

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

5'- AND 3'-NONCODING REGIONS IN FLAVIVIRUS RNA

Lewis Markoff

Laboratory of Vector-borne Virus Diseases
Center for Biologics Evaluation and Research
Food and Drug Administration
Bethesda, Maryland 20892

- I. Introduction
 - II. Conserved Linear Sequence Features of NCRs in Flavivirus RNA
 - A. Complementary Nucleotides at the 5' and 3' Termini of NCRs
 - B. Cyclization Sequences
 - C. Additional Conserved Linear Sequences in 3' NCRs of Mosquito-Borne Flaviviruses
 - D. Conserved Linear Sequence Features of 3' NCRs of the TBE Subgroup of Tick-Borne Flaviviruses
 - III. Secondary Structure of NCRs
 - A. Secondary Structure of the 5' NCR
 - B. The 3' SL
 - IV. Mutational Changes in the 5' or 3' NCR and the 3' SL May Result in Attenuation of Virus Virulence
 - V. Hypothesis for Hydrogen Bonding between 5' and 3' NCRs
 - VI. Additional Secondary Structure Predictions in the 3' NCR
 - VII. Summary and Future Directions
- References

I. INTRODUCTION

The major human pathogens among the flaviviruses are yellow fever (YF) virus, the four serotypes of dengue (DEN) viruses, Japanese encephalitis (JE) virus, and the tick-borne encephalitis (TBE) viruses. Flaviviruses contain a positive-stranded RNA genome that is approximately 10,500 nucleotides in length. The RNA contains a type 1 cap at its 5' terminus (Cleaves and Dubin, 1979) but is not polyadenylated (Wengler and Wengler, 1981), and it encodes a single, long open reading frame (ORF) (reviewed in Chambers *et al.*, 1990). All evidence indicates that virus-specific proteins are derived by co- or posttranslational cleavage of the polyprotein encoded by the ORF. These include three structural proteins [capsid (C), premembrane (preM), and envelope (E)] and at least seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5, in that gene order). Most cleavages of the polyprotein are mediated by the host cell enzyme signal peptidase or by a viral protease complex requiring a central hydrophilic

domain in NS2B and the amino terminus of NS3 (Chambers *et al.*, 1993; Falgout *et al.*, 1991). Genomic RNA contains noncoding regions (NCR) 5' and 3' to the ORF. The 5'-NCR is about 100 nucleotides in length, and the 3'-NCR is 400 to nearly 800 nucleotides in length (Table I). The conserved structural and nucleotide sequence elements of these NCRs and their function in RNA replication and translation are the subjects of this review.

TABLE I
COMPARISON OF THE LENGTH IN NUCLEOTIDES OF 5' AND 3' NCRs IN GENOMES
OF SELECTED FLAVIVIRUSES

Virus genome ^a	5' NCR ^b	3' NCR	Total genome length	Method ^c	Reference ^d
DEN1, strain WP74	97	466	10,736	RNA	NC 001477
DEN2, strain NGC	89	451	10,723	DNA	NC 001474
DEN3, strain H87	93	432	10,696	DNA	NC 001475
DEN4, strain 814669	101	384	10,649	RNA	NC 002640
JE, strain JaOAR5982	95	585	10,976	DNA, RNA	NC 001437
WN, strain NY99	96	631	11,029	DNA, RNA	AF 196835
MVE	95	614	11,014	DNA, RNA	NC 000943
YF, strain 17D	118	511	10,862	DNA	Rice <i>et al.</i> (1985)
<i>TBE, strain Sofjin</i>	<i>131</i>	<i>518</i>	<i>10,894</i>	<i>DNA, RNA</i>	<i>AB 062064</i>
<i>TBE, strain Neudorfl</i>	<i>132</i>	<i>764</i>	<i>10,469</i>	<i>DNA, RNA</i>	<i>NC 001672</i>
<i>POW</i>	<i>111</i>	<i>480</i>	<i>10,839</i>	<i>DNA, RNA</i>	<i>Mandl et al.</i> (1993)
CFA ^e	113	559	10,695	DNA, RNA	NC 001564

^a Names of viruses are abbreviated as follows: DEN, dengue (numbers indicate dengue serotypes 1 to 4); JE, Japanese encephalitis; WN, West Nile; YF, yellow fever; MVE, Murray Valley encephalitis; TBE, tick-borne encephalitis; POW, Powassan; and CFA, cell-fusing agent. Strain names are abbreviated as follows: WP74, Western Pacific, 1974; NGC, New Guinea C; NY99, New York, 1999; and 17D, attenuated YF vaccine strain. Data for mosquito-borne viruses are indicated in plain type. Data for tick-borne viruses are indicated in italics.

^b Number of nucleotides assigned by sequence analysis to the 5' NCR and 3' NCR for each virus genome is indicated where noted, and the total genome size in nucleotides is also indicated where noted.

^c DNA sequence data were obtained by sequencing a series of overlapping cloned DNA fragments spanning the genome. RNA sequence data were obtained by direct sequencing of RNA or by sequencing of uncloned DNA fragments generated from RNA.

^d Either a published manuscript listed in the references or a Genbank accession number is given.

^e Has no known vector.

The NCRs in genomic RNA must be involved in the initiation of negative-strand synthesis and in any processes related to switching from negative-strand synthesis to the production of progeny virion RNAs. NCRs may also play a role in the packaging of nascent virions and contain nucleotide sequences required to initiate translation. Of the viral proteins, NS1, NS2A, NS3, NS4A, and NS5 have been implicated directly or indirectly in RNA replication. The role of NS1 in this process is not defined, but mutations in NS1 have been shown to affect the initiation of minus-strand synthesis (Lindenbach and Rice, 1997, 1999; Muylaert *et al.*, 1997). NS2A is a small hydrophobic protein localized to the replication complex that may target it to membrane organelles (Mackenzie *et al.*, 1998). NS3 contains an RNA helicase and nucleotide triphosphatase activities in addition to its protease activity (Utama *et al.*, 2000; Warrener *et al.*, 1993; Wengler and Wengler, 1991), and NS4A may form a required association with NS1 during replication (Lindenbach and Rice, 1999). NS5 functions as an RNA-dependent RNA polymerase (Ranjith-Kumar *et al.*, 2001; Tan *et al.*, 1996) and also contains a potential methyl-transferase activity, possibly required for the capping of progeny viral genomes (Koonin, 1993).

The currently accepted model for flavivirus RNA replication is a semiconservative one. Input genome RNA must first be copied to form full-length minus strands in order to generate stable double-stranded replicative forms (RF). Free minus-strand genomic RNA is not detected (Chu and Westaway, 1985; Khromykh and Westaway, 1997), and the plus and minus RNA strands in the RF appear to be exact copies of each other, such that there are no overhanging unpaired nucleotides at the 3' termini of either strand (Wengler and Wengler, 1981). Plus-strand RNA synthesis is thought to proceed in replicative intermediates, using RFs as a template. No subgenomic virus-specific RNAs have been detected in flavivirus-infected cells, in support of the concept that the genomic plus-strand RNA is the only virus-specific mRNA. The RNA replication process appears to be localized to perinuclear membrane organelles, probably in complexes containing nascent RNA and NS proteins in association with cellular proteins (Chu and Westaway, 1985, 1987; Ng *et al.*, 1989; Westaway *et al.*, 1999).

Table I shows a list of the lengths of the respective 5' and 3' NCRs in representative flavivirus genomes. The total number of nucleotides in each genome is also shown for comparison. Flaviviruses have been speciated, and species have been subgrouped, by results of virus cross-neutralization assays (Calisher *et al.*, 1989). The results of such assays depend on antigenic differences among the respective envelope glycoproteins of the different viruses. Seven subgroups were thus defined,

but at least 15 viruses remained ungrouped, including YF virus. For example, the four serotypes of dengue viruses form one subgroup or “complex” among the mosquito-borne flaviviruses. Similarly, the Japanese encephalitis virus complex includes JE, West Nile (WN), and Murray Valley encephalitis (MVE) viruses. The tick-borne flaviviruses (e.g., TBE virus strains Sofjin and Neudorfl and Powassan virus, see [Table I](#)) form a complex that is serologically distinct from the mosquito-borne flaviviruses. Cell-Fusing Agent virus (CFA) was isolated from cultured mosquito cells and appears to have no serological relationship to other flaviviruses in the cross-neutralization assay, has no vector in nature, and causes no illness in humans or animals ([Cammissa-Parks *et al.*, 1992](#)).

Flaviviruses have also been grouped by phylogenetic analysis, based on sequence data obtained for a 1000-nucleotide cDNA generated from the 3' termini of the NS5 genes of several dozen isolates ([Kuno *et al.*, 1998](#)). The hierarchical levels of relatedness defined by this technique were organized in descending order as follows: “cluster,” “clade,” and “species,” where a clade was defined as a group of viruses that share 69% or higher pairwise nucleotide sequence identity among the members. Results of the analysis for grouping flaviviruses in clades were quite comparable to those obtained by the cross-neutralization assay as it was used to define antigenic complexes or subgroups, suggesting a parallel evolution of genome sequences and antigenic character. These findings will be seen to be relevant as we discuss differences in 5' and 3' NCR nucleotide sequences among viruses of different antigenic subgroups or clades.

Data for viruses listed in [Table I](#) show that the 5' NCRs of flavivirus RNAs are relatively short compared to the 5' NCRs of other positive-strand RNA viruses. Lengths of the 5' NCR sequences are between 89 nucleotides, for dengue serotype 2 (DEN2) strain New Guinea C (NGC), and 132 nucleotides, for TBE strain Sofjin. Tick-borne flaviviruses (strain Sofjin, strain Neudorfl, and Powassan) tend to have slightly longer 5' NCRs than mosquito-borne viruses. For comparison, the 5' NCR of a typical isolate of the hepacivirus, hepatitis C virus, is 341 nucleotides in length ([Fukushi *et al.*, 1994](#)), whereas that of a picornavirus, poliovirus type 1, is 742 nucleotides in length ([Rezapkin *et al.*, 1998](#)). This discrepancy in length of the 5' NCRs of these otherwise related positive-strand RNA virus genomes is due primarily to the fact that the hepatitis C and poliovirus 5' NCRs each contain an internal ribosome entry site (IRES) required for translation initiation ([Bergamini *et al.*, 2000](#); [Brown *et al.*, 1992](#); [Poyry *et al.*, 2001](#); [Wang *et al.*, 1995](#)). IRESs are fairly large tRNA-like structures composed of

hundreds of nucleotides. In contrast, translation of flavivirus RNAs is thought to initiate by ribosome scanning from the 5' cap structure (Ruiz-Linares, 1989) so possibly there is an adaptive advantage related to having a short 5' NCR.

DEN virus genomes had the shortest 3' NCRs among the selection of flavivirus genome sequences surveyed, ranging from 384 nucleotides for DEN4, strain 814669, to 466 nucleotides, for DEN1, strain WP74. The 3' NCRs in viruses of the JE group shown in Table I ranged from 573 to 631 nucleotides in length. The length of the YF vaccine strain 17D 3' NCR was in between that of the DEN and JE complex viruses at 511 nucleotides. The 3' NCRs of the TBE viruses, other than strains Neudorfl and Sofjin, vary greatly in length due to the presence of long poly(A) tracts in some genomes that are not present in others and due to deletions of sequences flanking the poly(A) tract within the variable region (see later for details). Many dozens of complete genome sequences of flavivirus have been entered in Genbank. In no case do the lengths of the 5' and 3' NCRs vary significantly from those noted for viruses listed in Table I.

II. CONSERVED LINEAR SEQUENCE FEATURES OF NCRs IN FLAVIVIRUS RNA

A. Complementary Nucleotides at the 5' and 3' Termini of NCRs

The first complete flavivirus genome sequence to be published was that of the YF virus, strain 17D (Rice *et al.*, 1985). The authors noted that the extreme 5'- and 3'-terminal sequences of the YF genome were partially homologous to those of the WN virus, which had been determined separately (Wengler and Wengler, 1981), and that for both genomes (i) the 5'-terminal two nucleotides of the plus-strand sequence (5'-AG) were complementary to the 3'-terminal two nucleotides (UC-3') and (ii) there was a short (five-nucleotide) region of identity between a segment near the 3' termini of plus-strand RNAs and the predicted 3' termini of minus-strand RNAs for each of the two virus genomes due to the presence of a complementary sequence in the 5' NCR of the plus strand (the sequence 5'-UGUGU in the 5' NCR and the sequence 5'-ACACA in the 3' NCR, as shown in Fig. 1). For the YF genome, the sequence ACACA initiated at the 13th nucleotide from the 3' end of the RNA and spanned nucleotides -13 through -9 (reading in the 3' to 5' direction is indicated by a minus sign), and for the WN genome the ACACA sequence spanned nucleotides -11 through -6 (Fig. 1).



FIG 1. Nucleotide sequences of 5' and 3' termini of a selection of mosquito-borne flavivirus positive-strand (+) genome RNAs are shown aligned with the nucleotide sequences of 3' and 5' termini of the respective minus (-) strands, predicted to result when the positive strand is used as a template to form a replicative intermediate during virus replication. Dotted lines indicate the intervening nucleotides of the respective full-length genomes. Flavivirus conserved 5'- and 3'-terminal dinucleotides and their complementary nucleotides are boxed. The semiconserved pentanucleotide sequence ACACA is shown in bold in the 3' termini of positive- and negative-strand RNAs, respectively, where present. Genome sequences are as follows: YF, yellow fever ([Rice et al., 1985](#)); WNV, West Nile ([Wengler and Wengler, 1981](#)); MVE, Murray Valley encephalitis (Genbank number NC00943); DEN3, dengue serotype 3 (Genbank number NC001475); KUN, Kunjin (Genbank numbers L24511 and L24512); and JE, Japanese encephalitis (Genbank number NC001437).

The identical five-nucleotide sequences near the 3' termini of both plus- and minus-strand RNAs were posited as recognition sites for the viral replicase that could be used in common during either positive- or negative-strand RNA synthesis ([Rice et al., 1985](#)).

An examination of a selection of genome sequences for other flaviviruses reveals that the complementary pairs at the ends of the RNA (5'-AG/UC-3') are completely conserved among mosquito- and tick-borne viruses. The CFA virus genome is unique among the flaviviruses in that it terminates with GC-3' and its 5'-terminal two nucleotides (5'-GU) are not complementary. The survey also revealed that the sequence ACACA and its complement are not well conserved in analogous positions in 3' and 5' NCRs, respectively, of flavivirus genomes ([Fig. 1](#)). For example, JE and Kunjin virus genomes do contain such sequences in the predicted locations, whereas the related MVE virus genome does not, and viruses of the DEN subgroup do not. Therefore, the significance of the "ACACA-type" sequences near the 3' termini of

plus and minus strands of the YF, WN, JE, KUN, and perhaps other flavivirus genomes must be regarded as uncertain, especially since the functionality of other better conserved segments of the NCRs in RNA replication has been demonstrated.

B. Cyclization Sequences

The most conserved linear sequence feature of the 5' termini of flavivirus genomes is the cyclization sequence (5'CS in Fig. 2; Hahn *et al.*, 1987; Khromykh *et al.*, 2001). The motif is so named because of the existence of a complementary sequence in the 3' NCR. Base pairing between 5' and 3' end sequences would in theory permit the formation of cyclized plus strands, perhaps during the initial phases of RNA replication. This conformation was initially predicted to permit the viral replicase to bind to both ends of the template RNA simultaneously in order to assure the generation of full-length copies (Hahn *et al.*, 1987). Similar "panhandle" structures are thought to form, for example, during the replication of alphavirus (Hsu *et al.*, 1973) and bunyavirus (Hewlett *et al.*, 1977) RNAs and of phage QB RNA (Blumenthal and Carmichael, 1979).

For mosquito-borne virus genomes, the 5' cyclization sequence is known as the "5'CS" [indicating the 8 nucleotide sequence UCAAUAUG, spanning nucleotides +137 through +144 of the Kunjin virus (KUN) genome, for example, where the plus sign indicates nucleotides are numbered from the 5' terminus]. The 5'CS actually lies within the ORF, about 34 to 40 nucleotides downstream from the start codon (Fig. 2). For tick-borne virus genomes, two potential 5' cyclization sequences have been identified (Khromykh *et al.*, 2001; Mandl *et al.*, 1993), here designated "C1" and "C2" (Fig. 3). C1 and C2 differ markedly in linear sequence from each other and from that of the 5'CS in mosquito-borne virus genomes (Table II). Tick-borne virus 5'-NCRs contain a very short ORF that is terminated with a stop codon within the NCR (Fig. 3). The most 5'-terminal potential cyclization sequence (C1; the 11 nucleotide sequence GGAGAACAAGA, spanning nucleotides +115 through +125 for TBE viruses) lies downstream from this mini-ORF and wholly within the 5' NCR (Mandl *et al.*, 1993). C2 (the 11 nucleotide sequence GGGGCGGUCCC, spanning nucleotides +164 through +174 for TBE) lies downstream from C1, within the true ORF, in a position analogous to that of the 5'CS in mosquito-borne virus genomes (Khromykh *et al.*, 2001).

The 3' extreme of the 8 nucleotide 3' cyclization sequence (CYC; Fig. 2) in mosquito-borne virus RNAs is variably located 99 to 112

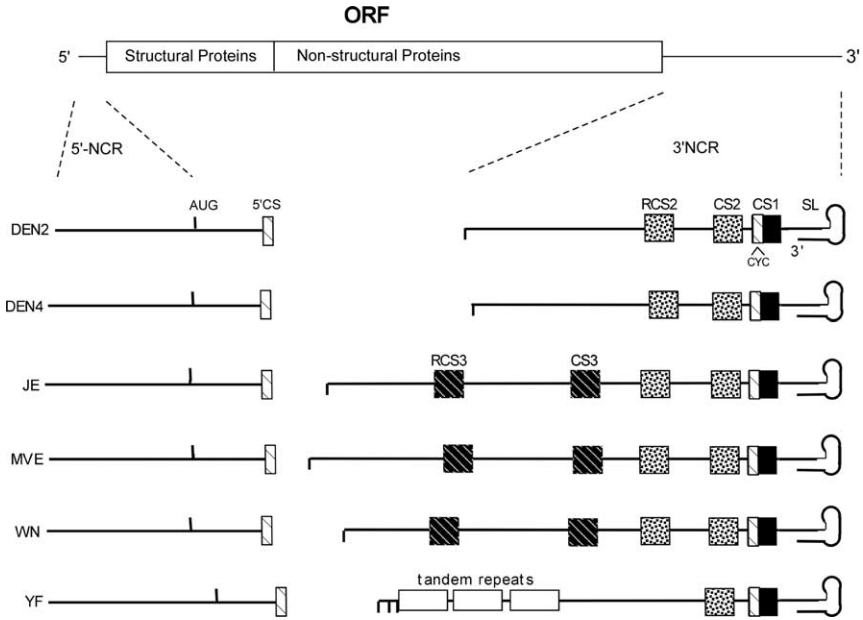


FIG 2. Some nucleotide sequence features of 5' and 3' NCRs of mosquito-borne flavivirus genomes are represented. Horizontal lines represent nucleotide sequences of DEN2, DEN4, JE, MVE, WN, and YF genome 5' NCRs (aligned in vertical order on the left) and 3' NCRs (aligned in vertical order on the right). Upward vertical ticks in the 5' NCR indicate the location of the 5'-most start codon for translation of the single ORF in the flavivirus genome. Downward vertical ticks indicate the location(s) of translational stop codons at the 5' ends of 3' NCRs. The flavivirus-conserved stem and loop secondary structure predicted to form at the 3' end of all genomes is indicated by a loop at the end of the linear sequence (SL). The relative locations of the conserved eight nucleotide cyclization sequences, 5'CS and CYC, respectively, at the 5' and 3' ends of the genome are indicated by open boxes with cross-hatching. The relative location of the conserved sequence, CS2, in all genomes is indicated by a stippled box, and the location of the CS2 tandem repeat sequence, RCS2, in all but the YF genome is indicated by a second stippled box upstream from CS2. Similarly, the relative locations of the conserved tandem repeat sequences, CS3 and RCS3, are indicated by black, cross-hatched boxes in genomes of viruses of the JE subgroup. The locations of three tandem repeat sequences in the upstream portion of the YF genome 3' NCR are indicated by open rectangles.

nucleotides from the 3' terminus of the genome (Table II). This locus is just upstream from the conserved 3' terminal stem and loop structure common to all flavivirus genomes (see later). The CYC sequence is included within the upstream portion of a longer sequence element, "CS1" (Figs. 2 and 4) that is also semiconserved among mosquito-borne virus genomes (Fig. 5). For tick-borne viruses, the 11 nucleotide 3'

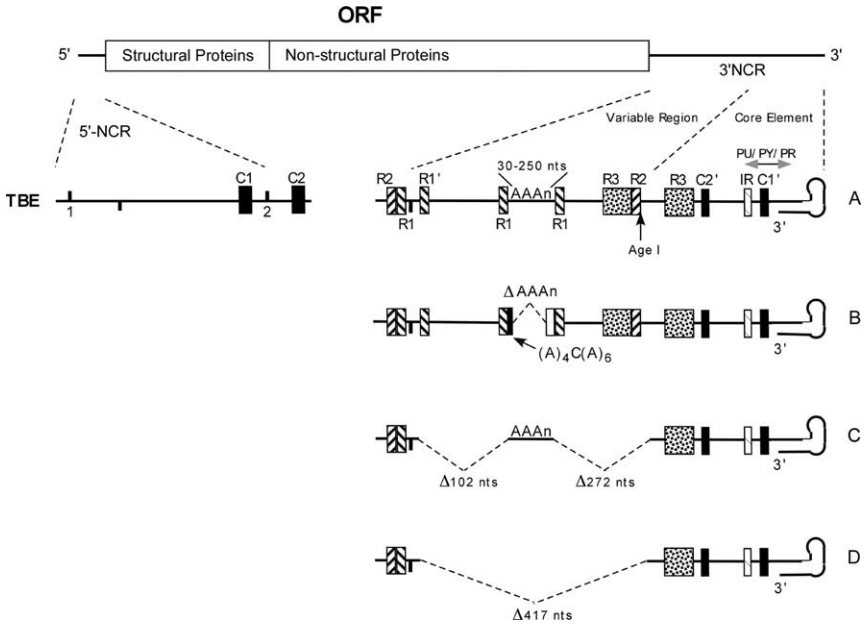


FIG 3. Some nucleotide sequence features of 5' and 3' NCRs of TBE virus strain genomes are represented. The 5' NCR of TBE virus is represented on the left as a horizontal line. The start codon for translation of a very short ORF in the 5' NCR is represented by a single upward vertical tick, numbered one (1); the start codon for translation of the long ORF in TBE genomes is represented by an upward tick, numbered two (2). Stop codons for the short ORF in the 5' NCR and for the long ORF at the 5' end of the 3' NCR are indicated by a downward vertical tick. Four of six different possible nucleotide sequence motifs for 3' NCRs of TBE viruses are indicated on the right as horizontal lines (adapted from Wallner *et al.*, 1995). Motif A, strains Neudorf1 and 263; B, strain Ljubljana I; C, strain 132; and D, strain RK1424. The flavivirus-conserved 3' stem and loop structure is indicated by a loop in the horizontal line and is labeled "SL." The location of the restriction endonuclease site *AgeI* in TBE genomic DNA is indicated by an arrow defining the boundary between the variable region and the core element in the strain Neudorf1 3' NCR. Black boxes indicate the relative locations of cyclization sequences C1 and C2 in the 5' NCR and C1' and C2' in the 3' NCR. Relative locations of TBE virus-conserved repeat sequences R1, R1', R2, and R3 in the 3' NCR are indicated by cross-hatched or stippled boxes. IR, conserved inverted repeat. PU/PY/PR, conserved homopurine, pyrimidine, and purine-rich segments in 3' NCR RNA, respectively. Dotted lines indicate the locations of deletions in the genomes of some TBE strains relative to the motif found in strain Neudorf1 RNA. The notation "AAA_n" indicates the location of a poly(A)tract in the strains Neudorf1, 263, and 132 genomes. The location of a severely truncated poly(A)tract in the strain Ljubljana 1 genome is specifically indicated by an arrow.

TABLE II
RELATIVE LOCATIONS OF PROPOSED CYCLIZATION SEQUENCES (CS) IN GENOMIC RNA OF *FLAVIVIRUS* SPECIES

Virus	Translation initiation ^a	Proposed CS and locations ^b	Genbank accession No.	Reference for CS
Mosquito borne		5'-UCAAUAUG 3'-AGUUAUAC		Hahn et al. (1987) ; 5'CS/CYC in Fig. 2
Kunjing	97	5' (137) (- 104) 3'	D00246, L24511, L24512	Khromykh et al. (2001)
JE	96	5' (136) (- 111) 3'	M10370	Fig. 2
MVE	96	5' (136) (- 111) 3'	NC000943	Hahn et al. (1987)
WN	97	5' (137) (- 105) 3'	M12294	Hahn et al. (1987)
DEN1	81	5' (118) (- 103) 3'	M87512	Fig. 2
DEN2	97	5' (134) (- 103) 3'	M20558	Fig. 2
DEN3	94	5' (132) (- 103) 3'	M93130	Fig. 2
DEN4	102	5' (136) (- 99) 3'	M14931	Fig. 2
YF	119	5' (156) (- 112) 3'	NC002031	Fig. 2
Cell-fusing agent	114	5' (169) CCCC <u>GUUC</u> GG 3' GGGGCA-GGUC (- 134)	M91671	Cammisa-Parks et al. (1992)
	114	5' (149) GCCAGGG 3' CGGUCCC (- 471)	M91671	CS "D"/CS"C" in Khromykh et al. (2001)

Tick borne					Khromykh et al. (2001)
TBE	133	5' (164) GGGGC ^{GG} UCCC 3' CCCCC _G AGGGG (- 193)	U27495		C2/C2' in Fig. 3
Powassan	112	5' (136) GGGGGC ^{GG} UCC 3' CCCCCG-AGG (- 189)	L06436		C2/C2' in Fig. 3
Louping ill	130	5' (160) GGGGGC ^{GG} UCC 3' CCCCC _G AGG (- 193)	Y07863		C2/C2' in Fig. 3
TBE	133	5' (115) GGAGAACAAGA 3' CCUCUUGUUCU (- 81)	U27495		Mandl et al. (1993) C1/C1' in Fig. 3
Powassan	112	5' (88) GGAGAACAAGA 3' CCUCUUGUUCU (- 81)	L06436		Mandl et al. (1993)
Louping ill	130	5' (112) GGAGAACAAGA 3' CCUCUUGUUCU (- 81)	Y07863		Gritsun et al. (1997)

^a Numbered position of first nucleotide in the AUG codon.

^b The 5' and 3' CS (cyclization sequence) shown are base paired. The first nucleotide of each CS is shown in bold. Additional but nonconserved 5'/3' *trans* complementary basepairs flanking the CSs are not shown. Positions of the first nucleotide of each CS in relation to the 5' or the 3' terminus (indicated by a minus sign) of each genome are shown in bold within parentheses. Reference is made to [Figs. 2 and 3](#) for mapping of the CSs in mosquito- and tick-borne virus genomes.

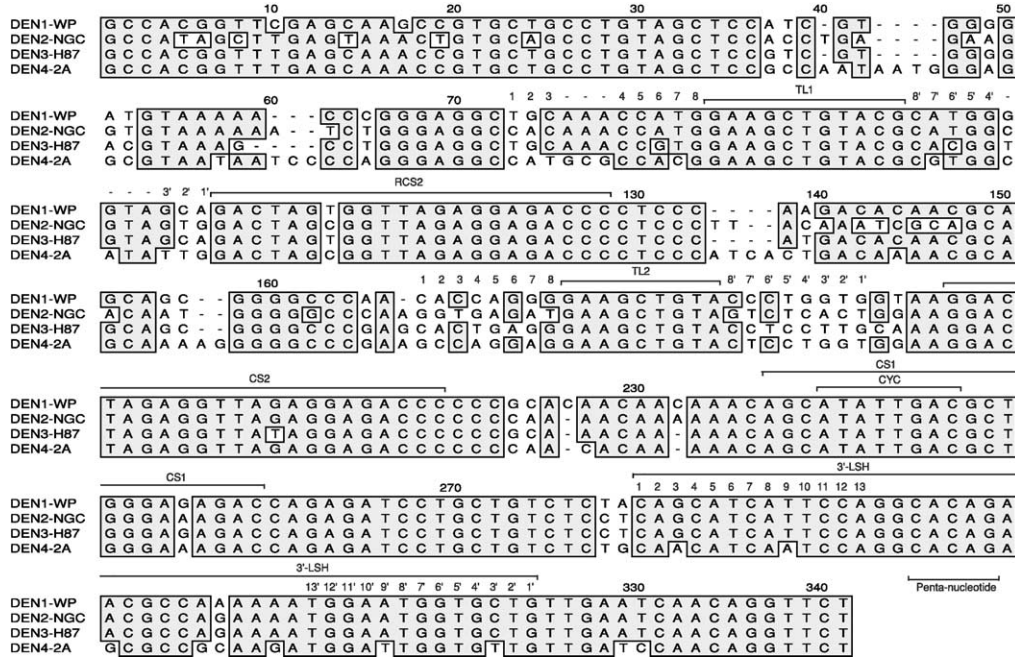


FIG 4. Nucleotide sequences of 3' NCRs of dengue virus genomes representing each of the four serotypes are shown. Shaded areas indicate sequence homology. Dashes indicate gaps introduced into a sequence in order to maximize its alignment with the other sequences. Nucleotides that form the long stem and loop within the flavivirus-conserved 3'-SL (see text) are indicated by a horizontal line above sequence data, labeled 3'-LSH. The location of the flavivirus-conserved pentanucleotide sequence, 5'-CACAG-3', within the 3'-SL, is indicated by a line below sequence data. The conserved sequences, CS1, the cyclization sequence, CS2, and RCS2 are indicated by horizontal lines above relevant sequence data. The conserved

CS1:

AGC**CAU**AUUGACACCUGGGAAAAGAC **MVE**
 AGC**CAU**AUUGACACCUGGGAU. AGAC **WN**
 AGC**CAU**AUUGACACCUGGGAAUAGAC **JE**
 AGC**CAU**AUUGACGC. UGGGAG. AGAC **DEN1**
 AGC**CAU**AUUGACGC. UGGGAG. AGAC **DEN3**
 ACC**CAU**AUUGACCCAGGGAA. AGAC **YF**

CS2:

GGACUAGAGGUUAGAGGAGACCC **MVE, WN, JE**
 GGACUAGAGGUUAGAGGAGACCC **DEN1**
 GGACUAGAGGUUAUAGGAGACCC **DEN3**
 GGUCUAGAGGUUAGAGGAGACCU **YF**

CS3 (1) and RCS3 (2) in MVE (JE subgroup) 3'NCR:

1 CCCCAGGAGGACUGGGUUACCAAGCUG
 2 CCCCAGGAGGACUGGGUAACAAGCCG

Upstream Tandem Repeats in the YF 17D 3' NCR:

1 ***UAA**CCGGGAUACAAACCACGGGUGGAGAACCGGACUCCCCACA
 2 GAAACCGGGAUUAAAAACACGGCUGGAGAACCGGCUCCCCACU
 3 GAAACCGGGAUAAAAACUACGGAUGGAGAACCGGACUCCCACACA

Fig 5. Nucleotide sequences of some conserved motifs in 3' NCRs of mosquito-borne flavivirus genomes. Deletions in CS1 with respect to the sequence for MVE (Hahn *et al.*, 1987) are indicated by a period. Nucleotide substitutions with respect to the sequences of CS1 and CS2 in the MVE genome are indicated by underlining the substituted nucleotides. Nucleotides comprising the 3' cyclization sequence (CYC) in CS1 are shown in bold print. The stop codon that comprises the first three nucleotides of the most upstream tandem repeat in the YF genome 3' NCR (Rice *et al.*, 1985) is also shown in bold and indicated by an asterisk. Nucleotide sequence data are from Genbank (Table I) or from the aforementioned references.

sequence complementary to C1 [C1', the sequence CCUCUUGUUCU (Mandl *et al.*, 1993)] is localized to the most upstream portion of nucleotides required to form the conserved stem-loop, spanning nucleotides -81 to -71 (Table II; Fig. 3). The 10 nucleotide 3' segment complementary to C2 [C2', the sequence CCCCGGAGGG (Khromykh *et al.*, 2001)] is located further upstream from the 3' terminus of the genome, spanning nucleotides -184 to -193, for example, in the TBE genomes.

Hahn and colleagues (1987) further predicted that the double-stranded panhandle structure potentially resulting from base pairing of the 5' and 3' cyclization sequences in the mosquito-borne virus genomes that were under analysis could be further stabilized by the presence of additional nonconserved complementary nucleotides



small stem and loop structures, TL1 and TL2 (Proutski *et al.*, 1997b), are also so indicated. Nucleotides presumed to base pair in the formation of the long stem in the 3'-SL and in the formation of the short TL1 and TL2 stems are numbered.

upstream from the 5'CS at the 5' end of the genome in a pyrimidine-rich segment of the genome and downstream from the 3' CYC sequence at the 3' end of the genome in a purine-rich region. Thus there were a total of 11 or 12 contiguous base pairings possible for cyclization of YF, MVE, WN, and DEN2 RNAs, resulting in a predicted thermal stability of from -9.1 kcal (for MVE RNA) to -12.3 kcal (for YF RNA). This was deemed to be sufficient free energy to cyclize flavivirus RNAs based on the previous observation that alphavirus RNAs could cyclize under physiological conditions (Hsu *et al.*, 1973) and that the free energy of cyclization of alphavirus RNA measured thermodynamically was determined to be -13.5 kcal at 25°C (Frey *et al.*, 1979). Additional but non-contiguous complementarity further upstream from the 5'CS and downstream from the 3' CYC sequence was predicted to stabilize the panhandle further for some of the RNAs, resulting in an increase of potential thermal stability to -20.5 , -25.5 , and -33.7 kcal, respectively, for the panhandle form of the WN, MVE, and YF genomes (Hahn *et al.*, 1987). The predicted thermal stability of the double-stranded regions potentially formed by base pairing of C1 with C1' and of C2 with C2' in the TBE virus genome was -19.8 and -30.9 kcal, respectively (Khromykh *et al.*, 2001). Potential cyclization sequences have also been identified in the CFA genome (Cammisa-Parks *et al.*, 1992; Table II).

Several bits of evidence demonstrate that base pairing of the cyclization sequences is required for RNA replication. First, in a study of the requirements for nucleotide sequences of the DEN4 3' NCR for virus replication, Men and co-workers (1996) created a series of mutant DEN genomic RNAs by the transcription of DEN4 infectious DNAs containing internal deletions centered about a unique *ApaI* restriction endonuclease site located between the conserved tandem repeat sequences CS2 and RCS2 (see later) in wild-type DEN4 DNA (Figs. 2 and 4). Eight mutant RNAs containing deletions ranging in size from 30 to 262 nucleotides, in which all or portions of CS2, RCS2, and CS1 were deleted, were evaluated for their infectivity. Although all the RNAs gave rise to viruses that were reduced in replication competence compared to wild type, only one RNA (mutant 3'd172-83) had a lethal phenotype. The deletion in this RNA was unique in that it included all of the 23 nucleotide CS1 segment in DEN4 RNA, whereas the maximum downstream extent of any of the other seven deletions spared the entirety of CS1. These results demonstrated that CS1, including the 8-nucleotide cyclization sequence, was indispensable for virus replication.

Second, Khromykh and Westaway (1997) evaluated the replication competence of several in-frame internal deletion mutant RNAs derived from a full-length infectious DNA copy of the Kunjin virus genome

(“replicons”). They demonstrated that an RNA with a deletion starting at codon 20 of the capsid gene and extending to include downstream sequences encoding the premembrane protein and the entire ectodomain of the envelope glycoprotein (mutant C20rep) replicated efficiently in hamster kidney (BHK) cells, whereas an otherwise identical deletion mutation, in which the 5' end of the deletion extended upstream to codon 2 of the capsid gene (mutant C2rep), was lethal for RNA replication. Similarly, another mutant RNA in which the capsid start codon and sequences downstream encoding the remainder of capsid, all of the premembrane protein, and the amino-terminal majority of the envelope protein (mutant Δ CME) was also replication defective. C20rep RNA contained the 5'CS for the Kunjin genome, whereas the mutants C2rep and Δ CME RNAs did not. This suggested that either the RNA sequence (including the 5'CS) within the ORF encoding the amino terminus of capsid or translation of the amino terminus of capsid was necessary for RNA replication. The results of follow-up experiments favored the former hypothesis, that nucleotide sequences comprising the 5' terminus of the ORF, including the 5'CS, were required for RNA replication.

Subsequent studies specifically addressed the requirement for both the 5'CS and the 3'CYC sequence for RNA replication. [You and Padmanabhan \(1999\)](#) reported on an *in vitro* assay for RNA-dependent RNA polymerase (RdRP) activity in DEN2-infected mosquito cell extracts using exogenous subgenomic RNAs containing the 5'-terminal 230 nucleotides and/or the 3'-terminal 373 nucleotides of the DEN2 genome as templates. 5'-terminal DEN2 sequences containing the 5'CS sequence, either linked covalently to the 3'-terminal DEN2 RNA sequence or supplied in *trans*, were required for replication to initiate on the 3'-terminal RNA segment. Mutagenesis of either the 5' or the 3' cyclization sequence such that base pairing between them was abrogated resulted in total loss of *in vitro* RdRP activity, but when both sequences were mutagenized such that the mutant cyclization sequences were complementary, RdRP activity was detected. This demonstrated that cyclization sequences were required for synthesis of a covalently closed double-stranded hairpin molecule *in vitro*. Free progeny RNA molecules were not detected in this system. However, [Khromykh and colleagues \(2001\)](#) addressed the relevant question *in vivo* using the KUN replicon system. They initially created a 5'CS deletion mutation in KUN replicon RNA and demonstrated that the resulting RNA was no longer competent for replication. Next, they created substitution mutations in either the 5'CS or the 3'CYC sequence in order to disrupt base pairing between them. As seen in the *in vitro* study, abrogation of

their complementarity by mutagenesis of either cyclization sequence was also lethal for RNA replication. However, when both cyclization sequences were mutagenized with respect to the wild-type sequence such that their capacity to base pair was maintained, RNA replication, including *de novo* minus and plus strand synthesis, was restored.

These results clearly demonstrated that complementarity, but not the respective wild-type nucleotides, of the 5'CS and 3'CYC sequences was necessary for flavivirus RNA replication to proceed. Future studies will likely be designed to prove that the input plus strand RNA actually does cyclize *in vivo* prior to or as a requirement for initiation of RNA replication and to determine how localization of the cyclization sequences, at nearly equal distances from the 5' and 3' termini of the genome, is related to their function in mediating the initiation of RNA synthesis. Moreover, it will be of interest to learn whether both pairs of complementary cyclization sequences present in the genomes of tick-borne viruses are required for the replication of tick-borne virus RNA or whether only one complementary pair is sufficient. Cellular and viral proteins that participate in the cyclization process are also yet to be identified. For further discussion of a model for panhandle formation related to additional complementarity of nonconserved sequences in the 5' and 3' NCRs (Khromykh *et al.*, 2001), see later.

C. Additional Conserved Linear Sequences in 3' NCRs of Mosquito-Borne Flaviviruses

1. The Pentanucleotide

The 3' NCRs of mosquito-borne flaviviruses contain a number of conserved sequences that are localized primarily to the 3'-terminal portion of this region of the genome. The most 3'-terminal of these sequences is the pentanucleotide 5'-CACAG, which forms part of an unpaired region forming a closed loop within the conserved stem and loop structure formed by the 3'-terminal ~95 nucleotides of all flavivirus genomes (Hahn *et al.*, 1987). For example, in DEN virus genomes, the pentanucleotide sequence occurs between nucleotides -43 and -47 (Figs. 4 and 6). This sequence and its location in the genome are completely conserved among mosquito- and tick-borne flaviviruses for which sequence data are available. It is also incompletely conserved in the 3' stem and loop structure predicted to be present in CFA genome RNA; the fourth nucleotide is C instead of A (Cammisa-Parks *et al.*, 1992). The functional requirement for the conserved pentanucleotide sequence in the virus life cycle is not known, but it has been suggested that the pentanucleotide may play an indispensable role in the binding of cellular or

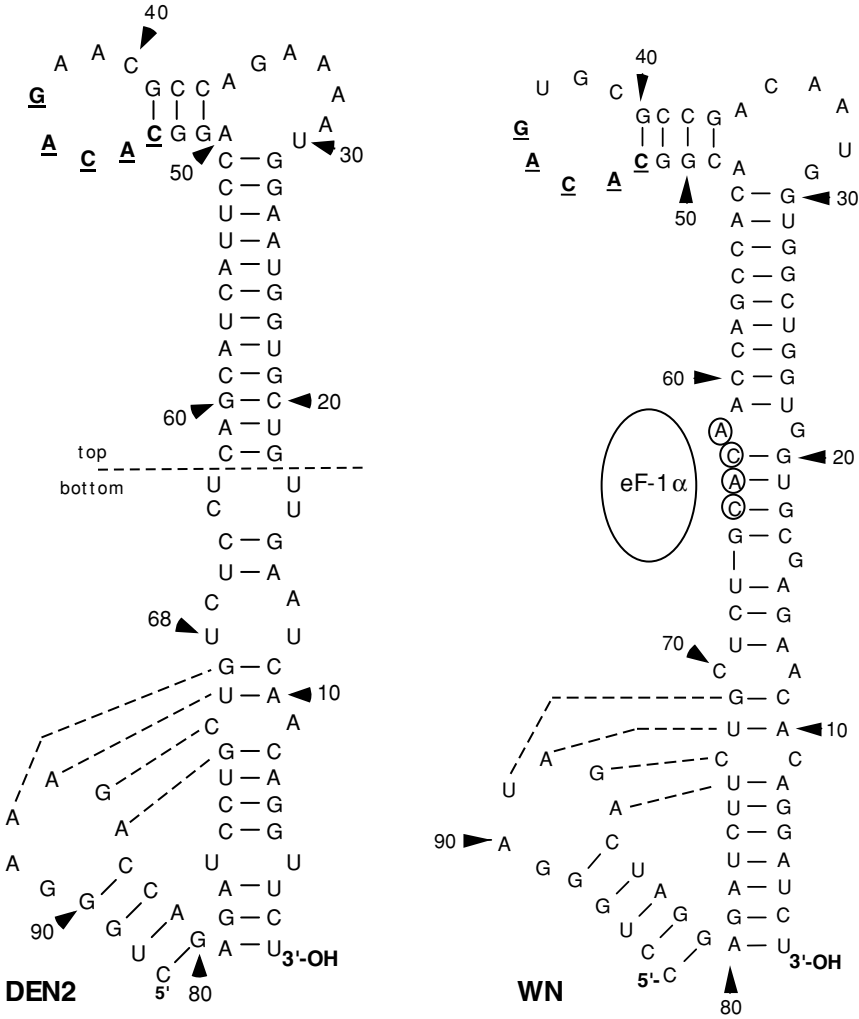


FIG 6. Nucleotide sequences of 3'-SLs in WN strain E101 and DEN2 strain NGC genomes (Zeng *et al.*, 1998) are shown in the expected conformation and numbered from the 3' end of each genome. The flavivirus-conserved pentanucleotide sequence 5'-CACAG-3' is indicated in bold underlined type. The "top" and "bottom" portion of the DEN2 3'-SL as designated in a previous study with respect to the WN 3'-SL is shown by a labeled horizontal dashed line (Zeng *et al.*, 1998). Essential nucleotides of the putative binding site for the translation elongation factor, eF-1 α , to the WN 3'-SL are circled (Blackwell and Brinton, 1997). Nucleotides potentially involved in pseudoknot formation in both DEN2 and WN 3'-SLs are linked by dashed lines (Shi *et al.*, 1996a).

viral proteins to the conserved 3' stem and loop structure (see later) during RNA replication (Khromykh *et al.*, 2001).

2. CS1

The next upstream conserved sequence in mosquito-borne flavivirus RNA is CS1, which has already been mentioned earlier in describing localization of the CYC sequence (Figs. 2, 4, and 5). The CS1 sequence is not present in 3' NCRs of tick-borne flavivirus genomes, although, as will be shown, the characteristic 3' cyclization sequence(s) in those genomes also occurs in the context of a larger conserved sequence motif (Wallner *et al.*, 1995). CS1, as originally described by Hahn and colleagues (1987), consisted of a 25 nucleotide segment in the MVE 3' NCR that was nearly but not totally conserved among viruses of the same subgroup and between subgroups. For example, CS1 for the MVE genome differs from that of the WN genome by one nucleotide substitution and the deletion of a single nucleotide, whereas CS1 in the JE virus genome differs from that of MVE by only a single nucleotide substitution. The sequence in the MVE genome differs from that found in the DEN1 and DEN3 genomes in length in that the DEN sequences appear to contain two discrete single nucleotide deletions and two discrete single nucleotide A to G substitution mutations with respect to the reference. DEN2 and DEN4 genomes contain the identical deletion mutations but only one of the two substitution mutations with respect to the MVE sequence. The CS1 sequence in the YF virus genome contains three substitution mutations and one deletion with respect to the MVE sequence. The 8-nucleotide CYC sequence is contained in the upstream half of CS1, between nucleotides -95 and -102 in DEN genomes. The CYC sequence itself is completely conserved among these and other mosquito-borne virus genomes.

3. CS2 and RCS2 and Analogous Upstream Tandem Repeat Sequences

The conserved tandem repeat sequences CS2 and RCS2 were also initially described by Hahn and co-workers (1987). CS2 and RCS2 are not found in 3' NCRs of tick-borne flavivirus genomes. For mosquito-borne viruses, CS2 is located closest to the 3' terminus of the genome, 12 to 22 nucleotides upstream from CS1 in the DEN, YF, and JE subgroup genomes (Figs. 2, 4, and 5). For the Genbank genome sequences of individual strains surveyed, the 20 nucleotide sequence of CS2 is completely conserved for viruses in the JE subgroup (MVE, WN, and KUN) and among all four serotypes of DEN viruses, except for a single internal nucleotide substitution (G to U) found in the DEN3 genome. The CS2 sequence in the YF 17D genome (Rice *et al.*, 1985) deviated from that

seen in genomes of viruses of the JE subgroup by two nucleotide substitutions, resulting in A to U and C to U changes in the sequence.

In the MVE genome, and similarly in genomes of the other JE subgroup viruses, RCS2 is 51 nucleotides upstream from the 5' terminus of CS2 and represents an exact duplication of the 20-nucleotide sequence of CS2. In the DEN genome sequences, RCS2 is located 62 (for DEN1) to 68 nucleotides (for DEN4) upstream from CS2 (see [Fig. 4](#)). CS2 is not repeated upstream in the YF genome. Instead, the YF genome contains three unique tandem repeats at the 5' terminus of its 3' NCR. The first of these repeats is 40 nucleotides long in the YF strain 17D genome, and it initiates downstream from the third of three in-frame stop codons found at the 3' terminus of the YF 17D ORF ([Figs. 2 and 5](#)). The second and third iterations of the repeat are 44 nucleotides in length. Save for the fact that the second repeat is 4 nucleotides longer than the first, there are only 5 nucleotides different between the two segments. Similarly, the third repeat differs from the first by only 4 nucleotides in 40. The second repeat initiates 4 nucleotides downstream from the 3' terminus of the first one, and the third repeat in turn initiates 6 nucleotides downstream from the 3' terminus of the second.

Not all YF virus genomes contain an analogous set of three tandem repeat sequences, as found in vaccine strain 17D. A study of partial nucleotide sequence data for genomes of 13 YF field isolates ([Wang *et al.*, 1996](#)) revealed that only the West African strain genomes contained a 511-nucleotide 3' NCR, including three tandem repeat sequences. Genomes of strains from central and east Africa and from South America had shorter 3' NCRs (443 to 469 nucleotides) due to the lower number of YF-specific repeat sequences. Central and east African strains had only two YF-specific repeats, and those from South America had only one copy of the "repeat" sequence. On the basis of overall nucleotide sequence relatedness, it was speculated that duplication of the repeated sequence took place in west Africa after introduction of the YF virus into South America. The resulting variation in lengths of 3' NCRs of the genomes of YF virus isolates, due to duplication and/or deletion of redundant nucleotide sequences within the 5' proximal portion of the 3' NCR, is reminiscent of that observed among the genomes of different strains of TBE virus (see later).

The JE subgroup viruses have an additional pair of repeats upstream from CS2 and RCS2 in the 3'NCR (CS3 and RCS3 in [Figs. 2 and 4](#)). For the MVE genome ([Hahn *et al.*, 1987](#)), these sequences are each 28 nucleotides in length and are situated 120 nucleotides apart in the MVE 3'NCR. The most upstream of these repeat segments has its 5' terminus 153 nucleotides downstream from the stop codon in the

MVE ORF. There are only three nucleotides different between the sequences of CS3 and RCS3 in the MVE 3' NCR.

4. Function of Conserved Tandem Repeat Sequences in Mosquito-Borne Flavivirus Replication

Relatively little is known regarding this subject. Perhaps more work will be done in this area in the future, as the use of full-length infectious DNAs to effect site-directed mutagenesis of the flavivirus genome has become more common. However, it is clear that deletion of conserved repeat sequences alters virus growth properties. As mentioned previously, [Men and co-workers \(1996\)](#) created a series of eight different mutant dengue genomic RNAs derived from DEN4-infectious DNAs containing internal deletions centered about a unique *ApaI* restriction endonuclease site located at nucleotide -171 in the DEN4 genome, about midway between CS2 and RCS2 ([Figs. 2 and 4](#)). Thus, all or portions of CS2 and/or RCS2 were deleted in six of the resulting seven viable deletion mutant viruses. (The eighth mutant RNA contained a deletion of CS1, which proved to be lethal, as discussed previously.) The maximum 3' boundary of any deletion that gave rise to virus was 113 nucleotides from the 3' end of the genome, preserving CS1 and the 3' stem and loop (SL). One DEN4 mutant containing a 121-nucleotide deletion that included all of RCS2 and most of the upstream portion of the 3' NCR (mutant 3'd303-183) was only slightly impaired for replication in cultured monkey kidney cells compared to wild-type virus. In contrast, a mutant containing a 60-nucleotide deletion that included almost all of CS2 (mutant 3'd172-113) was reduced in peak titer by about 10-fold with respect to wild type in both experiments. This suggested that CS2 may be more critical for viability than RCS2. Interestingly, a mutant that contained a deletion of only 30 nucleotides (3'd172-143) between CS2 and RCS2, but sparing both conserved sequences, was more impaired for replication than either mutant 3'd172-113 or 3'd303-183 in that its peak titers in the two assays were reduced compared to wild type by about 100-fold. This indicates that nonconserved nucleotides between CS2 and RCS2 are important for virus viability, possibly because of a requirement for spacing between the two conserved sequences or possibly because this deletion removed a DEN and JE subgroup conserved SL structure (TL2) predicted by [Proutski and colleagues \(1997b\)](#). Mutant virus 3'd172-143 is currently under study as a potential DEN vaccine candidate ([Durbin et al., 2001](#)).

The DEN4 genome could tolerate deletion of more than half its 3' NCR, provided the deleted sequences were upstream from nucleotide

–113, without complete loss of viability. A deletion of 202 nucleotides, extending from the stop codon (at nucleotide –384) to nucleotide –183 (3'd384-183) produced a viable virus that formed plaques in monkey cells and was only a little more reduced in peak titer compared to the vaccine candidate, mutant 3'd172-143, in the two growth assays, and a 262-nucleotide deletion of all but 10 nucleotides between the stop codon and nucleotide –113 produced viable virus after transfection of mosquito cells. However, this virus did not form plaques in monkey cells.

Khromykh and Westaway (1997) addressed the requirement for CS3 and RCS3, found in the JE subgroup virus genomes, using the KUN replicon system. They started with replicon Δ ME, which was stable and replicated efficiently *in vivo*. A 76-nucleotide deletion (KUN nucleotides 10,422 to 10,499) was created in Δ ME DNA between the stop codon and the upstream boundary of RCS3 (Fig. 2) to create mutant Δ ME/76 RNA. A second mutant RNA derived from Δ ME DNA, Δ ME/352, contained a 352-nucleotide deletion of KUN nucleotides 10,422 to 10,775, which included all of RCS3 and CS3, nonconserved sequences between these two boxes, and additional nucleotides extending upstream to overlap the upstream boundary of mutant Δ ME/76 RNA. Whereas Δ ME/76 RNA replicated efficiently in BHK cells, comparable in level to that of the “wild-type” Δ ME replicon RNA, only very low-level replication of Δ ME/352 RNA was detected above the threshold of sensitivity of the assay. Unfortunately, data are lacking relative to the phenotype of either deletion mutation in the context of a replication-competent KUN virus genome. Proutski and colleagues (1997b) offered the hypothesis that the phenotype of mutant Δ ME/352 RNA was related to the deletion of nonconserved nucleotides that form a conserved secondary structure in all JE subgroup virus genomes.

D. Conserved Linear Sequence Features of 3' NCRs of the TBE Subgroup of Tick-Borne Flaviviruses

1. The Pentanucleotide

The pentanucleotide 5'-CACAG-3' is the only linear nucleotide sequence that is conserved in both mosquito- and tick-borne flavivirus genomes (see earlier discussion).

2. Features Unique to TBE Viruses

The absolute lengths of the 3' NCRs in TBE virus genomes vary much more than those of mosquito-borne viruses and even compared to those of other tick-borne viruses, such as Powassan virus

(Table I). For example, the 3' NCR in TBE strain Neudorfl virus is 764 nucleotides in length (Table I), whereas that of TBE strain RK 1424 virus is only 350 nucleotides in length (Fig. 3; Wallner *et al.*, 1995). Accumulated nucleotide sequence data also show that this variability is due to differences in length of the upstream portion of the 3' NCR; the most 3'-terminal 325 nucleotides are conserved in all strains and share a high degree of sequence identity. Wallner and co-workers (1995) referred to the conserved 3'-terminal segment as the "core element" of the 3' NCR and designated any upstream sequences as part of a "variable region." Their findings were remarkable in that the viruses under study are all strain variants of a single serotype. In contrast, the 3' NCRs of the four serotypes of DEN viruses, for example, are nevertheless highly conserved with respect to both length and sequence content, perhaps because the DEN 3' NCR is minimal in the functional sequence content required for a viable flavivirus (Fig. 2). Thus the DEN 3' NCR may be analogous to the core element of the 3' NCR in TBE virus genomes. A comparison of the conserved sequences in 3' NCRs of TBE virus genomes to those of the Powassan virus genome (Mandl *et al.*, 1993) revealed that the latter genome has conserved features similar to TBE strain Neudorfl genome (to be described), except that the poly(A) tract found in strain Neudorfl and other TBE genomes is not present in the Powassan virus genome.

The core element of the TBE 3' NCR appears to be sufficient for a viable TBE virus, as the 3' NCR of strain RK1424 virus is virtually devoid of a variable region (Fig. 3). Essential elements of the core are as follows. (1) The approximately 100 3'-terminal nucleotides of the core region are predicted to form the flavivirus-conserved SL structure, which has been mentioned previously and which is discussed in greater detail later. Among the TBE virus genomes, the nucleotide sequence within the 3' stem and loop varies only in segments that are not required for base pairing to form the stem region of the secondary structure, with the exception of the conserved CACAG pentanucleotide box. (2) There is a short inverted repeat sequence (IR in Fig. 3) about 150 nucleotides upstream from the 3' terminus. The sequence has a very high GC content and a correspondingly high predicted thermal stability of ~ 19 kcal. Therefore, Wallner and co-workers (1995) suggested that there is a high probability that the IR actually forms a hairpin structure *in vivo*. (3) Adjacent to the IR sequence and overlapping it by three nucleotides is a 20-nucleotide stretch of purine residues, extending toward the 3' terminus of the RNA. This is the longest homopurine (PU) sequence within the entire genome of the TBE strain Neudorfl virus, apart from the poly(A) tract

in the variable region of the 3' NCR. (4) A few nucleotides further downstream from the PU sequence is a 14-nucleotide homopyrimidine (PY) sequence. (5) Thirty nucleotides still further downstream from the PY box exists a second conserved 14-nucleotide pyrimidine-rich (PR) sequence (consisting of 13 pyrimidine residues and one purine residue). The 3' cyclization sequence, C1', lies within the PR box. The PU, PY, and PR boxes and of course the sequence of C1', among other features of the TBE 3' NCR, are semiconserved in the Powassan virus genome (Mandl *et al.*, 1993), suggesting their fundamental importance in replication. Homopurine and -pyrimidine stretches have been found to function in processes related to the regulation of translation (Avni *et al.*, 1994; Behe, 1995), stabilization of RNA (Czyzyk-Krzeska *et al.*, 1994), or in (retrovirus) RNA replication (Hungnes *et al.*, 1992). It is perhaps noteworthy that the PU, PY, and PR boxes are found at loci in the genome analogous to those occupied by CS2 and CS1 in the mosquito-borne virus genomes, although there is no significant sequence homology between TBE or Powassan virus-specific domains and those of mosquito-borne viruses.

The remainder of the core region in TBE virus 3' NCRs contains one copy of sequence R3. This is a 76-nucleotide segment. The 3' terminus of R3 occurs 45 nucleotides upstream from IR. R3 is a tandem repeat sequence in the genomes of virus strains Neudorfl, 263, Ljubljana I, and Aina, where the repeat is found upstream from the core element in the variable region of the respective 3' NCRs. However, R3 is not repeated in the variable regions of the genomes of virus strains Hypr, Crimea, and RK 1424 (Fig. 3).

The variable region contains additional tandem repeat sequences. R1 is a 26-nucleotide segment that constitutes the 3' terminus of the ORF, including the stop codon. Well-conserved R1 repeat sequences flank the poly(A) tracts in the 3' NCRs of strains Neudorfl, 263, and Ljubljana I genomes. R1' is a 26-nucleotide semiconserved repeat of the R1 sequence that is found 14 nucleotides downstream from the stop codon at the 3' terminus of the ORF. R1' contains an additional UAA stop codon that is in-frame with the one designated to be functional at the 3' terminus of the most upstream R1 segment. R1' was identified in the genomes of virus strains Neudorfl, 263, Ljubljana, I, Aina, Hypr, and Crimea but not in those of strains 132 and RK 1424. The designation "R2" identifies a conserved 26-nucleotide sequence upstream from R1 in the 3' terminus of the ORF. R2 is repeated once in the variable region of the 3' NCR, just downstream from R3 in the genomes of virus strains Neudorfl, 263, Ljubljana I, and Aina. An R2 repeat sequence was also identified in an analogous location in the 3' NCRs of virus

strains Hypr and Crimea, despite the lack of an R3 repeat sequence in the variable region of the 3' NCRs in those genomes.

The poly(A) tract in the variable region of the 3' NCR in the TBE virus genome was first described for strain Neudorfl virus (Mandl *et al.*, 1991) and was believed to constitute the 3' terminus of the entire genome. It was later shown to be an internal part of an otherwise more conventional 3' NCR sequence in the strain Neudorfl genome, and similar tracts were also identified in the genomes of strains 263 and 132 (Wallner *et al.*, 1995). The strain Ljubljana I genome contains an oligo(A) tract at the analogous locus in its 3' NCR, the sequence (A)₄C(A)₆ (Fig. 3). The poly(A) tracts in the strain Neudorfl genome appeared to be heterogeneous in length, ranging from 30 to 250 nucleotides, but the degree of heterogeneity was difficult to assess due to the possibility that some of it was created artifactually either by stuttering of reverse transcriptase and/or *Taq* polymerase used in generating sequence data. A full-length cDNA copy of the strain Neudorfl genome containing a poly(A) tract only 49 nucleotides in length produced infectious RNA *in vitro* (Mandl *et al.*, 1997).

Wallner and co-workers (1995) made an effort to define a mechanism for the evolution of the diverse lengths of the 3' NCRs in these closely related TBE virus strains. They attempted to relate their findings (1) to other strain-specific parameters, such as the year, geographic origin, or source of the virus isolate, (2) to a dendrogram showing evolutionary relationships among the strains, derived by comparing sequence data for a 375-nucleotide segment of the NS5 gene, and (3) to a second dendrogram derived by comparing sequence data for the core elements of the various 3' NCRs. These analyses failed to yield any workable hypothesis to explain their observations. However, the close agreement between the dendrogram obtained by comparing the sequences of the NS5 coding regions to that obtained by comparing those of the core elements suggested that the sequences evolved in parallel. This implied that the heterogeneity in length of the variable regions did not evolve by intragenic recombination. Additional consideration of the details of the results of the sequence analysis led to the conclusion that the observed diversity in lengths of the 3' NCRs in these genomes was most likely due to deletion events and that therefore the viruses were likely to have evolved from a common ancestor with a long 3' NCR, such as those found in the genomes of strains Neudorfl and 263. Furthermore, these "long" 3' NCRs were thought to have arisen from an ancestral virus with a "short" 3' NCR by duplication events, as the variable region consists in a large part of direct repeats of sequence elements found in the core element and in the ORF.

Other unique sequences in the variable region may have arisen by intragenic recombination over short segments within subdomains.

Size variation of the 3' NCRs among closely related strains of other positive-strand RNA viruses has also been observed. (1) Among alphaviruses the 3' NCR in the genome of an avirulent variant of Semliki Forest virus was 334 nucleotides longer than that of the wild-type variant. The additional sequences consisted largely of tandem repeats (Santagati *et al.*, 1994). (2) The 3' NCRs of pestiviruses are also known to contain conserved and variable regions. One strain of bovine viral diarrhea virus was found to have a 41-nucleotide deletion in the variable region of its 3' NCR as compared to two other strains of this virus (Deng and Brock, 1993). There was at least a loose inverse correlation of the variation in length of the 3' NCRs in TBE strains to their virulence; strains Neudorfl and 263 were less virulent in mice than strain Hypr (which had a relatively short 3' NCR). However, strain Ljubljana caused severe disease in humans, despite having a genome with a "long" 3' NCR (Fig. 3).

Subsequently, Mandl and co-workers (1998) investigated the spontaneous occurrence of deletions in the 3' NCR of strain Neudorfl viruses that had been passaged many times in cultured hamster kidney (BHK) cells or of infectious DNA-derived strain Neudorfl non-3' NCR mutant viruses that had been passaged in both BHK and suckling or adult mouse brain. They determined the 3' NCR nucleotide sequence for 14 mutant viruses thus derived. The frequency with which 3' NCR deletions arose appeared to correlate best with the initial number of passages of virus in BHK cells. The spontaneous internal deletions that occurred in the 3' NCR had the following characteristics. (1) All deletions affected the variable region exclusively. (2) The deletions affected any and all sequence elements of the variable region [from nucleotide 10,394 to nucleotide 10,811, which is 15 nucleotides downstream from the *Age*I restriction endonuclease site (at nucleotide 10,796) that defines the boundary between the core element and the variable region (Fig. 3)]. (3) The poly(A) tract was removed in all deletion events, although short homoadenosine sequences were often retained. (4) There was some tendency for the deletions to cluster around certain nucleotides. The three largest deletions detected in this study were all of more than 300 nucleotides, and all of these resulted in the removal of the R2 repeat adjacent to the *Age*I site, the upstream R3 repeat sequence, the intervening heterogeneous sequence between R3 and the nearest R1 repeat, all of the poly(A) tract, and all or most of the R1 repeat that flanks the poly(A) tract at its upstream boundary.

These same investigators (Mandl *et al.*, 1998) also created six 3' NCR deletion mutants in the context of the strain Neudorfl infectious DNA. The smallest of these mutations deleted all of the variable region, from the AgeI site to the stop codon in the ORF (Fig. 3). The remaining five larger deletions also initiated on their upstream ends at the stop codon and extended to varying lengths on the downstream end into the core element. The longest deletion included all of the 3' NCR from the stop codon to the PU tract. All of these deletion mutant DNAs gave rise to viable mutant viruses, except for the longest of the deletions, just described (which included in its extent both potential 3' cyclization sequences, C1' and C2', and the conserved IR sequence). Among the viable mutant viruses, only the one with the largest deletion, extending from the stop codon to the downstream end of R3 in the core element, was reduced in peak titer compared to all other wild-type and mutant viruses by about 100-fold in a growth curve. This result was in contrast to that observed for internal deletion mutations of the 3' NCR in the DEN4 virus genome, where all deletion mutants had, to varying degrees, a reduced capacity for replication in tissue culture. However, it could be argued that the short DEN genome 3' NCRs are already functionally similar to the TBE core element, such that deletion mutations invariably remove sequences required for fully efficient replication.

The viable mutant TBE viruses were all comparable to wild type in lethality after intracerebral infection of suckling mice, although there was a correlation between the sizes of the deletions and the survival times after infection. However, there were marked differences in lethality of the viruses for 5-week-old mice after subcutaneous inoculation. This difference in virulence was directly correlated with the size of the deletion in the 3' NCR, such that the mutant virus with the largest deletion was more than 10,000-fold less virulent than wild-type strain Neudorfl. In contrast, the virus containing a deletion limited only to the entire variable region of the wild-type strain Neudorfl 3' NCR was indistinguishable from its parent in each of these assays. Thus the attenuation phenotype could be attributed to deletions of portions of the core element only. The variable region did not appear to have any relevant function for the growth of TBE virus in cell culture or in mice, as demonstrated by the occurrence of spontaneous deletions that affected almost all parts of this segment of the 3' NCR. Additional support for this conclusion arose from the finding that the engineered strain Neudorfl deletion mutant virus that contained a deletion of the entire variable region exhibited no change in its biology compared to the wild-type strain Neudorfl virus. These results also raised the

possibility that TBE strains with “short” 3' NCRs had undergone spontaneous deletion mutations during the processes of virus isolation and normal additional passaging in tissue culture.

It remains unclear why the variable region sequences exist and persist in naturally occurring TBE isolates. There may be some adaptive advantage associated with retention of the variable region, related perhaps to replication in ticks. [Mandl and co-workers \(1998\)](#) noted that the core element of the TBE strain Neudorfl 3' NCR contains 10 predicted SL secondary structures ([Gritsun *et al.*, 1997](#); [Rauscher *et al.*, 1997](#); see later), including the IR sequence and the 3'-SL. They suggested that differences in virulence among mutant viruses might be related to the effects of deletions in altering or abrogating the more upstream of these conserved secondary structures.

Similar observations regarding the requirement for the core element in tick-borne virus genomes were made by [Pletnev \(2001\)](#). Four internal deletions were created in the 3' NCR of an infectious DNA for an egg passage-attenuated Langat virus, a tick-borne virus highly antigenically related to TBE viruses. These all had their upstream boundaries at the stop codon for the ORF and extended downstream 320, 374, 449, and 471 nucleotides, respectively. All of these deletions eliminated short upstream portions of the core element, and only the 320 nucleotide deletion yielded viable virus when deletion mutant RNA was transfected into susceptible cells. Thus deletion of as little as 54 additional nucleotides of the core element (the difference between the 320-nucleotide deletion and the 374-nucleotide deletion) was lethal, even though the 3' SL, C1', C2', and other more downstream features of the TBE 3' NCR ([Fig. 3](#)) were intact in the 374-nucleotide deletion mutant. The viable 3' NCR deletion mutant virus was more attenuated in mice than its already attenuated parent virus.

III. SECONDARY STRUCTURE OF NCRs

A. *Secondary Structure of the 5' NCR*

[Brinton and Dispoto \(1988\)](#) determined the nucleotide sequences of the 5' NCRs of WN virus and of seven different isolates of the mosquito-borne flavivirus, St. Louis encephalitis (SLE) virus. They then compared the sequences to those of YF, MVE, and DEN viruses. While only short regions within the 5' NCRs were conserved among different subgroups, significant sequence homology was observed among viruses of the same subgroup, and an almost complete conservation of nucleotide sequence was observed among different strains of the same

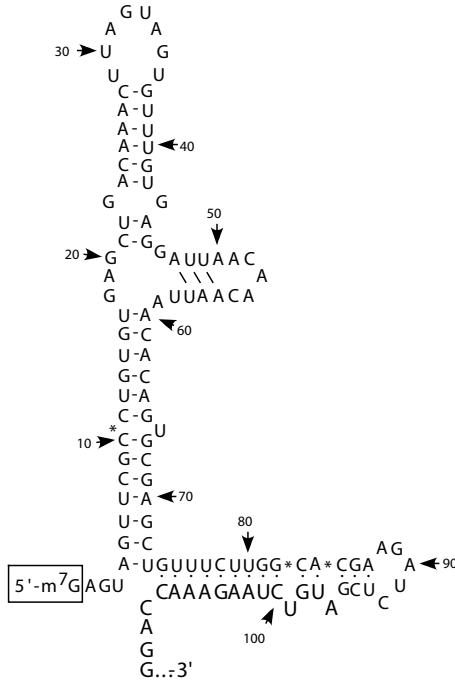


FIG 7. Proposed secondary structure of the WN 5' NCR (adapted from [Brinton and Disposto, 1988](#)). Nucleotides are numbered from the 5' terminus, excluding the cap, which is shown boxed. Nucleotides of the ORF, starting with nucleotide number 96, are shown in large, bold type (adapted from [Brinton and Disposto, 1988](#)).

virus, i.e., the seven isolates of SLE virus and two different strains of WN virus. For each 5' NCR sequence, secondary structures of similar size, shape, and predicted thermodynamic stability could be formed. In each case, the structures consisted of a stem with a small top loop and a larger side loop (Fig. 7). In most cases, a second short stem could be formed by 3'-terminal sequences of the 5' NCR, including the predicted start codon for the long ORF and downstream nucleotides in the ORF. The predicted thermal stability of the long stem structures ranged from -17.3 kcal (for the SLE sequence) to -27.8 kcal (for the WN strain E101 sequence, as shown). The conservation of this secondary structure in unrelated flavivirus genomes was taken as evidence of its possible importance in replication.

The relative importance of the predicted double-stranded regions in the 5' NCR was studied by deletion mutagenesis of an infectious DNA

copy of the DEN4 strain 814669 genome (Cahour *et al.*, 1995). The linear sequence of the 5' NCR is partially conserved between DEN and WN species (Fig. 7). For example, nucleotides 1–7, 11–16, 21–25, 55–60, and 69–77, numbering from the 5' terminus of the WN strain E101 sequence, are present in the 5' NCRs of both the DEN4 strain 814669 and the WN strain E101 genomes. Conservation of sequence and predicted secondary structure are much higher when any two DEN 5' NCRs are compared. Nine distinct deletions of from 5 to 12 nucleotides each were created between nucleotides 50 and 87 of the DEN4 sequence in the context of a DEN4 infectious DNA. In addition, two deletions of 5 and 26 nucleotides, respectively, were created downstream (involving nucleotides 94 to 98; d94-98) and upstream (d18-43) from the central targeted domain. Nucleotide 98 is four nucleotides upstream from the start codon of the DEN4 ORF. Mutants d55-60, d62-68, d69-75, and d82-87 contained deletions of nucleotides that were base paired in the long stem region of the 5' NCR, whereas mutants d50-54, d58-64, d76-81, d73-77, d76-87, and d94-98 contained deletions in the shorter stem or non-base-paired regions within the long stem. Mutant d18-43 contained a deletion of the upper loop and the adjacent upper portion of the long stem, as shown for the WN 5' NCR in Fig. 7. Five mutations were deemed lethal, in that infectious virus could not be recovered after transfection of cultured monkey kidney cells with RNA derived *in vitro* by the transcription of mutagenized DNAs. These were mutants d55-60, d58-64, d62-68, d69-75, and d94-98. Nucleotides 55 to 75 in the DEN4 genome are analogous to nucleotides 60 to 80 in the WN 5' NCR secondary structure shown in Fig. 7. In general, failure of the four contiguous deletion mutations (spanning DEN4 nucleotides 55 to 75) to yield viable viruses was ascribed to a requirement for those nucleotides for formation of the base pairs that constitute the bottom-most portion of the long stem and the short stem, which also requires nucleotides in the ORF for its formation.

The viable mutant viruses derived in this study were distinguishable from each other by plaque size in cultured mosquito and monkey cells, growth curves on these two cell types, relative levels of RNA in infected cells as assessed by dot-blot hybridization, and translation efficiency of the ORF in an *in vitro* system. Mutant d82-87 had a host range-restricted phenotype in that it uniquely failed to replicate in mosquito cells and in adult mosquitoes, apparently due to a defect in RNA synthesis. Translation of a portion of the DEN4 ORF containing the d82-87 5' NCR was also least efficient among all the mutants tested. However, the cell-free translation assay was not predictive of the viability of mutant viruses; RNAs containing some of the

lethal deletion mutations, e.g., d55-60, exhibited enhanced activity in translation compared to wild-type control RNA. Therefore, the dominant effect of deletion mutations on the viability of mutant viruses appeared to be at the level of RNA synthesis. The host range-restricted phenotype of mutant d82-87 also suggested that binding to RNA of host-specific factors, presumably cellular proteins, was a secondary determinant of the viability of mutant viruses.

Shi and co-workers (1996a) performed a computer-based analysis of the 3' negative-sense (–) RNA predicted to be synthesized in the process of replication of the 5' NCR of WN strain E101 virus. The presence of an SL secondary structure in 3' (–) NCR RNA, formed by the sequence complementary to the SL predicted to occur in the 5' NCR, was subsequently confirmed by spectroscopy, analysis of thermal melting curves, and selective RNase digestion of an *in vitro*-transcribed 75-nucleotide RNA representing the SL within the 96 nucleotide 3' NCR of the WN-negative strand. The 75-nucleotide transcript was then shown to bind specifically to four discrete proteins in three RNA–protein complexes in lysates of BHK cells and in lysates of embryonic fibroblasts derived from C3H/RV and C3H/He mice. C3H/RV mice are homozygous for the flavivirus resistance gene *Flv*, whereas C3H/He mice are susceptible to flavivirus infection (Groschel and Koprowski, 1965). Resistance of mice to the pathologic effects of encephalitic flavivirus infection had been shown previously to relate to reduced production of progeny plus strand RNAs at the cellular level. Because RNA–protein complexes formed in lysates of C3H/RV cells were indistinguishable in size from those formed in C3H/He cells (and in BHK cells), an investigation was conducted to determine whether the complexes between C3H/RV proteins and WN 3' (–) SL RNA were different in stability from those formed between C3H/He proteins and WN 3' (–) SL RNA. Complexes 1 and 3 formed in lysates of C3H/RV cells were shown to have a significantly longer half-life than the analogous complexes formed in C3H/He cell lysates.

B. The 3' SL

Analysis of the YF genome sequence originally revealed the presence of a potential SL structure formed by the 3'-terminal 87 nucleotides (Rice *et al.*, 1985). Similar stem–loop secondary structures with similar predicted thermal stability have been subsequently proposed for the 3' termini of all flavivirus genomes for which nucleotide sequence data are available (see, e.g., Brinton *et al.*, 1986; Hahn *et al.*, 1987; Irie *et al.*, 1989; Wengler and Castle, 1986). In fact, similar structures are

predicted to form at the 3' termini of the genomes of the other Flaviviridae, the hepac- and pestiviruses (Blight and Rice, 1997; Deng and Brock, 1993). The stem consists of about 30 hydrogen-bonded base pairs and is variably interrupted by bulges due to the predicted apposition of nucleotides that are unable to hydrogen bond. Sequence analysis also predicted a smaller adjacent stem-loop structure that is a consistent feature of the flavivirus 3' NCR, upstream from the long stem-loop. Together the two adjacent structures involve 90 to 100 nucleotides at the 3' end of the genome and are referred to collectively, for purposes of this discussion, as the 3' SL (Fig. 6). The 3' SL has a predicted thermal stability of -40 to -45.2 kcal, largely due to the free energy state of the long stem region.

The nucleotide sequence of the 3' SL is only partially conserved among unrelated flavivirus species (e.g., Fig. 6), although conserved segments do exist. For example, as already noted, the distal loop region of the long stem contains the flavivirus-conserved CACAG box, and in general there is a similarity among the nucleotide sequences of the predicted loop regions, both the double loop that terminates the long stem and in the single loop atop the adjacent short stem (the sequence 5' GANAGA-3') (Shi *et al.*, 1996b), nucleotides -83 to -89 in the DEN2 3' SL and nucleotides -85 to -91 in the WN 3' SL. In addition, there is similarity among flavivirus genomes at nucleotides -1 to -6 , which includes the 3' terminal 3' UC dinucleotide that is conserved among all known flavivirus genome sequences except that of the CFA virus. For example, Fig. 6 shows the similarity in sequence between 3' termini of the DEN genomes (the sequence 3'-UCUUGG-5', conserved in all four DEN serotypes) compared to the 3' terminus of the WN genome (the sequence 3'-UCUAGG-5', conserved in KUN, JE, and MVE genomes). Upstream from these six nucleotides, forming one side of the lower half of the long stem, one finds the semi-conserved pentanucleotide ACACA, which has been discussed previously. The 3' SL nucleotide sequence is best conserved among the closely related members of the same subgroup or clade, for example, among the four DEN serotypes and among species assigned to the JE subgroup.

Physical and biochemical evidence for the existence of the 3'-terminal long stem and loop in WN virus RNA was provided by Brinton and co-workers (1986). In their efforts to determine the nucleotide sequence of the genome at its 3' end, they subjected virus RNA to digestion with five different ribonucleases. Comparison of the gel mobility of the end products of these reactions to that of a sample of RNA that was sequenced by selective chemical degradation revealed that most of the

3'-terminal 84 nucleotides were relatively resistant to one or more of the RNases. This suggested that the nucleotides in question were involved in a secondary partially double-stranded structure that conferred the observed resistance to degradation. Drawing on the previous prediction of a 3'-terminal SL structure in the YF virus genome (Rice *et al.*, 1985), these workers hypothesized that the last 84 nucleotides of the WN genome were involved in a similar motif. WN virus RNA did not exhibit significant resistance to RNase digestion upstream from nucleotide -84.

Further analysis of interactions among nucleotides of the 3' SL suggested the existence of a pseudoknot tertiary structure (Shi *et al.*, 1996a), potentially formed by hydrogen bonding between nucleotides comprising one strand of the long stem (nucleotides -71 to -74 in the WN 3'-SL; Fig. 6) and nucleotides contained in the loop structure atop the adjacent short stem (nucleotides -86 to -89 in the WN 3' SL). In order for the pseudoknot to form in the 3' SL for most flavivirus species, one must posit the occurrence of G-A and G-U base pairs in some genomes (see DEN2 3'-SL; Fig. 6). Such atypical hydrogen bonding has been described for other RNA species (Heus and Pardi, 1991; Wimberly *et al.*, 1993), and pseudoknots have been described at the 3'-ends of other viral genomes (Dreher and Hall, 1988; Jacobson *et al.*, 1993; Pleij *et al.*, 1985). Replication of phage QB RNA was shown to be dependent on interaction of the replication complex with a pseudoknot formed by internal base pairing of the 3' terminus of the genome to a single-strand RNA segment 1200 nucleotides upstream (Klovins and van Duin., 1999).

WN and DEN3 3' model RNAs that included the lower half of the long stem and the upstream adjacent small stem and loop (nucleotides -1 to -11 linked covalently via four U residues to nucleotides -71 to -96, for the WN sequence; Fig. 6) were synthesized *in vitro* (Shi *et al.*, 1996a). Results from RNase probing, circular dichroism spectral analysis, UV-melting experiments, mutagenesis of the RNA sequence, and molecular modeling were consistent with the existence of the predicted pseudoknot in each of the two RNAs that were studied. A consequence of this prediction was that the RNA in a pseudoknot was predicted to form a cleft that could be functional in protein or metal binding. A potential weakness of this work, borne of necessity, was the fact that it was conducted using *in vitro*-transcribed RNA representing a small, discontinuous segment of the 3' NCR. Thus, possible interactions between and among the segments that were studied and more upstream sequences in the genome could not be assessed.

1. *Binding of Cellular Proteins to the 3' SL*

Because it likely contains RdRP promoter elements, the 3' SL is likely to interact with both cellular and viral proteins required for viral RNA replication, which must initiate with the synthesis of full-length minus-strand RNAs. The host range of most flaviviruses, both in nature and in the laboratory in tissue culture, is very broad. For example, flaviviruses can replicate efficiently in mammalian, avian, insect, reptile, and amphibian hosts, as well as in cell cultures from these hosts (Brinton, 1986). Therefore, one might expect that host cell proteins required for flavivirus replication are highly conserved or that domains required for this function may be conserved within a variety of multifunctional proteins of disparate size in different cell types. Cell proteins that bind 5' and/or 3' NCRs of genomes of several different RNA viruses have been described, including those of mouse hepatitis virus (Furuya and Lai, 1993), bovine coronavirus (Spagnolo and Hogue, 2000), measles virus (Leopardi *et al.*, 1993), Rubella virus (Nakhasi *et al.*, 1990, 1991), Sindbis virus (Pardigon *et al.*, 1993), hepatitis A virus (Schultz *et al.*, 1996; Yi *et al.*, 2000), and hepatitis C virus (Luo, 1999; Spangberg and Schwartz, 1999). In many cases, the cited reports refer to binding of cellular translation factors to IRES elements found in the 5' NCRs of hepaciviruses and picornaviruses, among others, but not present in the flavivirus 5' NCR. For further information, the reader is referred to reviews of this subject with emphasis on positive-strand RNA viruses in this volume (Brinton, 2001; Strauss and Strauss, 2002).

In general, the literature on binding of proteins to 3' NCRs of positive-strand RNA virus genomes does not shed light on the functional requirements associated with protein binding. However, two reports [for mouse hepatitis virus RNA (Huang and Lai, 2001) and poliovirus RNA (Herold and Andino, 2001)] suggest that cellular proteins may be involved in linkage of 5' and 3' termini of replicating RNA virus genomes through the formation of ribonucleoprotein (RNP) complexes. For poliovirus, formation of this RNP was required for the initiation of negative-strand RNA synthesis.

Cell proteins required for flavivirus replication have not yet been identified with certainty. However, the specific interaction of a radiolabeled RNA representing the nucleotide sequence of the WN 3' SL with proteins in extracts of uninfected BHK cells was reported by Blackwell and Brinton (1995). (BHK cells are a common laboratory cell substrate for growth of the WN virus to very high titers.) Gel mobility shift assays, UV cross-linking, Northwestern blotting, and ion-exchange chromatography were initially used to identify 56-, 84-, and 105-kDa

proteins that bound to a truncated *in vitro*-synthesized fragment of the WN and DEN3 3' SLs. The unidentified 105-kDa protein species also had significant affinity for an RNA representing the sequence of the 3' (negative-sense) SL (see earlier discussion). The hypothesis was advanced that this protein may play a role in the initiation of both plus- and minus-strand RNA synthesis. Similar specificity has been reported for other viral systems. For example, one of the cellular proteins that binds rubella virus RNA appears to have specific affinity for both minus- and plus-sense SL structures at the 3'-termini of the respective RNAs (Nakhasi *et al.*, 1991).

Blackwell and Brinton (1997) subsequently identified the 56-kDa protein that bound the WN 3' SL in BHK cell lysates as the translation elongation factor, eF1- α . The site for eF1- α binding was localized to nucleotides -62 to -65, the sequence 3'ACAC-5' (Fig. 6). Li and Brinton (2001) made the possibly relevant observation that an *in vitro*-synthesized RNA containing the nucleotide sequence of the WN 3' SL inhibited *in vitro* translation of WN RNAs containing 5'-terminal portions of the WN ORF and of foreign mRNAs. In contrast, an RNA containing nucleotides representing the 3'-terminal secondary structure of rubella virus RNA enhanced *in vitro* translation. These contrasting results may indicate that the affinity of the WN 3' SL for translation elongation factors is significant for upregulating WN viral translation or downregulating translation of cellular mRNAs *in vivo*. Plant and animal elongation factors eF1- α , β , and γ are recognized components of replicases in other systems. Das and co-workers (1998) demonstrated that the polymerase of the negative-strand virus, vesicular stomatitis virus (VSV), which replicates in both insects and mammals, binds eF1- α tightly, and that the resulting complex binds eF1- β and γ . All three elongation factors are required for VSV replicase activity in an *in vitro* system. In earlier work, Blumenthal *et al.* (1976) demonstrated that RNA bacteriophage Q β uses three cellular proteins as components of its replicase, the bacterial translation elongation factors, Ts and Tu, and ribosomal protein S1. Further work is required to establish with certainty the physiological significance of the observation that the WN 3' SL RNA binds translation elongation factors.

In a more recent study, an *in vitro*-synthesized, 83-nucleotide RNA representing the JE virus 3' SL nucleotide sequence was evaluated for binding of specific proteins in lysates of neonatal mouse brain (Ta and Vрати, 2000). Three RNP complexes thus derived were stable in the presence of high-salt buffer. UV cross-linking and Northwestern blotting analysis identified three proteins involved in RNPs, with

apparent molecular masses of 32, 35, and 50 kDa. Gene products derived from a neonatal mouse brain cDNA library were screened for binding to the JE 3' SL, and a 36-kDa protein, MOV34, was thus identified. Subsequent experiments showed that the MOV34 protein could also bind to the 3' SL in native full-length JE virus genomic RNA. MOV34 is a 26S proteasome subunit protein that belongs to a family also thought to be involved in RNA transcription and translation. A computer-based comparison of the amino acid sequences of translation elongation factor eF1- α and MOV34 revealed no detectable sequence similarity between them, even when constraints for detection were set very low (L. Markoff, unpublished data). However, both of these proteins contain a high proportion of positively and negatively charged amino acids, which could effect an association with RNA and/or cell membrane organelles. The translation elongation factor sequence is 463 amino acids in length and includes 23 Asp residues, 31 Glu residues, 17 Arg residues, and 47 Lys residues for a net positive charge of +10. MOV34 is 309 amino acids in length and contains 18 Asp residues, 23 Glu residues, 11 Arg residues, and 23 Lys residues for a net negative charge of -8.

2. Binding of Viral Proteins to the 3' SL and the Upstream Proximal Portion of the 3' NCR

As mentioned previously, the viral nonstructural proteins NS3 and NS5 contain domains with helicase and polymerase activities, respectively, required for flavivirus RNA synthesis. It is logical to expect that the viral helicase and RdRP both bind the 3' termini of plus-strand RNA in order to initiate negative-strand synthesis. [Chen and co-workers \(1997\)](#) made lysates of BHK cells infected with JE virus and first demonstrated that these lysates retained viral RNA synthetic activity. Next, they probed active lysates for proteins that could bind an exogenous 585-nucleotide *in vitro*-synthesized RNA representing the JE 3' NCR by UV cross-linking. They focused on proteins that specifically formed in lysates of infected cells, as compared to uninfected cells. Two proteins, p71 and p110, were thus identified. Mapping experiments revealed that protein p110 required only the 3'-terminal 80 nucleotides of the JE genome for binding, whereas protein p71 appeared to have a more diffuse binding site on the 585-nucleotide probe. The two proteins were subsequently identified as NS5 and NS3, respectively. The results of immune precipitation assays suggested further that NS5 and NS3 formed a complex with each other in association with the 3' NCR.

3. Nucleotide Sequence Specificity of the 3' SL

As mentioned previously, the nucleotide sequence of the 3' SL in the genomes of different species of flaviviruses is only partially conserved. Zeng and colleagues (1998) asked whether nucleotide sequences of the WN 3' SL could replace those of the DEN2 3' SL, in the context of the DEN2 genome. This was done by site-directed substitution mutagenesis of sequences encoding the 3' SL in a DEN2-infectious DNA (Polo *et al.*, 1997). RNA derived from "D2/WN-SL" DNA was infectious in that DEN2-specific immune fluorescence was observed in transfected cells, but the resulting virus was highly impaired for replication; it never achieved a titer greater than 100 pfu/ml in monkey cells. In contrast, DEN2 wild-type virus derived from RNA transcribed *in vitro* typically grew to titers >10E6 pfu/ml. This demonstrated that the nucleotide sequence, not merely the secondary structure of the 3' SL, was a determinant of replication efficiency.

An attempt was then made to define which DEN2 3' SL nucleotides were required for efficient virus replication by restoring DEN2 3' SL-specific sequence elements in the context of the D2/WN-SL mutant DNA. "Top" and "bottom" portions of the DEN2 (and WN) 3' SLs were defined arbitrarily according to previous work by Blackwell and Brinton (1995) (Fig. 6). Six additional DEN2 3' SL mutant DNAs were created, and the growth phenotypes of the resulting viruses in LLCMK2 cells are summarized in relation to their 3' SL structure in Fig. 8. The mutations were "lethal," meaning no positive immunofluorescence could be detected in transfected cells; "sublethal," meaning the viruses replicated about as well as the parent mutant, D2/WN-SL; or "viable," meaning peak titers in LLCMK2 cells were equal to or no worse than 10- or 100-fold reduced compared to the wild-type DEN2 virus. Results identified an 11-bp segment, comprising the uppermost portion of the bottom of the wild-type DEN2 3' SL sequence, as being essential in viable virus mutants (DEN2 nucleotides 7–17 and 63–73 in Fig. 6). In addition, the study demonstrated that the DEN2-specific nucleotide sequence of the "top" portion of the DEN2 3' SL was not required for efficient replication, but that its base-paired structure was essential. A mutant in which the entire top portion of the WN 3' SL was substituted for that of the DEN2 3' SL grew as well as the wild-type virus, but a second mutant, in which base pairing in the top portion of the long stem was abrogated by mutagenesis of the DEN2 nucleotide sequence, had a lethal phenotype. Defects in replication of all mutant viruses appeared to be at the level of RNA replication, suggesting a critical role of the 3' SL for initiation of negative-strand RNA synthesis.

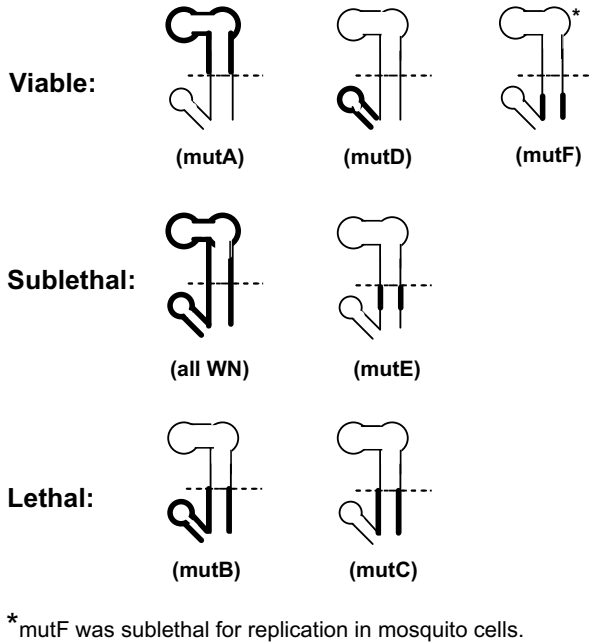


FIG 8. Chimeric DEN2/WN 3' SL nucleotide sequences in DEN2 genomic DNA. The 3' SL secondary structure is depicted as a line drawing. DEN2 nucleotide sequences are depicted as a thin line, and WN nucleotide sequences are depicted as a thick line. "Top" and "bottom" portions of the 3' SL are indicated by horizontal dashed lines bisecting the long stem in the 3' SL. The chimeric nucleotide sequences indicated were substituted for the wild-type nucleotide sequence of the 3' SL in DEN2 infectious DNA, and the resulting mutant DNAs were used to derive RNAs that yielded DEN2 viruses with the indicated phenotypes after transfection of cultured monkey kidney cells (see text). Viable viruses all retained the DEN2-specific 11-bp segment comprising the top portion of the bottom section of the long stem in the 3' SL. The DEN2mutF virus (asterisk) was notable for its host range-restricted phenotype in mosquito cells, where its replication was 100,000-fold reduced, compared to the wild-type DEN2 virus (Zeng *et al.*, 1998).

Mutant D2/WN-SL(mutF) or "mutant F" virus (Fig. 8) was of special interest in that it displayed a viable phenotype in LLCMK2 cells but was severely impaired for replication in cultured mosquito cells. Thus, site-directed mutagenesis of either the 5' NCR (Cahour *et al.*, 1995) or the 3' SL (Zeng *et al.*, 1998) may result in the genesis of host range-restricted mutant DEN viruses. These findings support the hypothesis that host cell proteins play an important role in virus replication. Zeng and co-workers (1998) also demonstrated that DEN2 mutant F virus was defective in RNA synthesis at early times after infection of

mosquito cells using an assay that did not distinguish between plus- and minus-strand synthesis. However, a more specific mechanism for the host range-restricted phenotype of DEN2 mutant F virus has not been elucidated thus far. For example, using an *in vitro* assay for RdRP activity (described previously), [You and co-workers \(2001\)](#) demonstrated that an RNA containing the DEN2 mutant F 5' and 3' NCR nucleotide sequence was not defective for replication in a mosquito cell extract compared to its activity in an extract of monkey kidney cells. Because this assay detected only negative-strand synthesis, perhaps the results indicate that DEN2 mutant F replication is defective at the level of progeny positive-strand RNA synthesis. Other efforts to detect differential binding of mosquito cell proteins to the DEN2 mutant F 3' SL, using techniques pioneered by [Blackwell and Brinton \(1995\)](#), have thus been unsuccessful (Yu and Markoff, unpublished results).

Introduction of mutant F mutational changes ([Markoff *et al.*, 2002](#)) into infectious DNAs for human virulent DEN1 and DEN4 viruses has resulted in the production of mutant viruses that have a host cell-restricted phenotype similar to that of the DEN2 mutant F virus. The DEN1 mutant F virus was subsequently shown to be attenuated and highly immunogenic in a monkey model for DEN virus infection of humans and is currently a candidate for DEN vaccine development.

IV. MUTATIONAL CHANGES IN THE 5' OR 3' NCR AND THE 3' SL MAY RESULT IN ATTENUATION OF VIRUS VIRULENCE

Mutational changes, either in the ORF or in the 5' and/or 3' NCRs, could be expected to alter virus growth properties and therefore virulence. For example, [Chiou and Chen \(2001\)](#) reported on the properties and genome sequences of two naturally occurring isolates of JE virus found in an area of Taiwan where local residents have high levels of JE antibodies. One of these isolates, strain T1P1, exhibited 10,000-fold less neurovirulence in mice compared to the other strain, CH1932. A comparison of the complete genome sequences of these two strains revealed that they differed by seven nucleotides. Two of these sequence differences resulted in amino acid changes in NS3, and there were two others in the 3' NCR, upstream from the 3' SL. The other three nucleotide sequence differences between the two strains were "silent" ones in the M and NS5 genes. The inference was made that the nucleotide sequence differences noted in NS3 and in the 3' NCR were responsible for the marked differences in neurovirulence between strain T1P1 and

strain CH1932. No attempt was made to determine whether the 3' SL mutational difference between strains T1P1 and CH1932 resulted in altered binding of virus- or cell-specific proteins to the 3' NCR. See also discussions of the work of Cahour (1995), Durbin *et al.* (2001), Mandl *et al.* (1998), Men *et al.* (1996), Pletnev (2001), and Zeng *et al.* (1998) in conjunction with Markoff *et al.* (2002) for other examples of attenuation of flavivirus virulence resulting from mutagenesis of the 5' or 3' NCR.

V. HYPOTHESIS FOR HYDROGEN BONDING BETWEEN 5' AND 3' NCRs

The existence *in vivo* of the the previously predicted secondary structures formed by *cis* interactions of nucleotides within the 5' NCR (Brinton and Disposito, 1988) and within the 3' NCR (Brinton *et al.*, 1986; Hahn *et al.*, 1987; Rice *et al.*, 1985) must be reconsidered in light of a more recent disclosure of a possible interaction between nucleotides of the 5' and 3' NCRs, including the cyclization sequences, in *trans* to form double-stranded panhandle-like structures (Khromykh *et al.*, 2001). Sequences representing the 5'-terminal 150 to 170 nucleotides, an intervening poly(A)tract, and the 3' terminal 110 to 120 nucleotides of the KUN and other mosquito-borne virus genomes, the TBE genome, and the CFA genome were folded by computer using the "MFOLD" program (Mathews *et al.*, 1999) and were based on the assumption of hydrogen bonding between the respective 5' and 3' cyclization sequences. Results of this analysis revealed that the 5' and 3' NCRs in all flavivirus genomes contained *trans* complementary sequences upstream and downstream from the conserved cyclization sequences, such that secondary structures of very similar shape and very high thermal stability could be formed between them.

Khromykh *et al.* predict an extensive *trans* interaction between the KUN genome 5' and 3' NCRs. Very similar structures were predicted for the JE, YF, DEN, TBE, and CFA genomes. The model predicts for the KUN genome that (1) 3' nucleotides -1 to -8 hydrogen bond to 5' nucleotides +45 to +52 and (2) 3' nucleotides -11 to -14 and -68 to -106 hydrogen bond with 5' nucleotides +77 to +150. As a result, (1) 3' NCR nucleotides -15 to -67 remain available to form a 3' *cis* stem and loop structure, a truncated 3' SL composed of nucleotides previously comprising the top portion of the long stem and including the distal loop regions (Fig. 6), and (2) 5' NCR nucleotides +1 to +44 and +53 to +76 are available in *cis* to form short stem-loop structures,

conserving part of the 5' SL predicted by [Brinton and Dispoto \(1988\)](#) ([Fig. 7](#)), including its distal loop. Therefore, the distal loop in the 5' NCR could still form. The pseudoknot predicted by [Shi *et al.* \(1996a\)](#) and involving a portion of the 3' SL conserved in the 5'/3' panhandle structure could also still form, as the semiconserved nucleotides that form the loop atop the short stem in the conventional model of the 3' SL secondary structure (5'-CANAG-3'; [Fig. 6](#)) are available for hydrogen bonding with the conserved remnant of the long SL in the model proposed by [Khromykh and co-workers \(2001\)](#). It is worthy to note that the predicted thermal stability of the panhandle exceeds those of the earlier predicted 5' SL and 3' SL by a wide margin. In addition, the predicted thermal stability of mutant KUN RNA 5'/3' hydrogen-bonded complexes that lacked complementary cyclization sequences was very little different from that of the wild-type complex. This suggests that the requirement for the cyclization sequences relates to their specific functional necessity, e.g., in initiation of RNA synthesis or translation or packaging, rather than to their additive effect on thermal stability of the panhandle. One must view all of these hypotheses for base pairing with an open mind for two reasons: (1) Computer power to predict folding and base pairing in nucleic acids is increasing rapidly and therefore any of these hypotheses are subject to evolution based on the state of the art in computer programs and computer power and (2) the "wild card" in any computer-based prediction is the unknown but certain effect of cellular and possibly viral proteins on the secondary and tertiary structure of the RNA.

VI. ADDITIONAL SECONDARY STRUCTURE PREDICTIONS IN THE 3' NCR

[Proutski and colleagues \(1997b\)](#) analyzed the distal 330 to 400 nucleotides of published flavivirus 3' NCR sequences for predictable secondary structure using a genetic algorithm that purportedly simulates the natural folding pathway that takes place during RNA elongation and allows prediction of tertiary interactions as well as secondary structure. They then used a second program to detect compensatory mutations (meaning mutations in both strands of a base-paired region that conserve hydrogen bonding) in the 3' NCRs of related viruses. The presence of compensatory mutations was taken as an indication that the base pairing in that region was likely to occur *in vivo*, as random mutational events conserved the double strandedness.

As a result of this analysis, three regions of secondary structure could be defined within what could be described as the core element

(Wallner *et al.*, 1995) of 3' NCRs in mosquito-borne as well as tick-borne viruses (Figs. 2 and 3). Region I consisted of nucleotides upstream from CS1, CS2, and RCS2 in mosquito-borne virus genomes or the 5'-most extreme of the core element in tick-borne virus genomes. This is a region of high variability in linear nucleotide sequences. However, all flavivirus sequences in region I formed very similar long hairpin structures with a branching stem-loop side structure or, in some cases, a bulge-loop on the 5' side of the main hairpin. Downstream from this long hairpin in region I was a conserved shorter hairpin of similar configuration. Nucleotides in region II included those assigned to the CS2 and RCS2 boxes plus intervening sequences. Unlike region I, where a single consensus secondary structure was predicted, region II nucleotides exhibited differences in folding patterns that separated the viruses into three groups: (1) DEN and JE subgroup viruses, (2) tick-borne viruses, and (3) YF virus. Region II in genomes of viruses in the DEN and JE subgroups was shown to contain tandem repeat sequences (TL1 and TL2; Fig. 4) that potentially form conserved short SL structures. TL1 is located upstream from RCS2, and TL2 is located between RCS2 and CS2. The secondary structures in region II for tick-borne viruses were so different from those predicted for mosquito-borne viruses that no one element was conserved between the two groups. Region III consisted of nucleotides traditionally believed to constitute CS-1 and the 3' SL. A 3'-terminal SL structure was predicted by this analysis, just as it was predicted by the MFOLD program used by Khromykh *et al.* (2001). As in the Khromykh model for pan-handle formation, the length of the long stem at the genome 3' terminus in the Proutski model (Proutski *et al.*, 1997b) is shorter than that predicted by older methods due to the alternative hydrogen bonding of nucleotides that form the lower portion of the long stem shown in Fig. 6. The Khromykh model (Khromykh *et al.*, 2001) and predictions for the secondary structure within the 3' NCR put forward by Proutski and colleagues (1997b) are not incompatible if one posits that sequences not included in the 3' SL in the Proutski model are base paired with nucleotides of the 5' NCR.

The model for the secondary structure of the 3'-terminal 380 nucleotides of the YF genome was subsequently applied to the sequences of several different YF isolates and to the sequences of vaccine strains 17D-204 and 17DD (Proutski *et al.*, 1997a). An association between secondary structure of the 3' NCR and virus virulence was claimed in that all wild strain genomes were likely to fold similarly and in a significantly different way from the attenuated vaccine strains. In addition, strain 17DD had an intermediate folding pattern that was less

different from that of wild strain genomes than was the strain 17D-204 3' NCR. The authors noted that strain 17DD is more virulent in mice than strain 17D-204 and suggested that mutagenesis of the primary nucleotide sequence of the 3' NCR to alter the secondary structure could be a useful strategy for vaccine development. The point may be well taken with respect to mice, but after more than 60 years of experience, there is no indication that the strain 17DD vaccine is any less attenuated in humans than is the strain 17D-204 vaccine (Monath, 1999).

The same group (Proutski *et al.*, 1999) applied similar reasoning to an analysis of the results of Men *et al.* (1996) and Khromykh and Westaway (1997), which were described earlier, in which each of these groups of investigators created internal deletions in the 3' NCRs of DEN4 and KUN virus genomes, respectively. Consideration of the resulting alterations of the 3' NCR secondary structure resulting from deletions in the DEN4 genome, in particular, could help explain the observed discordance between the size of the deletion and the growth phenotype of the resulting mutant DEN4 viruses.

Olsthoorn and Bol (2001) also sought to detect a secondary and tertiary structure in the 3' NCR, and they performed a computer analysis of 191 flavivirus genomic sequences representing four serogroups (DEN, JE, YF, and TBE), also using the MFOLD program. This program yielded a set of suboptimal structure predictions, which were then checked manually for the presence of pseudoknots and for phylogenetic consistency. The analyses confirmed the presence of conserved serogroup-specific secondary structures for each of the four serogroups in the 3' NCR upstream from the 3' SL. As in the work of Proutski and colleagues (1997b), the existence of conserved SL structures that involved CS2 and RCS2 was confirmed for viruses of the DEN and JE subgroups. These were paired “dumbbell”-shaped structures that included the conserved stem and the respective TL1 and TL2 loop sequences identified previously. Olsthoorn and Bol (2001) proposed that TL1 and TL2 loops could each form pseudoknots by hydrogen bonding with conserved tetranucleotide sequences pk1 and pk2 present in predicted linear regions of the DEN and JE subgroup 3' NCRs. For example, in the DEN serogroup, nucleotides in the TL1 sequence (5'-GAAG**CUGUA**-3'; Fig. 4), which forms a loop within a secondary structure that includes CS2, could form hydrogen bonds (bold letters in the TL1 sequence) with conserved sequence pk1 (5'-ACAG**C**-3'), which lies downstream from CS2. Similarly, in the upstream dumbbell-shaped secondary structure that includes RCS2, the nucleotides 5'-GCUG**C**-3' in TL2 could hydrogen bond with linear

sequence pk2 (5'-GCAGC-3'), which also lies between RCS2 and CS2. For JE serogroup viruses, pk1 and pk2 were both located downstream from CS2. In the YF genomes, which lack RCS2 (Fig. 2), three pseudoknots were predicted, one of which was similar in its genesis to the more downstream pseudoknot in DEN and JE serogroup genomes. As in the analyses of Proutski *et al.* (1997a), the TBE genomes were folded very differently from those of the mosquito-borne viruses, resulting in the formation of a single predicted pseudoknot within the core element. The internally consistent results of this study and the compatibility of the results with the work of others (Proutski *et al.*, 1997b; Rauscher *et al.*, 1997) constitute additional evidence that pseudoknot formation may be required for proper function of the 3' NCR in flavivirus RNA replication.

VII. SUMMARY AND FUTURE DIRECTIONS

The flavivirus genome is a capped, positive-sense RNA approximately 10.5 kb in length. It contains a single long ORF, flanked by a 5' NCR, which is about 100 nucleotides in length, and a 3' NCR ranging in size from about 400 to 800 nucleotides in length. Replication is thought to occur by a semiconservative mechanism requiring double-stranded replicative form RNA to generate replicative intermediate RNA containing nascent single strands. The 5' and 3' NCRs must play a role in the initiation of negative-strand synthesis on virus RNA released from entering virions, switching from negative-strand synthesis to synthesis of progeny plus strand RNA at late times after infection, and possibly in the initiation of translation and in the packaging of virus plus strand RNA into particles. However, none of these mechanisms is as yet well defined. The presence of conserved and non-conserved complementary nucleotide sequences near the 5' and 3' termini of flavivirus genomes suggests that "panhandle" or circular RNA structures are formed transiently by hydrogen bonding at some stage during RNA replication.

Genomes of flaviviruses that differ according to vector (tick-borne viruses vs mosquito-borne viruses) and by antigenic subgroup may also contain unique or characteristic nucleotide sequence features in their noncoding regions. However, a few such features are conserved in all flavivirus genomes. These are as follows: (1) All vector-borne flavivirus genomes begin with the dinucleotide sequence 5'-AG and end with the complementary nucleotide sequence UC-3'. The functional significance of these sequences is not known but may play a role in

cyclization of the genome during or as a prerequisite for replication. (2) While the nucleotide sequence of the 5'NCR is not well conserved among flavivirus species, the 5'NCRs of all flavivirus genomes do contain a predicted stable stem and loop structure. (3) All flavivirus genomes contain a 3'-terminal approximately 100-nucleotide sequence that is predicted to form a highly stable stem and loop structure, the 3' SL. The actual nucleotide sequence of this domain is well conserved among viruses of the same subgroup but is much less well conserved across subgroups. The 3' SL is known to bind the virus-coded RNA-dependent RNA polymerase protein NS5 and to be required for binding of a virus-coded helicase, NS3, to the 3'NCR. In addition, the 3' SL apparently binds several cellular proteins required for virus replication. One of these proteins has been identified as the translation elongation factor eF1- α . Mutations of the 3' SL can be shown to affect the host range of the virus, suggesting that binding of host cell proteins is critically important for replication. (4) Within the nucleotide sequence forming the distal loop of the 3' SL in all vector-borne flavivirus genomes, there is a conserved nucleotide sequence, 5'-CACAG. The function of this conserved pentanucleotide sequence in replication is unknown.

Functionally analogous group- and subgroup-specific nucleotide sequence elements have been identified in the 5' and 3' NCRs of flavivirus genomes. For example, mosquito-borne flavivirus genomes contain conserved complementary 8-nucleotide sequences. One is located in the 3' NCR, referred to as the cyclization sequence (CYC). The other is located in the 5' terminus of the long ORF, just downstream from the start codon that initiates translation of the capsid protein gene segment, referred to as the "5'CS." The CYC sequence and the 5'CS are thought to hydrogen bond, helping to form a panhandle structure during RNA replication. Tick-borne flavivirus genomes contain two analogous pairs of 5'/3' complementary sequences, each 11 nucleotides in length. The 3'-terminal motifs in tick-borne virus genomes are herein referred to as C1' and C2', and the 5'-terminal motifs are referred to as C1 and C2. The CYC sequence in mosquito-borne virus genomes and C1' are analogous in their position in the 3' NCR, just upstream from nucleotides comprising the 3' SL. However, the complement of C1', C1, lies wholly within the 5' NCR in tick-borne virus genomes, not in the ORF. C2' lies upstream from C1', and its complement at the 5' end of the genome, C2, lies within the ORF, in a location analogous to that of the 5'CS in mosquito-borne virus genomes. The 5' and 3' complementary sequences are necessary for viral RNA replication, but the specific nucleotide sequence of,

for example, the CYC sequence and the 5'CS in mosquito-borne virus genomes can be substituted by a random pair of complementary sequences 8 nucleotides in length without much loss of replication efficiency.

Mosquito-borne viruses of the dengue subgroup contain the shortest 3' NCRs in this group (~400 nucleotides), with a minimum of conserved features. CS1 is a 20-nucleotide segment just upstream from the 3' SL. The 8-nucleotide CYC sequence comprises the upstreammost part of CS1. Deletion of CS1 nucleotide sequences was specifically lethal for DEN4 virus replication, but the functional requirement for the 12 nucleotides of CS1 that do not compose the CYC sequence is not known. CS2 is a 23-nucleotide conserved element that lies 12 to 22 nucleotides upstream from CS1. CS2 is one of a pair of repeated sequence elements in the 3' NCR; the repeat sequence RCS2 lies 51 to 62 nucleotides upstream from CS2 in the genome of mosquito-borne viruses. All or parts of CS2 and RCS2 were dispensible for DEN-4 virus replication, but the respective mutant viruses were mildly to markedly reduced in replication efficiency.

3' NCRs of viruses in the JE subgroup are larger in size than those of the DEN subgroup (~500 to ~600 nucleotides) and contain additional conserved repeat sequence motifs, in addition to all the conserved motifs described for the 3' NCR in DEN virus genomes. These additional repeat sequences are referred to as CS3 and RCS3, respectively, in relation to their increasing distance from the 3' terminus of the genome compared to CS2 and RCS2. CS3 and RCS3 are each 28 nucleotides in length and are separated by 120 nucleotides in the 3' NCRs of JE subgroup viral genomes.

The YF virus genome is unique in several ways from that of the other mosquito-borne flaviviruses, including its organization of the 3' NCR, which contains CS1 and CS2 but lacks the repeat segment RCS2 found in the other genomes of the mosquito-borne virus group. The YF strain 17D vaccine virus 3' NCR also contains a series of three unique 40-nucleotide tandem repeats at the 5' end of its 3' NCR. Also, unlike other flavivirus genomes, the yellow fever 3' NCR contains three in-frame stop codons for translation of the ORF, the last of which lies within the upstream boundary of the first of the tandem repeats.

For tick-borne flaviviruses, the 3' NCRs of tick-borne encephalitis virus strains have been analyzed extensively. The genome of TBE strain Neudorfl virus genome contains the longest 3' NCR in this group, at 764 nucleotides. This segment has been divided, by comparative analysis of the nucleotide sequences of strain Neudorfl and other TBE viruses and the results of mutagenesis experiments, into a "core

element” and a “variable region.” The core element is directly analogous to the entirety of the DEN virus 3' NCR and includes the most distal portion of the 3' NCR, containing the 3' SL. The cyclization sequence, C1', lies within a mixed purine/pyrimidine-rich tract upstream from the 3' SL, which is analogous in position and sequence to CS1 in mosquito-borne flavivirus genomes. Upstream from this PR segment exists consecutive pyrimidine- and purine-rich segments, each about 14 nucleotides in length. These could be said to be analogous to CS2 found in mosquito-borne virus genomes. The second cyclization sequence motif, C2', lies upstream from C1', also within the core element. The core element also includes one of a pair of long tandem repeats in the archetypal strain Neudorfl 3' NCR, R3, which is 76 nucleotides in length.

The structure of the variable region, which comprises the upstream portion of the 3' NCR in strain Neudorfl and related TBE virus genomes, is complex. The variable region contains a second iteration of the R3 sequence found in the core element. It also contains additional repeated sequences, R1 and R2. R1 is repeated three times in the strain Neudorfl genome. One of these repeats lies within the 3' terminus of the ORF. The single repeat of R2 also lies within this same distal segment of the ORF. The most striking feature of the variable region in TBE genomes is the presence of a poly(A) segment that can be as long as 250 nucleotides. The function of the variable region remains a puzzle because there exist naturally occurring highly related TBE virus variants with much shorter 3' NCRs compared to that of strain Neudorfl, completely lacking a variable region. In addition, the variable region nucleotide sequence can be completely removed from the strain Neudorfl genome by site-directed mutagenesis or by spontaneous mutation that occurs during repeated *in vitro* tissue culture passage, with no effect on virus replication in tissue culture or on mouse neurovirulence.

In recent years, computer-based methods have been used to predict the existence of additional secondary and tertiary structures within the 3' NCR and to predict additional interactions between 5' and 3' NCRs based on the potential for hydrogen bonding of complementary segments in these two regions of the genome, in addition to that afforded by the known cyclization sequences. The significance of this work remains to be seen and requires better knowledge and understanding of the possible effects of viral and cellular proteins on folding of the RNA, which can only be gained by further *in vitro* studies with engineered RNAs and virus infectious clones.

REFERENCES

- Avni, D., Shama, S., Loreni, F., and Meyuhas, O. (1994). Vertebrate mRNAs with a 5'-terminal pyrimidine tract are candidates for translational repression in quiescent cells: Characterization of the translational cis-regulatory element. *Mol. Cell. Biol.* **14**(6):3822-3833.
- Behe, M. J. (1995). An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. *Nucleic Acids Res.* **23**(4):689-695.
- Bergamini, G., Preiss, T., and Hentze, M. W. (2000). Picornavirus IREs and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA* **6**(12):1781-1790.
- Blackwell, J. L., and Brinton, M. A. (1995). BHK cell proteins that bind to the 3' stem-loop structure of the West Nile virus genome RNA. *J. Virol.* **69**(9):5650-5658.
- Blackwell, J. L., and Brinton, M. A. (1997). Translation elongation factor eF-1 α interacts with the 3' stem-loop region of West Nile Virus genomic RNA. *J. Virol.* **71**(9):6433-6444.
- Blight, K. J., and Rice, C. M. (1997). Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *J. Virol.* **71**(10):7345-7352.
- Blumenthal, T., and Carmichael, G. G. (1979). RNA replication: Function and structure of Qbeta-replicase. *Annu. Rev. Biochem.* **48**:525-532.
- Blumenthal, T., Young, R. A., and Brown, S. (1976). Function and structure in phage QB RNA replicase: Association of EF-Tu-Ts with the other enzyme subunits. *J. Biol. Chem.* **251**(9):2740-2743.
- Brinton, M. A. (1986). Replication of flaviviruses. In "The Togaviridae and Flaviviridae: Biology, Structure, Replication" (S. Schlesinger and M. J. Schlesinger, eds.), pp. 327-365. Plenum, New York.
- Brinton, M. A. (2001). Host factors involved in West Nile virus replication. *Ann. N. Y. Acad. Sci.* **951**:207-219.
- Brinton, M. A., and Dispoto, J. H. (1988). Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology* **162**:290-299.
- Brinton, M. A., Fernandez, A. V., and Dispoto, J. H. (1986). The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* **153**:113-121.
- Brown, E. A., Zhang, H., Ping, L. H., and Lemon, S. M. (1992). Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res.* **20**:5041-5045.
- Cahour, A., Pletnev, A., Vazeille-Falcoz, M., Rosen, L., and Lai, C.-J. (1995). Growth-restricted dengue virus mutants containing deletions in the 5' noncoding region of the RNA genome. *Virology* **207**:66-76.
- Calisher, C. H., Karabatsos, N., Dalrymple, J. M., Shope, R. E., Porterfield, J. S., Westaway, E. G., and Brandt, W. E. (1989). Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. *J. Gen. Virol.* **70**:37-43.
- Cammisa-Parks, H., Cisar, L. A., Kane, A., and Stollar, V. (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.
- Chambers, T. J., Hahn, C. S., Galler, R., and Rice, C. M. (1990). Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* **44**:649-688.

- Chambers, T. J., Nestorowicz, A., Amberg, S. M., and Rice, C. M. (1993). Mutagenesis of the yellow fever virus NS2B protein: Effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *J. Virol.* **67**:6797–6807.
- Chen, C.-J., Kuo, M.-D., Chien, L.-J., Hsu, S.-L., Wang, Y.-M., and Lin, J.-H. (1997). RNA-protein interactions: Involvement of NS3, NS5, and 3' noncoding regions of Japanese encephalitis virus genomic RNA. *J. Virol.* **71**(5):3466–3473.
- Chiou, S.-S., and Chen, W.-J. (2001). Mutations in the NS3 gene and 3'-NCR of Japanese encephalitis virus isolated from an unconventional ecosystem and implications for natural attenuation of the virus. *Virology* **289**:129–136.
- Chu, P. W., and Westaway, E. G. (1985). Replication strategy of Kunjin virus: Evidence for recycling role of replicative form RNA as template in semiconservative and asymmetric replication. *Virology* **140**:68–79.
- Chu, P. W., and Westaway, E. G. (1987). Characterization of Kunjin virus RNA-dependent RNA polymerase: Reinitiation of synthesis in vitro. *Virology* **157**:330–337.
- Cleaves, G. R., and Dubin, D. T. (1979). Methylation status of intracellular dengue type 2 40S RNA. *Virology* **96**:159–165.
- Czyzyk-Krzaska, M. F., Dominski, Z., Kole, R., and Millhorn, D. E. (1994). Hypoxia stimulates binding of a cytoplasmic protein to a pyrimidine-rich sequence in the 3'-untranslated region of rat tyrosine hydroxylase mRNA. *J. Biol. Chem.* **269**(13):9940–9945.
- Das, T., Mathur, M., Gupta, A. K., Janssen, G. M., and Banerjee, A. K. (1998). RNA polymerase of vesicular stomatitis virus specifically associates with translation elongation factor-1 alphabeta gamma for its activity. *Proc. Natl. Acad. Sci. USA* **95**:1449–1458.
- Deng, R., and Brock, K. V. (1993). 5' and 3' untranslated regions of pestivirus genome: Primary and secondary structure analyses. *Nucleic Acids Res.* **21**(8):1949–1957.
- Dreher, T. W., and Hall, T. C. (1988). Mutational analysis of the sequence and structural requirements in brome mosaic virus RNA for minus strand promoter activity. *J. Mol. Biol.* **201**(1):31–40.
- Durbin, A. P., Karron, R. A., Sun, W., Vaughn, D. W., Reynolds, M. J., Perreault, J. R., Thumar, B., Men, R., Lai, C. J., Elkins, W. R., Chanock, R. M., Murphy, B. R., and Whitehead, S. S. (2001). Attenuation and immunogenicity in humans of a live dengue virus type-4 vaccine candidate with a 30 nucleotide deletion in its 3'-untranslated region. *Am. J. Trop. Med. Hyg.* **65**(5):405–413.
- Falgout, B., Pethel, M., Zhang, Y. M., and Lai, C. J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J. Virol.* **65**:2467–2475.
- Frey, T., Gard, D. L., and Strauss, J. H. (1979). Biophysical studies on circle formation by Sindbis virus 49 S RNA. *J. Mol. Biol.* **132**:1–18.
- Fukushi, S., Katayama, K., Kurihara, C., Ishiyama, N., Hoshino, F. B., Ando, T., and Oya, A. (1994). Complete 5' noncoding region is necessary for the efficient internal initiation of hepatitis C virus RNA. *Biochem. Biophys. Res. Commun.* **199**(2):425–432.
- Furuya, T., and Lai, M. M. (1993). Three different cellular proteins bind to complementary sites on the 5'-end-positive and 3'-end-negative strands of mouse hepatitis virus RNA. *J. Virol.* **67**(12):7215–7222.
- Gritsun, T. S., Venugopal, K., Zanutto, P. M. de A., Mikhailov, M. V., Sall, A. A., Holmes, E. C., Polkinghorne, I., Frolova, T. V., Pogodina, V. V., Lashkevich, V. A., and Gould, E. A. (1997). Complete sequence of two tick-borne flaviviruses isolated from Siberia and the UK: Analysis and significance of the 5' and 3'-UTRs. *Virus Res.* **49**:27–39.

- Groschel, D., and Koprowski, H. (1965). Development of a virus-resistant inbred mouse strain for the study of innate resistance to Arbo B viruses. *Arch Gesamte Virusforsch.* **17**:379–391.
- Hahn, C. S., Hahn, Y. S., Rice, C. M., Lee, E., Dalgarno, L., Strauss, E. G., and Strauss, J. H. (1987). Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *J. Mol. Biol.* **198**:33–41.
- Herold, J., and Andino, R. (2001). Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol. Cell* **7**(3):581–591.
- Heus, H. A., and Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**:191–194.
- Hewlett, M. J., Petterson, R. F., and Baltimore, D. (1977). Circular forms of Uukuniemi virion RNA: An electron microscopic study. *J. Virol.* **21**:1085–1093.
- Hsu, M. T., Kung, H. J., and Davidson, N. (1973). Secondary structure of tRNA. *Cold Spring Harbor Symp. Quant. Biol.* **38**:943–950.
- Huang, P., and Lai, M. M. (2001). Heterogeneous nuclear ribonucleoprotein a1 binds to the 3'-untranslated region and mediates potential 5'-3'-end cross talks of mouse hepatitis virus RNA. *J. Virol.* **75**(11):5009–5017.
- Hungnes, O., Tjøtta, E., and Grinde, B. (1992). Mutations in the central polypurine tract of HIV-1 result in delayed replication. *Virology* **190**:440–442.
- Irie, K., Mohan, P. M., Sasaguri, Y., Putnak, R., and Padmanabhan, R. (1989). Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene* **75**:197–211.
- Jacobson, S. J., Konings, D. A., and Sarnow, P. (1993). Biochemical and genetic evidence for a pseudoknot structure at the 3' terminus of the poliovirus RNA genome and its role in viral RNA amplification. *J. Virol.* **67**:2961–2971.
- Khromykh, A., Meka, H., Guyatt, K., and Westaway, E. (2001). Essential role of cyclization sequences in flavivirus RNA replication. *J. Virol.* **75**:6719–6728.
- Khromykh, A. A., and Westaway, E. G. (1997). Subgenomic replicons of the flavivirus Kunjin: Construction and applications. *J. Virol.* **71**(2):1497–1505.
- Klovins, J., and van Duin, J. (1999). A long-range pseudoknot in QB RNA is essential for replication. *J. Mol. Biol.* **294**:875–884.
- Koonin, E. V. (1993). Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses lambda 2 protein of reovirus. *J. Gen. Virol.* **74**(Pt.4):733–740.
- Kuno, G., Chang, G., Tsuchiya, R., Karabatsos, N., and Cropp, B. (1998). Phylogeny of the genus *Flavivirus*. *J. Virol.* **72**:73–83.
- Leopardi, R., Hukkanen, V., Vainionpaa, R., and Salmi, A. A. (1993). Cell proteins bind to sites within the 3' noncoding region and the positive-strand leader sequence of measles virus RNA. *J. Virol.* **67**:785–790.
- Li, W., and Brinton, M. A. (2001). The 3' stem loop of the West Nile virus genomic RNA can suppress translation of chimeric mRNAs. *Virology* **287**:49–61.
- Lindenbach, B., and Rice, C. (1999). Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *J. Virol.* **73**:4611–4621.
- Lindenbach, B., and Rice, C. M. (1997). trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. *J. Virol.* **71**:9608–9617.
- Luo, G. (1999). Cellular proteins bind to the poly(U) tract of the 3' untranslated region of hepatitis C virus RNA genome. *Virology* **256**:105–118.
- Mackenzie, J. M., Khromykh, A. A., Jones, M. K., and Westaway, E. G. (1998). Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. *Virology* **245**:203–215.

- Mandl, C., Aberle, W., Holzmann, H., Meixner, T., Rauscher, S., Stadler, P., Allison, S., and Heinz, F. (1998). Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: Construction of highly attenuated mutants of a flavivirus. *J. Virol.* **72**:2132–2140.
- Mandl, C., Kunz, C., and Heinz, F. (1991). Presence of poly(A) in a flavivirus: Significant differences between the 3' noncoding regions of the genomic RNAs of tick-borne encephalitis virus strains. *J. Virol.* **65**:4070–4077.
- Mandl, C. W., Ecker, M., Holzmann, H., Kunz, C., and Heinz, F. X. (1997). Infectious cDNA clones of tick-borne encephalitis virus European subtype prototypic strain Neudoerfl and high virulence strain Hypr. *J. Gen. Virol.* **78**:1049–1057.
- Mandl, C. W., Holzmann, H., Kunz, C., and Heinz, F. (1993). Complete genomic sequence of Powassan virus: Evaluation of genetic elements in tick-borne versus mosquito-borne flaviviruses. *Virology* **194**:173–184.
- Markoff, L., Pang, X., Houg, H., Falgout, B., Olsen, R., Jones, E., and Polo, S. (2002). Derivation and characterization of a dengue type 1 host range-restricted mutant virus that is attenuated and highly immunogenic in monkeys. *J. Virol.* **76**:3318–3328.
- Mathews, D. H., Sabrina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**:911–940.
- Men, R., Bray, M., Clark, D., Chanock, R., and Lai, C. (1996). Dengue type 4 virus mutants containing deletions in the 3' noncoding region of the RNA genome: Analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. *J. Virol.* **70**:3930–3937.
- Monath, T. P. (1999). Yellow fever. In "Vaccines" (S. Plotkin and W. Orenstein, eds.), pp. 815–879. Saunders, Philadelphia.
- Muyllaert, I. R., Galler, R., and Rice, C. M. (1997). Genetic analysis of the yellow fever virus NS1 protein: Identification of a temperature-sensitive mutation which blocks RNA accumulation. *J. Virol.* **71**:291–298.
- Nakhasi, H. L., Cao, X. Q., Rouault, T. A., and Liu, T. Y. (1991). Specific binding of host cell proteins to the 3'-terminal stem loop structure of rubella virus negative-strand RNA. *J. Virol.* **65**:5961–5967.
- Nakhasi, H. L., Rouault, T. A., Haile, D. J., Liu, T. Y., and Klausner, R. D. (1990). Specific high-affinity binding of host cell proteins to the 3' region of rubella virus RNA. *New Biol.* **2**:255–264.
- Ng, M. L., and Hong, S. S. (1989). Flavivirus infection: Essential ultrastructural changes and association of Kunjin virus NS3 protein with microtubules. *Arch. Virol.* **106**:103–120.
- Olsthoorn, R., and Bol, J. (2001). Sequence comparison and secondary structure analysis of the 3' noncoding region of flavivirus genomes reveals multiple pseudoknots. *RNA* **7**:1370–1377.
- Pardigon, N., Lenches, E., and Strauss, J. H. (1993). Multiple binding sites for cellular proteins in the 3' end of Sindbis alphavirus minus-sense RNA. *J. Virol.* **67**:5003–5011.
- Pleij, C. W., Rietveld, K., and Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13**:1717–1731.
- Pletnev, A. (2001). Infectious cDNA clone of attenuated Langkat tick-borne flavivirus (strain E5) and a 3' deletion mutant constructed from it exhibit decreased neuroinvasiveness in immunodeficient mice. *Virology* **282**:288–300.
- Polo, S., Ketner, G., Levis, R., and Falgout, B. (1997). Infectious RNA transcripts from full-length dengue virus type 2 cDNA clones made in yeast. *J. Virol.* **71**:5366–5374.

- Poyry, T. A., Hentze, M. W., and Jackson, R. J. (2001). Construction of regulatable picornavirus IRESes as a test for current models of the mechanism of internal translation initiation. *RNA* **7**:647–660.
- Proutski, V., Gaunt, M., Gould, E., and Holmes, E. (1997a). Secondary structure of the 3'-untranslated region of yellow fever virus: Implications for virulence, attenuation and vaccine development. *J. Gen. Virol.* **78**:1543–1549.
- Proutski, V., Gould, E., and Holmes, E. (1997b). Secondary structure of the 3' untranslated region of flaviviruses: Similarities and differences. *Nucleic Acids Res.* **25**:1194–1202.
- Proutski, V., Gritsun, T., Gould, E., and Holmes, E. (1999). Biological consequences of deletions within the 3'-untranslated region of flaviviruses may be due to rearrangements of RNA secondary structure. *Virus Res.* **64**:107–123.
- Ranjith-Kumar, C. T., Gajewski, J., Gutshall, L., Maley, D., Sarisky, R. T., and Kao, C. C. (2001). Terminal nucleotidyl transferase activity of recombinant flaviviridae RNA-dependent RNA polymerase: Implication for viral RNA synthesis. *J. Virol.* **75**:8615–8623.
- Rauscher, S., Flamm, C., Mandl, W., Heinz, F., and Stadler, P. (1997). Secondary structure of the 3'-noncoding region of flavivirus genomes. Comparative analysis of base pairing probabilities. *RNA* **3**:779–791.
- Rezapkin, G. V., Dragunsky, A., Parker, M., Pomeroy, K., Asher, D. M., and Chumakov, K. M. (1998). Genetic stability of Sabin 1 strain of poliovirus: Implications for quality control of oral poliovirus vaccine. *Virology* **245**:183–187.
- Rice, C., Lenches, E., Eddy, S., Shin, S., Sheets, R., and Strauss, J. (1985). Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* **229**:726–733.
- Ruiz-Linares, A., Bouloy, M., Girard, M., and Cahour, A. (1989). Modulations of the in vitro translational efficiencies of Yellow Fever virus mRNAs: Interactions between coding and noncoding regions. *Science* **17**:2463–2476.
- Santagati, M. G., Itaranta, P. V., Koskimies, P. R., Maatta, J. A., Salmi, A. A., and Hinkkanen, A. E. (1994). Multiple repeating motifs are found in the 3'-terminal non-translated region of Semliki Forest virus A7 variant genome. *J. Gen. Virol.* **75**:1499–1504.
- Schultz, D. E., Hardin, C. C., and Lemon, S. M. (1996). Specific interaction of glyceraldehyde 3-phosphate dehydrogenase with the 5'-nontranslated RNA of hepatitis A virus. *J. Biol. Chem.* **271**:14134–14142.
- Shi, P., Brinton, M., Veal, J., Zhong, Y., and Wilson, W. D. (1996a). Evidence for the existence of a pseudoknot structure at the 3' terminus of the flavivirus genomic RNA. *Biochemistry* **35**:4222–4230.
- Shi, P., Wei, L., and Brinton, M. (1996b). Cell proteins bind specifically to West Nile virus minus-strand 3' stem-loop RNA. *J. Virol.* **70**:6278–6287.
- Spagnolo, J. F., and Hogue, B. G. (2000). Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. *J. Virol.* **74**:5053–5065.
- Spangberg, K., and Schwartz, S. (1999). PolyC-binding protein interacts with the hepatitis C virus 5' untranslated region. *J. Gen. Virol.* **80**:1371–1376.
- Strauss, J., and Strauss, E. (2002). With a little help from the host. *Science* **282**:802–808.
- Ta, M., and Vrati, S. (2000). Mov34 protein from mouse brain interacts with the 3' noncoding region of Japanese encephalitis virus. *J. Virol.* **74**:5108–5115.

- Tan, B. H., Fu, J., Sugrue, R., Yap, E. H., Chan, Y. C., and Tan, Y. H. (1996). Recombinant dengue type 1 virus NS5 protein expressed in *Escherichia coli* exhibits RNA-dependent RNA polymerase activity. *Virology* **216**:317–325.
- Utama, A., Shimizu, H., Morikawa, S., Hasebe, F., Morita, K., Igarashi, A., Hatsu, M., Takamizawa, K., and Miyamura, T. (2000). Identification and characterization of the RNA helicase activity of Japanese encephalitis virus NS3 protein. *FEBS Lett.* **465**:74–78.
- Wallner, G., Mandl, W., Kunz, C., and Heinz, F. (1995). The flavivirus 3'-non-coding region: Extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology* **213**:169–178.
- Wang, C., Le, S. Y., Ali, N., and Siddiqui, A. (1995). An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. *RNA* **1**:526–537.
- Wang, E., Weaver, S., Shope, R., Watts, D., and Barrett, A. (1996). Genetic variation in yellow fever virus: Duplication in the 3' noncoding region of strains from Africa. *Virology* **225**:274–281.
- Warrener, P., Tamura, J. K., and Collett, M. S. (1993). RNA-stimulated NTPase activity associated with yellow fever virus NS3 protein expressed in bacteria. *J. Virol.* **67**:989–996.
- Wengler, G., and Castle, E. (1986). Analysis of structural properties which possibly are characteristic for the 3'-terminal sequence of the genome RNA of Flaviviruses. *J. Gen. Virol.* **67**:1183–1188.
- Wengler, G., and Wengler, G. (1981). Terminal sequences of the genome and replicative-form of the flavivirus West Nile virus: Absence of poly(A) and possible role in RNA replication. *Virology* **113**:544–555.
- Wengler, G., and Wengler, G. (1991). The carboxy-terminal part of the NS3 protein of the West Nile flavivirus can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated ATPase. *Virology* **184**:707–715.
- Westaway, E. G., Khromykh, A. A., and Mackenzie, J. M. (1999). Nascent flavivirus RNA colocalized in situ with double-stranded RNA in stable replication complexes. *Virology* **258**:108–117.
- Wimberly, B., Varani, G., and Tinoco, I., Jr. (1993). The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* **32**:1078–1087.
- Yi, M., Schultz, D. E., and Lemon, S. M. (2000). Functional significance of the interaction of hepatitis A virus RNA with glyceraldehydes 3-phosphate dehydrogenase (GAPDH): Opposing effects of GAPDH and polypyrimidine tract binding protein on internal ribosome entry site function. *J. Virol.* **74**:6459–6468.
- You, S., Falgout, B., Markoff, L., and Padmanabhan, R. (2001). In vitro RNA synthesis from exogenous dengue viral RNA templates requires long range interactions between 5'- and 3'-terminal regions that influence RNA structure. *J. Biol. Chem.* **276**:15581–15591.
- You, S., and Padmanabhan, R. (1999). A novel in vitro replication system for dengue virus. *J. Biol. Chem.* **274**:33714–33722.
- Zeng, L., Falgout, B., and Markoff, L. (1998). Identification of specific nucleotide sequences within the conserved 3' SL in the Dengue type 2 virus genome required for replication. *J. Virol.* **72**:7510–7522.