

# RNApdbee 2.0: multifunctional tool for RNA structure annotation

Tomasz Zok<sup>1,2</sup>, Maciej Antczak<sup>1</sup>, Michal Zurkowski<sup>1</sup>, Mariusz Popenda<sup>3</sup>, Jacek Blazewicz<sup>1,3</sup>, Ryszard W. Adamiak<sup>1,3</sup> and Marta Szachniuk<sup>1,3,\*</sup>

<sup>1</sup>Institute of Computing Science, and European Centre for Bioinformatics and Genomics, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland, <sup>2</sup>Poznan Supercomputing and Networking Center, Jana Pawla II 10, 61-139 Poznan, Poland and <sup>3</sup>Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

Received January 31, 2018; Revised March 30, 2018; Editorial Decision April 12, 2018; Accepted April 14, 2018

## ABSTRACT

**In the field of RNA structural biology and bioinformatics, an access to correctly annotated RNA structure is of crucial importance, especially in the secondary and 3D structure predictions. RNApdbee webserver, introduced in 2014, primarily aimed to address the problem of RNA secondary structure extraction from the PDB files. Its new version, RNApdbee 2.0, is a highly advanced multifunctional tool for RNA structure annotation, revealing the relationship between RNA secondary and 3D structure given in the PDB or PDBx/mmCIF format. The upgraded version incorporates new algorithms for recognition and classification of high-ordered pseudoknots in large RNA structures. It allows analysis of isolated base pairs impact on RNA structure. It can visualize RNA secondary structures—including that of quadruplexes—with depiction of non-canonical interactions. It also annotates motifs to ease identification of stems, loops and single-stranded fragments in the input RNA structure. RNApdbee 2.0 is implemented as a publicly available webserver with an intuitive interface and can be freely accessed at <http://rnapedbee.cs.put.poznan.pl/>**

## INTRODUCTION

Elucidation of the three-dimensional structure of RNA is essential to understand a plethora of its functions in the living cells. X-ray crystallography, NMR, and cryo-microscopy—major experimental methods to achieve this—are currently successfully accompanied by 3D structure prediction (1). The secondary structure which encodes both canonical and non-canonical base pairing schemes constitutes an important intermediate information to be

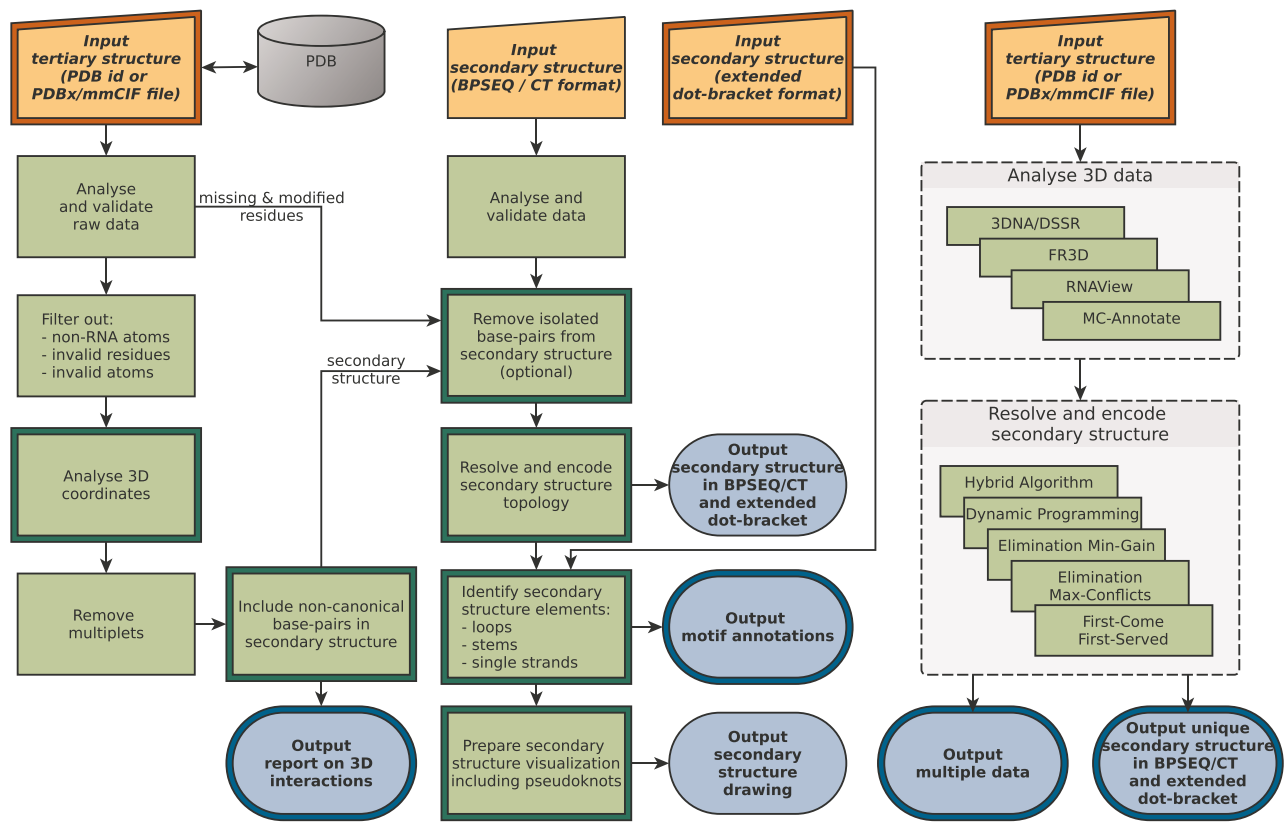
positioned between RNA sequence and 3D structure. For large RNAs, decrypting the secondary structure from sequence requires adjustment of *in silico* prediction results with the involvement of experimental data derived mostly from chemical probing.

In 2014, we introduced RNApdbee webserver (2) allowing for extraction of RNA secondary structure from 3D structure data encoded in the PDB file, and its visualization that showed RNA chain topology. The computational procedure to derive the secondary structure from the atom coordinate set was composed of two major steps. First, canonical and non-canonical base pairs were identified, extracted from the PDB file, and listed. Next, RNA secondary structure topology was encoded in the dot-bracket format and visualized. In this respect, RNApdbee complemented other programs to identify and classify RNA base pairs (3–5). Soon, our tool was directed to such important tasks like: (i) quality assessment of algorithms for RNA secondary structure prediction and validation (6), (ii) comparison of RNA folds on the secondary structure level (7), (iii) analysis of predicted or experimentally determined RNA 3D models via their conversion to the secondary structure (8,9). The latter application was an important checkpoint in the evaluation of automated prediction of large RNA 3D structures by RNAComposer (10).

Here, we present RNApdbee 2.0, a highly advanced multifunctional tool for RNA structure annotation and revealing the relationship between RNA secondary and tertiary structure. At present, analysis of this correlation needs progression from a classical PDB format (no longer developed by Protein Data Bank) to PDBx/mmCIF. All procedures in the upgraded version of our tool are based on both of these formats.

RNApdbee 2.0 features the following novelties: additional methods for base pair identification and secondary structure visualization, new algorithms to recognize and classify high-order pseudoknots, monitoring of isolated

\*To whom correspondence should be addressed. Tel: +48616652999; Fax: +48618771525; Email: [marta.szachniuk@cs.put.poznan.pl](mailto:marta.szachniuk@cs.put.poznan.pl)



**Figure 1.** Workflow scheme of RNApdbee 2.0. Double framed boxes indicate where new functionalities have been added.

base pairs impact on RNA topology, new multiple output scenario that allows inferring consensus secondary structure, classification of non-canonical interactions and their annotation in graphic view, new way to display secondary structure of quadruplexes, an option to annotate secondary structure motifs in the input RNA structure, an acceleration in 3D structure data processing. RNApdbee 2.0 supports all the steps of RNA structure prediction and analysis. It complements recent computational tools that process and annotate RNA secondary structure, like ClaRNA (11), FreeKnot (12) and CompAnnotate (13). ClaRNA provides information about nucleotide interactions based on the input RNA 3D structure. It runs the classification process to score ribonucleotide doublets against a reference set of interactions derived from experimentally solved RNAs. CompAnnotate (13) aims at improving annotation quality of low-resolution RNA structures by making use of comparative geometric assessments from high-resolution homologues. FreeKnot (12), equipped with new evaluation and optimization routines, has completed a set of methods for removal of pseudoknots from the RNA secondary structure (14).

RNApdbee 2.0 incorporates several structure annotation and pseudoknot encoding algorithms. It accepts inputs in both legacy and latest formats. The output is presented in complementary ways including textual and graphical rep-

resentations. It is implemented as publicly available web-server with a user-friendly interface, freely accessible at <http://rnapdbee.cs.put.poznan.pl/>

## METHOD OUTLINE

RNApdbee 2.0 offers four usage scenarios. Two of them, *3D* and *2D scenario*, introduced in the first version of the tool, have been now enriched with new options. *Image* and *Multiple scenario* are entirely new. Figure 1 presents the scheme of data processing in all scenarios. An overview of novelties is given in the following paragraphs.

*3D scenario* is run on the  $3D \rightarrow (\dots)$  tabpage of RNApdbee 2.0. Here, the secondary structure of RNA is obtained from its 3D structure in three steps: (i) base pair identification, (ii) structure topology resolving and encoding, (iii) structure visualization. In all the steps, the user has several choices between the algorithms dedicated to relevant tasks. Base pairs can be identified by 3DNA/DSSR (default) (4), RNAView (5), MC-Annotate (3) or newly added FR3D (15). The user can also decide how to treat non-canonical and isolated base pairs which strongly impact pseudoknot orders and structure topology. The second step has been recently introduced and has five algorithms, Hybrid Algorithm (default), Dynamic Programming, Elimination Min-Gain, Elimination Max-Conflicts, First-Come-First-Served (16), to support high-order pseudoknot pro-

cessing. They follow different routines to find the optimum nested substructure, identify pseudoknot orders and encode the secondary structure topology in extended dot-bracket notation. For nested RNAs, all of them provide the same result. In the third step, structure visualization can be done by our procedure based on VARNA (default) (17), PseudoViewer (18), or newly added R-chie (19). The new drawing function supports multi-stranded RNAs, e.g. quadruplexes. Additionally, a new option allows identification of structural elements, like stems, loops and single-stranded fragments in the input structure. Their location and dot-bracket representation can be also viewed at the result page.

*2D scenario* is available on the *2D*→(....) tabpage. It aims to process the input list of base pairs and elucidate secondary structure topology proceeding through (i) structure topology resolving and encoding (new), and (ii) visualization. New options for isolated base pair manipulation and motif identification can be also found here. Both steps and options work in the same way as in the *3D scenario*.

*Image scenario* runs on the (....)→*Image* tabpage and it is reserved for drawing RNA secondary structure based on the input dot-bracket string. As in previous scenarios, the user can select from three visualization methods. Motif identification option is also available here.

*Multiple scenario* is found on the *3D*→*multi 2D* tabpage. This new scenario allows running all algorithms for base pair identification and structure topology resolution with just one click, and compare their outcome. As a result, a compressed rectangular diagram (where each paired region is represented by single connection) is drawn to show consensus for pseudoknots. The differences between secondary structures given by various algorithms for topology resolving are shown in extended dot-bracket. An option to include or skip non-canonical and isolated base pairs expands the possibility of studying the differences between methods used in secondary structure processing. This scenario provides information from which the most likely secondary structure can be identified.

## WEB APPLICATION

RNApdbee 2.0 has been implemented in a two-layer architecture with distinguished computational backend and web-accessible frontend. Both components are based on Java 8 with third-party libraries. BioJava (20) provides methods to parse mmCIF files. The backend is hosted on a machine with 12GB RAM and 6 CPU units Intel(R) Xeon(R) CPU E5-4640 2.40 GHz. This ensures fast processing of concurrent workloads, even for large RNAs. The frontend is available as a web-page that works with all modern browsers and platforms, including mobile devices. It is based on Spring MVC, Apache Maven, Bower, jQuery, and open-sourced framework called Apache Tiles. RNApdbee web server host computer provides disk space for a periodically updated copy of PDB-deposited data. This speeds up computation and optimizes network bandwidth, which becomes important as increasingly large RNAs are solved and submitted to PDB. RNApdbee 2.0 web service is hosted and maintained by the Institute of Computing Science, Poznan University of Technology, Poland.

## Input and output description

In the *3D scenario* and *Multiple scenario*, the input 3D structure can be provided by the user in gzip compressed or plain format (PDB or PDBx/mmCIF), or downloaded from PDB server given the PDB id. The secondary structure can be loaded from a file in BPSEQ or CT format in the *2D scenario*. In the *Image scenario*, it should be given in the extended dot-bracket notation. Similar to RNA FRABASE (21), the input sequence is coded with one-letter format where unmodified RNA residues are represented by capital letters while all modified RNA and non-RNA units (e.g. DNA units) are recorded in small letters. In each scenario, several examples for all the supported formats are available for processing. The uploaded input structure can be visualized by clicking *Show file contents* and modified on site.

RNApdbee 2.0 generates a variety of output data. One of the most important is graphical view of RNA secondary structure for which a lot of novelties have been implemented. First, the output structure can be drawn by R-chie method (19) in a form of arc diagram. The colors used in these diagrams are consistent with the color code set for drawings generated by the other procedures working within RNApdbee webserver. This is particularly important in color encoding of pseudoknot orders. Non-canonical interactions are annotated with pictograms related to Leontis-Westhof classification (22) in the images generated by VARNA-based procedure (17). A new algorithm that prepares data for visualization also determines which isolated base pairs do not interfere with the overall topology of RNA structure (Supplementary Figure S1). The secondary structures of multi-stranded RNAs are now clearly visualized. The novelties in textual output include: the information about stacking, base-phosphate and base-ribose interactions and motifs annotated in the input data.

The results can be downloaded from the *Result page* in a single archive including files in common, user-friendly formats. The tabular data are saved in CSV files, all images are in both vector and raster graphics, and structural elements—in the text representation and as PDB files with 3D coordinates.

## RESULTS AND DISCUSSION

Here, we present examples of RNApdbee 2.0 application where major new features of the tool have been highlighted.

Crystal structure of cyanocobalamin (vitamin B12) aptamer (PDB id: 1DDY) (23) represents a complex, multi-strand RNA architecture in which duplex structure stacks perpendicularly to a locally folded triplex, stabilized by a novel three-stranded zipper. The formed cleft functions as the vitamin B12 binding site. RNApdbee 2.0 gives a detailed picture of the secondary structure of this molecule. Cyanocobalamin aptamer consists of four structurally similar strands, each containing 35 residues. Although the basic topology – determined by eight canonical interactions – is similar for all strands, the differences can be observed within a set of non-canonical base pairs. In every strand, eight non-canonical interactions have been recognized by the RNApdbee 2.0 procedure. However, they involve various nucleotide residues and, thus, have been differently





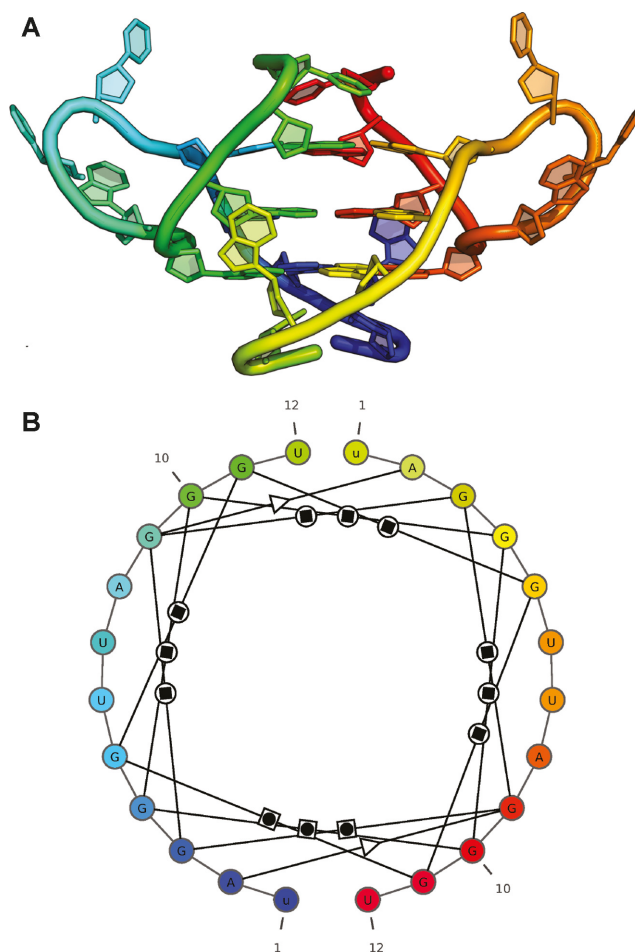
stranded fragments in the input structure, and its result depends on how pseudoknots—if present—are processed. Supplementary Figures S3-S4 show the vitamin B12 aptamer structure divided into elements and their textual annotation provided in the *3D scenario*. In the first case (Supplementary Figure S3), pseudoknot-involved residues are considered unpaired. Such approach is consistent with that used in RNAComposer (10,24), whose upgraded version (24) allows including own structure elements which can be obtained by *Identify structural elements* function of RNAPdbec 2.0. If pseudoknot-forming residues are treated as pairs, an analysis identifies more structural elements in the analyzed RNA which clearly shown on the provided example (Supplementary Figure S4).

Guanine-rich sequences in RNA can assemble into quadruplex structures that involve G-quartets linked by loop nucleotide residues. Currently, RNA (and DNA) quadruplexes represent a group of actively studied structures (25). Secondary structure visualization plays a vital role in the study of their features. Till today, there was no adequate method to visualize the canonical and non-canonical interactions within the secondary structures and various schemes of base-base tetrads. In RNAPdbec 2.0, we have implemented a new clear way of portraying quadruplex secondary topology.

The structure of G-quadruplex of human telomeric RNA (PDB id: 2KBP) (26) has been selected as our second example. This structure is build of 12-nucleotide strands and involves three G-tetrads. An image generated by RNAPdbec 2.0 procedure (Figure 3) gives a detailed view of its interactions and enables their analysis. A circular form of the structure clearly reflects tetrads and looped out residues. Each G-tetrad forms a tetragon with guanosine nodes. Twelve W/H (Watson-Crick/Hoogsteen) base pairs are created between the guanines. In addition, two non-canonical base pairs of S/S type (Sugar/Sugar) are revealed between adenosine and guanosine residues. The study indicates that quadruplex composed of three G-tetrads is parallel and includes 3-nucleotide UUA loops. Since RNAPdbec 2.0 accepts DNA as an input, it can be also applied to analyze DNA quadruplexes, like oxytricha telomeric DNA structure (PDB id: 1JPQ) (27) displayed in Supplementary Figure S5. This molecule consists of four G-tetrads and its structure is anti-parallel.

## CONCLUSIONS

We have introduced RNAPdbec 2.0 which significantly enhances the functionality of our webserver's first version. New functions allow a more detailed analysis of RNA 3D folds via its corresponding secondary structure. For pseudoknots, they give the basics to establish the order and complexity depending on the adopted criteria. To our knowledge, RNAPdbec 2.0 is the only tool able to visualize secondary structures determined by non-canonical base pairs only. This feature greatly helps in interpretation of the wide range of RNA and DNA quadruplex folds, and characterization of their interactions. We believe, RNAPdbec 2.0 will prove useful in the study of both secondary and tertiary structures of RNA molecules.



**Figure 3.** (A) 3D structure of G-quadruplex of human telomeric RNA (PDB id: 2KBP) visualized in PyMOL and (B) its secondary structure diagram given by RNAPdbec 2.0 run with the default settings (the original black and white diagram was manually colored in the rainbow scale to depict the correspondence with the 3D model).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Science Centre, Poland [2016/23/B/ST6/03931 to M.S., 2016/23/N/ST6/03779 to T.Z.]. Funding for open access charge: Institute of Bioorganic Chemistry, Polish Academy of Sciences.

*Conflict of interest statement.* None declared.

## REFERENCES

- Miao,Z. and Westhof,E. (2017) RNA structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.*, **46**, 483–503.
- Antczak,M., Zok,T., Popenda,M., Lukasiak,P., Adamiak,R., Blazewicz,J. and Szachniuk,M. (2014) RNAPdbec - a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Res.*, **42**, W368–W372.
- Gendron,P., Lemieux,S. and Major,F. (2001) Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.*, **308**, 919–936.

4. Lu, X.-J. and Olson, W. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
5. Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
6. Rybarczyk, A., Szostak, N., Antczak, M., Zok, T., Popena, M., Adamiak, R., Blazewicz, J. and Szachniuk, M. (2015) New in silico approach to assessing RNA secondary structures with non-canonical base pairs. *BMC Bioinformatics*, **16**, 276.
7. Micheletti, C., DiStefano, M. and Orland, H. (2015) Absence of knots in known RNA structures. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 2052–2057.
8. Miao, Z., Adamiak, R., Antczak, M., Batey, R., Becka, A., Biesiada, M., Boniecki, M., Bujnicki, J., Chen, S.-J., Cheng, C. *et al.* (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.
9. Gómez, Ramos L., Degtyareva, N., Kovacs, N., Holguin, S., Jiang, L., Petrov, A., Biesiada, M., Hu, M., Purzycka, K., Arya, D. *et al.* (2017) Eukaryotic ribosomal expansion segments as antimicrobial targets. *Biochemistry*, **56**, 5288–5299.
10. Popena, M., Szachniuk, M., Antczak, M., Purzycka, K., Lukasiak, P., Bartol, N., Blazewicz, J. and Adamiak, R. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.
11. Waleń, T., Chojnowski, G., Gierski, P. and Bujnicki, J. (2014) ClaRNA: a classifier of contacts in RNA 3D structures based on a comparative analysis of various classification schemes. *Nucleic Acids Res.*, **42**, e151.
12. Chiu, J. and Chen, Y.-P. (2015) Efficient conversion of RNA pseudoknots to knot-free structures using a graphical model. *IEEE Trans. Biomed. Eng.*, **62**, 1265–1271.
13. Islam, S., Ge, P. and Zhang, S. (2017) CompAnnotate: a comparative approach to annotate base-pairing interactions in RNA 3D structures. *Nucleic Acids Res.*, **45**, e136.
14. Smit, S., Rother, K., Heringa, J. and Knight, R. (2008) From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. *RNA*, **14**, 410–416.
15. Sarver, M., Zirbel, C., Stombaugh, J., Mokdad, A. and Leontis, N. (2007) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.*, **56**, 215–252.
16. Antczak, M., Popena, M., Zok, T., Zurkowski, M., Adamiak, R. and Szachniuk, M. (2018) New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation. *Bioinformatics*, **34**, 1304–1312.
17. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
18. Byun, Y. and Han, K. (2006) PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures. *Nucleic Acids Res.*, **34**, W416–W422.
19. Lai, D., Proctor, J., Zhu, J. and Meyer, I. (2012) R-CHIE: a web server and R package for visualizing RNA secondary structures. *Nucleic Acids Res.*, **40**, e95.
20. Prlic, A., Yates, A., Bliven, S., Rose, P., Jacobsen, J., Troshin, P., Chapman, M., Gao, J., Koh, C., Foisy, S. *et al.* (2012) BioJava: an open-source framework for bioinformatics in 2012. *Bioinformatics*, **28**, 2693–2695.
21. Popena, M., Blazewicz, M., Szachniuk, M. and Adamiak, R. (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res.*, **36**, D386–D391.
22. Leontis, N. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
23. Sussman, D., Wilson, C. and Nix, J. (2000) The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nat. Struct. Biol.*, **7**, 53–57.
24. Antczak, M., Popena, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., Adamiak, R. and Szachniuk, M. (2016) New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim. Pol.*, **63**, 737–744.
25. Kwok, C. and Merrick, C. (2017) G-Quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.*, **35**, 997–1013.
26. Martadinata, H. and Phan, A. T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K<sup>+</sup> solution. *J. Am. Chem. Soc.*, **131**, 2570–2578.
27. Haider, S., Parkinson, G. N. and Neidle, S. (2002) Crystal structure of the potassium form of an oxytricha nova G-quadruplex. *J. Mol. Biol.*, **320**, 189–200.