Toscana virus cap-snatching and initiation of transcription

Abdennour Amroun, Stéphane Priet and Gilles Querat*

Abstract

Toscana virus (TOSV) is an arthropod-borne phlebovirus within the family Phenuiviridae in the order Bunyavirales. It seems to be an important agent of human meningoencephalitis in the warm season in the Mediterranean area. Because the polymerase of Bunyavirales lacks a capping activity, it cleaves short-capped RNA leaders derived from the host cell, and uses them to initiate viral mRNA synthesis. To determine the size and nucleotide composition of the host-derived RNA leaders, and to elucidate the first steps of TOSV transcription initiation, we performed a high-throughput sequencing of the 5’ end of TOSV mRNAs in infected cells at different times post-infection. Our results indicated that the viral polymerase cleaved the host-capped RNA leaders within a window of 11–16 nucleotides. A single population of cellular mRNAs could be cleaved at different sites to prime the synthesis of several viral mRNA species. The majority of the mRNA resulted from direct priming, but we observed mRNAs resulting from several rounds of prime-and-realign events. Our data suggest that the different rounds of the prime-and-realign mechanism result from the blocking of the template strand in a static position in the active site, leading to the slippage of the nascent strand by two nucleotides when the growing duplex is sorted out from the active site. To minimize this rate-limiting step, TOSV polymerase cleaves preferentially capped RNA leaders after GC, so as to greatly reduce the number of cycles of priming and realignment, and facilitate the separation of the growing duplex.

INTRODUCTION

The Bunyavirales is the largest and most diverse order of pathogenic negative single-stranded RNA viruses. Present on five continents, it is divided into nine families, which infect animals, arthropods and plants. Toscana virus (TOSV) is an arthropod-borne virus that belongs to the genus Phlebovirus in the family Phenuiviridae [1]. TOSV was isolated in 1971 from collected Phlebotomus perniciosus in Grosseto province in central Italy, and then from Phlebotomus perfiliewi and Sergentomyia minuta sandflies [2–5]. TOSV is considered to be an emerging pathogen that frequently infects vertebrate and seems to be an important agent of human meningoencephalitis in the warm season in endemic regions [6]. Like all members of the genus Phlebovirus, TOSV possesses a negative-strand RNA genome divided into three segments: S, M and L for small, medium and large segments, respectively [7]. The S segment encodes two proteins in opposite orientations: the nucleoprotein N and the nonstructural protein NSs [8]. The M segment encodes the glycoprotein precursor in a single ORF [9] that will mature into the viral glycoproteins Gn and Gc, and a second non-structural protein, the NSm [10, 11]. The L segment encodes the viral RNA-dependent RNA polymerase (RdRp) [12], which is in charge of the replication and transcription of the viral RNA in the cytoplasm of the infected cell (reviewed in [13]).

All eukaryotic mRNAs contain a 7-methylguanosine cap structure (\(^{5'}\)mGppp) at their 5’ extremity. Messenger RNA capping is an evolutionarily conserved mechanism that occurs in the nucleus and represents the first modification of RNA polymerase II-transcribed mRNAs [14, 15]. The cap structure is directly implicated in protein translation, and is also involved in nuclear export, protection of mRNA from 5’ to 3’ exonuclease, polyadenylation, splicing and the looping of mRNA during translation (reviewed in [16]). All viruses depend entirely on the cellular translational machinery to produce their own proteins. However, many viruses are not able to use the mRNA-capping machinery of the cell, and so have implemented other strategies to cap their mRNA. The polymerase of Bunyavirales bypasses the lack of virally encoded capping activity using a mechanism called cap snatching. This capping strategy was first described for influenza virus [17] and is actually restricted to segmented negative-strand RNA viruses, including members of the...
order Bunyavirales and the families Orthomyxoviridae and Arenaviridae [18–21]. This mechanism involves the cleavage of short 5’-capped RNA leaders derived from the host mRNAs by the endonuclease activity of the viral RdRp [22–24]. Then, the polymerase domain of the viral RdRp uses them to prime the viral mRNA synthesis, leading to chimeric host–viral mRNAs at their 5′ extremities [24]. The host-derived caps are generally cleaved at 10–20 nucleotides downstream from the 5′ cap [25–35]. However, members of Nairoviridae [36] and Arenaviridae [19] use a relatively short host-derived cap.

Studies of the cap-snatching mechanism have revealed some heterogeneity in the repeated sequence downstream from the cap. This heterogeneity ranges from truncated to additional repeats when compared to the 3′ viral template [25–29, 36]. Truncated sequences result from direct elongation after internal priming, while partial reiterations of sequence result from the so-called prime-and-realign mechanism, which occurs during the initiation of transcription [29, 33, 34, 37]. This mechanism was first proposed for Hantaan virus [27], an orthohantavirus from the family Hantaviridae [1], and explains how both transcription and replication start with a host-capped RNA leader ending with a 3′ G residue and a single GTP, respectively. In this model, the host cap-derived primer is positioned facing the 3′ end of the template with a single base pairing, and after the addition of few nucleotides the small nascent chain is realigned backwards by a slippage of three nucleotides, allowed by the presence of terminal sequence repeats. Thereafter, a processive elongation of the mRNA takes place. The exact reason for the occurrence of prime and realign in mRNA synthesis remains unclear. Indeed, the proportion of mRNA synthesized with or without a prime-and-realignment event varies from one virus to another; it is employed a great deal in some viruses [29] and very little in others [28, 34, 38].

A study performed on TOSV 25 years ago revealed the presence of approximately 9–15 nucleotides (nt) at the 5′-end of N and NSs mRNAs that did not belong to the viral genome [39]. Here, we investigated the host-derived sequences present at the 5′ end of TOSV mRNA during human embryonic kidney 293 (HEK) cell infection using modern tools. To evaluate the size distribution, nucleotide composition and frequency of the prime-and-realign mechanism, we performed a high-throughput sequencing of N, NSs and GP mRNAs (for the glycoprotein precursor resulting from the transcription of the M segment). Our results showed that the RdRp cleaves the host RNA leaders within a window of 11–16 nucleotides preferentially after GC. The cleavage of the host RNA leader can occur at different sites to support the base pairing with the RNA template. The polymerase can use the same host-derived cap to prime any viral mRNA species. Our data suggested that 73% of mRNA was synthesized from direct initiation, whereas 27% appeared to result from prime-and-realignment events.

RESULTS AND DISCUSSION

Host RNA leader size

To determine the precise size of the host-snatched caps, an external primer was ligated to enzymatically decapped viral mRNAs and the 5′-ends of viral mRNAs were specifically amplified and analyzed using deep-sequencing technology. We analyzed and quantified the sequences containing both the ligated external primer and the viral sequence from the most frequent viral mRNA species: the N mRNAs at 7, 24 and 72 h post-infection (p.i.), the NSs mRNAs at 24 and 72 h p.i., and the GP mRNAs at 24 h p.i., separated by heterogeneous host-derived sequences (Fig. 1a and Table 1). Out of more than 8, 4×10⁵ validated reads, we found 4192 unique 5′ end mRNA extremities, with 3508 unique cell-derived cap sequences. A large number of different individual heterogeneous host-derived sequences were characterized, ranging from 96 unique sequences for the N mRNA at 7 h p.i. to 1482 different sequences for the N mRNA at 24 h p.i. (Table 1). The size and sequences of the 5′ ends of viral mRNA downstream of the host-derived segment were also found to vary (Fig. 1a), resulting in slightly more heterogeneous 5′ ends of viral mRNA (107 different N mRNA 5′-ends at 7 h p.i. and 1636 different N mRNA 5′-ends at 24 h p.i.). Looking at the size distribution for the host-snatched caps, there was a marked peak at around 80% of the heterogeneous sequence, 11 to 16 nucleotides in size (Fig. 1b, c). Less than 2.5% of the cap extensions had a size between 1–8 nt. They were excluded because they might have resulted from amplifications of degraded RNAs. On the other side of the distribution range, 4.81% of the heterogeneous sequences shared a size between 26 nt and 100 nt. Since the presence of such long host-derived caps has never been described, they were also excluded because we wanted to rule out the possible presence of artefactual cellular mRNAs or DNA fragment ligation products in our analysis. There was a restricted distribution for the size of the host-derived capped mRNAs that were added to the 5′ end of Toscana virus mRNAs, with most of the additions being 12–15 nt (Fig. 1c).

Host RNA leader cleavage site

The viral mRNA 5′-ends were highly variable in size and sequence and exhibited additions or truncations of the CA/AC repeat when compared to the viral template (Fig. 1a). These additions are believed to result from prime-and-realign mechanisms [27] that were also observed for many other segmented negative-strand RNA viruses. The cleavage of the host-capped RNA leaders is performed in a way that ensures base pairing complementarity with the viral template, as proposed for influenza virus [32, 40–43] and tenuiviruses [44]. Other works that studied Tomato spotted wilt virus (Orthotospovirus) suggested that the cleavage of the host-capped RNA leader seems to occur further downstream from the heterogeneous sequence, and at least one base pairing is required between the host-capped leader and the template for transcription initiation [34, 38]. Five conserved nucleotides (UGUGU) are present in all 3′
Fig. 1. Host-derived 5’-end of Toscana virus mRNAs are mostly 11 to 16 nucleotides in size. (a) Typical organization of sequenced N, NSs and GP viral mRNA 5’-ends. The external primer was ligated to the 5’ end of the total extracted RNA after decapping. All reads, starting with the external primer (blue box), are followed by the host-snatched mRNA leader (orange box) and the most frequent 5’ viral extremities (boxed in yellow). Only the first nine nucleotides belonging to the virus are represented. Nucleotides that match the...
viral template are in blue. Added nucleotides are in red. Dashes represent gaps. Truncated nucleotides are represented by a blue gap. The nucleotide R at position 6 is A for L, GP and NSs mRNAs, and G for N mRNAs. (b) Histogram representation of the mean frequency of unique host-derived cap sizes for all Toscana-derived mRNAs at all time-points, post-infection. The percentage of each population is indicated and the error bars are represented. (c) Size distribution of the heterogeneous sequences derived from HEK-infected cells at 24 h post-infection. Each graph shows the size distribution (in per cent) for all unique caps for N, NSs and GP mRNAs.

extremities of TOSV genomes that serve as templates for the initiation of transcription. Watson–Crick base pairing requires the presence of either A or C at the 3’ end of each host RNA leader to prime transcription. Therefore, we considered that the first C in CA or A in AC repeats, respectively, must belong to the cell-derived cap. In return, this allowed us to pinpoint the last nucleotide from the cell-derived cap primer. Analysis of the host-derived cap sequences (Fig. 2a–f) revealed that no preference was observed for an 8 nucleotide-long consensus sequence in the 5’ end of the snatched-cap and that about 81 % of the host-derived caps finished by C, regardless of the mRNA species (Fig. 2g, h). These data revealed that the polymerase of TOSV cleaved the host RNA leader preferentially after GC (53 %) or UC (19 %) (Fig. 2h).

We also observed that unique cellular mRNA could provide several different cap sizes at different times post-infection and for several viral mRNA species. Indeed, about 18 % of the caps derived from the same cellular mRNA, but were cleaved at different sites (Fig. 3), suggesting the possibility that (i) the endonuclease preferentially targets some specific cleavage sites, and consequently favours some mRNA 5’ end over others, or (ii) the endonuclease is able to reuse some of its own transcripts. The alignments of the shorter caps relative to the longer ones allowed us to identify the sequence motifs of some of the cleavage sites in the longer caps. These cleavage sites allowed us to identify at least two downstream nucleotides that could not belong to viral sequences. The analysis of two nucleotides upstream and two nucleotides downstream of the cleavage sites resulted in the identification of five preferential cleavage motifs. GC-GC (23.05 %) are the most likely motifs to induce the cleavage by the viral endonuclease, followed by GC-UC (8.55 %). However, 32 % of the identified cleavage sites occurred at the (AC)n and (CA)n motifs. In these cases, the cleavage can happen anywhere between the first and after the last nucleotide of each motif, which makes the exact identification of the last cleaved nucleotide impossible. The presence of this unidentified cleavage site fits with the possibility of multiple base pairing during the initiation of transcription. Although we could not exclude the possibility that caps being cleaved at different sites might be the result of recycling of the same cap by the viral endonuclease, we do not favour this explanation, because this phenomenon seems to be minor (it concerns no more than 18 % of sequenced caps), and for some additional reasons we discuss below.

**Most of the caps are taken once for a viral mRNA**

Sequencing of the cap-snatching repertoire of TOSV revealed the presence of 140 unique cap primers, presented in two or more Toscana virus mRNA species at different time points [Fig. 3 and Table S1 (available in the online Supplementary Material)]. Looking at the 24 h p.i. time point, there were 1482 unique caps for N mRNAs and 1478 unique caps of NSs mRNAs, 112 of which were held in common between N and NSs mRNA (Table S1). Regarding the 271 unique caps of GP mRNA, 5 were found to be held in common between the GP and N mRNA species. Overall, these data suggest that in the majority of the cases, the mRNA cap differs in the three viral mRNA species, although the cap-snatching is sometimes able to target the same pool of cellular mRNAs, regardless of the Toscana mRNA.

We looked at whether the unique cap leaders present at 7 h and 24 h p.i. might be over-represented at late time points of infection (72 h), but we found very few caps that were common to both the early and late time points (Table S1). Moreover, the profiles of AC/CA repeats are similar at 24 and 72 h p.i. (data not shown). These data do not support a model in which viral mRNA ends can be recycled to initiate transcription at later times in the infection. Indeed, even if recycling cannot not be excluded, this phenomenon seems to be very limited. It is noteworthy that the influenza polymerase was also reported to avoid targeting its own mRNAs [45, 46].

**Direct extension of cap leader is preferred over prime-and-realign events**

We found many identical host-derived cap sequences, but with various numbers of viral-derived AC/CA repeats downstream of the capped RNA leader (Table 1). Indeed, the same host RNA leader was able to generate several different rounds of prime-and-realign initiations, either on the same viral mRNA or for any other viral mRNA species (Fig. 3), suggesting that the number of rounds of prime-

| Table 1. Results for deep sequencing of the 5’ end of Toscana virus mRNAs |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Viral mRNAs type**       | **No. of reads** | **Unique viral mRNA 5’ end sequenced** | **Unique viral 5’ end (9–25 nt)** | **Unique host RNA leader end (9–25 nt)** |
| N 7 h                      | 15 978          | 113             | 107             | 96              |
| N 24 h                     | 183 183         | 1707            | 1636            | 1482            |
| N 72 h                     | 65 289          | 247             | 235             | 231             |
| NSs 24 h                   | 530 448         | 1691            | 1524            | 1478            |
| NSs 72 h                   | 21 994          | 143             | 122             | 120             |
| GP 24 h                    | 27 690          | 290             | 276             | 271             |
| Total                      | 844 582         | 4192            | 3900            | 3508            |
Fig. 2. Sequence analysis of the nucleotide composition of HEK-derived host cap primers at different times post-infection. Panels (a)–(f) shows a Logo representation of the relative frequency of the last nine nucleotides of the host cap leader for each position in the 5'→3' orientation using WebLogo 3.5.0 online software. Viral mRNA species and time post-infection are indicated. Numbers +1 represent the priming nucleotide. (g) Histograms of the frequency of the last two nucleotides, upstream of the hypothetical cleavage site. Cleavage after A is represented on the left of the panel, whereas cleavage after C is on right of the panel. (h) Pie chart representing the percentage of the two last nucleotides of all sequenced caps. Only caps in the size range 9–25 nucleotides were considered.
Fig. 3. Selected 5’ ends of Toscana virus N, NSs and GP mRNAs. Each panel is framed and represents the complete HEK-derived capped RNA sequences (in black), while each RNA leader is followed by the viral mRNA sequence, separated by a dash at different times post-infection (p.i.). 7mGppp represents the cap at the 5’ end of the HEK-derived RNA leader. Red nucleotides match with the viral template (in green). Additional nucleotides (in blue) are added during the prime-and-realign mechanism. The underlined
nucleotides belong to the host cap and mark the hypothetical cleavage site(s); in some cases, the presence of two or three underlined residues is due to the impossibility of determining the cleavage site with precision. Common caps are highlighted with the same colour (grey, yellow, green and blue). We assigned a score of 0 to the viral mRNAs with exact complementary at their 5′ ends relative to the viral template downstream of the host-derived cap (Fig. 4). We focused on scores ranging from −7 to +7, which include 99.95 % of sequenced mRNAs. The mRNAs with additional nucleotides (C and A) resulting from prime and realign events had positive scores corresponding to the number of added nucleotides. The sum of all positive scores gives the occurrence of all prime and realignments events, which is about 27 % of viral messenger RNAs. Conversely, negative scores were assigned according to the number of truncated nucleotides at the 3′ end of the viral sequence. They corresponded to internal initiations of the transcription, without any realignment. Together with score 0, mRNAs resulting from initiation at the first genomic nucleotide and direct initiation of transcription account for the majority of mRNA priming, i.e. 73 %.

Scores of −1 (48 %) and +1 (15 %), corresponding to the extensions of cap leaders ending with a C, represented the vast majority of viral ends. This suggested that the most frequent C priming required pairing with the first G of the viral genome and direct elongation without any realignment step (score −1, Fig. 5a). The second most frequent C priming event (+1 score) corresponded to pairing with this same first G in the 3′ end of the viral genome and one realignment step (Fig. 5a). Viral RNA ends exhibiting scores of −3 or less corresponded to internal priming.

Regarding the A priming (odd scores), we found similar results. The most frequent (12.7 %) event was direct priming on the first U of the viral genome with no realignment step (score 0, see also (Fig. 5a)).

In conclusion, studying the Toscana cap-snatching repertoire revealed that 73 % of mRNA resulted from direct initiation on the first two nucleotides of the genome. Analysis of the results of a previous cap-snatching study performed on the Uukuniemi virus in 1991, another phlebovirus [28], showed that among (only) 25 sequenced caps (from N and NSs mRNAs), 76 % of the host cap started with C and 40 % of caps also had −1 scores. Together, these results supported the idea that there could be a common mechanism of transcription initiation in phleboviruses.

Size limit of the template

A score of −4 seemed to be the limit for nucleotide truncations and might have resulted from the alignment of the last A of the host cap RNA with the U of the viral template, followed by direct elongation without realignment. These data suggested that only the first five conserved nucleotides in the 3′ end of each genome (3′ UUGUGU…) could serve as a template for the base pairing with the host cap, pointing to an intrinsic limiting mechanism. Indeed, all of the negative-strand RNA virus polymerase structures contained an element that partially or totally blocked the exit of the template–product duplex [47]. In La Crosse virus (LACV), an orthobunyavirus, the emerging templates clash with the C-terminal lid domains of the polymerase [48]. It is likely that a similar structure in TOSV polymerase (whose structure has not yet been resolved) also plays the role of a physical barrier limiting the number of nucleotides overshooting the active site at a maximum of five nucleotides, and consequently preserves the integrity of the genomic information (the AUG codons of GP and L mRNAs are at the seventeenth and eighteenth positions of their respective templates). Moreover, no repeated sequences that included the nucleotide Y0 or had a score of −5 were found. Indeed, the size distribution of the total length of the cellular leader and the viral repeats is very similar in shape to that of the cellular leader, but shifted by five nucleotides towards the 16–22 nucleotides (see Fig. 5). Hence, we concluded that the entry of the sixth nucleotide of the template (Y0) into the active site marks the arrest of prime-and-realign mechanism, while the RdRp switches to the elongation mode until it crosses a transcription termination signal.

Refined model of the initiation of transcription mechanism

The 3D structure of a LACV RdRp revealed a common overall architecture of RNA polymerases in negative-strand RNA viruses, and the presence of four channels that connect the internal cavity of the polymerase (including the active site) with its exterior environment: the template entry and exit channels, the nucleotide entry channel and the nascent strand (product) exit channel [48]. In its inactive state, the 3′ end of the template binds the polymerase in a specific binding site, but remains inaccessible to the template entry channel. However, to initiate RNA synthesis, the polymerase must perform conformational changes in order to relocate the 3′ end of the
template from the binding site to the active site, as observed for LACV [48]. Once relocated, the 3' end of the template of LACV overshoots the active site by 4–5 nucleotides, which allows the base pairing with the incoming host-capped primer during the initiation of transcription [48].

The cap-snatched host RNA leader will probably join the active site via the exit product channel [48, 49]. In the order Bunyavirales, the presence of partial sequence reiteration, which occurs during the initiation of RNA synthesis, is explained by the prime-and-realign mechanism, as first proposed for the Hantaan virus [27]. During the elongation step, the addition of nucleotides to the nascent strand is done in the active site cavity, and both of the nascent and the template strands form a duplex in the active site cavity [50] (see Fig. 5). After the addition of a few nucleotides and a complete double-helical turn, the emerging template strand clashes with the C-terminal lid domains, while the nascent strand clashes with the thumb of the polymerase [48]. This situation obliges the polymerase to force each strand to separate towards two distinct exit channels.

The majority of the TOSV mRNAs resulted from direct priming, but we observed mRNAs resulting from several rounds of prime-and-realign events. Thus, based on both these previous data and our own results, we formulate the hypothesis that the different rounds of prime-and-realign observed during the initiation of 27% of TOSV transcripts are probably the result of unsuccessful attempts by the polymerase to engage the 3' end of the template into its exit channel. The blocking of the template in a static position, and the unilateral movement of the nascent strand, will result in a slippage by two nucleotides of the host-capped primer, and induce an additional round of prime-and-realign that results in the addition of two supplementary nucleotides to the nascent chain, which can represent a rate-limiting step in viral mRNA synthesis. According to the number of template nucleotides overshooting the active site [1–5], the nature of the last base of the host-capped leader with an AC/CA repeat at the 3' end.

<table>
<thead>
<tr>
<th>Rounds of prime-and-realign mechanism</th>
<th>Ultimate/penultimate initiation</th>
<th>Prime and realign extension</th>
<th>Score</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td>47.76</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>7mGppp → C</td>
<td>-1</td>
<td>3.82</td>
</tr>
<tr>
<td>2</td>
<td>7mGppp → A</td>
<td></td>
<td>-2</td>
<td>7.76</td>
</tr>
<tr>
<td>3</td>
<td>7mGppp → C</td>
<td></td>
<td>-3</td>
<td>1.87</td>
</tr>
<tr>
<td>4</td>
<td>7mGppp → G</td>
<td></td>
<td>-4</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>7mGppp → R</td>
<td></td>
<td>-5</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>7mGppp → A</td>
<td></td>
<td>-6</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>7mGppp → C</td>
<td></td>
<td>-7</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Fig. 4. Toscana virus mRNAs result mainly from direct initiation on the first two nucleotides of the viral genome. The class of sequences most frequently appended at the 5' end of each Toscana virus transcript in HEK-infected cells. The score (number of additional or truncated nucleotides ranging from –7 to +7) is indicated at the right of each sequence. HEK-derived capped primers are represented by 7mGppp—A/C in black. Sequence matching with the viral template is represented in red, while added nucleotides from additional rounds of realignment are shown in blue. Only the first eight nucleotides belonging to the virus are represented. R residue represents A for L, GP and NSs mRNAs, and G for N mRNAs. The frequency of each score is indicated in the histogram. Scores equal to or less than 0 are probably due to the absence of a realignment step, while a score of more than 0 requires at least one round of realignment (indicated at the left). Up to: represents the number of rounds of prime-and-realign, including those resulting from a host-capped leader with an AC/CA repeat at the 3' end.
3’ termini of the S segment of Rift Valley fever virus (phlebovirus) showed that the 3’ U1 can be removed without affecting transcription, but deletion of the first 3’ repeat (UG) reduces transcription to 29.5%, while the deletion of three or more nucleotides (UGU) completely abolished the transcription [51]. The RdRp of Machupo virus (arenavirus) [52] and LACV [48, 53] binds the 3’ and 5’ extremities of the viral RNA in two separate binding sites, which raises questions about the panhandle structure. It is no coincidence that repetitive sequences are conserved at the 3’ termini of the genome of segmented negative-strand RNA viruses that use cap-snatching to initiate their transcription, such as Orthomyxoviridae [18], Arenaviridae [19] and members of the order Bunyavirales [20, 21, 24]. The deletion of this repetitive sequence resulted in the inability of this polymerase to initiate transcription.

However, our data show that 61% of sequenced viral mRNAs (scores from $/C_0^1$ to $/C_0^4$) lack at least one nucleotide at the 5’ end of the viral sequence. Similar truncated mRNAs were also reported for other negative-strand RNA viruses [28, 29, 44]. These data clearly show that an entire 5’

---

**Fig. 5.** Model of transcription initiation of Toscana virus: for all of the panels, a portion of the polymerase is represented schematically by a grey cartoon. In step 1 of (a), the template and the nucleotide (NTP) entry channels of the polymerase are indicated by incoming arrows, while the template and product exit channels are shown by outgoing arrows. The hypothetical location of the polymerase active site is shown in yellow inside the cavity of the polymerase, which is in blue (the 3D structure of TOSV polymerase has not yet been resolved). Only the first nine nucleotides of the 3’ end template are shown. The positions of the template nucleotides are indicated at the top of the polymerase. The Y at the sixth position (red) represents the difference between the four mRNA species of Toscana virus: in L, GP and NSs templates, there is a U, while an N template exhibits a C. The Y also marks the end of the UG repeat at the 3’ end of the template. Only the last nucleotide of the host-capped primer (7mGppp—A/C) is represented. Blue arrows indicate the movement of the different strands inside the polymerase. At the bottom of each panel, the scores represent the difference in size between the viral template and the viral nascent mRNA (highlighted in green). (a) Transcription initiation with C nucleotide at the penultimate nucleotide of the template. (b) Internal initiation with C nucleotide. For (a) in priming step (1), the last nucleotide of cap C is aligned with the template. In step (2), initial elongation by two nucleotides in the active site ensues. In step (3), the RNA duplex is successfully pressed outside of the active site, resulting in the movement of both the template and the small nascent mRNA. In this case, the engagement of the template strand toward its exit channel results in the entry of the Y in the active site, and signals the end of ACACA repeating. This allows the addition of two nucleotides in step (4) and the beginning of processive elongation by the polymerase in step (5), leading to a score of −1. If the polymerase fails to pull the template into its exit channel, steps (3 bis), this will lead to a slippage and the realignment of the nascent chain by two nucleotides; another (AC) repeat will then be added to the nascent strand, resulting in the increase of the score by +2. (b) internal initiation. The relocation of the 3’ template’s end in the polymerase cavity overshoots the active site by four nucleotides and directly engages the 3’ end of the template into the exit template channel, thus making the Y accessible in the active site. The last C of the cap is aligned with the second G of the template (1). The initial elongation in step (2) includes the Y residue and is followed by steps (3)–(5), resulting in a score of −3.
viral mRNA end is not required for the translation of viral protein.

**Role of the penultimate nucleotides of the host-capped RNA leader**

Our data revealed that in most cases the penultimate nucleotide of the host-capped RNA is a G (58% of cases) or a U (28%)(Fig. 2). In addition, the most frequent last dinucleotides of sequenced caps are GC (53% of cases)(Fig. 2h). To determine what advantage there might be for the virus to have a G in the penultimate nucleotide of the host-derived cap, we analysed the frequencies of the two last nucleotides for each score (Fig. 6a). Our results showed that 70% of the host caps ending with GC had a −1 score, regardless of the nature of the mRNA species. We also found that about 50% of the sequenced caps had a score of −1(Fig. 4), which may correspond to the best compromise for the RdRp to initiate the synthesis of its own mRNAs by direct elongation. Indeed, additional rounds of prime-and-realign can represent a rate-limiting step that requires resources and energy that the virus would otherwise retain.

Thus, one explanation for the preference observed for GC might be that it allows the formation of both G:U wobble and Watson–Crick base pairing with the 3′ viral terminal nucleotide (U) and penultimate (G) nucleotides, respectively. Consequently, it might promote the formation of the RNA duplex and the addition of the first nucleotides within the active site. Indeed, the thermodynamic stability of the G:U wobble is comparable to that of a Watson–Crick A/U

---

![Diagram](https://via.placeholder.com/150)

**Fig. 6.** Distribution of the two last nucleotides of the cap in each score. (a) Histogram bars represent the percentage of the different combinations of the last two nucleotides upstream of the hypothetical cleavage site of the cap. (b, c) Schematic representations of base pairing between the 3′ end of the template strand and nascent mRNA, with scores of −1 and 0, respectively. The host cap RNA leader is represented in black and the two last nucleotides are underlined. Nascent mRNA is in red, while the template RNA is in blue. Watson–Crick base pairing is represented by vertical lines. A mismatch between two bases is represented by a black dot, with it leading to a pre-open conformation that makes the separation of the two RNA strands easy. Y represents a pyrimidine nucleotide (U or C) and N represents any nucleotide.
pair, and it exhibits a strong affinity for metal ions that are involved in polymerase activities [54, 55]. A second nonexclusive explanation might be that a G:U wobble base pairing can stabilize the U1 of the viral template in a pre-open conformation (Fig. 6b), compared to mRNA with 0 score (Fig. 6c). This pre-open conformation might facilitate the engagement of the template strand towards its exit tunnel and make the separation of the two RNA strands easier. This as yet undescribed strategy, to the best of our knowledge, might well also be employed in Uukuniemi virus, after reanalysis of the results of the previous study by Simons and Pettersson [28]. Interestingly, the examination of another study of mixed infection, in which the reovirus rice ragged stunt virus was used as a cap donor for rice stripe virus and rice grassy stunt virus (plant-infecting viruses within the genus *Tenuivirus* that have 3'-UGUGU extremities) [44], which branch from an old common ancestor with phleboviruses [56], revealed many similarities to our results, despite a very different context. The most important of these is the fact that 21 out of 39 (53.8%) sequenced clones (5'-mRNA extremities of the two viruses) ending with GC led to a score of −1.

**Conclusion**

In this study, we sequenced 3678 different HEK-derived sequences that were present at the 5' extremities of TOSV N, NSs and GP mRNAs. The viral polymerase cleaves short cellular capped RNAs (11–16 nucleotides). Seventy-three per cent of TOSV mRNA results from direct initiation, whereas 27% requires prime and realignment events. The preference for cleavage after the GC motif could be the result of an adaptation to reduce the frequency of the prime-and-realign mechanism. Moreover, additional rounds of prime-and-realign can be also considered to be an initial adaptation to the internal structure of the polymerase, while being a rate-limiting step for the transcription of viral mRNAs. In other words, we may be witnessing a period in which TOSV is improving its replicative performance until the total disappearance of the prime-and-realign mechanism.

**METHODS**

**Virus propagation and cell culture**

Toscana virus T152 strain was isolated from *Phlebotomus larroussius* spp. collected in Tunisia [57] and provided by the European Virus Archive (ref: 1449). HEK cells were propagated in Dulbecco's modified Eagle's medium high glucose 4500 mg l⁻¹ (Gibco), supplemented with 7% foetal bovine serum (FBS), 1% penicillin/streptomycin and 1% glutamine at 37 °C with 5% CO₂. Infected cells were propagated in the same medium mixture, but with 2.5% FBS.

**RNA treatments**

Sub-confluent HEK cells in a 25 cm² flask were infected by TOSV at an m.o.i. of 8. At different times post-infection, cells were washed with 5 ml of HBSS and recovered before total RNAs were extracted using the NucleoSpin RNA II kit (Macherey-Nagel) according to the manufacturer’s protocol. Purified RNAs were dephosphorylated using FastAP thermosensitive alkaline phosphatase (Thermo Scientific) according to the manufacturer's instructions to convert all uncapped RNA (cellular and genomic TOSV RNA) into 5'-hydroxyl RNA. The total RNAs were then treated with tobacco acid pyrophosphatase from EpiCenter (ref: T19050) according to the manufacturer’s instructions in order to remove the 5' cap from viral and cellular mRNA, leaving 5' monophosphorylated mRNAs. A 5'-external primer (5'-GAATAAGGGCCACCGGAAATGT-3') was ligated to the 5'-end of the RNAs using T4 RNA ligase 1 (New England Biolabs).

**Next-generation sequencing of the 5’ end of the N, NS and GP mRNA of TOSV**

Ligation products were reverse -and amplified (20 cycles) using the SuperScriptIII One-Step RT-PCR Platinum Taq HiFi kit (Invitrogen) according to the manufacturer’s protocol, with the external primer and outer gene-specific primer (GSP): 5'-CCCCGCCTGTCTGAAAC-3', 5'-TAGCATCAA TAGTGGTCCACAGAC-3' and 5'-AATTGGTATAGGC TTCAATTTGT-3' for NSs, N and GP mRNAs, respectively. A second round of semi-nested amplification (30 cycles) was performed using Platinum PCR SuperMix High Fidelity (Invitrogen) with the external primer and inner GSP, resulting in −200 bp amplicons – 5'-AGGGTG TCCAAATCAGCTATGTAGAAC-3', 5'-TTGGGTCT AAATCCCTTGTAAGC-3' and 5'-GGCTCCCCAAGCT GGTTCATTG-3' for NSs, N and GP mRNAs, respectively. Sequencing of the RT-PCR product was performed using the Ion PGM sequencer (Ion PGM Sequencing 200 kit v2 and the Ion 316 Chip, Life Technologies) (Fig. S2). Automated read datasets provided by Torrent software suite 5.0.2 were trimmed according to quality score (99.9%) and size (reads shorter than 50 bp were removed) using CLC genomics workbench software (CLC bio-Qiagen). Reads were mapped on the Toscana virus T152 reference strain (GenBank accession numbers: JX867535 and JX867536 for the M and S segments, respectively) using CLC. The primers used for RT-PCR and the (CA)₉/(AC)₉, repeat, except that the first C or A, which we considered as belonging to the host cap segments, were removed using Excel software.

**Funding information**

This work was supported by the European Union Seventh Framework Programme (FP7/2007–2013) under a SILVER grant (no. HEALTH-F3-2010-260644). A.A. was supported by SILVER.

**Acknowledgements**

We thank Geraldine Piorkowski and Karine Barthelemy for their help with NGS sequencing.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**References**

A unique cap (m7GpppXm)-dependent influenza virion endonuclease cleaves USA transcription of influenza viral RNA in vitro.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.

The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology 1993;197:808–812.


Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.