

Complete genome sequence of Montana *Myotis* leukoencephalitis virus, phylogenetic analysis and comparative study of the 3' untranslated region of flaviviruses with no known vector

Nathalie Charlier,¹ Pieter Leyssen,¹ Cornelis W. A. Pleij,³ Philippe Lemey,² Frédérique Billoir,⁴ Kristel Van Laethem,² Anne-Mieke Vandamme,² Erik De Clercq,¹ Xavier de Lamballerie⁴ and Johan Neyts¹

^{1,2}Laboratory of Virology and Chemotherapy¹ and Laboratory of Clinical and Epidemiological Virology², Rega Institute for Medical Research, Minderbroedersstraat 10, B-3000 Leuven, Belgium

³Leiden Institute of Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

⁴Unité des Virus Emergents, Faculté de Médecine de Marseille, 27 Boulevard Jean Moulin, 13005 Marseille cedex 5, France

Montana *Myotis* leukoencephalitis virus (MMLV), a virus isolated from bats, causes an encephalitis in small rodents reminiscent of flavivirus encephalitis in humans. The complete MMLV genome is 10690 nucleotides long and encodes a putative polyprotein of 3374 amino acids. The virus contains the same conserved motifs in genes that are believed to be interesting antiviral targets (NTPase/helicase, serine protease and RNA-dependent RNA polymerase) as flaviviruses of clinical importance. Phylogenetic analysis of the entire coding region has confirmed the classification of MMLV in the clade of the flaviviruses with no known vector (NKV) and within this clade to the Rio Bravo branch (both viruses have the bat as their vertebrate host). We have provided for the first time a comparative analysis of the RNA folding of the 3' UTR of the NKV flaviviruses (Modoc, Rio Bravo and Apoi viruses, in addition to MMLV). Structural elements in the 3' UTR that are preserved among other flaviviruses have been revealed, as well as elements that distinguish the NKV from the mosquito- and tick-borne flaviviruses. In particular, the pentanucleotide sequence 5' CACAG 3', which is conserved in all mosquito- and tick-borne flaviviruses, is replaced by the sequence 5' C(C/U)(C/U)AG 3' in the loop of the 3' long stable hairpin structure of all four NKV flaviviruses. The availability of this latter sequence motif allows us to designate a virus as either an NKV or a vector-borne flavivirus.

Introduction

The *Flavivirus* genus (family *Flaviviridae*) consists of nearly 80 viruses, which can be grouped into vector-borne (mosquito- and tick-borne) flaviviruses and flaviviruses with no known arthropod vector (NKV) (Chambers *et al.*, 1990; Monath & Heinz, 1996). Several flaviviruses cause severe encephalitis in humans. These include Japanese encephalitis virus (JEV), tick-

borne encephalitis virus (TBEV), West Nile virus (WNV) and others (Han *et al.*, 1999; Heinz & Mandl, 1993). In recent years, the genomic sequences of an increasing number of flaviviruses have been determined. Almost all of the sequence data available, however, are from mosquito-borne and tick-borne flaviviruses. Recently, the sequences of the NKV flaviviruses Apoi virus (APOIV), Rio Bravo virus (RBV) and Modoc virus (MODV) have been reported (Billoir *et al.*, 2000; Leyssen *et al.*, 2002). Montana *Myotis* leukoencephalitis virus (MMLV) was first isolated in 1958 from a mouse bitten by a naturally infected little brown bat (*Myotis lucifugus*) captured in western Montana. The virus was subsequently isolated from saliva, brain and various other tissues from other bats of the same species. The biological and serological properties of the virus suggested that it belonged to the flaviviruses (Bell & Thomas,

Author for correspondence: Johan Neyts.

Fax +32 16 33 73 40. e-mail johan.neyts@rega.kuleuven.ac.be

The MMLV sequence has been submitted to EMBL (accession no. AJ299445). The MODV sequence is available at accession no. AJ242984. The nucleotide sequences of the 3' UTR of APOIV and RBV are available at accession nos AF452049 and AF452050.

1964). Based on both antigenic and molecular relationships (Kuno *et al.*, 1998), it is currently classified in the Rio Bravo virus group within the genus *Flavivirus* (Heinz *et al.*, 2000). We have used this virus to establish a small animal model of flavivirus encephalitis (accompanying paper: Charlier *et al.*, 2002).

We present here: (i) the complete genome sequence (coding and non-coding regions) of MMLV; (ii) the particular characteristics (including the phylogeny) of this genome; and (iii) a detailed and comparative study of the organization and the secondary structure of the 3' UTRs of MMLV and three other NKV flaviviruses (RBV, MODV and APOIV). Furthermore, we report that the pentanucleotide sequence CACAG, which is conserved in the 3' UTR of all arboflaviviruses, is replaced by the sequence C(C/U)(C/U)AG in NKV flaviviruses.

Methods

■ **Virus.** The original MMLV strain (Montana, 1958) was obtained from the ATCC (ATCC VR-537) and grown in Vero cells.

■ **Generation of PCR fragments.** Viral RNA was extracted from 140 µl of supernatant medium of virus-infected cells, using the QIAamp Viral RNA kit (Qiagen). A reverse transcription (RT) reaction was designed employing the reverse primer 5' GGGTCTCCTAACCTC-TAG 3'. Three sets of degenerated primers were designed based on the alignment of full-length genome sequences of different flaviviruses (Table 1). Other primers were designed based on: (i) the sequence of the amplicons generated by the first primer sets (a list of the primers is available upon request); and (ii) the sequence of a fragment of MMLV of

approximately 1 kb (nt 8929–9939; Kuno *et al.*, 1998) (Fig. 1). All PCR amplifications were achieved under standard conditions using *Taq* polymerase (HT Biotechnology) and 30 cycles including long polymerization steps (1–2 min depending on the expected size of the amplicons).

■ **Amplification of genomic termini.** The genomic 5' and 3' ends of MMLV RNA were determined by RACE (rapid amplification of cDNA ends). Amplification of the 3' UTR was achieved by the following method. An RNA oligonucleotide (5' AAGGAAAAAGCGGCCGCA-AAAGGAAAA 3') was ligated to the 3' end in a reaction mixture containing 50 µl total RNA (10 µg), 15 µl 10 × T4 RNA ligase buffer (Roche Molecular Biochemicals), 150 U T4 RNA ligase (Roche Molecular Biochemicals), 6 µl RNA oligonucleotide (2 µg), 142.5 U human placenta RNase inhibitor (HPRI; Amersham) and 62.5 µl RNase-free water. Fifty µl of the reaction product was denatured at 65 °C for 10 min (to remove secondary structures) in the presence of 150 pmol of reverse primer (5' TTTTCCTTTTGCGGCCGCTTTTTTCCTT 3') and 14 µl RNase-free water, and chilled on ice. For the RT reaction, the following were then added: 20 µl 5 × RT buffer (Amersham), 1 mM each of dATP, dTTP, dGTP and dCTP, 95 U HPRI and 40 U RAV-2 reverse transcriptase (Amersham). The mixture was incubated at 45 °C for 1.5 h and immediately chilled on ice. The resulting cDNA was used in a 50 µl PCR reaction with 60 pmol reverse primer (5' TTTTCCTTTTGCGGCCGCTTTTTTCCTT 3') and 60 pmol forward primer (5' CAGCAGTTCAGCCAACTGGGTT 3'). The 5' RACE was performed with the GeneRacer kit (Invitrogen).

■ **Cloning and sequence analysis.** PCR products were cloned into the TOPO Cloning vector (Invitrogen) or the pGEM-T Vector System I (Promega) and One Shot competent *E. coli* cells (Invitrogen) were used for transformation. The cloned inserts were sequenced in a cycle sequencing reaction with fluorescent dye terminators and analysed using an ABI 373 automatic sequencer (Perkin-Elmer).

Table 1. Flaviviruses included in the phylogenetic analysis

Virus	Vector*	Strain	GenBank no.
Apoi virus	NKV	–	AF160193
Montana <i>Myotis</i> leukoencephalitis virus	NKV	MML	AJ299445
Rio Bravo virus	NKV	RiMAR	AF144692
Modoc virus	NKV	M544	AJ242984
Langat virus	T	TP21	AF253419
Powassan virus	T	LB	L06436
Tick-borne encephalitis virus	T	Neudoerfl	U27495
Tick-borne encephalitis virus	T	Hyp	U39292
Tick-borne encephalitis virus	T	Vasilchenko	AF069066
Dengue virus type 1	M	Singapore	M87512
Dengue virus type 2	M	New Guinea	AF038403
Dengue virus type 3	M	H87	M93130
Dengue virus type 4	M	–	M14931
Japanese encephalitis virus	M	JaOArS982	M18370
Kunjin virus	M	MRM61C	D00246
Murray Valley encephalitis virus	M	Australia	NC 000943
St Louis encephalitis virus	M	MSI-7	AF160194
West Nile virus	M, T	EG-101	M12294
Yellow fever virus	M	17D	X03700
Cell-fusing agent virus	†	–	M91671

* M, Mosquito-borne; T, tick-borne; NKV, no known vector.

† Isolated only from mosquito cells (*Aedes aegypti*).

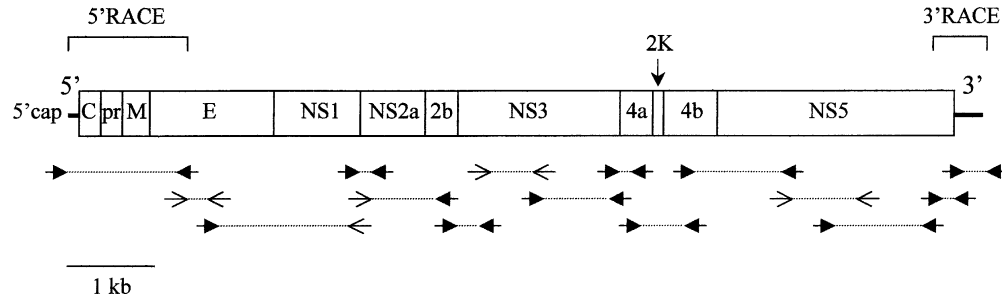


Fig. 1. Strategy employed for the sequencing of the MMLV genome. The locations of PCR fragments (dotted lines), specific primers (solid arrows) and random primers (open arrows) are indicated, along with the genomic organization of the virus.

■ **Prediction of RNA secondary structure.** The RNA secondary structure of the 3' UTRs of MMLV and of the three other NKV flaviviruses (MODV, RBV and APOIV) was analysed using the STAR program (Gultyaev *et al.*, 1995). After folding each of the four sequences separately, the resulting structures were searched for common structural elements as shown by the occurrence of covariations in the stem regions in one or more of the other three sequences. These sequences, containing proven structures, were excised and replaced by five non-pairing nucleotides. The shortened sequences were then submitted to a new cycle of folding and this was repeated until no further common elements were detected.

■ **Alignments and phylogenetic analysis.** The complete amino acid sequences of the flaviviruses listed above were aligned using the ClustalW (1.74) software (Monath & Lipman, 1988) and default alignment parameters, and manually edited in McClade (Maddison & Maddison, 1989). Conserved motifs allowed an unambiguous control of validity for alignment as previously reported (Billoir *et al.*, 2000).

Genetic distances were estimated using maximum-likelihood calculation in TreePuzzle-5.0 (Strimmer & Von Haeseler, 1996) based on the Blosum62 (Henikoff & Henikoff, 1992) model of substitution and taking into account a rate heterogeneity among sites with a discrete gamma distribution of eight categories. An unrooted phylogenetic tree based on the inferred distance matrix was constructed with NEIGHBOR in PHYLIP 3.5. Bootstrap analysis was performed according to the following algorithm: 1000 replicates were generated by SEQBOOT (PHYLIP) and redirected with Puzzleboot to TreePuzzle 5.0 where distance matrices were estimated. These distance matrices were subsequently used to infer phylogenetic trees in NEIGHBOR (PHYLIP) and a bootstrap consensus tree was generated with CONSENSE (PHYLIP).

Results

Genome and amino acid sequence

The entire genome of MMLV (EMBL accession no. AJ299445) is 10690 nucleotides long and encodes one long ORF extending from the AUG start codon at nt 109 to the first in-frame stop codon at position 10231, thereby encoding a 3374 amino acid long polyprotein (10122 nucleotides). The ORF is flanked by 5' and 3' UTRs that are, respectively, 108 and 457 nucleotides long.

Comparison of the amino acid sequence of MMLV with that of other flaviviruses revealed the presence of homologous protease cleavage sites (Table 2), internal signal sequences and transmembrane sequences (the C-terminal domains of the C, M, E and NS4A proteins are hydrophobic). As is the case for

APOIV and RBV, the mature MMLV virion C protein, envelope and NS4A genes are markedly shorter than the corresponding genes of arthropod-borne flaviviruses (Table 3).

We identified sequence motifs in the MMLV virus NS3 sequence that are known to be associated with RNA helicase activity (1748-DEAH-1751) and nucleoside triphosphate (NTP)-binding activity (1661-GSGKT-1665) (Pletnev *et al.*, 1990). The locations of these motifs proved to be perfectly conserved among the flaviviruses and identical to the motifs found in APOIV and RBV. The NS3 proteins of flaviviruses also possess the active components of a serine protease (Chambers *et al.*, 1990). The location of such sequences is also conserved in the NS3 protein of MMLV (1506-EGSFHTMW-HVTRG-1518, 1532-WANITEDLSYNGG-1545, 1590-PLD-FPPGTSGSPIITSSG-1609).

The flavivirus NS5 protein (the largest of the flavivirus-encoded proteins) encodes an RNA-dependent RNA polymerase and also contains a putative methyltransferase domain (Koonin, 1993). A heptapeptide sequence, containing the characteristic GDD sequence motif (Kamer & Argos, 1984; Poch *et al.*, 1989), is also conserved in the MMLV NS5 gene (3136-SGDDCVV-3142). The same heptapeptide motif is also present in APOIV and RBV.

Predicted cleavage sites in the MMLV polyprotein

The N-termini of the proteins were defined, based on the position of the cleavage sites of other flaviviruses. Cleavage by the viral protease generally occurs following two dibasic residues and before an amino acid with a short side chain, whereas processing with a host protease occurs at sites obeying the (-3, -1) rule (Von Heijne, 1984). Table 2 summarizes the cleavage sites for the processing of the MMLV polyprotein. At the N-termini of prM, E and NS1 of MMLV, predicted signalase cleavage sites are detected, which are also contributed by the C-terminal hydrophobic regions of anchored C, prM and E, respectively. A signal sequence also precedes the N-terminus of NS4B, suggesting that this hydrophobic protein is processed in association with endoplasmic membranes. The N-terminus of NS2A follows a cleavage site defined by the sequence Val-X-Ala (where X is Ser, Thr, Gln, Asn or Asp) (Cammisa-Parks *et al.*, 1992; Von

Table 2. Predicted protease cleavage sites in MMLV as compared with another NKV virus (RBV), a mosquito-borne virus (YFV) and a tick-borne virus (TBEV)

Cleavage site (protease)*	Cleavage after amino acid no.	MMLV	RBV†	YFV‡	TBEV‡
VirC ↓ CTHD	90	RKKQR ↓ SAKTV	KKQRR ↓ GGTES	SRKRR ↓ SHDVL	RGKRR ↓ SATDW
AnchC ↓ prM	107	ALMVA ↓ MEIEQ	TGLMA ↓ MQVSQ	LMTGG ↓ VTLVR	GMTLA ↓ ATVRK
Pr ↓ M	194	ERAKR ↓ SLVIQ	HRLKR ↓ SLSIT	RRSRR ↓ AIDL P	SRTRR ↓ SVLIP
M ↓ E	269	APNLA ↓ TNCVS	APSYS ↓ TQCVN	GPAYS ↓ AHCIG	APVYA ↓ SRCTH
E ↓ NS1	753	TGVMG ↓ DQGCV	TGVMG ↓ DHGCA	LGVGA ↓ DQGCA	LGVGA ↓ DVGCA
NS1 ↓ NS2A	1107	GLVSA ↓ QNEMS	GLVYA ↓ GSMTA	SWVTA ↓ GEIHA	SMVVA ↓ DNGEL
NS2A ↓ NS2B	1330	MKGQP ↓ SKRAT	HRGQR ↓ ATDYT	IFGRR ↓ SIPVN	HRGRR ↓ SFSEP
NS2B ↓ NS3	1463	DGKRR ↓ SLYLL	DATQR ↓ SIIVF	RGARR ↓ SGDVL	RSSRR ↓ SDLVF
NS3 ↓ NS4A	2079	AEKRR ↓ SSVLT	AQMRR ↓ SGVLL	AEGRR ↓ GAAEV	ASGRR ↓ SFGDV
NS4A ↓ 2K	2200	QGMQR ↓ TQIDT	EGMQR ↓ TQIDS	PGQQR ↓ SIQDN	AGKQR ↓ SSDDN
2K ↓ NS4B	2223	LLVFA ↓ NEMRW	VTVVA ↓ NEMRL	SAVAA ↓ NELGM	GLVAA ↓ NEMGF
NS4B ↓ NS5	2477	SPGRR ↓ GLSLS	RSDRR ↓ GIVTS	KTGRR ↓ GSANG	SGGRR ↓ GGSEG

* VirC, mature virion C protein; CTHD, C-terminal hydrophobic domain; AnchC, anchored C protein (mature virion C protein + CTHD).

† Data for RBV taken from Billoir *et al.* (2000).

‡ Data for YFV and TBEV taken from Chambers *et al.* (1990).

Table 3. Putative processing of viral polyproteins of MMLV as compared with another NKV virus (RBV), a mosquito-borne virus (YFV) and a tick-borne virus (TBEV)

The putative lengths of the 5' UTR and 3' UTR and of the respective genes are listed. The number of amino acids in each gene product is given in parentheses.

Protein	MMLV	RBV	YRV	TBEV
5' UTR	108	NA	118	132
VirC	270 (90)	264 (88)	309 (103)	297 (99)
CTHD	51 (17)	51 (17)	54 (18)	51 (17)
Pr	261 (87)	261 (87)	267 (89)	267 (89)
M	225 (75)	225 (75)	225 (75)	225 (75)
E	1452 (484)	1452 (484)	1479 (493)	1488 (496)
NS1	1062 (354)	1059 (353)	1044 (348)	1044 (348)
NS2A	669 (223)	687 (229)	693 (231)	705 (235)
NS2B	399 (133)	390 (130)	381 (127)	390 (130)
NS3	1848 (616)	1857 (619)	1869 (623)	1863 (621)
NS4A	363 (121)	357 (119)	378 (126)	378 (126)
2K	69 (23)	69 (23)	69 (23)	69 (23)
NS4B	762 (254)	774 (258)	750 (250)	756 (252)
NS5	2694 (897)	2691 (897)	2715 (905)	2709 (903)
3' UTR	457	NA	511	767
Total ORF	10122 (3374)	10137 (3379)	10233 (3411)	10242 (3414)
Total genome	10690	NA	10862	11141

NA, Not available.

Heijne, 1984). In the case of MMLV, the sequence consists of Val-Ser-Ala.

Five of the flavivirus polyprotein cleavages take place after two basic amino acids (either Lys-Arg or Arg-Arg or Arg-Lys) (Chambers *et al.*, 1990): i.e. anchored C-virion C,

NS2A-NS2B, NS2B-NS3, NS3-NS4A and NS4B-NS5. In the case of MMLV, an Arg-Arg sequence is present at the C-termini of NS2B, NS3 and NS4B. For the other two cleavages (anchored C-virion C and NS2A-NS2B), suitable dibasic sequences could not be identified. For anchored C-virion C, we

Table 4. The conserved pentanucleotide sequence (5' CACAG 3') in the last 61 nucleotides of the 3' UTR of mosquito- and tick-borne flaviviruses is a C(C/U)(C/U)AG sequence for NKV flaviviruses

Strain	Sequence alignment	Position of first nt of CACAG motif
TBEV 263	CCCAGA CAC AG AUAGUCUGACAA -GGA---	11092
TBEV ^{NEU}	CCCAGA CAC AG GUAGUCUGACAA -GGA---	11092
TBEV ^{HYPR}	CCCAGA CAC AG AUAGUCUGACAA -GGA---	10786
TBEV ^{VAS}	CCCAGA CAC AG ACAGUCUGACAA -GGA---	10878
LGTV	CCUAGA CAC AG AUAGUCUGAAAA -GGA---	10894
POWV	UCCAGC CAC AG AUAGCCUGACAA -GGA---	10790
YFV 17D-213	GUGAG- CAC AG UUUGCUCUAGAA -UAA---	10811
YFV 17DD	GUGAG- CAC AG UUUGCUCUAGAA -UAA---	10811
YFV neurotropic	GUGAG- CAC AG UUUGCUCUAGAA -UAA---	10811
YFV viscerotropic	GUGAG- CAC AG UUUGCUCUAGAA -UAA---	10811
YFV Trinidad	GUGAG- CAC AG UUUGCUCUAGAA -UAA---	10709
YFV 85-82H	GUGAG- CAC AG UUUGCUCUAGAA -UAA---	10811
DENV-1	UCCAGC CAC AG AACGCCAGAAAA UGGAAU-	10689
DENV-2	UCCAGC CAC AG AACGCCAGAAAA UGGAAU-	10677
DENV-3	UCCAGC CAC AG AACGCCAGAAAA UGGAAU-	10650
DENV-4	UCCAGC CAC AG AGCGCCGCAAGA UGGAAU-	10603
LIV	CCCAGA CAC AG AUAGUCUGACAA -GGA---	10822
WNV	CACGG- CAC AG UGCGCC-GACAU AGGUG--	10916
KUNV	CACGG- CAC AG UGCGCC-GACAA UGGUG--	11345
JEV	ACUAGC CAC AG AGCGCCGAAGUA UGUA---	10928
MVEV	-CAAGC CAC C GAGCGCCGAACAC UGUG---	10966
CFAV	--CCUCC CAC C GUUAG--GGAGUU UUGA---	10651
MMLV	--AAGA CUC AG AUUGUCUCAUGA -----	10644
RBV	U---G- CCC AG AUUGC---AUGC UGG----	NA
MODV	UCAAG- CUC AG AUUG-CUUACUA UGUA---	10552
APOIV	GUAAUC CUC AG GUUGGAUUAAAA UUAUCCU	NA

suggest a cleavage site following a Gln-Arg pair. For NS2A-NS2B cleavage may take place immediately after Gln-Pro (Gln is also present in the DEN2 and DEN4 NS2A-NS2B cleavage site) (Mandl *et al.*, 1998). These sites were chosen based on the sequence alignment with the polyproteins of other flaviviruses (Table 1) and on the notion that the dibasic sequences are usually flanked by amino acids with short side-chains, most commonly Gly, Ser or Ala. The prM protein is a glycoprotein precursor, which undergoes delayed cleavage to form M and the N-terminal 'pr' segment. Akin to all flavivirus sequences, the N-terminus of the M protein of MMLV immediately follows a pair of basic amino acids believed to represent a cleavage site for either a viral or

a host protease. The two amino acids are flanked by an amino acid with a short side-chain (Chambers *et al.*, 1990).

Characteristics of the 5'- and 3'-terminal nucleotide sequences

The ORF of the flavivirus genome is flanked by short non-coding regions, which may contain elements involved in the regulation of essential functions such as translation, replication or encapsidation of the genome (Cammissa-Parks *et al.*, 1992). The 5' UTR of MMLV is 108 nucleotides long. The MMLV 3' UTR contains 460 nucleotides. As in other flaviviruses, the 3' UTR is not extended by a poly(A) tract. At each end of the genome, two terminal nucleotides, which are conserved among

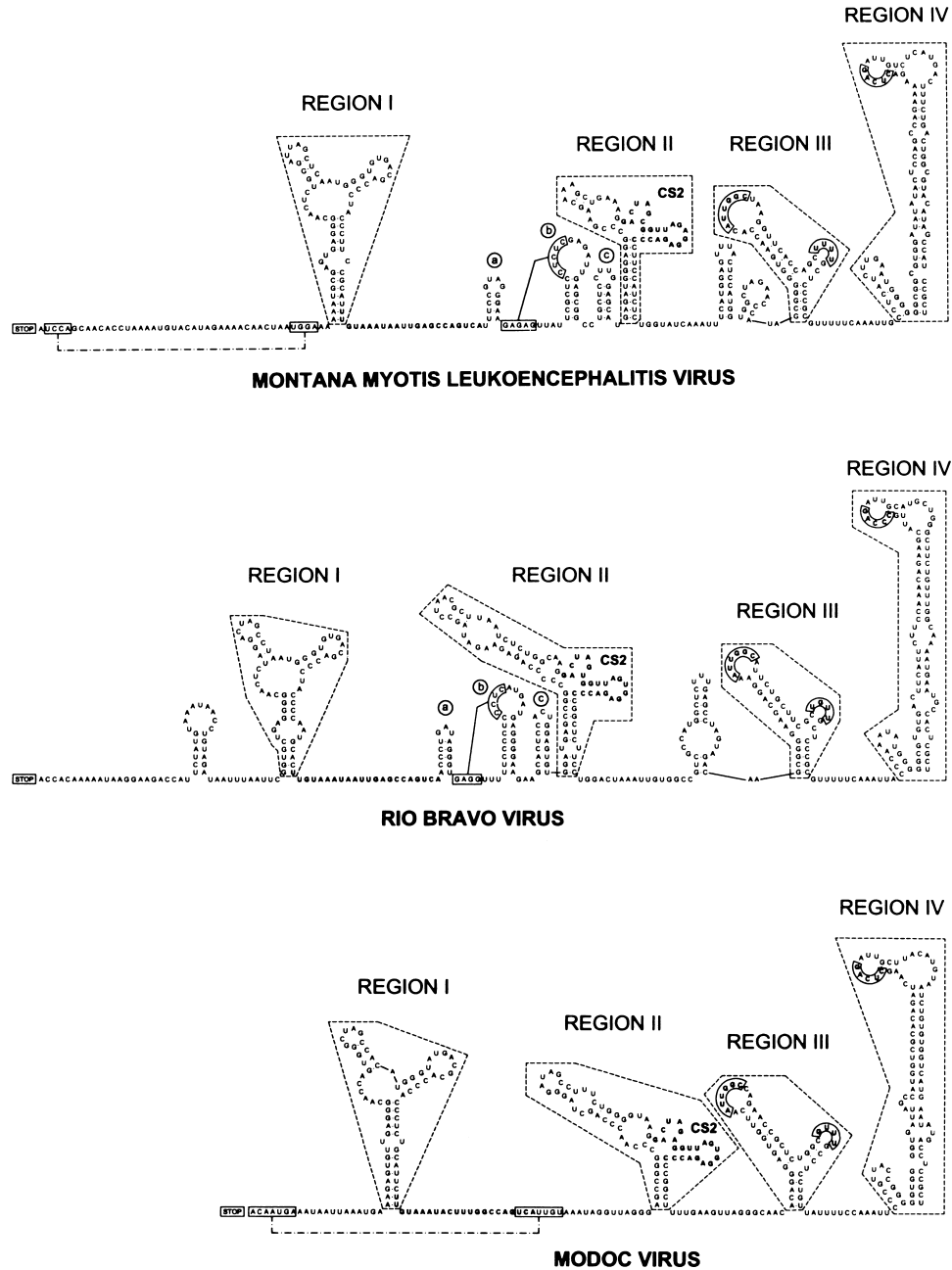


Fig. 2. For legend see facing page.

members of the whole *Flavivirus* genus, were detected, i.e. 5' AG and CU 3'. Besides these conserved terminal nucleotides, there is only one nucleotide sequence motif conserved among the mosquito- and tick-borne flaviviruses described. This conserved motif is a pentanucleotide sequence (5' CACAG 3') located approximately 45–61 nucleotides from the 3' terminus (Wengler & Castle, 1986). It is predicted that it is located on a side-loop of a conserved 3'-terminal secondary structure, suggesting that this motif would induce the formation of a circular RNA molecule, which could be important during replication or encapsidation (Chambers *et al.*, 1990; Khromykh

et al., 2001). From all 21 vector-borne flaviviruses (Table 4) that were analysed, only Murray Valley encephalitis virus (MVEV) had a different pentanucleotide sequence, i.e. an A → C change at position 4 (CACCG). We confirmed the presence of this deviating pentanucleotide sequence in the MVEV genome by sequencing this particular area of the genome of this virus. An A → C change was also noted at position 4 of this pentanucleotide sequence in the genome of cell fusing agent virus. The second position of this pentanucleotide (at an analogous position, i.e. within the loop of a 3'-terminal stem and loop structure) was either a U or a C instead of an A for all four NKV

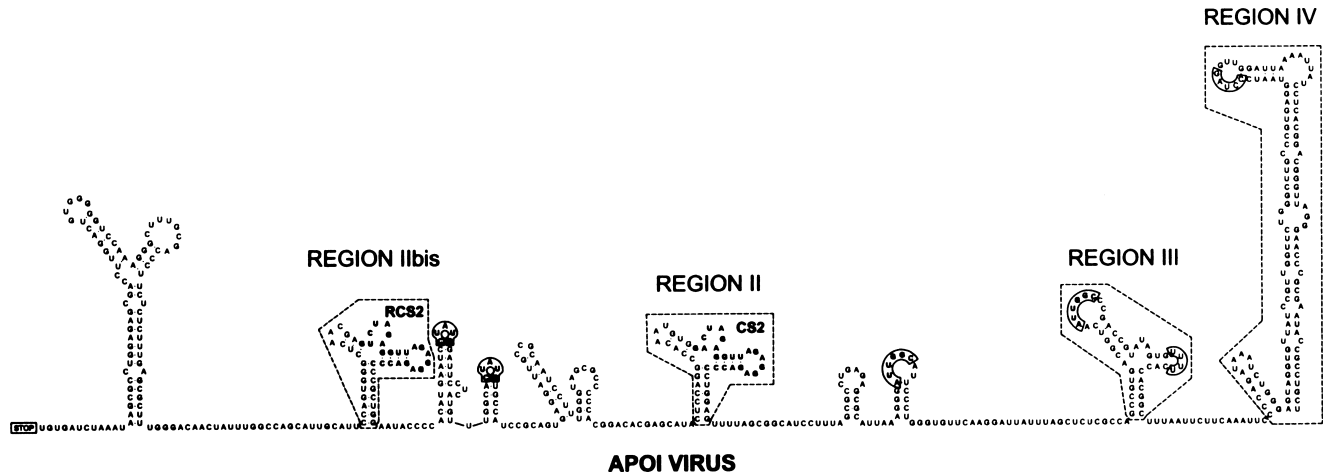


Fig. 2. Proposed secondary structure of the 3' UTR of four NKV flaviviruses. The four regions (labelled I to IV) are delineated by boxes. Conserved motifs are shown in bold and boxed. For MMLV and RBV, the predicted pseudoknot is shown by connecting boxes. For MMLV and MODV, possible stem-loops are connected by dotted/dashed lines.

flaviviruses. APOIV had, in addition, a C → U change at position 3. This pentanucleotide motif thus allows us to discriminate between NKV and vector-borne flaviviruses.

Comparative study of the folding of the 3'-terminal sequences of four NKV flaviviruses

Strong support was found for the presence of four different RNA regions (designated I, II, III and IV) in the 3' UTR of MMLV, MODV and RBV, but not in the 3' UTR of APOIV (Fig. 2). The latter is assumed not to have a region I equivalent.

Region I of MMLV, RBV and MODV consists of a long hairpin with a branching stem-loop structure. At the 3' end of region I of MMLV, RBV and MODV, a conserved motif of 22 nucleotides [5' UUGUAAUA(C/A)UU(U/G)(G/A)GCCAGUCA 3'] (labelled in bold in Fig. 2) was observed. MMLV and RBV contain exactly the same sequence, whereas MODV has A → C, G → U and A → G changes at positions 10, 13 and 14 of this sequence, respectively. This motif is not present in APOIV, which lacks region I.

The region between the stop codon and region I of the 3' UTRs of NKV flaviviruses is variable in length and is most probably single-stranded.

Region II of the 3' UTR is predicted to form a Y-shaped structure. The 3' arm of the Y structure contains the CS2 sequence [5' G(A/U)CUAGAGGUUAGAGGAGACCC 3'], which was present in all four of the NKV viruses studied. The 5' hairpin on the main stem of region II varies considerably in length for the four NKV viruses, in contrast with other flaviviruses. Upstream from region II, a repeated structure (Ilbis) is formed in the 3' UTR of APOIV, which contains the repeated conserved sequence 2 (RCS2) [5' GACUAG(A/C)GGUUAGAGGAGACCC 3']. Except for the CS2 sequence, the primary structure of region Ilbis is not identical to the sequence of region II; however, the secondary structure is well conserved.

Three very similar hairpins (a, b and c) are predicted between regions I and II for MMLV and RBV, the two NKV flaviviruses that are also, according to the phylogenetic analysis (see below), most related. These consist of a stem-loop (b), flanked by two shorter stem-loops (a and c); loop b may form a pseudoknot at the 5' side. The existence of this pseudoknot is not only supported by its prediction with the STAR program but also by the presence of one covariation in each of the two stems of the pseudoknot.

Region III folds into a Y shape for all four NKV viruses studied. The two loops of the Y structure of region III of the 3' UTRs are formed by a conserved stretch of nucleotides. However, the sequences of the stems carrying these loops are not conserved, but rather show a large number of compensatory base changes, strongly supporting the proposed secondary structure.

In region IV, the 3'-terminal nucleotides of the 3' UTR of the NKV flaviviruses form a long stable hairpin structure (3' LSH), which preserves its shape despite significant differences in sequence. This 3' LSH was calculated to fold in the genome of the four NKV flaviviruses with a similar position of the conserved C(C/U)(C/U)AG motif (45–61 nucleotides from the 3' terminus). At the 5' side of the 3' LSH, a small stem-loop (belonging to region IV and probably coaxially stacking with the long 3'-terminal hairpin) is calculated for the four NKV flaviviruses.

Inspection of the 3' UTRs revealed the existence of a 69–79 nucleotide long sequence motif (e.g. 5' GCUUUUGCUCGGCGUUUUUCAAAUUGCCUCAUCUUGAAUGG-GG GCGGCGUGGAUAUAUACUCCAGCC 3' for MMLV) located approximately 50 nucleotides away from the 3' terminus and representing an inverted repeat of another conserved sequence element located approximately 54 nucleotides from the 5' terminus (including the last 40 nucleotides of the 5' UTR and the first 29 nucleotides of the coding region).

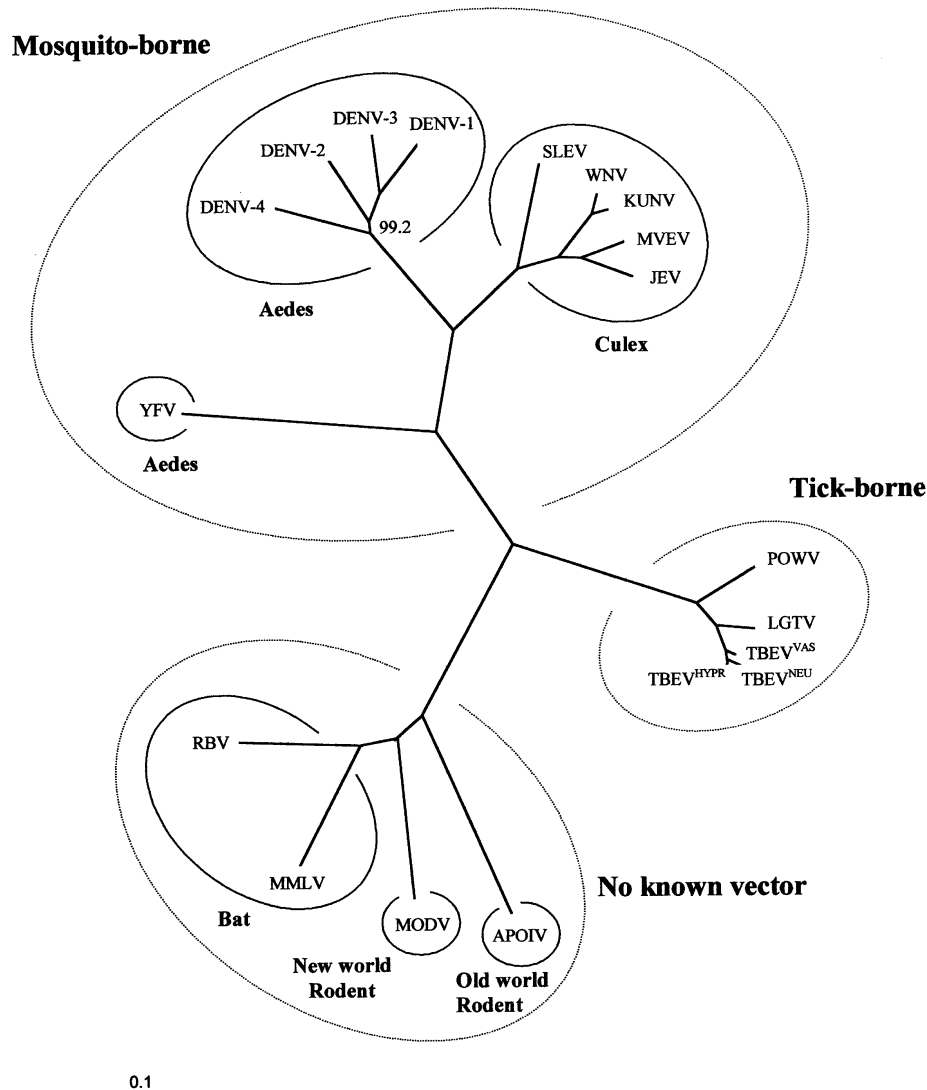


Fig. 3. Phylogenetic tree based on the complete coding region of 20 flaviviruses by the neighbour-joining method. Bootstrap statistical analysis was applied with 1000 bootstrap samples (only bootstrap values below 100% are marked). The main vectors or hosts from which viruses were isolated are indicated.

This may suggest a role in genome circularization. A similar but much smaller cyclization sequence has been observed for the tick-borne flaviviruses (Khromykh *et al.*, 2001). The predicted folding of the four regions in the 3' UTR as reported here is supported by the fact that a large number of covariant and semi-covariant sites occur in base-paired regions.

Phylogenetic analysis

A phylogenetic analysis was performed using complete coding sequences (Fig. 3). Compared with the other flaviviruses, CFAV showed a similarity below 30% and was therefore unreliable as an outgroup for phylogenetic analysis of the complete genome of the flaviviruses. An unrooted phylogenetic tree including the complete ORF sequences of 19 flaviviruses, and constructed with the neighbour-joining

method, was supported by high bootstrap values (ranging from 99.2 to 100%). Three major branches were observed: (i) the mosquito-borne virus branch; (ii) the tick-borne virus branch; and (iii) the NKV virus branch. This confirms the presence of MMLV in the group of the NKV viruses, as predicted by Kuno *et al.* (1998) using a 1 kb fragment in NS5. The bootstrap value of 100% allows us to conclude that MMLV belongs to the RBV branch, which is consistent with the fact that both viruses have the bat as their vertebrate host.

Discussion

We have determined the complete sequence of MMLV, a flavivirus with no known vector. The analysis of specific amino acid or nucleotide patterns and the phylogenetic reconstructions based on the entire flavivirus polyprotein confirmed

the taxonomic assignment of MMLV to the flaviviruses with no known vector, as previously suggested by the analysis of a 1 kb fragment of the NS5 gene (Kuno *et al.*, 1998).

Moreover, the maximum bootstrap value allowed us to conclude that MMLV belongs to the RBV branch, which is consistent with the fact that both viruses have the bat as their vertebrate host. APOIV and MODV (both isolated from rodents) are located in two distinct evolutionary branches, MODV being more closely related to MMLV and RBV than to APOIV.

The deduced amino acid sequence of MMLV revealed conservation of the main features of flaviviruses, i.e. cleavage and glycosylation sites of virus-specific proteins, and the presence of highly conserved motifs important for protease, helicase, methyltransferase and RNA-dependent RNA polymerase activity (Monath & Heinz, 1996). The fact that the genome of MMLV has the same organization as flaviviruses that are infectious to humans, as well as the same conserved regions in genes that can be considered as antiviral targets, further points to the relevance of this model in antiviral studies. Indeed, using MMLV we have established a convenient infection model for flavivirus encephalitis in SCID mice. This model may be particularly attractive for the *in vivo* evaluation of agents with anti-flavivirus activity (accompanying paper: Charlier *et al.*, 2002).

We have studied the particular characteristics of the 3' UTR of NKV flaviviruses and therefore also included the 3' UTR secondary structures of MODV, RBV and APOIV in our analysis. The 3' UTR structures of flaviviruses have previously been suggested to be organized into two distinct regions: (i) the 3'-terminal core element (approximately 330–400 nucleotides in length for mosquito- and tick-borne flaviviruses), which is, within the different serogroups, highly conserved in its primary sequence and RNA folding pattern (Wallner *et al.*, 1995; Proutski *et al.*, 1999); and (ii) the variable region, which is inserted between the core element and the coding region of the genome. This distinction corresponds with functional differences between these two regions (Proutski *et al.*, 1999). It was suggested that the variable region could possibly act as a spacer separating the folded 3' UTR structure from the rest of the genome (Blackwell & Brinton, 1995). Mandl *et al.* (1998) showed that deletion mutants of TBEV, which lack the entire variable region, replicate as efficiently as wild-type virus in cell culture and mice. Akin to the situation in mosquito- and tick-borne flaviviruses, the NKV flaviviruses also contain a variable region in their 3' UTR, and this region also varies substantially in length. Most of this variability is probably due to deletions or duplications in the region immediately following the NS5 stop codon, as suggested for the mosquito-borne flaviviruses (Shurtleff *et al.*, 2001). The core element, with its stems and loops, would constitute specific binding sites recognized by the virus-encoded replicase, cellular proteins or viral capsid proteins, and would play an important role in virus-specific transcription, translation and encapsidation (Mandl *et al.*,

1998; Gritsun *et al.*, 1997; Proutski *et al.*, 1997b; Blackwell & Brinton, 1995, 1997; Chen *et al.*, 1997). Using deletion mutants of DEN4, Proutski *et al.* (1999) proposed two parts within the core element: (i) the most 3'-terminal structures/sequences that would act as a viral promoter critical for the initiation of minus-strand RNA synthesis; and (ii) more 5' proximal structures/sequences that may function as enhancers of viral RNA replication. MODV has the shortest 3' UTR (366 nucleotides) of the NKV flaviviruses discussed here and of all flaviviruses sequenced so far. The folding pattern of this virus contains possibly the (almost) basic 3' UTR pattern that is necessary for replication of an NKV flavivirus.

The phylogenetic tree based on the UTRs of NKV flaviviruses showed similar topology to those constructed from the coding regions (data not shown), indicating that the genetic information in these regions reflects the evolutionary history of MMLV and the other NKV flaviviruses. Analysis of the folding of the 3' UTR points to the relatedness of MMLV, MODV and RBV through a common folding pattern. Folding of the 3' UTR of these three viruses revealed four conformationally conserved structural elements (regions I–IV) that are supported by compensatory mutations, which is suggestive for their functional importance. Six of the eight loops expose conserved sequence motifs. Furthermore, at the 3' end of region I, a conserved motif of 22 nucleotides [5' UUGUAA-AUA(C/A)UU(U/G)(G/A)GCCAGUCA 3'] was observed. This motif has not been described for the mosquito- or the tick-borne flaviviruses and may be a particular characteristic of New World NKV flaviviruses (MMLV, MODV, RBV). APOIV, which can be considered an NKV flavivirus of the Old World, lacks region I and thus this motif. Phylogenetically, APOIV is the most distantly positioned flavivirus within the NKV flavivirus cluster. The particular characteristics of the secondary structure of the 3' UTR of APOIV (absence of region I and presence of a duplicated region II) corroborates this observation. Interestingly, MMLV and RBV, which both have the bat as their natural host, share a common pseudoknot structure (located between regions I and II). Moreover, the sequence of region I in the 3' UTR of MMLV and RBV is very similar, whereas the stems contain compensatory mutations.

Several characteristics of the 3' UTRs of the NKV flaviviruses are comparable with those of either mosquito-borne or tick-borne flaviviruses or both. A characteristic feature similar for mosquito-borne (with the exception of YFV), but not tick-borne, flaviviruses is the presence of duplicated conserved RNA sequences (called CS2 and RCS2) (Chambers *et al.*, 1990; Proutski *et al.*, 1997b). It was assumed that they play an important role in initiating viral transcription as *cis*-acting signals, either by virtue of their exact nucleotide sequence (Hahn *et al.*, 1987; Mangada & Igarashi, 1997) or through the interaction of secondary RNA structures with cellular proteins (Blackwell & Brinton, 1995). The CS2 sequence is present in region II of the 3' UTR folding pattern of NKV flaviviruses and is located in a loop, as in the mosquito-borne flaviviruses.

According to Proutski *et al.* (1999), a single copy of either CS2 may be sufficient for normal virus replication of DENV4. However, deletion of both stem-loop structures containing CS2 and RCS2 sequences led to an inability of the mutants to replicate in mammalian cells. As is the case for the mosquito-borne flaviviruses, APOIV contains both CS2 and RCS2, whereas MMLV, MODV and RBV carry only one such sequence. This would be in line with the observation of Proutski *et al.* (1999) that only one CS2 sequence is required for the efficient replication of mosquito-borne flaviviruses. The mosquito-borne and NKV flaviviruses thus share a common factor that may possibly be important for replication in mammalian cells.

The 3' UTR of NKV flaviviruses share also particular characteristics with tick-borne flaviviruses. Like the tick-borne viruses, NKV flaviviruses lack the small stem-loop located in region I of the 3' UTR of mosquito-borne flaviviruses (Proutski *et al.*, 1997b, 1999). Deletion of this structure led to a reduced efficiency of replication of DENV4 in mosquito cells. It was suggested that this structure may function as an enhancer of virus replication in mosquito cells. The fact that NKV flaviviruses probably do not (or inefficiently) replicate in mosquito cells may reinforce this hypothesis.

Region III of the NKV flaviviruses folds into a structure similar to the one predicted in the 3' UTR of tick-borne flaviviruses [where it is part of a larger structure with three different branches of hairpins (Mandl *et al.*, 1998; Proutski *et al.*, 1997b)], but that is not present in the 3' UTR of mosquito-borne flaviviruses (Hahn *et al.*, 1987; Shurtleff *et al.*, 2001). As for the tick-borne flaviviruses, the two loops of the Y structure of region III of the NKV 3' UTRs are formed by a conserved stretch of nucleotides. These can be detected at analogous positions (5' AUUGGC 3' and 5' (G/U)(G/U)UU 3') (Gritsun *et al.*, 1997; Mandl *et al.*, 1993; Proutski *et al.*, 1997a, b).

The very 3' terminus of the 3' UTR (region IV) folds in a manner typical for all flaviviruses, forming the 3' LSH structure and a small stem-loop (belonging to region IV and probably coaxially stacking with the long 3'-terminal hairpin). The pseudoknot, which is predicted between the small stem-loop and the 3' LSH of mosquito-borne but not tick-borne flaviviruses, is probably not formed in the 3' UTR of the NKV flaviviruses. The fact that the 3' LSH is detected in mosquito-, tick-borne (Grange *et al.*, 1985; Brinton *et al.*, 1986; Hahn *et al.*, 1987; Mohan & Padmanabhan, 1991; Mandl *et al.*, 1993; Wallner *et al.*, 1995; Shi *et al.*, 1996; Proutski *et al.*, 1997a, b) and NKV flaviviruses strongly suggests that it plays a crucial role in the replication of all flaviviruses. In particular, the presence of the highly conserved pentanucleotide 5' CACAG 3' in the top loop has been suggested to play an important role in virus replication (Khromykh *et al.*, 2001). Analysis of the 3' UTRs of the NKV flaviviruses revealed a pentanucleotide sequence motif [5' C(C/U)(C/U)AG 3'] at an analogous position. However, this pentanucleotide sequence is, at positions 2 and 3, different from the 5' CACAG 3' motif

of mosquito- and tick-borne flaviviruses and appears to be unique to NKV flaviviruses. Indeed, whereas the conserved CACAG motif was detected in all flaviviruses analysed so far [with the exception of MVEV (and CFAV), which carries an A/C change at position 4], all four NKV flaviviruses analysed carry a C(C/U)(C/U)AG pentanucleotide motif. From a taxonomic point of view, knowledge of the sequence of this pentanucleotide may thus be sufficient to allocate a flavivirus either to the vector-borne or to the NKV flaviviruses. It would be of interest to unravel whether this pentanucleotide sequence plays a role at the molecular level in the fact that NKV flaviviruses are not vector-borne. Construction of an infectious full-length clone, and the introduction of mutations within this pentanucleotide motif, may provide more insight into this fascinating observation. It would also be worth determining whether vector-borne viruses, in which the pentanucleotide sequence CACAG has been replaced by the pentanucleotide sequence that is typical for NKV flaviviruses, would have an altered efficiency of replication in either mosquito or tick cells.

We thank L. Brullemans for excellent editorial assistance and S. Van Dooren and Y. Schrooten for their skilful technical assistance. N. Charlier and P. Leyssen are Research Assistants from the Instituut voor Wetenschappelijk Onderzoek & Technologie (IWT/SB/991056/Charlier and IWT/SB/981025/Leyssen) and J. Neyts is supported by the Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (FWO). This work was funded by a grant from the Geconcentreerde Onderzoeksacties – Vlaamse Gemeenschap (GOA: Project no 00/12) and the Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (G. 0122-00).

References

- Bell, J. F. & Thomas, L. A. (1964).** A new virus, 'MML', enzootic in bats (*Myotis lucifungus*) of Montana. *American Journal of Tropical Medicine and Hygiene*, 607–612.
- Billoir, F., De Chesse, R., Tolou, H., De Micco, P., Gould, E. A. & De Lamballerie, X. (2000).** Phylogeny of the genus *Flavivirus* using complete coding sequences of arthropod-borne viruses and viruses with no known vector. *Journal of General Virology* 81, 781–790.
- Blackwell, J. L. & Brinton, M. A. (1995).** BHK cell proteins that bind to the 3' stem-loop structure of the West Nile virus genome RNA. *Journal of Virology* 69, 5650–5658.
- Blackwell, J. L. & Brinton, M. A. (1997).** Translation elongation factor-1 alpha interacts with the 3' stem-loop region of West Nile virus genomic RNA. *Journal of Virology* 71, 6433–6444.
- Brinton, M. A., Fernandez, A. V. & Disposito, J. H. (1986).** The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* 153, 113–121.
- Cammisa-Parks, H., Cisar, L. A., Kane, A. & 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. & Rice, C. M. (1990).** Flavivirus genome organization, expression, and replication. *Annual Review of Microbiology* 44, 649–688.
- Charlier, N., Leyssen, P., Paeshuyse, J., Drosten, C., Schmitz, H., Van Lommel, A., De Clercq, E. & Neyts, J. (2002).** Infection of SCID mice

- with Montana *Myotis* leukoencephalitis virus as a model for flavivirus encephalitis. *Journal of General Virology* **83**, 1887–1896.
- Chen, C. J., Kuo, M. D., Chien, L. J., Hsu, S. L., Wang, Y. M. & Lin, J. H. (1997).** RNA–protein interactions: involvement of NS3, NS5, and 3′ noncoding regions of Japanese encephalitis virus genomic RNA. *Journal of Virology* **71**, 3466–3473.
- Grange, T., Bouloy, M. & Girard, M. (1985).** Stable secondary structures at the 3′-end of the genome of yellow fever virus (17 D vaccine strain). *FEBS Letters* **188**, 159–163.
- Gritsun, T. S., Venugopal, K., Zanotto, P. M., Mikhailov, M. V., Sall, A. A., Holmes, E. C., Polkinghorne, I., Frolova, T. V., Pogodina, V. V., Lashkevich, V. A. & 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 Research* **49**, 27–39.
- Gulyaev, A. P., Van-Batenburg, F. H. & Pleij, C. W. (1995).** The computer simulation of RNA folding pathways using a genetic algorithm. *Journal of Molecular Biology* **250**, 37–51.
- Hahn, C. S., Hahn, Y. S., Rice, C. M., Lee, E., Dalgarno, L., Strauss, E. G. & Strauss, J. H. (1987).** Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclization sequences. *Journal of Molecular Biology* **198**, 33–41.
- Han, L. L., Popovici, F., Alexander, J. J., Laurentia, V., Tengelsen, L. A., Cernescu, C., Gary, J. H., Ion, N. N., Campbell, G. L. & Tsai, T. F. (1999).** Risk factors for West Nile virus infection and meningoencephalitis, Romania, 1996. *Journal of Infectious Diseases* **179**, 230–233.
- Heinz, F. X. & Mandl, C. W. (1993).** The molecular biology of tick-borne encephalitis virus. *Acta Pathologica, Microbiologica et Immunologica Scandinavica* **101**, 735–745.
- Heinz, F. X., Collet, M. S., Purcell, R. H., Gould, E. A., Howard, C. R., Houghton, M., Moorman, R. J. M., Rice, C. M. & Thiel, H.-J. (2000).** *Flaviviridae*. In *Virus Taxonomy. Seventh Report of the International Committee for the Taxonomy of Viruses*, pp. 859–878. Edited by M. H. V. Van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.
- Henikoff, S. & Henikoff, J. (1992).** Amino acid substitution matrices for protein blocks. *Proceedings of the National Academy of Sciences, USA* **89**, 10915–10919.
- Kamer, G. & Argos, P. (1984).** Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269–7282.
- Khromykh, A. A., Meka, H., Guyatt, K. J. & Westaway, E. G. (2001).** Essential role of cyclization sequences in flavivirus RNA replication. *Journal of Virology* **75**, 6719–6728.
- Koonin, E. V. (1993).** Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus. *Journal of General Virology* **74**, 733–740.
- Kuno, G., Chang, G. J., Tsuchiya, K. R., Karabatsos, N. & Cropp, C. B. (1998).** Phylogeny of the genus *Flavivirus*. *Journal of Virology* **72**, 73–83.
- Leyssen, P., Charlier, N., Lemey, P., Billoir, F., Van Damme, A.-M., De Clercq, E., De Lamballerie, X. & Neyts, J. (2002).** Complete genome sequence, taxonomic assignment, and comparative analysis of the untranslated regions of the Modoc virus, a flavivirus with no known vector. *Virology* **293**, 125–140.
- Maddison, W. P. & Maddison, D. R. (1989).** Interactive analysis of phylogeny and character evolution using the computer program MacClade. *Folia Primatologica* **53**, 190–202.
- Mandl, C. W., Holzmann, H., Kunz, C. & Heinz, F. X. (1993).** Complete genomic sequence of Powassan virus: evaluation of genetic elements in tick-borne versus mosquito-borne flaviviruses. *Virology* **194**, 173–184.
- Mandl, C. W., Holzmann, H., Meixner, T., Rauscher, S., Stadler, P. F., Allison, S. L. & Heinz, F. X. (1998).** Spontaneous and engineered deletions in the 3′ noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. *Journal of Virology* **72**, 2132–2140.
- Mangada, M. N. & Igarashi, A. (1997).** Sequences of terminal non-coding regions from four dengue-2 viruses isolated from patients exhibiting different disease severities. *Virus Genes* **14**, 5–12.
- Mohan, P. M. & Padmanabhan, R. (1991).** Detection of stable secondary structure at the 3′ terminus of dengue virus type 2 RNA. *Gene* **108**, 185–191.
- Monath, T. P. & Heinz, F. X. (1996).** Flaviviruses. In *Fields Virology*, 3rd edn, vol. 1, pp. 961–1034. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.
- Monath, W. R. & Lipman, D. J. (1988).** Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**, 2444–2448.
- Pletnev, A. G., Yamshchikov, V. F. & Blinov, V. M. (1990).** Nucleotide sequence of the genome and complete amino acid sequence of the polyprotein of tick-borne encephalitis virus. *Virology* **174**, 250–263.
- Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989).** Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO Journal* **8**, 3867–3874.
- Proutski, V., Gaunt, M. W., Gould, E. A. & Holmes, E. C. (1997a).** Secondary structure of the 3′-untranslated region of yellow fever virus: implications for virulence, attenuation and vaccine development. *Journal of General Virology* **78**, 1543–1549.
- Proutski, V., Gould, E. A. & Holmes, E. C. (1997b).** Secondary structure of the 3′-untranslated region of flaviviruses: similarities and differences. *Nucleic Acids Research* **25**, 1194–1202.
- Proutski, V., Gritsun, T. S., Gould, E. A. & Holmes, E. C. (1999).** Biological consequences of deletions within the 3′ untranslated region of flaviviruses may be due to rearrangements of RNA secondary structure. *Virus Research* **64**, 107–123.
- Shi, P. Y., Brinton, M. A., Veal, J. M., Zhong, Y. Y. & Wilson, W. D. (1996).** Evidence for the existence of a pseudoknot structure at the 3′ terminus of the flavivirus genomic RNA. *Biochemistry* **35**, 4222–4230.
- Shurtleff, A. C., Beasley, D. W., Chen, J. J., Ni, H., Suderman, M. T., Wang, H., Xu, R., Wang, E., Weaver, S. C., Watts, D. M., Russell, K. L. & Barrett, A. D. (2001).** Genetic variation in the 3′ non-coding region of dengue viruses. *Virology* **281**, 75–87.
- Strimmer, K. & Von Haeseler, A. (1996).** Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13**, 964–969.
- Von Heijne, G. (1984).** How signal sequences maintain cleavage specificity. *Journal of Molecular Biology* **173**, 243–251.
- Wallner, G., Mandl, C. W., Kunz, C. & Heinz, F. X. (1995).** The flavivirus 3′-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology* **213**, 169–178.
- Wengler, G. & Castle, E. (1986).** Analysis of structural properties which possibly are characteristic for the 3′-terminal sequence of the genome RNA of flaviviruses. *Journal of General Virology* **67**, 1183–1188.

Received 14 December 2001; Accepted 10 March 2002