

RESEARCH ARTICLE

RNATHOR – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis

Julita Gumna¹, Tomasz Zok², Kacper Figurski², Katarzyna Pachulska-Wieczorek^{1*}, Marta Szachniuk^{1,2*}

1 Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, **2** Institute of Computing Science, Poznan University of Technology, Poznan, Poland

* kasiapw@ibch.poznan.pl (KPW); mszachniuk@cs.put.poznan.pl (MS)



OPEN ACCESS

Citation: Gumna J, Zok T, Figurski K, Pachulska-Wieczorek K, Szachniuk M (2020) RNATHOR – fast, accurate normalization, visualization and statistical analysis of RNA probing data resolved by capillary electrophoresis. PLoS ONE 15(10): e0239287. <https://doi.org/10.1371/journal.pone.0239287>

Editor: Danny Barash, Ben-Gurion University, ISRAEL

Received: April 7, 2020

Accepted: September 3, 2020

Published: October 1, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0239287>

Copyright: © 2020 Gumna et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Abstract

RNAs adopt specific structures to perform their functions, which are critical to fundamental cellular processes. For decades, these structures have been determined and modeled with strong support from computational methods. Still, the accuracy of the latter ones depends on the availability of experimental data, for example, chemical probing information that can define pseudo-energy constraints for RNA folding algorithms. At the same time, diverse computational tools have been developed to facilitate analysis and visualization of data from RNA structure probing experiments followed by capillary electrophoresis or next-generation sequencing. RNATHOR, a new software tool for the fully automated normalization of SHAPE and DMS probing data resolved by capillary electrophoresis, has recently joined this collection. RNATHOR automatically identifies unreliable probing data. It normalizes the reactivity information to a uniform scale and uses it in the RNA secondary structure prediction. Our web server also provides tools for fast and easy RNA probing data visualization and statistical analysis that facilitates the comparison of multiple data sets. RNATHOR is freely available at <http://rnathor.cs.put.poznan.pl/>.

Introduction

Structural features are of importance for the biological functions of RNA molecules. Specific RNA structures are recognized by RNA binding proteins, ligands, and other RNAs—these interactions impact almost every aspect of cell life or viral replication. Therefore, there is a great interest in developing novel approaches for proper and rapid RNA structure modeling. The computational methods enable the obtaining of good quality models of short RNAs based on sequence only, but the accuracy of structure prediction decreases with the length of RNA molecules [1–5]. The inclusion of RNA structure probing data as pseudo-energy constraints into the thermodynamic folding algorithms significantly improves the accuracy of RNA structure prediction [6, 7]. Among chemical and enzymatic methods, SHAPE (selective 2'-hydroxyl

Funding: This work was funded by the National Science Centre Poland (<https://ncn.gov.pl/?language=en>) in the form of grants awarded to MS (2016/23/B/ST6/03931, 2019/35/B/ST6/03074) and KPW (2016/22/E/NZ3/00426). Funding for the open access charge was provided by the statutory funds of Poznan University of Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

acylation analyzed by primer extension) [8] and DMS (dimethyl sulfate) mapping [9] are the best validated and most widely used techniques of RNA structure probing *in vitro* and *in vivo* [10, 11]. Besides, the pipelines of SHAPE and DMS probing data incorporation into RNA structure prediction software are well established [6, 12]. DMS modifies the Watson–Crick edge of unpaired adenosines or cytosines, whereas SHAPE reagents create covalent adducts at the 2'-OH group on the RNA sugar ring in a flexibility-sensitive manner [8, 9, 13]. Several SHAPE reagents that differ in their half-life and solubility have been developed until now [14–16]. They act independently from nitrogen base and, consequently, one probing reagent can be used instead of a combination of base-specific chemicals.

Effective detection and quantitative measurement of modification sites are critical for all RNA probing experiments. Typically, RNA chemical modification is followed by a reverse transcription to cDNA that is truncated or mutated at the adducts position [8, 17]. The sites of RT stops in the cDNA can be read-out using the capillary electrophoresis (CE) or next-generation sequencing (NGS) but only the second technique can be used for the detection of adduct-induced mutations [17, 18]. The NGS-based techniques allow genome-wide and transcriptome-wide profiling of RNA structure. The CE is widely used for resolving reactivity data from medium- and low-throughput RNA probing experiments. There are many examples of SHAPE-CE usage for analysis of the structure of many important RNAs, including ribosomal RNAs [19, 20], long noncoding RNAs [21–23], viral RNAs [24–30], and retrotransposon RNAs [31–33]. Besides, CE can also be used for the analysis of RNA probing experiments utilizing other chemical reagents such as CMCT, kethoxal, hydroxyl radicals, and RNases [34–37].

The extraction of quantitative data from CE electropherograms is challenging and requires complicated, multistep analysis of fluorescence signals. Several computational tools can process electropherograms from SHAPE-CE experiments [38–42]. Among them, ShapeFinder [41] and QuShape [42] are the most widely used and yield high-quality SHAPE reactivity data for 300–600 nucleotides in one experiment. Before the incorporation of probing information into the thermodynamic RNA folding algorithms, the reactivity values must be normalized to a uniform scale that is valid for diverse RNAs. Additionally, visual inspection of nonspecific RT strong-stops (non-induced by adduct formation) is required.

Normalization and other quality control steps are very important aspects of structure probing data analysis. Therefore, we developed RNAtbor, a user-friendly tool for fast, automatic normalization, and analysis of the CE-based RNA probing data (Fig 1). Features of our tool include (i) normalization of data from several experiments in the box-plot scheme at once, (ii) automatic detection of strong-stops of reverse transcriptase, (iii) reactivity data visualization, (iv) statistical analysis of the results to compare multiple data sets, and (v) RNA secondary structure prediction based on reactivity data.

Materials and methods

RNAtbor workflow

In the RNAtbor workflow, we distinguish five general stages: validation of the input data (ShapeFinder or QuShape file(s) and optionally RNA sequence), exclusion of unreliable data, normalization of probing data, prediction of the secondary structure (optional), and statistical analysis of the normalized data (optional) (Fig 1).

Validation of the input data. Initially, the user-uploaded files, resulting from ShapeFinder or QuShape, are parsed, and the basic validation of their format is executed. If, additionally, a sequence is entered, RNAtbor checks whether it is RNA and whether it is at least as long as the sequence in the input file(s). A positive validation results in the next step of the

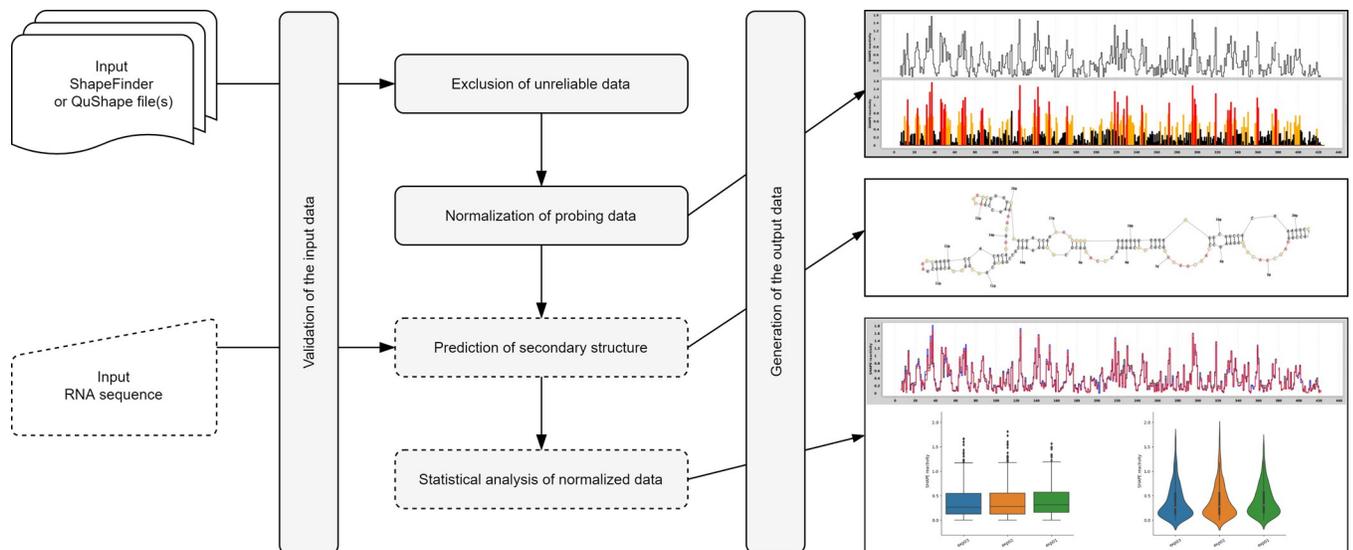


Fig 1. Workflow in the RNAthor system.

<https://doi.org/10.1371/journal.pone.0239287.g001>

computational process. Otherwise, the user receives an error message and is asked to provide correct data.

Exclusion of unreliable data. Unreliable data usually correspond to premature terminations of primer extension reaction due to reasons other than the formation of the adduct (e.g., preexisting cleavage or modification in RNA). These nucleotide positions are called RT strong-stops. RNAthor offers two ways of detecting such data and excluding them from further processing: a fully automated algorithm and an interactive procedure requiring manual selections. The automated procedure (Fig 2) was implemented based on our experience with analyzing data from RNA chemical probing experiments *in vitro* and *in vivo*. It was optimized for the analysis of SHAPE and DMS probing data. It eliminates the data, which meet one of the following criteria: the absolute reactivity value is negative; the background peak area is at least five times larger than the average background peak area; the difference in peak areas between background and reaction is less than 35% of the average background peak area, and the background peak area in this position is equal to or greater than this average. The alternative is a manual procedure, recommended especially for processing the data from RNA probing experiments other than SHAPE or DMS-probing. In this approach, users can identify unreliable data according to their own experience. They define the negative reactivity threshold, and indicate how to treat the negative reactivity values—they can be left as negative values, changed to 0, or marked as no data. RNAthor displays the histogram with peak areas for modification reaction and background for each nucleotide. Based on this view, users manually select RT strong-stops' positions. All identified RT strong-stops are next excluded from the normalization step.

Data normalization. In this step, the data are brought into proportion with one another, and outliers are removed, to provide users with easy to interpret reactivity data on a uniform scale. RNAthor applies the standard box-plot scheme, recommended to normalize the SHAPE-CE data [43, 44]. The normalization process involves: identifying outliers, determining the effective maximum reactivity, and calculating the normalized reactivity values. The initial task is to determine the first (Q1) and the third (Q3) quartile, the interquartile range (IQR), and compute the upper extreme: $UP = Q3 + 1.5(IQR)$. Reactivities greater than UP are considered outliers and not taken into account in subsequent calculations following the principle: for

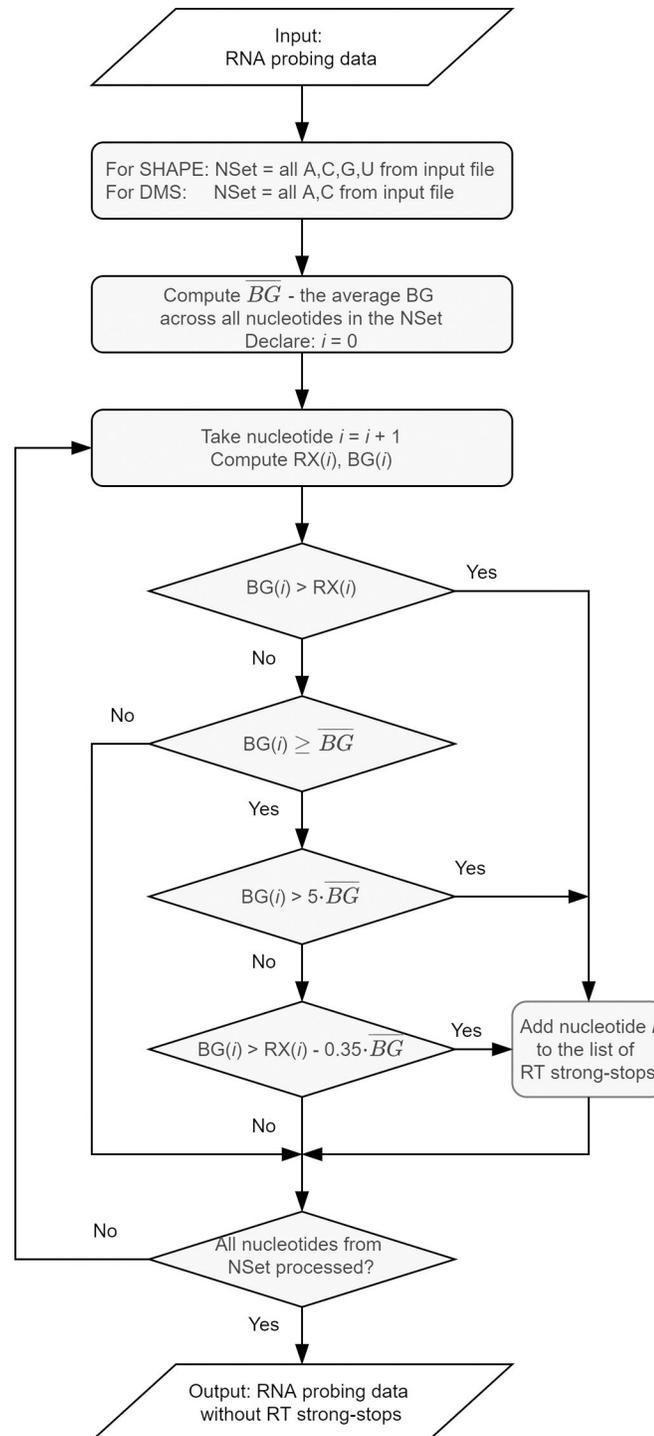


Fig 2. Scheme of the RNAthor algorithm for unreliable data exclusion.

<https://doi.org/10.1371/journal.pone.0239287.g002>

RNAs longer than 100 nucleotides, no more than 10% of the data are identified as outliers; for shorter RNAs, maximum 5% of data are removed. The remaining values are used to compute the effective maximum reactivity, i.e., the average of the top 8% of reactivity values. Finally, all absolute reactivity values are divided by the effective maximum reactivity. It results in

obtaining the normalized reactivity data on a uniform scale. Values close to 0 indicate no reactivity (and highly constrained nucleotides), while values greater than 0.85 correspond to high reactivity (and flexible nucleotides).

Secondary structure prediction. Optionally, users can obtain the secondary structure predicted for the RNA sequence provided at the input. If the sequence is given, RNATHOR automatically executes the incorporated RNAstructure algorithm [45] that supports SHAPE / DMS data-driven prediction. It takes the RNA sequence and the normalized probing data and generates the respective secondary structure. The graphical diagram of the structure is colored according to the color scheme defined for the default reactivity ranges. The output structure is also encoded in the dot-bracket notation.

Statistical analysis of the normalized data. Logged users can perform additional statistical analysis of the normalized probing data. The analysis includes 2–5 experiments selected by the user. It consists of running the Shapiro-Wilk test for normal data distribution, Bartlett test of variance homogeneity, non-parametric Mann-Whitney test (if the user selected 2 experiments), and Kruskal-Wallis rank-sum test (if the user selected 3–5 experiments). Two latter tests are performed if the probing data departure from the normal distribution. As a result of the analysis, users receive numerical, textual, and graphical data—among others, the comparative step plot, the box-and-whisker plot, and the violin plot.

Experimental setup

RNA probing data for the RNATHOR validation were obtained from SHAPE-CE and DMS-CE experiments performed in our laboratory for Ty1 RNA (+1–560). The results of SHAPE-based manual analysis were already published in [32]. The DMS experiment was performed especially for this work; its details are presented below. Electropherograms from SHAPE and DMS probing were processed using ShapeFinder software according to the authors' instructions [41].

For RNA probing with DMS, RNA (8 pmol) was refolded in 30 μ l of renaturation buffer (10 mM Tris-HCl pH 8.0, 100 mM KCl and 0.1 mM EDTA) by heating for 3 minutes at 95°C, slow cooling to 4°C, then adding 90 μ l of water and 30 μ l of 5x folding buffer (final concentration: 40 mM Tris-HCl pH 8.0, 130 mM KCl, 0.2 mM EDTA, 5 mM MgCl₂), followed by incubation for 20 minutes at 37°C. The RNA sample was divided into two tubes and treated with DMS dissolved in ethanol (+) or ethanol alone (-), and incubated at RT for 1 minute. The reaction was quenched by the addition of 14.7 M β -mercaptoethanol. RNA was recovered by ethanol precipitation and resuspended in 10 μ l of water. Primer extension reactions were performed using fluorescently labeled primer [Cy5 (+) and Cy5.5 (-)] as described previously [32]. Sequencing ladders were prepared using primers labeled with WellRed D2 (ddA) and LicorIRD-800 (ddT) and a Thermo Sequenase Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's protocol. Samples were analyzed on a GenomeLab GeXP Analysis System (Beckman-Coulter).

Web application

RNATHOR, implemented as a publicly available web server, has a simple and intuitive interface. It runs on all major web browsers and is accessible at <http://rnathor.cs.put.poznan.pl/>. The web service is hosted and maintained by the Institute of Computing Science, Poznan University of Technology, Poland.

Implementation details

The architecture of RNATHOR comprises two components: the computational engine (backend layer) and the web application (frontend layer). The backend layer, implemented in Java

OpenJDK 8.0, applies selected modules of the Spring Framework: Spring Boot 2.1.6 enables fast configuration of the application; Spring Security ensures user authentication and basic security; Spring MVC allows compatibility with the Model View Controller and Apache Tomcat server; Spring Test, via Junit and Mockito libraries, enables unit tests and integration; Spring Data allows for comprehensive database services, including transaction management. The user interface (frontend layer) is implemented in Angular technology. User data and basic information about the experiments are collected in the PostgreSQL relational database, the input and output data are saved on the server's hard drive. The tool uses the Apache License 2.0.

Input and output description

At the input, RNAthor accepts ShapeFinder or QuShape output files in a tab-delimited text format. Users upload their data via the *New experiment* page by selecting 1–15 files from the local folder. All files in the multiple-input should come from several repetitions of the RNA probing experiment performed for the same RNA. Repetitions increase the reliability of structural data for RNA secondary structure prediction. RNAthor processes all the input files in a single run. It starts after data uploading and setting additional parameters for the normalization process (algorithm for RT strong-stops detection, probing reagent, color settings). Additionally, users can provide an RNA sequence that is used to predict the RNA secondary structure.

RNAthor generates a selection of output data. First of all, users obtain the output file in the SHAPE format (*.shape) that is compatible with the RNAstructure software [45]. The file comprises two columns with nucleotide positions and normalized SHAPE / DMS-CE reactivity data. For the multiple-input, the generated SHAPE file contains averaged reactivities from all normalized data. Nucleotides for which there is no reactivity data are assigned -999 values as recommended in [43]. If the user uploaded the sequence of the analyzed RNA molecule, RNAthor provides the RNA secondary structure in dot-bracket notation and the graphical diagram. Additionally, RNAthor generates files that can facilitate the analysis of RNA probing experiments. One of them is the MS Excel file with spreadsheets containing the input data, the normalized reactivity values, and averaged normalized reactivity data with standard deviation (the average and standard deviation are calculated separately for each nucleotide across samples). Each spreadsheet with the input data contains a histogram, identical to this created during manual removal of RT strong-stops. Rows with normalized reactivity values are colored depending on the user's settings. In the processing of large RNAs, this file can help to combine probing data from overlapping reads (with a different set of primers). RNAthor also prepares a graphical output: step plot and bar plot presenting a reactivity profile for one experiment or averaged data from several repetitions. The bar plot is colored depending on the settings: black for reactivities in [0, 0.4), orange for reactivities in [0.4, 0.85), and red for reactivities in [0.85, ∞) by default. Logged users that run statistical analysis of experimental data also obtain comparative step plot, box-and-whisker plot, violin plot, and summary of test results. The latter one, available for download in .txt file, informs whether the uploaded data come from a normal distribution, whether they have equal variances, what statistical test was performed, and what is the p-value. The comparative step plot shows reactivity profiles of all compared experiments in one chart. The box-and-whisker plot displays the distribution of data based on the position measures, such as quartiles, minimum, and maximum. The violin plot presents the shape of the distribution and probability density of normalized reactivity values. All generated plots can be saved in PNG or EPS format. Users download the output files separately or in a single zipped archive. They can also obtain them as an email attachment—if the email was provided at the input. Additionally, the email contains a unique link to the result page. The results are

stored in the system for 3 days (for guest users) or 3 months (for logged users). Logged users can extend the storage time by an additional month.

Results

RNAtThor allows for efficient, automated processing and analysis of RNA probing data from SHAPE-CE and DMS-CE experiments and their use in data-driven RNA secondary structure prediction. It was tested on multiple datasets, containing data from SHAPE and DMS probing experiments resolved by capillary electrophoresis. The tests confirmed the reliability of the results and showed the utility of the tool. Here, we describe the experiments performed to compare the results of RNA probing data analysis carried out manually by an expert, and automatically by RNAtThor. For the experiments, we chose SHAPE-CE and DMS-CE probing data obtained for RNA of yeast retrotransposon Ty1. The structure of the 5'-end of Ty1 RNA was extensively studied and determined under different experimental conditions and biological states [31, 32, 46].

In the first test, we executed RNAtThor for the ShapeFinder-generated files containing the probing data obtained from three independent replicates of SHAPE experiment (raw data used in this experiment are provided in the [S1 File](#)). We ran RNAtThor with the default settings and the automated algorithm for the identification of RT strong-stops. The generated normalized reactivity data were next compared to the corresponding data published in [32], resulting from manual analysis of the same input. We aligned the obtained bar plots ([Fig 3A](#)), and we computed the correlation between normalized reactivity values ([Fig 3B](#)). In the second test, we repeated the same procedure for data obtained from the DMS experiment (unpublished data; the experiment was carried out for this work especially). “Blind” human experimentalist analyzed the DMS data preprocessed using ShapeFinder, normalized reactivity values, manually identified unreliable data and applied OriginPro to generate the bar plot presenting the reactivity profile. The results of this manual processing were compared to the output generated by RNAtThor that was executed with DMS reagent selected and automated identification of RT strong-stops ([Fig 3C and 3D](#)).

From these experiments, we observe that all RT strong-stops identified manually by the expert are also selected for exclusion by the automatic algorithm implemented in RNAtThor. On the other hand, few data assigned as RT strong-stops by RNAtThor can be considered reliable in the human-dependent analysis. This is due to the rigid criteria for determining RT strong-stops adopted in the algorithm. [Table 1](#) presents the results of the detailed analysis we did by comparing manual, expert-driven, and automatic, RNAtThor-performed detection of unreliable data. We computed basic measures used to evaluate the quality of binary classification: true positives (TP)–data classified as reliable by both expert and RNAtThor, true negatives (TN)–data classified as unreliable by both expert and RNAtThor, false positives (FP)–data indicated as unreliable by the expert but classified as reliable by RNAtThor, false negatives (FN)–data indicated as reliable by the expert but classified as unreliable by RNAtThor. Using these measures, we calculated the accuracy (ACC), sensitivity (TPR, true positive rate), specificity (TNR, true negative rate), and precision (PPV, positive predictive value) of the automatic algorithm implemented in RNAtThor. All these measures were determined for three datasets: SHAPE probing data separately, DMS probing data separately, and data from both sets together. They prove the high quality of the tested algorithm for all datasets. Accuracy and sensitivity equal 0.99, where accuracy, $ACC = (TP+TN)/(TP+TN+FP+FN)$, represents the ratio of correct classifications to the total number of input data, and sensitivity, $TPR = TP/(TP+FN)$, indicates what part of the actual reliable data has been correctly classified by RNAtThor. Specificity and precision are both equal to 1, which is because of $FP = 0$. Specificity, $TNR = TN/(TN$

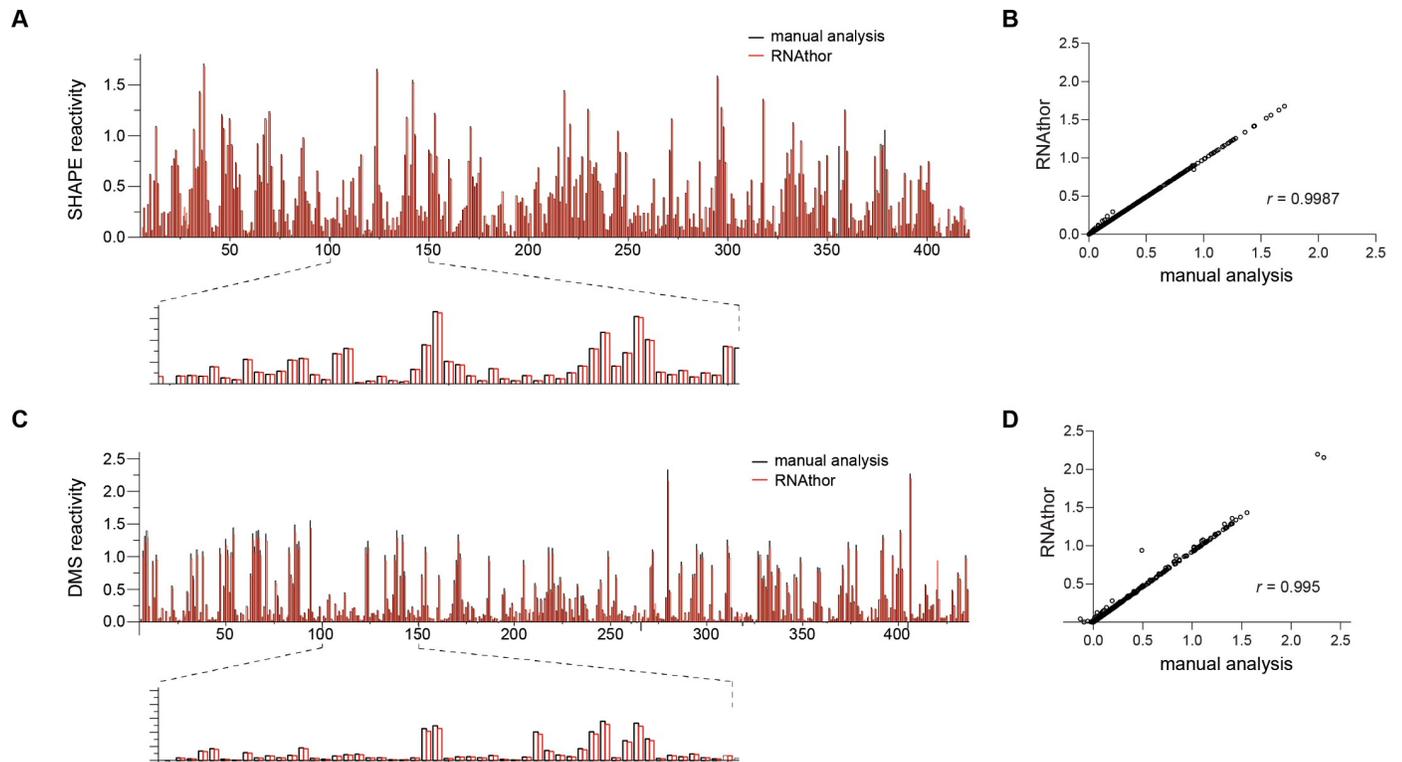


Fig 3. Automatic and manual normalization of RNA probing data. SHAPE (A) and DMS (C) reactivity profiles calculated by RNAtbor (red) and manually (black). Correlation between RNAtbor and manual analysis per nucleotide reactivity estimated for SHAPE (B) and DMS (D).

<https://doi.org/10.1371/journal.pone.0239287.g003>

+FP) is a fraction of correctly classified unreliable data, while the precision, $PPV = TP / (TP + FP)$, informs about the fraction of unreliable data classified as reliable. Finally, the experiments show that—despite some differences between expert- and RNAtbor-driven analysis—the normalized RNA probing reactivity values obtained in both approaches are highly similar. The comparison of the reactivity profiles indicates the conformity of manual and automatic procedures. The averaged results from three independent probing experiments yield a Spearman correlation coefficient equal to 0.9987 for SHAPE and 0.995 for DMS-based analysis (Fig 3).

In the testing phase, we also executed statistical analysis to verify the repeatability of obtained results for each nucleotide between the replicates, and compare the reactivity profiles. Fig 4 shows an example of such verification for selected DMS-CE experiments previously performed by RNAtbor (raw data used in these experiments are provided in the S1 File). Experiments 1 and 2 (denoted as DMSexp1 and DMSexp2 in Fig 4) were performed under identical experimental conditions, while the higher concentration of DMS was used in experiment 3 (denoted as DMSexp3). We observed a high similarity between reactivity profiles generated for experiments 1 and 2 (Fig 4A), whereas a significant difference was visible for experiment 3

Table 1. The results of validation of RNAtbor algorithm for unreliable data identification.

dataset	TP	TN	FP	FN	ACC	PPV	TPR	TNR
SHAPE	2462	38	0	32	0.99	1	0.99	1
DMS	2431	35	0	25	0.99	1	0.99	1
ALL	4893	73	0	57	0.99	1	0.99	1

<https://doi.org/10.1371/journal.pone.0239287.t001>

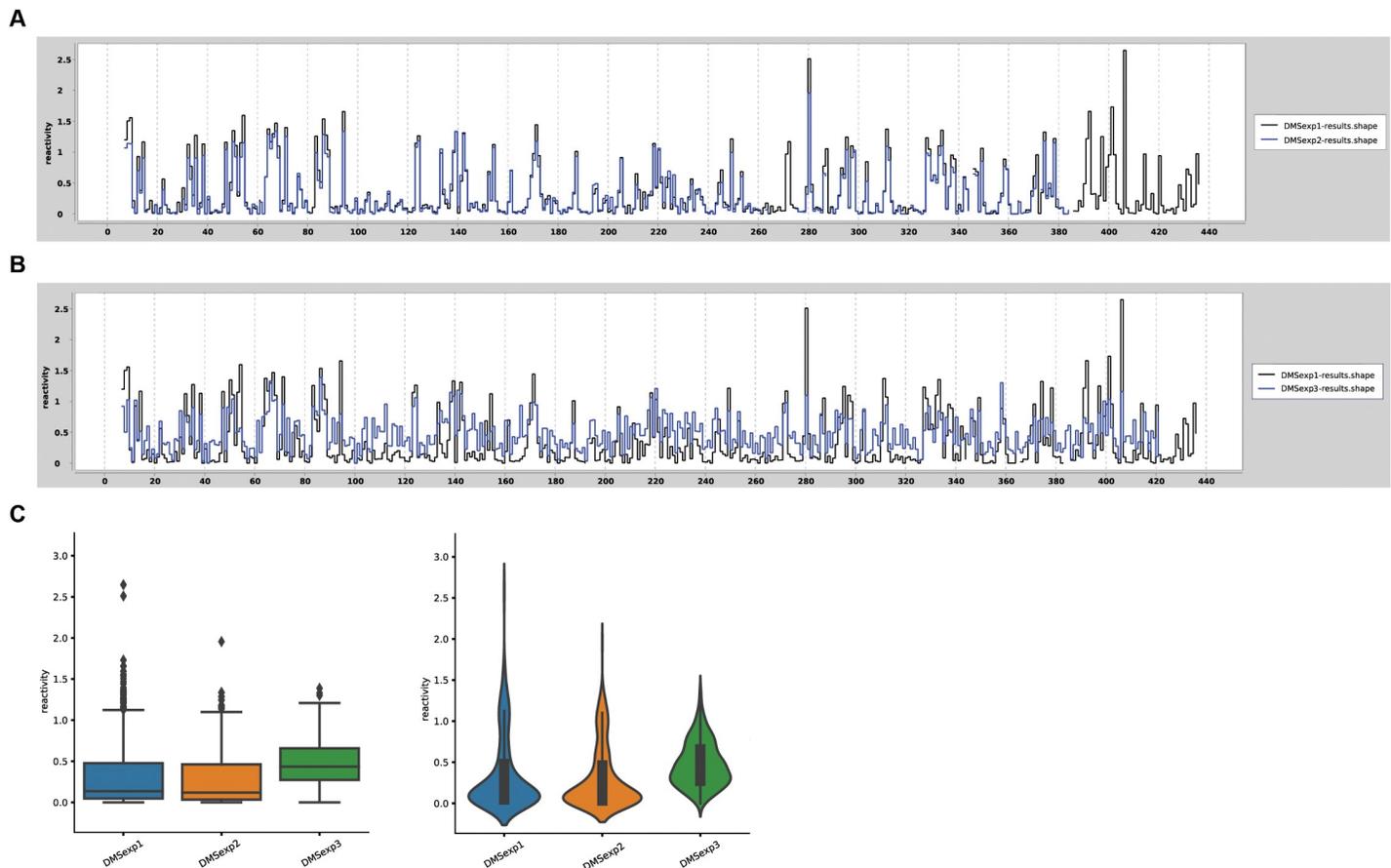


Fig 4. Example verification of the repeatability of RNA chemical probing experiments. Comparative step plot for repeatable (A) and non-repeatable (B) replicates of Ty1 RNA probing with DMS. (C) The box-and-whisker plot and violin plot presenting the differences in reactivity data distribution obtained for repeatable and non-repeatable experiments.

<https://doi.org/10.1371/journal.pone.0239287.g004>

(Fig 4B). As expected, the box plot and violin plot present the comparable DMS data distributions for experiments 1 and 2 (Fig 4C). The statistical plots for experiment 3 clearly show a significant increase in the number of more reactive nucleotides, and a concurrent decrease of unreactive nucleotides, consequently, the overall median reactivity is higher (Fig 4C). From these examples, we can see that additional options of RNAtor can be used for fast and easy comparative and statistical analysis for RNA chemical probing experiments.

Conclusions

In this work, we presented RNAtor, the new computational tool dedicated to the study of RNA structures that enriched the set of web-interfaced bioinformatics systems available within the RNapolis project [47]. RNAtor was designed for a fully automatic, quick normalization, and analysis of SHAPE / DMS-CE data. Although several programs can process the results of CE-based RNA probing, so far, no automatic procedure could identify unreliable data, and this step of the analysis was usually done manually. RNAtor incorporates the algorithm for the automatic exclusion of RT strong-stops to minimize user involvement in the probing data analysis. The tool can be applied to analyze data from other RNA probing methods if capillary electrophoresis and ShapeFinder or QuShape were used for data collection. RNAtor also visualizes the results of RNA probing data normalization, runs the data-driven prediction of

RNA secondary structure, and performs the statistical tests. The latter option facilitates the comparative study of multiple probing experiments, allows to assess the compatibility between experiments, and compare whole data sets of RNAs probed in different experimental conditions (e.g., in vitro, in vivo, ex vivo, in virio, ex virio), or in the absence or presence of protein/ligand. Compared to manual or semi-automated data processing, RNAtbor significantly reduces the time needed for data analysis; thus, it can highly improve the study and interpretation of data obtained from RNA chemical probing experiments.

In the future, we plan to extend the functionality of RNAtbor by implementing procedures combining RNA probing data from overlapping CE reads to facilitate the structural analysis of large RNAs.

Supporting information

S1 File.
(PDF)

Author Contributions

Conceptualization: Julita Gumna.

Data curation: Julita Gumna, Katarzyna Pachulska-Wieczorek.

Funding acquisition: Marta Szachniuk.

Methodology: Tomasz Zok.

Project administration: Katarzyna Pachulska-Wieczorek, Marta Szachniuk.

Software: Tomasz Zok, Kacper Figurski.

Supervision: Marta Szachniuk.

Validation: Julita Gumna, Tomasz Zok, Katarzyna Pachulska-Wieczorek, Marta Szachniuk.

Visualization: Kacper Figurski.

Writing – original draft: Julita Gumna.

Writing – review & editing: Julita Gumna, Katarzyna Pachulska-Wieczorek, Marta Szachniuk.

References

1. Waterman MS, Smith TF (1978) Rna Secondary Structure—Complete Mathematical-Analysis. *Mathematical Biosciences* 42: 257–266.
2. Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* 5: 105. <https://doi.org/10.1186/1471-2105-5-105> PMID: 15296519
3. Popena M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, et al. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40: e112. <https://doi.org/10.1093/nar/gks339> PMID: 22539264
4. Biesiada M, Pachulska-Wieczorek K, Adamiak RW, Purzycka KJ (2016) RNAComposer and RNA 3D structure prediction for nanotechnology. *Methods* 103: 120–127. <https://doi.org/10.1016/j.ymeth.2016.03.010> PMID: 27016145
5. Lukasiak P, Antczak M, Ratajczak T, Szachniuk M, Popena M, et al. (2015) RNAssess—a web server for quality assessment of RNA 3D structures. *Nucleic Acids Res* 43: W502–506. <https://doi.org/10.1093/nar/gkv557> PMID: 26068469

6. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, et al. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* 101: 7287–7292. <https://doi.org/10.1073/pnas.0401799101> PMID: 15123812
7. Ge P, Zhang S (2015) Computational analysis of RNA structures with chemical probing data. *Methods* 79–80: 60–66.
8. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127: 4223–4231. <https://doi.org/10.1021/ja043822v> PMID: 15783204
9. Inoue T, Cech TR (1985) Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci U S A* 82: 648–652. <https://doi.org/10.1073/pnas.82.3.648> PMID: 2579378
10. Spitale RC, Flynn RA, Torre EA, Kool ET, Chang HY (2014) RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip Rev RNA* 5: 867–881. <https://doi.org/10.1002/wrna.1253> PMID: 25132067
11. Mitchell D 3rd, Assmann SM, Bevilacqua PC (2019) Probing RNA structure in vivo. *Curr Opin Struct Biol* 59: 151–158. <https://doi.org/10.1016/j.sbi.2019.07.008> PMID: 31521910
12. Sloma MF, Mathews DH (2015) Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol* 553: 91–114. <https://doi.org/10.1016/bs.mie.2014.10.053> PMID: 25726462
13. McGinnis JL, Dunkle JA, Cate JH, Weeks KM (2012) The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 134: 6617–6624. <https://doi.org/10.1021/ja2104075> PMID: 22475022
14. Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129: 4144–4145. <https://doi.org/10.1021/ja0704028> PMID: 17367143
15. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, et al. (2013) RNA SHAPE analysis in living cells. *Nat Chem Biol* 9: 18–20. <https://doi.org/10.1038/nchembio.1131> PMID: 23178934
16. Busan S, Weidmann CA, Sengupta A, Weeks KM (2019) Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* 58: 2655–2664. <https://doi.org/10.1021/acs.biochem.8b01218> PMID: 31117385
17. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11: 959–965. <https://doi.org/10.1038/nmeth.3029> PMID: 25028896
18. Mailler E, Paillart JC, Marquet R, Smyth RP, Vivet-Boudou V (2019) The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip Rev RNA* 10: e1518. <https://doi.org/10.1002/wrna.1518> PMID: 30485688
19. Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* 106: 97–102. <https://doi.org/10.1073/pnas.0806929106> PMID: 19109441
20. Abeysirigunawardena SC, Kim H, Lai J, Rangunathan K, Rappe MC, et al. (2017) Evolution of protein-coupled RNA dynamics during hierarchical assembly of ribosomal complexes. *Nat Commun* 8: 492. <https://doi.org/10.1038/s41467-017-00536-1> PMID: 28887451
21. Garcia GR, Goodale BC, Wiley MW, La Du JK, Hendrix DA, et al. (2017) In Vivo Characterization of an AHR-Dependent Long Noncoding RNA Required for Proper Sox9b Expression. *Mol Pharmacol* 91: 609–619. <https://doi.org/10.1124/mol.117.108233> PMID: 28385905
22. Liu F, Somarowthu S, Pyle AM (2017) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat Chem Biol* 13: 282–289. <https://doi.org/10.1038/nchembio.2272> PMID: 28068310
23. Owens MC, Clark SC, Yankey A, Somarowthu S (2019) Identifying Structural Domains and Conserved Regions in the Long Non-Coding RNA lncTCF7. *Int J Mol Sci* 20.
24. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6: e96. <https://doi.org/10.1371/journal.pbio.0060096> PMID: 18447581
25. Soszynska-Jozwiak M, Michalak P, Moss WN, Kierzek R, Keszy J, et al. (2017) Influenza virus segment 5 (+)RNA—secondary structure and new targets for antiviral strategies. *Sci Rep* 7: 15041. <https://doi.org/10.1038/s41598-017-15317-5> PMID: 29118447
26. Michalak P, Soszynska-Jozwiak M, Biala E, Moss WN, Keszy J, et al. (2019) Secondary structure of the segment 5 genomic RNA of influenza A virus and its application for designing antisense oligonucleotides. *Sci Rep* 9: 3801. <https://doi.org/10.1038/s41598-019-40443-7> PMID: 30846846
27. Mahmud B, Horn CM, Tappich WE (2019) Structure of the 5' Untranslated Region of Enteroviral Genomic RNA. *J Virol* 93.

28. Newburn LR, White KA (2020) A trans-activator-like structure in RCNMV RNA1 evokes the origin of the trans-activator in RNA2. *PLoS Pathog* 16: e1008271. <https://doi.org/10.1371/journal.ppat.1008271> PMID: 31905231
29. Khoury G, Mackenzie C, Ayadi L, Lewin SR, Branlant C, et al. (2020) Tat IRES modulator of tat mRNA (TIM-TAM): a conserved RNA structure that controls Tat expression and acts as a switch for HIV productive and latent infection. *Nucleic Acids Res* 48: 2643–2660. <https://doi.org/10.1093/nar/gkz1181> PMID: 31875221
30. Kendall C, Khalid H, Muller M, Banda DH, Kohl A, et al. (2019) Structural and phenotypic analysis of Chikungunya virus RNA replication elements. *Nucleic Acids Res* 47: 9296–9312. <https://doi.org/10.1093/nar/gkz640> PMID: 31350895
31. Purzycka KJ, Legiewicz M, Matsuda E, Eizenstat LD, Lusvarghi S, et al. (2013) Exploring Ty1 retrotransposon RNA structure within virus-like particles. *Nucleic Acids Res* 41: 463–473. <https://doi.org/10.1093/nar/gks983> PMID: 23093595
32. Gumna J, Purzycka KJ, Ahn HW, Garfinkel DJ, Pachulska-Wieczorek K (2019) Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biol* 16: 1749–1763. <https://doi.org/10.1080/15476286.2019.1657370> PMID: 31469343
33. Blaszczyk L, Biesiada M, Saha A, Garfinkel DJ, Purzycka KJ (2017) Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression. *Viruses* 9.
34. McGinnis JL, Duncan CD, Weeks KM (2009) High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol* 468: 67–89. [https://doi.org/10.1016/S0076-6879\(09\)68004-6](https://doi.org/10.1016/S0076-6879(09)68004-6) PMID: 20946765
35. Novikova IV, Hennelly SP, Sanbonmatsu KY (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* 40: 5034–5051. <https://doi.org/10.1093/nar/gks071> PMID: 22362738
36. Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, Saha A, Gumna J, et al. (2015) Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic Acids Res* 43: 7414–7431. <https://doi.org/10.1093/nar/gkv695> PMID: 26160887
37. Pachulska-Wieczorek K, Blaszczyk L, Biesiada M, Adamiak RW, Purzycka KJ (2016) The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag. *Retrovirology* 13: 18. <https://doi.org/10.1186/s12977-016-0245-1> PMID: 26987314
38. Pang PS, Elazar M, Pham EA, Glenn JS (2011) Simplified RNA secondary structure mapping by automation of SHAPE data analysis. *Nucleic Acids Res* 39: e151. <https://doi.org/10.1093/nar/gkr773> PMID: 21965531
39. Yoon S, Kim J, Hum J, Kim H, Park S, et al. (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* 27: 1798–1805. <https://doi.org/10.1093/bioinformatics/btr277> PMID: 21561922
40. Cantara WA, Hatterschide J, Wu W, Musier-Forsyth K (2017) RiboCAT: a new capillary electrophoresis data analysis tool for nucleic acid probing. *RNA* 23: 240–249. <https://doi.org/10.1261/ma.058404.116> PMID: 27821510
41. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* 14: 1979–1990. <https://doi.org/10.1261/ma.1166808> PMID: 18772246
42. Karabiber F, McGinnis JL, Favorov OV, Weeks KM (2013) QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* 19: 63–73. <https://doi.org/10.1261/ma.036327.112> PMID: 23188808
43. Low JT, Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods* 52: 150–158. <https://doi.org/10.1016/j.ymeth.2010.06.007> PMID: 20554050
44. Lusvarghi S, Sztuba-Solinska J, Purzycka KJ, Rausch JW, Le Grice SF (2013) RNA secondary structure prediction using high-throughput SHAPE. *J Vis Exp*: e50243. <https://doi.org/10.3791/50243> PMID: 23748604
45. Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11: 129. <https://doi.org/10.1186/1471-2105-11-129> PMID: 20230624
46. Huang Q, Purzycka KJ, Lusvarghi S, Li D, LeGrice SF, et al. (2013) Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA* 19: 320–332. <https://doi.org/10.1261/ma.035535.112> PMID: 23329695
47. Szachniuk M (2019) RNApolis: Computational Platform for RNA Structure Analysis. *Foundations of Computing and Decision Sciences* 44: 241–257.