic human diseases such as multiple sclerosis (10, 13, 14) and, more recently, Paget's disease (21). Safe and effective vaccines are available for all these diseases but, owing to lack of coordinated and sustained vaccination campaigns, even measles virus remains a problem in developed countries (19). Rinderpest virus was almost eradicated from the African subcontinent in the 1960's but has reappeared, with devastating results to the economies of many African and Asian countries (26).

Until very recently nothing was known about the detailed genetic structure of these viruses. They were known to be typical paramyxoviruses with a pleiomorphic enve-
loped structure containing a non-segmented negative-strand genome RNA. The virus was known to encode six structural proteins and possibly one non-structural protein. The virus structural proteins consist of a nucleocapsid protein (NP) which surrounds the genome RNA, a large polymerase protein (L) and a smaller polymerase-associated protein (P), a matrix protein (M) associated with the virus envelope and two envelope glycoproteins, the haemagglutinin (H) and fusion (F) proteins. The viruses all show similar protein patterns in infected cells but variation in mobility is seen both within one virus group and between the virus groups (25). Because of the difficulty in getting these viruses to grow well in tissue culture cells, little was known until fairly recently about their detailed biochemistry and molecular biology. Advances in cDNA cloning techniques have enabled double-stranded DNA copies of most of the virus mRNAs of MV and CDV to be produced (4, 6, 7, 27, 28, 29, 30). These cDNA clones are now being used to study the genetic organisation and structure of these viruses. In this paper we summarise our knowledge to date of this important virus group and indicate relationships at the molecular level.

MESSENGER RNAs INDUCED IN INFECTED CELLS

When virus-infected cells are labelled with $^{32}$P-orthophosphate in the presence of high doses of actinomycin D to inhibit labelling of cell mRNA, ten virus-induced polyadenylated RNA species are detected. These RNAs do not separate well on polyacrylamide gels or on non-denaturing agarose gels, but they can be well separated on agarose gels containing either formaldehyde or methyl mercury (see Fig. 1). All the morbilliviruses induce similar mRNA patterns with almost identical mobilities on denaturing agarose gels. Only mRNA 5 of CDV runs slightly faster than the corresponding mRNA band of the other viruses. Six of these virus-induced mRNAs are required to code for the six virus structural proteins and the coding function of most of these is known for MV and CDV. By analogy, it is assumed that mRNAs of similar size will code for the corresponding proteins in RV and PPRV. RNA band 1 is known to code for the matrix protein of MV and CDV (27, 29, 30). The heavily labelled band 2/3 is known to contain two separate mRNAs, one coding for the NP protein and the other for the P protein in both MV and CDV (29, 30). Band 4 is known to code for the H protein of MV (1, 9). This leaves band 5 as the probable mRNA for the F protein and, because of the large size of the L protein, band 10 is assumed to code for the L protein. RNA bands 6-9 are known to contain genetic information from more than one virus gene and are thought to be readthrough products formed by the failure of the polymerase to detach, or to stop and restart, at the intergenic region, which is thought to control the synthesis of discrete mRNA species for each protein. Such bicistronic mRNA species have been described in other paramyxoviruses (3, 8, 23) and it is not known what role, if any, these bicistronic mRNA species play in virus replication. They may be fortuitous events which are totally irrelevant to virus replication. There is evidence, however, that a change in such a control region leading to the suppression of the production of one class of monocistronic mRNA may be one of the mechanisms involved in the persistence of MV in one case of SSPE (2).

Sequence data are available for a number of mRNAs of MV and CDV (see Table I). These data have enabled the mRNA for the non-structural protein to be identified. As is the case with Sendai virus and SV5, two other paramyxoviruses, the non-structural protein of both MV and CDV is encoded by the same mRNA as the P protein in an overlapping reading frame beginning at an AUG 19 base downstream of the first AUG which encodes the initiator methionine of the P protein (5, 6). By
Agarose-formaldehyde gel electrophoresis of [³²P]-labelled mRNAs synthesised in virus-infected Vero cells

RNA was labelled in the presence of 10 µg/ml Actinomycin-D to eliminate host cell background. The RNA was then purified by phenol-chloroform extraction and the polyadenylated RNA separated by oligo [dT]-cellulose chromatography. The virus infection from which the RNA was derived is indicated above each track. The virus-induced mRNAs are labelled 1-10.
### TABLE I

Sequence data of mRNA of measles virus (MV) and canine distemper virus (CDV)

<table>
<thead>
<tr>
<th>Virus gene</th>
<th>mRNA size (approx.)</th>
<th>Nucleotides Number sequenced</th>
<th>Nucleotides Coding capacity</th>
<th>Molecular weight (actual)</th>
<th>Molecular weight (PAGE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleocapsid protein (N)</td>
<td>MV 1850 CDV 1850</td>
<td>MV 1631 CDV 1597</td>
<td>MV 523 CDV ?#</td>
<td>MV 58.1K CDV ?</td>
<td>MV 60K CDV 58K</td>
</tr>
<tr>
<td>P protein</td>
<td>MV 1850 CDV 1850</td>
<td>MV 1657 CDV 1633</td>
<td>MV 507 CDV 507</td>
<td>MV 53.9K CDV 54.9K</td>
<td>MV 70K CDV 73K</td>
</tr>
<tr>
<td>Non-structural protein (C)</td>
<td>MV 1850 CDV 1850</td>
<td>MV 558 CDV 522</td>
<td>MV 186 CDV 174</td>
<td>MV 21.0K CDV 20.3K</td>
<td>MV 20K CDV 20K</td>
</tr>
<tr>
<td>Matrix protein (M)</td>
<td>MV 1500 CDV 1500</td>
<td>-</td>
<td>MV ? CDV ?</td>
<td>MV ? CDV ?</td>
<td>MV 36K-37K CDV 34K</td>
</tr>
</tbody>
</table>

† Data extracted from references 1, 4, 5, 6, 25, 28 and 29.
# Sequence data do not extend to the end of the coding region.
* Molecular weight of unglycosylated precursor.
analogy with the non-structural protein of Sendai virus, it is known as the C protein. Confirmation that this was the mRNA for the C protein was obtained by preparing antibodies to synthetic peptides predicted to be encoded by this part of the RNA and using them to immunoprecipitate the virus-induced non-structural protein (6). No other non-structural proteins are predicted from the sequence data so far obtained.

**GENE ORDER ON THE VIRION RNA**

The 3' end sequences of MV have been cloned and from these results it is known that there is a 56 base leader sequence before the start of the NP gene (7). Intergenic regions of parts of the MV genome have also been sequenced and these, in addition to confirming the gene order deduced by other means, have shown that the intergenic region between all virus genes sequenced is a trinucleotide sequence GAA. This sequence is preceded by a U-rich sequence (AAUAUUUUU) in the untranslated 3' end of the mRNA, which probably acts as a polyadenylation signal (7, 24). This U-rich sequence is not found at the end of the leader RNA and so it is predicted that any leader RNA produced in infected cells would not be polyadenylated. Each gene begins with the sequence UCCU. These sequences, along with the intergenic GAA sequence and the 5' sequences of the next mRNA along the genome, may be responsible for controlling the stop-start mechanism of virus mRNA transcription. The bicistronic mRNAs which contain sequences from two adjacent mRNAs on the genome have enabled the linear gene order of both MV and CDV to be deduced. The gene order is identical to that found in Sendai virus and is 3' NP-P/C-M-F-H-L-5' (31). The order of the first three genes of MV has been confirmed using synthetic peptides produced to sequences in known regions of the virus genome and relating them to known virus proteins (9, 24, 30). Fig. 2 summarises these results.

---

**UUAU(A 6)CUUAGGA**

**UGCAGCUUAGGA**

**MV Genome RNA**

**FIG. 2**

**Schematic representation of the linear order of genes on the virion RNA of MV**

The nucleotide sequences around the gene junctions shown are given in the positive (mRNA) sense. The intergenic trinucleotide GAA is not transcribed in the process of mRNA synthesis and its complementary sequence (CUU) is underlined. The exact numbers given under the N, P/C and H genes indicate the number of nucleotides sequenced for each of these genes. The round figures under the other genes indicate the approximate size of their mRNA. The 3' leader sequence is represented as a box.
GENETIC HOMOLOGY WITHIN THE MORBILLIVIRUS GROUP

Double-stranded cDNA copies of all the structural genes, except the L gene, of MV and CDV have been used to study the homology between all members of the morbillivirus group. DNA specific for each virus gene was bound to nitrocellulose discs and used to select labelled mRNA synthesised in vivo in cells infected with different virus species. As expected, DNAs from the homologous virus could select the corresponding labelled mRNA species from the total pool of labelled RNA. In each case only the P gene could select mRNA efficiently from the heterologous virus (Fig. 3). These selections were carried out under conditions of high stringency, i.e. 50°C in 0.1 x SSC. This showed that a high degree of homology exists between the P gene of all members of the group.

Under conditions of low stringency it is possible to show cross-hybridization between the NP genes of MV and CDV (28). The P gene clone used in this study was full-length. Other clones represent at least 50% of the sequences from the 3' end of the mRNA. In the case of the NP gene, homology between the NP genes of MV and CDV is lowest at the 3' end of the message (28). No cross-hybridization could be shown between the other structural genes studied (M, H, F genes), although in none of these cases were full-length genes available for this study. Sequence data are available for the NP and P genes of MV and CDV (5, 6, 7, 28).

Fig. 4 shows a computer analysis of the NP and P genes to demonstrate the regions of high and low homology at the nucleotide level. As shown by Rozenblatt et al. (28), the NP gene shows three distinct regions of homology when MV and CDV are compared. The first 500 nucleotides show a moderate degree of homology (59%), the middle region shows a high degree of homology (77%) and the 3' end shows little homology (33%). In contrast, the P genes of these viruses show an evenly distributed degree of homology with approximately 55% of the nucleotides in similar positions all along the molecule. There is no greater degree of conservation in the region which contains the overlapping C gene sequences than in the region which contains only the P gene sequences (5).

In each part of the molecule, regions of high and low homology are interspersed and this is reflected in the protein sequence which shows similar regions of high and low homology.

DISCUSSION

Molecular studies have confirmed the close relationships that were known to exist, based on structural and immunological studies, among the members of the morbillivirus group. Similar protein patterns are reflected in similar mRNA patterns and sequence studies suggest that the viruses are derived from a common ancestor. More detailed sequence analysis of the genomes of all the morbilliviruses will be needed before their exact evolutionary relationships can be deduced. Based on a survey of the viruses using a battery of monoclonal antibodies, Norrby et al. (20) have suggested that rinderpest virus is the common ancestor of the group (see article by McCullough et al. in this volume). It will be interesting to see whether this finding can be confirmed by a more detailed genetic analysis of the viruses.
FIG. 3
The use of cDNA clones of CDV genes to select $^{32}$P-labelled mRNAs from homologous and heterologous virus infections

Plasmid DNAs containing virus-specific gene cDNA copies were bound to nitrocellulose discs and used to select complementary mRNAs by hybridisation under suitable conditions. The specifically bound RNA was eluted in distilled water and separated by agarose-formaldehyde gel electrophoresis. The virus infection from which the $^{32}$P-labelled RNA was derived is indicated below each panel. pAT, vector plasmid pAT153 without a virus-specific insert; M, M gene cDNA; N, N gene cDNA clone; P, P/C gene cDNA clone; mRNA 4, H gene cDNA clone; mRNA 5, F gene cDNA clone; rRNA, 18S and 28S ribosomal RNA marker.

The over-exposed tracks contained total unselected mRNA.
FIG. 4
Diagonal plots showing homology at different stringencies between the NP and P genes of MV and CDV
Computer analysis was carried out using the Staden programmes. The odd span length was set at 101. The 5' end (message sense) is at the top left-hand corner in each panel. A, B, C: comparison of MV and CDV NP genes at 50%, 60% and 70% homology, respectively. D, E, F: comparison of the P/C genes at 50%, 60% and 70% homology, respectively.
It is not surprising that the greatest degree of homology should be seen with the polymerase-associated protein. It is probably the need to preserve important functional regions of the molecule for interaction with the other component of the polymerase complex, i.e. the L protein, in addition to the need to retain the coding capacity of the C protein unchanged, that has ensured that a high degree of sequence conservation is necessary for this gene to survive. Other genes may retain smaller regions of structural and functional significance which are preserved but which may or may not be preserved at the nucleotide sequence level. Strong cross-reacting antigenic sites do exist for the other virus genes and it is likely that these regions represent such structurally and functionally important regions.

The gene order of the morbilliviruses is shared with the rhabdoviruses and some, but not all, other members of the paramyxovirus group. The order of genes is identical to that found in Sendai virus and Newcastle disease virus. A similar gene order is also seen with SV5 virus (23) but, in the latter case, a small, possibly membrane-bound protein gene is inserted between the F and HN genes. This is in addition to the non-structural V gene encoded on the P mRNA. Another paramyxovirus, respiratory syncytial virus, shows a totally different arrangement of genes with two non-structural protein genes preceding the nucleocapsid protein gene on the virion RNA (8).

Leader RNA sequences at the 3' end of the virion RNA are a feature shared with other paramyxoviruses and with rhabdoviruses (18), as are short intergenic regions. It is probable that a similar replication strategy with individual mRNAs synthesised either in a stop-start fashion or by cleavage of a readthrough product is also shared among these viruses. Bicistronic mRNAs have been described in all the paramyxoviruses so far studied, and also in the rhabdoviruses. In the case of the rhabdoviruses, long poly-A tracts were found between the genes in bicistronic mRNAs (15), whereas no such A-rich regions were found between the genes in bicistronic mRNAs of Sendai virus (12). This may reflect a different mechanism of virus transcription in these viruses. It would appear that synthesis of bicistronic mRNAs is achieved by failure of a cleavage mechanism after the addition of the initial poly-A sequences. In the paramyxoviruses, on the other hand, polyadenylation may occur on the newly synthesised message after cleavage as the polymerase moves to the next gene on the template. More detailed molecular studies on these viruses may shed light on their wider relationships and may help to identify their origin.

* * *

ORGANISATION DU GÉNOME ET RELATIONS GÉNÉTIQUES PARMI LES MORBILLIVIRUS. — T. Barrett et B.W.J. Mahy.

Résumé: Les clones codant l'ADN des virus de la rougeole et de la maladie de Carré ont été utilisés pour étudier le groupe des morbillivirus dans la famille des Paramyxoviridae. Il existe une forte homologie des nucléotides entre les gènes N des virus de la rougeole et de la maladie de Carré, et entre les gènes P de tous les membres du groupe. Chaque membre induit des espèces semblables d'ARN messagers (ARNm) lors d'infection sur cellules Vero, et leur position sur le génome peut être déduite par l'analyse des ARNm bicistroniques produits par leur transcription. Récemment, la fonction codante des ARNm a été mise en évidence et il a été constaté que l'ordre des gènes était 3' N-P/C-M-F-H-L-5', comme celui observé dans d'autres paramyxovirus.
ORGANIZACIÓN DEL GENOMA Y RELACIONES GENÉTICAS ENTRE LOS MORBILLIVIRUS. — T. Barrett y B.W.J. Mahy.

Resumen: Se utilizaron los clones que codifican el ADN de los virus del sarampión y de la enfermedad de Carré para estudiar el grupo de morbillivirus en la familia de los Paramyxoviridae. Existe gran homología de los nucleótidos entre los genes N de los virus del sarampión y de la enfermedad de Carré y entre los genes P de todos los miembros del grupo. Cada miembro induce especies semejantes de ARN mensajeros (ARNm) durante la infección en células Vero, pudiendo deducirse su posición en el genoma analizando los ARNm bicistónicos producidos por su transcripción. Recientemente, se evidenció la función codificadora de los ARNm, comprobándose que el orden de los genes era 3’ N-P/C-M-F-H-L-5’, como el observado en otros paramixovirus.

PALABRAS CLAVE: Estructura del antígeno - Genes - Morbillivirus - Nucleótidos - Paramyxoviridae - Virus de la enfermedad de Carré - Virus del sarampión.

REFERENCES