Functions and requirements of conserved RNA structures in the 3' untranslated region of Flaviviruses

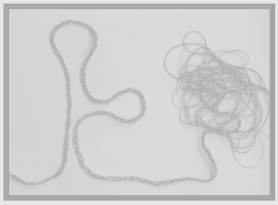
Agostinho Gonçalves Costa da Silva, P.

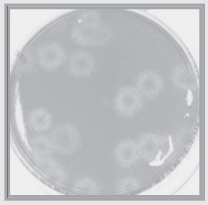
Chapter 4

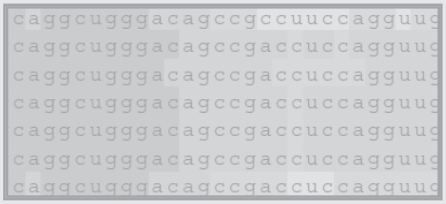
Characterization of the sfRNAs that are produced in cells infected with flaviviruses with no known vector and cell fusing agent

Patrícia A. G. C. Silva
Tim. J. Dalebout
Alexander P. Gultyaev
Rene C. Olsthoorn
Peter J. Bredenbeek

This manuscript has been submitted for publication in "RNA"







ABSTRACT

A virus-specific, non-coding RNA of 0.3 – 0.5 kb, co-linear with the genomic 3' UTR can be detected in cells and mice infected with arthropod-borne flaviviruses. This small flavivirus RNA (sfRNA) results from incomplete degradation of the viral genome by the host 5' – 3' exonuclease XRN1 and was shown to be important for viral pathogenicity. To determine whether sfRNA production is a unique feature of the Flavivirus genus or only restricted to vector-borne flaviviruses, flaviviruses with no known vector (NKV) and the insect flavivirus cell fusing agent virus (CFAV) have been analyzed for the production of the sfRNA. The data presented in this study clearly demonstrate that the XRN1-mediated production of sfRNA is not limited to the vector-borne flaviviruses and most likely is an unique common feature of all flaviviruses, implying that it could be considered an additional determinant to assign viruses to this genus. Computer-aided RNA structure predictions combined with *in vitro* XRN1 assays and cell culture experiments defined an RNA pseudoknot as the XRN1 stalling site for the production of these sfRNAs in NKV flaviviruses and CFAV. These data imply that the sfRNA is likely to be important for the flavivirus life cycle in both the mammalian host and the arthropod vector, as NKV flaviviruses restricted to mammalian hosts and the mosquito-restricted CFAV produce an sfRNA.

INTRODUCTION

The genus Flavivirus of the *Flaviviridae* family contains nearly 80 viruses, including many important human pathogens such as dengue virus (DENV), yellow fever virus (YFV), and West Nile virus (WNV). Based on phylogenetic analysis, flaviviruses were grouped into three major clusters that correlate with the type of vector used for their transmission: (i) mosquito-borne, (ii) tick-borne, and (iii) no known vector (NKV) flaviviruses [1,2]. No arthropod vector has yet been implicated in the transmission of NKV viruses. NKV flaviviruses have been isolated exclusively from rodents or bats and are divided into three groups: i) the Entebbe bat virus group which includes viruses like the Entebbe bat and Yokose virus (YOKV), ii) the Modoc virus group that comprises Modoc virus (MODV) and Apoi virus (APOIV) and iii) the Rio Bravo virus group which includes viruses like Rio Bravo virus (RBV) and Montana myotis leukoencephalitis virus (MMLV) [3]. In contrast to the MODV and RBV groups, whose members are unable to replicate in the mosquito C6/36 cell line, viruses belonging to the Entebbe bat group can replicate in these cells albeit to low titers [4].

Apart from the viruses that are assigned to one of the clusters within the Flavivirus genus, there are viruses like cell fusing agent virus (CFAV) that are considered tentative flaviviruses [5]. CFAV was isolated from a cell line derived from laboratory-reared *Aedes aegypti* mosquitoes [6] and has been classified as a tentative insect flavivirus with genome organization and gene expression strategy similar to that of the flaviviruses. However, CFAV can only be propagated in mosquito cells and not in cell lines of vertebrate origin [6,7]. Although CFAV has never been found in nature, CFAV-related viruses like Kamiti River virus (KRV) have been isolated from field-collected mosquitoes [8-10].

All flaviviruses have a positive single-stranded RNA genome of approximately 11 kb, with a 5′ cap structure and a 3′ non-polyadenylated end. The genome encodes one large open reading frame that is flanked by 5′ and 3′ untranslated regions (UTRs) that contain several conserved RNA sequences and structures that are involved in the regulation of translation and viral genome amplification. Translation of the viral genome results in a polyprotein that is co- and post-translationally processed by viral and cellular proteases into the individual viral proteins [11]. Northern blot analysis of viral RNA isolated from mammalian and insect cell lines or mice infected with arthropod-borne flaviviruses has revealed the production of a small, positive-stranded, non-coding flavivirus RNA (sfRNA) in addition to the viral genome [12-17]. This sfRNA is 0.3 – 0.5 kb long, co-linear with the 3′ end of the viral genome and originates from incomplete degradation of the viral genomic RNA by the host 5′-3′ exonuclease XRN1, due to stalling of this nuclease upstream an RNA pseudoknot located in the viral 3′ UTR [15,17,18].

Although the precise role of the sfRNA in the viral life cycle still needs to be elucidated, current data suggest that it is important for viral pathogenicity in the mammalian host [15,18], (Silva, Pereira, Dalebout, and Bredenbeek, unpublished results). Despite the fact that

sfRNA production has also been described in mosquito cells infected with mosquito-borne flaviviruses [13,15,17], nothing is known about the potential role of the sfRNA in the arthropod host. If production of the sfRNA is only required for efficient completion of the viral life cycle in either the mammalian or the arthropod host, it is not unlikely that the ability to produce an sfRNA might be lacking in either the NKV flaviviruses or CFAV. To address this hypothesis, the production of sfRNA in mammalian cells infected with several NKV flaviviruses and of mosquito cells infected with CFAV was analyzed. Surprisingly, all the flaviviruses that were included in this study produced at least one sfRNA that was co-linear with the 3' end of the viral genome. As has been shown for arthropod-borne flaviviruses, production of sfRNA by these NKV viruses and CFAV is also mediated by the host 5' – 3' exoribonuclease XRN1, which is well conserved in eukaryotes. In addition, the minimal sequence within the viral 3' UTR required for the stalling of XRN1 on the genome of MODV, MMLV and also CFAV was determined and found to form an RNA pseudoknot.

MATERIAL AND METHODS

Cell culture

The origin and culture conditions of the BHK-21J cells have been described before [19,20]. C6/36 cells [21] were obtained from the ATCC and grown in EMEM supplemented with 8% fetal calf serum (Bodinco, The Netherlands) and 5% none-essential amino acids.

Recombinant DNA techniques and plasmid constructions

Unless described in more detail, standard nucleic acid methodologies were used [22,23]. Chemically competent *E. coli* DH5α cells [24] were used for cloning. The nucleotide numbering was according to the sequence files for which the accession numbers can be found in table 1.

Infections were performed essentially as described before [20]. Total RNA was isolated with Trizol (Invitrogen) at 30 hr p.i. from BHK-21J cells infected with MODV, APOIV, MMLV or RBV or at 36 hr p.i. from CFAV infected C6/36 cells. RNA was dissolved in 30 μl H_2O and 5 μg was used for first strand cDNA synthesis using M-MuLV reverse transcriptase (Fermentas). The PCR was performed with GoTaq Flexi DNA Polymerase (Promega) as described by the manufacturer. Oligonucleotides used in the PCR contained either a *Mlu I* site (forward primer) or a *Sph I* site (reversed primer). The RT-PCR products were cloned using the TOPO TA Cloning system (Invitrogen). Inserts with the correct sequence were isolated after digestion of the plasmids with *Mlu I – Sph I* and cloned into Sinrep5eGFP

Table 1.	**Oligonucleotides that were used to identify and characterize the sfRNAs of NKV flaviviruses and CFAV.** The oligonucleotides that were used in this study, the virus to which they were directed, the NCBI accession number used to obtain the sequence and the actual nucleotide sequence are indicated. All oligonucleotides are complementary to the viral genome. Abbreviations in the column "Purpose" refer to Northern blotting and hybridizations (Hyb.), primer extension (Prim.Ex.) and DNA sequencing (Seq.).			
Oligo	**Virus**	**NCBI Number**	**Sequence (5' to 3')**	**Purpose**
NKV2	MMLV	AJ299445	CCGCTCAATCTCGAGAGGAGCGA	Hyb/Prim.Ex.
NKV3	APOIV	AF452050	CTCAGGCGCTAAAGGATGCCGCTA	Hyb.
NKV4	MODV	AJ242984	GGGTCTCCACTAACCTCTAGTCCT	Hyb.
NKV6	APOIV	AF452050	CGCTCAAAGAGAGAAGGGTCGC	Hyb.
NKV19	RBV	AF452049	ACTCGGTCAGTTGGGATCATCCCAC	Hyb.
NKV20	MODV	AJ242984	CCCTAACCTATTTACAATGACTGGC	Hyb./Prim.Ex.
NKV21	CFAV	NC_008604	AGATGGGCCGCCACCACCATCTTAG	Hyb./Prim.Ex.
NKV24	YOKV	NC_005039	TCCATGCGTAGGAGAGGGTCTCC	Hyb.
NKV31	RBV	AF452049	CACCCTATCAGGGTTGACTGGCTCA	Prim.Ex.
NKV33	APOIV	AF452050	CCCACTGGAATGCAATGCTGGCC	Prim.Ex.
YFV1632	YFV-17D	X03700	ACCCCGTCTTTCTACCACC	Seq.
YFV1676	SinV	NC_001547	GTACCAGCCTGATGCATTATGCACATC	Hyb.

[25,17]. Plasmid DNAs of these pSinrep5eGFP recombinants containing either a MODV, MMLV or CFAV insert were linearized with *Not I* and used for *in vitro* RNA transcription [20].

In vitro XRN1 assay

Plasmid DNA of the pSinrep5eGFP recombinants was prepared for *in vitro* RNA transcription without the addition of a cap analog as described above. The RNA transcripts were pre-treated with tobacco acid pyrophosphatase (TAP, Epicentre) to create a 5' mono-phosphate and incubated with 1 unit of XRN1 (available as Terminator 5'-phosphate-dependent exonuclease, Epicentre) as described before [17].

RNA transfection and analysis of viral RNA

BHK-21J cells were transfected with 5 or 20 µg of Sinrep5eGFP and recombinant RNAs as described before [20]. In general, 2.5 ml (approximately 1.5 x 10^6 cells) of the transfected BHK-21J cell suspension was seeded in a 35 mm plate. Total RNA was isolated from the transfected cells at 8 hr post electroporation (p.e.). Trizol (Invitrogen) was used for cell lysis and subsequent RNA purification.

For Northern blotting, samples containing 7.5 – 10 µg of total RNA isolated from either infected or electroporated cells, or obtained from *in vitro* XRN1 assays, were denatured using formaldehyde and separated on a formaldehyde containing 1.5% agarose gel and blotted to Hybond-N$^+$ (GE-Healthcare) [23]. The blots were hybridized with random hexameer primed cDNA probes or ^{32}P-labelled oligonucleotides [26,23,27] that were targeted at the 3' UTR of the virus under study.

Primer extension assay

Primer extension analysis was performed as reported by Sambrook *et al.* [23] with minor modifications [17]. Briefly, 5 – 7 µg total RNA from virus infected cells or XRN1-treated recombinant SINrep5 transcripts, were annealed to a ^{32}P-labeled oligonucleotide that was specific for the studied viral RNA. The primer extension products were analyzed on a denaturing 5% polyacrylamide/8M urea sequence gel. A ^{33}P-labeled Cycle Reader sequence reaction (Fermentas) using oligonucleotide 1632 primed pBlsrcptSK-YFV$_{9845\text{-}10861}$ as a template served as a sequence marker.

RNA structure prediction

RNA structure was predicted as described by Olsthoorn and Bol [28]. The viruses included in this analysis were: MODV, MMLV, RBV, APOIV and CFAV. The NCBI accession numbers for the sequences can be found in table 1.

RESULTS

Production of a small 3' subgenomic RNA is a unique feature of all Flaviviruses

Recently it was shown that many, if not all, of the arthropod-borne flaviviruses produce an sfRNA that is collinear with the distal part of the viral 3' UTR [13,15-17]. These sfRNAs were generated in infected mammalian as well as in insect cells. BHK-21J cells were infected with the NKV flaviviruses MODV, APOIV, RBV, MMLV and YOKV to determine whether such sfRNAs were also produced by the flaviviruses that lack an arthropod vector. Total RNA was isolated from the infected cells at 30 hr p.i. and analyzed for sfRNA production by Northern blot analysis using ^{32}P-labelled oligonucleotides directed against the distal part of the 3' UTR of the NKV viruses as a probe. In addition to the viral genomic RNA,

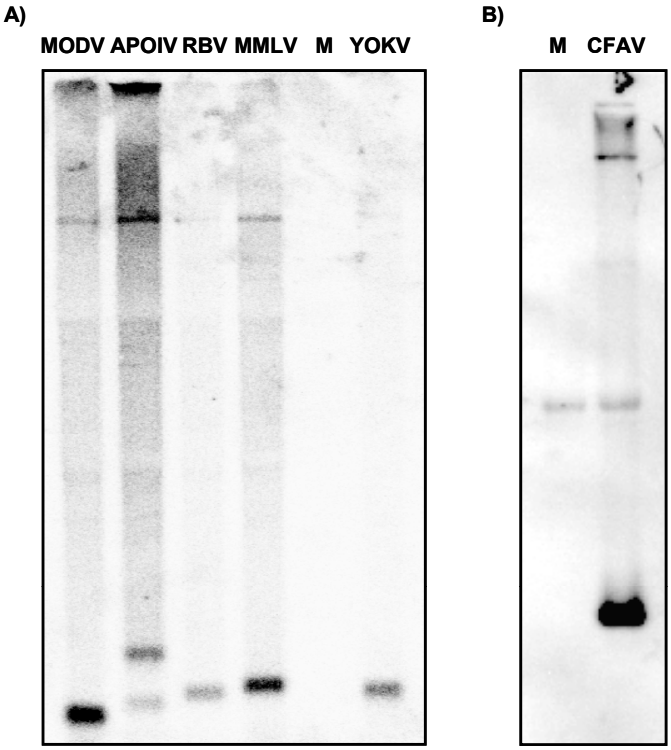


Fig. 1. sfRNA production in mammalian and insect cells infected with NKV flaviviruses and CFAV. A) Northern blot analysis of viral RNAs isolated at 30 hr p.i. from BHK-21J cells infected with MODV, APOIV, RBV, MMLV and YOKV respectively. Kinased oligonucleotides complementary to the distal part of the respective virus 3' UTR were mixed and used as probes. **B)** Northern blot analysis of CFAV RNAs isolated from infected C6/36 cells at 36 hr p.i. A ^{32}P-labelled oligonucleotide complementary to the distal part of the CFAV 3' UTR was used as a probe. Lane M corresponds to total RNA isolated from mock-infected cells.

a small virus-specific RNA was detected for the NKV flaviviruses MODV, RBV, MMLV and YOKV (fig. 1.A; lanes 1, 3, 4 and 6), whereas two small virus-specific RNAs were detected in cells infected with APOIV (fig. 1.A; lane 2).

Apart from these NKV flaviviruses, the tentative insect flavivirus CFAV was also tested for the production of a small virus specific RNA originating from the viral 3' UTR. Mosquito C6/36 cells were infected with CFAV and at 36 hr p.i. total RNA was isolated and analyzed by Northern blotting and hybridization. As shown in fig. 1.B, a small RNA was readily detected in CFAV-infected C6/36 cells. From these results it was concluded that, similar to the arthropod-borne flaviviruses, MODV, APOIV, RBV, MMLV and YOKV, representing the three different groups of NKV flaviviruses, produced at least one sfRNA. These data demonstrated that sfRNA production is a distinguishing feature for all Flaviviruses. Even CFAV, a virus tentatively assigned to the Flavivirus genus, was shown to produce an sfRNA.

Determining the 5′ end of the sfRNAs from flaviviruses with no known vector and CFAV

Primer extension analysis on total RNA isolated from infected cells was used to determine the 5′ end of the sfRNA of MODV, APOIV, MMLV, RBV and CFAV. As shown in fig. 2.A, primer extension on total RNA isolated from MODV-infected BHK-21J cells (lane 1) resulted in the production of two unique cDNA products that were only one nucleotide apart in length and not present in total RNA isolated from uninfected cells (lane 2). Using the sequence ladder that was run in parallel as a marker, the 5′ end of the MODV sfRNA was mapped to nt position 10.262 or 10.263. Based on these results, the MODV sfRNAs

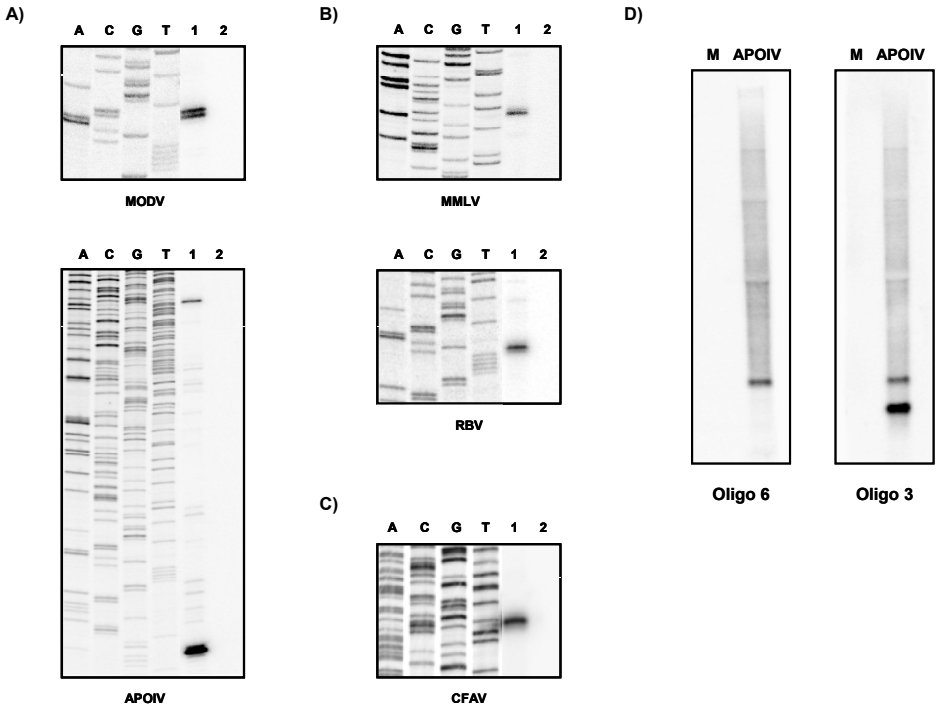


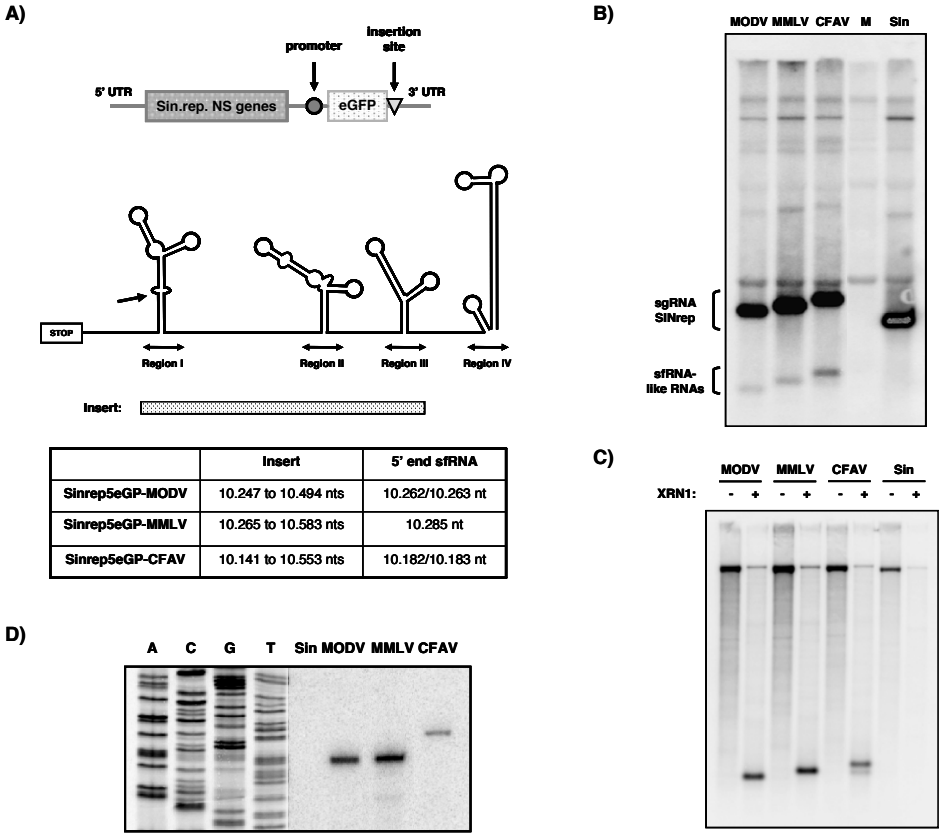
Fig. 2. Mapping the relative position of the NKV and CFAV sfRNAs to the viral genome. Primer extension analysis was performed to determine the 5′ end of the sfRNAs produced by the rodent NKV viruses MODV and APOIV (panel **A**), the bat NKV viruses MMLV and RBV (panel **B**), and the tentative insect flavivirus CFAV (panel **C**). RNA was isolated from infected BHK-21J cells at 30 hr p.i. for the NKV flaviviruses and at 36 hr p.i. from CFAV-infected C6/36 cells. For panels A to C, lanes 1 and 2 correspond to primer extension on total RNA isolated from infected and uninfected cells, respectively. Information on the oligonucleotides that were used as probes for these viruses is presented in table 1. A sequence reaction using oligonucleotide YFV1632 on pBluescript-$YFV_{9,845\text{-}10,861}$ [17] was used as a DNA size marker. **D)** Northern blot analysis of viral RNA isolated from APOIV-infected BHK-21J cells to determine the relative orientation on the viral genome of the two APOIV sfRNAs using oligonucleotides NKV3 and NKV6 (see table 1) as a probe. Lane M corresponds to total RNA isolated from mock-infected cells.

were calculated to be 337 to 338 nts in length. Primer extension on APOIV RNA also resulted in two cDNA products; however, in contrast to MODV, these products showed a significant size difference. This was actually expected given the results of the hybridization presented in fig. 1.A. Based on the length of the primer extension products, the longest sfRNA was calculated to be approximately 566 nts, whereas the smaller sfRNA was predicted to have a length of approximately 371 nts. Primer extension analysis on MMLV, RBV and CFAV RNA resulted in unique products (fig. 2, panels B and C). The 5′ ends were mapped to positions 10.285 in the MMLV genome, 61 nts into the 3′ UTR of RBV, and positions 10.182 – 10.183 in the genome of CFAV. Based on these primer extension results, the sfRNAs of MMLV, RBV and CFAV were calculated to be 405 nt, 425 nt, and 512 – 513 nt, respectively.

The combined results of the Northern blot (fig. 1.A, lane 2) and primer extension (fig. 2.A), suggested that the two detected APOIV sfRNAs would form a 3′ nested set. To determine whether this hypothesis was correct, the position of the APOI sfRNAs relative to the viral genome was analyzed by Northern blotting using oligonucleotides NKV3 and NKV6 as probes. Oligonucleotide NKV3 is complementary to the 3′ end of the viral genome and will recognize both APOIV sfRNAs if they form a 3′ nested set. Oligonucleotide NKV6 hybridizes to a position upstream of the determined 5′ end of the smaller APOIV sfRNA and is predicted to detect only the larger sfRNA if the hypothesis is correct. The results presented in fig. 2.D clearly demonstrated that the APOIV sfRNAs form a 3′ nested set. Both sfRNAs hybridized to the 3′ end-specific oligonucleotide NKV3, whereas only the largest APOIV sfRNA hybridized with oligonucleotide NKV6.

The host exoribonuclease XRN1 is required for sfRNA production of NKV flaviviruses and CFAV

It has now been firmly established that the sfRNAs of WNV and YFV are produced by incomplete degradation of the viral genome by the host 5′ – 3′ exoribonuclease XRN1 [15,17]. To analyze whether XRN1 was also required for the production of the NKV flaviviruses sfRNAs, MODV and MMLV were selected to represent two different groups of NKV flaviviruses that are associated with rodents and bats, respectively. In addition, whenever possible, the tentative flavivirus species CFAV was included in these studies. Unfortunately, no full-length cDNA clone for the transcription of infectious RNA is available for any of these three viruses. To circumvent this handicap, cDNA fragments encompassing the XRN1 stalling site were cloned into the Sindbis virus derived RNA driven expression vector Sinrep5eGFP [25]. This strategy allowed the use of the same RNA templates for both *in vivo* and *in vitro* studies. A schematic representation and relevant details of these constructs is shown in fig. 3.A. BHK-21J cells were transfected with *in*



	Insert	5' end sfRNA
Sinrep5eGP-MODV	10.247 to 10.494 nts	10.262/10.263 nt
Sinrep5eGP-MMLV	10.265 to 10.583 nts	10.285 nt
Sinrep5eGP-CFAV	10.141 to 10.553 nts	10.182/10.183 nt

Fig. 3. Insertion of the MODV, MMLV and CFAV 3' UTR sequence directs the *in vivo* and *in vitro* production of an sfRNA-like RNA in the context of a Sindbis replicon RNA. A) Schematic representation of the pSinrep5eGFP vector characteristics and the predicted general RNA folding of the NKV 3' UTR according to Charlier and colleagues [29]. The 5' end of the sfRNAs relative to this RNA structure is indicated by an arrow. The box below the structure indicates the region of the MODV, MMLV or CFAV 3' UTR cloned into the Sinrep5eGFP vector. The names of the constructs and the exact nucleotide numbers of the cDNA fragments cloned into Sinrep5eGFP are indicated in the table in panel A. **B)** Northern blot analysis of total RNA isolated from BHK-21J cells transfected with Sinrep5eGFP-MODV (lane 1), Sinrep5eGFP-MMLV (lane 2), Sinrep5eGFP-CFAV (lane 3), mock transfected cells (lane 4) and Sinrep5eGFP (lane 5). Oligonucleotide 1676 complementary to Sinrep5eGFP nucleotides downstream of the insertion site of flavivirus 3' UTR was used as probe. **C)** *In vitro* production of the MODV, MMLV and CFAV sfRNA-like RNA by incubation with XRN1. The "–" symbol refers to incubation without the enzyme, while the "+" signal corresponds to the addition of XRN1. **D)** Primer extension using oligonucleotide 1676 to determine the 5' end of the sfRNA-like RNAs produced by *in vitro* incubation of RNA transcribed from the Sinrep5eGFP constructs with XRN1. "Sin" represents the control reaction on pSinrep5eGFP.

vitro transcribed RNA of these Sinrep constructs to demonstrate that an sfRNA-like RNA could be produced in cell culture. As shown in fig. 3.B, Northern blotting of total RNA isolated from these transfected cells using oligonucleotide 1676 as a probe revealed that Sinrep5eGFP-MODV, Sinrep5eGFP-MMLV and Sinrep5eGFP-CFAV produced an sfRNA-like RNA. This additional RNA was not detected in cells that were transfected with Sinrep5eGFP that did not contain any flavivirus sequences (lane Sin) or mock transfected cells (lane M). RNA transcripts of Sinrep5eGFP, Sinrep5eGFP-MODV, Sinrep5eGFP-MMLV and Sinrep5eGFP-CFAV, were incubated *in vitro* with commercially available purified XRN1 and analysed by Northern blotting to demonstrate that this enzyme was also responsible for the production of the NKV and CFAV sfRNAs. The results showed that an sfRNA-like RNA was generated upon incubation with XRN1 of the Sinrep5eGFP transcripts that contain either the MODV, MMLV or CFAV insert encompassing the predicted XRN1 stalling site (fig. 3.C). Primer extension analysis on these sfRNA-like RNAs produced *in vitro* demonstrated that the 5' ends that were produced using this heterologous expression system were identical to the 5' ends of the sfRNAs as detected in BHK-21J cells infected with either MODV, MMLV or CFAV (fig. 3.D). The combined data of the *in vivo* and *in vitro* experiments using the Sinrep5eGFP templates provided strong evidence that XRN1 is also responsible for the production of the sfRNA in cells infected with NKV flaviviruses or CFAV.

An RNA pseudoknot in the 3' UTR of NKV flaviviruses and CFAV is required for the production of the sfRNAs

Current data for YFV and WNV strongly suggest that the stalling of XRN1 that is required for the production of sfRNA in arthropod-borne flaviviruses is directed by an RNA pseudoknot within the viral 3' UTR [17,18]. Therefore, the nucleotide sequences immediately downstream of the predicted 5' end of the NKV flaviviruses and CFAV sfRNAs were analyzed for their potential to fold into an RNA pseudoknot structure. An alignment of the primary sequence in this region of the NKV flaviviruses genomes showed significant sequence similarity interspaced by insertions or deletions of a few nucleotides (fig. 4.A). A similar result was obtained when CFAV was compared to KRV and Aedes virus. The latter two viruses were recently isolated from mosquitoes and shown to be closely related to CFAV. Subsequent extensive RNA structure modelling predicted the formation of RNA pseudoknots for all the NKV flaviviruses as well as for CFAV and the related KRV and Aedes virus. The predicted RNA pseudoknot structures for MODV, MMLV and CFAV that could potentially serve as stalling sites for XRN1 are depicted in fig. 4.B. The structure for MODV and MMLV was predicted to be very similar. Compared to the structure for YFV [17], the sequence indicated by pk that was predicted to base pair with the pk' sequence

A)

```
MODV:   5' ugaaagagugu-u-gaggg-caaccaguggggc-uagccac--------augggu-----augacgcaccc-acccucu-gca-uucuuguaaa-uacuuug-gccagucauugu----- 3'
MMLV:   5' aaaaaaugcgagu-gaggg-caac-ucugggu-uagcuca--------augggu-----gugacg-acccuacccuuccgca-uuu--guaaa-uaa-uugagccagucauuuc----- 3'
RBV:    5' -----uucggugc-u-gaggggcaac-ucuaggacuagccua-------augggu-----gugacg-acccuaccccaaugcacuu---guaaa-uaa-uugagccagucaacc------- 3'
APOIV:  5' ---aauaagcgccuggagag-cgac-cuu-ggacuguggggguccaaagggcuugcgacccuucucucuuugagcgcuuugggacaacuauuug-gccag-cauugcauuc 3'

CFAV:   5' ccacaggagcagggcaugaaaaugucgggcaugacgaacccgcucccccgaguc-cccug-gcaa-caggg- 3'  (1st putative stalling signal)
        5' ---guucgagcagggcacauuggugucgggcgugacgcacccgcucccuc-aguc-cccuguguaa-caggg- 3'  (2nd putative stalling signal)
KRV:    5' --------agagcagggcacgaaagugucgggcaugacgcacccgcucccccgaguc-cccug-aaaa-uaggg- 3'
Aedes:  5' --------agagcagggcacaauagugucgggccugacgacccgcucccccgagucgcccaacggaguuuggc 3'
```

B)

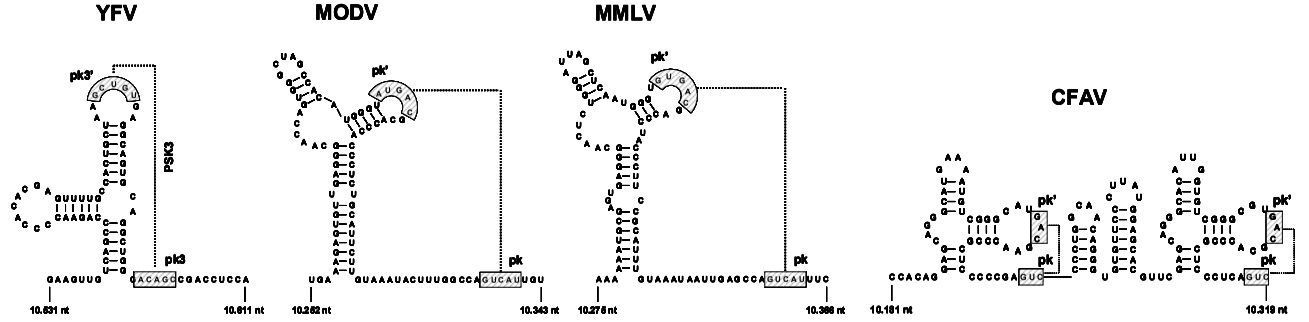
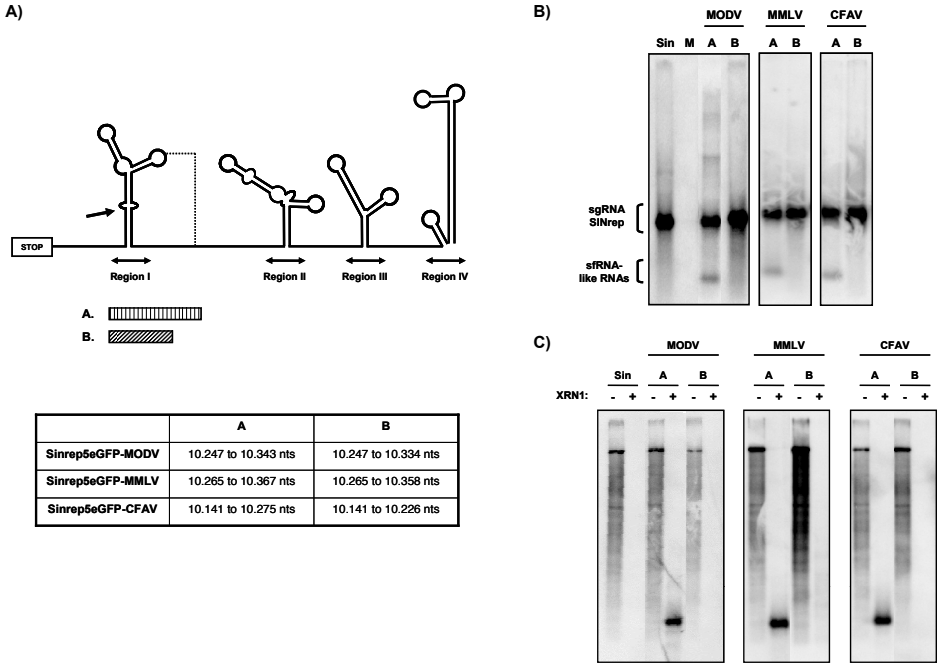


Fig. 4. Sequence alignment and predicted RNA structures of the region that is involved in stalling XRN1 in MODV, MMLV and CFAV. A) Alignment of the sequences downstream the 5' end of the sfRNAs of the NKV flaviviruses MODV, MMLV, RBV and APOIV (top panel) and the tentative flaviviruses CFAV, KRV and Aedes virus (bottom panel). The second predicted stalling signal for XRN1 in CFAV is also included. Base paired nucleotides are underlined. Conserved nucleotides are indicated in bold. Putative pk and pk' sequences are in italics and indicated by boxes. Gaps were introduced to maximize the nucleotide identity. The NCBI accession numbers for KRV and Aedes virus are NC_012671 and NC_012932, respectively. Accession numbers for the other viruses can be found in table 1. **B)** RNA structures were predicted using a combinatorial approach involving Mfold, phylogeny and manual sequence-structure analysis. Potential RNA pseudoknot interactions between the pk and pk' sequences [28] are indicated by grey boxes. For comparison, the YFV-17D RNA pseudoknot that has been shown to stall XRN1 for the production of the YFV sfRNA is depicted [17].

(nomenclature according to Olsthoorn and Bol [28]), is located further downstream of the stem. For both MODV and MMLV, the presumed pk' – pk interaction comprised five nucleotides, which is similar to the YFV pseudoknot that was proven to act as a stalling signal for XRN1 [17]. The initial impression of the folding for the comparable region of the CFAV RNA suggested that it was rather different from the proposed RNA pseudoknots for YFV, MODV and MMLV. However, closer inspection revealed that the overall folding was actually relatively similar to the proposed structures for NKV flaviviruses and YFV. The main differences of the CFAV structure versus that of the other viruses were in the length of the depicted stem structures and the pk' – pk interaction, which was proposed to involve only three nucleotides. Interestingly, the formation of a second, very similar RNA pseudoknot structure was predicted for the nucleotide sequence just downstream of the first predicted pseudoknot (fig. 4.B).

To provide support for the actual formation of these predicted RNA pseudoknots, two cDNA fragments of MODV, MMLV and CFAV were cloned into Sinrep5eGFP (fig. 5.A). The cDNA fragment A of MODV, MMLV and CFAV contained all the nucleotides that were predicted to be involved in the formation of the RNA pseudoknot required to stall XRN1. Fragment B contained a 3' deletion compared to the sequences contained in fragment A. This 3' truncation was expected to disrupt the pseudoknot formation by deletion of the pk sequence. The ability of the cDNA fragments A and B of MODV, MMLV and CFAV to direct the production of an sfRNA-like RNA was analyzed *in vivo* as well as *in vitro* using the Sinrep5eGFP based expression system. RNA transcripts of the NKV virus and CFAV Sinrep5eGFP constructs containing cDNA fragment A or B were electroporated into BHK-21J cells. At 8 hrs. p.e., total RNA was isolated and analyzed by Northern blotting using the Sindbis virus 3' UTR specific oligonucleotide 1676 as a probe. As shown in fig. 5.B, all the cells transfected with Sinrep5eGFP RNA containing cDNA fragment A of MODV, MMLV or CFAV were able to produce an sfRNA-like RNA in the transfected cells. As expected, this sfRNA-like RNA was not detected in BHK cells that were either mock transfected (lane M) or electroporated with Sinrep5eGFP RNA lacking any flavivirus insert (lane Sin). More importantly, this sfRNA-like RNA was not detected in cells that were transfected with RNA of the Sinrep5eGFP recombinants containing cDNA fragment B of MODV, MMLV, and CFAV.

In vitro experiments were performed with purified XRN1 and RNA transcribed from the Sinrep5eGFP recombinants containing the MODV, MMLV or CFAV cDNA fragments A and B to provide additional evidence for the role of the predicted RNA pseudoknot in the stalling of XRN1. Sinrep5eGFP RNA containing the NKV flavivirus or CFAV cDNA fragment A, did produce an sfRNA-like RNA after incubation with XRN1, whereas Sinrep5eGFP constructs encompassing cDNA fragment B were completely degraded and did not yield this additional subgenomic RNA (fig. 5.C). The outcome of the *in vivo* and in *vitro* experiments presented in fig. 5 supports the predicted model in which an RNA



	A	B
Sinrep5eGFP-MODV	10.247 to 10.343 nts	10.247 to 10.334 nts
Sinrep5eGFP-MMLV	10.265 to 10.367 nts	10.265 to 10.358 nts
Sinrep5eGFP-CFAV	10.141 to 10.275 nts	10.141 to 10.226 nts

Fig. 5. RNA sequences that are predicted to be involved in the formation of an RNA pseudoknot in the 3′ UTR of MODV, MMLV and CFAV are required for the production of an sfRNA. A) Generalized overall RNA folding of the NKV 3′ UTR based on Charlier and colleagues [29]. The relative position of the RNA pseudoknot that is predicted to be involved in sfRNA production is indicated by a dashed line. The 5′ end of the sfRNAs relative to this RNA structure is indicated with an arrow. The boxes below the structure indicate the regions of the flavivirus 3′ UTR that were cloned into the Sinrep5eGFP vector from MODV, MMLV and CFAV. The exact nucleotide numbers of the cDNA fragments cloned into Sinrep5eGFP are indicated in the table. **B)** Northern blot analysis of RNA isolated from BHK-21J cells transfected with Sinrep5eGFP (Sin) and Sinrep5eGFP with the 3′ UTR A and B inserts of MODV, MMLV and CFAV, respectively. M corresponds to uninfected BHK-21J cells. Oligonucleotide 1676 was used as probe. **C)** Northern blot analysis of RNA transcripts of Sinrep5eGFP and derivatives containing fragment A or B of MODV, MMLV or CFAV incubated with XRN1. The "–" symbol refers to incubation without the enzyme, while the "+" signal corresponds to the addition of XRN1. Oligonucleotide 1676 was used as a probe.

pseudoknot is required to stall XRN1 for the production of an sfRNA-like RNA. In contrast to the Sinrep5eGFP recombinants containing MODV, MMLV or CFAV cDNA fragment A, formation of the RNA pseudoknot that stalls XRN1 was no longer possible in fragment B due to deletion of the pk sequence, explaining why an sfRNA-like RNA was not produced.

DISCUSSION

The family *Flaviviridae* comprises three genera: Flaviviruses, Pestiviruses, and Hepaciviruses [5]. All the members of this virus family are enveloped, positive-stranded RNA viruses with a similar genome organization and expression strategy. Despite their evolutionary relatedness, these viruses show significant differences in host range, tropism, pathogenicity, and various aspects of their molecular biology. Over the last few years it has been shown that arthropod-borne Flaviviruses produce a small flavivirus RNA (sfRNA) in addition to their genome-length negative- and positive-stranded RNAs [12-17]. This RNA results from partial 5′ to 3′ degradation of the viral genome by the host exoribonuclease XRN1 [15,17]. No such small viral subgenomic RNA has been detected in cells infected with Pestiviruses and Hepaciviruses [15], suggesting that the production of the sfRNA might be a new feature specific for Flaviviruses. However, flaviviruses are divided into three groups and the previous studies only analyzed sfRNA production in cells and mice infected with either mosquito- or tick-borne viruses. In this study, viruses that belong to the third, poorly studied cluster of Flaviviruses with no known vector, were analyzed for their ability to direct sfRNA synthesis. The analysis included MODV and APOIV from the rodent-associated MODV-related flaviviruses and MMLV and RBV, along with YOKV, representing the bat-associated RBV and Entebbe bat virus subgroup of the NKV flaviviruses, respectively [3]. In addition to these NKV flaviviruses, the mosquito cell-infecting virus CFAV, which has been tentatively assigned to the Flavivirus genus, was also included in this study. The results of this analysis showed that all NKV flaviviruses as well as CFAV are able to produce at least one sfRNA in infected cells. APOIV even generates two sfRNAs that form a 3′ nested set. Primer extension analysis on total RNA mapped the 5′ end of the NKV sfRNAs to an internal bulge in the Y-structure that is indicated as region I in the 3′ UTR of NKV flaviviruses [29]. Due to recombination, deletions and sequence duplications, the length of the Flavivirus 3′ UTR is rather heterogeneous (reviewed in [30]). This size heterogeneity is also reflected in the length of the sfRNAs. In general, the sfRNAs are co-linear with the 3′ distal part of the viral 3′ UTR and contain all the conserved RNA elements that are present in this region of the genome. These results demonstrate that production of a subgenomic sfRNA is a new feature of viruses that belong to the genus Flavivirus and that it can be considered as an additional characteristic to the established ICTV criteria for classification of newly discovered viruses into the genus Flavivirus.

Using mutagenesis of the available infectious cDNA clones, *in vitro* assays and RNA silencing experiments, previous studies on WNV and YFV have demonstrated that the sfRNAs of the arthropod-borne flaviviruses are produced by incomplete degradation of the viral genomic RNA by the host 5′-3′ exoribonuclease XRN1 [15,17]. No infectious cDNA clone is currently available for any of the NKV flaviviruses or CFAV. Therefore, a

Sindbis virus-based expression system [25] was used to determine whether XRN1 was also required for the production of sfRNA in cells infected with NKV flaviviruses or CFAV. The cDNA fragments that contained the nearly complete distal part of the MODV, MMLV or CFAV 3' UTR were cloned into the Sinrep5-eGFP and analyzed for their ability to direct the synthesis of an sfRNA-like RNA *in vitro* and *in vivo*. Transfection of BHK-21J cells with RNA transcribed from these constructs did indeed result in the production of an sfRNA-like RNA. An identical RNA was detected upon incubation of the Sinrep5eGFP recombinants containing the MODV, MMLV or CFAV insert with purified XRN1. Furthermore, the *in vitro* produced sfRNA-like RNAs have an identical 5' end as the sfRNAs that are produced during infection of cell cultures with MODV, MMLV or CFAV. These results strongly suggest that XRN1 is also required for the production of the sfRNA in NKV flaviviruses and CFAV. These results are actually not very surprising since XRN1 is well conserved among eukaryotes as it plays a vital role in 5' to 3' mRNA decay (reviewed in [31-34]), and therefore in the homeostasis of the host. Several experiments have been performed to provide additional proof for the role of XRN1 in the sfRNA production of MODV and MMLV by using lentiviruses expressing small-hairpin RNAs to silence human XRN1 expression. However, silencing of XRN1 has a significant impact on the condition of the cells and this, together with the relatively poor replication of MODV and MMLV in the tested SW13 and Huh7 cells, unfortunately resulted in inconclusive data (Silva, Dalebout and Bredenbeek, unpublished results).

It has previously been shown that XRN1 is stalled by an RNA pseudoknot in the 3' UTR of YFV and WNV to produce the sfRNA [17,18]. However, no RNA pseudoknot structure has been suggested in the 3' UTR region predicted to stall XRN1 in either the NKV flaviviruses or CFAV [29,7]. Sequence alignment combined with RNA structure modelling have been used in this study to predict an RNA pseudoknot just downstream of the mapped 5' end of the sfRNA for every NKV virus, similar to what has been encountered in WNV and YFV. Unfortunately, the limited amount of sequence data available for these NKV flaviviruses does not allow for co-variance analysis to obtain additional support for the predicted structures. An RNA pseudoknot that can serve as a stalling site for XRN1 has also been predicted for CFAV. On first sight, this RNA structure may seem rather unlikely, due to the relatively short stem-loop structures and a pseudoknot interaction that only involves three nucleotides. A nearly perfect duplication of this pseudoknot was predicted downstream of the first pseudoknot (fig. 4, panels A and B). Careful inspection of the data presented in fig. 3, where the stalling of XRN1 was analyzed in the background of the Sindbis expression system, reveals a minor, slightly smaller sfRNA-like RNA that can be explained by the XRN1 stalling on the second, more downstream located pseudoknot. This minor band is no longer detected in the data presented in fig. 5, because the sequence involved in the formation of this second pseudoknot is not present in the CFAV inserts of those particular Sindbis constructs. So these experiments actually

provide support for the proposed CFAV pseudoknot. The sequence of the CFAV, KRV and Aedes virus 3' UTR in the region in which the pseudoknot structure is predicted is nearly identical (fig. 4.A); therefore any co-variance in the primary sequence to support this structure is very limited.

The sequence that encompasses the 5' end of the sfRNA and the predicted RNA pseudoknots of MODV, MMLV and CFAV was cloned into the Sinrep5eGFP expression vector and shown to be capable of producing an sfRNA-like RNA. These results demonstrate that the ability for stalling XRN1 is contained within a relatively small sequence of MODV, MMLV, and CFAV and that no other sequences are required to produce the sfRNA. Deletion of the pk sequence, which is predicted to interact with the pk' sequence to form the RNA pseudoknot, from the flavivirus insert in the Sinrep5eGFP constructs abolished the production of the sfRNA-like RNA. This result supports the model in which, similar to what was found in arthropod-borne flaviviruses, RNA pseudoknots are required to stall XRN1. In addition, N-methylisatoic anhydride (NMIA) based chemical probing of the RNA structure required for stalling XRN1 showed that the pk' and pk sequences were in a double-stranded conformation, which is in agreement with the predicted RNA pseudoknots. The RNA probing results of the sequences that encompass the XRN1 stalling sites of MODV, MMLV, and CFAV, however, did not fully support the stem-loop part of the predicted pseudoknots, suggesting that a more complex structure might actually be present (Silva, Dalebout and Bredenbeek, unpublished results). Unfortunately, NMIA-based probing only discriminates between single- and double-stranded nucleotides and does not allow identification of individual nucleotides within a structure.

Although the precise function of the sfRNA in the virus life cycle remains to be elucidated, current evidence for WNV and YFV have shown that it is an important determinant for the pathogenicity of these viruses in their mammalian hosts [15,18] (Silva, Pereira, Dalebout, and Bredenbeek, unpublished results). Nothing is known about the requirement of the sfRNA of these viruses in the arthropod host. The fact that the ability to produce the sfRNA is maintained in flaviviruses that have a more limited host range, like the NKV flaviviruses and CFAV, strongly suggest that the sfRNA is essential for successful survival of these viruses in the mammalian or the insect host. However, this does not necessarily imply that the function of the NKV virus or CFAV sfRNA is similar to that of the sfRNA of the arthropod-borne viruses in either their vertebrate or arthropod host. Determining the function of these Flavivirus sfRNAs in the various virus-host systems and unravelling the link between the sfRNA and viral pathogenicity will be an interesting challenge for further research.

REFERENCE LIST

1. **Kuno, G., G. J. Chang, K. R. Tsuchiya, N. Karabatsos, and C. B. Cropp**. 1998. Phylogeny of the genus Flavivirus. J. Virol. **72**:73-83.
2. **Cook, S. and E. C. Holmes**. 2006. A multigene analysis of the phylogenetic relationships among the flaviviruses (Family: Flaviviridae) and the evolution of vector transmission. Arch. Virol. **151**:309-325.
3. **ICTV**. 2005. Virus Taxonomy - Eight Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego.
4. **Varelas-Wesley, I. and C. H. Calisher**. 1982. Antigenic relationships of flaviviruses with undetermined arthropod-borne status. Am. J. Trop. Med. Hyg. **31**:1273-1284.
5. **Thiel, H. J., M. S. Collet, E. A. Gould, F. X. Heinz, G. Meyers, R. H. Purcell, C. M. Rice, and M. Houghton**. 2005. Flaviridae, p. 981-998. *In* L. A. Ball (ed.), Virus Taxonomy - Eight Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego.
6. **Stollar, V. and V. L. Thomas**. 1975. An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. Virology **64**:367-377.
7. **Cammisa-Parks, H., L. A. Cisar, A. Kane, and V. Stollar**. 1992. The complete nucleotide sequence of cell fusing agent (CFA): homology between the nonstructural proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. Virology **189**:511-524.
8. **Cook, S., S. N. Bennett, E. C. Holmes, C. R. de, G. Moureau, and L. de, X**. 2006. Isolation of a new strain of the flavivirus cell fusing agent virus in a natural mosquito population from Puerto Rico. J. Gen. Virol. **87**:735-748.
9. **Sang, R. C., A. Gichogo, J. Gachoya, M. D. Dunster, V. Ofula, A. R. Hunt, M. B. Crabtree, B. R. Miller, and L. M. Dunster**. 2003. Isolation of a new flavivirus related to cell fusing agent virus (CFAV) from field-collected flood-water Aedes mosquitoes sampled from a dambo in central Kenya. Arch. Virol. **148**:1085-1093.
10. **Crabtree, M. B., R. C. Sang, V. Stollar, L. M. Dunster, and B. R. Miller**. 2003. Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. Arch. Virol. **148**:1095-1118.
11. **Lindenbach, B. D., H. J. Thiel, and C. M. Rice**. 2007. Flaviviridae: The Viruses and Their Replication *In* D. M. Knipe and P. M. Howley (eds.), Fields Virology. Lippincott Williams & Wilkins, Philadelphia.
12. **Urosevic, N., M. M. van, J. P. Mansfield, J. S. Mackenzie, and G. R. Shellam**. 1997. Molecular characterization of virus-specific RNA produced in the brains of flavivirus-susceptible and -resistant mice after challenge with Murray Valley encephalitis virus. J. Gen. Virol. **78 (Pt 1)**:23-29.
13. **Lin, K. C., H. L. Chang, and R. Y. Chang**. 2004. Accumulation of a 3'-terminal genome fragment in Japanese encephalitis virus-infected mammalian and mosquito cells. J. Virol. **78**:5133-5138.
14. **Scherbik, S. V., J. M. Paranjape, B. M. Stockman, R. H. Silverman, and M. A. Brinton**. 2006. RNase L plays a role in the antiviral response to West Nile virus. J. Virol. **80**:2987-2999.
15. **Pijlman, G. P., A. Funk, N. Kondratieva, J. Leung, S. Torres, L. van der Aa, W. J. Liu, A. C. Palmenberg, P. Y. Shi, R. A. Hall, and A. A. Khromykh**. 2008. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. Cell Host. Microbe **4**:579-591.
16. **Liu, R., L. Yue, X. Li, X. Yu, H. Zhao, Z. Jiang, E. Qin, and C. Qin**. 2010. Identification and characterization of small sub-genomic RNAs in dengue 1-4 virus-infected cell cultures and tissues. Biochem. Biophys. Res. Commun. **391**:1099-1103.

17. **Silva, P. A., C. F. Pereira, T. J. Dalebout, W. J. Spaan, and P. J. Bredenbeek**. 2010. An RNA Pseudoknot Is Required for Production of Yellow Fever Virus Subgenomic RNA by the Host Nuclease XRN1. J. Virol. **84**:11395-11406.
18. **Funk, A., K. Truong, T. Nagasaki, S. Torres, N. Floden, M. E. Balmori, J. Edmonds, H. Dong, P. Y. Shi, and A. A. Khromykh**. 2010. RNA structures required for production of subgenomic flavivirus RNA. J. Virol. **84**:11407-11417.
19. **Bredenbeek, P. J., E. A. Kooi, B. Lindenbach, N. Huijkman, C. M. Rice, and W. J. Spaan**. 2003. A stable full-length yellow fever virus cDNA clone and the role of conserved RNA elements in flavivirus replication. J. Gen. Virol. **84**:1261-1268.
20. **Silva, P. A., R. Molenkamp, T. J. Dalebout, N. Charlier, J. H. Neyts, W. J. Spaan, and P. J. Bredenbeek**. 2007. Conservation of the pentanucleotide motif at the top of the yellow fever virus 17D 3′ stem-loop structure is not required for replication. J. Gen. Virol. **88**:1738-1747.
21. **Igarashi, A.** 1978. Isolation of a Singh's Aedes albopictus cell clone sensitive to Dengue and Chikungunya viruses. J. Gen. Virol. **40**:531-544.
22. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl**. 2000. Current Protocols in Molecular Biology. Wiley Interscience, New York.
23. **Sambrook, J., T. Fritsch, and T. Maniatis**. 1989. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
24. **Inoue, H., H. Nojima, and H. Okayama**. 1990. High efficiency transformation of Escherichia coli with plasmids. Gene **96**:23-28.
25. **Bredenbeek, P. J., I. Frolov, C. M. Rice, and S. Schlesinger**. 1993. Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. J. Virol. **67**:6439-6446.
26. **Meinkoth, J. and G. Wahl**. 1984. Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138**:267-284.
27. **van der Most, R. G., P. J. Bredenbeek, and W. J. Spaan**. 1991. A domain at the 3′ end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. J. Virol. **65**:3219-3226.
28. **Olsthoorn, R. C. and J. F. Bol**. 2001. Sequence comparison and secondary structure analysis of the 3′ noncoding region of flavivirus genomes reveals multiple pseudoknots. RNA. **7**:1370-1377.
29. **Charlier, N., P. Leyssen, C. W. Pleij, P. Lemey, F. Billoir, L. K. Van, A. M. Vandamme, C. E. De, L. de, X, and J. Neyts**. 2002. Complete genome sequence of Montana Myotis leukoencephalitis virus, phylogenetic analysis and comparative study of the 3′ untranslated region of flaviviruses with no known vector. J. Gen. Virol. **83**:1875-1885.
30. **Gritsun, T. S. and E. A. Gould**. 2007. Origin and evolution of 3′UTR of flaviviruses: long direct repeats as a basis for the formation of secondary structures and their significance for virus transmission. Adv. Virus Res. **69**:203-248.
31. **Sheth, U. and R. Parker**. 2003. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science **300**:805-808.
32. **Anderson, P. and N. Kedersha**. 2006. RNA granules. J. Cell Biol. **172**:803-808.
33. **Eulalio, A., I. Behm-Ansmant, and E. Izaurralde**. 2007. P bodies: at the crossroads of post-transcriptional pathways. Nat. Rev. Mol. Cell Biol. **8**:9-22.
34. **Garneau, N. L., J. Wilusz, and C. J. Wilusz**. 2007. The highways and byways of mRNA decay. Nat. Rev. Mol. Cell Biol. **8**:113-126.