Nucleotide Sequence of the Gene Encoding the Matrix Protein of a Recent Measles Virus Isolate

By MARTIN D. CURRAN AND BERT K. RIMA*
Department of Biology, The Queen's University of Belfast, Belfast BT9 7BL, U.K.

(Accepted 13 May 1988)

SUMMARY
The sequence of the M gene of the Hu2 strain of measles virus has been determined. In the coding region of the gene, six nucleotide replacements had occurred with respect to the sequence of the M gene of the Edmonston strain. Two of these were silent and only four led to amino acid replacements in the protein sequence even though in SDS-PAGE the mobility of the protein would indicate an Mr increase of 2000. No changes were observed in the 5' untranslated regions of the gene. Two changes were observed in the 3' untranslated region of the gene. The sequences are compared to those recently published for subacute sclerosing panencephalitis viruses.

Measles virus (MV) causes a comparatively harmless childhood disease which on rare occasions is complicated by encephalitis and in even rarer cases by a slow persistent virus infection called subacute sclerosing panencephalitis (SSPE). It has long been of interest whether specific MV strains are involved in the SSPE cases and a large number of comparative studies between MV and SSPE-derived viruses have been carried out. Although all MV isolates are serologically monotypic, strain differences have been observed in the mobilities of viral proteins in SDS-PAGE (Wechsler & Fields, 1978; Campbell et al., 1980) and proteolytic digest fingerprints (Hall et al., 1979; Rima et al., 1983). The ability of certain monoclonal antibodies to bind to MV isolates also varies from strain to strain (Sheshberadaran et al., 1983). Because of the importance of the matrix (M) protein of paramyxoviruses in the maturation of the virus, particular attention has been paid to the comparison of M proteins of various strains in attempts to explain the defectiveness of SSPE viruses. Differences in the mobility of M protein of SSPE-derived virus strains have been observed, and although these have been suggested to be distinct markers of such strains, later more extensive comparisons have indicated this not to be valid. However, it has been observed that a number of MV strains that have a low passage history in tissue culture do display a decreased mobility of the M protein (Rima et al., 1979, 1983; T. F. Wild, unpublished observations).

Recently, the nucleotide sequences of the M genes of the Onderstepoort strain of canine distemper virus (CDV) (Bellini et al., 1986) and the laboratory-adapted Edmonston strain of MV (Bellini et al., 1986; Wong et al., 1987) have been published. Furthermore, the sequence of the M coding region of the virus from one SSPE case and the entire M gene of another have been reported (Cattaneo et al., 1986, 1988). In this paper we report the sequence of the M gene of the Hu2 strain of MV. This strain has undergone only a small number of tissue culture passages, and the mobility of its M protein in SDS-PAGE is retarded as is the case in many, but not all, recent isolates of the virus. This strain was isolated from a child with dysgammaglobinaemia in Belfast in 1971 (Mawhinny et al., 1971).

cDNA clones pMV1096 and pMV485, the former encompassing the entire coding region of the M gene, were generated by reverse transcription of MV genomic RNA by random priming with a mixture of hexanucleotides (Rima et al., 1986). cDNA clones pMV250 and pMV608,
overlapping the above clones, were generated from oligo(dT)-primed MV-infected Vero cell mRNA copied into cDNA by the method of Gubler & Hoffman (1983). Collectively these clones provided the DNA restriction fragments for subcloning into the M13tg130 and M13tg131 vectors, and subsequent sequencing by the dideoxy chain termination method. Fig. 1 shows a restriction map of the M gene of the MV Hu2 strain and the sequencing strategy used in this project. The four clones obtained when taken together comprise two independent copies of the entire sequence of the M gene. Both copies covering the coding region of the gene were sequenced completely and in both directions; no differences were found between them. The 3' non-coding region was also sequenced from two clones but in one direction only. The sequence of the gene is given in Fig. 2. In the coding sequence only six nucleotide differences were observed between the Hu2 sequence and the earlier published sequence of the Edmonston strain at positions 62 (G to A), 222 (C to T), 294 (A to G), 457 (G to A), 797 (C to G) and 889 (A to G). Those at positions 62 and 797 are silent. The others led to the changes in the amino acid sequence in the M protein detailed in Fig. 3, discussed below. In the 5' untranslated region of the mRNA, no differences were found and in the long 3' untranslated region only one sequence difference from that published by Wong and coworkers (1987) was observed. At position 1417 we found in Hu2 a single G residue as opposed to two Gs in the Edmonston sequence (Wong et al., 1987). This change would lead to the potential protein products of the two open reading frames (ORFs) in this region being more similar in size to those suggested by Bellini et al. (1986) than to those suggested by Wong et al. (1987).

A number of interesting points emerge from the comparison of the Hu2 M gene sequence with those of CDV, the two sequences of the Edmonston strain of MV and the two SSPE sequences. First, amino acid replacements take place in Hu2 only at positions that are also variable between CDV and MV. The latter two are identical at 76.5% or have conservative replacements at 7.5% of the positions (Bellini et al., 1986). Furthermore, the amino acid replacements at positions 88 and 142, K to E and S to N respectively, bring the Hu2 sequence in line with that of CDV. In the sequences of the M proteins of Sendai virus, human and bovine parainfluenza type 3 viruses and in Newcastle disease virus, there is a glutamate residue at this position as compared to the lysine residue in the Edmonston strain sequences. Previously it has been shown that the Edmonston M protein is different from that of most other MV and SSPE virus strains (Hall et al., 1979; Rima et al., 1983). The E (88) and N (142) residues are probably the normal residues at these positions as they are also present in the two SSPE sequences (Cattaneo et al., 1986, 1988).

It is difficult to establish which mutations are responsible for the major mobility changes that occur in the M proteins of the morbilliviruses. Although the calculated Mr values of the M proteins deduced from sequence data are nearly the same for the Edmonston and Hu2 strains and for CDV, the M protein of Hu2 migrates with an apparent Mr 2000 higher than that of the Edmonston strain, whereas the CDV M protein has an Mr approximately 3000 lower than that of the Edmonston strain of MV. The apparent Mr change in the Hu2 protein is clearly not related to an extension in the N- or C-terminal part of the protein, since both the start and stop codons for translation of the M mRNA are unchanged. Both the change at position 64 (P to S) and at
position 287 (D to G) involve the removal or introduction of helix-breaking residues, although none of the amino acid replacements alters the secondary protein structure predicted by the algorithm of Garnier et al. (Microgenie, Beckman). However, it is difficult to predict how any change will affect the mobility of a protein in SDS-PAGE, since a change from V to G or A to T in the ras oncogene protein increases its apparent Mr by 2000 (McGrath et al., 1984). The mobility of the IP-3CA M protein (Cattaneo et al., 1988) appears to be similar to that of the Hu2 strain, so that the replacements of E (88) and N (142) may be involved in the mobility change.

One further interesting point in relation to the sequence of the Hu2 M gene is the apparent immutability of the 3' untranslated region of the gene. In MV but not in CDV the large untranslated part of the mRNA contains two ORFs one of which continues right up to the start of translation in the F gene and encodes a possible product of Mr 36115 (Wong et al., 1987), although smaller products of 13.3K and 14-2K are predicted from the data presented here and those of Bellini et al. (1986). There is at present no evidence for translation of these ORFs (Bellini et al., 1986) but it is interesting to note that there must clearly be some selective pressure to maintain the same nucleotide sequence in the two lytically growing viruses, the Edmonston and Hu2 strains. This may suggest a functional role for this sequence. In the case of the SSPE-derived strain IP-3CA, the M gene shows 23 nucleotide changes in the 3' untranslated region, whereas in the second sequence published by Cattaneo et al. (1988) in 52 nucleotides in the 3' untranslated region no differences were found.
### Short communication

**Fig 3** Comparison of the deduced primary protein sequences of the M proteins of morbilliviruses. **MEdm**: sequence of the Edmonston strain of MV (Bellini et al., 1986). **MED**: sequence of the Edmonston strain of MV (Wong et al., 1987). **MHu2**: this paper. **SK**: sequence of an SSPE virus (Cattaneo et al., 1986). **Sip3**: sequence of an SSPE virus (Cattaneo et al., 1988). **ConD**: sequence of the Onderstepoort strain of canine distemper virus (Bellini et al., 1986). Asterisk indicates a premature stop codon in the SK sequence.

In conclusion, the sequence of the M gene of the Hu2 strain of MV has revealed three interesting observations. First, it appears that a number of sequence changes between the Edmonston strain and Hu2 sequence are also present in the SSPE-derived sequences, indicating that the Hu2 strain sequence probably provides a better reference standard for recent measles virus isolates than the published Edmonston sequence. Secondly, the relatively low level of variability between the Edmonston strain and the Hu2 strain in this gene (0.5% Yo) is much lower than the recently described variation with respect to SSPE strains (3 Yo), suggesting the latter to be significant. Thirdly, the 3' untranslated area of the gene is highly conserved between the two lytically growing strains, indicating that some form of functional constraint probably operates on the sequence.

Support for this work provided under grant number G8406017CA by the U.K. Medical Research Council is gratefully acknowledged.

**REFERENCES**


(Received 7 March 1988)