# Computational prediction of non-enzymatic RNA degradation patterns

Agnieszka Rybarczyk[1,2,3], Paulina Jackowiak[1,4], Marek Figlerowicz[1,2,3] and Jacek Blazewicz[1,2,3]✉

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; [2]Institute of Computing Science, Poznan University of Technology, Poznań Poland; [3]European Centre for Bioinformatics and Genomics, Poznań, Poland; [4]Institute of Chemical Technology and Engineering, Poznan University of Technology, Poznań, Poland

Since the beginning of the 21st century, an increasing interest in the research of ribonucleic acids has been observed in response to a surprising discovery of the role that RNA molecules play in the biological systems. It was demonstrated that they do not only take part in the protein synthesis (mRNA, rRNA, tRNA) but also are involved in the regulation of gene expression. Several classes of small regulatory RNAs have been discovered (e.g. microRNA, small interfering RNA, piwiRNA). Most of them are excised from specific double-stranded RNA precursors by enzymes that belong to the RNaseIII family (Drosha, Dicer or Dicer-like proteins). More recently, it has been shown that small regulatory RNAs are also generated as stable intermediates of RNA degradation (the so called RNA fragments originating from tRNA, snRNA, snoRNA etc.). Unfortunately, the mechanisms underlying biogenesis of the RNA fragments remain unclear. It is thought that several factors may be involved in the formation of the RNA fragments. The most important are the specific RNases, RNA-protein interactions and RNA structure. In this work, we focus on the RNA primary and secondary structures as factors influencing the RNA stability and consequently the pattern of RNA fragmentation. Earlier, we identified the major structural factors affecting non-enzymatic RNA degradation. Now, based on these data, we developed a new branch-and-cut algorithm that is able to predict the products of large RNA molecules' hydrolysis *in vitro*. We also present the experimental data that verify the results generated using this algorithm.

**Key words**: RNA degradation, non-enzymatic RNA hydrolysis, branch-and-cut algorithm

**Received**: 30 May, 2016; revised: 21 July, 2016; accepted: 22 July, 2016; available on-line: 19 October, 2016

## INTRODUCTION

Over the last decades, it has become increasingly clear that the RNA molecules play a much more important role in the cellular processes than it had been expected earlier. They do not only associate with a set of proteins to form ribosomes (rRNA), serve as templates (mRNA) or adaptors (tRNA) in protein synthesis, but can also display catalytic functions previously believed to be restricted to proteins (Ross, 1995).

It was demonstrated that apart from large RNAs (e.g. rRNA or mRNA) there also exist plenty of short RNAs, called small regulatory RNAs or riboregulators, some of which have already verified, while others only suggested, roles in almost all stages of gene expression (Szweykowska-Kulińska *et al.*, 2003; Kim *et al.*, 2009,

Aalto *et al.*, 2012; Morris *et al.*, 2014; Szostak *et al.*, 2014). They regulate responses to changes in the environmental conditions, influence genome structure, mRNA stability and downregulate translation (Mattick, 2006; Mattick, 2009; Czech *et al.*, 2011; Mickiewicz *et al.*, 2016). There is an increasing evidence that the range of known, potentially functional, non protein coding short RNAs is likely to expand and also includes RNA degradation intermediates that have not been taken into account so far (Cole *et al.*, 2009; Jackowiak *et al.*, 2011; Nowacka *et al.*, 2013).

RNA degradation, in all organisms across all kingdoms of life, is a prevalent cellular activity and major component of RNA metabolism (Schweingruber *et al.*, 2013; Hui *et al.*, 2014). It is responsible for securing a balance between RNA synthesis and RNA decay pathways and provides cellular homeostasis (Houseley *et al.*, 2009; Jackowiak *et al.*, 2011). For fulfilling its role, RNA degradation machinery has to distinguish between a set of molecules that are unnecessary under given conditions or that are dysfunctional and thus should be digested, and those essential for proper cell functioning that should stay intact. Consequently, it has been demonstrated that the stability of individual transcripts differs significantly and their half-lives range from seconds to hours, with proportional effects on protein synthesis (Lorentzen *et al.*, 2006; Hui *et al.*, 2014). This is ensured by precise and complex control mechanisms. Unfortunately, at present our knowledge about how the RNA degradation pathways control the longevity of individual RNAs is very limited (Ross, 1995).

RNA degradation is a highly efficient process that rapidly eliminates transcripts that are not functional or no longer necessary (Schweingruber *et al.*, 2013). Due to this high efficacy, it protects the gene expression pathways from being affected by transiently or accidentally formed RNA molecules (Houseley *et al.*, 2009; Hui *et al.*, 2014). Interestingly, it has been shown that some mature functional RNAs are not digested into single nucleotides but become a source of stable intermediates (Thompson *et al.*, 2008; Taft *et al.*, 2009; Zhang, *et al.* 2009; Pederson, 2010; Hurto, 2011; Jackowiak *et al.*, 2011; Sobala *et al.*, 2011; Nowacka *et al.*, 2012; Maivali *et al.*, 2013), which can exhibit various regulatory functions through acting as translation inhibitors or signalling molecules (Hurto, 2011; Aalto *et al.*, 2012; Nowacka *et al.*, 2013). It is postulated that these molecules are products of both, enzymatic cleavage and non-enzymatic RNA structure-driven

spontaneous degradation (Hsieh *et al.*, 2009; 2010, Jackowiak *et al.*, 2011; Li *et al.*, 2012; Nowacka *et al.*, 2012).

Despite numerous reports indicating the pivotal role of the RNA structure in all processes involving these molecules (e.g. in gene expression: Grunberg-Manago *et al.*, 1999; Winkler *et al.*, 2005; Podkowinski *et al.*, 2009 or RNA recombination: Figlerowicz, 2000), to our knowledge very little has been published so far on this issue. Lack of methods that permit to determine or predict the highly dynamic RNA structure still remains the major problem. Molecular mechanisms that underlie the RNA structure-driven spontaneous RNA degradation have been of great interest over the last decade (Kazakov *et al.*, 1992; Keck *et al.*, 1995; Bibillo *et al.*, 2000; Kaukinen *et al.*, 2002). The detailed study of this process has been conducted by Kierzek and co-workers (Kierzek, 1992; Bibillo *et al.*, 1999; Bibillo *et al.*, 2000; Kierzek, 2001). They observed that short RNA oligoribonucleotides can be selectively and quantitatively cleaved in the absence of any protein enzymes. They also reported that this process is structure dependent and the cleavage preferentially occurs within single-stranded fragments of the RNA molecules. The most important for hydrolysis to take place is the presence of the labile phosphodiester bond such as YA or YC (Y being pyrimidine). Certainly, this process can be also affected by other factors, such as metal ions, organic polymers or polyamines, which stabilize or destabilize the RNA structure and influence the cleavage through nonspecific interactions (Kierzek, 1992; Ciesiolka *et al.*, 1998; Kierzek, 2001). It has been also shown that not only the primary, but also the secondary and tertiary structure of the substrate has a large influence on the non-enzymatic RNA hydrolysis (Ciesiolka *et al.*, 1998; Mikkola *et al.*, 2001; Kaukinen *et al.*, 2002).

These findings promote further research of RNA degradation which is essential for broadening our knowledge on RNA regulatory functions. In this work, we present a new branch-and-cut algorithm for the prediction of all products of RNA non-enzymatic hydrolysis. We demonstrate its performance in a model that involves artificial RNA molecules designed to be prone to spontaneous decay, according to the rules of degradation developed by Kierzek and co-workers (Kierzek, 1992; Kierzek, 2001; Blazewicz *et al.*, 2011). Our results show that the algorithm provides an efficient way to accurately predict the products of non-enzymatic RNA hydrolysis. In addition, our data provide insight into the mechanistic aspects of RNA decay, indicating that re-folding of degradation intermediates is the key determinant of the decay pattern. The analysis of the non-enzymatic degradation of RNA can provide the researchers with information necessary to design highly stable RNAs, or RNAs that could potentially be the source of stable intermediates functioning as regulatory molecules (riboregulators). It also constitutes a contribution to the studies on RNA processing in living organisms.

## MATERIALS AND METHODS

**Experimental analysis of non-enzymatic RNA degradation**. The experimental analysis of non-enzymatic RNA degradation was performed using two artificial molecules: RNA-A$_{108}$ (108 nt in length) and RNA-B$_{66}$ (66 nt in length) (Blazewicz *et al.*, 2011). They were designed taking into account the degradation rules developed by Kierzek and co-workers (Kierzek, 1992; Bibillo *et al.*, 1999; Bibillo *et al.*, 2000; Kierzek, 2001).

RNA-A$_{108}$ and RNA-B$_{66}$ were obtained by *in vitro* transcription and labelled with $^{32}$P at their 5' ends with the T4 polynucleotide kinase, as described in (Dutkiewicz *et al.*, 2005). Next, the products of non-enzymatic RNA hydrolysis were identified in accordance with the procedure proposed by Kierzek and co-workers (Kierzek, 1992; Bibillo *et al.*, 1999; Bibillo *et al.*, 2000; Kierzek, 2001). Briefly, labelled RNA molecules were incubated in the solution containing 50 mM Tris-HCl (pH 7.5), 2 mM EDTA, 1 mM spermidine, 50 mM NaCl, and 0.1% PVP at 37°C for 6 h. Then, samples were electrophoresed in a 16% denaturing polyacrylamide gel and visualized by autoradiography. The product lengths were determined by comparison with the length of the RNA markers, namely alkaline hydrolysis ladder and partial T1 nuclease digestion generated the same way as in (Dutkiewicz *et al.*, 2005).

**RNA secondary structure prediction**. To predict the canonical RNA secondary structure, we have used RNAstructure (Reuter *et al.*, 2010) and RNAfold (Hofacker *et al.*, 1994) with their default parameters.

In order to predict the extended RNA secondary structure (also containing the non-canonical interactions) we have applied the pipeline described in (Rybarczyk *et al.*, 2015), with sequence-based prediction. The idea of this approach is based on predicting the tertiary structure from a nucleotide sequence and then back-calculating the extended secondary structure from atom coordinate data. We have composed our pipeline (the RC/Rp pipeline) in the same way as in (Rybarczyk *et al.*, 2015). Briefly, we employed RNAComposer (Popenda *et al.*, 2012; Zok *et al.*, 2014), a fully automated and fast tool for 3D structure prediction, and RNApdbee (Antczak *et al.*, 2014) for extracting RNA 2D structure from atom coordinate data. If solely the sequence is provided as an input for RNAComposer (as in our case) then the canonical secondary structure is predicted by RNAstructure (default) or RNAfold (upon user selection). All these tools are incorporated into the RNAComposer system.

**RNA degradation rules**. It has been demonstrated that non-enzymatic RNA degradation is a structure dependent process (Kierzek, 1992; Bibillo *et al.*, 2000). The phosphodiester bonds that connect nucleotides are preferentially cleaved within single stranded regions of the RNA molecules. Analysis conducted with oligoribonucleotides had shown that different phosphodiester bonds are hydrolyzed with different rates. It was observed that only YA and YC (Y being pyrimidine) bonds are hydrolyzed with significant rates over 24 h at 37°C. The phosphodiester bonds in YA are 3–5 fold more susceptible to cleavage than YC. UA is 1.5–2.0 fold more sensitive to hydrolysis than CA. The phosphodiester bonds in YG and YU sequences are at least 20 fold more stable then YA and YC. Phosphodiester bonds in RR and RY (R denotes purine) are stable under the applied conditions. In general, the stability of phosphodiester bonds can be presented in the following order: YU>YG>YC>CA>UA (Kierzek, 1992; Bibillo *et al.*, 2000; Kierzek, 2001). The above RNA degradation rules proposed by Kierzek and co-workers (Kierzek, 2001) were determined using completely single- or double-stranded oligoribonucleotides (up to 10 nucleotides).

Let $P=\{UA, CA, UC, CC, UG, CG, UU, CU\}$ be a set of all possible cleavage sites according to Kierzek and co-workers (Kierzek, 1992; Kierzek, 2001). According to the rules of degradation, developed for different phosphodiester bonds, the following range of the degradation parameter [c, d], for each permissible cleavage site can be calculated:

In case of YG, YU:

$$c_{(YG,\ YU)}=0$$

$$d_{(YG,\ YU)}=0.1$$

For $YC$:

$$c_{(YC)}=20$$

$$d_{(YC)}=50$$

For $CA$:

$$c_{(CA)}=3\times c_{(YC)}$$

$$d_{(CA)}=5\times d_{(YC)}$$

For $UA$:

$$c_{(UA)}=1.5\times c_{(CA)}$$

$$d_{(UA)}=2\times d_{(CA)}$$

For each cleavage site from the $P$ set, the following degradation membership (Taguchi, 1987; Zhang et al., 1995) measures, based on the expert knowledge and reflecting measurement inaccuracy of the experiment, were defined:

$$\mu(UA)=\frac{\left(d_{(UA)}-c_{(UA)}\right)^{\frac{1}{2}}}{1+\left(d_{(UA)}-c_{(UA)}\right)^{\frac{1}{2}}}=0.953,$$

$$\mu(CA)=\frac{\left(d_{(CA)}-c_{(CA)}\right)^{\frac{1}{2}}}{1+\left(d_{(CA)}-c_{(CA)}\right)^{\frac{1}{2}}}=0.932,$$

$$\mu(YC)=\frac{\left(d_{(YC)}-c_{(YC)}\right)^{\frac{1}{2}}}{1+\left(d_{(YC)}-c_{(YC)}\right)^{\frac{1}{2}}}=0.846,$$

$$\mu(YG,\ YU)=0.1$$

## RESULTS AND DISCUSSION

To enable the prediction of the products of non-enzymatic RNA degradation, we designed a branch-and-cut algorithm and implemented it in C++ programming language. The algorithm simulates the degradation process of a given RNA molecule based on its sequence and the degradation rules developed by Kierzek and co-workers. In order to apply those rules, the secondary structure of the analysed RNA must be predicted with the use of a selected external program. Next, cleavage sites are recognized and for each individual cut two RNA fragments are generated. Importantly, we assumed that after initial cleavage both fragments can dissociate and adopt a new fold. Consequently, to predict further products of the spontaneous RNA degradation, the secondary structure must be predicted for each of the obtained RNA fragments (see Fig. 1).

The proposed algorithm takes as an input the sequence of the RNA molecule, the cut-off threshold for degradation measures ratio (default $\varepsilon=0.01$), minimal

length of the fragment that can be refolded and further analyzed (default $minLen$=20 nt) and the program used for the RNA 2D structure prediction (RNAstructure, RNAfold or RC/Rp pipeline). In addition, the following structure in case of each node is defined: list $\Omega_{node}$ of all RNA fragments $(x, y)$ that are longer than $minLen$ and can be further processed. Each element stored in the $\Omega_{node}$ is denoted by a "left" $(x)$ and "right" $(y)$ end which provides the information of the exact location of such fragment within the input RNA molecule. Each node also contains a degradation measure $\varepsilon_{node}$, which is calculated as a ratio of a parent node $\varepsilon_{node}$ and the current value of degradation measure assigned to a selected cleavage sites. The output of the algorithm is the $W$ set, which will be composed of all fragments being the result of the input RNA molecule degradation.

To find the solution, the algorithm builds and searches through a solution tree, whose leaves correspond to elements of the solution space of the problem. In the situation of each node of the solution tree, the element of the maximum length $\omega_{max}$ in $\Omega_{node}$ is selected (in case of the root of the tree, it will be the input RNA sequence). Then, the secondary structure for $\omega_{max}$ is predicted (using a chosen program; the outputs of the different programs are processed with Perl scripts). Next, all single-stranded substrings are extracted and for each of them the potential cleavage sites based on the set $P=\{UA, CA, UC, CC, UG, CG, UU, CU\}$ are found and assigned the values of degradation measures $\mu(YN)$ (see RNA degradation rules in Materials and Methods section for details). It is assumed that for effective cleavage, both nucleotides participating in the formation of the labile phosphodiester bond must be unpaired (Kierzek, 1992; Kierzek, 2001). All cleavage sites having the value of a degradation measure multiplied by a degradation measure $\varepsilon_{node}$ stored in a current node that are greater than the cut-off threshold $\varepsilon$ (which is a parameter given by the user) are further analysed. From among them, only the cleavage sites having the greatest value are taken into account. Finally, for each chosen cleavage site the cut is applied and the resultant RNA fragments longer than $minLen$ are inserted into the new node list $\Omega_{node}$. The RNA molecules shorter than $minLen$ are considered as final products of RNA degradation and are directly added to the set $W$.

In the situation, where for a considered RNA sequence no cleavage site could be selected (according to set $P$ or the value of degradation measure ratio, which
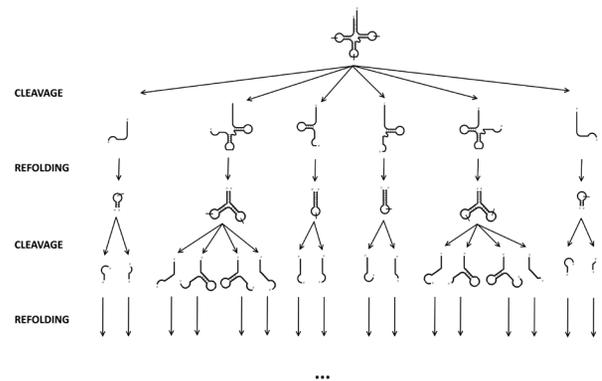


**Figure 1. General idea of the algorithm based on the example of a tRNA molecule.**
Let us assume, for example, that in the first step, RNA is cleaved at three sites. As a result six molecules are generated. Next, each fragment obtained is refolded and again subjected to cleavage. The whole procedure is repeated as long as fragments containing the allowed cleavage sites exist.

**A**



**B**
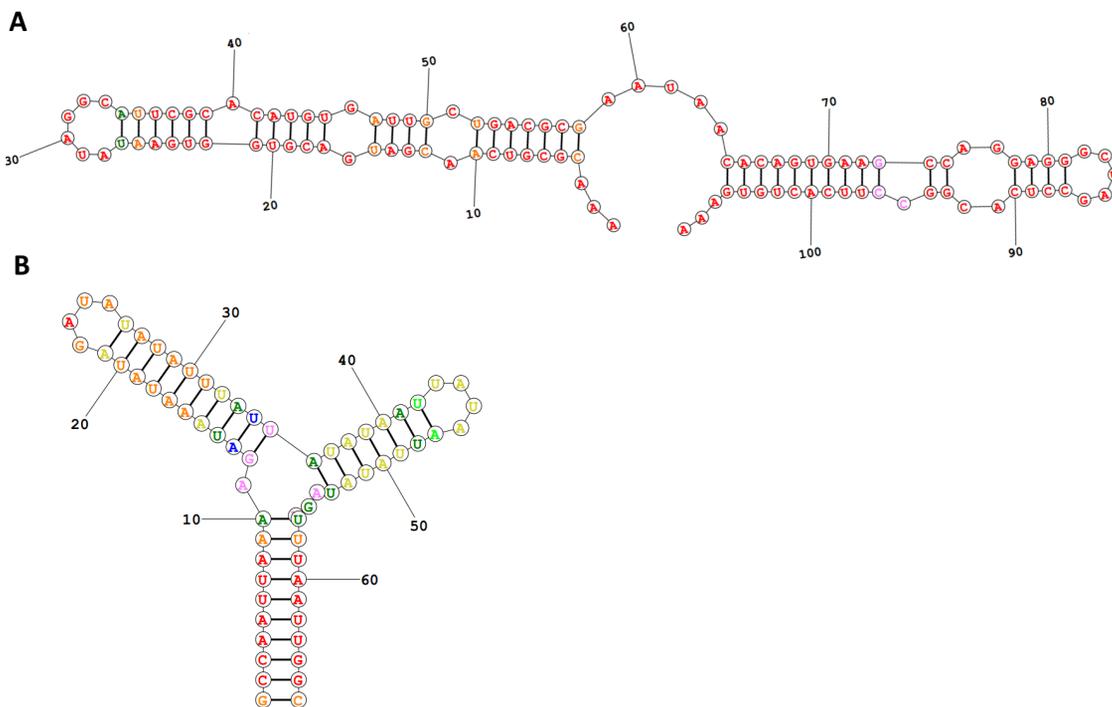


**Figure 2. Secondary structure of the two artificial RNA molecules: RNA-A$_{108}$ (A) and RNA-B$_{66}$ (B).**
The 2D structure of those molecules was predicted with the use of RNAstructure. The structures are colored according to base pair probabilities, where the highest probabilities are red (≥99%) and the lowest are purple (≤50%) (Reuter *et al.*, 2010).

occurred to be lower than $\varepsilon$), such a fragment is a final product of degradation and becomes an element of the resultant set $W$.

If there are no fragments containing the allowed cleavage sites, then the whole procedure ends with the final products of non-enzymatic RNA degradation in $W$.

To evaluate the performance of the algorithm, we used two artificial RNA molecules, RNA-A$_{108}$ and RNA-B$_{66}$. They were designed to have both, the primary and secondary structures, that would make them susceptible to degradation according to the rules proposed by Kierzek and co-workers. The secondary structures of those molecules, predicted with the use of RNAstructure (Reuter *et al.*, 2010) are presented in Fig. 2. We identified experimentally the products of their degradation (Fig 3, see also Materials and Methods section for details). We found that degradation of RNA-B$_{66}$ led to the formation of 12 radiolabeled fragments: 14, 18, 20, 21, 23, 24, 37, 39, 43, 45, 53 and 60 nucleotide long. Upon degradation of RNA-A$_{108}$, 8 fragments were detected: 12, 27, 28, 29, 30, 31, 61 and 84 nt long. Most of the cleavages occurred, as expected, between Y and A, but some of them were inconsistent with the rules proposed by Kierzek and co-workers, namely: A28-U29, A30-G31 and G31-G32, as far as RNA-A$_{108}$ is concerned, and A21-G22, A23-U24, A60-A61 in case of RNA-B$_{66}$. In parallel, the degradation products of both model molecules were predicted *in silico*. To this end, the developed branch-and-cut algorithm was run on a PC with Intel(R) Core(TM) i7-2600K, 3.40GHz processor and 8 GB RAM in a Linux environment. The algorithm correctly predicted the cleavage sites that were consistent with the incorporated rules. However, as mentioned above, experimental data revealed cleavage of phosphodiester bonds located within RR and RY motifs that are considered

as stable under the applied conditions (Kierzek, 1992, 2001). Hence, they are not present in the $P$ set and consequently could not be predicted by the algorithm. Unfortunately, to our knowledge, Kierzek and co-workers are the only ones that quantitatively analyzed the reactivity of the phosphodiester bonds within RNA and developed on this basis the rules displaying their stability. The cleavage rules in the case of RR and RY motifs have not been analyzed in depth so far.

Our data clearly showed that based on the input RNA 2D structure itself (the structure of RNA subjected to analysis), it is not possible to recognize all degradation products. For example, in case of 2D structure of RNA-B$_{66}$ (see Fig. 2), it is possible to indicate only the following potential cleavage sites: U24-A25, U43-A44 and U45-A46, while in case of RNA-A$_{108}$: U29-A30, U61-A62 and U84-A85. They constitute only a subset of the experimentally detected cleavages. Hence, refolding, which is applied by the algorithm, is necessary to see all products resulting from the input RNA degradation. This observation provides evidence that the spontaneous RNA degradation is very dynamic and involves multiple steps of structural rearrangements. Moreover, the structure adopted by each of the sequentially generated degradation intermediates sets the direction of the entire process and determines the final degradation pattern.

Because it was shown that the cleavages occur only within single-stranded regions of the RNA structure and the refolding is repeated by the algorithm many times, the selection of the appropriate method for 2D structure determination is crucial. Therefore, we performed the tests of the algorithm using different methods dedicated to the RNA secondary structure prediction. We applied RNAstructure, RNAfold and RC/Rp pipeline executed with each option for base-pair identification, namely

**Table 1. Results obtained with the algorithm for artificial RNA molecules: RNA-B$_{66}$ and RNA-A$_{108}$.**
The cleavage sites predicted by the algorithm that are consistent with experimental results are shown in bold.

| Method for RNA 2D structure prediction | RNA fragments identified in experiment and predicted by the algorithm | |
|---|---|---|
| | RNA-B$_{66}$ | RNA-A$_{108}$ |
| None (results of experimental analysis of non-enzymatic RNA degradation) | 14, 18, 20, 21, 23, 24, 37, 39, 43, 45, 53, 60 | 12, 27, 28, 29, 30, 31, 61, 84 |
| RNAfold | **14**, **18**, **20**, **24**, 26, 35, **37**, **39**, **43**, **45** | **27**, **29**, **61**, **84** |
| RC$_{RNAfold}$/Rp-1 | **14**, **18**, **20**, **24**, 35, **37**, **43**, **45** | **12**, 23, **27**, **29**, **61**, **84** |
| RC$_{RNAfold}$/Rp-2 | **18**, **20**, **24**, 35, **37**, **43**, **45**, **53** | **12**, 23, **27**, **29**, **61**, **84** |
| RC$_{RNAfold}$/Rp-3 | **14**, **18**, **20**, **24**, 35, **37**, **39**, **43**, **45**, **53** | **27**, **29**, **61**, **84** |
| RNAstructure | **24**, 26, 35, **37**, **39**, **43**, **45**, **53** | **27**, **29**, **61**, **84** |
| RC$_{RNAstructure}$/Rp-1 | **14**, **18**, **20**, **24**, 35, **37**, **43**, **45** | **12**, 23, **27**, **29**, **61**, **84** |
| RC$_{RNAstructure}$/Rp-2 | **18**, **20**, **24**, 35, **37**, **43**, **45**, **53** | **12**, 23, **27**, **29**, **61**, **84** |
| RC$_{RNAstructure}$/Rp-3 | **14**, **18**, **20**, **24**, 35, **37**, **39**, **43**, **45**, **53** | **27**, **29**, **61**, **84** |
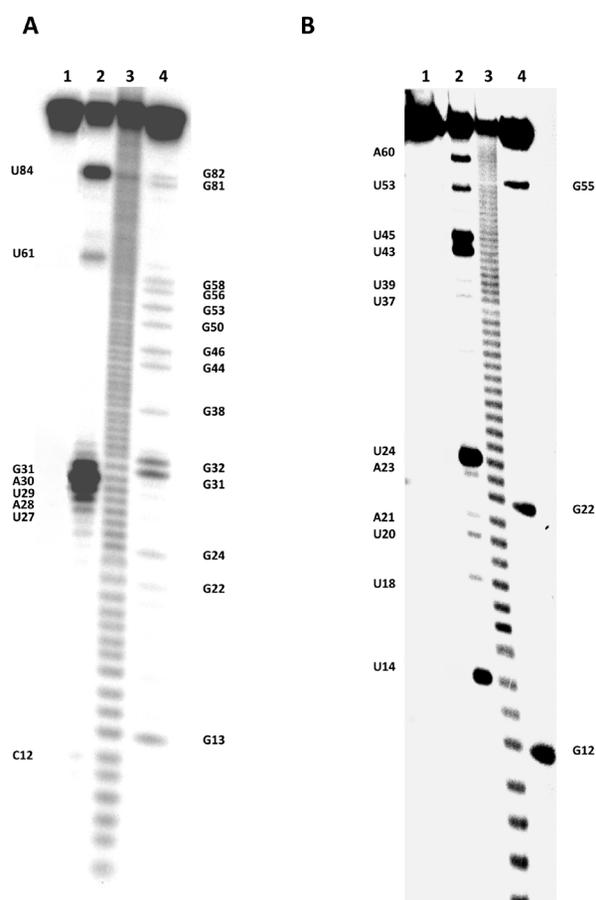


**Figure 3. The degradation pattern of two artificial RNA molecules: RNA-A$_{108}$ (A) and RNA-B$_{66}$ (B).**
RNA molecules labelled with $^{32}$P at their 5′ ends were incubated for 6 h at 37°C in the presence of 50 mM Tris-HCl (pH 7.5), 2 mM EDTA, 1 mM spermidine, 50 mM NaCl, and 0.1% PVP. Lane 1, reaction control; lane 2, incubated for 6 h; lane 3, formamide ladder; lane 4, limited hydrolysis by RNase T1 (guanine residues are labelled on the right).

**Table 2. Average computing times for artificial RNA molecules: RNA-B$_{66}$ and RNA-A$_{108}$ (in seconds).**

| Method for RNA 2D structure prediction | Average computational time [s] |
|---|---|
| RNAfold | 0.238 |
| RC$_{RNAfold}$/Rp-1 | 503.684 |
| RC$_{RNAfold}$/Rp-2 | 584.373 |
| RC$_{RNAfold}$/Rp-3 | 564.474 |
| RNAstructure | 0.758 |
| RC$_{RNAstructure}$/Rp-1 | 503.372 |
| RC$_{RNAstructure}$/Rp-2 | 583.459 |
| RC$_{RNAstructure}$/Rp-3 | 564.474 |

be used for canonical RNA secondary structure prediction. Each time RC/Rp was selected, RNAComposer was applied with RNAfold or RNAstructure, according to the user decision (RC$_{RNAstructure}$/Rp-1, RC$_{RNAstructure}$/Rp-2 or RC$_{RNAstructure}$/Rp-3 if RNAstructure was chosen to run within RNAComposer system and RC$_{RNAfold}$/Rp-1, RC$_{RNAfold}$/Rp-2 or RC$_{RNAfold}$/Rp-3 if RNAfold was concerned). Table 1 presents the results obtained *in vitro* and predicted by the algorithm. In case of RNA-B$_{66}$ all RNA structure predictors were successful in identifying the following cleavage sites: 24, 37, 43 and 45. All RC/Rp versions correctly predicted more cleavage sites than RNAstructure and RNAfold. Additionally, all RC/Rp versions falsely reported only one cleavage site, namely 35, while RNAstructure and RNAfold two: 26 and 35. The only tool that was able to predict almost all cleavage sites (apart from those not compatible with the rules provided by Kierzek and co-workers) was RC/Rp-3, which was consistent with the analysis conducted in (Rybarczyk *et al.*, 2015). In case of RNA-A$_{108}$, three out of eight cleavage sites generated in the *in vitro* experiment were incompatible with the rules of degradation developed by Kierzek and co-workers, thus, they could not be predicted by the algorithm. In case of all RNA structure predictors, the following cleavage sites were recognized: 27, 29, 61 and 84. Additionally, RC/Rp-1 and RC/Rp-2 correctly identified the cleavage site 12, but at the same time falsely reported the 23 site.

Table 2 summarizes running time results for the branch-and-cut algorithm tested on the two artificial

RNAView (Yang *et al.*, 2003), MC-Annotate (Gendron *et al.*, 2001) and 3DNA/DSSR (Lu *et al.*, 2003) (hereinafter RC/Rp-1, RC/Rp-2, RC/Rp-3, respectively). RNAfold and RNAstructure are incorporated into the RNAComposer system and the user can decide which of them will

RNA molecules. Due to differences between the approach for the automated assessment of extended RNA secondary structure (Rybarczyk *et al.*, 2015) and the methods that directly predict secondary structure, computing times of the RC/Rp pipeline are longer than those of RNAfold and RNAstructure. However, the RC/Rp pipeline is still fast when compared to the tools dedicated to the extended RNA secondary structure prediction and produces high-quality results that should be worth a longer wait (Rybarczyk *et al.*, 2015).

By analyzing the obtained results, we noticed that the algorithm performs quite efficiently and quickly, even when the RC/Rp pipeline is used for the refolding. The algorithm has been able to correctly identify most of the products of the non-enzymatic RNA degradation in both examples considered. All cleavage sites that were not recognized by the algorithm were not consistent with the rules developed by Kierzek and co-workers.

## CONCLUSIONS

To our knowledge, the proposed algorithm is the first approach that allows to simulate the non-enzymatic RNA hydrolysis. Because the RNA fragment refolding has to be used by the algorithm, the selection of the method for RNA 2D structure prediction is crucial. Hence, the algorithm allows the user to choose the external tool dedicated to canonical or extended secondary structure determination. Since the programs for 2D structure determination are run by external Perl scripts that are also responsible for parsing their output, it is easy to incorporate other methods.

The algorithm was tested with two artificial molecules: RNA-A$_{108}$ and RNA-B$_{66}$, whose non-enzymatic degradation was examined experimentally. The algorithm successfully identified all cleavage sites that were consistent with the rules provided by Kierzek and co-workers. Experimental data revealed, however, the presence of additional cleavages in the RR and RY motifs. The latter were considered stable according to the adopted rules and thus could not be predicted by the algorithm. To resolve this issue, additional chemical experiments involving large RNA molecules would have to be conducted to quantitatively analyze the stability of those phosphodiester bonds in relation to YY and YR. Obviously, the degradation rules of the algorithm can be modified to include any new findings concerning the RNA stability.

As a continuation of the research reported in this paper, one may consider for example the analysis of very long molecules (e.g. eukaryotic rRNAs) and/or molecules containing plenty of potential cleavage sites. Certainly, it may turn out that in such a case, the computing time of the branch-and-cut method would significantly increase, making it not acceptable for the user, especially when taking into account large testing sets. In such a situation, it would be worth considering the heuristic approach, similar to that presented in (Blazewicz *et al*, 2002; Blazewicz *et al*, 2005).

### Acknowledgements

## REFERENCES

Aalto AP, Pasquinelli AE (2012) Small non-coding RNAs mount a silent revolution in gene expression. *Curr Opin Cell Biol.* **24**: 333–340. doi: 10.1016/j.ceb.2012.03.006

Antczak M, Zok T, Popenda M, Lukasiak P, Adamiak RW, Blazewicz J, Szachniuk M (2014) RNApdbee – a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Res* **42**: W368–W372. doi: 10.1093/nar/gku330

Bibillo A, Figlerowicz M, Ziomek K, Kierzek R (2000) The nonenzymatic hydrolysis of oligoribonucleotides. VII. Structural elements affecting hydrolysis. *Nucleosides Nucleotides Nucleic Acids* **19**: 977–994. doi: 10.1080/15257770008033037

Bibillo A, Figlerowicz M, Kierzek R (1999) The non-enzymatic hydrolysis of oligoribonucleotides VI. The role of biogenic polyamines. *Nucleic Acids Res* **27**: 3931–3937. doi: 10.1093/nar/27.19.3931

Blazewicz J, Figlerowicz M, Kasprzak M, Nowacka M, Rybarczyk A (2011) RNA partial degradation problem: motivation, complexity, algorithm. *J Comput Biol* **18**: 821–834. doi: 10.1089/cmb.2010.0153

Blazewicz J, Formanowicz P, Guinand F, Kasprzak M (2002) A heuristic managing errors for DNA sequencing. *Bioinformatics* **18**: 652-660.

Blazewicz J, Szachniuk M, Wojtowicz A (2005) RNA tertiary structure determination: NOE pathway construction by tabu search. *Bioinformatics* **21**: 2356–2361. doi:10.1093/bioinformatics/bti351

Ciesiolka J, Michalowski D, Wrzesinski J, Krajewski J, Krzyzosiak WJ (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J Mol Biol* **275**: 211–220.

Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, Green PJ, Barton GJ, Hutvagner G (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* **15**: 2147–2160.

Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* **12**: 19–31. doi: 10.1038/nrg2916

Dutkiewicz M, Ciesiolka J (2005) Structural characterization of the highly conserved 98-base sequence at the 3' end of HCV RNA genome and the complementary sequence located at the 5' end of the replicative viral strand. *Nucleic Acids Res* **33**: 693–703. doi: 10.1093/nar/gki218

Figlerowicz M (2000) Role of RNA structure in non-homologous recombination between genomic molecules of brome mosaic virus. *Nucleic Acids Res* **28**: 1714–1723.

Gendron P, Lemieux S, Major F (2001) Quantitative analysis of nucleic acid three-dimensional structures. *J Mol Biol* **308**: 919–936. doi: 10.1006/jmbi.2001.4626

Grunberg-Manago M (1999) Messenger RNA Stability and its role in control of gene expression in bacteria and phages. *Annu Rev Genet.* **33**: 193–227. doi: 10.1146/annurev.genet.33.1.193

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte Für Chem Chem Mon* **125**: 167–188. doi: 10.1007/BF00818163

Houseley J, Tollervey D (2009) The many pathways of RNA degradation. *Cell* **136**: 763–776. doi: 10.1016/j.cell.2009.01.019

Hsieh LC, Lin SI, Shih AC, Chen JW, Lin WY, Tseng CY, Li WH, Chiou TJ (2009) Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. *Plant Physiol* **151**: 2120–2132. doi: 10.1104/pp.109.147280

Hsieh LC, Lin SI, Kuo HF, Chiou TJ (2010) Abundance of tRNA-derived small RNAs in phosphate-starved Arabidopsis roots. *Plant Signal Behav* **5**: 537–539. doi: 10.4161/psb.11029

Hui MP, Foley PL, Belasco JG (2014) Messenger RNA degradation in bacterial cells. *Annu Rev Genet* **48**: 537–559. doi: 10.1146/annurev-genet-120213-092340

Hurto RL (2011) Unexpected functions of tRNA and tRNA processing enzymes. *Adv Exp Med Biol* **722**: 137–155. doi: 10.1007/978-1-4614-0332-6_9

Jackowiak P, Nowacka M, Strozycki PM, Figlerowicz M (2011) RNA degradome-its biogenesis and functions. *Nucleic Acids Res* **39**: 7361–7370. doi: 10.1093/nar/gkr450

Kaukinen U, Lyytikainen S, Mikkola S, Lonnberg H (2002) The reactivity of phosphodiester bonds within linear single-stranded oligoribonucleotides is strongly dependent on the base sequence. *Nucleic Acids Res* **30**: 468–474. doi: 10.1093/nar/30.2.468

Kazakov S, Altman S (1992) A trinucleotide can promote metal ion-dependent specific cleavage of RNA. *Proc Natl Acad Sci USA* **89**: 7939–7943

Keck MV, Hecht SM (1995) Sequence-specific hydrolysis of yeast tRNA(Phe) mediated by metal-free bleomycin. *Biochemistry* **34**: 12029–12037. doi: 10.1021/bi00037a046

Kierzek R (1992) Nonenzymatic hydrolysis of oligoribonucleotides. *Nucleic Acids Res* **20**: 5079–5084

Kierzek R (2001) Nonenzymatic Cleavage of Oligoribonucleotides, *Methods Enzymol* **341**: 657–675. doi: 10.1016/S0076-6879(01)41183-9

Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**: 126–139. doi: 10.1038/nrm2632

Li Z, Ender C, Meister G, Moore PS, Chang Y, John B (2012) Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res* **40**: 6787–6799. doi: 10.1093/nar/gks307

Lorentzen E, Conti E (2006) The exosome and the proteosome: nanocompartments for degradation. *Cell* **125**: 651–654. doi: 10.1016/j.cell.2006.05.002

Lu X-J, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**: 5108–5121. doi: 10.1093/nar/gkg680

Maivali U, Paier A, Tenson T (2013) When stable RNA becomes unstable: the degradation of ribosomes in bacteria and beyond. *Biol Chem* **394**: 845–855. doi: 10.1515/hsz-2013-0133

Mattick JS, Makunin IV (2006) Non-coding RNA. *Hum Mol Genet* **15**: R17–R29.

Mattick JS (2009) The genetic signatures of noncoding RNAs. *PLoS Genet* **5**: e1000459. doi: 10.1371/journal.pgen.1000459

Mickiewicz A, Rybarczyk A, Sarzynska J, Figlerowicz M, Blazewicz J (2016) AmiRNA Designer – new method of artificial miRNA design. *Acta Biochim Pol* **63**: 71–77. doi: 0.18388/abp.2015_989

Mikkola S, Kaukinen U, Lonnberg H (2001) The effect of secondary structure on cleavage of the phosphodiester bonds of RNA. *Cell Biochem Biophys* **34**: 95–119. doi: 10.1385/CBB:34:1:95

Morris KV, Mattick JS (2014) The rise of regulatory RNA. *Nature Reviews Genetics* **15**: 423–437. doi:10.1038/nrg3722

Nowacka M, Jackowiak P, Rybarczyk A, Magacz T, Strozycki PM, Barciszewski J, Figlerowicz M (2012) 2D-PAGE as an effective method of RNA degradome analysis. *Molecular Biology Reports* **39**: 139–146. doi: 10.1007/s11033-011-0718-1

Nowacka M, Strozycki PM, Jackowiak P, Hojka-Osinska A, Szymanski M, Figlerowicz M (2013) Identification of stable, high copy number, medium-sized RNA degradation intermediates that accumulate in plants under non-stress conditions. *Plant Mol Biol* **83**: 191–204. doi: 10.1007/s11103-013-0079-3

Pederson T. (2010) Regulatory RNAs derived from transfer RNA? *RNA* **16**: 1865–1869. doi: 10.1261/rna.2266510

Podkowinski J, Zmienko A, Florek B, Wojciechowski P, Rybarczyk A, Wrzesinski J, Ciesiolka J, Blazewicz J, Kondorosi A, Crespi M, Legocki A (2009) Translational and structural analysis of the shortest legume ENOD40 gene in Lupinus luteus. *Acta Biochim Pol* **56**: 89–102

Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* **40**: e112. doi: 10.1093/nar/gks339

Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**: 129. doi: 10.1186/1471-2105-11-129

Ross J (1995) mRNA stability in mammalian cells. *Microbiol Rev* **59**: 423–450

Rybarczyk A, Szostak N, Antczak M, Zok T, Popenda M, Adamiak R, Blazewicz J, Szachniuk M (2015) New *in silico* approach to assessing RNA secondary structures with non-canonical base pairs. *BMC Bioinformatics* **16**: 276. doi: 10.1186/s12859-015-0718-6

Sobala A, Hutvagner G (2011) Transfer RNA-derived fragments: Origins, processing, and functions. *Wiley Interdiscip Rev RNA* **2**: 853–862. doi: 10.1002/wrna.96

Szostak N, Royo F, Rybarczyk A, Szachniuk M, Blazewicz J, del Sol A, Falcon-Perez JM (2014) Sorting signal targeting mRNA into hepatic extracellular vesicles. *RNA Biol* **11**: 836–844. doi: 10.4161/rna.29305

Schweingruber C, Rufener SC, Zund D, Yamashita A, Muhlemann O (2013) Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta* **1829**: 612–623. doi: 10.1016/j.bbagrm.2013.02.005

Szweykowska-Kulinska Z, Jarmolowski A, Figlerowicz M (2003) RNA interference and its role in the regulation of eukaryotic gene expression. *Acta Biochim Pol* **50**: 217–229.

Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS (2009) Small RNAs derived from snoRNAs. *RNA* **15**: 1233–1240. doi: 10.1261/rna.1528909

Taguchi G. (1987) System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs, *UNIPUB/Kraus International Publications*

Thompson DM, Lu C, Green PJ, Parker R (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14**: 2095–2103. doi:10.1261/rna.1232808

Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**: 487–517. doi: 10.1146/annurev.micro.59.030804.121336

Yang H, Jossinet F, Leontis NB, Chen L, Westbrook J, Berman H, Westhof E (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* **31**: 3450–3460. doi: 10.1093/nar/gkg529

Zhang JC, Styblinski M.A. (1995) Yield and Variability Optimization of Integrated Circuits. Springer Science, New York: 199-202. ISBN 978-1-4613-5935-7

Zhang S, Sun L, Kragler F (2009) The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation. *Plant Physiol* **150**: 378–387. doi: 10.1104/pp.108.134767

Zok T, Popenda M, Szachniuk M (2014) MCQ4Structures to compute similarity of molecule structures. *Central Eur J Operations Res* **22**: 457–474. doi: 10.1007/s10100-013-0296-5