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

Adriano de Bernardi Schneider [1] and Michael T. Wolfinger [2,3*]

September 3, 2018

[1]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, North Carolina, 28223, United States of America
[2]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria [3]Center for Anatomy and Cell Biology, Medical University of Vienna, Währingerstraße 13, 1090 Vienna, Austria

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 `UAG` motifs in the 3'UTR of RNA. We systematically analyzed the properties of Musashi binding elements (MBEs) in the 3'UTR of flaviviruses based on 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 shows that ZIKV MBEs are maximally accessible among all mosquito-borne flaviviruses. Our study addresses the broader question whether other emerging arboviruses can cause similar neurotropic effects through the same mechanism in the developing fetus. In this line, we establish 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.

1 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 [80, 50, 30].

*To whom correspondence should be addressed. Email: michael.wolfinger@univie.ac.at

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 vertrebrate 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 [25, 84].

Initially isolated in 1947 from a sentinel rhesus macaque in the Zika 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 [78, 14]. There are two main ZIKV lineages, the original African (type strain MR766) and an Asian (type strain FSS13025) [24, 77], the latter also comprising American strains such as PE243.

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 [44]. While MBFVs are typically transmitted by host-vector interaction, vertical transmission from mother to child during pregnancy via transplacental infection has been reported [63].

The neurotropic potential of ZIKV-related flaviviruses has been known since the 1970s, when Saint Louis encephalitis virus (SLEV) has been attributed to severe neurological disorder in infected mice [4, 5]. Vertical transmission has been observed with JEV in mice [22] and human [11] and a case of human fetal infection has been reported after YFV vaccination [81]. Other transmission pathways of ZIKV include blood transfusions and sexual transmission [21, 56]. Despite enormous efforts in studying ZIKV infections in the last years, the biological reasoning and mechanisms behind arbovirus congenital neurotropism remain elusive.

1.1 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

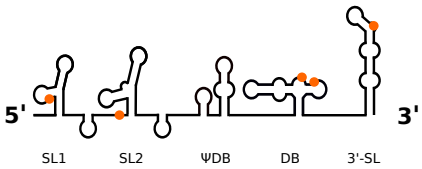

Figure 1: Schematic representation of the ZIKV 3'UTR. Conserved RNA elements include two stem-loop structures (SL1 and SL2), a ΨDB and canonical DB element as well as the terminal 3' stem-loop structure (3-SL). Positions of Musashi-binding `UAG` motifs in the Asian/American ZIKV lineage are highlighted in orange.

(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 [27, 55, 83, 16].

1.2 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 noncoding RNA (lncRNA) has been observed upon infection. These lncRNAs, also known as subgenomic flaviviral RNAs (sfRNAs) [62], are stable decay intermediates derived from exploiting the host's mRNA degradation machinery [1].

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 [36, 6]. 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 [23, 10]. Xrn1-resistant RNAs and sfRNAs appear to be ubiquitously present in many flaviviruses. They have been described in MBFVs, including DENV [45], YFV [76], JEV [19], and ZIKV [2], TBFVs [62, 74], and recently in ISFVs and NKVs [49]. There is typically more than one xrRNA, which is reflected in the varied molecular architecture of different flavivirus 3'UTRs. Earlier studies in our group have identified conserved RNA structural elements in viral 3'UTRs [66, 28, 87, 29, 79], some of which have later been attributed to xrRNA functionality [62]. 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 [43].

The inhibition of Xrn1 by viral RNA yields sfRNAs that affect many cellular processes, both in the vector and the host [68]. In mosquitoes, sfRNA

interacts directly with the predominant innate immune response pathway, RNA interference (RNAi), by serving as a template for microRNA (miRNA) biogenesis [32]. Conversely, in host cells sfRNA modulates the anti-viral interferon response [75], e.g., by binding proteins to inhibit the translation of interferon-stimulated genes [51]. Moreover, sfRNA has been shown to inhibit Xrn1 and Dicer activity, thereby altering host mRNA levels [54, 73].

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. [68] 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.

1.3 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 [69, 70, 33] and overexpressed in tumors and leukemias [42], they are absent in differentiated tissue. Moreover, Msi1 has been shown to be involved in the regulation of blood-testis barrier proteins and spermatogenesis in mice [18]. Musashi proteins have two RNA recognition motif (RRM) domains, whose sequence specificity has been determined by an *in vitro* selection method and NMR spectroscopy [33, 59, 34]. The trinucleotide sequence `UAG` has been identified as core Musashi binding element (MBE) and its thermodynamic binding specificity was determined by fluorescence polarization assays [91]. Uren et al. [82] performed iCLIP experiments with Msi1 in human glioblastoma cells and confirmed the preferential binding of Msi1 to single-stranded (stem-loop) `UAG` sequences in 3'UTRs. Conversely, they could not determine a preference for `UAG` binding in coding regions.

1.4 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 [44], de Bernardi Schneider et al. [14] 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. [12] 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. [64] 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.

2 MATERIALS AND METHODS

2.1 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 lists all viruses analyzed in this work and gives the number of Musashi binding elements (`UAG` trinucleotides) present in 3'UTR regions.

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 [26], we were looking to include this sequence into our analysis. Nikos Vasilakis (Univ. of Texas Medical Branch, Galveston, TX) 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.

2.2 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 [46]. 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 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

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

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

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

The partition function Z can be computed efficiently via dynamic programming [52] and allows calculation of individual base pair probabilities, even for large sequences [8]. 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) [47]. 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` [8] 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 [46] and can compute the accessibilities or single-strandedness of all intervals of an RNA in cubic time [9]. 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

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

where $E_{\text{open}}(x)$ is the opening energy of trinucleotide x in its genomic context, μ and σ are the mean and standard deviations, respectively, of the opening energies of x 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 x in place.

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` [46] scripting language interface for thermodynamics calculations, the `ViennaNGS` [88] suite for extraction of genomic loci and the uShuffle Perl bindings [35] 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.

2.3 Characterization of MBEs within xrRNAs

To localize MBEs within homologous substructures in flavivirus 3'UTRs we constructed `infernal` [57] covariance models for conserved xrRNA elements. The structural RNA alignments underlying the `infernal` models were computed with `locarna` [86] and further analyzed with `RNAalifold` [46].

3 RESULTS

3.1 MBEs are highly accessible in ZIKV 3'UTRs

The Musashi family of proteins preferentially bind single-stranded `UAG` motifs in 3'UTRs [82]. 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 x according to Eq. 4, thereby normalizing the opening energy of x 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 [12] by means of a thermodynamic model, thus underlining a possible role of Msi1 in ZIKV neurotropism. 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.

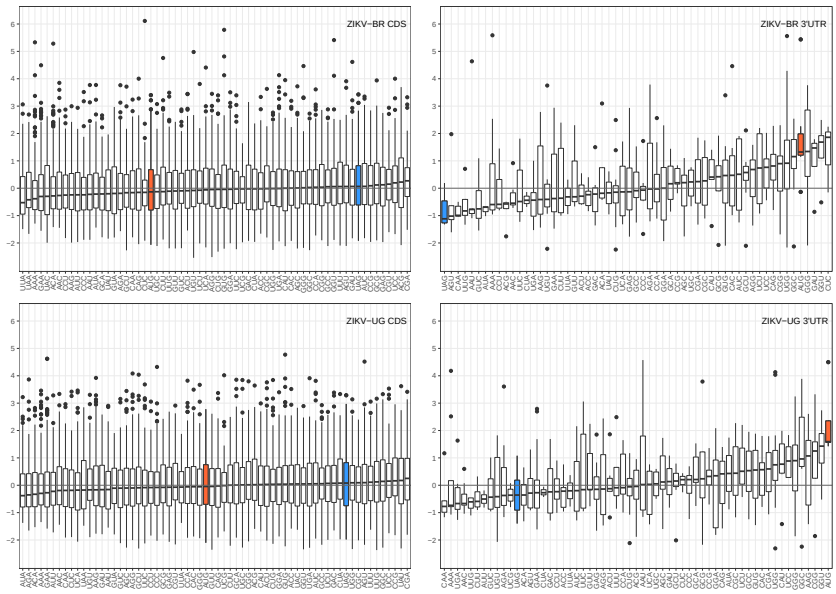

Figure 2: Distribution of z scores of opening energies for trinucleotides found in the coding region (CDS, left) and 3'UTR (right) of ZIKV from Brazil (top) and Uganda (bottom), sorted by median opening energy z score. The MBE motif is highlighted in blue and shows low overall z scores in the 3'UTR, 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.

3.2 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 out-

lined above (97 nt upstream/downstream windows, $n = 10,000$ dinucleotide shufflings).

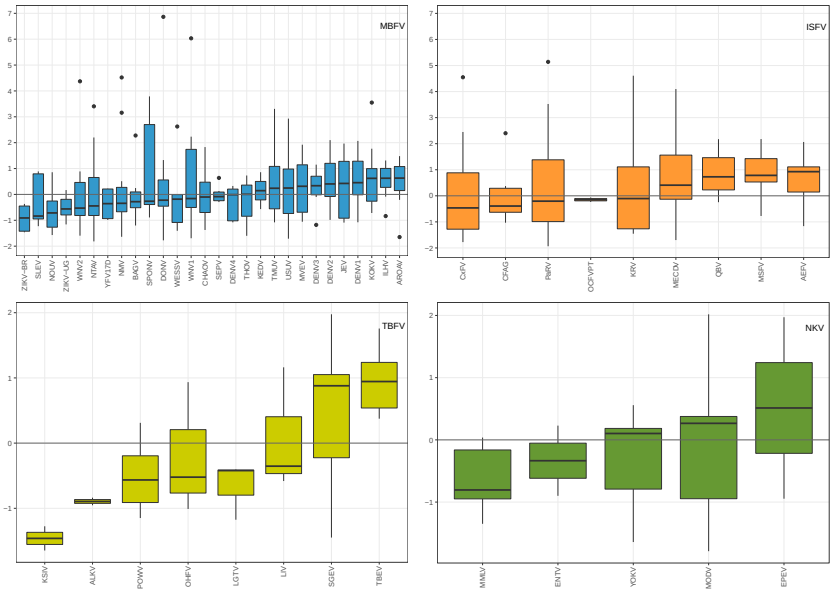

Figure 3: Distribution of MBE opening energy z scores in flavivirus 3'UTRs, grouped by vector specificity and sorted by median z scores. Top left: MBFVs, top right: ISFVs, bottom left: TBFVs and bottom right: NKVs. The Asian/American lineage ZIKV-BR isolate has the lowest median z score among all MBFV. 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.

Msi1 preferentially binds single stranded RNA [82], 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. Interestingly, 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. Among NKVs, Montana myotis leukoencephali-

tis 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.

3.3 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.

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.

4 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

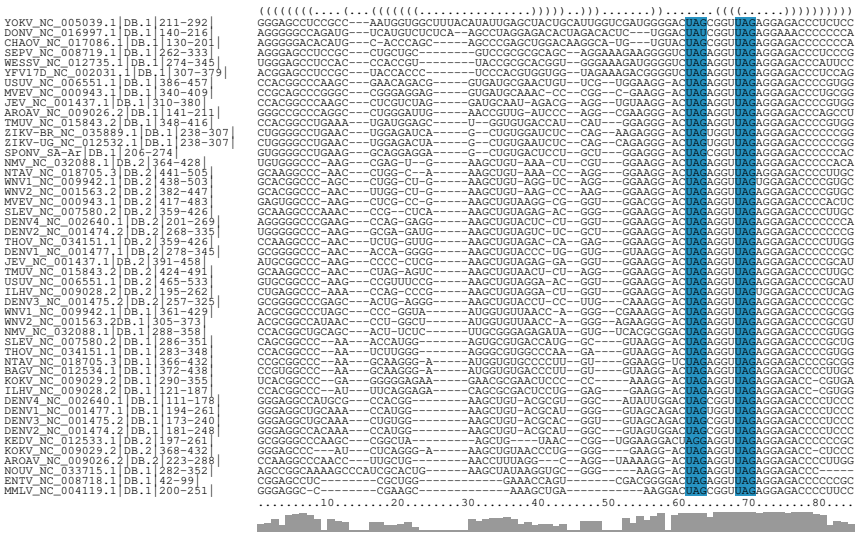

Figure 4: Structural alignment of conserved dumbbell (DB) elements in the 3'UTR of MBFV and NKV flaviviruses. Several species have two copies of DB elements in their 3'UTR, indicated by `DB.1` and `DB.2` in the sequence identifier. Coordinates are given relative to the 3'UTR start. A consensus structure 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 substructure (positions 60-80). Two conserved MBE motifs in the central multiloop and distal stem loop are highlighted in blue.

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 [15].

Previous experiments lead toward the idea that ZIKV is unique among flaviviruses regarding the clinical outcomes resulting from congenital infection [67, 39]. 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 Culicidae mosquito not known to harbor flaviviruses before [38]. While replication has been tested on human and non-human cell lines, vertebrate infection and pathogenesis could not be observed [31].

Within the TBFV serocomplex, KSIV has the lowest overall MBE opening energies. Originally isolated from *Ornithodoros papillipes* ticks in Uzbekistan in 1972 [48] it currently does not present history of infection in humans. Conversely, Powassan virus (POWV), another TBFV with negative MBE opening

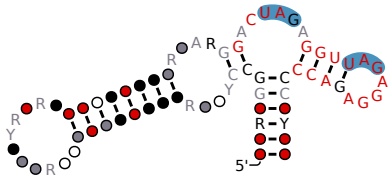

Figure 5: Consensus secondary structure of the flavivirus DB element with MBEs highlighted in blue. Figure generated from the MSA in Figure 4 with `R2R` [85]. Structure and color annotation inferred by `R2R`. Nucleotide symbols represent conserved nucleotides. Circles represent columns in the MSA that 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.

energy, originally isolated in Powassan, Ontario, Canada in 1959 from a child who died of acute encephalitis [53] can cause transplacental infection [64] and has been associated with severe neuropathology and death in mice and human [61].

Given that `UAG` can be regarded the primary Msi1 binding motif, we can argue that ZIKV has the highest affinity for binding Msi1 among all MBFVs. Platt et al. [64] 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 [4, 5], WNV [60, 64], YFV [81, 58] and POWV [64].

Bizarre neurological manifestations were also observed in patients infected by Ntaya virus (NTAV) [89], a neurotropic viruses from the Japanese encephalitis serocomplex, as well as WNV in humans [20, 3] and in mice [37], USUV [72, 7] and DENV [90, 65]. 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.

4.1 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.

Although the Musashi proteins could bind motifs other than `UAG`, `UAG` has been shown to be the motif contributing the highest binding free energy [91]. 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 `UAGs` within DB elements, but also those that overlap with SL elements show a negative opening energy z-score.

4.2 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 [12], this interaction could be a critical reason why ZIKV persists in sperm for a long time after the individual has been infected [71, 17], 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.

4.3 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 translation initiation factor eIF4G for binding to poly(A)-binding protein (PABP), thereby inhibiting the assembly of the 80S ribosomal unit [41]. Ribosome profiling experiments have corroborated this down-regulatory effect of Msi1, while keeping mRNA levels [40]. This allows for a speculative explanation of the findings by Chavali et al. [12], 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 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 [13, 51].

5 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 a similar neurotropic potential and thus might be potent emerging pathogens.

6 FUNDING

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

7 ACKNOWLEDGEMENTS

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

7.0.1 Conflict of interest statement.

None declared.

References

[1] Benjamin M Akiyama, Daniel Eiler, and Jeffrey S Kieft. Structured RNAs that evade or confound exonucleases: function follows form. *Curr Opin Struct Biol*, 36:40–47, 2016.

[2] Benjamin M Akiyama, Hannah M Laurence, Aaron R Massey, David A Costantino, Xuping Xie, Yujiao Yang, Pei-Yong Shi, Jay C Nix, J David Beckham, and Jeffrey S Kieft. Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science*, page aah3963, 2016.

[3] Samuel G Alpert, Jacqueline Fergerson, and Léon-Paul Noël. Intrauterine West Nile virus: ocular and systemic findings. *Am J Ophthalmol*, 136(4):733–735, 2003.

[4] A Andersen and R Hanson. Experimental transplacental transmission of St. Louis encephalitis virus in mice. *Infect Immun*, 2(3):320–325, 1970.

[5] A Andersen and R Hanson. Intrauterine infection of mice with St. Louis encephalitis virus: immunological, physiological, neurological, and behavioral effects on progeny. *Infect Immun*, 12(5):1173–1183, 1975.

[6] Sanja Antic, Michal T. Wolfinger, Anna Skucha, Stefanie Hosiner, and Silke Dorner. General and miRNA-mediated mRNA degradation occurs on ribosome complexes in Drosophila cells. *Mol Cell Biol*, pages MCB–01346, 2015.

[7] Maria Rosaria Bassi, Raquel Navarro Sempere, Prashansa Meyn, Charlotta Polacek, and Armando Arias. Extinction of Zika virus and Usutu virus by lethal mutagenesis reveals different patterns of sensitivity to three mutagenic drugs. *Antimicrob Agents Ch*, pages AAC–00380, 2018.

[8] Stephan H Bernhart, Ivo L Hofacker, and Peter F Stadler. Local RNA base pairing probabilities in large sequences. *Bioinformatics*, 22(5):614–615, 2005.

[9] Stephan H Bernhart, Ullrike Mückstein, and Ivo L Hofacker. RNA Accessibility in cubic time. *Algorithm Mol Biol*, 6(1):3, 2011.

[10] Erich G Chapman, Stephanie L Moon, Jeffrey Wilusz, and Jeffrey S Kieft. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *elife*, 3:e01892, 2014.

[11] UC Chaturvedi, A Mathur, A Chandra, SK Das, HO Tandon, and UK Singh. Transplacental infection with Japanese encephalitis virus. *J Infect Dis*, 141(6):712–715, 1980.

[12] Pavithra L Chavali, Lovorka Stojic, Luke W Meredith, Nimesh Joseph, Michael S Nahorski, Thomas J Sanford, Trevor R Sweeney, Ben A Krishna, Myra Hosmillo, Andrew E Firth, Richard Bayliss, Carlo L Marcelis, Ian Goodfellow, C Geoffrey Woods, and Fanni Gergely. Neurodevelopmental protein Musashi 1 interacts with the Zika genome and promotes viral replication. *Science*, page eaam9243, 2017.

[13] BD Clarke, JA Roby, A Slonchak, and AA Khromykh. Functional noncoding RNAs derived from the flavivirus 3' untranslated region. *Virus Res*, 206:53–61, 2015.

[14] Adriano de Bernardi Schneider, Robert W Malone, Jun-Tao Guo, Jane Homan, Gregorio Linchangco, Zachary L Witter, Dylan Vinesett, Lambodhar Damodaran, and Daniel A Janies. Molecular evolution of Zika virus as it crossed the Pacific to the Americas. *Cladistics*, 33(1):1–20, 2017.

[15] Adriano de Bernardi Schneider and Michael T. Wolfinger. Preventing disease outbreaks with computational biology, how far can we go? *NCT CBNW Newsletter*, 58, 2018.

[16] Luana de Borba, Sergio M Villordo, Nestor G Iglesias, Claudia V Filomatori, Leopoldo G Gebhard, and Andrea V Gamarnik. Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J Virol*, 89(6):3430–3437, 2015.

[17] Eric DOrtenzio, Sophie Matheron, Xavier de Lamballerie, Bruno Hubert, Géraldine Piorkowski, Marianne Maquart, Diane Descamps, Florence Damond, Yazdan Yazdanpanah, and Isabelle Leparc-Goffart. Evidence of sexual transmission of Zika virus. *New Engl J Med*, 374(22):2195–2198, 2016.

[18] Sun ErLin, Wei WenJie, Wang LiNing, Lu BingXin, Lei MingDe, Sun Yan, and Han RuiFa. Musashi-1 maintains blood–testis barrier structure during spermatogenesis and regulates stress granule formation upon heat stress. *Mol Biol Cell*, 26(10):1947–1956, 2015.

[19] Yi-Hsin Fan, Muthukumar Nadar, Chiu-Chin Chen, Chia-Chen Weng, Yun-Tong Lin, and Ruey-Yi Chang. Small noncoding RNA modulates Japanese encephalitis virus replication and translation in trans. *Virol J*, 8(1):492, 2011.

[20] Centers for Disease Control, Prevention (CDC, et al. Intrauterine West Nile virus infection–New York, 2002. *MMWR. Morb Mortal W*, 51(50):1135, 2002.

[21] Brian D Foy, Kevin C Kobylinski, Joy L Chilson Foy, Bradley J Blitvich, Amelia Travassos da Rosa, Andrew D Haddow, Robert S Lanciotti, and Robert B Tesh. Probable non–vector-borne transmission of Zika virus, Colorado, USA. *Emerg Infect Dis*, 17(5):880, 2011.

[22] Y Fujisaki, Y Miura, T Sugimori, Y Murakami, and K Miura. Experimental studies on vertical infection of mice with Japanese encephalitis virus. IV. Effect of virus strain on placental and fetal infection. *NatL I Anim Health Q*, 23(1):21–26, 1983.

[23] Anneke Funk, Katherine Truong, Tomoko Nagasaki, Shessy Torres, Nadia Floden, Ezequiel Balmori Melian, Judy Edmonds, Hongping Dong, Pei-Yong Shi, and Alexander A Khromykh. RNA structures required for production of subgenomic flavivirus RNA. *J Virol*, 84(21):11407–11417, 2010.

[24] Zhen Gong, Xiaoyu Xu, and Guan-Zhu Han. The diversification of Zika virus: Are there two distinct lineages? *Genome Biol Evol*, 9(11):2940–2945, 2017.

[25] DJ Gubler, G Kuno, and L Markoff. Flaviviruses. *Fields Virology*, 1:1153–1252, 2007.

[26] Andrew D Haddow, Farooq Nasar, Hilda Guzman, Alongkot Ponlawat, Richard G Jarman, Robert B Tesh, and Scott C Weaver. 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 Neglect Trop D*, 10(10):e0005083, 2016.

[27] Chang S Hahn, Young S Hahn, Charles M Rice, Eva Lee, Lynn Dalgarno, Ellen G Strauss, and James H Strauss. Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *J Mol Biol*, 198(1):33–41, 1987.

[28] Ivo L Hofacker, Martin Fekete, Christoph Flamm, Martijn A Huynen, Susanne Rauscher, Paul E Stolorz, and Peter F Stadler. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucleic acid res*, 26(16):3825–3836, 1998.

[29] Ivo L Hofacker, Peter F Stadler, and Roman R Stocsits. Conserved RNA secondary structures in viral genomes: a survey. *Bioinformatics*, 20(10):1495–1499, 2004.

[30] Peter J Hotez and Kristy O Murray. Dengue, West Nile virus, Chikungunya, Zika and now Mayaro? *PLoS Neglect Trop D*, 11(8):e0005462, 2017.

[31] Eili Huhtamo, Shelley Cook, Gregory Moureau, Nathalie Y Uzcátegui, Tarja Sironen, Suvi Kuivanen, Niina Putkuri, Satu Kurkela, Ralph E Harbach, Andrew E Firth, Olli Vapalahti, Ernest A Gould, and Xavier de Lamballerie. Novel flaviviruses from mosquitoes: Mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. *Virology*, 464:320–329, 2014.

[32] Mazhar Hussain, Shessy Torres, Esther Schnettler, Anneke Funk, Adam Grundhoff, Gorben P Pijlman, Alexander A Khromykh, and Sassan Asgari. 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(5):2210–2223, 2011.

[33] Takao Imai, Akinori Tokunaga, Tetsu Yoshida, Mitsuhiro Hashimoto, Katsuhiko Mikoshiba, Gerry Weinmaster, Masato Nakafuku, and Hideyuki Okano. The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol*, 21(12):3888–3900, 2001.

[34] Ryo Iwaoka, Takashi Nagata, Kengo Tsuda, Takao Imai, Hideyuki Okano, Naohiro Kobayashi, and Masato Katahira. 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(7):1207, 2017.

[35] Minghui Jiang, James Anderson, Joel Gillespie, and Martin Mayne. uShuffle: a useful tool for shuffling biological sequences while preserving the k-let counts. *BMC Bioinformatics*, 9(1):192, 2008.

[36] Christopher Iain Jones, Maria Vasilyevna Zabolotskaya, and Sarah Faith Newbury. The 5'-3' exoribonuclease Xrn1/Pacman and its functions in cellular processes and development. *Wires RNA*, 3(4):455–468, 2012.

[37] Justin G Julander, Quinton A Winger, Aaron L Olsen, Craig W Day, Robert W Sidwell, and John D Morrey. Treatment of West Nile virus-infected mice with reactive immunoglobulin reduces fetal titers and increases dam survival. *Antivir Res*, 65(2):79–85, 2005.

[38] Sandra Junglen, Anne Kopp, Andreas Kurth, Georg Pauli, Heinz Ellerbrok, and Fabian H Leendertz. A new flavivirus and a new vector: characterization of a novel flavivirus isolated from uranotaenia mosquitoes from a tropical rain forest. *J Virol*, 83(9):4462–4468, 2009.

[39] Angelina Kakooza-Mwesige, Abdul H Mohammed, Krister Kristensson, Sharon L Juliano, and Julius J Lutwama. Emerging viral infections in Sub-Saharan Africa and the Developing nervous System: A Mini Review. *Front Neurol*, 9:82, 2018.

[40] Yarden Katz, Feifei Li, Nicole J Lambert, Ethan S Sokol, Wai-Leong Tam, Albert W Cheng, Edoardo M Airoldi, Christopher J Lengner, Piyush B Gupta, Zhengquan Yu, Rudolf Jaenisch, and Christopher B Burge. Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state. *Elife*, 3, 2014.

[41] Hironori Kawahara, Takao Imai, Hiroaki Imataka, Masafumi Tsujimoto, Ken Matsumoto, and Hideyuki Okano. Neural RNA-binding protein

Musashi1 inhibits translation initiation by competing with eIF4G for PABP. *J Cell Biol*, 181(4):639–653, 2008.

[42] Michael G Kharas, Christopher J Lengner, Fatima Al-Shahrour, Lars Bullinger, Brian Ball, Samir Zaidi, Kelly Morgan, Winnie Tam, Mahnaz Paktinat, Rachel Okabe, et al. Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. *Nat Med*, 16(8):903, 2010.

[43] Jeffrey S Kieft, Jennifer L Rabe, and Erich G Chapman. New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: Conservation, folding, and host adaptation. *RNA Biology*, 12(11):1169–1177, 2015.

[44] Zachary A Klase, Svetlana Khakhina, Adriano De Bernardi Schneider, Michael V Callahan, Jill Glasspool-Malone, and Robert Malone. Zika fetal neuropathogenesis: etiology of a viral syndrome. *PLoS Neglect Trop D*, 10(8):e0004877, 2016.

[45] Ran Liu, Lei Yue, Xiaofeng Li, Xuedong Yu, Hui Zhao, Zhenyou Jiang, Ede Qin, and Chengfeng Qin. Identification and characterization of small sub-genomic RNAs in dengue 1–4 virus-infected cell cultures and tissues. *Biochem Bioph Res Co*, 391(1):1099–1103, 2010.

[46] Ronny Lorenz, Stephan H Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. Viennarna Package 2.0. *Algorithm Mol Biol*, 6(1):26, 2011.

[47] Ronny Lorenz, Michael T Wolfinger, Andrea Tanzer, and Ivo L Hofacker. Predicting RNA secondary structures from sequence and probing data. *Methods*, 103:86–98, 2016.

[48] DK Lvov, VM Neronov, VL Gromashevsky, TM Skvortsova, LK Berezina, GA Sidorova, ZM Zhmaeva, Yu A Gofman, SM Klimenko, and KB Fomina. "Karshi" virus, a new flavivirus (Togaviridae) isolated from *Ornithodoros papillipes* (Birula, 1895) ticks in Uzbek SSR. *Arch Virol*, 50(1-2):29–36, 1976.

[49] Andrea MacFadden, Zoe ODonoghue, Patricia AGC Silva, Erich G Chapman, René C Olsthoorn, Mark G Sterken, Gorben P Pijlman, Peter J Bredenbeek, and Jeffrey S Kieft. Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. *Nat Commun*, 9(1):119, 2018.

[50] Robert W Malone, Jane Homan, Michael V Callahan, Jill Glasspool-Malone, Lambodhar Damodaran, Adriano De Bernardi Schneider, Rebecca Zimler, James Talton, Ronald R Cobb, Ivan Ruzic, et al. Zika virus: medical countermeasure development challenges. *PLoS Neglect Trop D*, 10(3):e0004530, 2016.

[51] Gayathri Manokaran, Esteban Finol, Chunling Wang, Jayantha Gunaratne, Justin Bahl, Eugenia Z Ong, Hwee Cheng Tan, October M Sessions, Alex M Ward, Duane J Gubler, Eva Harris, Mariano A Garcia-Blanco, and Eng Eong Ooi. Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science*, 350(6257):217–221, 2015.

[52] John S McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7):1105–1119, 1990.

[53] DM McLean and WL Donohue. Powassan virus: isolation of virus from a fatal case of encephalitis. *Canad Med Assoc J*, 80(9):708, 1959.

[54] Stephanie L Moon, John R Anderson, Yutaro Kumagai, Carol J Wilusz, Shizuo Akira, Alexander A Khromykh, and Jeffrey Wilusz. A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability. *RNA*, 2012.

[55] Suchetana Mukhopadhyay, Richard J Kuhn, and Michael G Rossmann. A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol*, 3(1):13, 2005.

[56] Didier Musso, Claudine Roche, Emilie Robin, Tuxuan Nhan, Anita Teissier, and Van-Mai Cao-Lormeau. Potential sexual transmission of Zika virus. *Emerg Infect Dis*, 21(2):359, 2015.

[57] Eric P. Nawrocki and Sean R. Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935, 2013.

[58] De A Nishioka, Waldely P Pires, Flávia A Silva, Hégena L Costa, et al. Yellow fever vaccination during pregnancy and spontaneous abortion: a case-control study. *Trop Med Int Health*, 3(1):29–33, 1998.

[59] Takako Ohyama, Takashi Nagata, Kengo Tsuda, Naohiro Kobayashi, Takao Imai, Hideyuki Okano, Toshio Yamazaki, and Masato Katahira. Structure of Musashi1 in a complex with target RNA: the role of aromatic stacking interactions. *Nucleic acids res*, 40(7):3218–3231, 2012.

[60] Daniel R O'Leary, Stephanie Kuhn, Krista L Kniss, Alison F Hinckley, Sonja A Rasmussen, W John Pape, Lon K Kightlinger, Brady D Beecham, Tracy K Miller, David F Neitzel, et al. Birth outcomes following West Nile Virus infection of pregnant women in the United States: 2003-2004. *Pediatrics*, 117(3):e537–e545, 2006.

[61] Anne Piantadosi, Daniel B Rubin, Daniel P McQuillen, Liangge Hsu, Philip A Lederer, Cameron D Ashbaugh, Chad Duffalo, Robert Duncan, Jesse Thon, Shamik Bhattacharyya, Nesli Basgoz, Steven K Feske, and Jennifer L Lyons. Emerging cases of Powassan virus encephalitis in New

England: clinical presentation, imaging, and review of the literature. *Clin Infect Dis*, 62(6):707–713, 2015.

[62] Gorben P Pijlman, Anneke Funk, Natasha Kondratieva, Jason Leung, Shessy Torres, Lieke Van der Aa, Wen Jun Liu, Ann C Palmenberg, Pei-Yong Shi, Roy A Hall, et al. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe*, 4(6):579–591, 2008.

[63] Derek J Platt and Jonathan J Miner. Consequences of congenital Zika virus infection. *Curr Opin Virol*, 27:1–7, 2017.

[64] Derek J Platt, Amber M Smith, Nitin Arora, Michael S Diamond, Carolyn B Coyne, and Jonathan J Miner. Zika virus-related neurotropic flaviviruses infect human placental explants and cause fetal demise in mice. *Sci Transl Med*, 10(426):eaao7090, 2018.

[65] Ramani Ranjan, Kishore Kumar, and Nandini Nagar. Congenital dengue infection: Are we missing the diagnosis? *Pediatr Infect Dis J*, 8(4):120–123, 2016.

[66] Susanne Rauscher, Christoph Flamm, Christian W Mandl, Franz X Heinz, and Peter F Stadler. Secondary structure of the 3'-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities. *RNA*, 3(7):779–791, 1997.

[67] Audrey Stéphanie Richard, Byoung-Shik Shim, Young-Chan Kwon, Rong Zhang, Yuka Otsuka, Kimberly Schmitt, Fatma Berri, Michael S Diamond, and Hyeryun Choe. AXL-dependent infection of human fetal endothelial cells distinguishes Zika virus from other pathogenic flaviviruses. *Proc Natl Acad Sci USA*, page 201620558, 2017.

[68] Justin A Roby, Gorben P Pijlman, Jeffrey Wilusz, and Alexander A Khromykh. Noncoding subgenomic flavivirus RNA: multiple functions in West Nile virus pathogenesis and modulation of host responses. *Viruses*, 6(2):404–427, 2014.

[69] Shin-ichi Sakakibara, Takao Imai, Kayoko Hamaguchi, Masataka Okabe, Jun Aruga, Kazunori Nakajima, Daisuke Yasutomi, Takashi Nagata, Yasuyuki Kurihara, Seiichi Uesugi, et al. Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev Biol*, 176(2):230–242, 1996.

[70] Shin-ichi Sakakibara, Yuki Nakamura, Hitoshi Satoh, and Hideyuki Okano. RNA-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *J Neurosci*, 21(20):8091–8107, 2001.

[71] Shin-ichi Sakakibara, Yuki Nakamura, Tetsu Yoshida, Shinsuke Shibata, Masato Koike, Hiroshi Takano, Shuichi Ueda, Yasuo Uchiyama, Tetsuo Noda, and Hideyuki Okano. 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(23):15194–15199, 2002.

[72] Sara Salinas, Orianne Constant, Caroline Desmetz, Jonathan Barthelemy, Jean-Marc Lemaitre, Ollivier Milhavet, Nicolas Nagot, Vincent Foulongne, Florence E Perrin, Juan-Carlos Saiz, et al. Deleterious effect of Usutu virus on human neural cells. *PLoS Neglect Trop Dis*, 11(9):e0005913, 2017.

[73] Esther Schnettler, Mark G Sterken, Jason Y Leung, Stefan W Metz, Corinne Geertsema, Rob W Goldbach, Just M Vlak, Alain Kohl, Alexander A Khromykh, and Gorben P Pijlman. Non-coding flavivirus RNA displays RNAi suppressor activity in insect and mammalian cells. *J Virol*, pages JVI–01104, 2012.

[74] Esther Schnettler, Hana Tykalová, Mick Watson, Mayuri Sharma, Mark G Sterken, Darren J Obbard, Samuel H Lewis, Melanie McFarlane, Lesley Bell-Sakyi, Gerald Barry, et al. Induction and suppression of tick cell antiviral RNAi responses by tick-borne flaviviruses. *Nucleic Acids Res*, 42(14):9436–9446, 2014.

[75] Andrea Schuessler, Anneke Funk, Helen M Lazear, Daphne A Cooper, Shessy Torres, Stephane Daffis, Babal Kant Jha, Yutaro Kumagai, Osamu Takeuchi, Paul Hertzog, et al. West Nile virus noncoding subgenomic RNA contributes to viral evasion of the type I interferon-mediated antiviral response. *J Virol*, 86(10):5708–5718, 2012.

[76] Patrícia AGC Silva, Carina F Pereira, Tim J Dalebout, Willy JM Spaan, and Peter J Bredenbeek. An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. *J Virol*, 84(21):11395–11406, 2010.

[77] Yannick Simonin, Debby van Riel, Philippe Van de Perre, Barry Rockx, and Sara Salinas. Differential virulence between Asian and African lineages of Zika virus. *PLoS Negl Trop D*, 11(9):e0005821, 2017.

[78] Byung-Hak Song, Sang-Im Yun, Michael Woolley, and Young-Min Lee. Zika virus: history, epidemiology, transmission, and clinical presentation. *J Neuroimmunol*, 308:50–64, 2017.

[79] Caroline Thurner, Christina Witwer, Ivo L Hofacker, and Peter F Stadler. Conserved RNA secondary structures in flaviviridae genomes. *J Gen Virol*, 85(5):1113–1124, 2004.

[80] J Tognarelli, S Ulloa, E Villagra, J Lagos, C Aguayo, R Fasce, B Parra, J Mora, N Becerra, N Lagos, et al. A report on the outbreak of Zika virus on Easter Island, South Pacific, 2014. *Arch Virol*, 161(3):665–668, 2016.

[81] TF Tsai, Rosemary Paul, MC Lynberg, and GW Letson. Congenital yellow fever virus infection after immunization in pregnancy. *J Inf Dis*, 168(6):1520–1523, 1993.

[82] Philip J Uren, Dat T Vo, Patricia Rosa de Araujo, Rebecca Pötschke, Suzanne C Burns, Emad Bahrami-Samani, Mei Qiao, Raquel de Sousa Abreu, Helder I Nakaya, Bruna R Correa, et al. RNA-binding protein Musashi1 is a central regulator of adhesion pathways in glioblastoma. *Mol Cell Biol*, 35(17):2965–2978, 2015.

[83] Sergio M Villordo, Diego E Alvarez, and Andrea V Gamarnik. A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA*, 16(12):2325–2335, 2010.

[84] Scott C Weaver, Federico Costa, Mariano A Garcia-Blanco, Albert I Ko, Guilherme S Ribeiro, George Saade, Pei-Yong Shi, and Nikos Vasilakis. Zika virus: History, emergence, biology, and prospects for control. *Antivir Res*, 130:69–80, 2016.

[85] Zasha Weinberg and Ronald R Breaker. R2R-software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*, 12(1):3, 2011.

[86] Sebastian Will, Kristin Reiche, Ivo L Hofacker, Peter F Stadler, and Rolf Backofen. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comp Biol*, 3(4):e65, 2007.

[87] Christina Witwer, Susanne Rauscher, Ivo L Hofacker, and Peter F Stadler. Conserved RNA secondary structures in picornaviridae genomes. *Nucleic acid res*, 29(24):5079–5089, 2001.

[88] Michael T Wolfinger, Jörg Fallmann, Florian Eggenhofer, and Fabian Amman. ViennaNGS: A toolbox for building efficient next-generation sequencing analysis pipelines. *F1000Research*, 4(50), 2015.

[89] AW Woodruff, ET Bowen, and GS Platt. Viral infections in travellers from tropical Africa. *Br Med J*, 1(6118):956–958, 1978.

[90] Xueru Yin, Xiaozhu Zhong, and Shilei Pan. Vertical transmission of dengue infection: the first putative case reported in China. *Rev Inst Med Trop SP*, 58, 2016.

[91] N Ruth Zearfoss, Laura M Deveau, Carina C Clingman, Eric Schmidt, Emily S Johnson, Francesca Massi, and Sean P Ryder. A conserved three-nucleotide core motif defines Musashi RNA binding specificity. *J Biolo Chem*, 289(51):35530–35541, 2014.

Group	Accession number	Acronym	Scientific name	MBEs
MBFV	NC_009026.2	AROAV	Aroa virus	7
MBFV	NC_012534.1	BAGV	Bagaza virus	13
MBFV	NC_017086.1	CHAOV	Chaoyang virus	6
MBFV	NC_001477.1	DENV1	Dengue virus 1	7
MBFV	NC_001474.2	DENV2	Dengue virus 2	10
MBFV	NC_001475.2	DENV3	Dengue virus 3	7
MBFV	NC_002640.1	DENV4	Dengue virus 4	5
MBFV	NC_016997.1	DONV	Donggang virus	10
MBFV	NC_009028.2	ILHV	Ilheus virus	9
MBFV	NC_001437.1	JEV	Japanese encephalitis virus	15
MBFV	NC_012533.1	KEDV	Kedougou virus	2
MBFV	NC_009029.2	KOKV	Kokobera virus	8
MBFV	NC_000943.1	MVEV	Murray Valley encephalitis virus	12
MBFV	NC_032088.1	NMV	New Mapoon virus	14
MBFV	NC_033715.1	NOUV	Nounané virus	7
MBFV	NC_018705.3	NTAV	Ntaya virus	15
MBFV	NC_008719.1	SEPV	Sepik virus	6
MBFV	NC_007580.2	SLEV	Saint Louis encephalitis virus	7
MBFV	NC_034151.1	THOV	T'Ho virus	13
MBFV	NC_015843.2	TMUV	Tembusu virus	16
MBFV	NC_006551.1	USUV	Usutu virus	19
MBFV	NC_012735.1	WESSV	Wesselsbron virus	6
MBFV	NC_009942.1	WNV1	West Nile virus lineage 1	17
MBFV	NC_001563.2	WNV2	West Nile virus lineage 2	16
MBFV	NC_002031.1	YFV17D	Yellow fever virus 17D	4
MBFV	NC_035889.1	ZIKV-BR	Zika Virus - Asian-American lineage	5
MBFV	NC_012532.1	ZIKV-UG	Zika Virus - African lineage	6
MBFV	NC_029055.1	SPONV	Spondweni virus	5^a
MBFV	NC_026623.1	APCV	Cacipacore virus	N/A
MBFV	NC_034018.1	YAOV	Yaounde virus	N/A
MBFV	NC_033693.1	BOUV	Bouboui virus	N/A
MBFV	NC_030289.1	EHV	Edge Hill virus	N/A
MBFV	NC_033699.1	JUGV	Jugra virus	N/A
MBFV	NC_033697.1	SABV	Saboya virus	N/A
MBFV	NC_033698.1	UGSV	Uganda S virus	N/A
TBFV	NC_004355.1	ALKV	Alkhumra hemorrhagic fever virus	2
TBFV	NC_006947.1	KSIV	Karshi virus	2
TBFV	NC_003690.1	LGTV	Langat virus	3
TBFV	NC_001809.1	LIV	Louping ill virus	3
TBFV	NC_005062.1	OHFV	Omsk hemorrhagic fever virus	3
TBFV	NC_003687.1	POWV	Powassan virus	5
TBFV	NC_027709.1	SGEV	Spanish goat encephalitis virus	5
TBFV	NC_001672.1	TBEV	Tick-borne encephalitis virus	6
TBFV	NC_023424.1	TYUV	Tyuleniy virus	1
TBFV	NC_023439.1	KAMV	Kama virus	0
TBFV	NC_033721.1	MEAV	Meaban virus	N/A
TBFV	NC_033726.1	SREV	Saumarez Reef virus	N/A
TBFV	NC_033724.1	KADV	Kadam virus	N/A
TBFV	NC_033723.1	GGV	Gadgets Gully virus	N/A
ISFV	NC_012932.1	AEFV	Aedes flavivirus	10
ISFV	NC_001564.2	CFAG	Cell fusing agent virus	9
ISFV	NC_008604.2	CxFV	Culex flavivirus	10
ISFV	NC_012671.1	QBV	Quang Binh virus	7
ISFV	NC_005064.1	KRV	Kamiti River virus	13
ISFV	NC_027819.1	MECDV	Mercadeo virus	11
ISFV	NC_021069.1	MSFV	Mosquito flavivirus	9
ISFV	NC_034242.1	OCFVPT	Ochlerotatus caspius flavivirus	2
ISFV	NC_027817.1	PaRV	Parramatta River virus	12
ISFV	NC_030401.1	HANV	Hanko virus	N/A
ISFV	NC_033694.1	PCV	Palm Creek virus	N/A
NKV	NC_008718.1	ENTV	Entebbe bat virus	2
NKV	NC_027999.1	EPEV	Paraiso Escondido virus	2
NKV	NC_004119.1	MMLV	Montana myotis leukoencephalitis virus	9
NKV	NC_003635.1	MODV	Modoc virus	9
NKV	NC_005039.1	YOKV	Yokose virus	9
NKV	NC_026624.1	SOKV	Sokoluk virus	N/A
NKV	NC_003676.1	APOIV	Apoi virus	N/A
NKV	NC_026620.1	JUTV	Jutiapa virus	N/A
NKV	NC_029054.2	POTV	Potiskum virus	N/A
NKV	NC_034007.1	PPBV	Phnom Penh bat virus	N/A
NKV	NC_003675.1	RBV	Rio Bravo virus	N/A
NKV	NC_003996.1	TABV	Tamana bat virus	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). The number of Musashi binding elements (`UAG` trinucleotides) found in the 3'UTR sequences is listed in the last column. ^{a}MBEs found in the 3'UTR of SPONV SA-Ar strain. $^{N/A}$3'UTR partial or not available in the `refseq` data set.

	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.