Evolution of Flavivirus sfRNA

Roman Ochsenreiter[1], Andrea Tanzer[1], Ivo L. Hofacker[1,2], Michael T. Wolfinger[1,3*]

[1]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Wien, Austria
[2]Bioinformatics and Computational Biology Research Group, University of Vienna, Währingerstraße 17, 1090 Wien, Austria
[3]Center for Anatomy and Cell Biology, Medical Unversity of Vienna, Währingerstraße 13, 1090 Wien, Austria

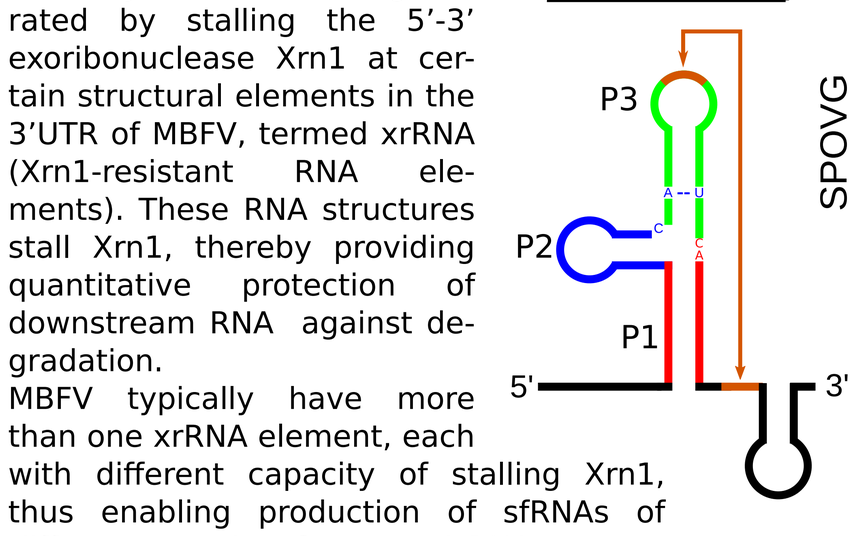
Contact: michael.wolfinger@univie.ac.at - http://www.tbi.univie.ac.at

1. Introduction

Arthropod-borne flaviviruses (FVs), including human pathogens such as Yellow fever virus (YFW), Japanese encephalitis virus (JEV) and Dengue virus (DENV) are a growing global health treat. The current outbreak of Zika virus (ZIKV) in Central and South America is supposedly linked to an increasing number of congenital microcephaly and Guillain-Barré syndromes. FV are small, single stranded positive-sense RNA viruses of 10-12kb length with highly structured untranslated regions (UTRs). The latter are associated with regulation of the viral life cycle, inducing processes such as genome circularization, viral replication, packaging, and modulating pathogenicity [1].
Here, we present a computational approach for identification and characterization of conserved structural elements in the UTRs of mosquito-borne flaviviruses (MBFV).

2. sfRNA / xrRNA

Upon FV infection, accumulation of stable long non-coding viral RNAs, termed subgenomic flaviviral RNAs (sfRNAs), which dysregulate cellular function with the aim of promoting viral infections is observed. Many MBFV produce sfRNAs by efficiently hijacking the host cell's mRNA degradation pathway. Mechanistically, sfRNAs are generated by stalling the 5'-3' exoribonuclease Xrn1 at certain structural elements in the 3'UTR of MBFV, termed xrRNA (Xrn1-resistant RNA elements). These RNA structures stall Xrn1, thereby providing quantitative protection of downstream RNA against degradation.
MBFV typically have more than one xrRNA element, each with different capacity of stalling Xrn1, thus enabling production of sfRNAs of different lengths. Conserved RNA structural elements in viral 3'UTRs have been identified in our group [4], some of which have later been attributed to xrRNA functionality, specifically stem-loop (SL) and dumbbell (DB) elements.



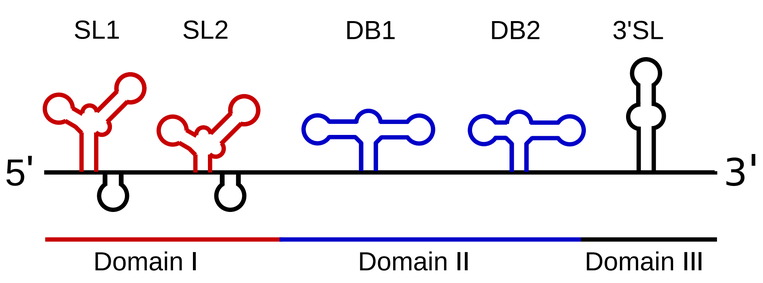


Fig. 1 Architecture of MBFV 3'UTRs. RNA elements attributed to halting Xrn1 are shown in red and blue. SL1 and SL2 correspond to previously described SLII and SLIV in West nile virus (WNV).

3. MBFV SL element classes

We have analysed the SL elements of MBFV with covariance models (CM) [6] and identified five unique families of SL structures. Initial structural alignments of SL elements gave strong evidence for heterogeneity even within single virus families. Consequently we built initial CMs for distinct SL1 and SL2 sequences and iteratively refined and extended the models by searching against MBFV 3'UTRs. Our CM classification revealed independent evolutionary conservation of SL hairpins YFVG, JEVG and DENVG.
To confirm our findings, we measured sequence similarity between these SL elements using a z-score approach [2], which has previously been applied in evolutionary studies of miRNAs. The resulting gene tree confirmed our CM analysis, and supported our hypothesis of independent deletion and duplication events in the evolution of MBFV SL elements.

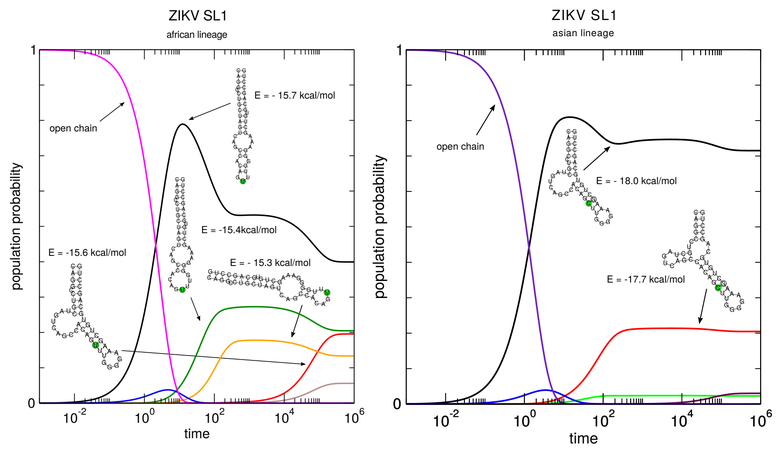


Fig. 3 Reduced folding kinetics based on numerical integration of a Markov process on the coarse grained energy landscapes [3] of african and asian lineage ZIKV SL1. Simulations are run from the open chain conformation to thermodynamic equilibrium. Each line corresponds to the temporal population of a macrostate, characterized by its minimum structure. The compensatory U to C mutation is highlighted in green.

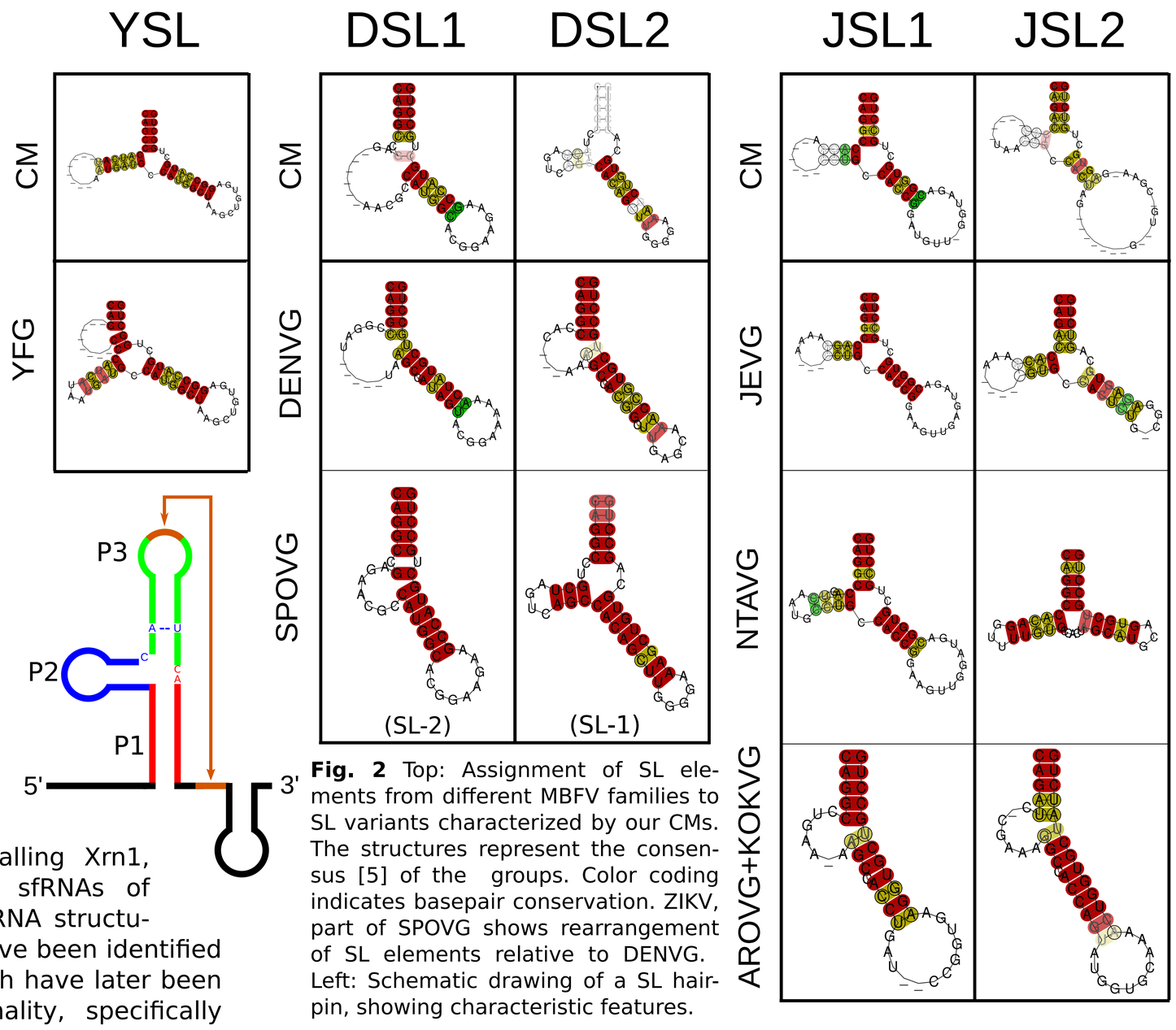


Fig. 2 Top: Assignment of SL elements from different MBFV families to SL variants characterized by our CMs. The structures represent the consensus [5] of the groups. Color coding indicates basepair conservation. ZIKV, part of SPOVG shows rearrangement of SL elements relative to DENVG. Left: Schematic drawing of a SL hairpin, showing characteristic features.

4. Case Study - Zika virus SL1

ZIKV, member of the Spondveni virus family is closely related to DENVG. We found strong evidence that SL elements in ZIKV were subject to a genomic reorganization, compared to other DENV viruses. Our analysis relates ZIKV SL1 to the DSL2 group, whereas ZIKV SL2 is more similar to DSL1.
A single compensatory U to C mutation in asian lineages of ZIKV critically influences the capacity of the RNA to fold into functional conformations. The less stable african lineage folds into an unfavourable stem-loop structure (Fig. 3). Contrary, the asian lineage folds predominantly into structures that are capable of building the three-way helix junction. The mutation results in thermodynamic stabilization of the functional form, thus allowing for efficient stalling of Xrn1 and possibly increased virulence.

5. Methods

Identification and characterization of SL elements was performed by two fundamentally independent methods:

The identification step utilized CMs of known SL elements. Extending classic Markov models, CMs can be understood as a variant of stochastic contetxt free grammars that allow for simultaneous identification of sequence and structure given a known element. The CM, as provided in the infernal toolsuite [6], also gives P-values, thus quantifying the credibility of a match.

Alternatively, we reconstructed the phylogenic history of SL elements via a z-scoring approach [2]. To identify similarities between SL elements of different families, we computed the significance of the alignment score as follows

$$z(I,J) = \frac{s(I,J) - m}{\sqrt{v}}$$

Where $s(I,J)$ is the identity score of 2 sequences in the alignment, m and $\sqrt{v}$ are mean and variance of identity scores, after repeated shuffling of both sequences.

6. Acknowledgements

This work was partly funded by the Austrian Science Fund FWF projects "RNA regulation of the transcriptome" (F43), "mRNAs von Viren: Evolution und Struktur-Funktionsbeziehungen" (FWF-I-1303) and the Austrian/French project "RNA-Lands" (FWF-I-1804-N28 and ANR-14-CE34-0011).

[1] Villordo, S.M., Carballeda, J.M.,Filomatori, C.V., Garmarnik, A.V. (2016), **RNA Structure Duplications and Flavivirus Host Adaptation**. *Trends Microbiol 24(4), 270-283.*
[2] Tanzer,A., Stadler, P.S. (2004), **Molecular evolution of a microRNA cluster.** J. Mol. Biol. 339:2, 327-35.
[3] Wolfinger, M.T., Svrcek-Seiler, A., Flamm, C., Hofacker, I.L., Stadler, P.S. (2004), E**fficient Computation of RNA Folding Dynamics**. J.Phys.A: Math.Gen. 37: 4731-4741
[4] Rauscher S., Flamm C., Mandl, C., Heinz, F., Stadler, P.F. (1997), **Secondary structure of the 3'-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities RNA**, 3(7):779–791
[5] Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L. (2011), **ViennaRNA Package 2.0**, Algorithms for Molecular Biology 6:1, 26
[6] Nawrocki, E.P., Eddy S.R., **Infernal 1.1: 100-fold faster RNA homology searches** (2013)**,** Bioinformatics, 29:2933-2935