jViz.Rna - An Interactive Graphical Tool for Visualizing RNA Secondary Structure Including Pseudoknots

Kay C. Wiese, Senior Member, IEEE and Edward Glen, Student Member, IEEE
Simon Fraser University
School of Computing Science
13450 102nd Avenue, Surrey B.C., V3T 5X3, Canada
wiese@cs.sfu.ca, eglen@cs.sfu.ca

Abstract

In order for structure prediction researchers to better understand the results of their algorithms and to enable life science researchers to interpret RNA structure easily, it is helpful to provide them with a flexible and powerful tool for RNA secondary structure visualization. jViz.Rna *is a multi-platform visualization tool capable of displaying RNA secondary structures encoded in a variety of file formats. A single structure can be shown using the linear Feynman, circular Feynman, dot plot, and classical structure visualization models. The resulting drawings are dynamic and can easily be further modified by the user. Any of the drawings produced can be saved to disk enabling easy dissemination. The unique usage of a spring model for classical structure drawing allows for clear visualization of pseudoknots with minimal overlaps. The addition of a locality tool allows for the isolation of pseudoknotted regions or other regions of interest.*

Availability: *http://jviz.research.iat.sfu.ca*

1. Introduction

Due to the importance of RNA to protein synthesis, there are extensive efforts to better understand its function. Part of this effort is focused on providing a means whereby the shape of these biomolecules can be predicted through the use of computer algorithms. Dynamic Programming [14] and Genetic Algorithms [11, 13], take a sequence of bases and attempt to find a secondary structure using free energy minimization as a guide. Comparative methods, on the other hand, use known sequence and folding patterns to generate a predicted secondary structure [6].

Visualization of the results of these algorithms can help researchers to better identify the strengths and weaknesses of their prediction methods as well as analyze the function of the biomolecule. There exists a number of tools which provide diverse capabilities, each with its own benefits and drawbacks. *jViz.Rna* combines a number of useful features into a single package. It currently supports a number of common file input and output formats and multiple dynamic visualization methods which allow straightforward, direct manipulation by the user. This ease of modification of displayed structures makes *jViz.Rna* a useful tool in a multiuser collaborative environment.

A key feature of *jViz.Rna* which has not been discussed in [12] is its ability to display pseudoknots in all of its visualization methods. Pseudoknots are complex interactions between structural elements of RNA for which there is little visualization capability in the majority of existing tools. A pseudoknot exists between two base pairs (i, j) and (i', j') if $(i < i' < j < j')$. This paper discusses how pseudoknots are displayed in all of the visualization methods supported by *jViz.Rna* and compares this functionality to that of other tools. The discussion focuses particularly on pseudoknot visualization in classical structure plots. This is a challenging problem and only supported by very few tools.

An overview of *jViz.Rna* is provided in Section 2 with a brief discussion of supported file formats and visualization modes in Sections 3 and 4. A detailed description of pseudoknot handling and multiuser interaction with the classical structure mode is given in Sections 5 and 6. Conclusions are given in Section 7.

2. Overview

jViz.Rna is a flexible tool intended to visualize RNA secondary structures using a variety of visualization techniques. As the tool is written in Java, it can be used on any platform with a compatible Java runtime environment (J2SE 1.4.2 or greater). It has been successfully used on Linux, MacOSX, and Windows platforms. The default usage is through a graphical interface, however, command-

Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)
0-7695-2517-1/06 $20.00 © 2006 IEEE

Table 1. Commonly available tools and their ability to display pseudoknots

	Classical Structure	Circular Feynman	Linear Feynman	Dot Plot
jViz.Rna	*	*	*	*
MFold		*		*
PseudoViewer	*			
RnaFold				
RnaViz	*			
XRna				

jViz.Rna [12], MFold [14], PsuedoViewer [4], RnaFold [5], RnaViz [8], XRna [7]

line options are available for most functions, enabling usage in a batch-processing environment. Full documentation of available features is provided with the tool.

A detailed description of an earlier version of *jViz.Rna*, and a comparison with existing visualization tools is provided in [12]. The work discussed here presents the current incarnation of *jViz.Rna* which allows for easy visualization of pseudoknots and pseudoknotted regions in the classical structure plot (a significant feature not included in [12]). A web site for download and additional information is available now and we anticipate that this will aid the dissemination of *jViz.Rna* amongst interested users. Also, a new GUI has been developed which improves the usability of *jViz.Rna*. This includes large display touchscreen visualization for multiple users.

3. File Formats

jViz.Rna can read three different types of files containing structural information: Connectivity Table (.ct) [14], Base Pair Sequence (.bpseq) [2], and an implementation of Dot Bracket Notation (.dbn) [4]. All visualizations can be saved to disk using either the Encapsulated Postscript (.eps) or Portable Network Graphics (.png) file formats, suitable for usage in publications and presentations. The popular .gif and .jpeg formats are not included due to licensing restrictions for the necessary libraries. As *jViz.Rna* is designed to be highly extensible, it is very easy to add new file formats as both input and output file types. Future work in this area includes support for the RNAML [10] file format.

The *Connectivity Table (.ct)* file format has many different, unpublished specifications. It has been used in a variety of programs, each one implementing its usage in a slightly different manner. This format is most commonly found as the output of the popular MFOLD web server [14]. As we were unable to determine a precise and universal specification for this format, we have based our implementation on a number of sample files given by the MFOLD web server.

The *Base Pair Sequence (.bpseq)* file format is designed in much the same way as the .ct format. Its simplicity stems from its restriction to having only a single structure per file as opposed to the .ct format which may contain multiple structures per file. This allows the .bpseq format to store less data than the .ct format. Due to its simplicity it is a consistent and stable format.

Dot Bracket Notation is very simple and can vary widely in implementation. Our implementation supports the 'Bracket View 1' format used by Pseudoviewer [4] in order to provide compatibility and testing with that application. It is a very obvious format in that by itself it is a simple visualization allowing for rapid identification of helices and pseudoknots.

4. Visualization Modes

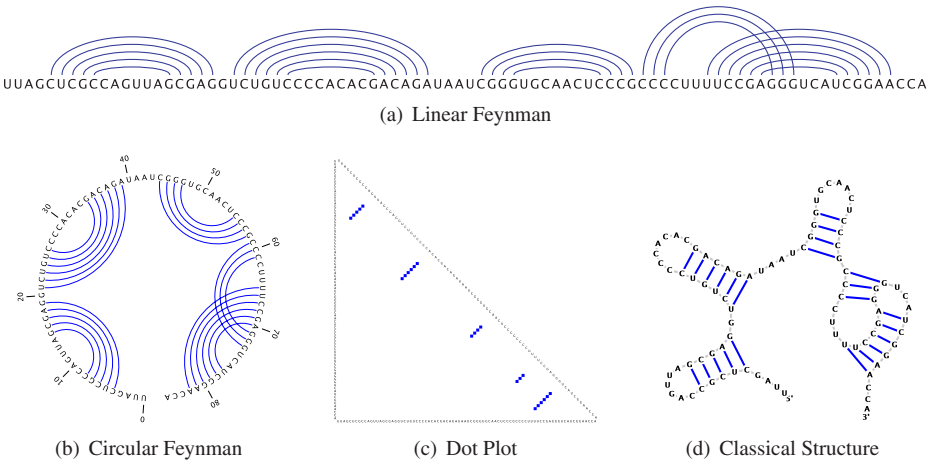
The secondary structure of RNA can become fairly complex as the quantity of nucleotides grows. Thus, several models of secondary structure visualization have been devised, each suited for a different type of analysis.

jViz.Rna provides the user with a choice of multiple visualization models. Given an RNA secondary structure, drawings can be created in linear Feynman, circular Feynman, dot plot, and classical structure formats as shown in Fig. 1. This gives the user the ability to select the most appropriate method for viewing the structural aspects they are interested in. Each view has many options available, allowing the user to dynamically adjust the visualization to suit the structure being shown or their own preferences. All of the models are also capable of displaying pseudoknots. The support for pseudoknots in other commonly used tools is summarized in Tab. 1.

The *Linear Feynman Diagram*, Fig. 1(a), shows the entire RNA strand on a horizontal line with base pairs represented by arcs above the structure. This graph can show structures of varying complexity but with larger structures it can rapidly grow quite wide and become difficult to view. Pseudoknots can be identified with ease in Feynman diagrams where an overlapping set of arcs indicates the pseudoknotted bonds.

The *Circular Feynman Diagram*, Fig. 1(b), is a variant of the linear Feynman where, instead of having the RNA strand stretched in a horizontal line, it is looped into a circle. This representation has all the benefits of the linked graph, but without the problem of rapidly growing too wide. It is therefore a good visualization technique for reviewing global interactions of a large RNA molecule as well as rapidly identifying pseudoknots.

The *Dot Plot*, Fig. 1(c), produces a graph where the RNA strand forms both the x and y axes. Thus, a point (i, j) on the grid corresponds to a bond formed between the i^{th} and j^{th} nucleotides. The structure is symmetrical across its



(a) Linear Feynman

(b) Circular Feynman (c) Dot Plot (d) Classical Structure

Figure 1. *jViz.Rna* **views of the pseudoknotted portion of** *Turnip Yellow Mosaic Virus*, **ssRNA, M58309, 86nt**

diagonal, so only the bottom triangle under the diagonal is shown. Although the depiction is not directly intuitive to the folding, with some experience, it is possible to identify structural elements (such as pseudoknots) in large, complex structures.

The *Classical Structure* [1], Fig. 1(d), illustrates the RNA strand such that bases which are paired are spatially placed near each other and connected with lines. The backbone structure is also shown with solid lines. Although this clearly displays structural elements, it can become complex when dealing with large structures. *jViz.Rna* uses a unique method for drawing and interacting with the classical structure. This is discussed in greater detail in the following section.

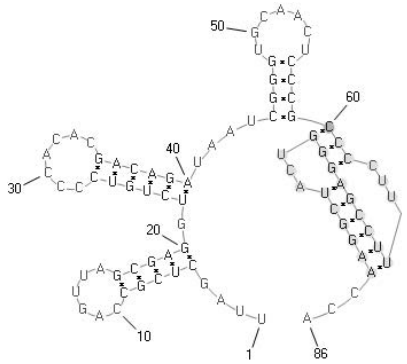


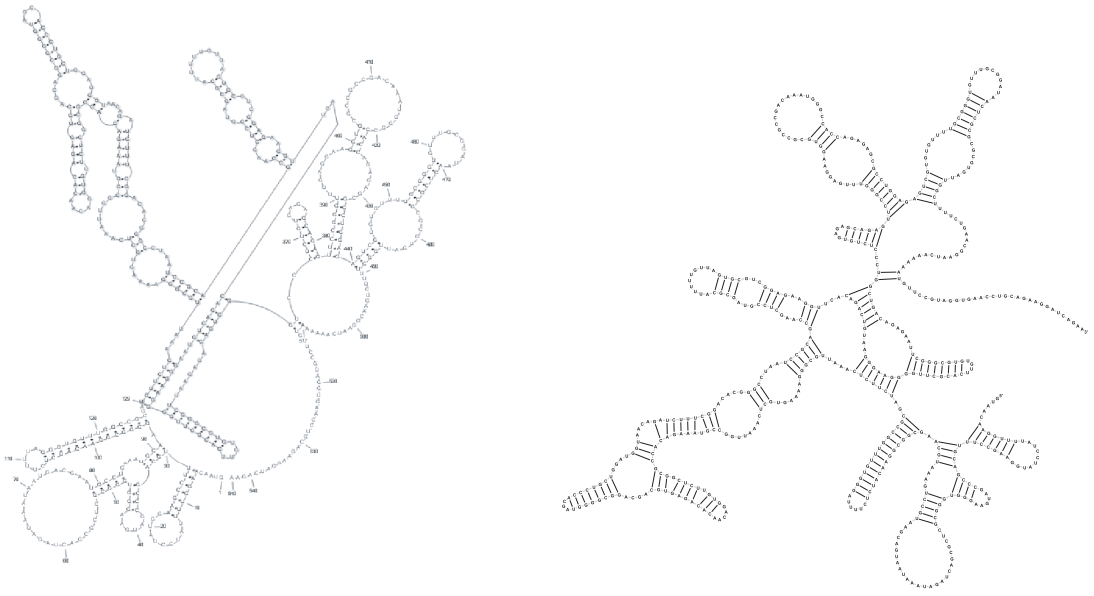
Figure 2. Pseudoviewer classical structure view of the pseudoknotted portion of *Turnip Yellow Mosiac Virus*, **ssRNA, M58309, 86nt**

5. Pseudoknots in Classical Structure

First discussed in [9], pseudoknots are tertiary structural elements formed by the interaction of bases of a single-stranded loop pair and complementary bases outside of the loop. While not prevalent in all RNA molecules, pseudoknots do have significant functional effects when present [3]. It is thus important that when these structural elements are found, they can be visualized in the same manner as non-pseudoknotted structures.

Providing pseudoknot capability in the classical structure model is more difficult than in the other previously discussed methods. This is due to the complex interactions that can occur between helices in a pseudoknotted structure. Typical classical structure layout algorithms internally represent RNA secondary structural elements as nodes in a tree. When pseudoknots are added to the structure, inner cycles can occur within the pseudoknot and outer cycles can occur between the pseudoknot and other structural elements. Due to the scarcity and difficulty in displaying pseudoknots, many applications do not allow for visualization of pseudoknotted structures when using the classical structure visualization model.

There are two notable exceptions in this area with RnaViz [8] and PseudoViewer [4] both being capable of displaying pseudoknots in classical structure plots (Tab. 1). While RnaViz is capable of displaying pseudoknots, initial classical structure plots containing pseudoknots in RnaViz must be manipulated by the user for easy interpretation. This manipulation tends to disconnect the structural elements, affecting the readability of the final drawing. Pseu-



(a) PseudoViewer (b) jViz.Rna

Figure 3. *jViz.Rna* **and PseudoViewer classic structure view of pseudoknotted structure of** *Hildenbrandia Rubra* **Group I intron, 16S rRNA, L19345, 543 nt**

doViewer has been built specifically for visualization of pseudoknots. Typically, the initial layout is reasonable, however, user modification is beneficial to create a well-presented structure. This user interaction does not disconnect structural elements as can occur in RnaViz. Another key difference is that PseudoViewer presents pseudoknots as planar graphs (Fig. 2). These pseudoknotted regions are not modifiable by the user and can become cumbersome in some structures. Note the artificially long lines used to represent the pseudoknot in Fig. 3(a). Examples of *jViz.Rna* displaying pseudoknots are given in Fig. 1(d) and Fig. 3(b).

The layout techniques used by PseudoViewer and *jViz.Rna* are significantly different with PseudoViewer applying a deterministic, algorithmic approach whereas *jViz.Rna* makes use of a non-deterministic spring model. In the spring model, each nucleotide becomes a vertex n_i $(1 \leq i \leq n)$ and the backbone is formed using edges $b_i = (n_i, n_{i+1})$ $(1 \leq i \leq n-1)$ to connect each node sequentially. Once the backbone is created, the hydrogen bonds are added as edges $h_j = (n_k, n_l)$ $(1 \leq k < l \leq n)$ with n_k and n_l forming a base pair in the structure which pulls the paired vertices (nucleotides) together, forming the final structure. Each node has a small repulsive force which helps to keep it distanced from the other nodes. The edges act as springs, pulling the connected nodes together with a user-selected amount of force. This results in an elastic classical structure which the user can then stretch and adjust into the preferred layout. By relying on the underlying spring model to shape the structure, *jViz.Rna* is able to visualize pseudoknots with ease. As well, all elements of the structure can be moved in a straightforward manner without disconnection of structural elements. An example of these differences can be seen in Fig. 3. With *jViz.Rna* (Fig. 3(b)) the structure is presented without overlaps or artificially long lines related to the pseudoknotted region and the pseudoknot is displayed with the same proportion as the rest of the structure.

5.1. Locality

Another benefit of the spring model is the ability to display only selected portions of the structure. This allows for the isolation of pseudoknotted regions or other regions of interest. *jViz.Rna* is equipped with a locality tool which limits the visibility of vertices at a specified distance d from the selected vertex n_s. A vertex n_i is displayed if the shortest path from n_s to n_i is less than or equal to d. For this calculation, the shortest path can include both base pair edges and backbone edges. In the case of larger structures, it can be beneficial to limit the viewed nucleotides to just the region of interest as shown in Fig. 4 where only the pseudoknotted region of L19345 is shown. This can also be beneficial for

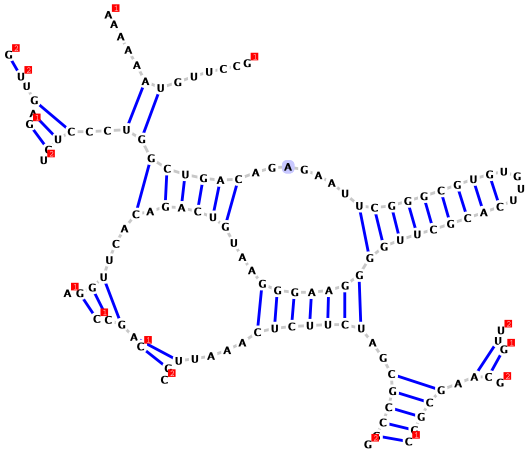


Figure 4. *jViz.Rna* **classic structure view of pseudoknotted portion of** *Hildenbrandia Rubra* **Group I intron, 16S rRNA, L19345, 543 nt. The shaded vertex indicates** n_s **with** $d = 16$

viewing structures in which there are complex interactions of which only a small portion is of interest.

6. Interaction

The dynamic nature of the spring model used to shape the classical structure plot lends itself well to investigation through interaction. While the spring model is effective at minimizing the number of overlaps in the final structure, there are situations in which overlap can be difficult to eliminate automatically. It is in these situations that the interactive capability of *jViz.Rna* is useful. In order to manually remove overlap, the user can drag a nucleotide to the desired location while the spring model ensures that the rest of the structure conforms to the new layout. This action essentially "untwists" the structure and removes the overlap.

Should there still be some areas with a complex layout, it is possible to actively shift the structure such that the motion yields a pseudo-3D view. Through this motion, certain nucleotides move behind other nucleotides and in fact can be shifted in such a way as to enable viewing of the complex region. While this cannot be captured in a still image, it is useful in explaining complex interactions in a multiuser environment.

Due to the straightforward method of interacting with *jViz.Rna*, it is well suited to a multiuser environment where any participant can interact with the visualization to describe a property of the structure as shown in Fig. 5. Simple touch-screen interaction allows the users to modify the structure intuitively.

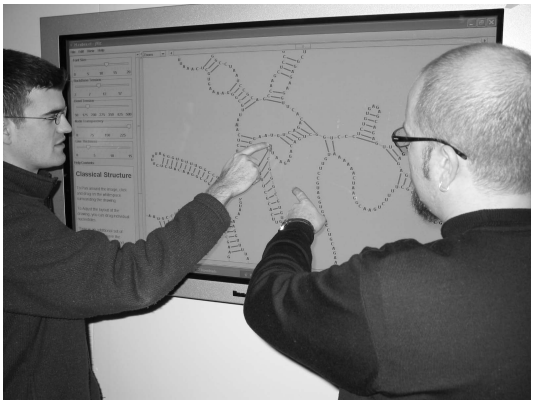


Figure 5. Interacting with *jViz.Rna* **on a plasma touch screen display.**

7. Conclusions

This paper presents several new features of *jViz.Rna*, a software tool for the visualization of RNA secondary structure. Specifically, the visualization of pseudoknots in the classical structure and the locality tool are significant additions. *jViz.Rna* has a number of advantages including 1) platform independence, 2) support for a wide range of common input and output file formats, 3) availability of different visualization methods, 4) dynamic output that can be further manipulated by the user, 5) the possibility to create high resolution static images of structures for dissemination, and 6) the ability to intuitively display pseudoknots in the classical structure plot without overlapping edges or artificially long edges. The provided locality tool allows for the isolation of pseudoknotted regions or other regions of interest. The dynamic nature of the spring model allows multiple users to intuitively interact with the structure to achieve a better understanding of the characteristics of the structure or to perform tasks such as the isolation of regions of interest.

References

[1] R. E. Bruccoleri and G. Heinrich. An improved algorithm for nucleic acid secondary structure display. *Bioinformatics*, 4:167–173, 1988.

[2] J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Müller, N. Pande, Z. Shang, N. Yu, and R. R. Gutell. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3:2, 2002.

[3] B. Deiman and C. W.A.Pleij. Pseudoknots: A vital feature in viral RNA. *Seminars in Virology*, 8:166–175, 1997.

[4] K. Han and Y. Byun. Pseudoviewer 2: visualization of RNA pseudoknots of any type. *Nucleic Acids Research*, 31(13), 2003.

[5] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh.Chem.*, 125:167–188, 1994.

[6] F. Major and R. Griffey. Computational methods for RNA structure determination. *Current Opinion in Structural Biology*, 11:282–286, 2001.

[7] U. of California Santa Cruz. XRNA. Website. `http://rna.ucsc.edu/rnacenter/xrna/xrna.html`.

[8] P. D. Rijk, J. Wuyts, and R. D. Wachter. RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics*, 19(2):299–300, August 2003.

[9] C. W.A.Pleij, K. Rietveld, and L. Bosch. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Research*, 13:1717–1731, 1985.

[10] A. Waugh, P. Gendron, R. Altman, J. W. Brown, D. Case, D. Gautheret, S. C. Harvey, N. Leontis, J. Westbrook, E. Westhof, M. Zuker, and F. Major. RNAml: A standard syntax for exchanging RNA information. *RNA*, 8:707–717, 2002.

[11] K. C. Wiese and E. Glen. A permutation-based genetic algorithm for the RNA folding problem: a critical look at selection strategies, crossover operators, and representation issues. *BioSystems - Special Issue on Computational Intelligence in Bioinformatics*, 72:29–41, 2003.

[12] K. C. Wiese and E. Glen. jViz.Rna - A Java tool for RNA secondary structure visualization. *IEEE Transactions on NanoBioscience*, 4(3):212–218, September 2005.

[13] K. C. Wiese and A. Hendriks. Comparison of P-RnaPredict and mfold - algorithms for RNA secondary structure prediction. *Bioinformatics*, 22(8):934–942, April 2006.

[14] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406 – 3415, 2003.