

RNApolis: Computational Platform for RNA Structure Analysis

Marta Szachniuk*

Abstract. In the 1970s, computer scientists began to engage in research in the field of structural biology. The first structural databases, as well as models and methods supporting the analysis of biomolecule structures, started to be created. RNA was put at the centre of scientific interest quite late. However, more and more methods dedicated to this molecule are currently being developed. This paper presents RNApolis - a new computing platform, which offers access to seven bioinformatic tools developed to support the RNA structure study. The set of tools include a structural database and systems for predicting, modelling, annotating and evaluating the RNA structure. RNApolis supports research at different structural levels and allows the discovery, establishment, and validation of relationships between the primary, secondary and tertiary structure of RNAs. The platform is freely available at <http://rnapolis.pl>

Keywords: Structural bioinformatics, RNA structure, Biological data processing, Algorithms for RNA

1. Introduction

Ribonucleic acid (RNA) belongs to the least recognized biological molecules. For many years, it has been treated as a poor relative in the family, when scientists' interests have focused primarily on DNA and protein research. The developing branch of structural biology, whose aim was to learn about the structure of molecules and to discover the structure-function relationship, only occasionally touched upon the subject of RNA. This was due not only to the low interest in this molecule but also to the difficulty of studying and determining the RNA structure with experimental methods [1, 21, 41, 52, 53].

RNA structure is hierarchical (Figure 1) [5, 14]. This hierarchy allows the architecture and properties of a molecule to be studied from partial knowledge, which is often sufficient to draw basic conclusions [13, 35]. The simplest way to describe the structure of an RNA is to specify a nucleotide sequence of its chain (or chains if the molecule contains more than one). A sequence is also known as a primary structure. The RNA sequence is coded, like

* Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland, and Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

DNA, using one-letter IUPAC codes [24, 25] to represent the individual nucleotides in the chain [55]. Experimental identification of the RNA sequence is not as challenging as determining a higher order structure. As a result, the number of known RNA sequences has been growing rapidly for several years. They are collected in various databases, led by GenBank [9], EMBL [27], and DDBJ [51].

The simplicity of representing the RNA structure only through the sequence is associated with low information content in the description. The sequence shows which nucleotides compose the molecule structure, but it does not reveal what interactions take place between them. The information about bonds and their types, and the resulting image of the arrangement of single and double-stranded fragments in the molecule is contained in the representation of its secondary structure. Determining the secondary structure of RNA is not an easy task, especially if experimental methods are considered. Thus, the most popular way to obtain information about this structure is to predict it computationally from the sequence or to annotate it from the tertiary structure. Since the 1980s, many *in silico* sequence-based prediction methods have been developed for RNA [33, 47, 49]. Recently, they have been strengthened by the possibility to incorporate constraints from chemical probing, primarily SHAPE method [20, 32, 38]. Automated methods of RNA secondary structure annotation are also developing [5, 18, 48, 59]. Both these paths, prediction and annotation, result in a set of different secondary structures obtained for the same RNA molecule. The identification of the appropriate native structure is a separate problem that is expected to be solved with the support of the experimental data.

Information about the shape of the molecule in three-dimensional space is encoded in the tertiary structure. The description of RNA molecule at this level contains the coordinates of all atoms from which it is composed. In this representation, the sequence, i.e. the primary structure, is also included, which cannot be said about the secondary structure. No direct, single-output translation procedure of the 3D structure into the secondary structure exists. The 3D structure of RNA can be assessed in the experimental or computational way. The first approach utilizes experiments like X-ray crystallography or NMR spectroscopy and applies computational tools that support processing and analysis of experimental data [1, 52]. In the second one, the structure is predicted from the sequence by *in silico* methods [4, 19, 34]. From the experimental point of view, the first RNA 3D structures were studied relatively early. As Protein Data Bank (PDB) [11] statistics show, already in 1978, i.e. seven years after the creation and publication of this database, the first known experimental structure of the RNA molecule was deposited in it. However, the growth of data on RNA tertiary structures has been slow. Currently (as of 01.02.2019), PDB contains 1447 RNA 3D structures, which is less than 1% of data stored in this database that collects a total of 148586 molecular structures (including 137832 proteins, 1789 DNAs and 7518 structures of molecular complexes) [56]. The growing gap between the number of known sequences and 3D RNA structures became the motivation to work on automatic structure prediction. Since the beginning of the 21st century, intensive work has been underway to develop better and better systems that can anticipate the 3D shape of the RNA molecule based on information about its sequence or the secondary structure.

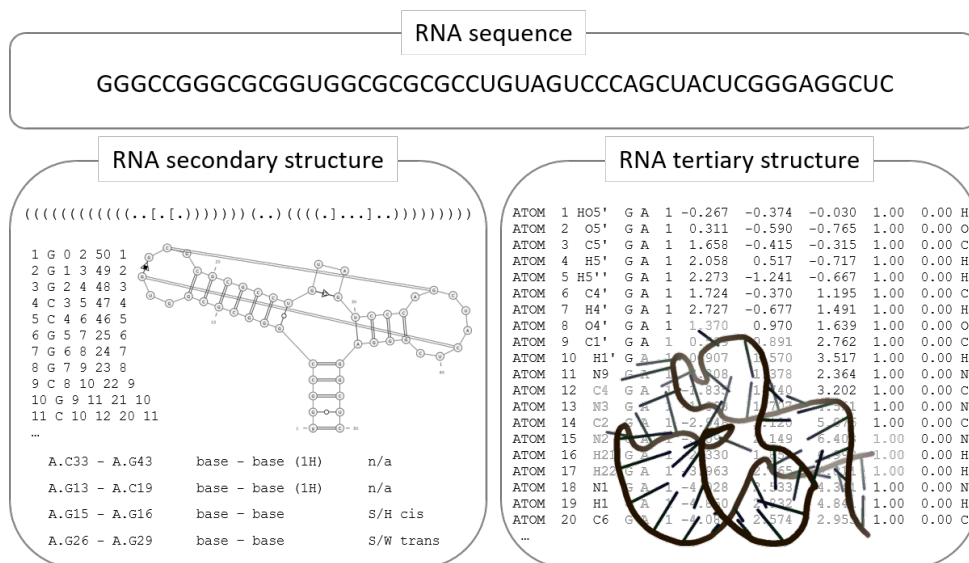


Figure 1. Sequence, secondary and tertiary structure data of an example RNA

In the previous paragraphs, some of the issues considered by structural bioinformatics have been indicated. They include, among others, modelling, prediction, annotation, refinement, visualization, validation and evaluation, motif analysis, similarity assessment, clustering, structure determination, a collection of structural data, their search and analysis. Many computer algorithms dedicated to solving these issues for RNA have been developed. For years, most of them have existed in separation from a range of similar methods implemented by other authors and offering competitive or complementary solutions. The first attempts to catalogue, on the websites, computational biology algorithms with links to their binaries or executables were made as part of Wikipedia. Then, the development of Internet technologies and easy access to the disk space offered by various companies has - in recent years - made it possible to create web portals collecting bioinformatics tools under one address. Most significant such initiatives are BioCatalogue and OmicX. BioCatalogue (<https://www.biocatalogue.org/>), created by the European Bioinformatics Institute in cooperation with myGrid project [50] at the University of Manchester, provides a curated catalogue of life science Web services [12]. From among 1168 offered services, nine are dedicated to RNA and mainly address the problem of modelling of RNA secondary structures. OmicX (created by OmicX start-up company), available at <https://omictools.com/> is a virtual analytical ecosystem for life science, covering research areas such as genomics, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics. Within transcriptomics, one can find rich collections of tools, each with a short commentary, concerning RNA, for example, RNA sequencing, RNA structure analysis, non-coding RNA databases. There are also initiatives like Bioinformatics.org available at <http://www.bioinformatics.org/>. It is a portal and repository for open source bioinformatics software. However, its content is not organized by a molecule type, and tools addressing RNA structures are not easy to find there. More specialized, RNA-centred online tool clusters are being created. Some of them are offered by teams creating their own

bioinformatic tools. Others accumulate all computational resources that can be used in research focused on specific topic. For example, University of Bielefeld maintains Bielefeld Bioinformatics Service (<https://bibiserv.cebitec.uni-bielefeld.de/>) that includes RNA chapter with own tools for RNA secondary structure processing. A selection of RNA secondary structure related computer programs is also available as ViennaRNA Package at <https://www.tbi.univie.ac.at/RNA/> run by the Theoretical Biochemistry Group from Vienna University [29]. The RNA Bioinformatics Center founded within the German Network of Bioinformatics Infrastructure offers own tools, services and databases for the analysis of RNA-based regulation (<https://github.com/bgruening/galaxy-rna-workbench/>) [8]. A set of 44 methods predicting RNA secondary structure can be found at the website of CompaRNA (<http://iimcb.genesilico.pl/comparna/>) that aims to benchmark these bioinformatics programs [47]. Another portal, miRToolsGallery (<http://www.mirtoolsgallery.org/>), collects information about computational tools that support the analysis of small non-coding RNAs [16]. The newly created RNA-Puzzles toolkit (<https://github.com/RNA-Puzzles/>) provides the resources (datasets, data formatting and model evaluation programs) used by the organizers and assessors of the RNA-Puzzles competition [17]. Other smaller RNA-dedicated toolkits are scattered over the internet and can be tracked if necessary.

This paper presents RNAPolis, a new computing platform that collects bioinformatics tools developed by the RNA bioinformatics group of researchers affiliated at the Institute of Computing Science, Poznan University of Technology and the Institute of Bioorganic Chemistry, Polish Academy of Sciences. The following RNA-dedicated tools are accessible via the platform: RNA FRABASE [42, 43], RNAComposer [4, 44, 46], RNApdbec [5, 59], RNAnalyzer [30], RNAssess [31], RNAfitme [5, 6, 58], and RNAvista [7]. They aim to support the research which focuses on the secondary and the tertiary structure of RNA molecules. With their help, one can study the relationship between these two structure levels of organization of RNA architecture. All of these tools have already gained recognition among scientists around the world who are engaged in the study of RNA structures. RNAPolis is available at <http://rnapolis.pl/>.

2. RNAPolis for RNA structure analysis

RNAPolis has been designed as a web portal allowing direct access to all bioinformatics tools developed by our group for RNA structure study. Currently, seven computational systems are available there. All of them have been designed as web services and can be accessed separately under unique URLs. Subsequent paragraphs familiarize the reader with individual tools, bringing their basic functionalities closer. The possibilities of using RNAPolis platform components are presented in terms of the input data the user possesses, and the goals that are to be achieved by this data processing.

2.1. Bioinformatics tools in RNAPolis

This section is devoted to a brief presentation of all bioinformatics systems currently included in the RNAPolis computational platform. Their descriptions, including an

introduction of the primary functions, are placed in the order in which the individual tools have been released and published.

RNA FRABASE 1.0 [42] was first published in 2008, and its version 2.0 appeared two years later [43]. It has been designed as a database of RNA structures, gathering information about the primary, secondary and tertiary structure of RNAs determined experimentally. Most of the data and metadata about RNA 3D structures uploaded to RNA FRABASE comes from Protein Data Bank (PDB) [11] and Nucleic Acid Database (NDB) [37], and is updated once a month. Additionally, information about torsional angles is computed by own scripts and saved in the database records. This data is supplemented by secondary structure information annotated with the use of own scripts and external programs. The database system gives the possibility to display the visualization of the secondary and tertiary structure of each deposited RNA. It contains links to associated records in PDB and NDB. The power of the system lies in the search engine, which enables quick and efficient searching of the repository content. The user can define the query pattern by providing either the RNA sequence, or secondary structure in dot-bracket notation, or both. In the dot-bracket notation, the secondary structure is represented by a sequence of characters (dots and brackets). Every unpaired nucleotide in the RNA chain is represented by a dot, and every base pair is encoded by a pair of brackets (an opening bracket for the first nucleotide in the pair, a closing bracket for the second nucleotide). RNA FRABASE engine searches all 3D structures in the database for the inclusion of the defined input pattern. As a result, the user obtains a list of all RNA structures that contain the searched sequence or secondary structure motif. For each found fragment, the user can view full information about its structure.

Since its release, RNA FRABASE has gained over 500 thousand users from more than 130 countries around the world. The total number of launches of the database search engine exceeds 3 million. Statistically, every 2 minutes a query is submitted to the database. RNA FRABASE is available at <http://rnafrabase.cs.put.poznan.pl/>.

RNAComposer [44] was released in 2012. Its first version was gradually enriched with new functionalities, which were summarized in a publication from 2016 [4]. RNAComposer was designed as a fully-automated system for RNA 3D structure prediction. It can predict the three-dimensional shape of the molecule on the basis of both the sequence and the secondary structure. The procedure creating 3D models uses incorporated database of RNA 3D structures which is an adjusted version of RNA FRABASE. RNAComposer works in two modes. The interactive mode allows to submit one input sequence or secondary structure at a time, run the prediction routine with a single click, and immediately obtain a single output 3D model. The batch mode has been provided to enable submitting multiple input data and obtain up to ten 3D models for every input. In this mode, the user can configure the system according to his/her preferences and include additional constraints to be considered by the structure refinement procedures. Finally, the batch mode has been enriched with the possibility to provide own 3D structure elements to be built into the composed model of the RNA molecule. This functionality allows the system to be used for remodelling and adjusting once predicted 3D structure in a semi-automatic way. Currently, RNAComposer is the most popular system for predicting RNA 3D structure. It has been used worldwide in over 5 million prediction tasks run by over 20 thousand users. There are two equal instances of the system, one available at <http://rnacomposer.cs.put.poznan.pl/> and the other at <http://rnacomposer.ibch.poznan.pl/>.

RNAlyzer dates back to 2013 [30]. This system aims to evaluate RNA 3D models by their comparison to the reference structure. The tertiary structures of both the assessed model and the reference (target) are uploaded by the user. The evaluation is based on the local and global similarity of the model to the target. By default, the similarity is estimated upon root mean square deviation (RMSD) computed between the corresponding sets of superimposed atoms in the three-dimensional Euclidean space [26]. The other similarity measures are also available, including Deformation Index (DI), and Interaction Network Fidelity (INF) [39]. The general concept behind RNAlyzer is based on a local neighborhood which is defined as a set of atoms found in the sphere centered around an atom of the user-defined type. The idea derives from similar solutions applied for assessing protein models submitted in CASP (the Critical Assessment of protein Structure Prediction) experiment [2, 36]. The structure comparison procedure is parameterized according to the sphere center, radius length, and precision. It guides the sphere along RNA chain, stopping it at successively fixed points which determine the position of the sphere's center. At each stop, an analysis of atoms within the sphere is carried out. The average distance of these atoms from their equivalents in the sphere moving along the target structure is computed. As a result, the user receives a list of RMSD values together with a multidimensional visualization. A visualization of the model-reference structure distance at different precision levels, from the individual residues to the entire molecules, is an exceptional feature of the system. It helps to identify structure fragments of high and low similarity easily. RNAlyzer proved its usefulness in the discrimination between the correct and incorrect RNA 3D structure models. It is the only system in RNApolis that works as a standalone application. It can be downloaded from <http://rnalyzer.cs.put.poznan.pl/>. However, in response to users' requests, we have also prepared a web server version of the system. It was published under the name **RNAAssess** in 2015 [31]. The computational engine of RNAAssess is the same as in RNAlyzer. A new feature is a convenient and user-friendly interface and the ease of displaying multidimensional graphs and maps representing the results of computation directly in the browser. RNAAssess is available at <http://rnassess.cs.put.poznan.pl/>.

RNApdbee 1.0 was published in 2014 [3]. In 2018, we have released version 2.0 of this webserver tool [59]. Primarily, RNApdbee was developed as the program to annotate the secondary structure of RNA from the 3D structure data saved in PDB files. In the following years, the needs of our users motivated us to expand the capabilities of this tool significantly. Today, RNApdbee is a multifunctional system for annotating RNA secondary structures, translating between different data formats, classifying pseudoknots and non-canonical base pairs in the secondary structure, comparing different state-of-the-art methods for base pair identification, determining the secondary structure topology, revealing a consensus secondary structure. RNApdbee runs in several usage scenarios. In the basic one, 3D-to-(...), the secondary structure is derived from the input data provided in PDB or mmCIF format. The second scenario, 2D-to-(...), accepts data in BPSEQ and CT files and provides a dot-bracket representation of the structure encoded in the input. In (...)-to-image scenario, the user can enter the secondary structure encoded in dot-bracket format and display its graphical view. Finally, the 3D-to-multi 2D scenario allows to run all the incorporated algorithms for base pair identification and structure topology encoding, and compare the results. The latter scenario allows finding a consensus secondary structure, which is not apparent especially in the case of large and complex RNA structures. It is essential that RNApdbee can visualize the secondary structure using several programs that

involve different algorithms to represent and draw the RNA topology. RNApdbee has already served in over 12 thousand runs. It can be accessed at <http://rnapdbee.cs.put.poznan.pl/>.

RNAfitme was launched in 2018 [6]. The system accepts input 3D structure data and modifies them according to the user-selected mode, i.e. nucleobase or nucleoside residue remodelling/reconstruction. The input structure can or cannot be complete (it must contain the backbone atoms which are not modified during data processing). In the first case, RNAfitme remodels the provided structure by changing either all or user-selected nucleobases/nucleotide residues. Remodelling consists in replacing selected elements with their equivalents from the built-in library of conformers, best suited to the rest of the structure. If a user has entered an incomplete RNA structure, the system reconstructs it by adding missing elements that it retrieves from the conformer library. Five libraries of conformers are available in RNAfitme [58]. The difference between them lies in the initial set of structures which the conformers come from, as well as similarity measures and methods used for their clustering. An important fact is that RNAfitme optimizes the modified RNA 3D structure concerning the Clash score, to avoid clashes between remodeled atoms [15]. As a result, it provides a complete, stereochemically correct model of RNA 3D structure. Currently, the number of the system executions counted from the release date reaches 600. RNAfitme is available at <http://rnafitme.cs.put.poznan.pl/>.

RNAvista is the youngest in the RNApolis platform. It was published in 2019 [7]. RNAvista aims to predict an extended secondary structure of RNA, that is the structure that includes both canonical and non-canonical base pairs. The reason to develop such a tool was that most existing methods for RNA secondary structure prediction could anticipate canonical base pairs only [23]. With the growing understanding of the significance of non-canonical interactions in RNAs, it was essential to have the method that could provide some clues about their presence in the structure as well. The concept behind RNAvista is based on a combination of selected functionalities of two previously described tools, RNAComposer and RNApdbee [48]. The prediction itself is a two-stage process. First, the 3D structure model is anticipated by RNAComposer-based procedure from sequence or the canonical secondary structure. The model is composed of experimentally determined structure fragments that already include non-canonical interactions. In the second stage, RNApdbee-based procedure proceeds to annotate the extended secondary structure from the 3D model. Since its publication, the system was tried 400 times already. RNAvista is available at <http://rnavista.cs.put.poznan.pl/>.

2.2. Sequence-based computing

One of the objectives of RNApolis is to enable the researchers to determine the relationship between RNA sequence, secondary and tertiary structure. Therefore, let us look at how the platform's tools can facilitate studying these relationships.

Having a given RNA sequence one can ask what shape the structure of such a sequence takes. This question can be answered by RNA FRABASE [42, 43]. The RNA sequence (with length >3) entered by the user becomes a pattern that is searched for in the database. The search engine of RNA FRABASE is based on the exact matching algorithm. If in any RNA structure deposited in the database there is a sequence exactly matching the pattern, this structure is added to the result list. Consequently, the user receives a list of RNAs

containing the searched sequence. For each found RNA (or a part of it) a secondary and tertiary structure is shown, both encoded in text format and visualized. The user can get acquainted with all the structural parameters that characterize the molecule (or its fragment) with the sequence equal to the searched one.

However, it may happen that none of the experimentally determined RNA structures contains the desired sequence. Then, the systems for predicting the structure come in handy. RNApolis offers two such systems, RNAComposer [4, 44] and RNAvista [7]. Both systems accept an input sequence in FASTA format, no longer than 500 nucleotides. A user interested primarily in the secondary structure should run RNAvista. Based on the input sequence of RNA this system predicts its secondary structure. The resulting structure is visualized and encoded in BPSEQ, CT and dot-bracket formats. As RNAvista was designed to take into account non-canonical interactions also, they are specially treated. All predicted non-canonical base pairs are listed separately, together with an annotation about the class to which they belong. Even in the diagram of the secondary structure, the pictograms defining the base pair classes, according to Leontis-Westhof classification [28], are drawn.

Users who are interested in how an RNA with a given sequence folds, should run RNAComposer. This system performs sequence-based prediction of RNA three-dimensional structure and provides detailed information on the resulting model. The data includes atom coordinates, total energy, structure elements from which the model was assembled, visualization of the model, and the secondary structure in dot-bracket format.

All tools listed here also accept (optionally) additional input data and allow the user to configure the system to suit his/her needs. Detailed information can be found on the Help page of each system. A general scheme showing the scenarios of the use of RNApolis tools for input data in the form of a sequence is shown in Figure 2.

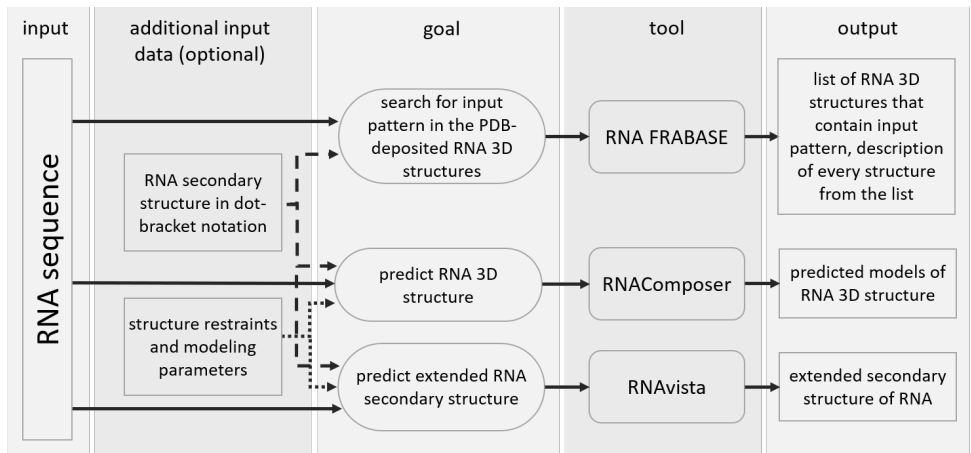


Figure 2. Sequence-based data processing in RNApolis

2.3. Secondary structure-based computing

This section presents possible scenarios for the use of RNApolis tools when the user knows the secondary structure of an RNA (Figure 3). Part of the computations performed on the basis of the secondary structure is similar to those described in Section 2.2. Therefore, having a secondary structure, it is possible to check whether such a structure pattern appears in RNA molecules that were experimentally solved. It can be done with RNA FRABASE [42, 43]. Note that the pattern of the secondary structure that is looked for, may or may not contain a sequence. In the latter case, the RNA FRABASE search engine identifies all fragments with a given secondary structure encoded in dot-bracket notation (e.g. apical loops, bulges, n-way junctions, etc.) having any sequence. The structural pattern in the dot-bracket representation is matched exactly. This allows the relationship between the secondary and tertiary structure to be precisely defined and analysed.

The secondary structure may also be the input to the RNAvista system [7]. Such use of RNAvista is quite unusual, but not meaningless. If the secondary structure held by the user is determined solely by canonical base pairs, RNAvista can be used to enrich it with non-canonical ones. Let us recall that RNAvista predicts an extended secondary structure of RNA. This task can be accomplished based on both the sequence as well as the canonical secondary structure provided at the input.

RNAComposer [4, 44] can predict the 3D model of RNA starting from both the sequence and the secondary structure represented in dot-bracket notation. In fact, the main algorithm of the system is based on the secondary structure and its fragmentation into structural motifs. Thus, predicting the 3D shape of an RNA molecule upon the information about base pairs is a procedure assumed by the system developers. The advantage of secondary structure-based prediction is that RNAComposer can model a structure containing pseudoknots if they are defined at the input. If RNAComposer prediction starts from the sequence only, the output models do not contain pseudoknots. Therefore, the system's authors recommend prediction based on a secondary structure as more reliable for large RNA structures in which there is a high probability of multiple pseudoknots.

Finally, let us note that the secondary structure can also be processed in RNApdbee [3, 59]. The tool allows entering data in BPSEQ or CT format and transforms it into the dot-bracket representation (see 2D-to-(....) scenario). This task is unambiguous and straightforward in the case of structures that do not contain pseudoknots. However, for entangled RNAs, the RNApdbee algorithm must analyse which of the conflicting base pairs are primary pairs and which ones form pseudoknots. In the case of complex structures, the tool also classifies pseudoknots by identifying their order. The pseudoknot order is reflected in the dot-bracket notation (by the corresponding type of bracket). It is also marked with the appropriate colour on the structure drawing. RNApdbee creates graphical representations of the input secondary structure introduced in BPSEQ, CT and dot-bracket (see (....)-to-image scenario) formats. According to the user's choice, the structure can be drawn in the form of a typical diagram or an arc graph.

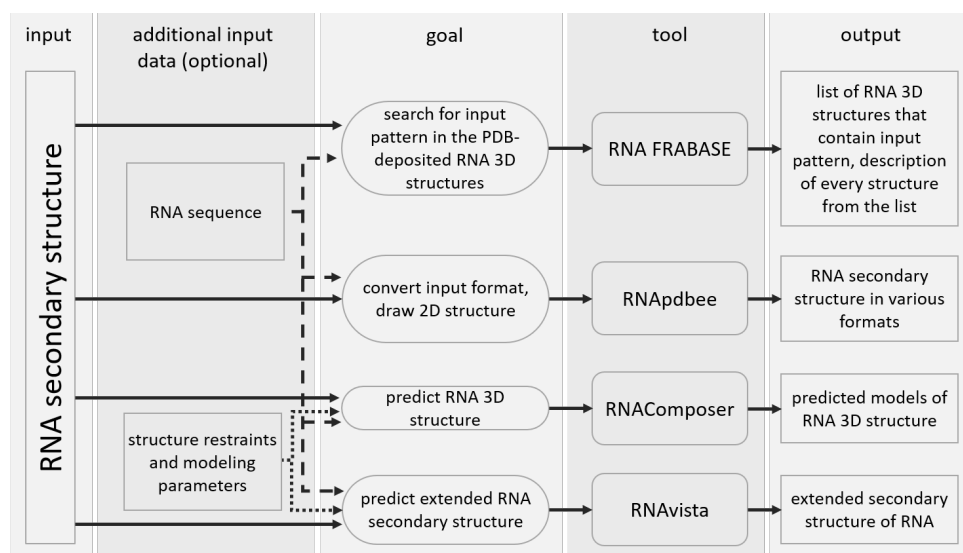


Figure 3. Secondary structure-based data processing in RNAPolis

2.4. Tertiary structure-based computing

Analysis of the tertiary structure is a demanding task. On the one hand, it seems that the knowledge of the RNA structure at this level allows us to answer every question about the molecule architecture. On the other side, the 3D structure is merely a collection of atoms. Only the identification and appropriate interpretation of the relationships between the elements of this set allows for a detailed study on RNA.

RNAPolis provides four bioinformatics tools to process the 3D structure of RNA. Let us start their brief description from RNAPdbec [3, 59] that facilitates studying the relationship between the tertiary and secondary structure. The main task of this tool is to annotate the secondary structure from the data on the three-dimensional shape of a molecule. In two scenarios dedicated to this task, 3D-to-(....) and 3D-to-multi 2D, RNAPdbec accepts input data in PDB and mmCIF formats [10, 22]. Input data are first processed by methods which, on the basis of the geometry (distances between the specific subset of atoms), identify base pairs in the structure. These methods often give different results, especially for large and complex structures. Therefore, the user can select one or run all of them and compare the results. The latter option is available in the 3D-to-multi 2D scenario. After base pair identification, the pseudoknots are analyzed, and the secondary structure topology determined. The output is both textual and graphical.

An essential issue in structural bioinformatics is the assessment of the similarity of 3D structures. This problem occurs whenever we compare different models of the same molecule or evaluate the predicted RNA structures against the target. RNAPolis proposes to use one of the two systems for assessing RNA 3D structure similarity, RNALyzer [30] and RNAssess [31]. Both work on the same principle, except that RNALyzer is a standalone application and RNAssess has been implemented as a webservice. Input data must contain at

least two RNA structures, one of which is a reference. The accepted format is PDB file format. In addition to the compared structures, the user determines the computation parameters: similarity measure (RMSD, DI, INF), sphere radius, sphere center and comparison mode (all atoms/selected subset). At the output, the user receives values of similarity measures in a global and local scale, as well as various visualizations showing the structural similarity depending on the size of the sphere and the precision of computation.

Finally, RNApolis offers a tool for RNA 3D structure reconstruction and remodelling. RNAfitme [6] works on the 3D structure input and outputs the optimized, complete 3D structure of the RNA molecule. The input data should be uploaded in the PDB format. The output data format is the same. Depending on a mode, input data should include coordinates of the backbone atoms, O5',C5',C4',C3',O3' (in the nucleoside residue mode), or backbone and sugar ring atoms, O5',C5',C4',C3',O3',O4',C1',C2',O2' (in the nucleobase mode). All the remaining atoms are reconstructed by RNAfitme. Reconstruction usually requires knowledge of the sequence that the user enters in FASTA format [40]. In addition to reconstructing a full-atom model based on backbone and sequence, RNAfitme is also used to remodel the complete RNA structure. In this case, two scenarios are possible. In the first one, the user inputs only the 3D structure and starts computation. The program tries to optimize the structure by removing clashes between atoms without changing the sequence. The second scenario makes it possible to modify selected nucleobases. To initiate such modifications, the user introduces a sequence in which the upper case letters indicate nucleobases to be remodeled while lowercase letters represent nucleobases that should remain intact. The output structure is optimized concerning the clash score, torsional angle values, and bond geometry.

The above-mentioned scenarios of processing the RNA tertiary structure within the RNApolis platform are schematically shown in Figure 4.

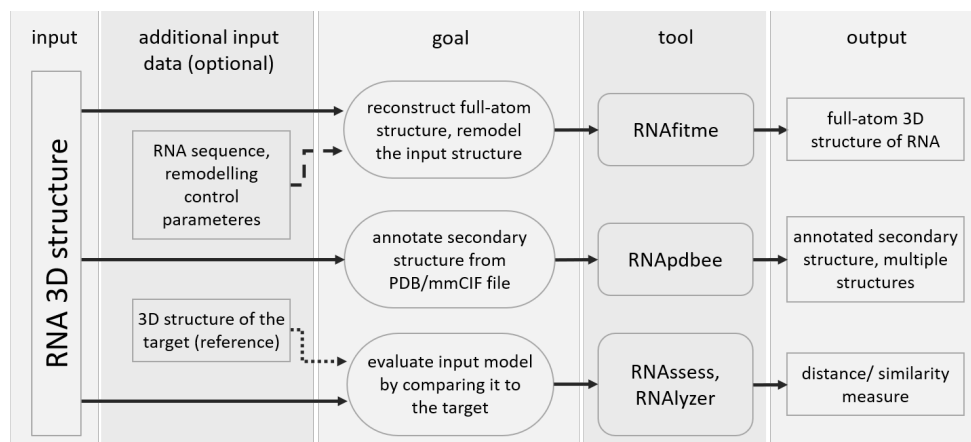


Figure 4. Tertiary structure-based data processing in RNApolis

3. Implementation

RNapolis (<http://rnapolis.pl>) provides a direct access to the latest versions of the following systems developed by our team: RNA FRABASE, RNAComposer, RNApdbee, RNAfitme, RNAvista, RNAnalyzer, and RNAssess. The platform is designed to work with most of the available web browsers: Microsoft Internet Explorer, Mozilla Firefox, Google Chrome, Opera. The latest versions of browsers are strongly recommended. It is platform independent and can also operate on mobile devices. The systems available within RNapolis work independently of each other. They have been implemented in a two-layer architecture with distinguished computational backend and web accessible frontend. All their components are based on Java with third-party libraries (i.a., BioJava [45]). RNapolis web service is hosted and maintained by the Institute of Computing Science, Poznan University of Technology, Poland.

4. Conclusions

The paper introduces RNapolis, a new computational platform that currently provides direct access to seven computational tools developed for RNA structural biology. The resources, as well as RNapolis itself, have a user-friendly interface and are easily operated via all common web browsers. The main tasks carried out by the RNapolis tools are as follows: searching for structure fragments and motifs in PDB-deposited RNA 3D structures, predicting the extended secondary and the tertiary structure, remodelling and refinement of RNA 3D structure, validating and evaluating of 3D structure models in three-dimensional Euclidean space, annotating the secondary structures of RNAs, including large, complex structures with pseudoknots and non-canonical interactions, comparison of various methods for RNA secondary structure annotation and visualization, processing and transforming between multiple formats for RNA secondary structure representation, secondary structure visualization with annotation of pseudoknot orders and classes of non-canonical base pairs. RNapolis is a newly launched project. In the future, its resources will be increased by further data and tools for RNA bioinformatics researchers. Notably, the platform will be enriched, among other things, with systems analysing the structure in the space of torsional angles [57], searching for similar fragments [54], supporting the study of nucleic acids quadruplexes, and identifying a consensus structure in a set of multiple RNA models. The ambition of the author of the RNapolis project is for the platform to become a virtual laboratory for the analysis of RNA structures enabling multidirectional bioinformatics research.

Acknowledgements

The author was supported by the National Science Centre, Poland, grant no. 2016/23/B/ST6/03931. My thanks go to Marcin Borowski for the preparation of RNApolis website. I acknowledge Maciej Antczak, Mariusz Popenda, and Tomasz Zok - the leading developers of bioinformatics tools offered within the platform, as well as Ryszard W. Adamiak and Jacek Blazewicz who have supported me in the field of RNA bioinformatics research for many years.

References

- [1] Adamiak R.W., Blazewicz J., Formanowicz P., Gdaniec Z., Kasprzak M., Popenda M., Szachniuk M., An algorithm for an automatic NOE pathways analysis of 2D NMR spectra of RNA duplexes, *Journal of Computational Biology*, **11**, 2004, 163-180.
- [2] Antczak M., Blazewicz J., Lukasiak P., Milostan M., Krasnogor N., Palik G., DomAns-Pattern based method for protein domain boundaries prediction and analysis, *Foundations of Computing and Decision Sciences*, **36**, 2011, 99-119.
- [3] Antczak M., Zok T., Popenda M., Lukasiak P., Adamiak R.W., Blazewicz J., Szachniuk M., RNApdbee - a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs, *Nucleic Acids Research*, **42**, 2014, W368-W372.
- [4] Antczak M., Popenda M., Zok T., Sarzynska J., Ratajczak T., Tomczyk K., Adamiak R.W., Szachniuk M., New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure, *Acta Biochimica Polonica*, **63**, 2016, 737-744.
- [5] Antczak M., Popenda M., Zok T., Zurkowski M., Adamiak R.W., Szachniuk M., New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation, *Bioinformatics*, **34**, 2018, 1304-1312.
- [6] Antczak M., Zok T., Osowiecki M., Popenda M., Adamiak R.W., Szachniuk M., RNAfitme: a webserver for modeling nucleobase and nucleoside residue conformation in fixed-backbone RNA structures, *BMC Bioinformatics*, **19**, 2018, 304.
- [7] Antczak M., Zablocki M., Zok T., Rybarczyk A., Blazewicz J., Szachniuk M., RNAvista: a webserver to assess RNA secondary structures with non-canonical base pairs, *Bioinformatics*, **35**, 2019, 152-155.
- [8] Backofen R., Engelhardt J., Erxleben A., Fallmann J., Grüning B., Ohlerd U., Rajewsky N., Stadler P.F., RNA-bioinformatics: Tools, services and databases for the analysis of RNA-based regulation, *Journal of Biotechnology*, **261**, 2017, 76-84.
- [9] Benson D., Karsch-Mizrachi I., Lipman D., Ostell J., Wheeler D., Genbank, *Nucleic Acids Research*, **35**, 2007, D21-D25.
- [10] Berman H.M., The protein data bank: a historical perspective, *Acta Crystallographica Section A*, **64**, 2007, 88-95.

- [11] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E., The Protein Data Bank, *Nucleic Acids Research*, **28**, 2000, 235-242.
- [12] Bhagat J., Tanoh F., Nzuobontane E., Laurent T., Orlowski J., Roos M., Wolstencroft K., Aleksejevs S., Stevens R., Pettifer S., Lopez R., Goble C.A., BioCatalogue: a universal catalogue of web services for the life sciences, *Nucleic Acids Research*, **38**, 2010, 689-694.
- [13] Blazewicz J., Figlerowicz M., Kasprzak M., Nowacka M., Rybarczyk A., RNA Partial Degradation Problem: Motivation, Complexity, Algorithm, *Journal of Computational Biology*, **18**, 2011, 821-834.
- [14] Brion P., Westhof E., Hierarchy and dynamics of RNA folding, *Annual Review of Biophysics and Biomolecular Structure*, **26**, 1997, 113-137.
- [15] Chen V.B., Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S., Richardson D.C., MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallographica. Section D , Biological crystallography*, **66**, 2010, 12-21.
- [16] Chen L., Heikkinen L., Wang C.L., Yang Y., Knott K.E., Wong G., miRToolsGallery: A microRNA bioinformatics resources database portal, *Database (Oxford)*, 2018, bay004.
- [17] Cruz J.A., Blanchet M.-F., Boniecki M., Bujnicki J.M., Chen S.-J., Cao S., Das R., Ding F., Dokholyan N.V., Flores S.C., Huang L., Lavender C.A., Lisi V., Major F., Mikolajczak K., Patel D.J., Philips A., Puton T., Santalucia J., Sijenji F., Hermann T., Rother K., Rother M., Serganov A., Skorupski M., Soltysinski T., Sripakdeevong P., Tuszynska I., Weeks K.M., Waldsich C., Wildauer M., Leontis N.B., Westhof E., RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction, *RNA*, **18**, 2012, 610-625.
- [18] Danaee P., Rouches M., Wiley M., Deng D., Huang L., Hendrix D., bpRNA: large-scale automated annotation and analysis of RNA secondary structure, *Nucleic Acids Research*, **46**, 2018, 5381-5394.
- [19] Dawson W.K., Bujnicki J.M., Computational modeling of RNA 3D structures and interactions, *Current Opinion in Structural Biology*, **37**, 2016, 22-28.
- [20] Deigan K.E., Li T.W., Mathews D.H., Weeks K.M., Accurate SHAPE-directed RNA structure determination, *Proceedings of National Academy of Sciences USA*, **106**, 2009, 97-102.
- [21] Gudanis D., Popena L., Szpotkowski K., Kierzek R., Gdaniec Z., Structural characterization of a dimer of RNA duplexes composed of 8-bromoguanosine modified CGG trinucleotide repeats: a novel architecture of RNA quadruplexes, *Nucleic Acids Research*, **44**, 2016, 2409-2416.
- [22] Hall S.R., Allen F.H., Brown I.D., The Crystallographic Information File (CIF): a new standard archive file for crystallography, *Acta Crystallographica*, **A47**, 1991, 655-685.
- [23] Honer zu Siederdissen C., Bernhart S.H., Stadler P.F., Hofacker I.L., A folding algorithm for extended RNA secondary structures, *Bioinformatics*, **27**, 2011, i129-i136.

- [24] IUPAC-IUB Commission on Biochemical Nomenclature, Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents, *Biochemistry*, **9**, 1970, 4022-4027.
- [25] Johnson A.D., An extended IUPAC nomenclature code for polymorphic nucleic acids, *Bioinformatics*, **26**, 2010, 1386-1389.
- [26] Kabsch W., A solution for the best rotation to relate two sets of vectors, *Acta Crystallographica*, **A32**, 1976, 922-923.
- [27] Kulikova T., Akhtar R., Aldebert P., Althorpe N., Andersson M., Baldwin A., Bates K., Bhattacharyya S., Bower L., Browne P., Castro M., Cochrane G., Duggan K., Eberhardt R., Faruque N., Hoad G., Kanz C., Lee C., Leinonen R., Lin Q., Lombard V., Lopez R., Lorenc D., McWilliam H., Mukherjee G., Nardone F., Pastor M.P., Plaister S., Sobhany S., Stoehr P., Vaughan R., Wu D., Zhu W., Apweiler R., EMBL nucleotide sequence database in 2006, *Nucleic Acids Research*, **35**, 2007, D16-D20.
- [28] Leontis N.B., Westhof E., Geometric nomenclature and classification of RNA base pairs, *RNA*, **7**, 2001, 499-512.
- [29] Lorenz R., Bernhart S.H., Höner zu Siederdissen C., Tafer H., Flamm C., Stadler P.F., Hofacker I.L., ViennaRNA Package 2.0, *Algorithms for Molecular Biology*, **6**, 2011, 26.
- [30] Lukasiak P., Antczak M., Ratajczak T., Bujnicki J.M., Szachniuk M., Popenda M., Adamiak R.W., Blazewicz J., RNAnalyzer – novel approach for quality analysis of RNA structural models, *Nucleic Acids Research*, **41**, 2013, 5978-90.
- [31] Lukasiak P., Antczak M., Ratajczak T., Szachniuk M., Popenda M., Adamiak R.W., Blazewicz J., RNAssess - a webserver for quality assessment of RNA 3D structures, *Nucleic Acids Research*, **43**, 2015, W502-W506.
- [32] Mathews D.H., Disney M.D., Childs J.L., Schroeder S.J., Zuker M., Turner D.H., Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proceedings of National Academy of Sciences USA*, **101**, 2004, 7287-7292.
- [33] Mathews D.H., Turner D.H., Prediction of RNA secondary structure by free energy minimization, *Current Opinion in Structural Biology*, **16**, 2006, 270-278.
- [34] Miao Z., Westhof E., RNA Structure: Advances and Assessment of 3D Structure Prediction, *Annual Review of Biophysics*, **46**, 2017, 483-503.
- [35] Miskiewicz J., Szachniuk M., Discovering structural motifs in miRNA precursors from Viridiplantae kingdom, *Molecules*, **23**, 6, 2018, 1367.
- [36] Moult J., Fidelis K., Kryshtafovych A., Schwede T., Tramontano A., Critical assessment of methods of protein structure prediction (CASP)-Round XII, *Proteins*, **86**, 2018, 7-15.
- [37] Narayanan B.C., Westbrook J., Ghosh S., Petrov A.I., Sweeney B., Zirbel C.L., Leontis N.B., Berman H.M., The Nucleic Acid Database: new features and capabilities, *Nucleic Acids Research*, **42**, 2014, D114–D122.

- [38] Pang P.S., Elazar M., Pham E.A., Glenn J.S., Simplified RNA secondary structure mapping by automation of SHAPE data analysis, *Nucleic Acids Research*, **39**, 2011, e151.
- [39] Parisien M., Cruz J.A., Westhof E., Major F., New metrics for comparing and assessing discrepancies between RNA 3D structures and models, *RNA*, **15**, 2009, 1875-1885.
- [40] Pearson W.R., Lipman D.J., Improved tools for biological sequence comparison, *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 1988, 2444-2448.
- [41] Popenda L., Bielecki L., Gdaniec Z., Adamiak R.W., Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C:G-A triple, *Arkivoc: Archive for Organic Chemistry*, **3**, 2009, 130-144.
- [42] Popenda M., Blazewicz M., Szachniuk M., Adamiak R.W., RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures, *Nucleic Acids Research*, **36**, 2008, D386-D391.
- [43] Popenda M., Szachniuk M., Blazewicz M., Wasik S., Burke E.K., Blazewicz J., Adamiak R.W., RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures, *BMC Bioinformatics*, **11**, 2010, 231.
- [44] Popenda M., Szachniuk M., Antczak M., Purzycka K.J., Lukasiak P., Bartol N., Blazewicz J., Adamiak R.W., Automated 3D structure composition for large RNAs, *Nucleic Acids Research*, **40**, 2012, e112.
- [45] Prlic A., Yates A., Bliven S.E., Rose P.W., Jacobsen J., Troshin P.V., Chapman M., Gao J., Koh C.H., Foisy S., Holland R., Rimsa G., Heuer M.L., Brandstätter-Müller H., Bourne P.E., Willis S., BioJava: an open-source framework for bioinformatics in 2012, *Bioinformatics*, **28**, 2012 2693–2695.
- [46] Purzycka K.J., Popenda M., Szachniuk M., Antczak M., Lukasiak P., Blazewicz J., Adamiak R.W., Automated 3D RNA structure prediction using the RNAComposer method for riboswitches, in: S.-J. Chen, D.H. Burke-Aguero (eds.), *Methods in Enzymology: Computational Methods for Understanding Riboswitches*, **553**, Elsevier, 2014, 3-34.
- [47] Puton T., Kozłowski L.P., Rother K.M., Bujnicki J.M., CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction, *Nucleic Acids Research*, **41**, 2013, 4307-4323.
- [48] Rybarczyk A., Szostak N., Antczak M., Zok T., Popenda M., Adamiak R.W., Blazewicz J., Szachniuk M., New in silico approach to assessing RNA secondary structures with non-canonical base pairs, *BMC Bioinformatics*, **16**, 2015, 276.
- [49] Seetin M.G., Mathews D.H., RNA structure prediction: an overview of methods, *Methods of Molecular Biology*, **905**, 2012, 99-122.
- [50] Stevens R.D., Robinson A.J., Goble C.A., MyGrid: Personalised bioinformatics on the information grid, *Bioinformatics*, **19**, 2003, i302-i304.

-
- [51] Sugawara H., Ogasawara O., Okubo K., Gojobori T., Tateno Y., Ddbj with new system and face, *Nucleic Acids Research*, **36**, 2008, D22-D24.
- [52] Szachniuk M., Assigning NMR Spectra of Irregular RNAs by Heuristic Algorithms, *Bulletin of the Polish Academy of Sciences Technical Sciences*, **63**, 2015, 329-338.
- [53] Turner D.H., Mathews D.H., *RNA Structure Determination: Methods and Protocols*, Springer, New York, 2016.
- [54] Wiedemann J., Zok T., Milostan M., Szachniuk M., LCS-TA to identify similar fragments in RNA 3D structures, *BMC Bioinformatics*, **18**, 2017, 456.
- [55] Wojciechowski P., Frohmberg W., Kierzyńska M., Zurkowski P., Blazewicz J., G-MAPSEQ—a new method for mapping reads to a reference genome, *Foundations of Computing and Decision Sciences*, **41**, 2016, 123-142.
- [56] wwPDB consortium, Protein Data Bank: the single global archive for 3D macromolecular structure data, *Nucleic Acids Research*, **47**, 2019, D520–D528.
- [57] Zok T., Popenda M., Szachniuk M., MCQ4Structures to compute similarity of molecule structures, *Central European Journal of Operations Research*, **22**, 2014, 457-474.
- [58] Zok T., Antczak M., Riedel M., Nebel D., Villmann T., Lukasiak P., Blazewicz J., Szachniuk M., Building the library of RNA 3D nucleotide conformations using clustering approach, *International Journal of Applied Mathematics and Computer Science*, **25**, 2015, 689-700.
- [59] Zok T., Antczak M., Zurkowski M., Popenda M., Blazewicz J., Adamiak R.W., Szachniuk M., RNApdbec 2.0: multifunctional tool for RNA structure annotation, *Nucleic Acids Research*, **46**, 2018, W30-W35.

Received 7.01.2019, Accepted 15.02.2019