

The RNA shapes studio

Stefan Janssen and Robert Giegerich*

Practical Computer Science, Faculty of Technology, Bielefeld University, D-33615 Bielefeld, Germany

Associate Editor: John Hancock

ABSTRACT

Motivation: Abstract shape analysis, first proposed in 2004, allows one to extract several relevant structures from the folding space of an RNA sequence, preferable to focusing in a single structure of minimal free energy. We report recent extensions to this approach.

Results: We have rebuilt the original RNASHAPES as a repository of components that allows us to integrate several established tools for RNA structure analysis: RNASHAPES, RNAALISHAPES and PKNOTSRG, including its recent extension PKISS. As a spin-off, we obtain heretofore unavailable functionality: e. g. with PKISS, we can now perform abstract shape analysis for structures holding pseudoknots up to the complexity of kissing hairpin motifs. The new tool PALIKISS can predict kissing hairpin motifs from aligned sequences. Along with the integration, the functionality of the tools was also extended in manifold ways.

Availability and implementation: As before, the tool is available on the Bielefeld Bioinformatics server at <http://bibiserv.cebitec.uni-bielefeld.de/rnashapesstudio>.

Contact: bibi-help@cebitec.uni-bielefeld.de

Received on August 18, 2014; revised on September 25, 2014; accepted on September 26, 2014

1 THE RNA SHAPES STUDIO

1.1 Integration of tools for RNA abstract shape analysis

The framework of algebraic dynamic programming (ADP) allows us to express dynamic programming algorithms for sequence analysis on a high level of abstraction. They are composed from signatures, tree grammars and evaluation algebras (Giegerich *et al.*, 2004a). Powerful product operations on algebras allow one to derive new types of analysis by the combination of available components, essentially with a single keystroke (Steffen and Giegerich, 2005). Relying on the recent BELLMAN'S GAP system (Sauthoff *et al.*, 2013), which implements the ADP framework, we have built a repository of components that allows us to integrate several established tools for RNA structure analysis: RNASHAPES, RNAALISHAPES and PKNOTSRG, including its recent extension PKISS. As a spin-off, we obtain heretofore unavailable functionality: e. g. with PKISS, we can now perform abstract shape analysis for structures holding pseudoknots up to the complexity of kissing hairpin motifs. The new tool PALIKISS can predict kissing hairpin motifs from aligned sequences. Along with the integration, the functionality

of the tools was also extended in manifold ways. Figure 1 gives an overview.

1.2 Integrated tools and their new functionality

1.2.1 Extensions to RNASHAPES It is generally agreed that predicting a single structure of minimal free energy does not adequately capture the subtlety and versatility of RNA structure formation. The RNASHAPES tool introduced the notion of abstract shapes (Giegerich *et al.*, 2004b; Voß *et al.*, 2006)—a (mathematically precise) characterization of structures by their arrangement of helices. For example, '[[] []]' indicates a cloverleaf shape, and '[_ [_] _]' a single stem-loop with a 5' bulge and an internal loop. Classical abstract shape analysis reports minimum free energy structures from *different* shape classes, or Boltzmann structure probabilities accumulated by shape. This gives synoptic information about the folding space of a given RNA sequence, without heuristics or sampling. Extending RNASHAPES, we added different modes of treating dangling bases (consistent with RNAFOLD options -d0, -d1 and -d2) (Janssen *et al.*, 2011; Lorenz *et al.*, 2011), computation of base pair probabilities and maximum expected accuracy (MEA) folding (Lu *et al.*, 2009).

1.2.2 Extensions to RNAALISHAPES The work of Voß (2006) combines the ideas of RNAALIFOLD and RNASHAPES and performs shape analysis based on pre-aligned RNA sequences. We added the computation of a structure conservation index, different dangling base models, MEA folding and a window mode. RIBOSUM scoring (Bernhart *et al.*, 2008) was added for evaluating sequence similarity.

1.2.3 Extensions to PKISS In Theis *et al.* (2010) the ideas of PKNOTSRG (Reeder and Giegerich, 2004) are extended to predict (aside from unknotted structures and H-type pseudoknots) RNA structures that exhibit kissing hairpin motifs in an arbitrarily nested fashion, requiring $O(n^4)$ time and $O(n^2)$ space. We added shape analysis, probabilities, different folding strategies and different dangling base models. The -cast option provides comparative prediction of pseudoknotted structures as in the RNACAST approach (Reeder and Giegerich, 2005). A window mode was also included.

1.2.4 New tool PALIKISS The program PALIKISS allows to predict pseudoknots, including kissing hairpins from aligned sequences. Being composed from the grammars and algebras of the other tools, it inherits all the features and options that make sense for it.

*To whom correspondence should be addressed.

		RNashapes										RNAalishapes										pKiss						pAliKiss									
		probs	shapes	sample	cast	mfe	subopt	mea	outside	eval	abstract	probs	shapes	sample	mfe	subopt	mea	outside	eval	abstract	mfe	subopt	enforce	local	shapes	probs	cast	eval	abstract	mfe	subopt	enforce	local	shapes	probs	eval	abstract
sub-optimal energy deviation	relative deviation	x	x	x								x	x								x	x	x	x													
	absolute deviation	x	x	x								x	x								x	x	x	x													
stochastic options	number of samples		x										x																								
	show samples		x										x																								
	low probability filter	x										x																									
	output probability filter	x	x									x	x																								
	decimals for probabilities	x	x									x	x																								
alignment options	structure conservation index	x	x	x	x	x	x					x	x	x	x	x																					
	consensus												x	x	x	x																					
	ribosum scoring												x	x	x	x																					
	pairing fraction												x	x	x	x																					
	cfactor												x	x	x	x																					
pseudoknot options	nfactor											x	x	x	x	x																					
	strategy																																				
	Htype penalty																																				
	Ktype penalty																																				
	maximal pseudoknot size																																				
folding options	minimal hairpin length																																				
	grammar																																				
	shape level	x	x	x	x	x	x				x	x	x	x	x	x																					
	temperature																																				
	thermodynamic parameters																																				
input style	lonely base pairs																																				
	window size	x	x	x		x	x					x	x	x	x	x																					
	window increment	x	x	x		x	x																														
		compared to RNashapes 2.1.5 (15.04.2008) by Peter Steffen										compared to RNAalishapes 1.0 (2006) by Bjoern Voss										compared to pknotsRG 1.3 (19.09.2006) by Jens Reeder															

Fig. 1. Parameters for the RNA shapes studio. New features are indicated by +. New parameters are highlighted in yellow. New analysis modes are shaded in green

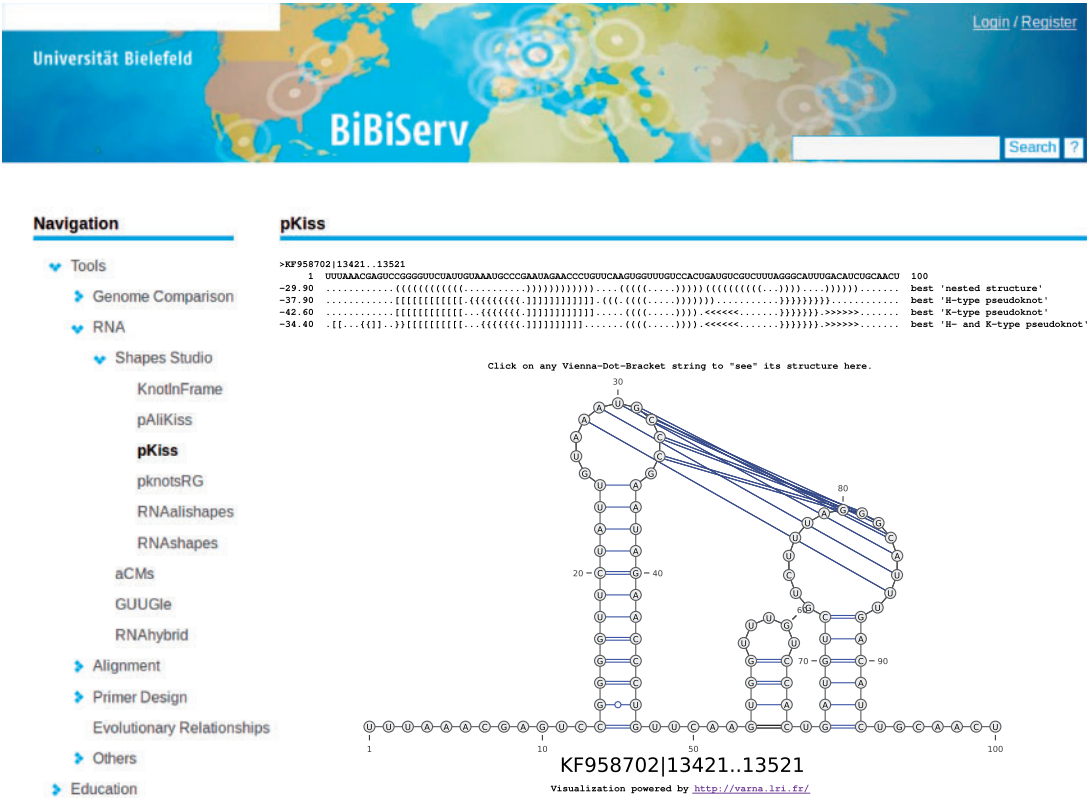


Fig. 2. RNA shapes studio result page for folding the MERS example with pKiss. Illustration by VARNAs (Darty *et al.*, 2009)

1.2.5 Utilities All tools were augmented with utilities to compute folding energy or abstract shape for sequences that are provided with a structure from an outside source, in a way consistent with the tools' energy model. The graphical motif description tool LOCOMOTIF (Reeder *et al.*, 2007) now uses modules from the RNA shapes studio. The KNOTINFRAME (Theis *et al.*, 2008) tool that predicts -1 ribosomal frameshifts has been updated as well.

2 APPLICATION CASE: A FRAMESHIFT STIMULATION ELEMENT IN MERS

The Corona virus family contains a frameshift stimulation element (Baranov *et al.*, 2005), where the frameshift is facilitated by a slippery site together with either an H-type or a K-type pseudoknot. RFAM (Burge *et al.*, 2013) holds the corresponding family model RF00507, although the tools of RFAM cannot explicitly model pseudoknots. PKISS (Theis *et al.*, 2010) in -enforce mode reveals that for 11 family members, minimal free energy structures are H-types, another 11 are K-type pseudoknots and for only one member a purely nested structure has the best energy.

The recently sequenced MERS genome (KF958702.1) is annotated with a homologous frameshift site, whereas the structure of the triggering element remains unclear. Structure prediction with PKISS for a 100 bp stretch downstream the slippery site attests a most stable K-type pseudoknot (see Fig. 2). A second run of PKISS, this time in probability mode, shows that the shape class of this particular K-type pseudoknot has an overwhelming Boltzmann probability of $\approx 99\%$; leaving not much probability mass for any other shape classes.

3 AVAILABILITY

The RNA shapes studio is available at <http://bibiserv.cebitec.uni-bielefeld.de/rnashapesstudio>. Users can access the BELLMAN's GAP source code of all components in the repository, and combine or extend them according to their own goals. This has been done, for example, in Reinkensmeier *et al.* (2011) for defining the CCUCCUCC-motif family in the *Rhizobiales*.

Conflict of interest: none declared.

REFERENCES

- Baranov,P.V. *et al.* (2005) Programmed ribosomal frameshifting in decoding the sars-cov genome. *Virology*, **332**, 498–510.
- Bernhart,S. *et al.* (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, **9**, 474.
- Burge,S.W. *et al.* (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.*, **41**, D226–D232.
- Darty,K. *et al.* (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
- Giegerich,R. *et al.* (2004a) A discipline of dynamic programming over sequence data. *Sci. Comput. Program.*, **51**, 215–263.
- Giegerich,R. *et al.* (2004b) Abstract shapes of RNA. *Nucleic Acids Res.*, **32**, 4843–4851.
- Janssen,S. *et al.* (2011) Lost in folding space? Comparing four variants of the thermodynamic model for RNA secondary structure prediction. *BMC Bioinformatics*, **12**, 429.
- Lorenz,R. *et al.* (2011) ViennaRNA package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
- Lu,Z.J. *et al.* (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA*, **15**, 1805–1813.
- Reeder,J. and Giegerich,R. (2004) Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics*, **5**, 104.
- Reeder,J. and Giegerich,R. (2005) Consensus shapes: an alternative to the Sankoff algorithm for RNA consensus structure prediction. *Bioinformatics*, **21**, 3516–3523.
- Reeder,J. *et al.* (2007) Locomotif: from graphical motif description to RNA motif search. *Bioinformatics*, **23**, i392.
- Reinkensmeier,J. *et al.* (2011) Conservation and occurrence of trans-encoded sRNAs in the rhizobiales. *Genes*, **2**, 925–956.
- Sauthoff,G. *et al.* (2013) Bellman's GAP - a language and compiler for dynamic programming in sequence analysis. *Bioinformatics*, **29**, 551–556.
- Steffen,P. and Giegerich,R. (2005) Versatile and declarative dynamic programming using pair algebras. *BMC Bioinformatics*, **6**, 224.
- Theis,C. *et al.* (2008) Knotinframe: prediction of -1 ribosomal frameshift events. *Nucleic Acids Res.*, **36**, 6013–6020.
- Theis,C. *et al.* (2010) Prediction of RNA secondary structure including kissing hairpin motifs. In: Moulton,V. and Singh,M. (eds), *Algorithms in Bioinformatics, Vol. 6293 of Lecture Notes in Computer Science*. Berlin Heidelberg, Springer, pp. 52–64.
- Voß,B. (2006) Structural analysis of aligned RNAs. *Nucleic Acids Res.*, **34**, 5471–5481.
- Voß,B. *et al.* (2006) Complete probabilistic analysis of RNA shapes. *BMC Biol.*, **4**, 5.