Molecular characterization of the Prospect Hill virus M RNA segment: a comparison with the M RNA segments of other hantaviruses

Mark A. Parrington, 1 Pyung-Woo Lee 2 and C. Yong Kang 1 *

1 Department of Microbiology and Immunology, University of Ottawa, Faculty of Medicine, Ottawa, Ontario, Canada K1H 8M5 and 2 Department of Microbiology, Korea University School of Medicine, Seoul, Korea

Complementary DNA representing the genomic M RNA segment of the Prospect Hill (PH) Hantavirus was cloned and its nucleotide sequence determined. The PH virus M RNA segment is 3707 nucleotides in length and has a long open reading frame in the viral complementary-sense RNA with a coding capacity of 1142 amino acids. The predicted gene product of the PH virus M segment was compared with the corresponding gene products of Hantaan virus strain 76-118 (Hantaan), Sapporo rat virus strain SR-11 (SR) and Puumala virus strain Hällnäs B1 (Hällnäs). The amino acid sequence identities between the G1 and G2 proteins of PH virus and Hällnäs virus are respectively 74% and 79%. In contrast, the amino acid sequence similarities between the G1 proteins of PH virus and SR virus or Hantaan virus are only 50%. However the G2 proteins of SR and Hantaan viruses were more closely related to the G2 protein of PH virus with amino acid sequence similarities of approximately 62%. The G1 proteins of all four viruses had three potential asparagine-linked glycosylation sites conserved and there was one conserved site in the G2 proteins. Hydrophilicity plots of the four virus glycoproteins were very similar. The region of greatest hydrophilicity was conserved in the Hällnäs, SR and Hantaan viruses, and was located near the C terminus of the G1 protein, whereas the region of greatest hydrophilicity in the PH virus glycoprotein precursor is located closer to the N terminus of the G1 protein. Our data demonstrate that despite differences in the serotypic profiles and virulence of PH and Hällnäs viruses, their G1 and G2 proteins are closely related. We conclude that PH and Hällnäs viruses may have evolved along a separate evolutionary pathway in the Hantavirus genus from that of SR and Hantaan viruses.

Introduction

Hantavirus is the genus in the Bunyaviridae family (Schmaljohn et al., 1985). Like other viruses in this family, hantaviruses possess a tripartite ssRNA genome with negative polarity (Schmaljohn & Dalrymple, 1983). The three genomic RNA segments, large (L), medium (M) and small (S), have Mr values of approximately 2.2×10^6, 1.3×10^6 and 0.5×10^6, respectively (Yoo & Kang, 1987a). The S RNA segment encodes the nucleocapsid protein (N) and the M RNA segment encodes a glycoprotein precursor that is cleaved into the envelope glycoproteins G1 and G2 (Schmaljohn et al., 1986, 1987; Yoo & Kang, 1987a). The M RNA segment encodes the glycoprotein precursor with a coding capacity of approximately 62%. The G1 proteins of all four viruses had three potential asparagine-linked glycosylation sites conserved and there was one conserved site in the G2 proteins. Hydrophilicity plots of the four virus glycoproteins were very similar. The region of greatest hydrophilicity was conserved in the Hällnäs, SR and Hantaan viruses, and was located near the C terminus of the G1 protein, whereas the region of greatest hydrophilicity in the PH virus glycoprotein precursor is located closer to the N terminus of the G1 protein. Our data demonstrate that despite differences in the serotypic profiles and virulence of PH and Hällnäs viruses, their G1 and G2 proteins are closely related. We conclude that PH and Hällnäs viruses may have evolved along a separate evolutionary pathway in the Hantavirus genus from that of SR and Hantaan viruses.

There are at least four antigenically distinct serotypes recognized in the Hantavirus genus (Lee et al., 1985b; Sugiyama et al., 1987). Each antigenically distinct serotype was isolated from a different rodent species. Serotype 1 includes Apodemus-derived strains associated with the severe form of the human disease, haemorrhagic fever with renal syndrome (HFRS) (Lee et al., 1978). Serotype 2 includes Rattus-derived strains associated with a moderate although potentially still fatal form of HFRS (Lee et al., 1978). Serotype 2 includes Rattus-derived strains associated with a moderate although potentially still fatal form of HFRS (Lee et al., 1982a; Kitamura et al., 1983). Serotype 3 includes Clethrionomys-derived strains associated with a milder form of disease (Brummer-Korvenkontio et al., 1980). Serotype 4 includes Microtus-derived strains that have not been linked to any disease in man (Lee et al., 1982b, 1985a; Yanagihara et al., 1984, 1987).

It has been demonstrated for members of the Bunyavirus genus that the M RNA segment is associated with virulence (Beatty et al., 1981, 1982; Shope et al., 1981; Gonzalez-Scarano et al., 1985; Janssen et al., 1986; Sundin et al., 1987). If this is true for hantaviruses as well,
a comparison between the M segments of pathogenic hantaviruses and a non-pathogenic hantavirus might reveal regions important for Hantavirus virulence. The M RNA segments of at least one disease-causing Hantavirus within each of the serotypes (1, 2 and 3) have been cloned and sequenced (Schmaljohn et al., 1987; Yoo & Kang, 1987; Giebel et al., 1989; Arikawa et al., 1990; Antic et al., 1991). Previously, a comparison of the deduced amino acid sequence of their G1 and G2 proteins demonstrated that Hantaan virus strain 76-118 (Hantaan) (serotype 1) and Sapporo rat virus strain SR-11 (SR) (serotype 2) are much more closely related to each other than to the Puumala virus strain Hällnäs B1 (Hällnäs) (serotype 3) (Arikawa et al., 1990). A comparison between these viruses and the non-pathogenic Hantavirus, Prospect Hill virus, strain Prospect Hill-1 (PH) (serotype 4) may reveal information about Hantavirus virulence.

In this paper we present the nucleotide sequence of the PH virus M RNA segment and the deduced amino acid sequence of its G1 and G2 proteins. This deduced amino acid sequence is compared with the corresponding sequences of Hällnäs, SR and Hantaan viruses. We found that PH and Hällnäs viruses are more closely related to each other than to either SR or Hantaan viruses. Our data further indicate that the Hantavirus genus has diverged into at least two branches and these branches contain genetically similar, but antigenically dissimilar viruses.

Methods

Viruses, cells and media. PH virus, provided by Dr J. M. Dalrymple (USAMRIID, Frederick, Md., U.S.A.), was propagated in Vero E6 cells (ATCC 1008, CRL 1586). Cells were grown in Dulbecco's modified Eagle's medium, supplemented with heat-inactivated foetal bovine serum (10%) and L-glutamine (2 mm). Virus was propagated at 35 °C in 100 mm tissue culture dishes and harvested 8, 12 and 15 days post-infection (p.i.).

Virus isolation and RNA extraction. Virus particles in clarified tissue culture fluid were pelleted by centrifugation for 2 h at 81 000 g at 4 °C in a Beckman SW28 rotor (Beckman). RNA was extracted and purified from virus pellets or cells 9 days p.i. by the guanidinium thiocyanate method (Chirgwin et al., 1979).

Molecular cloning of the PH virus M RNA segment. Random primed cDNA clones representing the PH virus M RNA segment were prepared as described previously (Parrington & Kang, 1990). The 3' and 5' termini of the PH virus M RNA were cloned using the polymerase chain reaction (PCR). To obtain 3'‐specific cDNA, 2 µg of total cellular RNA from PH virus‐infected Vero E6 cells was mixed with 500 ng of a primer complementary to 16 bases at the 3' end of viral M RNA (5'-CAGCTGCAGTGATTGACTGCACC; Fig. 1a, primer MP61), and 500 ng of a viral sense primer (5'-CAGCTGCAGTTGAAAGCATGCTC; Fig. 1a, primer MP63, nucleotide positions 608 to 625, measuring from the 3' end of the M RNA) in 100 µl of the reaction buffer (10 mm-Tris–HCl pH 8.3, 50 mm-KCl, 2.5 mm-MgCl₂, 200 µM of each dNTP, with 100 µg/ml of gelatin) containing 32 units of avian myeloblastosis virus (AMV) reverse transcriptase (Pharmacia LKB) and 5 units of Taq DNA polymerase (Perkin-Elmer Cetus). The sample was incubated at 42 °C for 60 min, heated to 94 °C for 2 min, and then subjected to 40 cycles of PCR (1 min at 94 °C, 2 min at 55 °C and 3 min at 72 °C) using a DNA thermal cycler (Perkin-Elmer Cetus). Both primers were constructed with PstI sites at their 5' ends (sequence shown underlined) for the purpose of cloning. However the resultant PCR fragment, M3', had an internal PstI site so it was blunt‐end ligated into the Smal site of pUC19 instead.

The 5'-terminal clone was made with PCR using primers MP62 and MP61 using the method described previously (Parrington & Kang, 1990). Primer MP62 was complementary to viral M RNA (5'-CAGCTGCAGTGATTGACTGCACC; Fig. 1a, nucleotide positions 2379 to 2396, measuring from the 3' end of M RNA) and was used for first‐strand cDNA synthesis, using viral RNA as template. Primer MP61 had 16 nucleotides identical to the predicted 5' terminus of virion M RNA (sequence shown above). The resultant PCR fragment, M5', was cloned into the PstI site of pUC19 using the PstI sites located at the 5' ends of these primers (sequence shown underlined).

All clones were sequenced in both directions by the dideoxynucleotide chain termination method of Sanger et al. (1977) using several synthetic primers. All primers were synthesized on an Applied Biosystems model 380B DNA synthesizer. Nucleotide sequences were determined by both manual sequencing with Sequenase (United States Biochemical) and automated sequencing using the DuPont Genesis 2000 automated DNA sequencer.

Computer analysis of nucleotide and amino acid sequences. Analysis of the secondary structure of the RNA, searches for potential open reading frames (ORFs), translation of a DNA sequence, locations of signal peptide cleavage sites, hydropathy plots, searches for potential antigenic sites and locations of potential transmembrane helices were completed using the PC/Gene program (Version 5.11, 1987, Department of Medical Biochemistry, University of Geneva, Switzerland, distributed by Intelligenetics).

Analysis of the secondary structure of the RNA was based on Zucker's method (Zucker & Stiegler, 1981) with modifications (Jacobsen et al., 1984). Hydropathy plots and searches for potential antigenic sites were done using the method of Hopp & Woods (1981); signal sequence cleavage sites were determined by the method of von Heijne (1986); the consensus sequence used to determine potential asparagine‐linked glycosylation sites was N-X-S/T-X, where X cannot be proline (P) (Bause, 1983); and prediction of transmembrane helices was based on the method of Rao & Argos (1986).

Results

Molecular cloning and sequence analysis of the PH virus M RNA segment. A cDNA library of clones representing the PH virus M RNA segment was made using random primers as described under Methods. The nucleotide sequence of six clones, P111, R355, T103, S110, P108 and P15 (Fig. 1a), was determined using the dideoxynucleotide chain termination method (Sanger et al., 1977). This nucleotide sequence was compared with that of Hällnäs virus (Giebel et al., 1989). The comparison revealed the following: clone P111 started approximately 340 bases
from the 3' end of the PH virus M RNA; there was a gap of approximately 500 bases between clones P108 and P15; and clone P15 ended approximately 20 bases from the 5' end of the PH virus M RNA. The 3' and 5' termini of the PH virus M RNA were cloned using PCR with Taq DNA polymerase. PCR of the 3' end was done with primers MP61 and MP63 as described under Methods and generated a 643 bp fragment (Fig. 1b, lane 1) that was ligated into the Smal site of pUC19 and designated M3'. Nucleotide sequence analysis of this clone revealed differences with the putative 3'-terminal sequence of the PH virus M RNA previously reported (Schmaljohn et al., 1985). The previously reported sequence had two extra bases at positions 17 and 23 that were not present in our sequence. When we used a primer that was complementary to the previously reported sequence (Schmaljohn et al., 1985) in a cDNA reaction no PH virus M-specific cDNA was produced. Our sequence was identical to the reported Y-terminal sequence of H/illn~is virus for the first 35 bases (Giebel et al., 1989). Therefore, clone M3' most likely represents the actual 3' end of the PH virus M RNA. The 5' end of the M RNA was also cloned using PCR with primers MP62 and MP61. The 1344 bp fragment (Fig. 1c, lane 2) was ligated into the PstI site of pUC19 and designated M5'. The nucleotide sequence of clones M3' and M5' was determined using the dideoxy-nucleotide chain termination method (Sanger et al., 1977). The combined and consensus sequence (3707 bases) of M3', P111, R355, T103, S110, P108, P15 and M5' are shown in Fig. 2. The base composition of the PH virus M RNA segment is 29.8% A, 18.8% G, 21.7% C, and 29.7% U. These values are similar to those of the H/illn~is, SR and Hantaan virus M segments (Schmaljohn et al., 1987; Giebel et al., 1989; Arikawa et al., 1990). Comparison of the 3' and 5' termini of the PH virus M segment in the genomic RNA sense revealed an inverse complementary sequence involving 38 of the terminal 45 nucleotides (Fig. 3). The calculated free energy of this structure is −196.9 kJ/mol at 25 °C. The stability of this structure is similar to that of the M segment of Hantaan virus (−177.2 kJ/mol at 25 °C) and to that of SR virus (−165.5 kJ/mol at 25 °C), but is much lower than that of the H/illn~is virus (−270.9 kJ/mol at 25 °C) (Giebel et al., 1989).

**Coding strategy of the PH virus M RNA**

The PH virus M RNA segment is 3707 bases long (Fig. 2). This is the largest Hantavirus M segment reported to date as compared to 3682 bases for H/illn~is virus, 3651 bases for SR virus, and 3616 bases for Hantaan virus (Schmaljohn et al., 1987; Yoo & Kang, 1987b; Giebel et al., 1989; Arikawa et al., 1990). There is a large ORF in frame 2 of the PH virus M cRNA. The PH virus M ORF initiates at nucleotides 50 to 52, terminates at nucleotides 3476 to 3478, and has a coding capacity of 1142 amino acids (Fig. 2). This is eight and six amino acids larger than the SR virus (1134 amino acids) and Hantaan virus...
Arikawa et al., but is six amino acids smaller than the H/illn~is virus (1148 amino acids) glycoprotein precursor (Schmaljohn et al., 1987; Yoo & Kang, 1987b; Giebel M. A. Parrington, P.-W. Lee and C. Y. Kang et al., 1989; Arikawa et al., 1990).

The deduced amino sequences of the G1 and G2 proteins of the PH, H/illn~is, SR and Hantaan viruses were compared (Fig. 4a, b). The cleavage site between the signal peptide and N terminus of G1 is known for SR and Hantaan viruses (Schmaljohn et al., 1987; Arikawa...
et al., 1990). However the signal peptide cleavage sites for PH and Hällnäs viruses were based on computer predictions. Conservation of cysteine residues between PH, Hällnäs, SR and Hantaan viruses is very high (G1 and G2 for PH and Hällnäs viruses were based on computer and Hantaan viruses to PH virus are summarized in comparison of the G1 and G2 proteins of Hällnäs, SR and Table 1. The PH virus G1 and G2 proteins have higher amino acid sequence identities with the Hällnäs virus G1 and G2 proteins than with the G1 and G2 proteins of SR and Hantaan viruses. The G2 protein of all four viruses has been conserved to a greater extent than the G1 protein (Fig. 4b). A possible variable region was identified in the G1 protein of these four viruses (PH virus amino acid positions 209 to 282) (Fig. 4a). More amino acid sequence heterogeneity between the four viruses was observed in this region as compared with other regions of the G1 proteins. Amino acid sequence similarity between all four viruses was 38-7% for the G1 protein and 52-2% for the G2 protein. If conservative amino acid substitutions are included these numbers increase to 54-6% for the G1 protein and 67-1% for the G2 protein.

Three of the four potential asparagine-linked glycosylation sites in the PH virus G1 protein were conserved in the other three viruses. A potential glycosylation site at amino acid positions 527 to 529 was not conserved in the other three viruses (Fig. 4a). However SR virus had two other potential sites (amino acid positions 233 to 235 and 560 to 562) whereas Hantaan virus had only one other potential site (amino acid positions 220 to 225, Fig. 5). This region is apparently associated with an amino acid deletion in the PH virus G1 protein as compared with the G1 proteins of the other three viruses (Fig. 4a). Interestingly a possible glycosylation site that is probably not used in Hällnäs virus (N-P-S, amino acid positions 585 to 587) is also conserved in PH virus (N-P-T, amino acid positions 581 to 583) (Fig. 4a). The one potential glycosylation site in the PH virus G2 protein is conserved in all four viruses (Fig. 4b). The second site in the Hällnäs virus G2 protein (amino acid positions 898 to 900) is not observed in the PH virus G2 protein (Fig. 4b).

The analysis of potential transmembrane helices (Rao & Argos, 1986) identified only one conserved region in the G1 protein and one conserved region in the G2 protein of all four viruses. The region in the G1 protein ends 132 amino acids from C terminus (PH virus amino acid positions 447 to 522) and was 76 amino acids in length in PH and Hällnäs viruses, and 71 amino acids in length in SR and Hantaan viruses (Fig. 4a). The conserved region in the G2 protein is near the C terminus (PH virus amino acid positions 1102 to 1132) and is 31 amino acids long in all four viruses (Fig. 4b). Hydrophilicity profiles (Hopp & Woods, 1981) of the four virus glycoprotein precursors were compared (Fig. 5). There were many similarities between the four virus hydrophilicity profiles. The region of greatest hydrophilicity in the Hällnäs virus profile was conserved in the profiles of the SR and Hantaan viruses, and is located near the C terminus of G1 beginning 14 amino acids after the potential transmembrane domain (Fig. 5). The SR and Hantaan virus profiles also have a second equally hydrophilic region near the C terminus of G1 (Fig. 5). In contrast the region of greatest hydrophilicity in the PH virus glycoprotein precursor is closer to the N terminus (amino acid positions 220 to 225, Fig. 5). This region is apparently associated with an amino acid deletion in the PH virus G1 protein as compared with the G1 proteins of the other three viruses (Fig. 4a). When hydrophilicity profiles of the G2 proteins were compared, the highest peaks of hydrophilicity in Hällnäs, SR-11 and Hantaan viruses was located at the C terminus. The region of greatest hydrophilicity in the PH virus G2 protein however, is located 12 amino acids prior to the transmembrane domain (amino acid positions 1084 to 1089, Fig. 4b).
Hantavirus isolates have been divided into at least four antigenically distinct serotypes by the immunofluorescent antibody, immune adherence haemagglutination, and plaque reduction neutralization tests (Lee et al., 1985b; Sugiyama et al., 1987). A comparison of the G1 proteins of Hantaan virus (serotype 1) and SR virus (serotype 2) revealed amino acid similarities of 75% or 83% if conservative changes were included. The G2 proteins of these viruses were even more highly conserved with amino acid similarities of 82% or 92% if conservative amino acid substitutions were included (Arikawa et al., 1990). These figures were similar to the ones we found when comparing the PH virus (serotype 4) and Hällnäs virus (serotype 3) G1 and G2 proteins (G1 74-3%, or 85-8% with conservative changes; G2 79%, or 88-4% with conservative changes). When the PH virus (serotype 4) G1 and G2 proteins were compared with those of Hantaan (serotype 1) and SR viruses (serotype 2), the amino acid similarity was much lower. These data suggest that the Hantavirus genus diverged into at least two evolutionary pathways, one composed of the viruses in serotypes 1 and 2, and another of viruses in serotypes 3 and 4. Other hantaviruses not belonging to these four serotypes may belong to one of these two lineages or may reveal further branches in the Hantavirus genus. This divergence may have been caused by virus adaptation to the host, since rodent species, not geography, determine the serotypic profile (Lee et al., 1985b; Sugiyama et al., 1987).

A distinguishing feature of some of the hantaviruses in serotypes 1, 2 and 3 is that they can cause disease in humans. Hantaan virus (serotype 1) is the aetiological agent of a severe form of HFRS (Lee et al., 1978), SR virus (serotype 2) is linked to a moderate although potentially still fatal form of HFRS (Lee et al., 1982a; Shope et al., 1981), and Hällnäs virus (serotype 3) is associated with a milder form, nephropathia epidemica (Brummer-Korvenkontio et al., 1980). In contrast, PH virus (serotype 4) has not been linked to any disease in humans (Yanagihara et al., 1984, 1987). In bunyaviruses the M RNA segment, specifically the G1 protein, has been identified as a major determinant of virulence and infectivity (Beaty et al., 1981, 1982; Shope et al., 1981; Gonzalez-Scarano et al., 1985; Janssen et al., 1986; Sundin et al., 1987). If this is true for hantaviruses as well, the large number of amino acid changes observed in the Hällnäs virus G1 and G2 proteins as compared with SR and Hantaan viruses may explain why this virus only causes the milder form of the disease. The similarity between the G1 and G2 proteins of PH and Hällnäs

<table>
<thead>
<tr>
<th>Virus</th>
<th>PH</th>
<th>Hällnäs</th>
<th>Hantaan</th>
<th>Hantaan</th>
<th>Hallnas</th>
<th>Hantaan</th>
<th>Hantaan</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>894</td>
<td>930</td>
<td>904</td>
<td>884</td>
<td>949</td>
<td>959</td>
<td>958</td>
</tr>
<tr>
<td></td>
<td>TGA</td>
<td>F</td>
<td>P</td>
<td>N</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>K</td>
<td>Q</td>
<td>T</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>K</td>
<td>S</td>
<td>T</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>KR</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>904</td>
<td>940</td>
<td>1007</td>
<td>1067</td>
<td>1099</td>
<td>1129</td>
<td>1129</td>
</tr>
<tr>
<td></td>
<td>TGA</td>
<td>F</td>
<td>P</td>
<td>N</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>K</td>
<td>Q</td>
<td>T</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>K</td>
<td>S</td>
<td>T</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>KR</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
</tr>
</tbody>
</table>

Fig. 4. Comparison of the predicted amino acid sequence of the PH virus, Hällnäs virus, SR virus and Hantaan virus G1 proteins (a) and G2 proteins (b). Amino acids are numbered on the left and right ends of each line. Non-identical amino acids are denoted by dots, conservative amino acid changes are underlined, and missing amino acids in any sequence are denoted with a dash. Small boxes denote potential N-linked glycosylation sites. The large box denotes the potential transmembrane domain common to all four viruses.

Discussion

Hantavirus isolates have been divided into at least four antigenically distinct serotypes by the immunofluorescent antibody, immune adherence haemagglutination, and plaque reduction neutralization tests (Lee et al., 1985b; Sugiyama et al., 1987). A comparison of the G1 proteins of Hantaan virus (serotype 1) and SR virus (serotype 2) revealed amino acid similarities of 75% or 83% if conservative changes were included. The G2 proteins of these viruses were even more highly conserved with amino acid similarities of 82% or 92% if conservative amino acid substitutions were included (Arikawa et al., 1990). These figures were similar to the ones we found when comparing the PH virus (serotype 4) and Hällnäs virus (serotype 3) G1 and G2 proteins (G1 74-3%, or 85-8% with conservative changes; G2 79%, or 88-4% with conservative changes). When the PH virus (serotype 4) G1 and G2 proteins were compared with those of Hantaan (serotype 1) and SR viruses (serotype 2), the amino acid similarity was much lower. These data support earlier work suggesting that serotypes 3 and 4 are antigenically closer to each other than to serotypes 1 or 2 (Sugiyama et al., 1987). These data suggest that the Hantavirus genus diverged into at least two evolutionary pathways, one composed of the viruses in serotypes 1 and 2, and another of viruses in serotypes 3 and 4. Other hantaviruses not belonging to these four serotypes may belong to one of these two lineages or may reveal further branches in the Hantavirus genus. This divergence may have been caused by virus adaptation to the host, since rodent species, not geography, determine the serotypic profile (Lee et al., 1985b; Sugiyama et al., 1987).

A distinguishing feature of some of the hantaviruses in serotypes 1, 2 and 3 is that they can cause disease in humans. Hantaan virus (serotype 1) is the aetiological agent of a severe form of HFRS (Lee et al., 1978), SR virus (serotype 2) is linked to a moderate although potentially still fatal form of HFRS (Lee et al., 1982a; Shope et al., 1981), and Hällnäs virus (serotype 3) is associated with a milder form, nephropathia epidemica (Brummer-Korvenkontio et al., 1980). In contrast, PH virus (serotype 4) has not been linked to any disease in humans (Yanagihara et al., 1984, 1987). In bunyaviruses the M RNA segment, specifically the G1 protein, has been identified as a major determinant of virulence and infectivity (Beaty et al., 1981, 1982; Shope et al., 1981; Gonzalez-Scarano et al., 1985; Janssen et al., 1986; Sundin et al., 1987). If this is true for hantaviruses as well, the large number of amino acid changes observed in the Hällnäs virus G1 and G2 proteins as compared with SR and Hantaan viruses may explain why this virus only causes the milder form of the disease. The similarity between the G1 and G2 proteins of PH and Hällnäs
viruses suggests that these viruses may have undergone amino acid changes that attenuated virulence in these viruses as compared with SR and Hantaan viruses. A possible variable region was identified in the G1 protein of these viruses and may be involved in the differences in virulence between these four viruses. Also, there was less conservation of the G1 protein as compared with the G2 protein. This could mean that Hantavirus virulence is associated with the G1 protein. The only major difference observed between the PH virus G1 and G2 proteins and those of the other three viruses was in the hydrophilicity profiles. The highest peak of hydrophilicity in the PH virus glycoprotein precursor is apparently associated with an amino acid deletion in the PH virus G1 protein as compared with the G1 proteins of the other three viruses. It is not presently known whether these differences are involved in the lack of virulence of PH virus. However, changes as small as a single amino acid substitution in the haemagglutinin protein were shown to alter the virulence of influenza A viruses (Kawaoka et al., 1984; Philpott et al., 1990).

Only one potential transmembrane region was conserved in the G1 proteins and one region in the G2 proteins of all four viruses. This suggests that these regions are possibly membrane-spanning domains that anchor the G1 and G2 proteins to the surface of the viral envelope.

In our results, we show the locations of potential N-glycosylation sites in the G1 and G2 proteins of the four viruses. One of the potential sites in the G1 proteins of the PH and SR viruses was on the cytoplasmic domain of the transmembrane part and was therefore unlikely to be glycosylated. Also, there is some doubt that the amino acid sequence N-D-S/T is recognized for glycosylation (Kornfeld & Kornfeld, 1985). This would eliminate the fourth site in PH virus and one potential site in Hantaan virus.

The sequence we report here for the 24 bases at the 3' end of the PH virus M RNA differs from the sequence reported using direct RNA sequencing (Schmaljohn et al., 1985). When we used a primer (21-mer) complementary to the sequence of Schmaljohn et al. (1985) we were unable to generate any PH virus M-specific cDNA. Any cDNA that was produced represented the PH virus L RNA segment. Therefore, we believe our sequence was the correct one for the PH virus M RNA segment, but we cannot explain the differences with the earlier results.

Our data show that antigenically distinct hantaviruses may be more closely related than serotypic studies would indicate. It also demonstrates the possibility that many antigenically distinct hantaviruses may have diverged along only two evolutionary pathways. Further work with the PH virus M RNA segment may eventually give us a better understanding of Hantavirus virulence.
We would like to thank Dr K. E. Wright, Dr E. G. Brown and Dr K. D. Dimock for their constructive review of this manuscript. We also would like to thank N. DeCellier for preparing oligonucleotide primers and B. Mah and D. McLean for operation of the automated DNA sequencer. This study is supported by grants from the Natural Sciences and Engineering Research Council of Canada, and the URF program from the Ontario Ministry of Colleges and Universities.

References


(Received 21 January 1991; Accepted 5 April 1991)