ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

VIROLOGY

Virology 313 (2003) 81–90

www.elsevier.com/locate/yviro

Analysis of the complete genome of the tick-borne flavivirus Omsk hemorrhagic fever virus

Di Lin,[a] Li Li,[a] Daryl Dick,[b] Robert E. Shope,[a] Heinz Feldmann,[b] Alan D.T. Barrett,[a] and Michael R. Holbrook[a,*]

[a] *Department of Pathology, Center for Biodefense and Emerging Infectious Diseases, and Sealy Center for Vaccine Development, University of Texas Medical Branch, Galveston, TX 77555-0609, USA*

[b] *Canadian Science Center for Human and Animal Health, National Microbiology Laboratory, Health Canada, Winnipeg, Manitoba, Canada, and University of Manitoba, Department of Medical Microbiology, Winnipeg, Manitoba R3E 3R2, Canada*

Received 8 January 2003; returned to author for revision 13 February 2003; accepted 13 March 2003

Abstract

Omsk hemorrhagic fever virus (OHF) is a tick-borne flavivirus endemic to Western Siberia. This virus is the only known tick-borne flavivirus to cause hemorrhagic disease in humans in the absence of encephalitis. OHF virus circulates within a small, defined niche in which other tick-borne complex flaviviruses are also present. The objectives of this study were to genetically classify OHF virus based on its complete genome and to identify genetic determinants that might be involved in tissue tropism and viral replication leading to the disease state caused by this virus. The OHF virus genome was sequenced and phylogenetic analysis demonstrated that OHF virus falls within the tick-borne encephalitis serocomplex of flaviviruses, yet is distinct from other members of the complex, including those closely associated geographically. OHF is also distinct from Alkhurma (ALK) and Kyasanur forest disease (KFD) viruses, both of which cause disease that includes hemorrhagic and encephalitic manifestations. Several amino acid residues were found to be distinct among OHF, KFD, and ALK viruses; these residues include E-76, which is closely associated with the viral envelope protein fusion peptide. In addition, variation between the viral 5′-untranslated region of OHF and other tick-borne flaviviruses suggests potential variability in viral replication. These data demonstrate that OHF is a unique virus among the tick-borne flaviviruses and also provide insight to viral biodiversity and tropism.

Keywords: Flavivirus; Omsk hemorrhagic fever; Tick-borne encephalitis; Genome sequence

Introduction

Omsk hemorrhagic fever (OHF) is a disease endemic to the Omsk and Novosibirsk Oblast region of Russia, in close proximity to the cities of Omsk and Novosibirsk (Kharitonova and Leonov, 1985). The causative agent of this disease is OHF virus, a flavivirus that is transmitted via the bite of an infected ixoid tick. The original isolate of OHF virus was from human serum during an epidemic in 1947 (Burke and Monath, 2001). Serological examination of subsequent viral isolates suggests two antigenic subtypes of OHF virus (Theiler and Downs, 1973). The principal arthropod vector of OHF virus is *Dermacentor reticulates*, though *Ixodes spp.* are believed to be involved in the maintenance of the virus in its sylvatic cycle (Kharitonova and Leonov, 1985). The sylvatic cycle of OHF virus involves muskrats (*Ondata zibethica*) and water voles (*Arvicola terrestris*), though most animals within endemic areas can be infected with the virus. Thus, this is a common disease of trappers who come in contact with infected animals. Although transovarial transmission is not believed to play a significant role in the persistence of the virus in the sylvatic cycle, OHF virus can persist for extended periods in ixoid ticks. The virus is apparently stable in adult ticks for the overwintering period and transstadial transmission also appears to be involved in viral persistence (Kharitonova and Leonov, 1985).

* Corresponding author. Department of Pathology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0609. Fax: +1-409-747-2415.

E-mail address: mrholbro@utmb.edu (M.R. Holbrook).

0042-6822/03/$ – see front matter
doi:10.1016/S0042-6822(03)00246-0

OHF virus is transmitted primarily via the bite of infected ticks. Gamasid mites are also thought to play a minor role in transmission within the sylvatic cycle (Kharitonova and Leonov, 1985). In comparison, transmission by mosquitoes, fleas, or biting midges has not been demonstrated. OHF virus has also been found to be transmitted by ingestion of contaminated goat and sheep milk and via processing of carcasses of infected animals (Kharitonova and Leonov, 1985).

OHF virus is classified as a biosafety level 4 virus and is a member of the tick-borne encephalitis (TBE) serocomplex, genus *Flavivirus*, family *Flaviviridae.* OHF virus is one of only three known tick-borne flaviviruses that cause a hemorrhagic disease, the other two being Alkhurma virus (ALK) (Zaki, 1997) and Kyasanur Forest disease virus (KFD) (Work, 1958). Unlike ALK and KFD viruses, however, OHF virus is the only known tick-borne flavivirus that causes hemorrhagic disease with minimal or infrequent evidence of neurological involvement. Clinical signs of the OHF are fever, headache, myalgia, dehydration, and hemorrhage (Pavri, 1989). OHF has a case fatality rate of 0.5–3% (Burke and Monath, 2001), which is lower than the 3–5% rate seen in KFD.

The objective of this study was to determine the complete genome sequence of OHF virus and to compare it to complete genome sequences of other tick-borne serocomplex viruses, including the hemorrhagic ALK virus (the complete genome sequence for KFD virus has not been reported), in an effort to identify molecular pathogenic determinants of the OHF virus, or hemorrhagic TBE viruses, that may be involved in the hemorrhagic manifestations of disease. Phylogenetic analysis of the complete viral genome found that OHF virus falls within the TBE serocomplex, but is distinct from other members of the serocomplex. These data support previously reported sequencing data limited only to the OHF virus E protein (Gritsun et al., 1993). Comparison of the E protein amino acid sequence of neurotropic TBE serocomplex viruses with those that cause hemorrhagic manifestations identified four amino acid residues that were specific to the hemorrhagic TBE viruses and others that were specific for OHF virus alone, suggesting specificity in tissue and vector tropism. These data represent the first analysis of differential hemorrhagic TBE viruses and comparison to genetically and geographically related viruses.

Results

Sequencing and phylogenetic analysis

The genome of OHF virus strain Bogoluvovska (Clarke, 1964) was determined to comprise 10,787 nucleotides with an open reading frame of 10,242 nucleotides encoding a 3414 amino acid polyprotein. The size of the OHF viral proteins, their calculated molecular weight, and their cleavage sites are given in Table 1. The cleavage sites between viral proteins within the OHF open reading frame are completely conserved with those described for other tick-borne flaviviruses (Table 1B), as described by Charrel et al. (2001). Multiple levels of phylogenetic analysis were performed on the OHF viral genome sequence using maximum likelihood analyses with supporting bootstrap analysis.

Table 1
Genetic characteristics of OHF virus

A. OHF gene and protein sizes

Gene	Nucleotides	Amino acids
5′ UTR	132	—
C	348	116
prM	492	164
(M)	(225)	(75)
E	1488	496
NS1	1059	353
NS2A	687	229
NS2B	393	131
NS3	1863	621
NS4A	378	126
2K	69	23
NS4B	756	252
NS5	2709	903
3′ UTR	413	—
Total	10,787	3414

B. Protease cleavage sites in OHF polyprotein

Proteins	Amino acid sequence
5′ UTR/C	gggg/MAGK
C/prM	SIALA/ATVRK
M/E	TRSRR/SVLIP
E/NS1	LGVGA/DVGCA
NS1/NS2A	MVVAD/NGELL
NS2A/NS2B	RRDRR/SFSEP
NS2B/NS3	RSARR/SDLVF
NS3/NS4A	ASGRR/SLGDM
NS4A/2K	AGKQR/SSDDN
2K/NS4B	GLVAA/NEMGF
NS4B/NS5	SGTRR/GGSEG
NS5/3′ UTR	SSII/taacca

Genome analysis

Maximum likelihood phylogenetic analysis of the complete genome of OHF virus and comparison with that of other flaviviruses (Table 2) found that OHF virus clusters among the tick-borne flaviviruses (Fig 1). However, OHF forms its own clade among the tick-borne flaviviruses, separating it from other TBE serocomplex viruses including the Siberian subtype of virus that is also found in the geographical region where OHF is endemic. Interestingly, OHF virus formed a clade that was separate and distinct from ALK virus, another TBE serocomplex virus that causes a disease with hemorrhagic manifestations (note that the entire genome sequence of KFD virus, highly homologous to ALK based on the E protein sequence, has not been determined).

Table 2
Viruses used in this study

Virus species	Abbreviation	GenBank Accession No.	Seq. type	Other info.
Alkhurma virus	ALK	AF331718	Complete	Human isolate, Saudi Arabia
Langat virus	LGT	AF253419	Complete	Strain TP21
Kyasanur Forest disease virus	KFD	X74111	Envelope	Human isolate, India
Yellow Fever virus	YF	Not in GenBank	Complete	Asibi strain
Japanese encephalitis virus	JE	M18370	Complete	Strain JaOArS982
Dengue virus type 1	DEN-1	U88535	Complete	Strain Western Pacific
Dengue virus type 2	DEN-2	U87411	Complete	Strain 16681
Dengue virus type 3	DEN-3	M93130	Complete	Strain H87
Dengue virus type 4	DEN-4	M14931	Complete	Strain 814669, Dominica
West Nile virus	WN	AF260969	Complete	Strain RO97-50
Rio Bravo virus	RB	AF144692	Complete	Strain RiMAR, no known vector
Powassan virus	POW	NC_003697	Complete	Strain, LB, Canada
Louping ill virus	LI	NC_001809	Complete	Strain 369/T2
Tick-borne encephalitis virus	TBE^{263}	U27491	Complete	Western subtype
Tick-borne encephalitis virus	TBE^{NEU}	U27495	Complete	Western subtype
Tick-borne encephalitis virus	TBE^{HYPR}	U39292	Complete	Western subtype
Tick-borne encephalitis virus	TBE^{OSH}	AB062063	Complete	Oshima5–10, Japan
Tick-borne encephalitis virus	TBE^{VAS}	AF069066	Complete	Siberian subtype, Siberia
Tick-borne encephalitis virus	TBE^{SOF}	AB062064	Complete	Far-Eastern subtype, Far-East Russia
Omsk hemorrhagic fever virus	OHF	This paper	Complete	Strain Bogoluvovska

To determine whether codon bias could explain the particular host, vector, or tissue tropism of OHF virus, codon usage analysis was also performed on several tick-borne flaviviruses including Western subtype TBE virus strain Neudorfl, Siberian TBE virus strain Vasilchenko, Eastern TBE virus strain Sofjin, and ALK virus. This analysis did not identify any particular codon bias in OHF virus polyprotein that distinguished it from other TBE serocomplex viruses or would suggest that differential codon usage is responsible for viral phenotype (data not shown).

E protein analysis

Analysis of the OHF virus E protein nucleotide sequence in the context of the E protein from a large number of flaviviruses grouped the OHF virus E protein clearly among the TBE serocomplex of viruses. Subsequently, maximum likelihood phylogenetic analysis of the E nucleotide sequence was focused to include only members of the TBE serocomplex (Fig. 2). As was seen in phylogenetic analysis of complete viral genomes, the OHF virus E protein gene did not clearly fall within the three major subtypes of TBE virus, but rather as its own distinct subtype, or clade. These results correlate with previously described results (Gritsun et al., 1993; Hayasaka et al., 1999). Comparison of nucleotide and amino acid identity of OHF virus E protein with other TBE viruses (Table 3A) found that OHF virus had a moderate degree of similarity with other TBE serocomplex viruses, especially when compared to the level of similarity between members of the Western/Siberian/Far-Eastern subtypes. These results indicate that while many of these viruses have considerable variation within their genetic sequence, they retain a large degree of sequence identity at the protein level, presumably allowing the viruses to retain major functional motifs within the viral E protein. Molecular modeling of E protein amino acid sequences from several TBE complex viruses using the Western subtype of TBE virus (strain Neudorfl) crystal coordinates (Rey et al., 1995) supports this notion (data not shown). Interestingly, however, OHF virus is very distinct at both the nucleotide and the amino acid level from the ALK and KFD viruses (Table 3A), the only other TBE viruses to cause disease with hemorrhagic manifestations.

NS5 analysis

Similar analyses were performed on the NS5 gene of OHF virus to determine if the degree of nucleotide and amino acid similarity was consistent within viral proteins with significantly different functions. Comparison of the OHF virus NS5 nucleotide and amino acid sequences with other TBE complex viruses supported the results from analyses of the E protein gene wherein the similarity of OHF virus NS5 with that of the Western, Siberian, or Far-Eastern TBE subtypes was virtually indistinguishable (Table 3B). Maximum likelihood phylogenetic analysis of the OHF virus NS5 gene supported the results found for the E protein gene (data not shown).

OHF virus is clearly a member of the TBE serocomplex, yet it is genetically distinct from geographically related viruses. OHF virus is also very distinct from the hemorrhagic ALK virus at the genome level, and ALK and KFD viruses at the E protein level. All told, these data support the hypothesis that there may be significant biological factors

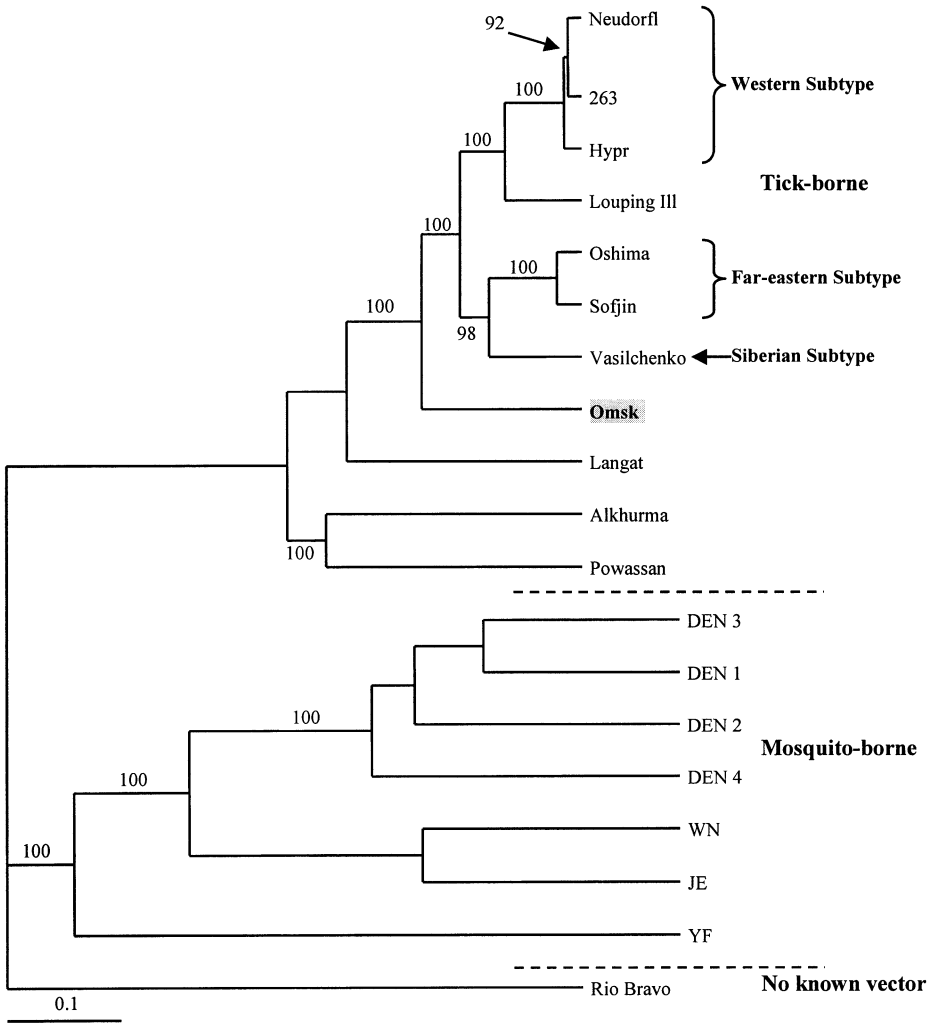

Fig. 1. Maximum likelihood phylogenetic analysis of the complete genome of several flaviviruses. Viruses and abbreviations are as noted in Table 2. The tree was rooted with the non-vector-borne flavivirus Rio Bravo. Bootstrap values, given as a percentage of 500 replicates, are noted on the tree for major points of divergence.

that confine the virus to the Omsk and Novosibirsk Oblast region and to its vertebrate and invertebrate hosts. This biological pressure may have allowed the OHF virus to evolve in a lineage distinct from other TBE viruses endemic to the same region of Siberia. Phylogenetic analysis of the TBE serocomplex viruses suggests an evolutionary cline from East to West across Asia with OHF virus representing an older lineage of virus (Gould et al., 2001; Zanotto et al., 1995). To further examine genetic factors that may contribute to this apparent geographical confinement and to the hemorrhagic manifestation of disease, structural comparisons of the E protein and untranslated regions (UTR) were made between OHF and other TBE viruses.

Structural protein similarities

The flavivirus E protein is the major immunogen for this genus of viruses and is also involved in cell receptor attachment and membrane fusion (Heinz and Allison, 2000, 2001). An examination of the OHF virus E protein was undertaken to determine if there were residues specific to OHF and other hemorrhagic tick-borne flaviviruses that might precipitate the hemorrhagic manifestations of disease. Comparisons of the E protein amino acid sequence were made between nine tick-borne flaviviruses: the three hemorrhagic fever viruses (OHF, KFD, and ALK), two Far-Eastern/Siberian subtype viruses (Sofjin and Vasilchenko), two Western subtype viruses (Hypr and Neudorfl), and Langat and Powassan viruses. The latter two viruses were included in the sequence comparison and casual examination of the amino acid alignments identified several residues that were common in the hemorrhagic fever viruses present in either Langat or Powassan virus.

When comparing amino acid alignments, 28 amino acids (of 496, 5.6%) were found to be specific to OHF virus E protein when compared to two Far-Eastern/Siberian TBE

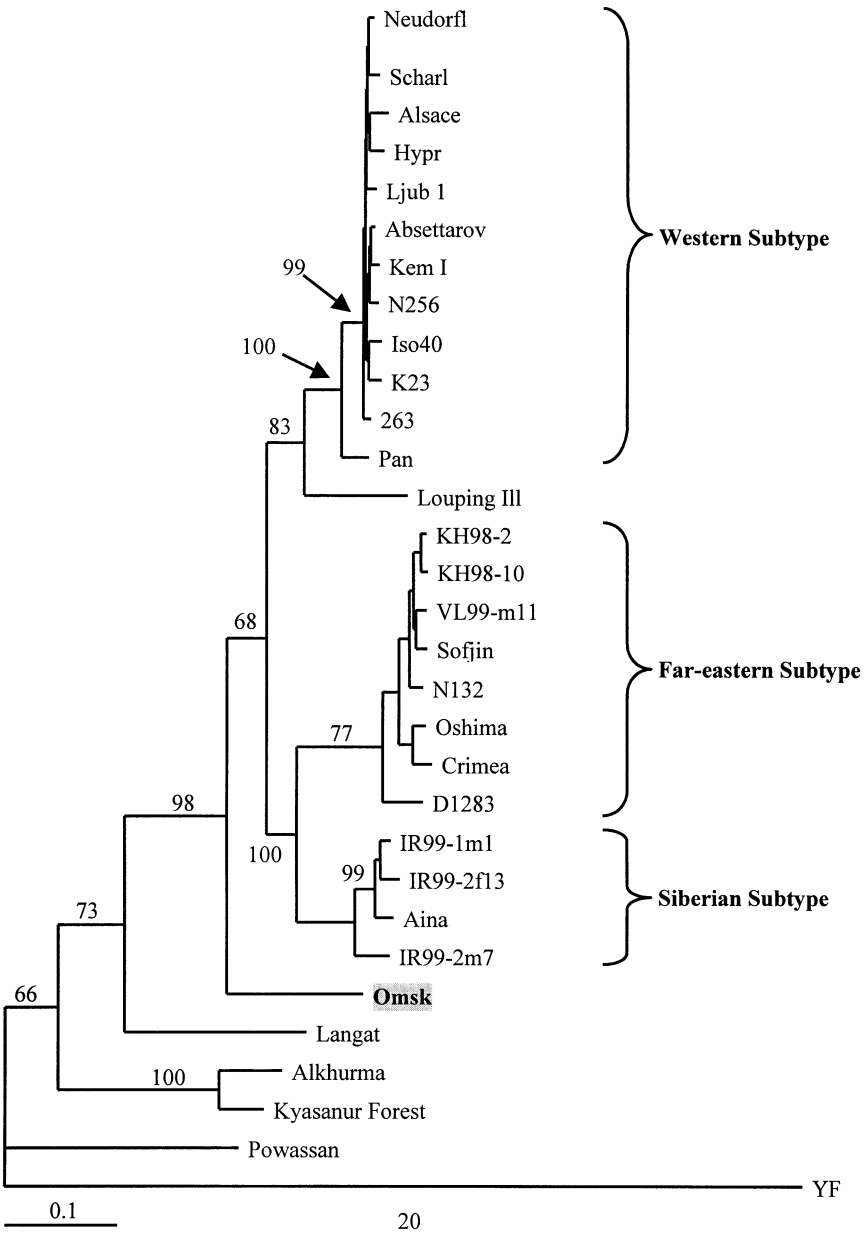

Fig. 2. Maximum likelihood phylogenetic analysis of the envelope protein gene of several tick-borne flaviviruses. In addition to the viruses listed in Table 2, the following viruses were included in this tree: Absettarov-F091005, Als I-AF091007, Crimea-AF091008, D1263-AB049347, IR 1m1- AB049348, IR 2fl3-AB049353, IR99 2m7-AB049351, Iso40-AF091009, K23-AF091010, Kem I-AF091011, KH98 2-AB022295, KH98 10-AB022297, Ljub-AF091012, N132-AF091013, N256-AF091014, Pan-AF091015, Scharl-AF091017, VL99 m11-AB049345. The tree was rooted with the mosquito-borne flavivirus yellow fever (Asibi strain). Bootstrap values, given as a percentage of 500 replicates, are noted on the tree.

viruses (data not shown). When including the eight viruses listed above in the comparison with OHF, this number was reduced to 13 (2.6%), and 3 (0.6%) when looking for residues specific for hemorrhagic fever viruses (ALK, KFD, and OHF) (Table 4). The degree of amino acid variation within the OHF virus E protein is actually less than in the other viral surface protein, the OHF virus M protein, which had a 10.6% variation from the Far-Eastern/Siberian viruses (data not shown) and 1.3% variation from nonhemorrhagic TBE complex viruses (Table 4). This result suggests that of the two viral surface proteins, the structural architecture of the viral E protein is more critical to proper viral function, whereas the structure of the M protein allows for some degree of structural plasticity while retaining proper function. The proposed role of the M protein is to function as a chaperone in its intracellular preprotein form (preM) and as a stabilizer of E protein conformation on the surface of the mature virion (Heinz and Allison, 2000, 2001).

Comparison of the E protein of the three hemorrhagic TBE viruses against that of several other TBE serocomplex viruses identified three residues that are common in all three hemorrhagic viruses and are not found in the other TBE

Table 3
A. Comparison of the E protein gene

AA Virus	Nucleotide identity											
	Sib.	Far-eastern		Western								
	TBE^{VAS}	TBE^{OSH}	TBE^{SOF}	TBE^{263}	TBE^{NEU}	TBE^{HYPR}	POW	LI	LGT	KFD	ALK	OHF
TBE^{VAS}	—	84.8	84.7	84.7	84.9	84.7	70.2	82.9	75.9	72.9	72.2	81.0
TBE^{OSH}	97.2	—	95.7	84.5	84.3	84.1	70.0	81.9	75.3	72.2	71.4	80.3
TBE^{SOF}	97.2	99.2	—	84.1	84.3	84.2	70.8	82.8	75.5	72.1	71.1	80.7
TBE^{263}	96.0	95.8	95.8	—	98.3	97.5	69.4	87.0	75.0	73.6	72.7	81.9
TBE^{NEU}	96.2	95.8	95.8	99.4	—	97.8	69.1	86.9	75.3	73.9	73.1	81.8
TBE^{HYPR}	95.8	95.4	95.4	98.6	98.8	—	69.9	87.0	75.5	73.3	73.0	82.4
POW	78.6	78.2	78.2	78.8	78.2	78.2	—	69.6	70.1	70.0	71.0	70.4
LI	92.5	91.3	91.3	94.0	93.8	93.1	76.8	—	74.4	73.1	72.2	79.4
LGT	88.1	88.7	88.3	87.9	87.5	87.5	78.0	84.9	—	72.6	72.5	74.7
KFD	81.7	81.3	80.8	80.8	81.3	80.6	77.4	79.4	79.8	—	91.0	71.7
ALK	81.3	80.4	80.0	80.2	80.6	80.2	77.8	79.2	79.8	97.0	—	71.3
OHF	92.9	94.0	94.0	93.1	93.1	92.9	77.6	89.7	86.9	80.6	79.8	—

B. Comparison of the NS5 gene

AA Virus	Nucleotide identity										
	Sib.	Far-eastern		Western							
	TBE^{VAS}	TBE^{OSH}	TBE^{SOF}	TBE^{263}	TBE^{NEU}	TBE^{HYP}	POW	LI	LGT	ALK	OHF
TBE^{VAS}	—	86.0	86.6	86.3	85.5	85.9	73.7	83.7	77.5	74.1	81.5
TBE^{OSH}	94.3	—	95.4	84.9	84.4	84.7	74.2	82.4	76.4	73.9	80.2
TBE^{SOF}	95.1	98.6	—	85.2	84.6	84.9	73.9	82.8	76.3	74.1	80.2
TBE^{263}	95.1	93.6	94.1	—	97.5	97.9	74.1	88.4	76.9	73.8	81.5
TBE^{NEU}	94.7	93.5	94.0	98.6	—	97.6	73.9	88.6	76.4	73.5	81.4
TBE^{HYP}	94.7	93.7	94.2	99.1	98.8	—	74.1	88.3	76.7	73.8	81.2
POW	84.1	83.3	83.1	83.7	83.8	83.9	—	73.1	73.9	71.7	73.1
LI	91.5	91.5	91.6	94.2	94.3	94.2	82.3	—	76.6	74.6	80.3
LGT	87.4	88.0	88.1	87.3	87.5	87.5	83.3	86.0	—	74.2	76.7
ALK	86.3	84.9	85.7	84.9	85.4	85.3	84.1	83.8	85.7	—	75.5
OHF	91.6	91.5	91.9	92.2	92.5	92.6	83.0	89.7	87.7	84.9	—

viruses examined. One of these residues, E-76, is situated in domain II of the E protein, while the other two, E-457 and E-489, are located in the stem-anchor region of the viral E-protein. Residue E-76 is situated near the putative membrane fusion peptide of the E protein (Allison et al., 2001), though not within the fusion peptide cd loop (Fig 3). Residue E-76 resides in a loop immediately adjacent to the cd loop and appears to have Van der Waals interactions with the fusion peptide. The change of residue E-76 from a Thr in encephalitic TBE viruses to an Ala in the viruses that cause hemorrhagic disease may alter the tertiary structure of the fusion peptide to alter the ability of the E protein to fuse with endosomal membranes. The physical location of Thr76 indicates that it may be directly involved in Van der Waals interactions with Leu107 of the cd loop, a loop that connects strands c and d in domain II of the Western subtype E protein (Rey et al., 1995). The presence of this change in a highly conserved loop that contains a cysteine and two flanking proline residues may allow a Thr → Ala mutation to have a stronger affect that would be immediately apparent. In addition, the loss of a hydrogen bond (hydroxyl group on Thr) may have an effect on the interaction with a second E protein in the E protein dimer. The amino acid sequence is highly conserved in both the E-76 loop and the cd loop, yet modeling of the OHF virus E protein based on the Western subtype TBE virus (strain Neudorfl) crystal structure indicates some shifting of side-chain orientation,

Table 4
Amino acids common to HF[a]-associated viruses or OHF specific amino acids

	HF[a] specific		OHF specific	
Protein	AA	%	AA	%
C	1	0.9	12	10.3
prM	1	0.6	13	7.9
(M)	(1)	(1.3)	(11)	(14.7)
E	3	0.6	13	2.6
NS1	4	1.1	14	4.0
NS2A	3	1.3	17	7.4
NS2B	1	0.8	7	5.3
NS3	4	0.6	12	1.9
NS4A	0	0.0	8	6.3
2K	0	0.0	0	0.0
NS4B	0	0.0	3	1.2
NS5	3	0.3	14	1.6
n	20	0.6	113	3.3

[a] Hemorrhagic fever associated TBE complex viruses.

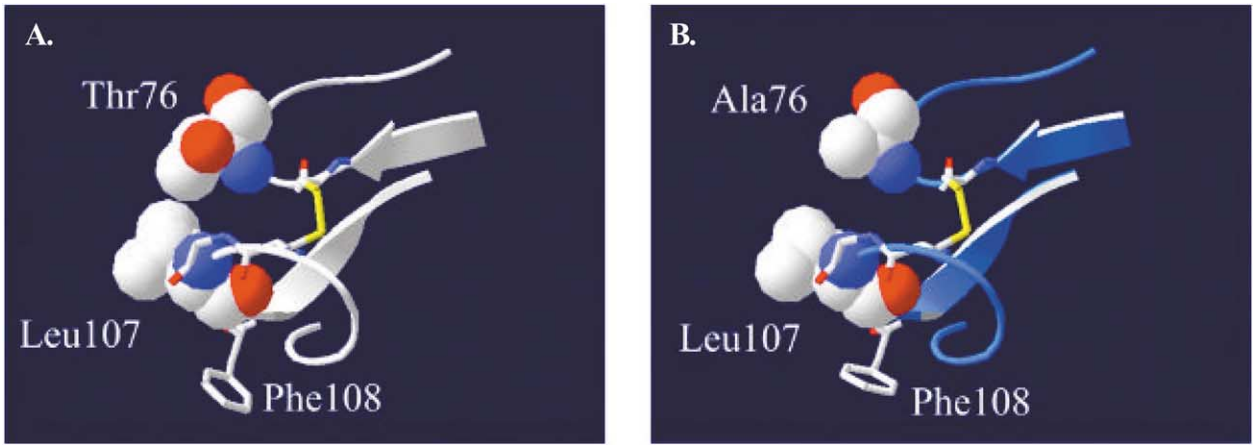

Fig. 3. Model of the TBE E glycoprotein fusion peptide. (A) Structure of the E protein fusion peptide as derived from the crystal coordinates of Rey et al. (1995) for central European encephalitis (CEE) virus. The structure indicates the fusion peptide loop defined by Leu107 and Phe108 and the adjacent Thr76 residue. (B) Model of the OHF virus fusion peptide based on the coordinates of the CEE crystal structure as generated by SwissModel. Note the loss of hydrogen binding potential with the presence of Ala76 in hemorrhagic TBE viruses rather that Thr76 as found in neurotropic TBE viruses.

in particular, Phe108. A similar conformational shift appears to be present in models of ALK and KFD viral E protein (not shown).

Further comparison of the OHF viral E protein sequence with residues that appear to specify specific lineages of Siberian subtype TBE viruses (Gritsun et al., 2003) found that OHF virus fell into neither of the proposed lineages based on comparison of a specific E protein amino acid pair at residues E-234 and E-431. The two proposed lineages had amino acid pairs His234-Ala431 and Gln234-Thr431. OHF virus was quite distinct with residues Gly234 and Ser431 at the specified positions. In addition, Asp83 was unique among all of the TBE viruses examined and those defined by Gritsun et al. (2003), who also suggested residue E-83 as a marker for differentiating TBE virus strains.

Nonstructural protein similarities

Considering the role of several nonstructural proteins in the viral life cycle and potential roles in viral pathogenesis, the amino acid sequence of OHF virus open reading frame was compared to that of other TBE complex viruses, including ALK virus, to identify residues that were associated with TBE serocomplex viruses that cause disease with hemorrhagic manifestations. Fifteen amino acids were identified in the nonstructural region of OHF and ALK viruses that were different from all of the other TBE complex viruses examined (Table 4). The percentage difference between OHF/ALK and the remaining viruses was relatively consistent throughout the nonstructural protein gene region of the open reading frame. As structural data for these viral proteins are lacking, correlations between these changes and their direct effect on viral propagation are difficult to determine.

Untranslated region similarities

The 5′ UTR of OHF virus is 132 nucleotides in length and appears to have secondary structure similar to that predicted for other tick-borne flaviviruses (Gritsun et al., 1997) (Fig. 4A). The OHF virus 5′ UTR has a small stem-loop within a highly conserved region (Fig. 4A, loop 3) in addition to a larger stem-loop structure at the 5′-terminus of the genome. Within this structure is a region, approximately nucleotides 58–87, in which the sequence of OHF virus is considerably different from 10 other TBE complex viruses examined (data not shown). This region is denoted by the opposing arrows in Fig. 4A. Despite the sequence variability, the predicted folding patterns in the 5′ UTR of OHF virus are nearly identical to those predicted by Gritsun et al. for the Vasilchenko and Powassan viruses (Gritsun et al., 1997). Similar to other TBE viruses, with the exception of POW virus, a single initiation codon signals the beginning of the ORF.

The 3′ UTR of the OHF virus is 413 nucleotides in length. The 3′ UTR consists of a 67-nucleotide variable region and a 346-nucleotide core region that is highly conserved, a finding that is similar to other tick-borne flaviviruses (Mandl et al., 1998; Wallner et al., 1995). The OHF virus 3′ UTR also lacks long poly(A)tracts within the 3′ UTR that define several TBE viruses of the Western subtype, including strains Neudorfl and 263 (Mandl and Kunz, 1991; Wallner et al., 1995). A number of conserved genetic elements have previously been described within the 3′ UTR of the tick-borne flaviviruses (Gritsun et al., 1997; Mandl et al., 1998; Mandl and Kunz, 1991; Pletnev et al., 1990; Rauscher et al., 1997; Wallner et al., 1995). The OHF virus 3′ UTR adheres to several of these elements including the intact 3′-terminal core element (nucleotides 342 to 413) (Fig 4B) (Rauscher et al., 1997). This region is nearly

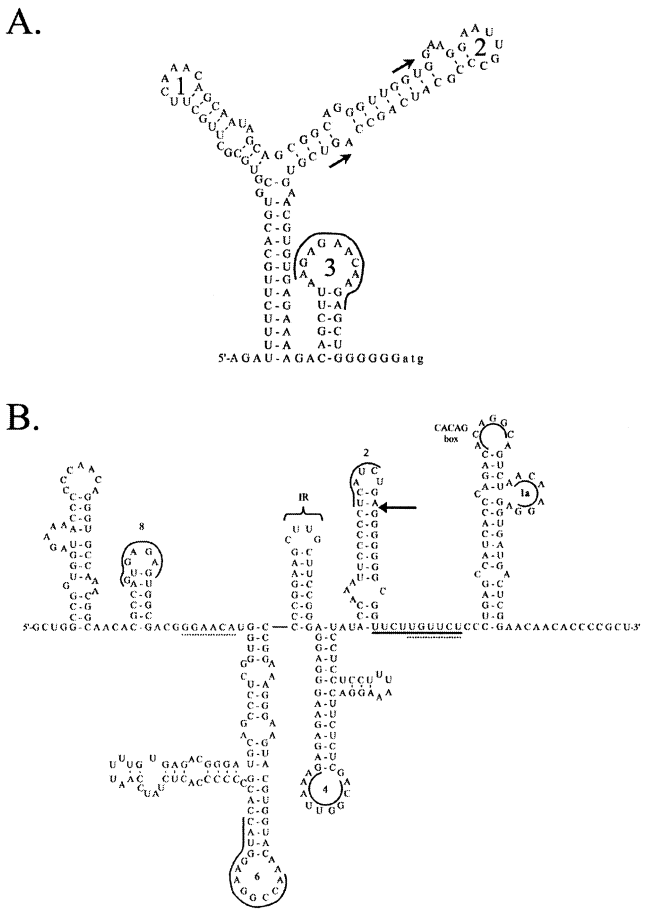

Fig. 4. Predicted folding of the untranslated regions of OHF virus. (A) The predicted 5′ UTR of OHF virus. Previously defined secondary structures common to TBE virus 5′ UTR regions are designated by numbers and an inverted complement region is denoted by the solid line (Gritsun et al., 1997). The arrows indicate the beginning and end of a region of the 5′ UTR where OHF virus is considerably different from other TBE complex viruses. The viral ORF initiation codon is designated by lowercase letters. (B) The predicted structure of the core region of the OHF virus 3′ UTR. Previously defined motifs common to TBE complex viruses are shown as numbered and underlined regions (Gritsun et al., 1997; Mandl et al., 1998; Mandl and Kunz, 1991; Pletnev et al., 1990; Rauscher et al., 1997; Wallner et al., 1995). A highly homologous inverted repeat (IR) and TBE complex specific sequence (solid bar) are also identified. The dotted bars indicate regions that are predicted to hybridize in energy minimization diagrams. The arrow indicates the beginning of the very highly conserved terminal stem-loop region of the genome.

completely conserved among the 11 TBE complex viruses examined, the notable exceptions being ALK virus, for which the 3′-terminal 71 nucleotides are missing, and POW virus. The core region of the 3′ UTR has been shown to be vital to viral infectivity (Mandl et al., 1998). Also identified in the core region of the OHF virus 3′ UTR were several conserved stem-loop structures as well as an inverted repeat region and a TBE-specific sequence element (Gritsun et al., 1997). Immediately 5′ of the core region is the variable region which has a high degree of heterogeneity in both nucleotide sequence and length. The variable region is apparently not vital to viral function (Mandl et al., 1998).

Discussion

OHF virus is a unique virus among the tick-borne flaviviruses. It is the only virus within this serocomplex that causes disease with little, if any, central nervous system involvement, and only one of three viruses to cause disease with hemorrhagic manifestations. OHF virus is found in the same geographic region as the highly pathogenic Russian spring–summer encephalitis (RSSE) virus that causes severe encephalitic disease. OHF virus is also unique among the European/Asian TBE serocomplex viruses as its principal vector is *D. reticulates*, whereas other tick-borne flaviviruses are transmitted primarily by *Ixodes spp.*, including *Ix. ricinus* (central European subtype) or *Ix. perculatus* (Siberian/Far-Eastern subtypes). The primary animal host for OHF virus is the muskrat (*O. zibethica*), a species that is not native to the OHF endemic area. Many other animals have been found to become infected by the virus, particularly the water vole (*A. terrestris*) (Andrewes et al., 1978; Kharitonova and Leonov, 1985), and may be involved in maintaining the virus within the ecosystem. The unique biological characteristics of the OHF virus suggest that there are specific molecular characteristics that define the viral phenotype, viral tropism, and disease presentation. To this end, the genetic sequence of OHF virus was determined and compared to other tick-borne flaviviruses in an effort to identify molecular markers that might play a role in the biology of this virus.

The current study elucidated and analyzed the first complete viral genome sequence of OHF virus. Phylogenetic analysis of the OHF viral genome grouped the virus among the TBE serocomplex of flaviviruses, yet set it in a separate clade from the other viruses, including those endemic to the same geographical region as OHF virus. These results support previous analysis of the viral E protein gene (Gritsun et al., 1993). Only two other tick-borne flaviviruses are known to cause disease with hemorrhagic manifestations, KFD and ALK viruses. Limited genetic analysis of KFD and ALK viruses demonstrates that they are very closely related and could possibly be strains of the same virus. KFD and ALK virus form a clade within the tick-borne flaviviruses that is separate from other viruses as well as OHF virus.

The ability of OHF, ALK, and KFD viruses to cause hemorrhagic disease suggests that molecular determinants within the viral genome may be responsible for this disease manifestation. We suggest that the E protein is a major determinant of tissue tropism for the flaviviruses and, therefore, a principal suspect to contain determinants of hemorrhagic disease. Subsequent analysis of the E protein from nine TBE serocomplex viruses identified three residues that may be related to cellular or tissue tropism. The most interesting of these is a change at residue E-76. In OHF, KFD, and ALK viruses, this residue is an alanine (Ala); in the encephalitic viruses examined, it is a threonine (Thr). The major biochemical difference between these amino acids is the presence of a hydroxyl group on the Thr residue.

Molecular modeling of OHF, KFD, and ALK E proteins demonstrate the loss of Van der Waals interactions between E-76 and E-107 (Leu) in the fusion peptide cd loop. This variation possibly affects the stability of the fusion peptide and may also affect the fusion properties of these viruses. It is also important to note that the E protein fusion peptide appears to interact with domain III in the second E protein of the E protein dimer (Rey et al., 1995). This interaction may function to stabilize the fusion peptide. The presence of an Ala residue at E-76 rather than a Thr eliminates hydrogen bonding potential and thus may destabilize the interaction between the E protein monomers and increase the chances of "accidental" triggering of the fusion event.

The 5′ UTR of OHF virus is considerably different from other TBE complex viruses through an approximately 30 nucleotide stretch, while the remainder of the 5′ UTR is highly homologous. As this result does not necessarily define tissue tropism, it may determine whether the virus can replicate in a cell or tissue once infection has occurred. The 5′ UTR is known to be involved in viral translation initiation as well as serving as the initiation site for (−)-strand synthesis (reverse complement genome); the high degree of variability in this region may alter viral translation and replication in mammalian hosts. The alternative possibility is that the variable sequence, though within a highly conserved stem-loop structure, may allow more efficient replication in *Demacentor* ticks rather than *Ixodes* by increasing the affinity of the viral genome for host replication factors. While these arguments are speculative, the degree of variability seen in the 5′ UTR of OHF virus suggests a functional role for the predicted stem-loop structure that is different in OHF virus than in other TBE complex viruses.

Unlike the 5′ UTR, the 3′ UTR of OHF virus is not particularly different from that of other TBE complex viruses. The 3′ UTR of OHF virus is slightly shorter than most other viruses examined and lacks a poly (A) tract, a characteristic that makes it more like the Far-Eastern and Siberian subtypes of TBE serocomplex viruses. This is not unexpected, as one would speculate that OHF evolved from the Siberian subtype viruses as they exist within similar geographic regions. The nearly complete homology of the 3′ stem-loop structure of OHF virus with other TBE complex viruses suggests that this portion of the viral genome does not play a significant role in cell- or vector-specific viral replication. As the 3′ UTR is thought to function as a promoter element for (−)-strand synthesis, the complete conservation of this element suggests that genome recognition elements in viral factors (e.g., NS5) that bind to this structure are highly conserved among the TBE serocomplex viruses.

Overall, OHF virus is unique among other TBE serocomplex viruses in that it typically uses a genus of tick for transmission that is different from other TBE serocomplex viruses. The virus also causes a severe hemorrhagic disease rather that one characterized by encephalitis. The objectives of this study were to sequence the genome of OHF virus in an effort to understand what molecular determinants might be involved in the vector and tissue tropism found with this virus. To this end, we have identified several residues within the viral E protein that may play a role in tissue tropism, one in particular at E-76, which could affect membrane fusion. We have also found that the 5′ UTR of OHF virus varies considerably from other TBE complex viruses and we suggest that this may define the ability of OHF virus to replicate in specific tissues or tick vectors. Efforts are currently under way to more closely examine the role of these molecular determinants upon viral infection and replication.

Materials and methods

Viruses used in this study are listed in Table 2. The OHF viral genome has been submitted to GenBank, under Accession No. AY193805.

Viral propagation and RNA extraction

Subconfluent monolayers of Vero E6 cells (ATCC) were infected with OHF virus strain Bogoluvovska at an m.o.i. of 0.01. After the appearance of viral cytopathic effects, RNA was isolated from infected cell lysates and purified virus particles using TRIZOL LS Reagent (Gibco/Invitrogen) following the instructions of the manufacturer.

RT-PCR

Viral RNA was reverse transcribed using Superscript II reverse transcriptase (Invitrogen) using specific antisense primers followed by Rnase H digestion. The resultant cDNA was amplified by PCR using Taq DNA polymerase (Roche). The viral genome was amplified as 16 overlapping PCR products excluding the 5′- and 3′-terminal regions. The 5′- and 3′-termini were amplified using the method described by Mandl et al., in which the 5′ cap is removed by tobacco acid pyrophosphatase and the termini are ligated with RNA ligase prior RT-PCR amplification across the ends.

Sequencing

PCR products were sequenced directly using specific primers or were cloned into the pGEM-T Easy cloning vector (Promega). Sequences generated from cloned PCR products were the result of sequencing three or more clones in both the forward and the reverse directions. Sequencing was performed on a PE Biosystem 373XL automated sequencing system in the UTMB protein chemistry core facility using ABI big dye termination protocols.

Sequence analysis

Sequence analysis and genome assembly were performed using the Vector NTI (InfoMax) sequence analysis

program. Phylogenetic analysis was continued using DNAStar, Paup, and applications within the Phylip package of genetic analysis software as follows: Initial sequence alignments were generated using ClustalW. Phylogenetic trees were constructed using maximum likelihood methods using both DNAmlk and FastDNAml programs within Phylip using default parameters. Resultant trees were identical using both programs. The robustness of the phylograms was evaluated by 500 bootstrap replicates. RNA structure predictions were made using RNAdraw (http://rnadraw.base8.se/) and the Mfold (http://www.bioinfo.rpi.edu/applications/mfold) secondary structure prediction application. Envelope protein structure modeling was performed using the crystal coordinates of central European TBE (SVB01.pdb) (Rey et al., 1995) as a template and using SwissModel (http://www.expasy.org/swissmod/) for prediction of the OHF E protein structure (Guex and Peitsch, 1997; Peitsch, 1995, 1996). Analysis and graphical development were performed using Swiss-Pdb Viewer and VMD pdb visualization software.

Acknowledgments

The authors thank the UTMB Protein Chemistry core facility for sequencing services. Funding was provided by the James W. McLaughlin Fellowship Fund (M.R.H.) and the Centers for Disease Control and Prevention (A.D.T.B.).

References

Allison, S.L., Schalich, J., Stiasny, K., Mandl, C.W., Heinz, F.X., 2001. Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. J. Virol. 75, 4268–4275.

Andrewes, C., Pereira, H.G., Wildy, P., 1978. Viruses of Vertebrates, fourth ed. Bailliere Tindall, London.

Burke, D.S., Monath, T.P., 2001. Flaviviruses, in: Knipe, D.M., Howley, P.M. (Eds.), Field's Virology, fourth ed. Lippincott Williams & Wilkins, Philadelphia, pp. 1043–1125.

Charrel, R.N., Zaki, A.M., Attoui, H., Fakeeh, M., Billoir, F., Youef, A.I., de Cheese, R., De Micco, P., Gould, E.A., de Lamballerie, X., 2001. Complete coding sequence of the Alkhurma virus, a tick-borne flavivirus causing severe hemorrhagic fever in humans in Saudi Arabia. Biochem. Biophys. Res. Commun. 287, 455–461.

Clarke, D.H., 1964. further studies on antigenic relationships among the viruses of the group B tick-borne complex. Bull. World Health Organ. 31, 45–56.

Gould, E.A., de Lamballerie, X., Zanotto, P.M., Holmes, E.C., 2001. Evolution, epidemiology, and dispersal of flaviviruses revealed by molecular phylogenies. Adv. Virus Res. 57, 71–103.

Gritsun, T.S., Frolova, T.V., Zhankov, A.I., Armesto, M., Turner, S.L., Frolova, M.P., Pogodina, V.V., Lashkevich, V.A., Gould, E.A., 2003. Characterization of a siberian virus isolated from a patient with progressive chronic tick-borne encephalitis. J. Virol. 77 (1), 25–36.

Gritsun, T.S., Lashkevich, V.A., Gould, E.A., 1993. Nucleotide and deduced amino acid sequence of the envelope glycoprotein of Omsk haemorrhagic fever virus; comparison with other flaviviruses. J. Gen. Virol. 74, 287–291.

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 Res. 49, 27–39.

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-Pdbviewer: an environment for comparative protein modelling. Electrophoresis 18, 2714–2723.

Hayasaka, D., Suzuki, Y., Kariwa, H., Ivanov, L., Volkov, V., Demenev, V., Mizutani, T., Gojobori, T., Takashima, I., 1999. Phylogenetic and virulence analysis of tick-borne encephalitis viruses from Japan and far-eastern Russia. J. Gen. Virol. 80, 3127–3135.

Heinz, F.X., Allison, S.L., 2000. Structures and mechanisms in flavivirus fusion. Adv. Virus Res. 55, 231–269.

Heinz, F.X., Allison, S.L., 2001. The machinery for flavivirus fusion with host cell membranes. Curr. Opin. Microbiol. 4, 450–455.

Kharitonova, N.N., Leonov, Y.A., 1985. Omsk Hemorrhagic Fever. Trans. S. Korzelska. Amerind Publishing Co., New Delhi.

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. J. Virol. 72 (3), 2132–2140.

Mandl, C.W., Kunz, C.H., F.X. 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(8), 4070–4077.

Pavri, K., 1989. Clinical, clinicopathologic, and hematologic features of Kyasanur forest disease. Rev. Infect. Dis. 11, S854–S859.

Peitsch, M.C., 1995. Protein modeling by e-mail. Biotechnology 13, 658–660.

Peitsch, M.C., 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. Biochem. Soc. Trans. 24 (1), 274–279.

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 (1), 250–263.

Rauscher, S., Flamm, C., Mandl, C.W., Heinz, F.X., Stadler, P.F., 1997. Secondary structure of the 3′-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities. RNA 3 (7), 779–791.

Rey, F.A., Heinz, F.X., Mandl, C.W., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis at 2 Å resolution. Nature 375, 291–298.

Theiler, M., Downs, W.G., 1973. The Arthropod-Borne Viruses of Vertebrates. Yale Univ Press, New Haven.

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 (1), 169–178.

Work, T., 1958. Russian spring-summer encephalitis virus in India. Kyasanur Forest disease. Prog. Med. Virol. 1, 248–279.

Zaki, A.M., 1997. Isolation of a flavivirus related to the tick-borne encephalitis complex from human cases in Saudi Arabia. Trans. R. Soc. Trop. Med. Hygiene 91 (2), 179–181.

Zanotto, P.M., Gao, G.F., Gritsun, T., Marin, M.S., Jiang, W.R., Venugopal, K., Reid, H.W., Gould, E.A., 1995. An arbovirus cline across the northern hemisphere. Virology 210 (1), 152–159.