Sequence analysis of the Marburg virus nucleoprotein gene: comparison to Ebola virus and other non-segmented negative-strand RNA viruses

Anthony Sanchez,1* Michael P. Kiley,2 Hans-Dieter Klenk3 and Heinz Feldmann3

1Department of Biology and Laboratory for Microbial and Biochemical Sciences, Georgia State University, Atlanta, Georgia 30302-4010, 2Research and Development Program, The Salk Institute, Government Services Division, Swiftwater, Pennsylvania 18370, U.S.A. and 3Institut für Virologie, Philipps-Universität Marburg, 3550 Marburg, Germany

The first 3000 nucleotides from the 3' end of the Marburg virus (MBG) genome were determined from cDNA clones produced from genomic RNA and mRNA. Identified in the sequence was a short putative leader sequence at the extreme 3' end, followed by the complete nucleoprotein (NP) gene. The 5' end of the NP mRNA was determined as was the polyadenylation site for the NP gene. The transcriptional start (3' UUCUUCUUAAAUU..) and termination (3' ..UAAUUCUUUU) signals of the MBG NP gene are very similar to those seen with Ebola virus (EBO). In comparison to other non-segmented negative-strand RNA viruses, filovirus transcriptional signals are most similar to members of the Paramyxovirus and Morbillivirus genera. In vitro translation of a run-off transcript containing the entire MBG NP coding region produced an authentic NP. Sequence comparisons of the 3' end of the MBG and EBO genomes revealed weak nucleotide sequence similarity, but the predicted sequence of the first 400 amino acids of these viruses showed a high degree. This homology is encoded in divergent nucleotide sequences through different codon usages and substitutions of similar amino acids. A small region in the middle of the MBG and EBO NP sequences was found to contain a significant amino acid homology with NPs of paramyxoviruses and to a lesser extent with rhabdoviruses. Specific sites of conserved sequence are contained in hydrophobic domains and may have a common function. Alignments of the entire NP amino acid sequences of these viruses also suggest that filoviruses are more closely related to paramyxoviruses than to rhabdoviruses.

Introduction

Marburg virus (MBG) is a 'Biosafety Level 4' agent (Richardson & Barkley, 1988) that was first identified in 1967 following human outbreaks of acute haemorrhagic fever in the cities of Marburg and Frankfurt, Germany, and Belgrade, Yugoslavia (Martini & Siegert, 1971). Initial infections occurred in persons working with blood, organs or cell culturing of tissues from infected African green monkeys (Cercopithecus aethiops) imported from Uganda. This pathogen received its name from the city of Marburg, where most of the cases occurred and where much of the initial work on the virus was performed. Three subsequent human outbreaks have been attributed to MBG (Gear et al., 1975; Smith et al., 1982; Kiley et al., 1988).

† Present address: Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control, 1600 Clifton Road, Mail Stop G14, Atlanta, Georgia 30333, U.S.A.

The nucleotide sequence data reported in this paper have been assigned GenBank accession number M72714.

MBG, Ebola virus (EBO) and Reston virus (RES) (Ebola-like monkey filovirus; Centers for Disease Control, 1989; Jahrling et al., 1990) are non-segmented negative-strand (NNS) RNA viruses and are members of the family Filoviridae. Together with the Paramyxoviridae and Rhabdoviridae, these families make up the order Mononegavirales, a status accorded to them in 1990 by the International Committee on Taxonomy of Viruses (Pringle, 1991). The filovirus virion is bacilliform in morphology and is composed of a helical nucleocapsid surrounded by a lipid envelope. Virions contain at least seven structural proteins and for MBG these proteins are an RNA-dependent RNA polymerase (L protein; \( M_r 267K \); unpublished data), a single surface glycoprotein (GP; \( M_r 170K \); Feldmann et al., 1991), a nucleoprotein (NP; \( M_r 94K \)) and four proteins ranging in \( M_r \) from 24K to 38K (Kiley et al., 1988).

The genetic features of filoviruses are similar to other NNS viruses in that (i) transcription of the infecting ribonucleoprotein complex, which contains a single negative-sense genomic RNA template, yields monocistronic polyadenylated mRNA species and (ii) for EBO
and other NNS RNA viruses more fully, we have defined the genetic relationship of MBG to EBO, RES and other NNS RNA viruses more fully, we have undertaken a project of cloning and sequencing the entire genome of a 1980 isolate of MBG (Musoke strain). In this report we present sequence data for the first 3000 nucleotides from the 3' end of the genomic RNA and comparisons of the nucleic acid and predicted amino acid sequences of the MBG NP gene to EBO, paramyxoviruses and rhabdoviruses.

Methods

Cells and viruses. Vero E6 cells were used to culture viruses as previously described (Sanchez & Kiley, 1987). The Musoke strain of MBG was used throughout this study and was derived from the serum of a fatal human infection in Nairobi, Kenya in 1980 (Smith et al., 1982). The virus was isolated and plaque-purified three times on Vero E6 cells, and then large seed stocks were prepared in the same cell line. For comparisons, a Zaire subtype of EBO (Mayinga strain) was used; its passage history is described elsewhere (Sanchez et al., 1989). To define the genetic relationship of MBG to EBO, RES and other NNS RNA viruses more fully, we have undertaken a project of cloning and sequencing the entire genome of a 1980 isolate of MBG (Musoke strain).

Preparation of viral RNAs, molecular cloning and sequencing. Preparation of MBG genomic RNA (vRNA) and mRNA, synthesis of cDNA, molecular cloning in pUC18, identification of virus-specific clones, and chemical and dideoxynucleotide sequencing were performed as previously described (Gubler & Hoffman, 1983; Sanchez et al., 1989; Maxam & Gilbert, 1980; Sanger et al., 1977; Zimmer & Kaesberg, 1978). A synthetic oligodeoxynucleotide primer complementary to the first 19 bases from the 3' end of the genome (5' AGACACACAAAAAACAAAAAACAGAGATG) was used to generate first-strand cDNA (vRNA template) in the production of one cDNA library, and also to probe cDNA libraries generated from vRNA and poly(A)-tailed vRNA. The 5' end of the MBG NP mRNA was sequenced by primer extension, using the primer 5' CCAACAAACTGTGTAAATCCAT (vRNA sense; bases 31 to 61) and 5' CAGGGTACCGCTGCATG-TATGATGAGTCCCACATTGTGA (vRNA sense; bases 2969 to 2940). These primers each contain a KpnI restriction endonuclease site positioned downstream of the T7 RNA polymerase promoter. The DNA from this construct was purified by banding in CsCl gradients, then uncapped run-off RNA transcripts were produced from this DNA after it had been linearized with XhoI. Transcription was performed using T7 RNA polymerase in a large-scale transcription reaction (Promega protocol). The resulting transcript was translated in vitro and labelled with [35S]methionine (New England Nuclear) in a rabbit reticulocyte lysate system (Promega). Translation products were immunoprecipitated with a human anti-MBG serum, subjected to SDS-PAGE and processed for fluorography as previously described (Sanchez & Kiley, 1987).

Results

Cloning and sequence analysis of the 3' end of the MBG genome

A schematic representation of the primary clones used in generating the 3' end sequence data for the MBG genome is shown in Fig. 1. The sequences obtained were double-strand data obtained by either chemical or dideoxynucleotide sequencing of cloned inserts. The clones MV-88 and MV-39 contain the exact 3' end of the genome and were generated by priming first-strand
cDNA synthesis with a 3' complementary synthetic oligonucleotide plus random priming, using vRNA as template. Clone MV-34 contains the 3' end plus a poly(A) tail added to vRNA with poly(A) polymerase (prior to first-strand cDNA synthesis) and primed with oligo(dT). In addition to sequencing of cloned inserts, the sequence from nucleotides 1650 to 3000 was verified through direct dideoxynucleotide sequencing of purified vRNA.

Hybridization of clones MV-88 and MV-17 to Northern blots of MBG vRNA and mRNA transcripts demonstrated their specificity for MBG sequences and identified the transcripts recognized by these clones (Fig. 2a). Clone MV-88 hybridizes to a single transcript, which corresponds to a large MBG mRNA species seen in Fig. 2(b), indicated by an asterisk, and is comparable in size to the EBO NP mRNA. As shown in Fig. 1, the MV-17 clone contains sequences that overlap the first and second genes, and hybridization analysis showed that this clone anneals to the NP transcript and to a second mRNA species that is transcribed from the next (adjacent) gene. The agarose electrophoresis pattern of oligo(dT)-selected MBG mRNA transcripts is very similar to that seen for EBO (Fig. 2b) and also shows them to be polyadenylated and monocistronic.

The 5' end of the NP mRNA was sequenced by primer extension and chemically sequencing the extension product, the results of which are shown in Fig. 2(c). The last band of this sequencing ladder is weaker than the next smaller band and may be due to (i) variability in the exact site in which transcription of the NP mRNA occurs, (ii) copying of a 5' cap structure by reverse transcriptase, (iii) premature termination of the extension product caused by a cap structure or (iv) an artefact of the sequencing chemistry. In any event, from the results obtained from 5' end sequence analysis, the transcriptional start site shown in Fig. 3 represents the longest possible transcript.

Fig. 3 shows the viral complementary sequence (mRNA sense) of the MBG genome, including the entire NP gene. The transcriptional start site and stop or poly(A) site are identified on the sequence and delineate the NP gene. The start site was determined by sequencing the 5' end of the NP mRNA, and the poly(A)
The ORF is initiated by the first AUG from the 5' end of three mRNA clones containing the poly(A) tail region of site was identified by the isolation and sequencing of amino acids in length and has a predicted Mr of 77.86K.

<table>
<thead>
<tr>
<th>350</th>
<th>A. Sanchez and others</th>
</tr>
</thead>
<tbody>
<tr>
<td>301</td>
<td>Ile Ala Leu Gly Val Ala Met Asn Leu TER</td>
</tr>
<tr>
<td>331</td>
<td>Asp Ala Glu Val</td>
</tr>
<tr>
<td>151</td>
<td>Ile Glu Lys Ala</td>
</tr>
<tr>
<td>121</td>
<td>Ser Thr Glu Ser</td>
</tr>
<tr>
<td>271</td>
<td>Arg His Gly Glu</td>
</tr>
<tr>
<td>241</td>
<td>Asp Ser Gly Val</td>
</tr>
<tr>
<td>211</td>
<td>Asp Ser Ile Ile</td>
</tr>
<tr>
<td>181</td>
<td>Val Ile Phe Gly</td>
</tr>
<tr>
<td>631</td>
<td>Asn Trp Pro Gln</td>
</tr>
<tr>
<td>691</td>
<td></td>
</tr>
<tr>
<td>511</td>
<td>Ser Thr Asn Arg</td>
</tr>
<tr>
<td>481</td>
<td>Gly Thr Thr Ser</td>
</tr>
<tr>
<td>601</td>
<td>Ser Ser Pro Ser</td>
</tr>
<tr>
<td>661</td>
<td>Thr Ala Leu Val</td>
</tr>
<tr>
<td>91</td>
<td>Ile Lys Asn Ala</td>
</tr>
<tr>
<td>1</td>
<td>Met Asp Leu Ser</td>
</tr>
</tbody>
</table>

![Fig. 3. Viral complementary sequence (mRNA sense) of the 3' end of the MBG genome. Identified in the figure are the putative leader sequence and the beginning of the NP gene in this contains 695 amino acids.](image)

The MBG NP ORF encodes a protein that is 695 amino acids in length and has a predicted Mr of 77.86K. The ORF is initiated by the first AUG from the 5' end of the mRNA and is flanked by sequences that place it in a favourable context (Kozak, 1986). The NP contains a large amount of negatively charged amino acids, resulting in a highly acidic protein (net charge = -28). As shown with the EBO NP, the majority of the acidic amino acids and proline residues are found in the C-
terminal half of the MBG NP (68% and 75%, respectively, between residues 348 and 695), and all three cysteine residues are found in the N-terminal third of the NP. The MBG NP is also similar to the EBO NP in its hydrophathy plot, seen in Fig. 4, which shows that the NP can be divided into a hydrophobic N-terminal half and a hydrophilic C-terminal half. For comparison, the hydrophathy plots of the NP proteins of Sendai virus (SEN) and vesicular stomatitis (VSV) are also shown in Fig. 4, and it appears that the SEN NP profile is closer to those of the MBG and EBO NPs than that of the VSV NP. This similarity in hydrophathic profiles is particularly prominent in the hydrophobic sequences from residue 130 to approximately 320 (160 to 350 for SEN).

**Sequence comparisons of the MBG and EBO NP genes**

Computer-aided matrix (dot plot) comparisons of the nucleotide and amino acid sequence homologies between the MBG and EBO NP genes are shown in Fig. 5(a). Some scattered similarity is seen in the nucleotide sequences, corresponding to bases 450 to 1100 for MBG and 950 to 1600 for EBO. In contrast, matrix comparison of amino acid sequences reveals a close resemblance between these two proteins in the N-terminal 400 (approximate) residues, with only a small break around residues 120 to 140. An alignment of the NP amino acid sequences of MBG and EBO is shown in Fig. 5(c). The alignment shows that the region from positions 130 to 392 of the MBG sequence has very strong similarity, and is highlighted by a run of 34 identical amino acids (MBG sequence 296 to 329). The strongest identity seen in the nucleic acid matrix comparison corresponds to the strongest regions of identity seen in the amino acid alignment. It should also be noted that two of three cysteines in this alignment are conserved and are nearest to the N terminus.

**Comparisons of the NP amino acid sequences of filoviruses and other NNS RNA viruses**

Comparisons of the predicted NP amino acid sequences of filoviruses and other NNS RNA viruses were performed to determine whether any conserved regions were present. These analyses revealed that in both the MBG and EBO NPs there is a short region which has a significant degree of identity with paramyxovirus NPs, and to a lesser extent with those of rhabdoviruses. This sequence in the MBG and EBO NP corresponds to the highly conserved region in the central part of the protein described above. These findings are illustrated in Fig. 6(a) and (b), which show matrix comparisons of the NP amino acid sequences of MBG and SEN, and MBG and VSV, respectively. The MBG amino acid sequence around the short region of similarity seen in the centre of the MBG/SEN matrix plot (see arrow, Fig. 6a) was aligned to the NP amino acid sequences of six paramyxoviruses and two rhabdoviruses (Fig. 6c). This alignment was initially computer-generated, and then a manual alignment was made to minimize the introduction of gaps. A consensus sequence was derived and asterisks beneath the consensus line mark the locations where both MBG and EBO sequences are identical or have amino acids similar to those of the consensus sequence. At these locations MBG and EBO sequences occurred in 79% of the consensus positions and showed a significantly greater likeness to paramyxoviruses than to VSV or rabies virus. This region of similarity contains the sequences previously identified by Elango (1989) as highly conserved within the family Paramyxoviridae (underlined positions in line with asterisks in Fig. 6c).

Fig. 7 shows a set of computer-generated dendrograms which schematically show the relatedness of the amino acid sequences of the entire and conserved region of the NPs of the viruses seen in Fig. 6(c). The homology of these viral NP sequences can be seen in the clustering of sequences and can be measured by the lengths of the branches in the horizontal axis, which is proportional to
the difference between sequences. These dendrograms are comparable in that the same viruses segregate into the same positions in the alignments. They differ in that the conserved region alignment shows a greater similarity (shorter horizontal lengths) than the entire NP sequence. As expected, rhabdoviruses, paramyxoviruses and filoviruses segregated into family groups, with the sole exception of respiratory syncytial virus (RSV) which surprisingly split into the filovirus branch of each dendrogram and may indicate a closer relationship to these viruses.

In vitro expression of the MBG NP

Fig. 8 shows the SDS–PAGE of immunoprecipitated in vitro translation products compared with purified MBG virion proteins. Lanes 1 and 3 contain translation products; their synthesis was primed by a T7 RNA polymerase-generated run-off transcript containing the MBG NP coding region. Comparison of these products with MBG virion proteins (lane 2) and MBG mRNA-primed and uninfected cell RNA-primed translation products; their synthesis was primed by a T7 RNA polymerase-generated run-off transcript containing the MBG NP coding region. Comparison of these products with MBG virion proteins (lane 2) and MBG mRNA-primed and uninfected cell RNA-primed translation products (lanes 4 and 5, respectively) demonstrates that the most prominent translation product comigrates with the MBG NP seen in virions and produced from mRNA translation. These data confirm that the ORF in Fig. 3 encodes the MBG NP, and in vitro expression of this region results in the synthesis of an authentic NP. In this in vitro system the NP transcript also primes the synthesis of many smaller proteins (which we have also seen with a similar EBO NP transcript) and may arise from the internal initiation of translation.
Discussion

The results obtained from sequence data and expression studies show that, as for most NNS RNA viruses, the MBG NP gene is positioned at the extreme 3' end of the genome and is preceded by a short leader RNA sequence. The putative leader sequences for MBG and EBO have yet to be demonstrated, but we assume that it is delineated by transcriptional signals that act to initiate transcription at the 3' end and terminate transcription [with the concomitant addition of a poly(A) tail] at the 5' end of the gene. The transcriptional signals of the MBG NP gene show a high degree of sequence relatedness with those of the EBO NP gene; the signals differ in three of the 14 bases in the transcriptional start site and are shorter by one U at the end of the poly(A) site (Sanchez et al., 1989). The start sites for the NP genes of MBG and EBO can be related to those of paramyxoviruses (except RSV) in the first and third bases from the 3' end (U and C, respectively) and the presence of the sequence CUU or UUC. The poly(A) sites of the NP genes of filoviruses are very similar to those of viruses in the Paramyxovirus and Morbillivirus genera. The transcriptional signals of the NP genes of filoviruses are distinct from those of other viruses in that they contain a common sequence, 3' UAAUU, that is positioned at the 5' end of the start site and the 3' end of the poly(A) site. This sequence is also seen in the other genomes of MBG and EBO (unpublished data), and its significance in interactions with the viral polymerase is
unknown. The MBG NP ORF is located 56 bases from the 5' end of its mRNA, similar to those described for other NNS RNA viruses, but this is far shorter than the 415 bases present in the EBO NP mRNA. Both the MBG and EBO NP transcripts do, however, have long untranslated regions of 656 and 341 nucleotides [exclusive of added poly(A) tail sequences] at their 3' ends, respectively. The function of these long regions of untranslated sequences is also unknown, but they may influence the level of NP expression or some other viral process.

The 3' end of the MBG genome (first 3000 nucleotides) is AU-rich (59.7%), which has an effect on the codon usage in the ORF (56.4% AU content). An A or a U is present in 43.8% of the first base positions of codons, compared to 63.9% of the second and 62.8% of the third. This bias towards these bases results in an increase in aspartic acid, glutamic acid, asparagine and glutamine, which are abundant in the MBG NP as well as the EBO NP. The ORF for the EBO NP, however, is not as AU-rich and utilizes a somewhat more balanced codon usage.

The composition of the predicted amino acid sequence of the MBG NP is very similar to that of the EBO NP. They can be divided into an N-terminal half which is hydrophobic and a C-terminal half which is decidedly hydrophilic (Fig. 4) and very acidic, a feature that is also characteristic of the NPs of certain paramyxoviruses, such as SEN and human parainfluenza virus type 3 (Sánchez et al., 1986). As seen with the EBO NP, the MBG NP has three cysteine residues that are localized in the N-terminal third and concentrations of proline and acidic residues in the C-terminal half of the molecule. The MBG NP has a predicted M, of 77.9K, much lower than the value calculated from SDS–PAGE, 94 K (Kiley et al., 1988). This difference is not unusual, since similar observations has been reported for the NP of EBO and other related viruses (Sanchez et al., 1989; Galinski et al., 1986; Sánchez et al., 1986; Gallione et al., 1981). The larger mass of filovirus NPs, compared to those of other NNS RNA viruses, may be due to additional sequences at the C terminus.

Computer matrix comparisons of the nucleic acid sequences of the MBG and EBO NP genes (including the 3' leader sequence) indicate that only in the central part of the coding regions is there any significant homology. This finding is in contrast to the homology seen in the predicted amino acid sequences, which show strong identity or similarity in the 400 residues at the N terminus, which is achieved through different codon usage. The greatest degree of identity in these amino acid sequences is located in the central part of the proteins (MBG residues 297 to 330) and corresponds to the region in which the nucleic acid sequences are most similar (MBG bases 990 to 1090). The varying lengths of the untranslated regions of the NP genes of MBG and EBO, together with their different codon usages in
generating very similar NP molecules, suggests that these agents may have diverged at some point in the distant past.

The lack of amino acid sequence homology in the C-terminal half of the NPs of MBG and EBO is analogous to that seen with certain closely related paramyxoviruses, in which the last 100 or more residues are very divergent (Rozenblatt et al., 1985; Galinski et al., 1986; Jambou et al., 1986; Matsuoka & Ray, 1991; Sakai et al., 1987; Lyn et al., 1991). The reason for the divergence of the C termini of these viruses has not been determined, but it has been suggested that this region is exposed to the environment and may interact with the viral matrix protein in the assembly process (Heqgeness et al., 1981; Rima, 1989; Kondo et al., 1990; Barr et al., 1991). If the size of filovirus NPs relative to the NPs of other NNS RNA viruses is due to an increase in length at the C terminus, then one might speculate that interactions of this region with the matrix protein and/or other structural proteins during the budding process could result in the characteristic bacilliform shape of the filovirus virion. Alternatively, this region might interact with a putative second NP found in both EBO and MBG (Kiley et al., 1988; Elliott et al., 1985).

The similarities in hydropathy and primary amino acid sequence between the MBG and EBO NP proteins are outwardly striking, but despite these likenesses the serological evidence indicates that these agents are antigenically unrelated. Localization of major antigenic epitopes in the C-terminal half could explain this phenomenon if cross-reactivity of antibodies to this region is dependent on a high degree of conservation (lacking between MBG and EBO). Immunodominance of the C terminus in stimulating antibody production has been described for the NPs of paramyxoviruses (Gill et al., 1988; Tanabayashi et al., 1990) and a similar condition could exist with filovirus NPs.

Results of matrix comparisons of the NP amino acid sequences of filoviruses and other NNS RNA viruses demonstrated that a small region in the central part of these NPs showed some conservation between MBG and SEN (Fig. 6a). Alignment of this NP region of MBG and EBO to that of other NNS RNA viruses identified certain sites that are conserved in these proteins, particularly between filoviruses and paramyxoviruses. Two areas within this region showing the greatest concentration of similar sequences are underlined in the consensus line of Fig. 6(c). This region has also been shown by others to contain sequences that are highly conserved within the Paramyxoviridae (Galinski et al., 1986; Sánchez et al., 1986; Elango, 1989; Kondo et al., 1990; Lyn et al., 1991; Morgan, 1991) and are contained in the alignment in Fig. 6(c). Recently, Barr et al. (1991) reported that pneumoviruses, paramyxoviruses, rabdoviruses and EBO contain three separate conserved regions, the largest of which is 50 bases and is contained within the alignment shown in Fig. 6(c) (box C; MBG sequence 255 to 304). The other two boxes are both 13 residues in length, one of which is contained in the N-terminal end of the alignment in Fig. 6(c) (box B; MBG sequence 213 to 225) and the other is positioned in the N-terminal third of the NP (box A; MBG sequence 132 to 144). It should be noted that box A corresponds to the beginning of the region of strong similarity seen between MBG and EBO and that box C overlaps the longest run of identity that ends this region. The authors stated that this region appears to begin with sequences that contain a high degree of α-helices and terminates with a β-sheet and reverse turn suggesting a similar conformation. They hypothesized that the first 350 to 400 residues of the NP proteins of these viruses may have retained a similar structure with divergent sequences, and that the regions of common structure (homology) may represent areas involved in RNA binding and interactions with other proteins in the nucleocapsid. Lyn et al. (1991) arrived at a similar conclusion following cross-reaction studies using anti-SEN NP monoclonal antibodies to detect epitopes on the human parainfluenza type 1 virus NP. Our alignment of the amino acid sequences of the MBG and EBO NP proteins (Fig. 5) supports the observations of maintained functional domains in the first 400 N-terminal residues. We have noted that these similarities in amino acid composition occur within an extremely hydrophobic domain, as seen in the hydropathy plots in Fig. 4. The bar above the plots identifies the conserved region aligned in Fig. 6(c) and the two arrows point to two hydrophobic peaks that contain the two most related sites. The peak nearest to the C terminus also corresponds to the sequence previously identified by Elango (1989) as a region highly conserved in paramyxoviruses. Galinski et al. (1986) also observed that conserved regions, identified in alignments of paramyxovirus NP sequences, resided in hydrophobic domains. N-terminal to the conserved region in the MBG, EBO and SEN hydropathy plots in Fig. 4 there is a prominent hydrophobic peak, identified by an asterisk, positioned approximately 80 amino acids from this region. Alignments of the sequences from these regions (MBG or EBO with SEN) failed to show any significant relationship, but they could represent a region whose structure/function is less sequence-dependent. The significance of these hydrophobic sequences in the biology of the NP proteins is unclear, but a role in either folding of the NP and/or binding to the vRNA has been postulated (Morgan, 1991). If these areas represent functional domains, then their characterization would contribute greatly to the basic understanding of the biology of the ribonucleoprotein complex of NNS RNA viruses.
Multiple amino acid sequence alignments of the NPs of representative members of the three families of NNS RNA viruses (seen in the form of dendrograms in Fig. 7) revealed three things. First, the alignment of the NP sequences, save that of RSV, segregated them into their respective families. Second, filoviruses appear more closely related to paramyxoviruses than to rhabdoviruses. And finally, the segregation of RSV with filoviruses may indicate that a closer evolutionary relationship exists between these agents. Using a different alignment procedure, Pringle (1991) generated a similar dendrogram and noted the uniqueness of pneumoviruses. In a paper that compared L amino acid sequences, Stec et al. (1991) concluded that RSV represents a distinct lineage from other paramyxoviruses and showed a dendrogram that is similar to our results. Comparisons of other filovirus genes and gene products should define further the relationship of filoviruses to other NNS RNA viruses, particularly in polymerase amino acid sequences which have been found to be highly conserved (Blumberg et al., 1988; Galinski et al., 1988; Tordo et al., 1988; Barik et al., 1990; Stec et al., 1991) and could provide a better means of gauging genetic relatedness.

In conclusion, sequence analysis of the 3' end of the MBG genome, including the entire NP gene, has shown it to be organized and structured in a manner that is consistent with those of EBO and other NNS RNA viruses. The similarity of filovirus NP genes and gene products to those of paramyxoviruses implies a closer evolutionary relationship than to rhabdoviruses.

References


the hydropathic character of a protein. Journal of Molecular Biology
157, 105–132.
nucleoproteins of human parainfluenza virus type 1 and Sendai virus
share amino acid sequences and antigenic and structural determinants.
MARTINI, G. & SIEGERT, R. (editors) (1971). The Paramyxoviridae:
complete nucleotide sequence of human parainfluenza virus type 3 NP mRNA.
nucleocapsid-associated proteins. In The Paramyxoviruses, pp. 163–
Press.
sequence of the Sendai virus NP gene from a cloned insert. Virology
135, 279–287.
applicable to the search for similarities in the amino acid sequence of
117, 137–140.
cal and Biomedical Laboratories. USPH, CDC. HHS Publication no.
88-8395.
RIMA, B. K. (1989). Comparison of amino acid sequences of the major
structural proteins of the paramyxo- and morbilliviruses. In Genetics and Pathogenicity of Negative-Strand Viruses, pp. 254–263. Edited by
O'MALLEY, B. (1975). Preparation and preliminary characterization
of purified ovalbumin messenger RNA from the hen oviduct. Biochemistry
14, 69–78.
ROZENBLATT, S., EBENBERG, O., BEN-LEVY, R., LAVIE, V. & BELLINI,
sequence of the bovine parainfluenza 3 virus genome: its 3' end and
the genes of NP, P, C and M proteins. Nucleic Acids Research 15,
2927–2944.