JOURNAL OF VIROLOGY, Dec. 1996, p. 8361–8367 Vol. 70, No. 12
0022-538X/96/$04.00+0

Nucleoprotein Viral RNA and mRNA of Thogoto Virus: a Novel "Cap-Stealing" Mechanism in Tick-Borne Orthomyxoviruses?

FRIEDEMANN WEBER, OTTO HALLER,* AND GEORG KOCHS

Abteilung Virologie, Institut für Medizinische Mikrobiologie und Hygiene, Universität Freiburg, D-79008 Freiburg, Germany

Received 26 June 1996/Accepted 15 August 1996

Tick-borne Thogoto virus (THOV) represents the prototype virus of a new genus in the *Orthomyxoviridae* family. Its genome consists of six segments of negative-sense, single-stranded RNA. We have cloned and sequenced the fifth genomic segment, which codes for the viral nucleoprotein (NP). The deduced amino acid sequence shows 43% similarity to the NP of Dhori virus, a related tick-transmitted orthomyxovirus, and about 14% sequence similarity to those of the influenza viruses. To reveal the mechanism by which THOV initiates mRNA synthesis, we characterized the 5′ ends of the NP mRNAs. Transcripts were recognized by a cap-specific monoclonal antibody, indicating that THOV mRNAs are capped. Surprisingly, no large heterogeneous extensions were found at the 5′ end, as would have been expected if THOV were using a classical "cap-stealing" mechanism. We therefore propose that THOV is stealing only the cap structure with one or two additional nucleotides from cellular mRNAs to generate appropriate primers for initiation of viral mRNA transcription.

Thogoto virus (THOV) was originally isolated from ticks and is now recognized to constitute a new genus in the family *Orthomyxoviridae*, dubbed Thogoto-like viruses (38). Like influenza virus, THOV is sensitive to the antiviral action of the interferon-induced mouse Mx1 protein (20) that accumulates in the host cell nucleus (13), suggesting that THOV replication has a nuclear phase. Unlike influenza virus, the THOV genome consists of only six distinct RNA segments (48). The three largest segments code for subunits of the viral polymerase complex: RNA segment 1 encodes a PB2-like protein (49a), segment 2 corresponds to PB1 of influenza virus (32a), and segment 3 codes for a PA-like protein (48). THOV contains a single surface glycoprotein encoded by RNA segment 4, as does influenza C virus. Interestingly, this glycoprotein is not related to the influenza virus surface glycoproteins but shows similarity to the glycoprotein of baculoviruses (37).

The nucleoprotein (NP) is the major structural protein that interacts with the RNA segments to form the nucleocapsids. It represents the main type-specific antigen and is used to distinguish between the different genera within the *Orthomyxoviridae* family (31). NP sequences have been widely used to study the evolution of influenza viruses (50), but sequence data of the prototype tick-transmitted orthomyxovirus have not been available so far. Here, we report the cloning and sequencing of the NP gene of THOV and present a complete phylogeny of the *Orthomyxoviridae* family.

Influenza viruses require host cell nuclear functions for at least two critical steps in their multiplication cycle: mRNA splicing and a unique process called cap stealing (29, 32). mRNA synthesis is hereby initiated with 10 to 15-nucleotide-long capped primers that are cleaved from a subset of host cell RNAs by a virally encoded endonuclease, PB2 (49). Initiation of transcription with these primers leads to the characteristic 5′ ends of influenza virus mRNAs that are heterogeneous in sequence (7, 12, 30). A feature common to all influenza viruses is their sensitivity to actinomycin D and α-amanitin (35, 44), because these inhibitors of cellular RNA polymerase II lead to a shortage of capped cellular RNAs in the nucleus (29). Sensitivity to actinomycin D and α-amanitin suggested that Thogoto-like viruses use the same mechanism for initiation of viral mRNA synthesis as influenza viruses (47). We therefore expected that THOV transcripts would similarly possess heterogeneous 5′ extensions consisting of 10 to 15 extra nucleotides. We show here that this is not the case. The 5′ ends of THOV mRNAs are capped but are, in most cases, exact copies of the 3′ end of the viral RNA (vRNA) template. We propose that Thogoto-like viruses use a novel cap-stealing mechanism to generate their mRNAs.

MATERIALS AND METHODS

Isolation of RNA. For isolation of viral mRNA, monolayers of BHK-21 (baby hamster kidney) cells were infected at an input multiplicity of infection of 10 with the Sicilian isolate (SiAr 126) of THOV (1) for 15 h. For isolation of viral genomic RNA, cells were infected with THOV at an input multiplicity of 0.05 and incubated for 48 h. Sedimented virus from the supernatant was purified in an Urographin 370 (Schering AG Pharmaceuticals, Berlin, Germany) gradient and finally pelleted through a 30% (vol/vol) glycerol cushion (37). Total RNA and vRNA were extracted by the acid phenol method (8), and poly(A)$^+$ RNA was enriched by oligo(dT)-cellulose chromatography (4).

Cloning of the nucleoprotein gene. Double-stranded cDNA was synthesized from mRNA of THOV-infected cells by using a Time-Saver cDNA synthesis kit (Pharmacia, Freiburg, Germany) as instructed by the manufacturer. Briefly, first-strand cDNA synthesis was performed by using murine reverse transcriptase with oligo(dT)$_{12–18}$ primer and second-strand synthesis by RNase H nicking and replacement synthesis with DNA polymerase I (19). The blunt-ended cDNA was ligated to an adapter containing an *Eco*RI overhang and an internal *Not*I site. After phosphorylation with T4 polynucleotide kinase, the cDNA was inserted into the *Eco*RI-digested and dephosphorylated ZAP Express lambda vector (Stratagene, Heidelberg, Germany) to construct a *lacZ* fusion. Lambda phage particles were made using the Gigapack II packaging extract and plated onto the *Escherichia coli* host strain XL1-Blue MRF′ (Stratagene). The lambda library was titrated, and the amount of recombinants was determined by blue/white color screening. About 75,000 recombinant plaques were lifted onto isopropyl-β-D-thiogalactopyranoside-drained nitrocellulose filters and treated by the standard protocol for immunoscreening of fusion proteins (4). Positive phages were identified by using mouse monoclonal antibody (MAb) 2 specific for the THOV NP (a gift from Patricia A. Nuttall, NERC Institute of Virology and Environmental Microbiology, Oxford, England) (43) with the ECL nonradioactive detection system (Amersham Life Science, Buckinghamshire, England) and purified by two further rounds of immunoscreening. The fragments cloned into the ZAP Express vector were in vivo excised with the ExAssist helper phage and *E. coli* XLOLR (Stratagene), resulting in chimeric plasmids (pBK-L1 to pBK-L10) with a prokaryotic *lacZ* promoter and a eukaryotic cytomegalovirus immediate-early promoter. Clone pBK-L3 was digested with *Eco*RI, and the insert was labelled radioactively by nick translation (Boehringer, Mannheim, Germany). Genomic RNA (0.5 μg) from purified virus preparations was separated by

* Corresponding author. Phone: 49-761-2036534. Fax: 49-761-2036626. Electronic mail address: haller@sun1.uk1.uni-freiburg.de.

electrophoresis on a 1.2% agarose gel containing 3.7% formaldehyde. Northern (RNA) blotting and hybridization were performed as described previously (4).

Standard RT-PCR. Reverse transcription (RT)-PCR was performed by a two-step protocol (26). Briefly, a 20-μl reaction volume contained the RNA template with 1 mM each deoxynucleoside triphosphate (dNTP), 20 U of RNasin, 10 mM dithiothreitol, 4 μl of 5× RT buffer (375 mM KCl, 250 mM Tris-HCl [pH 8.3], 15 mM $MgCl_2$), 80 ng of random hexanucleotides, and 200 U of Superscript reverse transcriptase (Gibco/BRL, Berlin, Germany). PCR was performed with 1 μl of the RT reaction mixture in 100 μl of 1× PCR buffer (10 mM Tris-HCl [pH 8.3], 1.5 mM $MgCl_2$, 50 mM KCl) with 0.1 mM each dNTP, 0.5 μM each upstream and downstream primers, and 2.5 U of *Taq* polymerase (Boehringer), overlaid with mineral oil.

Cloning of the RNA 5′ and 3′ ends. The 5′ end of genomic segment 5 of THOV was cloned by procedures described in the 5′ RACE system (Gibco/BRL) for rapid amplification of cDNA ends (15). Briefly, RT was performed by using an oligonucleotide corresponding to nucleotides 734 to 753 of segment 5 (5′ TGCTGCCTCTCTGCTGGCAA 3′). After removal of the RNA template by RNase H treatment, the first-strand cDNA was poly(dC) tailed with terminal deoxynucleotidyltransferase. PCR amplification of the vRNA 5′ end was performed by using a poly(dG) anchor primer with different restriction sites and an oligonucleotide corresponding to nucleotides 969 to 988 of segment 5 (5′ GAT CTGGGAGTTTTAGAGTG 3′). Following heating at 94°C for 5 min, the reaction mixture was cycled 30 times with 94°C for 30 s, 53°C for 1 min, and 72°C for 1 min.

The 3′ end of genomic segment 5 was cloned by intramolecular self-ligation of viral RNA, using T4 RNA ligase and subsequent RT-PCR across the junction of the 5′ and 3′ ends (34). Briefly, 1 μg of purified viral RNA was incubated overnight at 20°C in a volume of 200 μl of ligation buffer (50 mM Tris-HCl [pH 7.8], 10 mM $MgCl_2$, 1 mM β-mercaptoethanol, 1 mM ATP, 10 mM dithiothreitol, 100 U of RNasin) with 100 U of T4 RNA ligase (New England Biolabs, Schwalbach, Germany). The reaction product was purified by phenol-chloroform extraction and ethanol precipitation. RT was performed as described above with random hexanucleotide primers. The primers for the subsequent PCR were identical to nucleotides 1224 to 1243 of segment 5 (5′ TCTGGGTCTCAAAC CGTCAG 3′) and complementary to nucleotides 344 to 363 (5′ ACCATTGCC CCTTGTATTTC 3′). The PCR cycling conditions were the same as described above.

To determine the 5′ ends of the NP mRNAs, two different approaches were used. First, a 5′ RACE (Gibco/BRL) analysis was performed as described for the vRNA, using poly(A)$^+$ RNA and an oligonucleotide complementary to nucleotides 631 to 650 of the viral NP mRNA (5′ GTGCCCGACGATGCTCTTCC 3′) for RT. After RNase H digestion, the first-strand cDNA was dC or dG tailed with terminal deoxynucleotidyltransferase. To amplify the 5′ end by PCR, a poly(dG) or poly(dC) primer was used together with an oligonucleotide complementary to nucleotides 413 to 435 of the NP mRNA (5′ TGACGGGGTATTTCTGCCAC AAG 3′). The reaction mixture was heated for 5 min at 94°C and was then cycled 30 times for 30 s at 94°C, 1 min at 57°C, and 1 min at 72°C. In a second approach, we determined the sequence of the 5′-3′ end junction after circularization of the decapped mRNA and subsequent amplification of the ligated junction by RT-PCR as described by Mandl et al. (34). Briefly, the isolated poly(A)$^+$ RNA was treated with tobacco acid pyrophosphatase (Epicentre Technologies, Madison, Wis.) and intramolecularly ligated with T4 RNA ligase (New England Biolabs). The circularized product was then reverse transcribed by using a primer complementary to nucleotides 796 to 815 of the NP mRNA (5′ ATCCTCGCAGT TGGCTATCA 3′). PCR of the 5′-3′ junction was performed with the same primer pair and PCR conditions as described for the vRNA circularization. The amplified products were cloned into the pCRII vector and transformed into One Shot bacterial cells as instructed by the manufacturer (Invitrogen, Leek, The Netherlands).

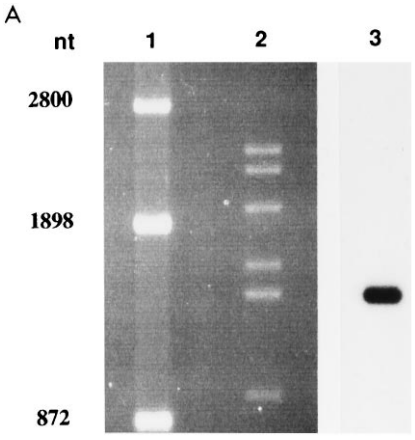
Sequence analysis. Sequencing was accomplished by the dideoxy method (45) on double-stranded plasmid DNA, using a T7 sequencing kit (Pharmacia) and different primers for both strands. Nucleotide sequence data were stored and edited with an Apple Power Macintosh 6100/60 and the LASERGENE Biocomputing software package (DNAStar Inc., Madison, Wis.). Screening of the Swiss-Prot protein database for similarities with the predicted amino acid sequence of segment 5 of THOV was conducted by using the BLASTP (3) program included in the HUSAR software package operated by the German Cancer Research Center, Heidelberg. The amino acid sequence data were analyzed by the Protean and MegAlign programs from the LASERGENE program package.

Immunoprecipitation of the cap structure. Poly(A)$^+$ RNAs from THOV-infected BHK-21 cells were incubated with a 5′ radiolabelled oligonucleotide primer complementary to nucleotides 37 to 54 of the THOV NP mRNA (5′ GAGGGCCCGAGATGTCCA 3′). RNAs were reverse transcribed with 400 U of Superscript reverse transcriptase in 40 μl of RT buffer (Gibco/BRL) according to standard protocols (4). The resulting RNA-cDNA hybrids were then immunoprecipitated with the cap-specific mouse MAb H-20 (a gift from Reinhard Lührmann, Institut für Molekularbiologie und Tumorforschung, Marburg, Germany) (5) bound to protein G-Sepharose beads (Pharmacia) as described by Garcin and Kolakowsky (17). Finally, the immunoselected labelled cDNA was separated on a 8% polyacrylamide sequencing gel next to the products of a sequencing reaction of the cloned 5′-3′ junction of circularized THOV segment 5 mRNA (using the same radiolabelled oligonucleotide). An unrelated MAb (MAb 2) (43) was used as a control.

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under accession number X96872.

RESULTS

RNA segment 5 of THOV encodes the nucleoprotein gene. To clone the NP gene, we constructed a cDNA expression library from poly(A)$^+$ RNA of cells infected with the Sicilian THOV isolate SiAr 126 (1). Ten positive clones of the recombinant lambda phage library were identified with MAbs detecting NP of THOV. After two additional rounds of immunoscreening, the recombinant lambda clones were converted into the chimeric plasmids pBK-L1 to pBK-L10 by in vivo excision. Seven clones contained a cDNA insert of the expected size. A Northern blot analysis with virion RNA revealed that the radiolabelled insert of pBK-L3 hybridized to segment 5 (Fig. 1A), indicating that this segment codes for NP. The cDNA insert of pBK-L3 was then sequenced by a primer-walking strategy. To determine the extreme 5′-end sequences of the vRNA template (which are not found on mRNA transcripts), a 5′ RACE analysis of genomic RNA was performed. In addition, the 3′ end of the genome segment was characterized by circularization of vRNA and amplification by RT-PCR across the 5′-3′ junction. Figure 1B shows that short untranslated regions of 20 and 33 nucleotides are present at the 3′ and the 5′ ends, respectively, of segment 5. Comparison with the cDNA sequence of pBK-L3 revealed that a stretch of 6 U's at



B

3′ UCGUUUUUGUCCGUCAGUUU ORF AAAAACAAAGGUUUUUUGACGGAACUAAAGAGA 5′

FIG. 1. vRNA segment 5 of THOV. (A) The cloned cDNA corresponds to vRNA segment 5. Genomic RNA from purified THOV virions was separated on a 1.2% agarose gel containing 3.7% formaldehyde and ethidium bromide. RNA markers (lane 1) and the six vRNA segments of THOV (lane 2) were detected under UV light. Blotted RNA was probed with the radiolabelled cDNA insert of plasmid pBK-L3 (lane 3). nt, nucleotides. (B) The noncoding 3′ and 5′ end sequences of THOV segment 5. The box indicates the putative ORF which is in antisense orientation, and the polyadenylation signal is underlined.

the 5′ end (positions 1397 to 1402) comprises the polyadenylation signal.

The complete RNA segment 5 has a length of 1,418 nucleotides and contains a single open reading frame (ORF) of 1,362 nucleotides that starts with an AUG at positions 21 to 23. This AUG is located within a context known to be favorable for initiation of protein synthesis (28). The long ORF is terminated by an UGA codon at nucleotides 1383 to 1385. No additional reading frames longer than 150 amino acids were detected. The small ORFs found are unlikely to be expressed since no subgenomic mRNAs related to segment 5 were detectable by Northern blotting (data not shown). The single long ORF has a coding capacity for 454 amino acids, giving rise to a polypeptide with a calculated molecular weight of 51,850. In vitro transcription and translation of the cDNA resulted in a product of the expected size (data not shown). The amino acid composition showed that the predicted protein is rich in lysine (8.59%) and arginine (4.85%). These charged and basic amino acids contribute to the net charge of +15.6 at neutral pH and to an isoelectric point of 9.0. No large clusters of charged residues were present. The basic amino acid residues were evenly distributed along the length of the polypeptide, as previously described for NPs of other negative-stranded RNA viruses, including influenza A virus (51) and Dhori virus (DHOV) (16). None of the known RNA-binding motifs (6) were found, but recombinant THOV NP exhibited RNA-binding capacity (data not shown).

To determine the subcellular localization of the expressed gene product, we transiently transfected Vero cells with clone pBK-L3 and detected NP by indirect immunofluorescence staining. As expected, it accumulated in the nucleus, exactly like the virally encoded NP in THOV infected cells (data not shown).

Comparison of orthomyxovirus nucleoproteins. To search for similarities, the deduced NP amino acid sequence of THOV was compared with entries in the Swiss-Prot database by the BLASTP program. Interestingly, the program detected the NP sequences of DHOV and influenza B virus but not of influenza A or C virus. To determine regions of importance shared by the five viruses, we used the Jotun-Hein algorithm (21), which is the preferred method for aligning gene families with unambiguous evolutionary relationships. Figure 2A shows the result of such an alignment. Four highly conserved central regions (marked by boldface letters) which resemble those previously identified by Fuller et al. (16) were detected. One of these regions included the nuclear accumulation sequence (c in Fig. 2A) that has been determined by Davey et al. (10). The RNA-binding domains that were originally identified by Albo et al. (2) and Kobayashi et al. (27) were less conserved (regions a and b, respectively, in Fig. 2A). The similarity matrix (Fig. 2B) revealed that the NP of THOV has 43% amino acid sequence similarity with that of DHOV and about 14% similarity with those of influenza A, B, and C viruses. On the basis of these alignments, a phylogenetic tree for the *Orthomyxoviridae* family was constructed (Fig. 2C). It can be seen that influenza A and B viruses are closely related to each other and share a common ancestor with the evolutionarily more distant influenza C virus (50). The two tick-transmitted viruses THOV and DHOV constitute a separate branch distinct from that of the influenza viruses.

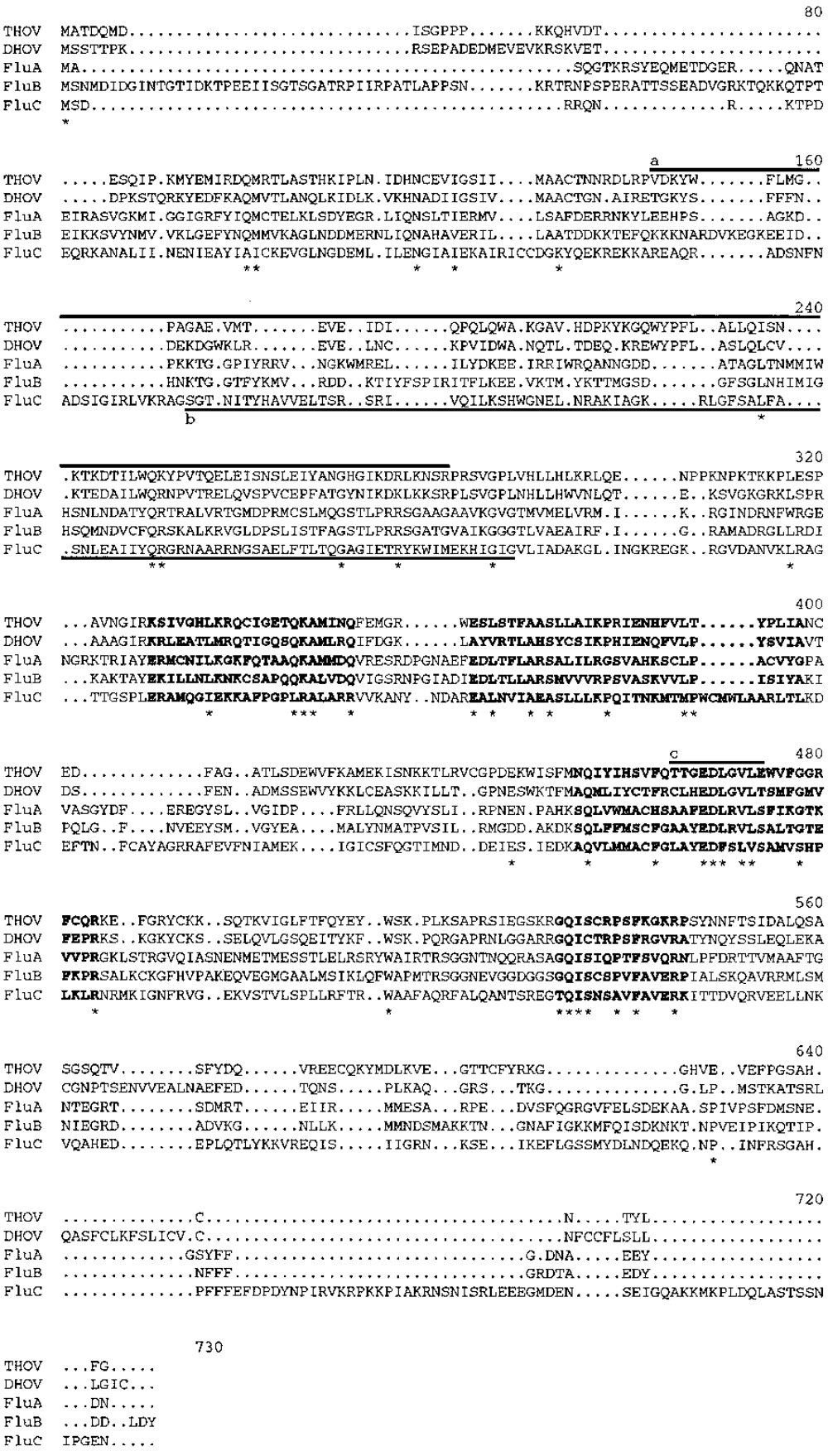
THOV mRNAs lack large heterogeneous sequences at their 5′ ends. The sequence at the 3′ end of THOV RNA segment 5 was found to be 3′ UCGUUU. . . (Fig. 1B). Given the overall similarity of THOV with the influenza viruses, we expected that the 5′ ends of the corresponding mRNA would possess extensions of approximately 10 to 15 nucleotides that are derived from host cellular mRNAs by a cap-stealing mechanism (29). We determined the 5′-end sequences of the NP mRNAs by a 5′ RACE analysis using either dC or dG homopolymer tailing. To obtain a representative set of sequences, 31 clones that differed in the lengths of their homopolymer tails were analyzed. Surprisingly, no heterogeneous sequences resembling those of influenza viral mRNAs (7, 12, 30) were detected (Table 1). Most 5′-end sequences perfectly matched the 3′ end of the NP genomic RNA. Sequences with one or two extra nucleotides were also found. It is not clear whether these sequences represent true products of the THOV transcription process or whether they are derived from nucleotides incorporated during the 5′ RACE procedure by mistake. Shorter products lacking one or two nucleotides were also observed. These may have been generated by premature termination of RT or by RNA degradation. To exclude artifacts due to contaminating cRNAs, we also used an mRNA circularization procedure. Decapped mRNA was circularized, and the sequence of the 5′-3′ end junction was determined. This provided direct proof for the presence of the poly(A) tail ligated to the mRNA 5′ end (Fig. 3, lanes A, C, T, and G). Taken together, these results indicate that THOV mRNAs lack large additional heterogeneous extensions at their 5′ ends.

THOV mRNAs have cap structures. The next question was whether THOV mRNAs contain a cap structure. To address this question, we used an anti-cap antibody to immunoselect mRNA-cDNA hybrids after primer extension. Poly(A)$^+$ RNA isolated from infected cells was annealed with a radiolabelled NP-specific oligonucleotide primer and was reverse transcribed. The extension product of this reaction is shown in Fig. 3, lane 1. The resulting RNA-cDNA hybrid was then immunoprecipitated with the cap-specific MAb H-20 (5) coupled to protein G-Sepharose beads. Finally, the immunoselected labelled RNA-cDNA hybrid was denatured and separated on a 8% polyacrylamide gel next to the products of a sequencing reaction (using the same radiolabelled oligonucleotide). Figure 3 shows that the anti-cap antibody specifically recognized the THOV NP mRNAs (lane 2), indicating that THOV mRNAs do have a m^7G cap structure. An unrelated antibody was unable to immunoselect the capped mRNAs, and therefore no radiolabelled primer extension products became visible (Fig. 3, lane 3). Furthermore, the dominant band of the primer extension product comigrated with the position of the first nucleotide at the 5′ end of the circularized NP mRNA (Fig. 3, lane T). Again, this result indicates that no large 5′ extensions were present, confirming the result of the 5′ RACE analysis. The surrounding weaker bands found at distinct positions represent extension products with one or two additional or missing nucleotides (compare with Table 1). The faint signals above the main band could also be reverse transcription products of the uncoded 5′ terminal cap G (22).

DISCUSSION

The orthomyxovirus family contains two tick-borne viruses, THOV and DHOV. Both are structurally and genetically similar to influenza viruses but are capable of infecting arthropod as well as vertebrate hosts. A distinguishing trait is their unique envelope glycoprotein that has no sequence similarity with the surface glycoproteins of the influenza viruses (14, 37). Instead, it is related to the gp64 glycoproteins of baculoviruses, a virus group infecting insects, suggesting that tick-borne orthomyxoviruses have acquired an arthropod cell gene at some point in their evolution (40). The NP of influenza virus is the major structural protein that associates with the genomic RNA segments to form the ribonucleoprotein particles. Our results

A


80
THOV MATDQMD.....ISGPPP.....KKQHVDT.....
DHOV MSSTTPK.....RSEPADEDMEVEVKRSKVET.....
FluA MA.....SQGTKRSYEQMETDGER.....QNAT
FluB MSNMDIDGINTGTIDKTPEEIISGTSGATRPIIRPATLAPPSN.....KRTRNPSPERATTSSEADVGRKTQKKQTPT
FluC MSD.....RRQN.....R.....KTPD
a
160
THOV .....ESQIP.KMYEMIRDQMRTLASTHKIPLN.IDHNCEVIGSII....MAACTNNRDLRPVDKYW.....FLMG..
DHOV .....DPKSTQRKYEDFKAQMVTLANQLKIDLK.VKHNADIIGSIV....MAACTGN.AIRETGKYS.....FFFN..
FluA EIRASVGKMI.GGIGRFYIQMCTELKLSDYEGR.LIQNSLTIERMV....LSAFDERRNKYLEEHPS.....AGKD..
FluB EIKKSVYNMV.VKLGEFYNQMMVKAGLNDDMERNLIQNAHAVERIL....LAATDDKKTEFQKKKNARDVKEGKEEID..
FluC EQRKANALII.NENIEAYIAICKEVGLNGDEML.ILENGIAIEKAIRICCDGKYQEKREKKAREAQR.....ADSNFN
240
THOV .....PAGAE.VMT.....EVE..IDI.....QPQLQWA.KGAV.HDPKYKGQWYPFL..ALLQISN....
DHOV .....DEKDGWKLR.....EVE..LNC.....KPVIDWA.NQTL.TDEQ.KREWYPFL..ASLQLCV....
FluA .....PKKTG.GPIYRRV...NGKWMREL.....ILYDKEE.IRRIWRQANNGDD.....ATAGLTNMMIW
FluB .....HNKTG.GTFYKMV...RDD..KTIYFSPIRITFLKEE.VKTM.YKTTMGSD.....GFSGLNHIMIG
FluC ADSIGIRLVKRAGSGT.NITYHAVVELTSR..SRI.....VQILKSHWGNEL.NRAKIAGK.....RLGFSALFA....
b
320
THOV .KTKDTILWQKYPVTQELEISNSLEIYANGHGIKDRLKNSRPRSVGPLVHLLHLKRLQE.....NPPKNPKTKKPLESP
DHOV .KTEDAILWQRNPVTRELQVSPVCEPFATGYNIKDKLKKSRPLSVGPLNHLLHWVNLQT.....E..KSVGKGRKLSPR
FluA HSNLNDATYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGVGTMVMELVRM.I.....K..RGINDRNFWRGE
FluB HSQMNDVCFQRSKALKRVGLDPSLISTFAGSTLPRRSGATGVAIKGGGTLVAEAIRF.I.....G..RAMADRGLLRDI
FluC .SNLEAIIYQRGRNAARRNGSAELFTLTQGAGIETRYKWIMEKHIGIGVLIADAKGL.INGKREGK..RGVDANVKLRAG
400
THOV ...AVNGIRKSIVGHLKRQCIGETQKAMINQFEMGR.....WESLSTFAASLLAIKPRIENHFVLT.....YPLIANC
DHOV ...AAAGIRKRLEATLMRQTIGQSQKAMLRQIFDGK.....LAYVRTLAHSYCSIKPHIENQFVLP.....YSVIAVT
FluA NGRKTRIAYERMCNILKGKFQTAAQKAMMDQVRESRDPGNAEFEDLTFLARSALILRGSVAHKSCLP.....ACVYGPA
FluB ...KAKTAYEKILLNLKNKCSAPQQKALVDQVIGSRNPGIADIEDLTLLARSMVVVRPSVASKVVLP.....ISIYAKI
FluC ...TTGSPLERAMQGIEKKAPPGPLRALARRVVKANY..NDAREALNVIAEASLLLKPQITNKMTMPWCNWLAARLTLKD
c
480
THOV ED.....FAG..ATLSDEWVFKAMEKISNKKTLRVCGPDEKWISFMNQIYIHSVFQTTGEDLGVLEWVFGGR
DHOV DS.....FEN..ADMSSEWVYKKLCEASKKILLT..GPNESWKTFMAQMLIYCTFRCLHEDLGVLTSMFGMV
FluA VASGYDF....EREGYSL..VGIDP....FRLLQNSQVYSLI..RPNEN.PAHKSQLVWMACHSAAFEDLRVLSFIKGTK
FluB PQLG..F....NVEEYSM..VGYEA....MALYNMATPVSIL..RMGDD.AKDKSQLFFMSCFGAAYEDLRVLSALTGTE
FluC EFTN..FCAYAGRRAFEVFNIAMEK....IGICSFQGTIMND..DEIES.IEDKAQVLMMACFGLAYEDFSLVSAMVSHP
560
THOV FCQRKE..FGRYCKK..SQTKVIGLFTFQYEY..WSK.PLKSAPRSIEGSKRGQISCRPSFKGKRPSYNNFTSIDALQSA
DHOV FEPRKS..KGKYCKS..SELQVLGSQEITYKF..WSK.PQRGAPRNLGGARRGQICTRPSFRGVRATYNQYSSLEQLEKA
FluA VVPRGKLSTRGVQIASNENMETMESSTLELRSRYWAIRTRSGGNTNQQRASAGQISIQPTFSVQRNLPFDRTTVMAAFTG
FluB FKPRSALKCKGFHVPAKEQVEGMGAALMSIKLQFWAPMTRSGGNEVGGDGGSGQISCSPVFAVERPIALSKQAVRRMLSM
FluC LKLRNRMKIGNFRVG..EKVSTVLSPLLRFTR..WAAFAQRFALQANTSREGTQISNSAVFAVERKITTDVQRVEELLNK
640
THOV SGSQTV.....SFYDQ.....VREECQKYMDLKVE...GTTCFYRKG.....GHVE..VEFPGSAH.
DHOV CGNPTSENVVEALNAEFED.....TQNS.....PLKAQ...GRS...TKG.....G.LP..MSTKATSRL
FluA NTEGRT.....SDMRT.....EIIR.....MMESA...RPE...DVSFQGRGVFELSDEKAA.SPIVPSFDMSNE.
FluB NIEGRD.....ADVKG.....NLLK.....MMNDSMAKKTN...GNAFIGKKMFQISDKNKT.NPVEIPIKQTIP.
FluC VQAHED.....EPLQTLYKKVREQIS.....IIGRN...KSE...IKEFLGSSMYDLNDQEKQ.NP..INFRSGAH.
720
THOV .....C.....N.....TYL.....
DHOV QASFCLKFSLICV.C.....NFCCFLSLL.....
FluA .....GSYFF.....G.DNA.....EEY.....
FluB .....NFFF.....GRDTA.....EDY.....
FluC .....PFFFEFDPDYNPIRVKRPKKPIAKRNSNISRLEEEGMDEN.....SEIGQAKKMKPLDQLASTSSN
730
THOV ...FG.....
DHOV ...LGIC...
FluA ...DN.....
FluB ...DD..LDY
FluC IPGEN.....

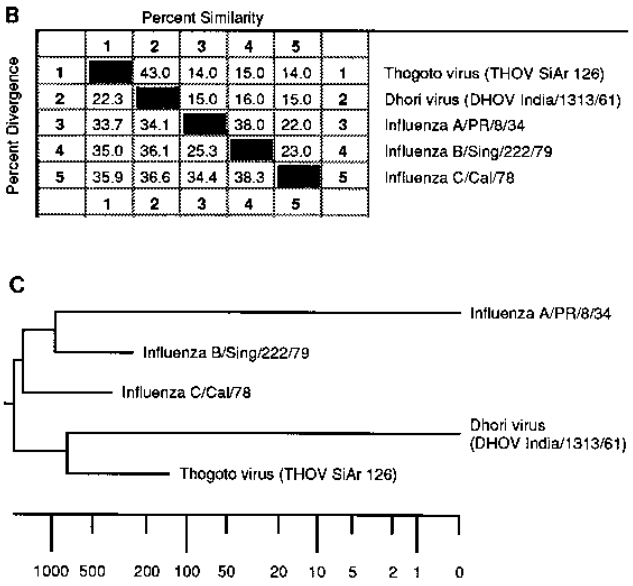


FIG. 2. Comparison of the THOV, DHOV, and influenza virus NP sequences. (A) Multiple alignment of amino acid sequences. The NP sequences of THOV strain SiAr 126 (THOV), Dhori/Ind/1313/61 (DHOV) (16), influenza A/PR/8/34 virus (FluA) (51), influenza B/Sing/222/79 virus (FluB) (33), and influenza C/Cal/78 virus (FluC) (39) are shown. The asterisks indicate residues with at least 80% identity between the five sequences compared. The horizontal lines designate regions that have previously been identified as important for either RNA binding (a [2] and b [27]) or nuclear accumulation (c [10]). Boldface letters indicate the four regions which are highly conserved between all orthomyxoviruses (see text). The alignment was obtained by using the MegAlign program with the Jotun-Hein algorithm (21). The residue weight table used was PAM 250, gap penalty was 11, k-tuple was 2, and gap length penalty was 3. (B) Similarity matrix of the NP sequences aligned in panel A. Whereas the similarity index compares two sequences directly, the divergency is calculated by comparing sequence pairs in relation to the complete phylogeny reconstructed by the MegAlign program. (C) Proposed phylogenetic tree for the two known Thogoto-like viruses THOV and DHOV and for members of the influenza A, B, and C viruses. The length of each of the branches is proportional to the number of amino acid substitutions, which is indicated by the scale beneath the tree.

show that the NP of THOV has many features in common with the NPs of influenza viruses, although the amino acid sequence similarity is only 14%. In contrast, the NP sequence of THOV shows a considerable degree of similarity (43%) with the putative NP of DHOV. Interestingly, four separate short regions (14 to 30 amino acids long) that are highly conserved between THOV, DHOV, and the influenza viruses were detected (Fig. 2A). They overlap with four regions identified by Fuller et al. (16) and may represent critical domains for conserved functions of this protein family. One of the blocks includes the nuclear accumulation sequence as defined by Davey et al. (10). The sequence similarities in another region previously shown to be responsible for RNA binding (2, 27) are less prominent. Nevertheless, the predicted secondary structure (consisting of two alpha helices connected by a loop-beta sheet-loop domain) of the RNA-binding domain described by Albo et al. (2) is preserved (data not shown). Taken together, these data suggest that all known orthomyxoviruses derive their NP genes from one common progenitor gene, unlike the glycoprotein gene. Yet, the tick-borne viruses constitute a separate evolutionary path, as demonstrated by the phylogenetic tree shown in Fig. 2C. Clearly, influenza A and B viruses share a common ancestor with the same origin as influenza C virus (50), whereas THOV and DHOV form a more distant branch with no direct relationship to any of the influenza viruses.

We have obtained the noncoding sequences at the 5′ and 3′

TABLE 1. Comparison of the 5′ ends of the NP mRNAs with the corresponding 3′ end of the genomic vRNA of THOV

Sequence	No. of clones obtained by:		
	5′ RACE		mRNA circularization
	Oligo(dC) tail	Oligo(dG) tail	
NP vRNA 3′ end			
UCGUUUUUGUCC...			
NP mRNA 5′ end			
AGCAAAAACAGG...	11	3	1
C**A**GCAAAAACAGG...	3		
UC**A**GCAAAAACAGG...	1		
U**U**GCAAAAACAGG...	1		
(G)CAAAAACAGG[a]...	6		
More than 1 nucleotide shortened	3	3	

[a] Sequences with (G) most likely start with a vRNA-templated G at position 2 which is indistinguishable from the first nucleotide of the homopolymeric tail.

ends of RNA segment 5 by two different approaches consisting of a 5′ RACE analysis and a vRNA circularization procedure. Our results show that the sequence at the 3′ end is 3′ UCG..., as already suggested for segment 3 and 4 of THOV (37, 48). The sequence is thus identical with the published 3′ ends of influenza A, B, and C viruses (11) and DHOV (9). This information enabled us to address the question whether THOV mRNAs contain extra nucleotides at their 5′ ends. In the case of influenza viruses, mRNA synthesis occurs in the cell nucleus and is primed by 5′ capped fragments derived from newly synthesized host cell RNA polymerase II transcripts (29). Transcription initiation with these 10- to 15-nucleotide-long capped primers leads to the distinctive heterogeneous 5′ ends of influenza virus mRNAs (7, 12, 30). Likewise, THOV has a nuclear phase (43, 47) and requires host cell mRNA synthesis for its multiplication (47). Moreover, we found that mRNAs of THOV were recognized by a cap-specific antibody, indicating that the mRNAs had a m^7G cap structure at their 5′ ends (Fig.

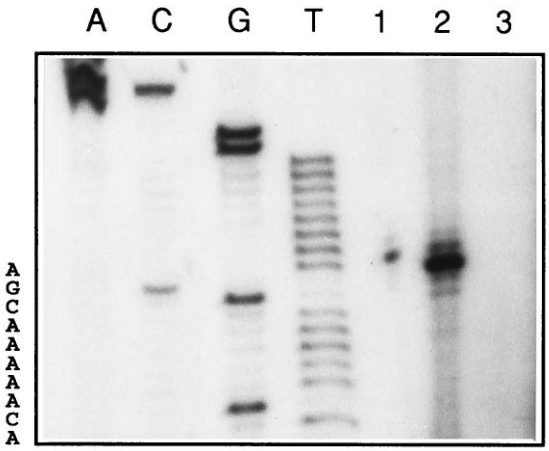


FIG. 3. THOV mRNA transcripts possess 5′-terminal cap structures. Autoradiogram of the primer extension product before (lane 1) and after (lanes 2 and 3) immunoselection. Primer extension was performed with poly(A)$^+$ RNAs from THOV-infected cells and a labelled primer complementary to 5′-end sequences of THOV NP mRNA. The resulting RNA-cDNA hybrid was immunoprecipitated (17) with the anti-cap MAb H-20 (5) (lane 2) or an unrelated control antibody (lane 3) and was analyzed on a 8% sequencing polyacrylamide gel. The same primer was used to generate a sequencing ladder from the cloned 5′-3′ junction of the NP mRNA (lanes A, C, G, and T). The plus-sense sequence of the mRNA 5′ end is indicated on the left.

3). We therefore expected that the mRNAs transcribed from THOV segment 5 would possess heterogeneous 5′ extensions of approximately 10 to 15 nucleotides that are not templated from vRNA. In contrast, we found that the 5′ ends of the capped THOV mRNAs were homogeneous in length and sequence and were exact copies of the vRNA 3′-end sequences. Interestingly, Portela and coworkers have recently obtained similar results for the mRNA transcripts of THOV segments 3 and 4 (42a). An earlier report by Staunton et al. (48) had mentioned that oligo(dT)-selected RNA transcripts of the third THOV segment lacked extra nucleotides when analyzed by primer extension.

The present data do not allow us to draw firm conclusions about the origin of the cap structure found at the 5′ end of THOV mRNAs. One possibility is that mRNA synthesis is initiated without a primer, in the same way as synthesis of vRNA or cRNA of influenza virus (29), and that a viral or cellular mechanism generates the cap. However, this is unlikely because (i) the RNA polymerases of all segmented, negative-strand RNA viruses studied so far seem to utilize capped cellular RNAs to prime mRNA synthesis (7, 12, 17, 18, 24, 36, 42, 46) and (ii) THOV has a homolog to the PB2 polymerase subunit of influenza virus (49a) which is implicated in cap binding and cleaving of the primer RNA (49). A more likely scenario is that cap stealing does occur, but that only the cap structure plus the first nucleotide is cleaved from a restricted subset of host mRNAs. Since nearly all 5′-end sequences characterized here have an adenosine residue at position 1, we propose that fragments derived from cellular mRNAs with this nucleotide at the 5′ end are used preferentially, as suspected for influenza virus (25, 41). In some cases, the polymerase may cleave one or two positions downstream of the cap structure, resulting in an extended 5′ sequence. Such minor products (Table 1) would account for the more slowly migrating signals detected by primer extension (Fig. 3). Since all full-length 5′ RACE-derived clones contained the guanosine at position 2, this residue is most likely the first nucleotide incorporated by the viral polymerase, as found for influenza virus (7, 41). The U at position 1 of the vRNA 3′ end is able to pair with bases other than A (23) and may thus serve as a loose adapter for the primer fragments. Strong hydrogen bonding of the G and C bases that are added subsequently may then hold the growing mRNA chain in a stable position for elongation. Alternatively, longer primers may be generated and subsequently trimmed to fit the 3′ end of the vRNA template by a hypothetical exonuclease. To experimentally confirm the novel cap-stealing mechanism proposed here, it will be necessary to characterize the PB2 polymerase subunit of THOV and to demonstrate that the THOV polymerase complex has cap-binding and endonuclease activities. Elucidating the mechanism by which tick-borne orthomyxoviruses differ from influenza viruses in initiating mRNA synthesis should contribute to a better understanding of the molecular biology of this important virus family.

ACKNOWLEDGMENTS

We thank Patricia A. Nuttall for providing THOV and MAbs and for helpful discussions and Reinhard Lührmann for the generous gift of anti-cap antibodies and for advice. We are grateful to Daniel Kolakoswky for helpful suggestions, Agustin Portela for communicating results prior to publication, and Peter Staeheli for critical comments.

This work was supported by grants from the Deutsche Forschungsgemeinschaft (Ko 1579/1-1), the EU Human Capital and Mobility Network program (grant ERBCHRXCT940453), and the Land Baden-Württemberg (ZKF-B1).

REFERENCES

1. **Albanese, M., C. Bruno-Smiraglia, G. Di Cuonzo, A. Lavagnino, and S. Srihongse.** 1972. Isolation of Thogoto virus from *Rhipicephalus bursa* ticks in western Sicily. Acta Virol. **267:**267.
2. **Albo, C., A. Valencia, and A. Portela.** 1995. Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein. J. Virol. **69:**3799–3806.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
4. **Ausubel, F., R. Brent, R. E. Kingston, D. D. Moore, J. S. Seidman, J. A. Smith, and K. Struhl (ed.).** 1992. Current protocols in molecular biology. John Wiley & Sons, Inc., Boston.
5. **Bochnig, P., R. Reuter, P. Bringmann, and R. Lührmann.** 1987. A monoclonal antibody against 2 2 7 trimethylguanosine that reacts with intact class U small nuclear ribonucleoproteins as well as with 7 methylguanosine-capped RNA. Eur. J. Biochem. **168:**461–467.
6. **Burd, C. G., and G. Dreyfuss.** 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science **265:**615–621.
7. **Caton, A. J., and J. S. Robertson.** 1980. Structure of the host-derived sequences present at the 5′ ends of influenza virus mRNA. Nucleic Acids Res. **8:**2591–2603.
8. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
9. **Clerx, J. P. M., F. Fuller, and D. H. Bishop.** 1983. Tick-borne viruses structurally similar to orthomyxoviruses. Virology **127:**205–219.
10. **Davey, J., N. J. Dimmock, and A. Colman.** 1985. Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in *Xenopus* oocytes. Cell **40:**667–675.
11. **Desselsberger, U., V. R. Racaniello, J. J. Zarza, and P. Palese.** 1980. The 3′ and 5′ terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. Gene **8:**315–328.
12. **Dhar, R., R. M. Chanock, and C.-J. Lai.** 1980. Nonviral oligonucleotides at the 5′ terminus of cytoplasmic influenza viral mRNA deduced from cloned complete genomic sequences. Cell **21:**495–500.
13. **Dreiding, P., P. Staeheli, and O. Haller.** 1985. Interferon-induced protein Mx accumulates in nuclei of mouse cells expressing resistance to influenza viruses. Virology **140:**192–196.
14. **Freedman-Faulstich, E. Z., and J. F. Fuller.** 1990. Nucleotide sequence of the tick-borne, orthomyxo-like Dhori/Indian/1313/61 virus envelope gene. Virology **175:**10–18.
15. **Frohmann, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:**8998–9002.
16. **Fuller, F. J., E. Z. Freedman-Faulstich, and J. A. Barnes.** 1987. Complete nucleotide sequence of the tick-borne, orthomyxo-like Dhori/Indian/1313/61 virus nucleoprotein gene. Virology **160:**81–87.
17. **Garcin, D., and D. Kolakofsky.** 1990. A novel mechanism for the initiation of Tacaribe arenavirus genome replication. J. Virol. **64:**6196–6203.
18. **Garcin, D., M. Lezzi, M. Dobbs, R. M. Elliott, C. Schmaljohn, C. Y. Kang, and D. Kolakofsky.** 1995. The 5′ ends of Hantaan virus (*Bunyaviridae*) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. J. Virol. **69:**5754–5762.
19. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25:**263.
20. **Haller, O., M. Frese, D. Rost, P. Nuttall, and G. Kochs.** 1995. Tick-borne Thogoto virus infection in mice is inhibited by the orthomyxovirus resistance gene product Mx1. J. Virol. **69:**2596–2601.
21. **Hein, J.** 1990. Unified approach to alignment and phylogenies. Methods Enzymol. **183:**626–645.
22. **Hirzmann, J., D. Luo, J. Hahnen, and G. Hobom.** 1993. Determination of messenger RNA 5′-ends by reverse transcription of the cap structure. Nucleic Acids Res. **21:**3597–3598.
23. **Holbrook, S. R., C. Cheong, I. Tinoco, and S. H. Kim.** 1991. Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. Nature (London) **353:**579–581.
24. **Huiet, L., P. A. Feldstein, J. H. Tsai, and B. W. Falk.** 1993. The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5′ termini. Virology **197:**808–812.
25. **Kawakami, K., K. Mizumoto, and A. Ishihama.** 1983. RNA polymerase of influenza virus. IV. Catalytic properties of the capped RNA endonuclease associated with the RNA polymerase. Nucleic Acids Res. **11:**3637–3649.
26. **Kawasaki, E. S., S. S. Clark, M. Y. Coyne, S. D. Smith, R. Champlin, O. N. Witte, and F. P. McCormick.** 1988. Diagnosis of chronic myeloid and acute lymphocytic leukemias by detection of leukemia-specific mRNA sequences amplified in vitro. Proc. Natl. Acad. Sci. USA **85:**5698–5702.
27. **Kobayashi, M., T. Toyoda, D. M. Adyshev, Y. Azuma, and A. Ishihama.** 1994. Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain. J. Virol. **68:**8433–8436.
28. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15:**8125–8148.

29. **Krug, R. M., F. Alonso-Caplen, I. Julkunen, and M. G. Katze.** 1989. Expression and replication of the influenza virus genome, p. 89–152. *In* R. Krug (ed.), The influenza viruses. Plenum Press, New York.
30. **Krug, R. M., B. B. Broni, and M. Bouloy.** 1979. Are the 5′ ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs? Cell **18:**329–334.
31. **Lamb, R. A.** 1989. Genes and proteins of the influenza viruses, p. 1–88. *In* R. M. Krug (ed.), The influenza viruses. Plenum Press, New York.
32. **Lamb, R. A., and C. M. Horvarth.** 1991. Diversity of coding strategies in influenza viruses. Trends Genet. **7:**261–266.
32a. **Leahy, M., F. Weber, and J. T. Dessens.** Unpublished data.
33. **Londo, D. R., A. R. Davis, and D. P. Nayak.** 1983. Complete nucleotide sequence of the nucleoprotein gene of influenza B virus. J. Virol. **47:**642–648.
34. **Mandl, C. W., F. X. Heinz, E. Puchhammer-Stoeckl, and C. Kunz.** 1991. Sequencing the termini of capped viral RNA by 5′-3′ ligation and PCR. BioTechniques **10:**485–486.
35. **Mark, G. E., J. M. Taylor, B. Broni, and R. M. Krug.** 1979. Nuclear accumulation of influenza viral RNA and the effects of cycloheximide actinomycin D and alpha-amanitin. J. Virol. **29:**744–752.
36. **Meyer, B. J., and P. J. Southern.** 1993. Concurrent sequence analysis of 5′ and 3′ RNA termini by intramolecular circularization reveals 5′ nontemplated bases and 3′ terminal heterogeneity for lymphotic choriomeningitis virus mRNAs. J. Virol. **67:**2621–2627.
37. **Morse, M. A., A. C. Marriott, and P. A. Nuttall.** 1992. The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein gp64. Virology **186:**640–646.
38. **Murphy, F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers.** 1995. Virus taxonomy. Springer-Verlag, Vienna.
39. **Nakada, S., R. S. Creager, M. Krystal, and P. Palese.** 1984. Complete nucleotide sequence of the influenza C/California/78 virus nucleoprotein gene. Virus Res. **1:**433–441.
40. **Nuttall, P. A., M. A. Morse, L. D. Jones, and A. Portela.** 1995. Orthoacariviruses. *In* A. J. Gibbs and C. H. Calisher (ed.), Molecular evolution of viruses, p. 416–425. Cambridge University Press, Cambridge.
41. **Plotch, S. J., M. Bouloy, I. Ulmanen, and R. Krug.** 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell **23:**847–858.
42. **Polyak, S. J., S. Zheng, and D. G. Harnish.** 1995. 5′ termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. J. Virol. **69:**3211–3215.
42a. **Portela, A.** Personal communication.
43. **Portela, A., L. D. Jones, and P. Nuttall.** 1992. Identification of viral structural polypeptides of Thogoto virus (a tick-borne orthomyxo-like virus) and functions associated with the glycoprotein. J. Gen. Virol. **73:**2823–2830.
44. **Rott, R., and C. Scholtissek.** 1970. Specific inhibition of influenza replication by alpha-amanitin. Nature (London) **228:**56.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
46. **Shaw, M. W., and R. A. Lamb.** 1984. A specific subset of host-cell mRNAs prime influenza virus mRNA synthesis. Virus Res. **1:**455–467.
47. **Siebler, J., O. Haller, and G. Kochs.** 1996. Thogoto and Dhori virus replication is blocked by inhibitors of cellular polymerase II activity but does not cause shutoff of host cell protein synthesis. Arch. Virol. **141:**1587–1594.
48. **Staunton, D., P. A. Nuttall, and D. H. L. Bishop.** 1989. Sequence analyses of Thogoto viral RNA segment 3: evidence for a distant relationship between an arbovirus and members of the *Orthomyxoviridae*. J. Gen. Virol. **70:**2811–2817.
49. **Ulmanen, I., B. A. Broni, and R. M. Krug.** 1981. Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription. Proc. Natl. Acad. Sci. USA **78:**7355–7359.
49a. **Weber, F., G. Kochs, and O. Haller.** Unpublished data.
50. **Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka.** 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. **56:**152–179.
51. **Winter, G., and S. Fields.** 1981. The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34. Virology **114:**423–428.