Completion of the Tula hantavirus genome sequence: properties of the L segment and heterogeneity found in the 3′ termini of S and L genome RNAs

Sami K. J. Kukkonen, Antti Vaheri and Alexander Plyusnin

Department of Virology, Haartman Institute, PO Box 21, FIN-00014 University of Helsinki, Finland

In this study the L segment and the 5′ and 3′ termini of the S, M and L segments of the prototype Tula hantavirus (TUL) were sequenced, thus completing the first determination of the genome sequence of a hantavirus that has not been linked to any human disease. The TUL L segment comprises 6541 nt with one ORF of 6459 nt in the antigenome sense. This ORF potentially encodes a 2153 aa protein with a predicted molecular mass of 247 kDa. The amino acid sequence includes all the motifs conserved in RNA-dependent RNA polymerases. The 5′ termini of all three genome RNAs (vRNAs) had the expected sequences conserved in hantaviruses. The 3′ termini of M vRNAs were also conserved. However, the 3′ termini of S and L vRNAs were heterogeneous as most of the sequenced 3′ termini had either deletions of 1 to 22 nt or an extra 1 to 3 nt. No increase in the level of heterogeneity was seen in vRNAs of virions collected 3, 6, 9 and 12 days post-infection, suggesting that the heterogeneity already exists at the early stages of infection. The S and L vRNAs of infected cells had more truncated 3′ termini than vRNAs from pelleted virus. Heterogeneity of the 3′ termini of genome RNAs could decrease the efficiency of antigencode and mRNA syntheses and contribute to the slow growth observed for TUL and other hantaviruses in cell culture.

Introduction

Tula virus (TUL) belongs to the genus Hantavirus, family Bunyaviridae (Plyusnin et al., 1994; Vapalahti et al., 1996). Hantaviruses are rodent-borne agents which have co-evolved with their natural hosts. Consequently, hantaviruses of closely related rodents share common features, including pathogenicity to humans (Plyusnin et al., 1996b; Schmaljohn & Hjelle, 1997). Hantaviruses carried by rodents of the genus Microtus, subfamily Microtinae, have not been found to be associated with any human disease. This is also the case for TUL, which has the European common vole (Microtus arvalis) and eastern vole (Microtus rossiaemeridionalis) as its natural hosts.

The genomes of hantaviruses consist of three negative-strand RNA segments: Large (6.5–6.6 kb), Medium (3.6–3.7 kb) and Small (1.7–2.1 kb) (Schmaljohn, 1996). They encode in the viral complementary sense the four structural proteins: the L segment encodes an RNA polymerase (~250 kDa), the M segment two glycoproteins (G1 ~ 70 kDa and G2 ~ 50 kDa, without the glycans) and the S segment a nucleocapsid protein (~50 kDa).

The 5′ and 3′ termini of hantavirus genome segments are highly conserved and complementary to each other (Schmaljohn, 1996). In hantaviruses the complementarity extends to at least 17 bases. In most hantaviruses 14 of the 17 bases from each terminus are identical in all three genome segments and the complementarity of the termini is incomplete, with a mismatch at position 9 and a non-canonical U–G pair at position 10 (Yoo & Kang, 1987; Parrington & Kang, 1990; Bowen et al., 1995, Chizhikov et al., 1995).

The 5′ and 3′ termini of vRNAs are thought to form a double-stranded promoter regulating RNA transcription and replication, similar to what has been demonstrated for Rift Valley fever virus of the family Bunyaviridae (Prehaud et al., 1997), and for influenza (Fodor et al., 1994, 1995; Lee & Seong, 1996) and Thogoto (Leahy et al., 1997) viruses, which are also segmented, negative-strand RNA viruses. The 3′ termini of the genome segments contain three trinucleotide repeats, AUC AUC AUC…., which may be involved in a prime-and-realign mechanism of initiation of RNA synthesis (Garcin et al., 1995).

Approximately 30 distinct hantavirus sero/genotypes are known. However, the complete genome sequence has been reported for only four hantavirus species: Hantaan (Schmaljohn...
et al., 1986, 1987; Schmaljohn, 1990), Seoul (Arikawa et al., 1990; Antic et al., 1991), Puumala (Giebel et al., 1987; Stohwasser et al., 1990, 1991; Vapalahti et al., 1992; Piparinen et al., 1997) and Sin Nombre (Spiropoulou et al., 1994; Chizhikov et al., 1995).

L is the longest fragment and thus the most difficult to sequence. As both the L segment gene product and terminal sequences of the genome segments are thought to be crucial for transcription and replication, obtaining the sequences of L and the termini is essential for establishing a reverse-genetic system, which is our goal. Such systems have been developed for Bunyamwera virus (Dunn et al., 1994, 1995; Bridgen & Elliott, 1996) and Rift Valley fever virus (Prehaud et al., 1997), representing two genera of the family Bunyaviridae, but not yet for hantaviruses.

Previously, the S and M segments of TUL were sequenced apart from the 5’ and 3’ termini, which were deduced from primer sequences (Plyusnin et al., 1994, 1995; Vapalahti et al., 1996); the L segment was partially sequenced (nt 1914–2909) (Vapalahti et al., 1996). The aim of the present study was to complete the genome sequence of TUL by determining the sequence of the L segment as well as the terminal sequences of all three genome segments.

Methods

■ Virus. The TUL prototype strain Tula/Moravia/Ma5302V, isolated from naturally infected Microtus arvalis (vole) and passaged in Vero E6 cells (Vapalahti et al., 1996), was used in the study. Passage 3 was used for sequencing of the L segment and the termini. The growth medium of infected cells was used for infecting fresh cell cultures.

■ Sequencing of the L segment. Vero E6 cells were infected with passage 3 of the above TUL strain at m.o.i. < 0·1. At 11–13 days post-infection (p.i.) cells were harvested, and total RNA was extracted using the guanidium thiocyanate–phenol–chloroform extraction method (Chomczynski & Sacchi, 1987). The RNA was reverse transcribed using Moloney murine leukaemia virus reverse transcriptase (MMLV RT) (Life Technologies). Primers for reverse transcription and PCR were initially designed based on known L segment sequences of other hantaviruses. As partial sequences of TUL L were obtained, this information was used to design TUL-specific L primers. (Sequences of all primers are available on request.) PCR amplifications were performed with AmpliTaq thermostable polymerase (Perkin-Elmer). PCR products were cloned using the pGEM-T Vector System (Promega) or the TA Cloning Kit (Invitrogen). The L segment was cloned in five parts, which were sequenced from at least two clones for each part by the dyeodeoxy chain termination method (Sanger et al., 1977) using the Sequenase version 2.0 DNA Sequencing Kit (United States Biochemical) according to the manufacturer’s recommendations. After completion of the initial sequencing the L segment was reverse transcribed and PCR amplified with the Expand High Fidelity PCR System (Boehringer Mannheim) in two parts to facilitate the obtaining of a full-length clone. These two parts were cloned and also sequenced from at least two clones confirming the sequence obtained initially.

■ Sequencing of the 5’ and 3’ termini of vRNAs. Vero E6 cells were infected with TUL passage 3 and at 3, 6, 9 and 12 days p.i. virus was pelleted by centrifuging the growth medium through a 30% (w/v) sucrose cushion. RNA was extracted from the virus preparation and the cells as described above. The 5’ and 3’ termini of the RNAs were ligated with T4 RNA ligase (New England Biolabs). After RNA ligation the reaction mixture was divided into three aliquots which were used as templates for reverse transcription with S-, M- and L-specific primers using MMLV RT. The ligation products of RNA from virions were reverse transcribed with primers specific for genome-sense RNA. Ligation products of RNA from infected cells were reverse transcribed with primers specific for genome- and antigenome-sense RNA in separate reactions. Reverse transcription was followed by nested PCR. For the L segment the amplified regions included nt 6265–340 in the first PCR and nt 6346–294 in the second, nt 3375–209 and nt 3446–120 for the M segment, and nt 1426–157 and nt 1507–107 for the S segment. Cloning and sequencing were done as described above. The complete nucleotide sequence of TUL L segment was deposited in the EMBL Database under accession number AJ05637.

■ Sequence analyses. Sequence analyses were done using the programs of the GCG package (Wisconsin Package version 9.0, Genetics Computer Group, Madison, WI, USA). Fitch–Margoliash phylogenetic trees based on the coding regions of the completely sequenced L segments were constructed using the PHYLIP program package (Felsenstein, 1993). Nucleotide sequences for multiple sequence alignments and calculations of similarities were obtained from the EMBL Database under the following accession numbers: Tula virus, strain Moravia/Ma5302V, M: Z69993 and S: Z69991 (Vapalahti et al., 1996); Puumala virus, strain Sotkamo, L: Z66548 (Piparinen et al., 1997), M: X61034, and S: X61035 (Vapalahti et al., 1992); Puumala virus, strain Bashkirtia/Cg18-20, L: M63194 (Stohwasser et al., 1991), M: M29079 (Giebel et al., 1989), and S: M32750 (Stohwasser et al., 1990); Sin Nombre virus, strain NMH10, L: L37901 (Chizhikov et al., 1995), M: L25783, and S: L25784 (Spiropoulou et al., 1994); Hantaan virus, strain 76-118, L: X55901 (Schmaljohn, 1990), M: M14627 (Schmaljohn et al., 1987), and S: M14626 (Schmaljohn et al., 1986); and Seoul virus, strain 80-39, L: X56402, M: S47716 (Antic et al., 1991), and S (strain SR-11): M34881 (Arikawa et al., 1990).

Results

Characterization of the L segment of TUL

RNA extracted from infected cells was amplified with L-specific primers in five overlapping parts: nt 1–953, 879–2954, 2863–4220, 4093–5919 and 5886–6541. The L segment was found to be 6541 bases long with only one large ORF in the noncoding region following the first stop codon consisted «». The conserved amino acids of noncoding region was 36 bases long and the predicted molecular mass of 246816 Da. The deduced amino acid sequence contains the motifs conserved in all RNA-dependent polymerases, from premotif A to motif E (Poch et al., 1989; Müller et al., 1994). The conserved amino acids of these motifs are thought to be located in the polymerase domain and to be essential for the enzymatic activity of the
Table 1. TUL L, M and S sequences: comparison with other hantaviruses

<table>
<thead>
<tr>
<th></th>
<th>TUL</th>
<th>PUUs</th>
<th>PUUb</th>
<th>SN</th>
<th>HTN</th>
<th>SEO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nt</td>
<td>6541</td>
<td>6550</td>
<td>6550</td>
<td>6562</td>
<td>6530</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>75</td>
<td>75</td>
<td>71</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>aa</td>
<td>2153</td>
<td>2156</td>
<td>2156</td>
<td>2153</td>
<td>2151</td>
</tr>
<tr>
<td>kDa</td>
<td>246</td>
<td>247</td>
<td>247</td>
<td>246</td>
<td>245</td>
<td>246</td>
</tr>
<tr>
<td>%</td>
<td>85</td>
<td>85</td>
<td>79</td>
<td>68</td>
<td>69</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>nt</td>
<td>3694</td>
<td>3662</td>
<td>3682</td>
<td>3696</td>
<td>3616</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>71</td>
<td>72</td>
<td>67</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>aa</td>
<td>653</td>
<td>658</td>
<td>658</td>
<td>652</td>
<td>648</td>
</tr>
<tr>
<td>kDa</td>
<td>72.1</td>
<td>72.7</td>
<td>73.1</td>
<td>72.0</td>
<td>72.5</td>
<td>71.9</td>
</tr>
<tr>
<td>%</td>
<td>76</td>
<td>75</td>
<td>64</td>
<td>50</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nt</td>
<td>1831</td>
<td>1830</td>
<td>1784</td>
<td>2059</td>
<td>1696</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>73</td>
<td>73</td>
<td>70</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>aa</td>
<td>429</td>
<td>433</td>
<td>433</td>
<td>429</td>
<td>429</td>
</tr>
<tr>
<td>kDa</td>
<td>49.5</td>
<td>49.5</td>
<td>49.5</td>
<td>48.2</td>
<td>48.1</td>
<td>48.2</td>
</tr>
<tr>
<td>%</td>
<td>79</td>
<td>80</td>
<td>73</td>
<td>63</td>
<td>62</td>
<td></td>
</tr>
</tbody>
</table>

polymerase (Müller et al., 1994). The amino acid sequence also includes two residues, glutamate and lysine, located between premotif A and motif A, and motif E with the tetrapeptide E(F/Y)XS, that are conserved in polymerases of viruses with segmented negative-strand RNA genomes (Müller et al., 1994).

L proteins of hantaviruses are highly conserved in the regions of the polymerase motifs. Most of the few amino acid changes that do occur are conservative substitutions. The L protein of TUL has the most extensive amino acid identity with the L proteins of the two strains of Puumala virus followed by the L proteins of Sin Nombre, Hantaan and Seoul viruses (Table 1).

Close to the carboxy terminus of the L protein of TUL there is an acidic region of 38 aa that has also been described for the L proteins of Puumala virus (Stohwasser et al., 1991; Piiparinen et al., 1997) and Sin Nombre virus (Chizhikov et al., 1995). In the L protein of TUL, 27 of the 38 amino acids are either aspartate or glutamate and the isoelectric point of the region is 2.93 compared to 7.00 for the complete L protein. This region is acidic in all hantavirus L proteins but the amino acid identity within this region is not particularly extensive. Gaps of 3 to 5 aa have to be introduced in the amino acid sequences of the L proteins of TUL, Sin Nombre, Hantaan and Seoul viruses in order to align them with the amino acid sequence of the L protein of Puumala virus.

Outside the motifs described there was no clear pattern to the distribution of substitutions with the exception of four variable regions surrounding aa 145, 280, 770 and 1670.

Completion of the TUL L sequence allowed comparison of the genome segments and proteins of the six completely sequenced hantaviruses (Table 1). The sizes of the genome segments are similar for all six hantaviruses with the exception of the S segments, in which the size of the 3′ noncoding region varies. The proteins encoded are of similar size. The coding region of the L segment is the most conserved followed by the coding regions of the S and M segments. At the amino acid level, L protein and G2 glycoprotein are the most conserved whereas G1 glycoprotein is the most variable.

In accordance with the phylogenetic relationships determined from the S and M segments, TUL is placed in the closest proximity to Puumala virus in a phylogenetic tree based on the coding regions of the L segments (Fig. 1). This reflects the fact that while hosts for both TUL and for Puumala viruses are A. collicola rodents, the other three hantaviruses are carried by hosts belonging to the rodent subfamilies Sigmodontinae (Sin Nombre virus) or Murinae (Hantaan and Seoul viruses).

Analysis of the 5′ and 3′ termini of TUL RNA segments

To obtain exact terminal sequences of TUL genome RNA segments, RNA was initially extracted at 11 days p.i.; the 5′ and 3′ termini of each RNA segment were ligated, and then reverse transcribed and PCR-amplified through the ligation
(a) Tula S termini

3 days p.i.

3'.........CAUGAAUCUGAGACGUACUGAU...
3'.........CUGAGAACUUCCUGA...
3'.........CAUGAAUCUGAGACGUACUGAU...
3'.........CUGAGAACUUCCUGA...
3'.........CAUGAAUCUGAGACGUACUGAU...
3'.........CUGAGAACUUCCUGA...

12 days p.i.

5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CUGAGAACUUCCUGA...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CUGAGAACUUCCUGA...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CUGAGAACUUCCUGA...

12 days p.i. RNA from cells

5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CUGAGAACUUCCUGA...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CUGAGAACUUCCUGA...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CUGAGAACUUCCUGA...

(b) Tula M termini

3 days p.i.

7'.........UGUGAAGAAUAGGACUGUGAU...
7'.........UGUGAAGAAUAGGACUGUGAU...
7'.........UGUGAAGAAUAGGACUGUGAU...
7'.........UGUGAAGAAUAGGACUGUGAU...

12 days p.i.

5'.........UGUGAAGAAUAGGACUGUGAU...
5'.........UGUGAAGAAUAGGACUGUGAU...
5'.........UGUGAAGAAUAGGACUGUGAU...
5'.........UGUGAAGAAUAGGACUGUGAU...

12 days p.i. RNA from cells

5'.........UGUGAAGAAUAGGACUGUGAU...
5'.........UGUGAAGAAUAGGACUGUGAU...
5'.........UGUGAAGAAUAGGACUGUGAU...
5'.........UGUGAAGAAUAGGACUGUGAU...

(c) Tula L termini

3 days p.i.

4'.........CAUGAAUCUGAGACGUACUGAU...
4'.........CAUGAAUCUGAGACGUACUGAU...
4'.........CAUGAAUCUGAGACGUACUGAU...
4'.........CAUGAAUCUGAGACGUACUGAU...

12 days p.i.

5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CAUGAAUCUGAGACGUACUGAU...

12 days p.i. RNA from cells

5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CAUGAAUCUGAGACGUACUGAU...
5'.........CAUGAAUCUGAGACGUACUGAU...

Fig. 2. For legend see facing page.
point. The PCR products were cloned and sequenced (see Methods). Surprisingly, heterogeneity was discovered in the termini of genome segments. It was decided to study this phenomenon in more detail to address three main questions: (i) does the heterogeneity depend on duration of infection in cell culture, (ii) are there differences in the heterogeneity of the termini of S, M, and L segments and which of the termini, the 5' or the 3' termini, are more heterogeneous, and (iii) are there any differences in the termini of RNAs from pelleted virus particles and from infected cells?

To answer the first question, virions were pelleted from the growth media of Vero E6 cell cultures at 3, 6, 9, and 12 days p.i.; vRNA was extracted from the virions and the 5' and 3' termini were sequenced (Fig. 2). For the first time-point, 10 S, 9 M, and 10 L clones were sequenced. For the fourth time-point, 10 S, 9 M and 9 L clones were sequenced. No difference was detected in the degree of heterogeneity from 3 to 12 days p.i. The sequences of 6 S clones and 9 M clones from the two intermediate points of infection (6 and 9 days) were in agreement with this conclusion (data not shown).

To study whether there are differences regarding heterogeneity of the termini between the three genome segments and which of the two termini, the 5' or the 3', is more heterogeneous, 20 S, 18 M and 19 L clones originated from virus particles were sequenced (Fig. 2). The 5' termini of all three vRNA segments appeared to have the expected sequences. The 3' termini of M vRNAs were also highly preserved with only 1 of 18 clones having the residues A and U missing from the 3' terminus. In contrast, the 3' termini of the other two segments were less conserved: only 7 of 20 S clones and 3 of 19 L clones had the expected vRNA 3'-terminal sequences. Of 20 S clones, 9 had from 1 to 8 nt missing and 9 of 19 L clones had from 3 to 12 nt missing. A number of sequences were extended beyond the expected 3'-terminal point: 4 of 20 S clones and 6 of 19 L clones had from 1 to 3 nt of an extra AUC repeat. One of the L clones had an oligo(A)_1 stretch in the vRNA 3' terminus.

To address the question whether there are any differences in the termini of vRNAs from virus particles and from infected cells, vRNAs collected from cells at 12 days p.i. were compared to vRNAs of virus particles collected at the same time-point. Of 10 S clones from infected cells 7 had truncated 3' termini with 4 to 22 nucleotides missing. In comparison, only 2 of 10 S clones from pelleted virus had truncated 3' termini and there were only 1 to 4 nucleotides missing. One of the 9 clones from infected cells had an oligo(A)_1 stretch in the 3' terminus. All 7 L vRNA clones from infected cells had truncated 3' termini with 3 to 8 nt missing compared to 5 of 9 L vRNA clones from pelleted virus with truncations of 3 to 4 nt. The 3' termini of M vRNAs were equally conserved irrespective of the RNA source, and none of the 17 clones had truncations or extra nucleotides in the 3' terminus. The 5' termini of all three segments were intact, with the exception of two L clones from infected cells that had 1 or 6 nt missing from the 5' terminus.

The sequences of the 5' and 3' termini of antigenome RNAs from infected cells could not be determined because the RT–PCR reactions were not successful with antigenome templates.

### Discussion

**Completion of the TUL genome sequence**

The L segment of TUL is similar to but clearly distinct from other hantavirus L segments. Based on results of multiple sequence alignments, pairwise comparison and phylogenetic analysis, the L segment of TUL is most closely related to the L segments of the two strains of Puumala virus, also carried by an *Arvicolinae* rodent, followed by the L segments of Sin Nombre, Seoul and Hantaan viruses, carried by *Sigmodontinae* and *Murinae* rodents. This pattern is the same as for the S and M segments (Plyusnin et al., 1994, 1996a) and conforms with the established view that hantaviruses have been co-evolving with their natural hosts (Plyusnin et al., 1996b; Morzunov et al., 1998).

With the L genome segment and the termini of all three genome segments sequenced, TUL is now the first completely sequenced hantavirus that has not been linked to any human disease. There are no striking differences between the genome segments of TUL and the other four completely sequenced hantavirus types, which are all known human pathogens. If TUL turns out to be apathogenic to humans, it would be a safe model to study hantaviruses.

**Heterogeneity of the 3' termini of S and L genome RNAs**

Before the study presented here, the published terminal sequences of genome segments have either been deduced from PCR primers or have been obtained by either sequencing PCR products directly or determining the consensus sequence of multiple clones (Bowen et al., 1995; Chizhikov et al., 1995; Rakhov et al., 1995). Both approaches give only the most abundant sequence. The 5' and 3' termini of mRNAs have been studied in more detail. The first UAG repeat has been missing.

---

**Fig. 2.** The 5' and 3' termini of TUL genome segments expressed as RNA sequences. The number of clones having each sequence and the expected positions of the 5' and 3' termini are indicated. In cases where there was an AU doublet between G and C residues in the ligation region the exact ligation point could not be determined with certainty. A and U residues placed in the 5' termini that could belong to the 3' termini are in bold. Missing nucleotides are shown by dashes. (a) S segment, (b) M segment and (c) L segment. Data for the intermediate time-points are not included in the figure.
Fig. 3. Hypothetical prime-and-realign mechanisms to explain extended and truncated 3′ termini. Genome RNAs are in bold. Extra nucleotides are shown by underlining and missing nucleotides by dashes. Genome RNA is first used as a template in the synthesis of antigenome RNA. Normally the realignment of nascent RNA is suggested to occur 3 bases upstream of position 3 of the template (Garcin et al., 1995) as shown in the middle. If the realignment step occurs 6 bases upstream, as shown on the left, antigenome RNA with extra nucleotides in the 5′ terminus will be produced. When this, in turn, is used as a template in genome synthesis, a genome with extra nucleotides in the 3′ terminus will result. A model in which no realignment takes place is shown on the right. This will result in genome RNAs that have truncated 3′ termini. Truncated 3′ termini could also result from the initial priming step occurring downstream of position 3 on the template or the realignment step occurring downstream of the normal site.

from the virally templated part of some of the sequenced 5′ ends of mRNAs (Garcin et al., 1995; Hutchinson et al., 1996) and a minority of Hantaan virus genome S segment 5′ termini have had a few nucleotides missing starting with the purine A at position 2 or 5 from the 5′ terminus (Garcin et al., 1995).

The 5′ termini of TUL S, M and L vRNAs and the 3′ termini of the M vRNAs had the expected sequences, identical to those reported for other hantaviruses (Chizhikov et al., 1995). However, the 3′ termini of S and L RNAs turned out to be heterogeneous. Both deletions and extensions of the 3′ termini could be caused by the prime-and-realign mechanism, which has been proposed to initiate the transcription of mRNA and replication of genome RNA (Garcin et al., 1995). According to this model, RNA synthesis is initiated with a G residue aligned with a C residue at position 3 from the 3′ terminus. It is elongated for a few nucleotides followed by an upstream realignment of the nascent RNA and its final elongation. This mechanism has the potential to generate heterogeneity in the 5′ termini of genome and antigenome RNAs and the virally templated part of the 5′ termini of mRNAs (Fig. 3).

There are several possible explanations why the 3′ termini of genome RNAs are more heterogeneous than the 5′ termini.

(i) The above mentioned prime-and-realign mechanism could for some reason make more mistakes in the initiation of antigenome RNA synthesis than genome RNA synthesis which would lead to heterogeneity in the 5′ termini of antigenome RNAs that are, in turn, templates for the 3′ termini of genome RNAs.
(ii) There could be premature termination of genome RNA synthesis due to secondary structures in the 5’ terminus of the template or due to the 5’ terminus of the template being bound by nucleocapsid proteins or L protein

(iii) The 3’ termini could be degraded by RNases in infected cells before packaging of the virus

(iv) The heterogeneity of the 3’ termini could be an artifact of the experimental procedures used, for example RNA degradation during RNA extraction before the 5’→3’ RNA ligation

The fourth possibility seems unlikely because the 3’ termini of the M RNAs were intact. Before reverse transcription, the S, M and L RNAs were in the same reaction mixture and the terminal regions of the segments were always reverse transcribed using the same RNA ligation mixture as template for each segment. RNases should, therefore, have degraded the M RNAs as well. Furthermore, RNA degradation could not explain the extra nucleotides found in the 3’ termini of S and L RNAs.

In experiments with members of other genera of the family Bunyaviridae, Bunyamwera virus (Dunn et al., 1995) and Rift Valley fever virus (Prehaud et al., 1997), genome-like RNAs with truncated 3’ termini have not efficiently templated RNA synthesis. Deletion of at least five 3’ terminal nucleotides has abolished reporter gene activity in both transcription systems. Therefore, it is likely that the TUL genome RNAs that have truncated 3’ termini will not be used as templates in RNA synthesis.

Truncated L RNAs have also been described in persistent infections of Vero E6 cells with the Seoul hantavirus, strain SR-11 (Meyer & Schmaljohann, 1997). Moreover, heterogeneous termini of all three RNAs have been found in bank voles (Clethrionomys glareolus) infected with Puumala virus (our unpublished results), indicating that this phenomenon is not restricted to viruses growing in cell culture. Truncated 3’ termini could contribute to the low virus yields observed with TUL and other hantaviruses.

Ms Leena Kostamovaara, Ms Raija Leveelahti, Ms Tytti Manni and Ms Anja Virtanen are acknowledged for excellent technical assistance. This work was supported by the Medical Research Council of the Academy of Finland and the Sigrid Jusélius Foundation, Helsinki.

References


Received 8 May 1998; Accepted 17 July 1998