Visualizing RNA Secondary Structure Base Pair Probabilities

William K. Jannen and Daniel P. Aalberts

Abstract— RNAbow diagrams are a versatile tool for visualizing and comparing ensembles of RNA secondary structures. Previously, RNAbows have proved useful when investigating individual ensembles of folds, performing cluster analysis, and identifying conformational changes caused by single nucleotide polymorphisms [Aalberts and Jannen, RNA, 19, 475–478 (2013)]. This contest submission highlights their usefulness in (1) making visual comparisons of the Minimum Free Energy structure to the full structural ensemble and (2) understanding of the effects of mutations on RNA folding stability.

Index Terms—RNA secondary structure, Visualization, Partition function, Mutations

1 Introduction

As authors of visualization tools, we make editorial decisions that — both implicitly and explicitly — attach relative importance to features in the data. Our challenge is to create visualizations that accurately represent physical reality and provide insight.

An RNA secondary structure identifies which bases form pairs, and which unpaired bases form hairpin or internal or multi-branch loops. The Minimum Free Energy (MFE) state is the most probable secondary structure. The MFE or any other single state can be depicted effectively with airport diagrams, bracket notation, and rainbow diagrams. The strengths of these representations — cleanliness and clarity — make them both popular and dangerous. Their use suggests a one-to-one mapping between sequence and structure. However, thermodynamic equilibrium samples many structures.

The partition function is a weighted average of all of the structures at a specified temperature [4]. Encoding the partition function probabilities as a heat-map atop an airport diagram's structure is a recent improvement, but the heat map measures the certainty of the MFE structure rather than suggesting the reality of thermal fluctuations among multiple structures.

There are fewer methods to visualize ensembles of states. Dot plots have often been used to display the partition function's base pairing information. Dot plots compactly represent the probabilities P_{ij} of pairing base i with base j. For coexisting multiple structural classes, however, dot plots often become puzzling. It is difficult to pick out compatible structures within an ensemble, and comparing structures across ensembles requires the viewer to translate between matrices or to reflect across the main diagonal.

Our RNAbow diagrams [1] approach to visualizing RNA secondary structure combines the intuitive qualities of rainbow diagrams with the information density of dot plots to encode the entire partition function in an easily-digestible graph. Further, RNAbows use color, weight, and brightness, along with vertical juxtaposition, to ease the comparison of different ensembles or clusters.

2 RNAbow Diagrams

The rainbow (or arc) diagram graphically represents a single secondary structure state. In a rainbow diagram, bases in the primary sequence form nodes along the graph axis, and an edge between base i and base j represents a bond.

- *William K. Jannen is at Stony Brook University. E-mail: wjannen@cs.stonybrook.edu*
- *Daniel P. Aalberts is Prof. of Physics at Williams College. E-mail: aalberts@williams.edu*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

The RNAbow diagram [1] is a generalization which represents an ensemble of states, using edge thickness and darkness to represent pair probabilities.

Our RNAbows webserver (http://rnabows.com) permits users to select from ViennaRNA, UNAFold, or RNAstructure to compute the partiton function [2, 3, 5] and provides several tools:

AllPairs is a generalization of the rainbow diagram and represents base pair probabilities with line thickness and darkness. *AllPairs* is a drop-in replacement for the dot plot. Because our visual processing system naturally groups parallel lines, the *AllPairs* method makes the compatible stems of multiple structures easy to identify.

Difference RNAbows facilitate the comparison of the folds of different ensembles in just a glance by coloring the regions of difference. *Difference RNAbows* are useful when comparing macrostates (Figure 1) and analyzing the effects of mutations (Figure 3).

Cluster RNAbows allow users to split the *AllPairs* partition function to reveal two sub-partition functions (or "macrostates") describing local minima. Figure 1 is an example.

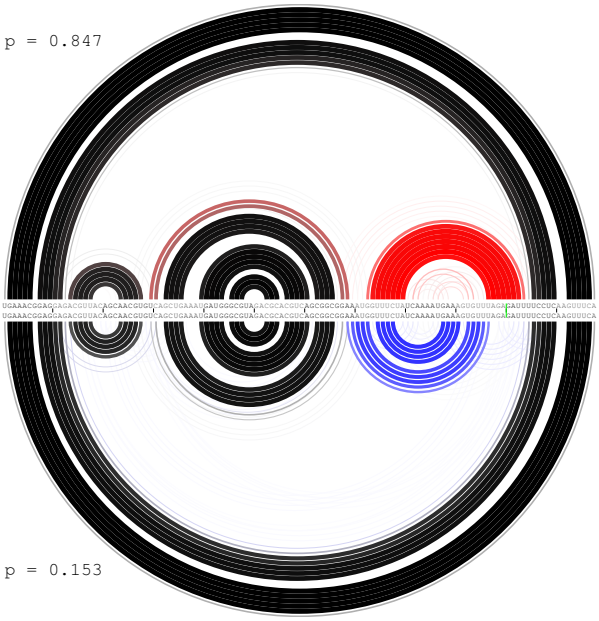

Fig. 1. This *Cluster* RNAbow shows two macrostates within the HAR1 ncRNA Human sequence, along with their relative probabilities. Base pairs unique to either cluster are colored proportional to their probability difference. Common pairs are in black.

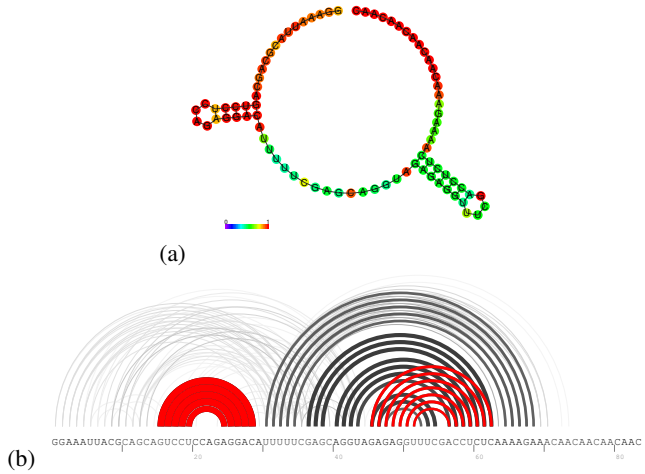

Fig. 2. (a) An "MFE structure drawing encoding base-pair probabilities" image from ViennaRNA [2]. It is clear that half of this structure is uncertain based on the partition function information, but the diagram does not identify why. (b) Our *AllPairsMFE RNAbow* diagram colors pairs from the MFE structure in red and depicts the other pairs in black. Line thickness and darkness are proportional to P_{ij}. The *AllPairsMFE RNAbow* diagram immediately reveals the second local minimum structure.

For this contest submission, we introduce an additional tool:

AllPairsMFE enhances the *AllPairs* representation by signifying the pairs of the MFE structure in red. See Figure 2(b). This tool is a replacement of the heat-map airport diagram, Figure 2(a). *AllPairsMFE* presents the MFE structure, but by also depicting competing structures, it more clearly identifies the sources of structural uncertainty.

3 Contest Challenge 1: Visualizing Uncertainty

Focus on MFE representations may mislead researchers into picturing RNA as a single structure, and not as a thermally fluctuating ensemble of structures. One first step towards visualizing the uncertainty of this approximation is to overlay pairs of the MFE structure with partition function information as a heat map. This approach is seen in Figure 2(a), a visualization from the ViennaRNA package [2]. This diagram makes clear that half of the predictions are fairly certain, but the other half are not. The cause of this uncertainty is not shown.

The *AllPairsMFE* RNAbow for the same sequence, shown in Figure 2(b), reveals the cause of this uncertainty to be a competing structure. The full thermal ensemble of folds is represented clearly, with the pairs of a particular structure (here the MFE, though it could be the consensus or any suboptimal structure) highlighted by color. We see the uncertainty of the pairs *within* the MFE structure, as well as the uncertainty from competing structures.

This example further highlights the dangers of single-state-centric representations. With MFOLD [6] we find $G_{\text{MFE}} = -18.8$ kcal/mol and a second local minimum $G_{MFE2} = -17.7$ kcal/mol. After including the entropic weight, $p_{\text{MFE}} = 0.41$ and $p_{MFE2} = 0.59$. In other words, the entropic weight of the second macrostate makes it more probable than the MFE's macrostate. *Cluster RNAbows* is our tool to resolve macrostates, Figure 1 shows the two Human HAR1 ncRNA macrostates and their probabilities.

4 Challenge 2: Visualizing Sequence Evolution

We investigated the HAR1 ncRNA sequences from Chimp and Human. The *Difference RNAbow* allows us to juxtapose the folds, and reports the folding free energy (here, as calculated by UNAFold [3]). We observe that mutations stabilize the Human sequence's folding free energy by $\Delta G = -11.6$ kcal/mol.

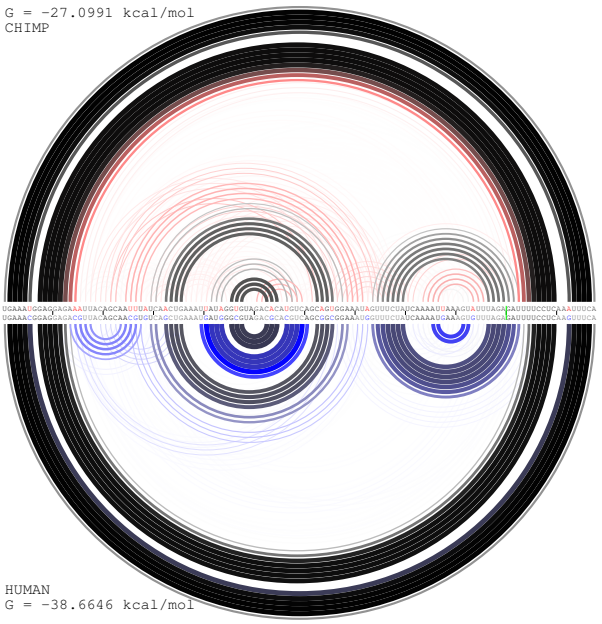

Fig. 3. A *Difference RNAbow* highlights the differences in the RNA secondary structures of the Chimp and Human HAR1 ncRNA sequences. Base pairs unique to either structure are colored proportional to the probability difference. Base substitutions are also color coded. The mutations substantially increase the RNA folding stability of the Human fold relative to the Chimp.

The *Difference RNAbows* tool not only allows for easy comparisons of two known sequences, but suggests where mutations would most influence the fold. Mutations that extend stems would stabilize the fold, while mutations that disrupt stems would destabilize the fold. For example, in Figure 3, the U41G mutation stabilizes the stem and permits the U40-A61 pair in the Human.

5 Conclusions

The *AllPairsMFE* RNAbow of Figure 2(b) is an intuitive visual representation of the uncertainty of RNA secondary structures — not just of pairs within single structures, but of entire structural classes within an ensemble. The *Cluster RNAbow* of Figure 1 isolates structural classes and shows the uncertainty within each.

Our *Difference RNAbow* diagram of Figure 3 facilitates comparisons of the structures of related sequences.

Acknowledgments

This work was supported by the National Institutes of Health [R15GM106372 to DPA]. Earlier development of RNAbows was supported by the National Institutes of Health [GM080690 to DPA] and the National Science Foundation [MCB-0641995 to DPA].

References

[1] D. P. Aalberts and W. K. Jannen. Visualizing RNA base-pairing probabilities with RNAbow diagrams. *RNA (New York, N.Y.)*, 19(4):475–8, 2013.

[2] R. Lorenz, S. H. Bernhart, C. H. Z. Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker. ViennaRNA package 2.0. *Algorithms for Molecular Biology*, 6.

[3] N. R. Markham and M. Zuker. *UNAFold - Software for nucleic acid folding and hybridization*, volume 453 of *Methods in Molecular Biology*, pages 3–31.

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

[5] J. S. Reuter and D. H. Mathews. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11.

[6] M. Zuker. On finding all suboptimal foldings of an rna molecule. *Science*, 244(4900):48–52, 1989.