

Analyzing the Branching Degree of RNA Viral Genomes: A Hepatitis C Case Study

Christine E. Heitsch ¹

Keywords: RNA secondary structure, branching degree, RNA viruses, Hepatitis C

1 Introduction.

Improved knowledge of RNA structure and function has many significant applications, including efforts to cure numerous RNA-related diseases. For instance, understanding the structural mechanism by which the RNA virus Hepatitis C controls the ribosome of an infected cell [6] could lead to new drug targets to treat a disease which alone kills up to 10,000 Americans annually. In this work, we present the preliminary results from our analysis of branching degrees in large RNA viral genomes, beginning with two examples of Hepatitis C secondary structures. The motivation for these results is our theoretical analysis of RNA folding which suggests that the locally favorable energies of branching loops are balanced by the energetic cost of increasing numbers of hairpin structures.

2 Theoretical Analysis of RNA Branching.

Like many RNA molecules, the overall structure and function of Hepatitis C viral genomes are significantly related to the base pairings of their secondary structure. Single-stranded RNA sequences are understood to self-bond with a complex interplay between energetically beneficial stacked base pairs and destabilizing single-stranded structures called “loops”. In calculating the minimal free energy, the predicted secondary structure will balance the trade-offs between minimizing loop energy costs while maximizing stacked base pairs. However, there has previously been only a very limited analysis of the trade-offs *within* the types of loops structures formed [1, 2, 3].

To begin, we observe an odd phenomenon regarding loops which contain $k \geq 3$ base pairs: the energy of these branching loops can be increasingly negative as a function of k . This is due to the favorable stacking interactions between the base pairs and adjacent single-stranded nucleotides in the loop, and also holds for the external loop (which has no closing base pair). On first inspection, this observation suggests that, energetically speaking, branching is locally favorable. But this contradicts the common understanding of loops as destabilizing elements in RNA secondary structures!

Using a combinatorial model for RNA sequence folding, we resolve this apparent contradiction by completely characterizing the minimal and maximal loop energy configurations. In this context, there are natural energy minima corresponding to structures which maximize the number of loops containing three base pairs. Intuitively, the locally favorable free energies associated with branching loops are offset by the cost of producing more energetically expensive hairpin structures. Concluding there is an energetic cost to the branching of RNA secondary structures, despite the apparent benefit indicated by the energy functions [7] and thermodynamic parameters [5], is consistent with observed RNA viral configurations [4].

¹Genome Center of Wisconsin and Mathematics Department, University of Wisconsin – Madison, Madison, WI, 53706-1580 E-mail: heitsch@math.wisc.edu

3 Preliminary Hepatitis C Branching Results.

To explore the validity of our theoretical conclusions, we computationally analyzed the predicted secondary structures of two Hepatitis C viral genomes: the China virus complete genome (HCV-gen, 9400 bases, Genbank Accession No. L02836) and individual isolate HC-J1 (HCV-j1, 9502 bases, No. D10749). Table 1 summarizes the breakdown of our preliminary branching degree results. As expected, in both cases the majority of branching loops contain $k = 3$ base pairs, and the numbers decrease rapidly with increasing k .

$k =$	1	2	3	4	5	6	7	8	9	10	...	31
HCV-gen	174	389	55	22	10	6	1	1*	0	1	...	0
HCV-j1	154	425	55	19	7	1	1	0	0	0	...	1*

Table 1: Number of loops containing k base pairs (* denotes the external loop).

We are currently investigating the correlation of loops with high branching and motifs with functional significance. A correlation could be interpreted as evidence of evolutionary pressure overcoming the tendency towards low degree branching and would be consistent with the fact that the viral backbone, which is known to be relatively undifferentiated, contains large numbers of loops with little branching. We expect this and other related work to advance understanding and applications by allowing us to distinguish RNA viral substructures whose base pairings encode significant structural and functional information.

4 Acknowledgments.

The author gratefully acknowledges the Hepatitis C Secondary Structure Prediction data made publicly available² by Prof. Ann C. Palmenberg³ and Dr. Jean-Yves Sgro⁴. This research of Christine E. Heitsch, Ph.D., is supported in full by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

References

- [1] W. Fontana, D. Konings, P. F. Stadler, and P. Schuster. Statistics of RNA secondary structures. *Biopolymers*, 33(9):1389–404, Sept 1993.
- [2] H. H. Gan, S. Pasquali, and T. Schlick. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucl. Acids. Res.*, 31(11):2926–43, June 2003.
- [3] T. Liu and R. Bundschuh. Analytical description of finite size effects for RNA secondary structures. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69, 2004.
- [4] A. C. Palmenberg and J.-Y. Sgro. Topological organization of picornaviral genomes: Statistical prediction of RNA structural signals. *S. Virol.*, 8:231–241, 1997.
- [5] M. Serra, D. Turner, and S. Freier. Predicting thermodynamic properties of RNA. *Meth. Enzymol.*, 259:243–261, 1995.
- [6] C. M. T. Spahn et al. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science*, 291(5510):1959–1962, March 9 2001.
- [7] M. Zuker, D. Mathews, and D. Turner. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In J. Barciszewski and B. Clark, editors, *RNA Biochemistry and Biotechnology*, NATO ASI Series, pages 11–43. Kluwer Academic Publishers, 1999.

²Online at http://www.virology.wisc.edu/acp/RNAFolds/RNAFolds_hep.html

³Dept. of Biochemistry & Institute of Molecular Virology, University of Wisconsin – Madison

⁴Biotechnology Center & Institute of Molecular Virology, University of Wisconsin – Madison