Musashi binding elements in Zika and related Flavivirus 3'UTRs: A comparative study *in silico*

Adriano de Bernardi Schneider[1] **and Michael T. Wolfinger**[2,*]

[1]Department of Medicine, University of California San Diego, 220 Dickinson St, Suite A, San Diego, CA 92103, United States of America
[2]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria
[*]To whom correspondence should be addressed: michael.wolfinger@univie.ac.at

ABSTRACT

Zika virus (ZIKV) belongs to a class of neurotropic viruses that have the ability to cause congenital infection, which can result in microcephaly or fetal demise. Recently, the RNA-binding protein Musashi-1 (Msi1), which mediates the maintenance and self-renewal of stem cells and acts as a translational regulator, has been associated with promoting ZIKV replication, neurotropism, and pathology. Msi1 predominantly binds to single-stranded motifs in the 3' untranslated region (UTR) of RNA that contain a *UAG* trinucleotide in their core. We systematically analyzed the properties of Musashi binding elements (MBEs) in the 3'UTR of flaviviruses with a thermodynamic model for RNA folding. Our results indicate that MBEs in ZIKV 3'UTRs occur predominantly in unpaired, single-stranded structural context, thus corroborating experimental observations by a biophysical model of RNA structure formation. Statistical analysis and comparison with related viruses show that ZIKV MBEs are maximally accessible among mosquito-borne flaviviruses. Our study addresses the broader question of whether other emerging arboviruses can cause similar neurotropic effects through the same mechanism in the developing fetus by establishing a link between the biophysical properties of viral RNA and teratogenicity. Moreover, our thermodynamic model can explain recent experimental findings and predict the Msi1-related neurotropic potential of other viruses.

Introduction

Flaviviruses are an emerging group of arboviruses belonging to the *Flaviviridae* family. Researchers have been describing recent outbreaks of these viruses that have not been previously detected for decades[1–3].

The genus *Flavivirus* comprises more than 70 species that are mainly transmitted by mosquitoes and ticks, typically classified into four groups: Mosquito-borne flaviviruses (MBFVs), tick-borne flaviviruses (TBFVs), insect-specific flaviviruses (ISFVs), that do not have vertebrate hosts, and no known arthropod vector flaviviruses (NKVs), which typically infect bats and rodents. Flaviviruses represent a global health threat, including emerging and re-emerging human pathogens such as Dengue (DENV), Yellow fever (YFV), Japanese encephalitis (JEV), West Nile (WNV), Tick-borne encephalitis (TBEV) and Zika (ZIKV) viruses[4,5].

Initially isolated in 1947 from a sentinel rhesus macaque in the Ziika forest, Uganda, ZIKV has not been associated with severe disease, apart from skin rashes, body pain, and fever. Likewise, ZIKV has been circulating across equatorial zones in Africa and Asia for 60 years, until the first outbreak was reported in Yap Island, Micronesia in 2007. Subsequently, the virus spread eastwards to French Polynesia and other Pacific islands in 2013 and reached the Americas in 2015[6,7]. There are two main ZIKV lineages, the original African (type strain MR766) and an Asian (type strain FSS13025)[8,9], the latter also comprising American strains such as PE243.

Background

The 2015-2017 outbreak in the Americas raised the possibility of a link between ZIKV infection and congenital abnormalities, which included placental damage, intrauterine growth restrictions, eye diseases and microcephaly in children as well as acute motor axonal neuropathy-type Guillain-Barré syndrome in adults [10]. While MBFVs are typically transmitted by host-vector interaction, vertical transmission from mother to child during pregnancy via transplacental infection has been reported[11].

The neurotropic potential of ZIKV-related flaviviruses has been known since the 1970s, when Saint Louis encephalitis virus (SLEV) has been attributed to a severe neurological disorder in infected mice[12,13]. Vertical transmission has been observed with JEV in mice[14] and human[15] and a case of human fetal infection have been reported after YFV vaccination[16]. Other transmission pathways of ZIKV include blood transfusions and sexual transmission[17,18]. Despite enormous efforts in studying ZIKV infections in the last years, the biological reasoning and mechanisms behind arbovirus congenital neurotropism remain

elusive.

Flavivirus genome organization

Flaviviruses have the structure of an enveloped sphere of approximately 50 nm diameter. They are single-stranded positive-sense RNA viruses of 10-12 kb in size, and their genomic RNA (gRNA) encodes a single open reading frame (ORF) flanked by highly structured untranslated regions (UTRs). Upon translation of the ORF, a polyprotein is produced which is processed by viral and cellular enzymes, yielding structured (C, prM, E) and unstructured proteins (NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, NS5). Both flavivirus UTRs are crucially related to regulation of the viral life cycle, mediating processes such as genome circularization, viral replication and packaging[19–22].

Flaviviruses hijack the host mRNA degradation pathway

The central role of flavivirus 3'UTR in modulating cytopathicity and pathogenicity became apparent when an accumulation of both gRNA and viral long non-coding RNA (lncRNA) has been observed upon infection. These lncRNAs, also known as subgenomic flaviviral RNAs (sfRNAs)[23], are stable decay intermediates derived from exploiting the host's mRNA degradation machinery[24].

sfRNAs are produced by partial degradation of viral gRNA by Xrn1, a host 5'-3' exoribonuclease that is associated with the endogenous mRNA turnover machinery[25,26]. The enzyme stalls at highly conserved RNA structures in the viral 3'UTR, so-called Xrn1-resistant RNAs (xrRNAs), resulting in sfRNAs of variable lengths[27,28]. Xrn1-resistant RNAs and sfRNAs appear to be ubiquitously present in many flaviviruses. They have been described in MBFVs, including DENV[29], YFV[30], JEV[31], and ZIKV[32], TBFVs[23,33], and recently in ISFVs and NKVs[34,35]. There is typically more than one xrRNA, given the diverse molecular architecture of different flavivirus 3'UTRs. Pseudoknot interactions have been proposed in some, but not all flavivirus xrRNAs[32,36]. While they may form transiently under certain conditions[28], conclusive validation of their ubiquitous presence is missing. Hence, we will exclude them in this work. Earlier studies in our group have identified conserved RNA structural elements in viral 3'UTRs [37–41], some of which have later been attributed to xrRNA functionality[23]. Stem-loop (SL) as well as dumbbell (DB) structures are found in 3'UTRs of flaviviruses in single or double copies (Figure 1) and have been associated with quantitative protection of downstream viral RNA[42].

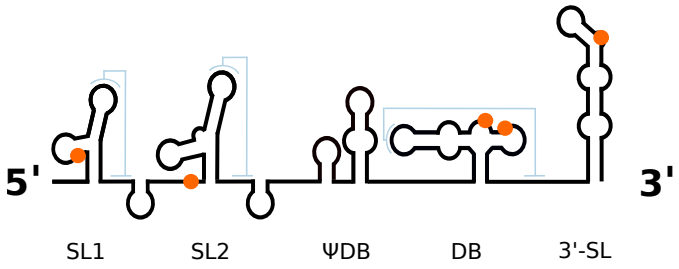

Figure 1. Schematic representation of the ZIKV 3'UTR. Conserved RNA elements include two stem-loop structures (SL1 and SL2), a ΨDB and canonical DB element as well as the terminal 3' stem-loop structure (3-SL). Positions of Musashi-binding *UAG* motifs in the Asian/American ZIKV lineage are highlighted in orange. Possible pseudoknot interaction sites (sketched in light blue) do not overlap with potential Musashi binding sites.

The inhibition of Xrn1 by viral RNA yields sfRNAs that affect many cellular processes, both in the vector and the host[43]. In mosquitoes, sfRNA interacts directly with the predominant innate immune response pathway, RNA interference (RNAi), by serving as a template for microRNA (miRNA) biogenesis[44]. Conversely, in host cells sfRNA modulates the anti-viral interferon response[45], e.g., by binding proteins to inhibit the translation of interferon-stimulated genes[46]. Moreover, sfRNA has been shown to inhibit Xrn1 and Dicer activity, thereby altering host mRNA levels[47,48].

At the same time, a variety of host proteins bind the 3'UTR of flaviviruses, thereby mediating viral replication, polyprotein translation or the anti-viral immune response (see Table 1 in ref.[43] for a comprehensive overview of host proteins that bind flavivirus 3'UTR/sfRNA). Although notoriously underrepresented in literature, one can expect that many of these proteins also bind sfRNA due to sequence and structure conservation.

Subgenomic flaviviral RNA interacts with Musashi

One of these groups of host factors is the Musashi (Msi) protein family. Msi is a highly conserved family of proteins in vertebrates and invertebrates that act as a translational regulator of target mRNAs and is involved in cell proliferation and differentiation. While the two Msi paralogs in mammals, Musashi-1 (Msi1) and Musashi-2 (Msi2), are expressed in stem

cells[49–51] and overexpressed in tumors and leukemias[52], they are absent in differentiated tissue. Moreover, Msi1 is involved in the regulation of blood-testis barrier proteins and spermatogenesis in mice[53]. Musashi proteins have two RNA recognition motif (RRM) domains, whose sequence specificity has been determined by an *in vitro* selection method and NMR spectroscopy[51,54,55]. The trinucleotide sequence *UAG*, whose thermodynamic binding specificity was determined by fluorescence polarization assays, has been identified as core Musashi binding element (MBE). Nucleotides enclosing the main MBE recognition motif make minor contributions to binding affinity[56]. While earlier SELEX experiments identified the binding aptamer sequence $(G/A)U_nAGU$ $(n = 1-3)$[51], iCLIP experiments with Msi1 in human glioblastoma cells confirmed the preferential binding of Msi1 to single-stranded (stem-loop) *UAG* sequences in 3'UTRs, but not in coding regions[57]. Zearfoss et al.[56] observed that both *GUAGU* and *AUAGU* are recognized by mouse Msi1, whereas Drosophila Msi1 has a higher affinity for *GUAGU*. NMR-derived structures of the two Msi1 RNA recognition motifs in complex with RNA also show that both RNA-binding domains bind *GUAGU* (PDB IDs 2RS2 and 5X3Z).

In summary, there is a strong consensus in the literature that *UAG* is central to all proposed Musashi binding motifs. Therefore, we focus our calculations around this trinucleotide, and provide evidence that the availability of *UAG* in pentanucleotides expands to the accessibility of the entire motif.

Musashi is involved in Flavivirus neurotropism

An interesting, yet understudied hypothesis is the possibility that the stem cell regulator protein Musashi could be related to ZIKV tropism. Based on the identification of a MBE in the 3'UTR of the ZIKV genome[10], de Bernardi Schneider et al.[7] reported the presence of the same element with a higher binding affinity for human Msi1 in all ZIKV sequences that belong to the Asia-Pacific-Americas clade in an *in silico* screen and implied that there could be a change of tropism for the viral lineage. Chavali et al.[58] tested the possibility of Msi1 interaction with the ZIKV genome *in vivo* and found that Msi1 not only interacts with ZIKV, but also enhances viral replication. They noted that ZIKV RNA could compete with endogenous targets for binding Msi1 in the brain of the developing fetus, thereby dysregulating the expression of genes required for neural stem cell development. Based on their data the authors concluded that Msi1 is involved in ZIKV neurotropism and pathology and raised the question whether MBEs present in other flavivirus genomes could exhibit similar functionality. In a recent study, Platt et at.[59] investigated whether ZIKV-related arboviruses can cause congenital infection and fetal pathology in utero in immunocompetent mice. They tested two emerging neurotropic flaviviruses, WNV, and Powassan virus (POWV), as well as two alphaviruses, Chikungunya virus (CHIKV) and Mayaro virus (MAYV). All four viruses caused placental infection, however, only WNV and POWV resulted in fetal demise, indicating that ZIKV is not unique among flaviviruses in its capacity to be transplacentally transmitted and cause fetal neuropathology.

In this contribution, we systematically analyze the Musashi-related neurotropic potential of well-curated flavivirus genomes *in silico*. We investigate structural features of MBEs in viral 3'UTRs by a thermodynamic model of RNA structure formation and work out the biophysical properties of conserved RNA structures harboring MBEs in order to build a theoretical ground for future *in vivo* studies.

Materials and Methods

Dataset

Sequence data for the present study was acquired from the public National Center for Biotechnology Information (NCBI) `refseq` database (https://www.ncbi.nlm.nih.gov/refseq/) on 15 December 2017. We filtered for all complete viral genomes under taxonomy ID 11051 (genus *Flavivirus*), resulting in 72 genomes, 51 of which had 3'UTR sequences and annotation available (Table 1).

The core Musashi binding element is only three nucleotides long, hence one can expect to observe a certain number of *UAG* trinucleotides by chance in any viral 3'UTR. Table 1 shows the number of MBEs present in 3'UTR regions of viral genomes analyzed here as well as the ratio $R_{UAG} = O_{UAG}/E_{UAG}$, i.e., observed versus expected frequencies. Assuming that all four nucleotides (A,U,G,C) occur independently and with equal probability, the expected probability to observe a subsequence of length l is equal to $(1/4)^l$. For $l = 3$, this is equal to $1/64$. More realistically, the frequency of each nucleotide $i \in \{A,U,G,C\}$ in an RNA sequence of length L is $F_i = N_i/L$, where, N_i is the nucleotide count of i. For any trinucleotide XYZ, the expected trinucleotide frequency E_{XYZ} is then computed from mononucleotide frequencies as $E_{XYZ} = F_X * F_Y * F_Z$.

The `refseq` genome for Spondweni virus (SPONV, accession number NC_029055.1) does not include a 3'UTR sequence. Since SPONV is phylogenetically closely related to ZIKV[60], we were looking to include this sequence into our analysis. Nikos Vasilakis (Univ. of Texas Medical Branch, Galveston, TX, USA) generously provided SPONV sequence data. The 338 nt 3'UTR sequence of the SA-Ar strain (see supplementary material) has been added to the set of flavivirus sequences analyzed here.

Kama virus (KAMV) does not contain *UAG* trinucleotides in the 3'UTRs, consequently it has been discarded from our dataset. The remaining virus species contain between 1 and 19 MBEs in their 3'UTRs.

Group	Accession number	Acronym	Scientific name	3'UTR length	MBE count	R_{UAG}
MBFV	NC_009026.2	AROAV	Aroa virus	421	7	1.18
MBFV	NC_012534.1	BAGV	Bagaza virus	566	13	1.40
MBFV	NC_017086.1	CHAOV	Chaoyang virus	326	6	1.25
MBFV	NC_001477.1	DENV1	Dengue virus 1	462	7	1.10
MBFV	NC_001474.2	DENV2	Dengue virus 2	451	10	1.47
MBFV	NC_001475.2	DENV3	Dengue virus 3	440	7	1.14
MBFV	NC_002640.1	DENV4	Dengue virus 4	384	5	0.94
MBFV	NC_016997.1	DONV	Donggang virus	343	10	1.75
MBFV	NC_009028.2	ILHV	Ilheus virus	388	9	1.87
MBFV	NC_001437.1	JEV	Japanese encephalitis virus	582	15	1.64
MBFV	NC_012533.1	KEDV	Kedougou virus	390	2	0.38
MBFV	NC_009029.2	KOKV	Kokobera virus	558	8	0.98
MBFV	NC_000943.1	MVEV	Murray Valley encephalitis virus	614	12	1.25
MBFV	NC_032088.1	NMV	New Mapoon virus	546	14	1.71
MBFV	NC_033715.1	NOUV	Nounané virus	347	7	1.29
MBFV	NC_018705.3	NTAV	Ntaya virus	565	15	1.67
MBFV	NC_008719.1	SEPV	Sepik virus	459	6	0.91
MBFV	NC_007580.2	SLEV	Saint Louis encephalitis virus	549	7	0.88
MBFV	NC_034151.1	THOV	T'Ho virus	556	13	1.45
MBFV	NC_015843.2	TMUV	Tembusu virus	618	16	1.56
MBFV	NC_006551.1	USUV	Usutu virus	665	19	1.62
MBFV	NC_012735.1	WESSV	Wesselsbron virus	478	6	0.90
MBFV	NC_009942.1	WNV1	West Nile virus lineage 1	631	17	1.70
MBFV	NC_001563.2	WNV2	West Nile virus lineage 2	573	16	1.83
MBFV	NC_002031.1	YFV17D	Yellow fever virus 17D	508	4	0.61
MBFV	NC_035889.1	ZIKV-BR	Zika Virus - Asian-American lineage	429	5	0.92
MBFV	NC_012532.1	ZIKV-UG	Zika Virus - African lineage	428	6	1.09
MBFV	NC_029055.1	SPONV	Spondweni virus	338^{a}	5^{a}	0.99
MBFV	NC_026623.1	APCV	Cacipacore virus	N/A	N/A	N/A
MBFV	NC_034018.1	YAOV	Yaounde virus	N/A	N/A	N/A
MBFV	NC_033693.1	BOUV	Bouboui virus	N/A	N/A	N/A
MBFV	NC_030289.1	EHV	Edge Hill virus	N/A	N/A	N/A
MBFV	NC_033699.1	JUGV	Jugra virus	N/A	N/A	N/A
MBFV	NC_033697.1	SABV	Saboya virus	N/A	N/A	N/A
MBFV	NC_033698.1	UGSV	Uganda S virus	N/A	N/A	N/A
TBFV	NC_004355.1	ALKV	Alkhumra hemorrhagic fever virus	320	2	0.46
TBFV	NC_006947.1	KSIV	Karshi virus	381	2	0.38
TBFV	NC_003690.1	LGTV	Langat virus	568	3	0.40
TBFV	NC_001809.1	LIV	Louping ill virus	500	3	0.49
TBFV	NC_005062.1	OHFV	Omsk hemorrhagic fever virus	410	3	0.60
TBFV	NC_003687.1	POWV	Powassan virus	480	5	0.76
TBFV	NC_027709.1	SGEV	Spanish goat encephalitis virus	493	5	0.83
TBFV	NC_001672.1	TBEV	Tick-borne encephalitis virus	764	6	0.55
TBFV	NC_023424.1	TYUV	Tyuleniy virus	273	1	0.22
TBFV	NC_023439.1	KAMV	Kama virus	282	0	-
TBFV	NC_033721.1	MEAV	Meaban virus	N/A	N/A	N/A
TBFV	NC_033726.1	SREV	Saumarez Reef virus	N/A	N/A	N/A
TBFV	NC_033724.1	KADV	Kadam virus	N/A	N/A	N/A
TBFV	NC_033723.1	GGV	Gadgets Gully virus	N/A	N/A	N/A
ISFV	NC_012932.1	AEFV	Aedes flavivirus	942	10	0.78
ISFV	NC_001564.2	CFAG	Cell fusing agent virus	553	9	1.16
ISFV	NC_008604.2	CxFV	Culex flavivirus	674	10	1.17
ISFV	NC_012671.1	QBV	Quang Binh virus	673	7	0.79
ISFV	NC_005064.1	KRV	Kamiti River virus	1208	13	0.82
ISFV	NC_027819.1	MECDV	Mercadeo virus	638	11	1.07
ISFV	NC_021069.1	MSFV	Mosquito flavivirus	674	9	1.01
ISFV	NC_034242.1	OCFVPT	Ochlerotatus caspius flavivirus	148	2	0.81
ISFV	NC_027817.1	PaRV	Parramatta River virus	629	12	1.21
ISFV	NC_030401.1	HANV	Hanko virus	N/A	N/A	N/A
ISFV	NC_033694.1	PCV	Palm Creek virus	N/A	N/A	N/A
NKV	NC_008718.1	ENTV	Entebbe bat virus	155	2	1.06
NKV	NC_027999.1	EPEV	Paraiso Escondido virus	316	2	0.37
NKV	NC_004119.1	MMLV	Montana myotis leukoencephalitis virus	460	9	1.18
NKV	NC_003635.1	MODV	Modoc virus	366	9	1.42
NKV	NC_005039.1	YOKV	Yokose virus	429	9	1.21
NKV	NC_026624.1	SOKV	Sokoluk virus	N/A	N/A	N/A
NKV	NC_003676.1	APOIV	Apoi virus	N/A	N/A	N/A
NKV	NC_026620.1	JUTV	Jutiapa virus	N/A	N/A	N/A
NKV	NC_029054.2	POTV	Potiskum virus	N/A	N/A	N/A
NKV	NC_034007.1	PPBV	Phnom Penh bat virus	N/A	N/A	N/A
NKV	NC_003675.1	RBV	Rio Bravo virus	N/A	N/A	N/A
NKV	NC_003996.1	TABV	Tamana bat virus	N/A	N/A	N/A

Table 1. Viral genomes analyzed in this study. Flaviviruses are categorized into the groups mosquito-borne flaviviruses (MBFV), tick-borne flaviviruses (TBFV), insect-specific flaviviruses (ISFV) and no known vector flaviviruses (NKV). 3'UTR lengths, number of Musashi binding elements found in 3'UTRs and relative MBE abundance R_{UAG} are listed. R_{UAG} values above 1 indicate relative enrichment of *UAG* trinucleotides, whereas values below 1 indicate relative depletion. ^{a}3'UTR length and MBE count of SPONV SA-Ar strain. $^{N/A}$3'UTR partial or not available in the `refseq` data set.

Opening energy directly relates to single-strandedness

The biophysical model employed here is based on a description of RNA at the level of secondary structures, building upon the thermodynamic *nearest neighbor* energy model as implemented in the `ViennaRNA` Package[61]. This allows for computing equilibrium properties of RNA such as the single most stable, minimum free energy (MFE) structure, as well as the partition function $\mathscr{Z}$. The latter makes an evaluation of the thermodynamic ensemble of RNA structures available and is defined as the sum over all Boltzmann factors of individual structures s

$$\mathscr{Z} = \sum_s e^{-E(s)/RT} \tag{1}$$

where $E(s)$ is the free energy of the structure, R the universal gas constant and T the thermodynamic temperature of the system. The equilibrium probability of a secondary structure s is then defined as

$$p(s) = \frac{e^{-E(s)/RT}}{\mathscr{Z}}\,. \tag{2}$$

The partition function $\mathscr{Z}$ can be computed efficiently via dynamic programming[62] and allows calculation of individual base pair probabilities, even for large sequences[63]. In this line, the *accessibility* (i.e., the probability that a region $i \ldots j$ along the RNA is single-stranded) can be derived from the partition function (Eq. 1)[64]. Likewise, the *opening energy* (i.e., the free energy required to force the region to be single-stranded) can be computed as

$$\Delta G_{\text{open}} = -RT \ln P(\text{unpaired})\,. \tag{3}$$

The opening energy of a region within an RNA is directly related to local RNA secondary structure. In this line, low opening energy is a reliable indicator for single-strandedness. We employ the sliding window approach of `RNAplfold`[63] to compute local pairing probabilities of *UAG* trinucleotide motifs to assess the likelihood of single-strandedness of and around MBEs. `RNAplfold` is part of the `ViennaRNA` Package[61] and can compute the accessibilities or single-strandedness of all intervals of an RNA in cubic time[65]. We select 97 nt windows upstream and downstream of MBEs in viral 3'UTRs and compute local pairing probabilities for base pairs within 100 nt windows. Opening energies for trinucleotides are then evaluated from averaged pairing probabilities with `RNAplfold`.

The significance of a calculated MBE opening energy is assessed by comparison with a large number of randomized sequences of the same length and same base or dinucleotide composition. We compute the opening energies of trinucleotides both in a genomic as well as a shuffled sequence context and apply a z score statistics. The normalized z score is defined as

$$z = \frac{E_{\text{open}}(XYZ) - \mu}{\sigma} \tag{4}$$

where $E_{\text{open}}(XYZ)$ is the opening energy of trinucleotide XYZ in its genomic context, μ and σ are the mean and standard deviations, respectively, of the opening energies of XYZ computed over a large sample of randomized sequences. Randomization with regard to keeping sequence composition is achieved here by applying dinucleotide shuffling to the 97 nt windows upstream and downstream of MBEs, while keeping XYZ in place. The same idea applies to calculations of pentanucleotide motifs.

The approach outlined above is implemented in the Perl utility `plfoldz.pl`, which is available from https://github.com/mtw/plfoldz. The script employs the `ViennaRNA`[61] scripting language interface for thermodynamics calculations, the `ViennaNGS`[66] suite for extraction of genomic loci and the uShuffle Perl bindings[67] for k-let shuffling. The tool reports for each requested trinucleotide the opening energy in a genomic context as well as an opening energy z score obtained from n shuffling events of upstream and downstream sequences. Here, $n = 10,000$ dinucleotide shuffling events were used.

Characteriztaion of MBEs within xrRNAs

To localize MBEs within homologous substructures in flavivirus 3'UTRs we constructed `infernal`[68] covariance models for conserved xrRNA elements. The structural RNA alignments underlying the `infernal` models were computed with `locarna`[69] and further analyzed with `RNAalifold`[61] and `RNAaliSplit`[70].

Results

MBEs are highly accessible in ZIKV 3'UTRs

The Musashi family of proteins preferentially bind single-stranded *UAG* motifs in 3'UTRs[57]. To evaluate the thermodynamics of Msi-*UAG* affinity more broadly, we set out to analyze the single-strandedness of all possible trinucleotides in ZIKV genomes.

To this end, we computed the opening energies of all trinucleotide motifs present in the coding sequence (CDS) and 3'UTR of the African (ZIKV-UG) and Asian/American (ZIKV-BR) Zika strains. A z score was calculated for each occurrence of trinucleotide XYZ according to Eq. 4, thereby normalizing the opening energy of XYZ in its genomic context with $n = 10,000$ dinucleotide-shuffled upstream and downstream regions of 97 nt, using 100 nt windows in `RNAplfold`.

Negative opening energy z scores indicate increased accessibility, i.e., *UAG* trinucleotides in viruses with overall low z scores are likely to occur in an unpaired structural context within the 3'UTR. Through the distribution of z scores, sorted by median z score (Figure 2) we were able to see three aspects standing out. First, the distribution of z scores is markedly divergent among CDS and 3'UTR. The interquartile ranges of opening energy z scores are homogeneous within the CDS region, while dispersion is varied within the 3'UTR. We hypothesize that this is caused by a different sequence composition that manifests in highly variable opening energies. It could, however, also be an artifact of the different sample sizes based on the divergent trinucleotide count in CDS and 3'UTR, respectively. Second, *UAG* is the most accessible trinucleotide in the 3'UTR of ZIKV-BR and among the highest accessible trinucleotides in the 3'UTR of ZIKV-UG. This is striking as it corroborates previous experimental evidence of Musashi affinity to ZIKV[58] by means of a thermodynamic model, thus underlining a possible role of Msi1 in ZIKV neurotropism. Moreover, the *UAG* trinucleotide is neither enriched nor depleted in the 3'UTRs of ZIKV-BR and ZIKV-UG (Table 1). Third, the canonical start codon *AUG* appears to the far right end of the scale in both ZIKB-BR and ZIKV-UG 3'UTRs, i.e., it is among the least accessible trinucleotides. This suggests evolutionary pressure on keeping the start codon in a paired structural context within the 3'UTR, thereby prohibiting accessibility to ribosomes and disabling undesirable leaky translation start from these *AUG* triplets.

We also tested the accessibility of larger Musashi recognition motifs. To this end, we employed the same approach outlined above for all pentanucleotides found in the 3'UTRs of ZIKV-BR and ZIKV-UG, respectively. The distributions of opening energy z scores (supplementary data, Figure S1) are in good agreement with our results derived for trinucleotides as well as previous experimental data, suggesting that core a *UAG* is among the most accessible motifs. In particular, our data shows that *NUAGN* is the most accessible pentanucleotide in the 3'UTR of ZIKV-BR and among the highest accessible pentanucleotides in the 3'UTR of ZIKV-UG, similar to the situation found for trinucleotides. *UAG* appears to be conserved in an unpaired structural context not only by itself but also in a larger sequence context of enclosing nucleotides, which exhibit high accessibility upon the presence of a central *UAG* Musashi recognition element. This finding is in line with the reported Musashi recognition pentamers *GUAGU* and *AUAGU*.

MBE accessibility in related viruses

To assess the Musashi-related neurotropic potential of other flaviviruses, we evaluated the accessibilities of MBEs in related species. To this end, all 435 *UAG* trinucleotide motifs within 3'UTRs in the `refseq` dataset were identified, grouped by vector specificity and subjected to the computational approach outlined above (97 nt upstream/downstream windows, $n = 10,000$ dinucleotide shufflings).

Msi1 preferentially binds single stranded RNA[57], consequently *UAG* motifs that contribute with low z scores have a high affinity for Msi1 binding. Within the MBFV group, the Asian/American lineage Zika virus (ZIKV-BR) has the lowest median z score, followed by Saint Louis encephalitis virus (SLEV), Nounané virus (NOUV) and the African lineage Zika virus (ZIKV-UG). Among others, two lineages of West Nile virus (WNV1, WNV2) and Yellow fever virus (YFV) appear with a negative median z score. ZIKV-BR turns out to be the only isolate among MBFVs that has just negative z score values in our simulations, i.e., all *UAG* motifs within the 3'UTR of the Brazilian ZIKV isolate appear in an unpaired structural context. Likewise, Karshi virus (KSIV), Alkhumra hemorrhagic fever virus (ALKV) and Langat virus (LGTV) have a strictly negative z scores distribution among the TBFV group. POWV, Omsk hemorrhagic fever (OHFV) and Louping ill virus (LIV) have negative mean opening energy z scores. Interestingly, *UAG* trinucleotides are relatively depleted in all TBFV species analyzed here (Table 1). Among NKVs, Montana myotis leukoencephalitis virus (MMLV) and Entebbe bat virus (ENTV) show negative mean opening energy z scores. Culex flavivirus (CxFV), Cell fusing agent virus (CFAG), Parramatta River virus (PaRV) and Ochlerotatus caspius flavivirus (OCFVPT) tend to have singe-stranded MBEs among the ISFVs. Here, OCFVPT is the only isolate with a strictly negative z score distribution (Figure 3).

The number of *UAG* trinucleotide motifs in 3'UTRs of the `refseq` dataset lies between 1 and 19 (Table 1). The overall range of opening energy z scores is not equal among different flavivirus groups. While the lower bound is between -1.65 and -1.93 among all groups, MBFVs and ISFVs show markedly higher upper bounds than TBFVs and NKVs, respectively. Absolute values of computed MBE opening energy z scores are listed in Table 2.

Conserved xrRNAs contain MBEs

Several species appear to the right of the plots in Figure 3 due to the sorting by median z score. However, they comprise a non-negligible number of accessible MBEs, as indicated by negative opening energy z scores. Examples are (re-) emerging species like JEV and Usutu virus (USUV), which contain 15 and 19 MBEs, respectively.

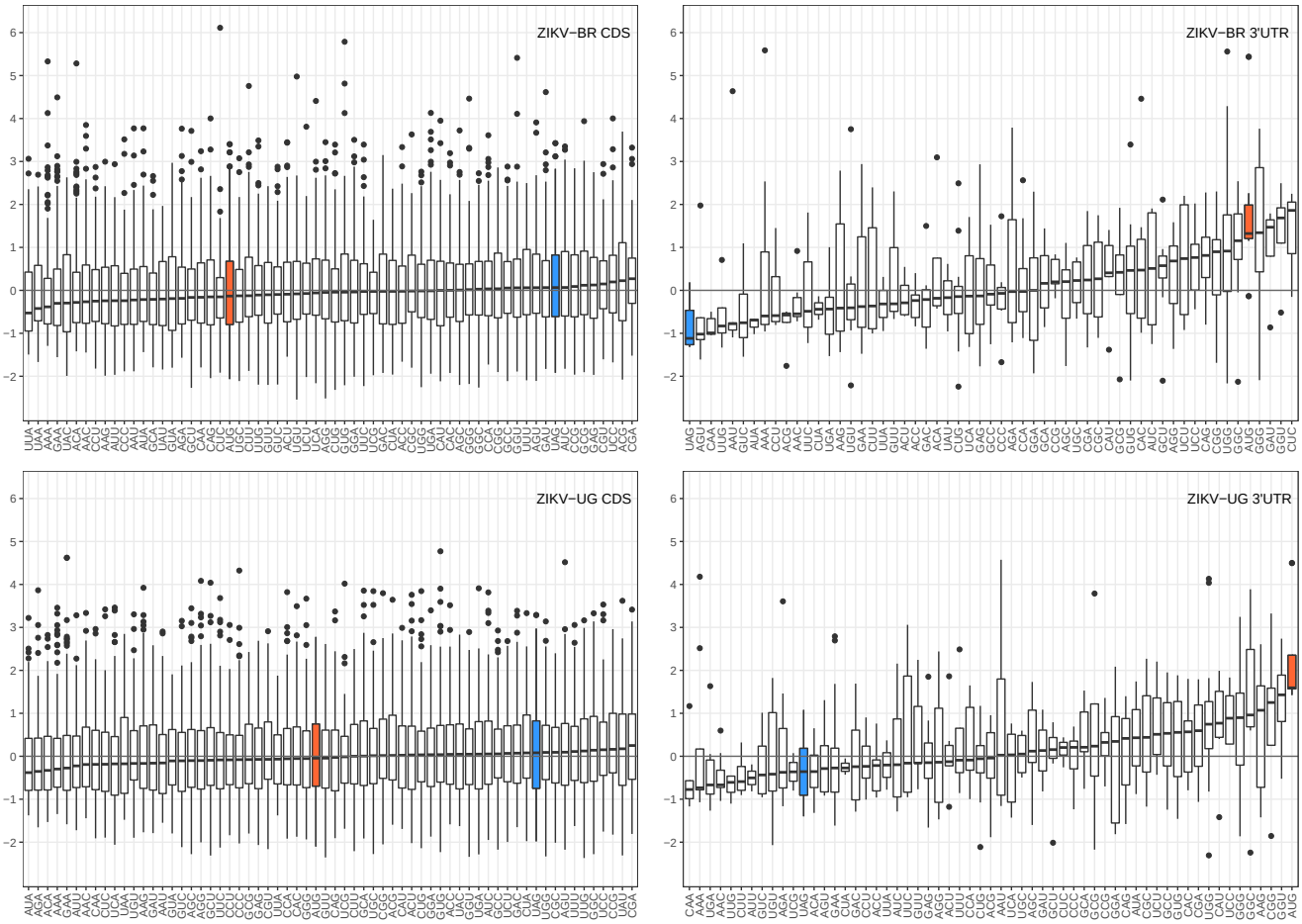

Figure 2. Distribution of z scores of opening energies for trinucleotides found in the coding region (CDS, left) and 3'UTR (right) of ZIKV from Brazil (top) and Uganda (bottom), sorted by median opening energy z score. The MBE motif is highlighted in blue and shows low overall z scores in the 3'UTR, indicating that this trinucleotide is more likely to appear in a single-stranded structural context. Contrary, the canonical start codon AUG (highlighted in orange) shows high opening energy z scores, indicating reduced accessibility within the 3'UTR region. Data for trinucleotides AUU, UAA, UCG and UUU are omitted because they only occur once within the 3'UTR of ZIKV-BR. Likewise, trinucleotides CGU, GUA, UAA and UAU are omitted in the ZIKV-UG 3'UTR plot.

	total		SL		DB	
	min	max	min	max	min	max
MBFV	−1.82	6.86	−1.35	2.10	−1.42	2.23
ISFV	−1.93	5.14	-	-	-	-
TBFV	−1.65	1.97	-	-	-	-
NKV	−1.79	2.02	-	-	−1.79	2.02

Table 2. Distribution ranges of opening energy z scores for MBEs in the 3'UTR of flaviviruses. The minimal and maximal z score is listed for all UAG motifs (total) within a the 3'UTR, and only for those that overlap one of the conserved xrRNA elements SL and DB. Dashes indicate that SL/DB elements are not conserved.

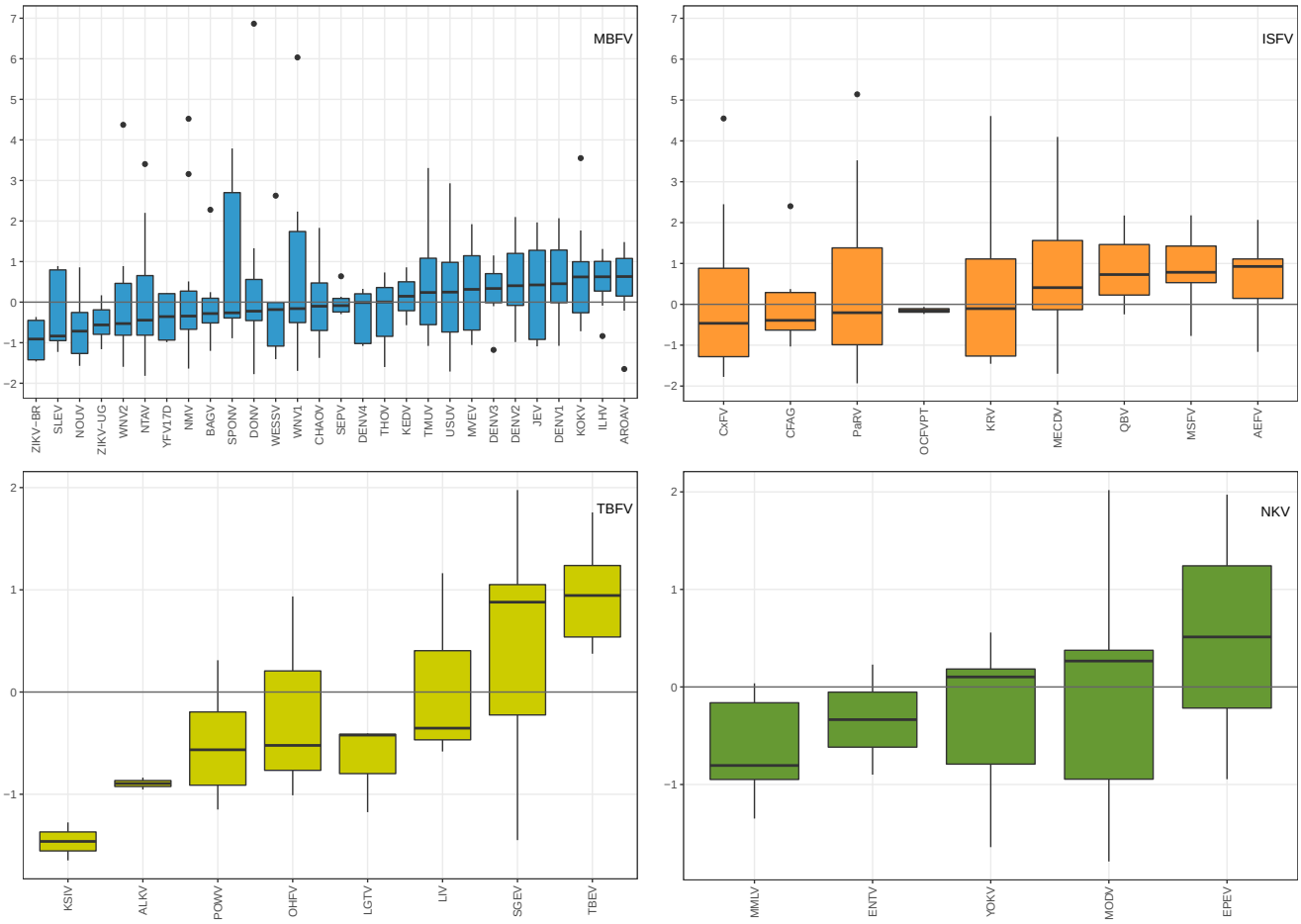

Figure 3. Distribution of MBE opening energy z scores in flavivirus 3'UTRs, grouped by vector specificity and sorted by median z scores. Top left: MBFVs, top right: ISFVs, bottom left: TBFVs and bottom right: NKVs. The Asian/American lineage ZIKV-BR isolate has the lowest median z score among all MBFV. Alkhurma virus (ALKV), Ochlerotatus caspius flavivirus (OcFV), ENTV and EPEV contain only two MBEs on the 3'UTR. Tyuleniy virus (TYUV) was excluded as it only contains a single MBE.

To investigate this further, we assigned each MBE in our dataset to one of the conserved elements stem-loop (SL), dumbbell (DB) and 3' stem-loop (3SL) (Figure 1) by means of covariance models. Analysis of RNA sequence and structure conservation revealed that the majority of virus isolates in our dataset contain only a single *UAG* motif within their SL and 3SL elements. Conversely, DB elements, which are conserved in MBFVs and NKVs (Figure 4), stand out among conserved RNA structures in flavivirus 3'UTRs. They contain a pair of MBEs, separated by a 4 nt spacer, within a perfectly conserved sequence motif of approx. 20 nt length in their distal stem-loop structure. We hypothesize that this pair of conserved *UAG* motifs interacts with the two RNA-binding domains in Musashi proteins. Figure 5 shows the consensus secondary structure of flavivirus DB elements.

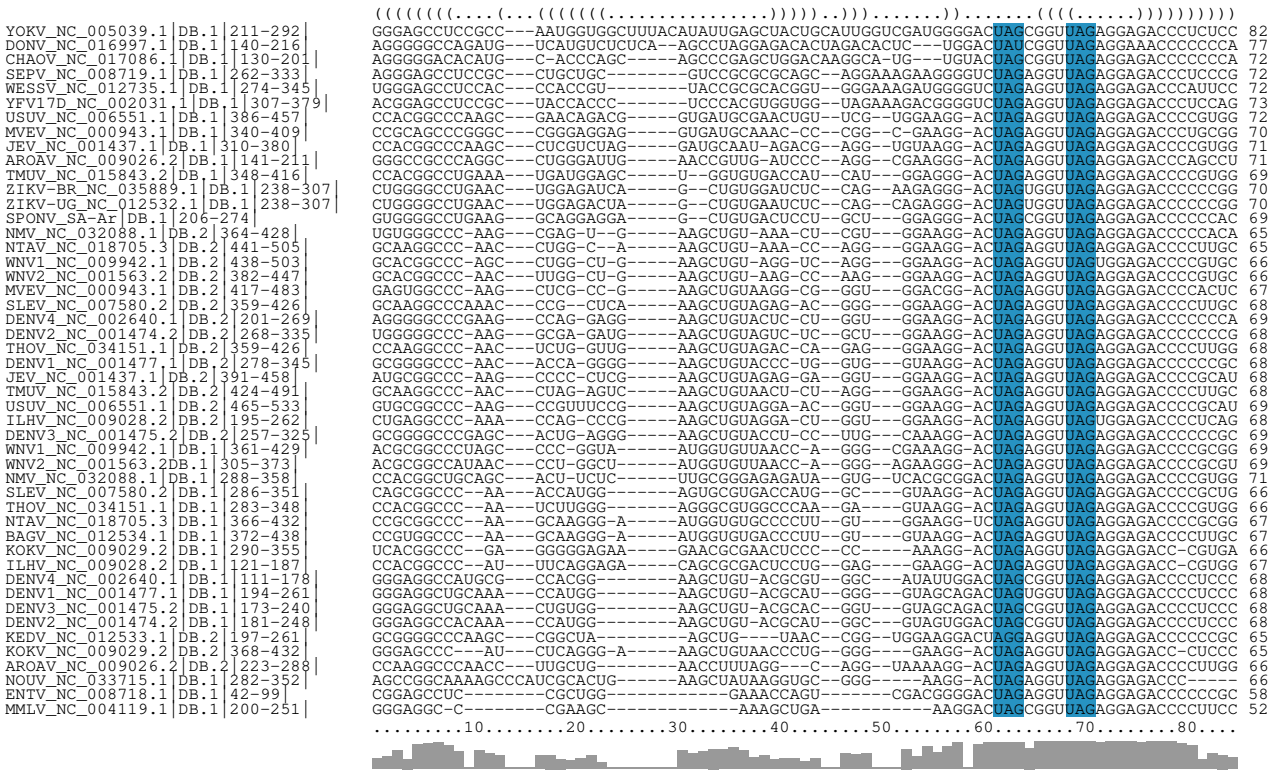

Figure 4. Structural alignment of conserved dumbbell (DB) elements in the 3'UTR of MBFV and NKV flaviviruses. Several species have two copies of DB elements in their 3'UTR, indicated by `DB.1` and `DB.2` in the sequence identifier. Coordinates are given relative to the 3'UTR start. A consensus structure in dot-bracket notation is plotted on top of the alignment. Gray bars at the bottom indicate almost perfect sequence conservation within the distal stem-loop sub-structure (positions 60-80). Two conserved MBE motifs in the central multiloop and distal stem loop are highlighted in blue.

Discussion

Our findings lead to the conclusion that the accessibility of *UAG* motifs calculated through opening energies in Flavivirus 3'UTRs is indicative of the Musashi-related neurotropic potential of virus species. Our computational analyses show that there is little difference in the distribution of opening energies for all trinucleotides within the polyprotein (CDS) region of ZIKV. When comparing CDS and 3'UTR regions, we see a difference in behavior, as different trinucleotides do possess different opening energies, *UAG* being highly accessible in ZIKV. Although not possible to quantify the impact of the accessibility on the patient phenotype, it is interesting to see the *UAG* motifs on the Brazilian ZIKV isolate more accessible than on the Ugandan ZIKV isolate. This result raises the question once again if the increased pathogenicity seen in ZIKV today is due to changes in the sequence over time or simply lack of better surveillance[72].

Previous experiments lead toward the idea that ZIKV is unique among flaviviruses regarding the clinical outcomes resulting from congenital infection[73,74]. Although our results indicate that this may be true for well-studied viruses such as DENV and WNV, other viruses which have not caused recent outbreaks may have been neglected.

Looking in depth at other viruses, Nounané virus (NOUV), a dual-host affiliated insect-specific flavivirus is found among the viruses with high MBE accessibility. NOUV was isolated in Cote d'Ivoire in 2004 from *Uranotaenia mashonaensis*, a

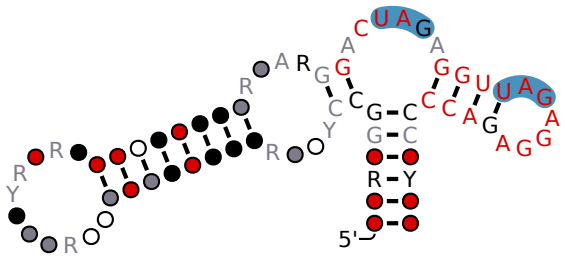

Figure 5. Consensus secondary structure of the flavivirus DB element with MBEs highlighted in blue. Figure generated from the MSA in Figure 4 with R2R[71]. Structure and color annotation inferred by R2R. Nucleotide symbols represent conserved nucleotides. Circles represent columns in the MSA that are typically or always present but do not conserve nucleotide identity. Red, black and gray colors indicate the level of nucleotide conservation, in decreasing order.

Culicidae mosquito not known to harbor flaviviruses before[75]. While replication has been tested on human and non-human cell lines, vertebrate infection and pathogenesis could not be observed[76].

Within the TBFV serocomplex, KSIV has the lowest overall MBE opening energies. Originally isolated from *Ornithodoros papillipes* ticks in Uzbekistan in 1972[77] it currently does not present history of infection in humans. Conversely, Powassan virus (POWV), another TBFV with negative MBE opening energy, first isolated in Powassan, Ontario, Canada in 1959 from a child who died of acute encephalitis[78] can cause transplacental infection[59] and has been associated with severe neuropathology and death in mice and human[79].

Given that *UAG* can be regarded as the primary Msi1 binding motif, we can argue that ZIKV has the highest affinity for binding Msi1 among all MBFVs. Platt et al.[59] showed that besides ZIKV, the neurotropic flaviviruses WNV and POWV, as well as the alphaviruses Chikungunya (CHIKV) and Mayaro (MAYV) infect placenta and fetus in immunocompetent, wild-type mice. However, only WNV was shown to infect the placenta and the fetal central nervous system, causing injury to the developing brain.

Congenital infection in humans is documented for WNV, JEV, YFV, and ZIKV. CHIKV and MAYV did not show this behavior. In this line, our results are in agreement with experimental studies that reported teratogenicity for SLEV [12,13], WNV[59,80], YFV[16,81] and POWV[59].

Bizarre neurological manifestations were also observed in patients infected by Ntaya virus (NTAV)[82], a neurotropic virus from the Japanese encephalitis serocomplex, as well as WNV in humans [83,84] and in mice[85], USUV [86,87] and DENV [88,89]. The fact that these viruses line up more on the positive side of the opening energy plots in Fig. 3 does not mean that they should not be neurotropic. It merely highlights that there might be additional mechanisms causing neuropathogenicity.

MBEs are conserved in flavivirus 3'UTR elements

Flavivirus DB elements do not only show structural conservation over the MBFV and NKV serocomplexes, but even maintain their primary sequence within a region of approx. 20 nt of the distal stem-loop (Figs. 4 and 5). The combination of covariation and primary sequence conservation within a single RNA element underlines the importance of DB elements in flavivirus pathogenicity. It could also be indicative of a special role of DB element regions in the minus-strand synthesis during flavivirus replication.

UAG trinucleotides are the core nucleotides within MBE motifs to contribute the highest binding energy[56]. Our current analysis underlines that there seems to be evolutionary pressure on keeping *UAG* motifs within the DB elements unpaired. In ZIKV we see that not only the *UAG*s within DB elements but also those that overlap with SL elements show negative opening energy z-score.

Msi1 presence in different cells

The presence of Msi1 proteins in both sperm and neural precursor cells highlights the importance of studying the Msi1-MBE interaction in flaviviruses. Given that Msi1 has been shown to enhance ZIKV replication[58], this interaction could be a critical reason why ZIKV persists in sperm for a long time after the individual has been infected[90,91], allowing the virus to be transmitted sexually and also why the virus would harbor itself in neuronal cells, allowing it to interfere with and dysregulate neurodevelopment.

A possible role of Musashi in the flavivirus life cycle

Msi1, which binds to the 3'UTR of target mRNAs, has been shown to repress translation initiation by competing with the translation initiation factor eIF4G for binding to poly(A)-binding protein (PABP), thereby inhibiting the assembly of the 80S

ribosomal unit[92]. Ribosome profiling experiments have corroborated this down-regulatory effect of Msi1, while keeping mRNA levels[93]. This allows for a speculative explanation of the findings by Chavali et al.[58], i.e., that Msi1 enhances ZIKV replication, and a possible role of Msi1 in the viral life cycle: Flaviviruses need to "donate" a few copies of the quasispecies ensemble for Xrn1 degradation and subsequent sfRNA production. In this line, Msi1 could serve as an agent that provides a reasonable amount of gRNAs that are not translated but subject to Xrn1 degradation. The resulting sfRNAs can then down-regulate the host response[46,94].

Conclusion

We studied a specific aspect of flavivirus congenital pathogenicity, i.e., the neurotropic effect inferred by the presence of MBEs in the 3'UTR of flavivirus genomes. Employing an established biophysical model of RNA structure formation, we analyzed the thermodynamic properties of MBEs *in silico*. Our results underline experimental studies suggesting that ZIKV is not alone in its capacity to cause severe neuropathology to infants through the MBE mechanism. While several tick-borne and mosquito-borne flavivirus species like Karshi virus (KSIV), Alkhumra hemorrhagic fever virus (ALKV) or Nounané virus (NOUV) line up with ZIKV in our theoretical model, their tropism might have been overseen due to the lack of reported significant outbreaks. However, some of them appear to have similar neurotropic potential and thus might be potent emerging pathogens.

The approach presented here could in principle be used for developing a tool to predict the Musashi-related neurotropic potential of novel viruses or (re-)emerging strains of known viruses. Combination of opening energy z scores with large scale epidemiologic data could be employed in a machine learning framework that also considers structural conservation and homology of flavivirus 3'UTR elements. Such a tool could play a role in categorizing viruses.

References

1. Tognarelli, J. *et al.* A report on the outbreak of Zika virus on Easter Island, South Pacific, 2014. *Arch Virol* **161**, 665–668 (2016).
2. Malone, R. W. *et al.* Zika virus: medical countermeasure development challenges. *PLoS Negl. Trop D* **10**, e0004530 (2016).
3. Hotez, P. J. & Murray, K. O. Dengue, West Nile virus, Chikungunya, Zika and now Mayaro? *PLoS Negl. Trop D* **11**, e0005462 (2017).
4. Gubler, D., Kuno, G. & Markoff, L. Flaviviruses. *Fields Virol.* **1**, 1153–1252 (2007).
5. Weaver, S. C. *et al.* Zika virus: History, emergence, biology, and prospects for control. *Antivir Res* **130**, 69–80 (2016).
6. Song, B.-H., Yun, S.-I., Woolley, M. & Lee, Y.-M. Zika virus: history, epidemiology, transmission, and clinical presentation. *J Neuroimmunol* **308**, 50–64 (2017).
7. de Bernardi Schneider, A. *et al.* Molecular evolution of Zika virus as it crossed the Pacific to the Americas. *Cladistics* **33**, 1–20 (2017).
8. Gong, Z., Xu, X. & Han, G.-Z. The diversification of Zika virus: Are there two distinct lineages? *Genome Biol Evol* **9**, 2940–2945 (2017).
9. Simonin, Y., van Riel, D., Van de Perre, P., Rockx, B. & Salinas, S. Differential virulence between Asian and African lineages of Zika virus. *PLoS Negl Trop D* **11**, e0005821 (2017).
10. Klase, Z. A. *et al.* Zika fetal neuropathogenesis: etiology of a viral syndrome. *PLoS Negl. Trop D* **10**, e0004877 (2016).
11. Platt, D. J. & Miner, J. J. Consequences of congenital Zika virus infection. *Curr Opin Virol* **27**, 1–7 (2017).
12. Andersen, A. & Hanson, R. Experimental transplacental transmission of St. Louis encephalitis virus in mice. *Infect Immun* **2**, 320–325 (1970).
13. Andersen, A. & Hanson, R. Intrauterine infection of mice with St. Louis encephalitis virus: immunological, physiological, neurological, and behavioral effects on progeny. *Infect Immun* **12**, 1173–1183 (1975).
14. Fujisaki, Y., Miura, Y., Sugimori, T., Murakami, Y. & Miura, K. Experimental studies on vertical infection of mice with Japanese encephalitis virus. IV. Effect of virus strain on placental and fetal infection. *NatL I Anim Heal. Q* **23**, 21–26 (1983).
15. Chaturvedi, U. *et al.* Transplacental infection with Japanese encephalitis virus. *J Infect Dis* **141**, 712–715 (1980).
16. Tsai, T., Paul, R., Lynberg, M. & Letson, G. Congenital yellow fever virus infection after immunization in pregnancy. *J Inf Dis* **168**, 1520–1523 (1993).

17. Foy, B. D. *et al.* Probable non–vector-borne transmission of Zika virus, Colorado, USA. *Emerg Infect Dis* **17**, 880 (2011).

18. Musso, D. *et al.* Potential sexual transmission of Zika virus. *Emerg Infect Dis* **21**, 359 (2015).

19. Hahn, C. S. *et al.* Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *J Mol Biol* **198**, 33–41 (1987).

20. Mukhopadhyay, S., Kuhn, R. J. & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol* **3**, 13 (2005).

21. Villordo, S. M., Alvarez, D. E. & Gamarnik, A. V. A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA* **16**, 2325–2335 (2010).

22. de Borba, L. *et al.* Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J Virol* **89**, 3430–3437 (2015).

23. Pijlman, G. P. *et al.* A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* **4**, 579–591 (2008).

24. Akiyama, B. M., Eiler, D. & Kieft, J. S. Structured RNAs that evade or confound exonucleases: function follows form. *Curr Opin Struct Biol* **36**, 40–47 (2016).

25. Jones, C. I., Zabolotskaya, M. V. & Newbury, S. F. The 5'-3' exoribonuclease Xrn1/Pacman and its functions in cellular processes and development. *Wires RNA* **3**, 455–468 (2012).

26. Antic, S., Wolfinger, M. T., Skucha, A., Hosiner, S. & Dorner, S. General and miRNA-mediated mRNA degradation occurs on ribosome complexes in Drosophila cells. *Mol Cell Biol* MCB–01346, DOI: 10.1128/MCB.01346-14 (2015).

27. Funk, A. *et al.* RNA structures required for production of subgenomic flavivirus RNA. *J Virol* **84**, 11407–11417 (2010).

28. Chapman, E. G., Moon, S. L., Wilusz, J. & Kieft, J. S. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *elife* **3**, e01892 (2014).

29. Liu, R. *et al.* Identification and characterization of small sub-genomic RNAs in dengue 1–4 virus-infected cell cultures and tissues. *Biochem. Bioph Res Co* **391**, 1099–1103 (2010).

30. Silva, P. A., Pereira, C. F., Dalebout, T. J., Spaan, W. J. & Bredenbeek, P. J. An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. *J Virol* **84**, 11395–11406 (2010).

31. Fan, Y.-H. *et al.* Small noncoding RNA modulates Japanese encephalitis virus replication and translation in trans. *Virol J* **8**, 492 (2011).

32. Akiyama, B. M. *et al.* Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* aah3963 (2016).

33. Schnettler, E. *et al.* Induction and suppression of tick cell antiviral RNAi responses by tick-borne flaviviruses. *Nucleic Acids Res* **42**, 9436–9446 (2014).

34. MacFadden, A. *et al.* Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. *Nat Commun* **9**, 119 (2018).

35. Ochsenreiter, R., Hofacker, I. L. & Wolfinger, M. T. Functional RNA Structures in the 3'UTR of Tick-borne, Insect-specific and No Known Vector Flaviviruses. *Viruses* (2019).

36. Sztuba-Solinska, J. *et al.* Structural complexity of Dengue virus untranslated regions: cis-acting RNA motifs and pseudoknot interactions modulating functionality of the viral genome. *Nucleic Acids Res* **41**, 5075–5089 (2013).

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

38. Hofacker, I. L. *et al.* Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucleic Acids Res* **26**, 3825–3836 (1998).

39. Witwer, C., Rauscher, S., Hofacker, I. L. & Stadler, P. F. Conserved RNA secondary structures in picornaviridae genomes. *Nucleic Acids Res* **29**, 5079–5089 (2001).

40. Hofacker, I. L., Stadler, P. F. & Stocsits, R. R. Conserved RNA secondary structures in viral genomes: a survey. *Bioinformatics* **20**, 1495–1499 (2004).

41. Thurner, C., Witwer, C., Hofacker, I. L. & Stadler, P. F. Conserved RNA secondary structures in flaviviridae genomes. *J Gen Virol* **85**, 1113–1124 (2004).

42. Kieft, J. S., Rabe, J. L. & Chapman, E. G. New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: Conservation, folding, and host adaptation. *RNA Biol.* **12**, 1169–1177 (2015).

43. Roby, J. A., Pijlman, G. P., Wilusz, J. & Khromykh, A. A. Noncoding subgenomic flavivirus RNA: multiple functions in West Nile virus pathogenesis and modulation of host responses. *Viruses* **6**, 404–427 (2014).

44. Hussain, M. *et al.* West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. *Nucleic Acids Res* **40**, 2210–2223 (2011).

45. Schuessler, A. *et al.* West Nile virus noncoding subgenomic RNA contributes to viral evasion of the type I interferon-mediated antiviral response. *J Virol* **86**, 5708–5718 (2012).

46. Manokaran, G. *et al.* Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science* **350**, 217–221 (2015).

47. Moon, S. L. *et al.* A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability. *RNA* (2012).

48. Schnettler, E. *et al.* Non-coding flavivirus RNA displays RNAi suppressor activity in insect and mammalian cells. *J Virol* JVI–01104 (2012).

49. Sakakibara, S.-i. *et al.* Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev Biol* **176**, 230–242 (1996).

50. Sakakibara, S.-i., Nakamura, Y., Satoh, H. & Okano, H. RNA-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *J Neurosci* **21**, 8091–8107 (2001).

51. Imai, T. *et al.* The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol* **21**, 3888–3900 (2001).

52. Kharas, M. G. *et al.* Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. *Nat Med* **16**, 903 (2010).

53. ErLin, S. *et al.* Musashi-1 maintains blood–testis barrier structure during spermatogenesis and regulates stress granule formation upon heat stress. *Mol Biol Cell* **26**, 1947–1956 (2015).

54. Ohyama, T. *et al.* Structure of Musashi1 in a complex with target RNA: the role of aromatic stacking interactions. *Nucleic Acids Res* **40**, 3218–3231 (2012).

55. Iwaoka, R. *et al.* Structural Insight into the Recognition of r(UAG) by Musashi-1 RBD2, and Construction of a Model of Musashi-1 RBD1-2 Bound to the Minimum Target RNA. *Molecules* **22**, 1207 (2017).

56. Zearfoss, N. R. *et al.* A conserved three-nucleotide core motif defines Musashi RNA binding specificity. *J Biolo Chem* **289**, 35530–35541 (2014).

57. Uren, P. J. *et al.* RNA-binding protein Musashi1 is a central regulator of adhesion pathways in glioblastoma. *Mol Cell Biol* **35**, 2965–2978 (2015).

58. Chavali, P. L. *et al.* Neurodevelopmental protein Musashi 1 interacts with the Zika genome and promotes viral replication. *Science* eaam9243 (2017).

59. Platt, D. J. *et al.* Zika virus-related neurotropic flaviviruses infect human placental explants and cause fetal demise in mice. *Sci Transl Med* **10**, eaao7090 (2018).

60. Haddow, A. D. *et al.* Genetic Characterization of Spondweni and Zika Viruses and Susceptibility of Geographically Distinct Strains of *Aedes aegypti*, *Aedes albopictus* and *Culex quinquefasciatus* (Diptera: Culicidae) to Spondweni Virus. *PLoS Negl. Trop D* **10**, e0005083 (2016).

61. Lorenz, R. *et al.* Viennarna Package 2.0. *Algorithm Mol Biol* **6**, 26 (2011).

62. McCaskill, J. S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**, 1105–1119 (1990).

63. Bernhart, S. H., Hofacker, I. L. & Stadler, P. F. Local RNA base pairing probabilities in large sequences. *Bioinformatics* **22**, 614–615 (2005).

64. Lorenz, R., Wolfinger, M. T., Tanzer, A. & Hofacker, I. L. Predicting RNA secondary structures from sequence and probing data. *Methods* **103**, 86–98, DOI: 10.1016/j.ymeth.2016.04.004 (2016).

65. Bernhart, S. H., Mückstein, U. & Hofacker, I. L. RNA Accessibility in cubic time. *Algorithm Mol Biol* **6**, 3 (2011).

66. Wolfinger, M. T., Fallmann, J., Eggenhofer, F. & Amman, F. ViennaNGS: A toolbox for building efficient next-generation sequencing analysis pipelines. *F1000Research* **4**, DOI: 10.12688/f1000research.6157.1 (2015).

67. Jiang, M., Anderson, J., Gillespie, J. & Mayne, M. uShuffle: a useful tool for shuffling biological sequences while preserving the k-let counts. *BMC Bioinforma.* **9**, 192 (2008).

68. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, DOI: 10.1093/bioinformatics/btt509 (2013).

69. Will, S., Reiche, K., Hofacker, I. L., Stadler, P. F. & Backofen, R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comp Biol* **3**, e65 (2007).

70. Wolfinger, M. T. Bio::RNA::RNAaliSplit 0.09, DOI: 10.5281/zenodo.2532826 (2019). Https://github.com/mtw/Bio-RNA-RNAaliSplit.

71. Weinberg, Z. & Breaker, R. R. R2R-software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinforma.* **12**, 3 (2011).

72. de Bernardi Schneider, A. & Wolfinger, M. T. Preventing disease outbreaks with computational biology, how far can we go? *NCT CBNW Newsl.* **58**, DOI: 10.5281/zenodo.1463018 (2018).

73. Richard, A. S. *et al.* AXL-dependent infection of human fetal endothelial cells distinguishes Zika virus from other pathogenic flaviviruses. *Proc Natl Acad Sci USA* 201620558 (2017).

74. Kakooza-Mwesige, A., Mohammed, A. H., Kristensson, K., Juliano, S. L. & Lutwama, J. J. Emerging viral infections in Sub-Saharan Africa and the Developing nervous System: A Mini Review. *Front Neurol* **9**, 82 (2018).

75. Junglen, S. *et al.* A new flavivirus and a new vector: characterization of a novel flavivirus isolated from uranotaenia mosquitoes from a tropical rain forest. *J Virol* **83**, 4462–4468 (2009).

76. Huhtamo, E. *et al.* Novel flaviviruses from mosquitoes: Mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. *Virology* **464**, 320–329 (2014).

77. Lvov, D. *et al.* "Karshi" virus, a new flavivirus (Togaviridae) isolated from *Ornithodoros papillipes* (Birula, 1895) ticks in Uzbek SSR. *Arch Virol* **50**, 29–36 (1976).

78. McLean, D. & Donohue, W. Powassan virus: isolation of virus from a fatal case of encephalitis. *Canad Med Assoc J* **80**, 708 (1959).

79. Piantadosi, A. *et al.* Emerging cases of Powassan virus encephalitis in New England: clinical presentation, imaging, and review of the literature. *Clin Infect Dis* **62**, 707–713 (2015).

80. O'Leary, D. R. *et al.* Birth outcomes following West Nile Virus infection of pregnant women in the United States: 2003-2004. *Pediatrics* **117**, e537–e545 (2006).

81. Nishioka, D. A., Pires, W. P., Silva, F. A., Costa, H. L. *et al.* Yellow fever vaccination during pregnancy and spontaneous abortion: a case-control study. *Trop Med Int Heal.* **3**, 29–33 (1998).

82. Woodruff, A., Bowen, E. & Platt, G. Viral infections in travellers from tropical Africa. *Br Med J* **1**, 956–958 (1978).

83. for Disease Control, C., (CDC, P. *et al.* Intrauterine West Nile virus infection–New York, 2002. *MMWR. Morb Mortal W* **51**, 1135 (2002).

84. Alpert, S. G., Fergerson, J. & Noël, L.-P. Intrauterine West Nile virus: ocular and systemic findings. *Am J Ophthalmol* **136**, 733–735 (2003).

85. Julander, J. G. *et al.* Treatment of West Nile virus-infected mice with reactive immunoglobulin reduces fetal titers and increases dam survival. *Antivir Res* **65**, 79–85 (2005).

86. Salinas, S. *et al.* Deleterious effect of Usutu virus on human neural cells. *PLoS Negl. Trop Dis* **11**, e0005913 (2017).

87. Bassi, M. R., Sempere, R. N., Meyn, P., Polacek, C. & Arias, A. Extinction of Zika virus and Usutu virus by lethal mutagenesis reveals different patterns of sensitivity to three mutagenic drugs. *Antimicrob Agents Ch* AAC–00380 (2018).

88. Yin, X., Zhong, X. & Pan, S. Vertical transmission of dengue infection: the first putative case reported in China. *Rev Inst Med Trop SP* **58** (2016).

89. Ranjan, R., Kumar, K. & Nagar, N. Congenital dengue infection: Are we missing the diagnosis? *Pediatr Infect Dis J* **8**, 120–123 (2016).

90. Sakakibara, S.-i. *et al.* RNA-binding protein Musashi family: roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation. *Proc Nat Acad Sci USA* **99**, 15194–15199 (2002).

91. D'Ortenzio, E. *et al.* Evidence of sexual transmission of Zika virus. *New Engl J Med* **374**, 2195–2198 (2016).

92. Kawahara, H. *et al.* Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. *J Cell Biol* **181**, 639–653 (2008).

93. Katz, Y. *et al.* Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state. *Elife* **3** (2014).

94. Clarke, B., Roby, J., Slonchak, A. & Khromykh, A. Functional non-coding RNAs derived from the flavivirus 3' untranslated region. *Virus Res* **206**, 53–61 (2015).

Funding

This work was partly funded by the Austrian science fund FWF project F43 "RNA regulation of the transcriptome".

Acknowledgements

We thank Nikos Vasilakis for providing Spondweni virus sequences. We further thank Ivo Hofacker for fruitful discussions.

Author contributions statement

A.B.S. and M.T.W. conceived the study, conducted the *in silico* experiments, analysed the results and wrote the manuscript. Both authors reviewed the manuscript.

Data availability

The `plfoldz.pl` Perl Utility for computing RNA opening energy z scores is available from https://github.com/mtw/plfoldz.

Additional information

Competing interests: None declared.