Characterization of Mapuera virus: structure, proteins and nucleotide sequence of the gene encoding the nucleocapsid protein

G. W. Henderson, C. Laird, E. Dermott and B. K. Rima

Division of Molecular Biology, School of Biology and Biochemistry and Department of Microbiology and Immunobiology, The Queen's University of Belfast, Belfast BT9 7BL, UK

The molecular biology of Mapuera virus was studied at both the protein and nucleic acid levels. Seven virus-encoded proteins were detected in infected Vero cells. The sizes and characteristics of each of the proteins determined from various radiolabelling experiments allowed preliminary identification of the proteins as the large (L; 190 kDa), haemagglutinin neuraminidase (HN; 74 kDa), nucleocapsid (N; 66 kDa), fusion (F; 63 kDa), phosphoprotein (P; 49 kDa), matrix (M; 43 kDa) and non-structural (V; 35 kDa) proteins. Western blot analysis showed that the HN, N and P proteins were major antigens recognized in the mouse. A cDNA library of total virus-infected cellular mRNA was created and screening of the library resulted in the detection of cDNA sequences representing the N mRNA transcript of Mapuera virus. The N mRNA sequence determined from the clones was 1731 nt in length and contained an ORF that encoded 537 amino acids, the complete 3' untranslated region and part of the 5' non-coding region. The calculated M of the N protein was 59 kDa, which is close to the 66 kDa protein observed by SDS-PAGE.

Introduction

Mapuera virus was isolated from the salivary glands of a fruit bat (Sturnira lilium) captured in the tropical rainforest of Brazil in 1979. When captured the bat showed no visible signs or symptoms of illness. The virus was registered in the International Catalogue of Arboviruses in 1982 (Karabatsos, 1985) where it was unclassified but noted as probably not an arbovirus. A preliminary characterization of Mapuera virus by Zeller et al. (1989) using electron microscopy suggested that the virus belonged to the Paramyxoviridae family, on the basis of its morphology and its ability to haemagglutinate guinea-pig erythrocytes.

With the exception of the Duvenhage rabies virus and the Tacaribe arenavirus, relatively little is known about viruses which naturally infect bats. The social nature of bats and the fact that they roost at high population densities in the confined area of caves should make them susceptible to many viruses, especially those that are transmitted by a respiratory route. There has been only one previous report of the isolation of a paramyxovirus from a bat (Hollinger & Pavri, 1971). This virus was isolated from another species of fruit bat (Rousettus leschenaulti) in India and was identified as a new animal subtype of parainfluenza virus (PIV) type 2.

The aim of this investigation was to characterize Mapuera virus more fully as regards its structure and the proteins induced in infected cells, and thereby determine the placing of the virus in the family Paramyxoviridae. The types of protein and M's were determined by in vivo radiolabelling methods. The nucleotide sequence of the gene encoding the nucleocapsid (N) protein was also determined. The results confirm the initial classification of the virus within the family Paramyxoviridae and indicate that Mapuera virus should be placed within the genus Rubulavirus, which includes mumps virus and simian virus 5 (SV5) (Rima et al., 1995a).

Methods

Cells and virus. Vero cells were obtained from Flow Laboratories and grown in Eagle's MEM (Glasgow modification) supplemented with 8% newborn calf serum and antibiotics, as described previously (Rima et al., 1980). Mapuera virus strain BeAnn 370284 was obtained as a Vero cell adapted stock from Professor C. R. Pringle (Warwick University, UK) and after plaque purification, the virus was used to infect confluent Vero cell monolayers maintained in Eagle's MEM supplemented with 2% newborn calf serum.

Electron microscopy. When CPE was extensive throughout a monolayer of infected Vero cells, the medium was removed and the
flask was frozen at \(-20\) °C. After thawing, several droplets of sterile distilled water were dropped onto a specific area of the thawed cell layer and left for 1–2 min to further lyse the cells. Several drops of this freeze fractured cell lysate were then spotted onto Parafilm and grids coated with a Formvar film were floated on these droplets for a few seconds. The grids were drained of excess fluid, stained with methylamine tungstate stain and examined using a Philips CM10 transmission electron microscope.

**Protein labelling.** Vero cells were grown to confluence in 50 mm diameter Petri dishes and infected with Mapuera virus at an m.o.i. of 0.3 as higher doses appeared to give rise to interference. At 18 h p.i., the cell monolayer was washed three times in pre-warmed PBS and 2–3 ml of warm (37 °C) methionine- or cysteine-free medium was added. The cells were incubated at 37 °C for 30 min, after which the cells were supplemented with medium containing either 50 \(\mu\)Ci/ml L-[\(\text{\textsuperscript{35}S}\)]methionine or 200 \(\mu\)Ci/ml L-[\(\text{\textsuperscript{35}S}\)]cysteine (specific activities of both > 1000 Ci/mmol; Amersham) and incubation continued for 1 h when labelling with methionine and 2 h when using cysteine. At the end of the labelling period, the medium was removed and the cell sheet was washed three times with ice-cold PBS. The cell monolayer was then scraped into 250 \(\mu\)l of cell lysis buffer [20 mm-Tris-HCl (pH 6.8), 6% (w/v) urea, 2% (w/v) SDS, 2% (v/v) 2-mercaptoethanol] and boiled for 3 min, after which it was either stored at \(-20\) °C or analysed immediately by SDS-PAGE (Laemmli, 1970).

For radiolabelling of cell lysates with \(\text{\textsuperscript{14}C}\)-labelled amino acids, infected Vero cells were starved of amino acids by incubation in 2 ml of amino acid-free medium at 37 °C for 30 min. After this period, the amino acid-free medium was supplemented with 5 \(\mu\)Ci/ml of \(\text{\textsuperscript{14}C}\)-labelled L-amino acid mixture (algal lysate mixture of 16 amino acids, specific activity > 50 mCi/mmol; Amersham) and incubated for a further 3 h at 37 °C.

For radiolabelling of cell lysates with \(\text{\textsuperscript{14}C}\)-H\(\text{\textsuperscript{3}}\)glucosamine hydrochloride, infected Vero cells were incubated at 37 °C in glucose-free medium for 30 min before supplementing the medium with 100 \(\mu\)Ci/ml \(\text{\textsuperscript{14}C}\)glucosamine hydrochloride (specific activity 20–40 Ci/mmol; Amersham) and incubating for a further 2 h.

For radiolabelling of cell lysate with \(\text{\textsuperscript{32}P}\)orthophosphate the infected cells were starved of phosphate by growth in 2 ml of phosphate-free medium (Gibco BRL) at 37 °C for 30 min. This medium was then replaced by a further 2 ml containing 200 \(\mu\)Ci/ml \(\text{\textsuperscript{32}P}\)orthophosphate (Amersham) and incubated for 3 h at 37 °C.

**cDNA production.** Total cellular RNA was extracted from Mapuera virus-infected Vero cells by the method of Chirgwin et al. (1979). cDNA was produced from the mRNA component of the RNA sample by the method of Gubler & Hoffman (1983) using the TimeSaver
cDNA synthesis kit (Pharmacia). EcoRI/NotI linkers were attached to the cDNA molecules which were then ligated into EcoRI restricted pBluescript (Stratagene). The plasmid was then used to transform highly competent *Escherichia coli* DH5αF’IQ cells.

For screening recombinant *E. coli*, cells were replica streaked onto two Hybond-N membranes (Amersham) and colony hybridizations were performed using 32P-labelled cDNA probes prepared by reverse transcription of poly(A)+ RNA from infected and mock-infected cells as described previously (Curran *et al.*, 1985).

**DNA sequencing.** Plasmid DNA was purified by the NaOH-SDS lysis method (Maniatis *et al.*, 1982). cDNA inserts were digested with appropriate restriction enzymes, the fragments were then subcloned into M13 phage vectors and sequenced in both directions by the dideoxy chain termination method using *Taq* DNA polymerase (Promega). Sequence analysis was done with the assistance of Beckman's Microgenie software.

**Results**

**Electron microscopy**

The published micrographs of Mapuera virus did not show enough detail about the structure of the virus to consider a preliminary classification. Hence, Mapuera virus-infected Vero cell cultures were examined for the presence of virus structures (Fig. 1). The Mapuera virions (Fig. 1 a, b) were found to be spherical, enveloped particles which ranged between 150–300 nm in diameter. Disrupted virions were shown to contain nucleocapsid structures (Fig. 1 a). Surface projections approximately 10 nm in length and approximately 7 nm apart were visible to a limited extent on the virions. Fig. 1 (c) shows an example of the nucleocapsid filaments which were observed in the infected cells. The 'herring-bone' morphology of the nucleocapsid filaments produced by the close association of the N protein with the genomic RNA, which is a characteristic of the *Paramyxoviridae* family, is clearly visible. From the measurement of eight nucleocapsid filaments the average length was found to be 1.1 μm with a diameter of 17.5 nm.

**Analysis of Mapuera virus-induced polypeptides**

Mapuera virus CPE on Vero cells formed only slowly and was not characterized by the extensive fusion observed with other paramyxoviruses. Individual 1 mm diameter plaques which could be enumerated formed at 5–6 days p.i. The highest titre obtained after plaque purification of the virus stocks was 4.0 × 10⁶. To determine the optimum time for the labelling of Mapuera virus proteins, a number of time course experiments were performed in which replicate monolayers of infected cells were labelled with [35S]methionine at various times post-infection. These experiments revealed that Mapuera virus proteins were synthesized rapidly after infection, reaching a maximum level at approximately 18 h p.i., and maintaining this level of production over a period of 62 h (Fig. 2). Based on these studies, further experiments to detect virus-induced proteins using different radio-labels were carried out at 18 h p.i.

Five virus-encoded polypeptides were readily observed in the infected cell lysates that were labelled by addition of [35S]methionine to cultures at 18 h p.i (Fig. 2). The viral proteins have apparent *M*ₚ₁ of 35 kDa, 43 kDa and 49 kDa and a heavy diffuse band in lanes 3–5 (Fig. 2) appeared to represent two virus-induced proteins with *M*ₚ₁ of 63 kDa and 66 kDa (Fig. 2, lanes 6–8). Labelling
polypeptides with $^{[35S]}$cysteine allowed a clearer visualization of two further virus-induced proteins with $M_s$ of 74 kDa and 190 kDa (Fig. 3), probably due to the reduced host background and the cysteine content of these proteins. The majority of the viral proteins were labelled to a lower intensity with $^{[35S]}$cysteine than with $^{[35S]}$methionine, with the exception of the 35 kDa protein which was much more strongly labelled and appeared to
represent a cysteine-rich protein. To investigate if any of the viral polypeptides which had been identified so far were phosphorylated, virus-infected Vero cells were exposed to $[\text{32}P]_{\text{orthophosphate}}$ (Fig. 3). Only one viral polypeptide, the 35 kDa protein, was found to be strongly phosphorylated. All seven already identified viral polypeptides were visible in the protein lysate with $^{14}\text{C}$-labelled amino acids (Fig. 3) although the 74 kDa and 190 kDa polypeptides appeared as faint bands. The profile produced after labelling with $[^3\text{H}]\text{glucosamine hydrochloride}$ revealed the presence of three virus-induced polypeptides (Fig. 3). These were a 63 kDa protein, the diffuse faint protein band of 74 kDa size and a band migrating slightly ahead of the 35 kDa protein. Two other bands (at 48 kDa and 42 kDa) present at enhanced levels in the infected cell in Fig. 3 were also present in mock-infected control cells.

Based on the similarity of the Mapuera virus protein profile with that of other recognized paramyxoviruses such as mumps virus and SV5 and on the results of the labelling experiments, the virus-induced proteins were each proposed to represent particular paramyxovirus proteins, as summarized in Table 1.

The 66 kDa (N) and 74 kDa haemagglutinin neuraminidase (HN) proteins detected in virus-infected cells were confirmed to be virus specific by the use of immunoprecipitation. Five monoclonal antibodies (MAbs) were produced by injection of virus into BALB/c mice by standard techniques. Virus inactivated by UV irradiation was used since live virus produced a fatal infection after approximately 25–30 days. The specificity of all MAbs appeared to be the same and due to the granular appearance of immunofluorescent foci in infected cells and the lack of surface staining (data not shown), they appeared to be directed against an internal antigen. All the MAbs precipitated the 66 kDa (N) protein which was the major protein in the infected cell. A polyclonal mouse serum obtained from the mice before splenic fusion precipitated this same 66 kDa (N) protein as well as the 49 kDa phosphoprotein (P) and 74 kDa (HN) protein (Fig. 4a). The 66 kDa (N) and the 49 kDa (P) proteins were also identified in Western blot analysis of the polyclonal serum (Fig. 4b).

**Figure 4. Immunoprecipitation and Western blot.** (a) Immunoprecipitation of a $[^3\text{S}]$methionine-labelled lysate of Mapuera virus-infected Vero cells. The immunoprecipitation products were analysed on a 10% polyacrylamide gel. Virus proteins are indicated by a dot. The designations are derived from Table 1. Lane 1, mock-infected cells (not precipitated); lane 2, virus-infected cells (not precipitated); lane 3, mock-infected cell lysate precipitated with MAb 68 (anti-N); lane 4, virus-infected cell lysate precipitated with MAb 68 (anti-N); lane 5, mock-infected cell lysate precipitated with polyclonal mouse serum; lane 6, virus-infected cell lysate precipitated with polyclonal mouse serum. (b) Western blot strips generated by transfer of virus-infected or mock-infected cell proteins separated on a 10% polyacrylamide gel allowed to react with polyclonal mouse serum. Strip 1, mock-infected cell protein; strip 2, virus-infected cell protein.

**Characterization of cDNA clones**

In order to classify further the position of Mapuera virus in the Paramyxoviridae family, reverse transcription and cloning of mRNA extracted from infected cells was carried out as described. This resulted in the isolation of 500 cDNA clones. As MAbs against other paramyxoviruses were unable to bind to Mapuera virus (data not shown) and since cDNA clones from viruses possibly related to Mapuera virus, such as mumps virus and human PIV type 2 (HPIV2), did not bind to nucleic acids extracted from Mapuera virus-infected cells, a plus/minus screening technique was used to determine which of the cDNA clones were virus specific. Colony hybridization with probes obtained by reverse transcription of virus-infected cell and mock-infected cell mRNA in the presence of $[\alpha-^3\text{P}]{\text{dCTP}}$ allowed the identification of potential virus-specific cDNA clones. These were further analysed by using their inserts as probes on Northern blots of infected and mock-infected cell RNA. Twelve clones appeared to contain Mapuera...
Fig. 5. For legend see opposite.
virus cDNA sequences. After isolation and purification, the plasmids of these clones were found to contain related cDNA inserts and further clones were obtained from re-screening the cDNA library made from infected cell mRNA. Nucleotide sequencing of the largest cDNA inserts resulted in the determination of the complete genome of Mapuera virus from re-screening the cDNA library made from infected cells. A comparison of the N ORF coding for the N protein of Mapuera virus and part of related cDNA inserts and further clones were obtained from re-screening the cDNA library made from infected cell mRNA. Nucleotide sequencing of the largest cDNA inserts resulted in the determination of the complete genome of Mapuera virus and part of the N non-coding region. A comparison of the N protein sequence of Mapuera virus with the N protein sequences of other paramyxoviruses belonging to the rubulavirus subgroup is shown in Fig. 5.

Discussion

The virus structures observed were similar to those seen in other paramyxovirus-infected cultures. The measurements of the viral nucleocapsid (approximately 1 μm long and 17.5 nm diameter) and virion particles (diameter approximately 250 nm) corresponded well with those of other paramyxoviruses (Kingsbury et al., 1974) and extended the findings of Zeller et al. (1989). These data indicate that the ribonucleoprotein and probably the genome of Mapuera virus are similar in length to those of other paramyxoviruses (Galinski & Wechsler, 1991).

Against a background of host protein synthesis, seven virus-induced proteins were detected (Table 1). Comparison of the Mapuera virus protein profile with the profiles of other paramyxoviruses revealed that the pattern of Mapuera virus proteins was more similar to the profile of mumps virus and SV5 than to that of Sendai virus or PIV type 3 (PIV3). In particular, the size of the proposed P protein was similar to the small P proteins encoded by members of the mumps/SV5 subgroup of viruses (Herrler & Comps, 1982; Rima et al., 1980; Lamb et al., 1976; Galinski & Wechsler, 1991; Lamb & Paterson, 1991). Labelling with [35S]methionine and [35S]cysteine allowed the identification of seven viral proteins which probably represent the large (L), HN, N, fusion (F₀), P, matrix (M) and non-structural (NS1 or V) proteins. Labelling polypeptides with a 14C-labelled amino acid mixture allowed and confirmed the detection of all seven viral proteins although the band representing the putative HN protein was faint due to its disperse nature. The intensity of the labelled proteins obtained with [35S]cysteine was reduced, with the exception of the 35 kDa protein. This was thus proposed to represent the Mapuera virus V protein, by analogy with other paramyxoviruses (Thomas et al., 1988; Vidal et al., 1990; Cattaneo et al., 1989).

The use of [32P]orthophosphate to detect phosphorylated proteins revealed that only the non-structural V protein was strongly phosphorylated and the P protein was very weakly phosphorylated, if at all. This is an unusual result as the P protein of all other paramyxoviruses has been shown to be the major phosphorylated protein of the virion with the N protein being phosphorylated to a lesser extent (Lamb & Choppin, 1977a, b; Rima et al., 1980; Hsu & Kingsbury, 1982). Hsu & Kingsbury (1982) mapped the majority of the phosphorylation sites on the synthesized P protein of Sendai virus to the N-terminal one-quarter of the protein by partial V8 protease digestion. By contrast, Vidal et al. (1988) reported the bulk of the phosphates to be contained within the second N-terminal one-quarter of the in vitro synthesized P protein of Sendai virus. Even accounting for the obvious discrepancy between the results, which could be due to differences in the source of the phosphoprotein, it is clear that the majority of the phosphorylation sites of Sendai virus P protein are in the N-terminal half of the protein. This would indicate that
the non-structural V protein should also be phosphorylated as it shares N-terminal sequence identity with the P proteins and differs only in the C-terminal sequence. Indeed, Gombart et al. (1992) reported that the non-structural V proteins of acute measles virus and subacute sclerosing panencephalitis virus strains were phosphorylated. The relative intensities of the P and V proteins with 3H-labelled amino acids indicates that if the P protein was phosphorylated to the same extent as V, it would have been detected in the 33P-labelled profiles.

Labelling proteins with [3H]glucosamine showed that three virus-encoded proteins appeared to be glycosylated, the proposed F0 and HN proteins and an apparent viral protein which migrated slightly ahead of the proposed V protein. There were no apparent bands representing the F1 and F2 subunits of the F1,2 glycoprotein. However, the absence of either band is probably caused by the fact that not enough time was available during the 2 h labelling period to allow for processing and transport of the fusion protein. Comparisons of the known amino acid sequences of a number of paramyxoviruses (Morrison, 1988) have shown the majority of potential glycosylation sites to be on the F1 and not the F2 subunit, in contrast with the morbilliviruses where the F2 subunit contains all the glycosylation sites. The apparent glycosylated viral polypeptide of Mr, 34–35 kDa is too small to represent the F1 subunit and too large for the F2 subunit, based on comparisons with the F proteins of established paramyxoviruses (Morrison, 1988). However, it has been reported previously that the HN glycoprotein of mumps virus can undergo proteolysis and proteins with Mr, s of 41 kDa and 32 kDa were resolved by SDS-PAGE (Waxham et al., 1986; Merz & Wolinsky, 1983). It is possible that the observed glycosylated protein of 34–35 kDa represents a breakdown product of the Mapuera virus HN protein.

The use of MAbs and polyclonal serum in immunoprecipitation studies confirmed that the proposed HN and N proteins were virus specific. The lack of reactivity of the HN protein with polyclonal serum in a Western blot indicates that as with other paramyxoviruses the majority of epitopes on the HN molecule are conformational.

The length of the N mRNA sequence of Mapuera virus determined from the various clones was 1731 nt. The 3' end of this sequence ended in a poly(A) tail and corresponded well with the polyadenylation signal (TTAAGAn) found in all other paramyxoviruses from the rubulavirus genus. The 3' UTR was 52 nt in length and was preceded by an ORF that encodes 537 amino acid residues and a stop codon at position 1678. This ORF started at position 64 with an AUG codon in a favourable context (ANNATGT) as defined by Kozak's rules (Kozak, 1986). The context was the same as for mumps virus. The comparison of this nucleotide sequence with the N gene sequence of mumps virus (Elango, 1989) indicated that there are probably approximately 23 nt missing from the 5' end of the Mapuera virus N mRNA sequence, which would make the gene approximately 1750 nt in length. This is in good agreement with the size of the major RNA band observed in infected cells radiolabelled in the presence of actinomycin D (results not shown). Since the actual 5' end of the mRNA sequence has not been determined, it is possible that there could be an AUG initiation codon before the first one determined from the available sequence. However, the position of the proposed AUG initiation codon fits well with the determined initiation sequences of the other paramyxoviruses belonging to the mumps/SV5 subgroup. The alignment of amino acid sequences determined for the N proteins of the numerous viruses belonging to the rubulavirus genus of paramyxoviruses (Fig. 5) shows that all available N protein sequences begin with the conserved MSSVL motif except for that of Newcastle disease virus (NDV), which has an F residue in place of the final L residue. The calculated Mr of the Mapuera virus N protein is 59 kDa, which is close to the 66 kDa protein observed by SDS-PAGE. It has been noted for other paramyxoviruses, such as mumps virus, that the N protein migrates as a larger sized protein than that predicted from its deduced amino acid sequence. This may be due to post-translational modification of the protein or an anomaly of the SDS-PAGE system (Elango, 1989).

A comparison of the Mapuera virus N protein sequence with that of the other paramyxoviruses revealed 55% identity with mumps virus (Elango et al., 1989), 54% with SV5 (Tsurudome et al., 1991), 51% with SV41 (Tsurudome et al., 1991) and HPIV2 (Yuasa et al., 1990), 50% with HPIV type 4a/b (Kondo et al., 1990) and only 35% with NDV (Ishida et al., 1986). Lower levels of identity were found when Mapuera virus N protein was compared to the measles virus N protein (28%) and Sendai virus N protein (21%). These results clearly indicate that on the basis of its N protein sequence Mapuera virus is a distinct virus and shows greater similarity to viruses of the rubulavirus subgroup than those of the Sendai virus/PIV type 1 and PIV3 subgroups of paramyxovirus.

From the comparison of paramyxovirus N protein sequences it is evident that the level of amino acid identity varies over the length of the N protein. The N-terminal one-third of the protein shows a low level of similarity between numerous paramyxoviruses. This N-terminal one-third of the N protein has been found to carry an overall positive charge and therefore could interact with the RNA genome of the virus (Morgan &
Kingsbury, 1984; Sanchez et al., 1986). The middle portion of the protein is the most conserved region within all the groups, as well as between the numerous subgroups of paramyxovirus. The degree of identity observed and the strict conservation of certain residues indicates that this region could have a catalytic role, which could involve interaction with the P and L subunits of the RNA polymerase (Homann et al., 1991). The degree of identity in the C-terminal one-third of the protein is very low and it has been described as a hypervariable region. It is difficult to align this region with confidence as large gaps need to be introduced. Various studies have indicated the C-terminal region of the N proteins (after residue 399) to be on the outside of the protein, based on trypsin digests (Heggeness et al., 1981; Mountcastle et al., 1974) and antigenic studies (Buckland et al., 1989; Gill et al., 1988; Tanabayashi et al., 1990). The C terminus of Mapuera virus N protein is acidic in nature, as has been reported for other paramyxoviruses, which could allow it to interact with the highly basic M protein (Buchholz et al., 1993). Homann et al. (1991) also reported the C terminus of the N protein of SV to be involved in the binding of P protein as deletion of the C-terminal 27 amino acids decreased binding of the P protein by 50% in protein overlay assays. Curran et al. (1993) reported that this region of the Sendai virus N protein was not required for the encapsidation of genomic RNA but was needed for the genomic RNA to function in RNA synthesis. Even though the C-terminal region is highly variable, the Mapuera virus sequence and the sequences of the other paramyxoviruses included in Fig. 5 contain a conserved GDXQ/D (where X represents a variable amino acid residue) motif at the extreme C terminus of the N protein, except for NDV, which has a TDXG motif at the same position. Small conserved motifs are also present at the C termini of members in the genus Morbillivirus (YN + LL; where ‘+ ’ and ‘−‘ represent positively and negatively charged amino acids, respectively) (Rima et al., 1995 b) and the genus Paramyxovirus.

G.W. H. was supported by a Department of Education for Northern Ireland postgraduate studentship.

References


(Received 10 March 1995; Accepted 22 May 1995)