

VARNA: Interactive drawing and editing of the RNA secondary structure

Kévin Darty¹, Alain Denise^{1,2} and Yann Ponty^{1,3*}

¹ LRI, UMR CNRS 8623, Université Paris-Sud 11 F91405 Orsay cedex, France

² IGM, UMR CNRS 8621, Université Paris-Sud 11 F91405 Orsay cedex, France

³ Génomique Analytique, INSERM U511, 75013 Paris, France

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

ABSTRACT

Description: VARNA is a tool for the automated drawing, visualization and annotation of the secondary structure of RNA, designed as a companion software for web servers and databases.

Features: VARNA implements four drawing algorithms, supports input/output using the classic formats dbn, ct, bpseq and RNAML and exports the drawing as five picture formats, either pixel-based (JPEG, PNG) or vector-based (SVG, EPS and XFIG). It also allows manual modification and structural annotation of the resulting drawing using either an interactive *point and click* approach, within a web server or through command-line arguments.

Availability: VARNA is a free software, released under the terms of the GPLv3.0 license and available at <http://varna.lri.fr>

Contact: ponty@lri.fr

1 INTRODUCTION

With the increasing interest in structure-based methods for the analysis of RNA, many web servers now provide tools ranging from the detection of non-coding RNAs (Washietl *et al.*, 2005) to the rational design of small interfering RNAs (Ding *et al.*, 2004). In parallel, databases listing RNA secondary structures associated with specific functions (Mokrejs *et al.*, 2006), or obtained through specific methods (Cannone *et al.*, 2002), are increasing rapidly in number.

Various tools have been proposed for the clear visualization of the results of such methods, and to produce publication-quality pictures of RNA. However, most available applications either rely on a specific OS and/or third-party libraries, or do not allow any user interaction (command-line tools). A visualization application is therefore needed that: can be easily accessed from within a web server to display its results; operates as a standalone application; is platform-independent and free of external dependencies; enables user-interaction; and exports publication-quality pictures of edited/displayed RNA structures.

Thus, we developed VARNA (Visualization Applet for RNA), a JAVA-based system implementing classic RNA drawing algorithms coupled with a set of edit/export/annotate features, allowing publication-quality pictures to be created and displayed.

*to whom correspondence should be addressed

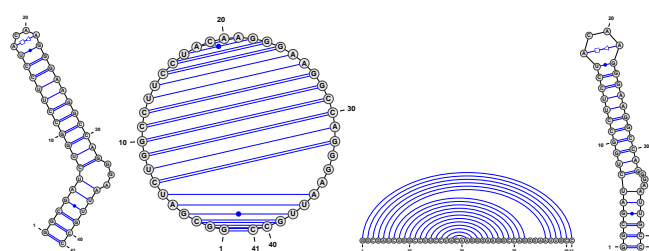


Fig. 1. Four representations (*Radial*, *Circular*, *Linear* and *NAView*) of the GAG-Pol -1 frameshift-inducing element in HIV-1 (PDBID:1ZC5) as rendered by VARNA under default settings.

2 MAIN FEATURES

VARNA can be used in three different ways: As a standalone JAVA application; as an applet featured on an HTML page, fully accessible through the HTML `param` applet option; or as a component that can be included in any JAVA software requiring a surface to display and annotate the secondary structure of RNA. A fully-documented API is provided for interfacing our software from within existing software.

2.1 Input

VARNA accepts most file formats traditionally used for the secondary structure of RNA. These include Vienna RNA dot-bracket notation (Washietl *et al.*, 2005) (.dbn), Mfold connect (Markham and Zuker, 2005) (.ct), Gutell's CRW format (Cannone *et al.*, 2002) (.bpseq) and the unifying RNAML (Waugh *et al.*, 2002). With the exception of RNAML, only accepted for input, each other format is supported both for input and output, allowing for file format conversions. Additionally, the software can process and display structures featuring non-canonical base-pairs (RNAML) and/or pseudoknots.

2.2 Output/Export

VARNA can export the resulting drawing to a variety of graphic formats, both pixel-based – Portable Network Graphics (PNG) or JPEG – and vector-based – Encapsulated PostScript (EPS), Scalable Vector Graphics (SVG) and XFIG. Although the resolution and compression level can be specified for pixel formats, vector

formats should be chosen for the production of publication-quality drawings, as they allow further editing with third-party vector-graphics software without any loss of quality. Generic libraries for the three vector-based formats were recoded from scratch for VARNA to be used within a minimal environment (no external software required), while remaining as lightweight as possible (no bundled third-party library).

2.3 Drawing algorithms

VARNA implements four distinct algorithms to draw the secondary structure of RNA in one of its three usual representations: The *linear* representation draws the backbone on a straight line and connects paired bases with arcs; the Feynman diagram-like *circular* representation draws the backbone on a circle while connecting partners with chords; the *planar graph* representation uses predefined distances between neighboring paired bases, while limiting the structural overlap. We implemented two algorithms for the planar graph representation. The *radial* strategy draws each helix and base in a multiloop at regular angular distances, as done by RNAviz (Rijk *et al.*, 2003). This gives a drawing which is potentially self-intersecting, but allows for richer user-interactions. Additionally, the exterior loop can be aligned to a baseline in this representation. The *NAVView* drawing algorithm (Brucoleri and Heinrich, 1988) uses a heuristic approach to generate a non-intersecting representation of the RNA secondary structure.

2.4 Non-canonical base-pairs/Pseudoknots

Non-canonical base-pairs can be read from an RNAML file and displayed using standard notations defined by the Leontis-Westhof classification system (Leontis and Westhof, 2001). Other representations, such as that used in RNAviz (Rijk *et al.*, 2003), may be the preferred choice.

Partial support for pseudoknots is also provided by VARNA. Upon loading a pseudoknotted secondary structure, a maximal non-crossing subset of its base-pairs is extracted using a dynamic programming algorithm (Xayaphoummine *et al.*, 2003). The resulting secondary structure is then used as a scaffold for additional base-pairs, resulting in a *planar graph* representation.

2.5 Editing/annotation

VARNA users can modify an automatically drawn structure using basic interactions. Namely, structures drawn using the *radial* algorithm can be modified by rotating or flipping (exterior loop featured only) an helix, together with all its subsequent elements, around the center of its supporting loop, allowing manual removal of potential self-intersections. The other drawing algorithms offer the possibility of post-editing modifications by means of moving single bases (*NAVView* or *circular*) or base-pairs (*planar*).

VARNA also offers base and base-pair annotation features, allowing selected bases, base-pairs or any structural element (Helix, loop, stem, ...), chosen from a contextual menu, to be highlighted and/or annotated. Once these elements are selected, individual bases and base-pairs can be customized to highlight a region of interest.

2.6 Web server features

As an applet, most features offered by VARNA can be accessed through HTML `param` tags. This allows a web server to

benefit from VARNA's visual impact, as previously shown for the IRESite (Mokrejs *et al.*, 2006), using a previous version of our software. Additionally, the VARNA applet can render several RNA structures simultaneously, allowing a visual comparison of RNA structures, as previously shown with the NestedAlign (Blin *et al.*, 2008) webserver (<http://nestedalign.lri.fr/>). User interactions can be restricted to circumvent undesired modifications.

3 REQUIREMENTS

A version (1.5+) of the JAVA plugin is required to run VARNA.

FUNDING AND ACKNOWLEDGEMENTS

This work was partially supported by the ANR projects BRASERO ANR-06-BLAN-0045 and GAMMA 07-2_195422. The authors are grateful to M. Weber, D. Castillo and the anonymous referees for helpful suggestions during early stages of VARNA's development.

REFERENCES

- Blin, G., Denise, A., Dulucq, S., Herrbach, C., and Touzet, H. (2008). Alignment of RNA structures. *IEEE/ACM Trans Comp Biol Bioinfo*. DOI: 10.1109/TCBB.2008.28.
- Brucoleri, R. E. and Heinrich, G. (1988). An improved algorithm for nucleic acid secondary structure display. *Comp Appl Biosci*, **4**(1), 167–173.
- Cannone, J., Subramanian, S., Schnare, M., Collett, J., D'Souza, L., Du, Y., Feng, B., Lin, N., Madabusi, L., Muller, K., Pande, N., Shang, Z., Yu, N., and Gutell, R. (2002). 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).
- Ding, Y., Chan, C., and Lawrence, C. (2004). SFold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res. - Web Server Issue*, **32**, 135–141.
- Leontis, N. and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
- Markham, N. R. and Zuker, M. (2005). Dinamelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, 577–581.
- Mokrejs, M., Vopalensky, V., Kolenaty, O., Masek, T., Feketova, Z., Sekyrova, P., Skaloudova, B., Kriz, V., and Pospisek, M. (2006). IRESite: the database of experimentally verified IRES structures (www.iresite.org). *Nucleic Acids Res*, **34**(S1), D125–130.
- Rijk, P. D., Wuyts, J., and Wachter, R. D. (2003). Rnaviz2: an improved representation of RNA secondary structure. *Bioinformatics*, **2**(19), 299–300.
- Washietl, S., Hofacker, I., and Stadler, P. F. (2005). Fast and reliable prediction of noncoding RNAs. *PNAS*, **102**(7), 2454–2459.
- Waugh, A., Gendron, P., Altman, R., Brown, J. W., Case, D., Gautheret, D., Harvey, S. C., Leontis, N., Westbrook, J., Westhof, E., Zuker, M., and Major, F. (2002). RNAML: a standard syntax for exchanging RNA information. *RNA*, **8**(6), 707–717.
- Xayaphoummine, A., Bucher, T., Thalmann, F., and Isambert, H. (2003). Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations. *PNAS*, **100**(26), 15310–15315.