# Sequence-specific endoribonucleases


## ABSTRACT

**Ribonucleases are nucleolytic enzymes that commonly occur in living organisms and act by cleaving RNA molecules. These enzymes are involved in basic cellular processes, including the RNA maturation that accompanies the formation of functional RNAs, as well as RNA degradation that enables removal of defective or dangerous molecules or ones that have already fulfilled their cellular functions. RNA degradation is also one of the main processes that determine the amount of transcripts in the cell and thus it makes an important element of the gene expression regulation system. Ribonucleases can catalyse reactions involving RNA molecules containing specific sequences, structures or sequences within a specific structure, they can also cut RNAs non-specifically. In this article, we discuss ribonucleases cleaving the phosphodiester bond inside RNA molecules within or close to particular sequences. We also present examples of protein engineering of ribonucleases towards the development of molecular tools for sequence-specific cleavage of RNA.**


## INTRODUCTION

Ribonucleases (RNases) belong to the class of transferases (phosphotransferases, EC 2.7), or to the class of hydrolases (esterases, EC 3.1), enzymes that catalyse the hydrolysis of phosphodiester bonds in the ribonucleic acid (RNA). RNases can be classified according to the location of the hydrolysed bond in the RNA polynucleotide chain into: exoribonucleases that cleave the bond connecting the terminal nucleotide residue in the chain, and endoribonucleases that cleave bonds inside the RNA molecule. Additionally, RNases are classified according to their substrates, hence RNases can be grouped into those acting on single-stranded RNA (ssRNA) or double-stranded RNA (dsRNA). Another way of RNase classification takes into account the degree of similarity of their sequences and structures, and their inferred evolutionary relationships. At the highest level of this classification, RNases are grouped according to types of three-dimensional folds (architectures) of their catalytic domains. Depending on the dominance of one type of secondary structures, one can distinguish folds that are dominated by α helices or β sheets, and if there is no clear dominance of one of these structures, mixed α/β or α+β folds [1]. The majority of RNases (also the majority of those described in this article) have folds of a mixed type, α/β or α+β. At the next level of classification, RNases can be grouped into superfamilies that encompass evolutionarily related proteins exhibiting structural similarity and often also functional one (similar biochemical properties) (Tab. 1). All organisms, whose genomes have been sequenced so far, contain genes encoding RNases that belong to various superfamilies and classes, and fold into various types of structures. It is postulated that a large number of RNases were also present in the last universal common ancestor of all living organisms, LUCA [1], which strongly suggests that the RNA cleavage process existed already at a very early stage of evolution of the cellular life.

RNases participate in many cellular processes and are of key importance for proper functioning of cells. Apart from RNA degradation, which is one of necessary elements of gene expression regulation, these enzymes take part in the maturation of all kinds of RNA, including both protein-coding mRNAs and the structurally and functionally diverse group of non-coding RNAs [2]. Moreover, RNases constitute an important component of the cellular defence system which is used, for instance, against viruses [3].

Despite significant progress in our knowledge of the relationship between the RNA sequence, structure and function, the research on RNA biology is still hampered due to the lack of many crucial enzymatic tools that would allow performing precise and planned manipulations of RNA molecules *in vitro*. RNases specifically cleaving RNA sequence seem to be ideal molecular tools that have the potential to revolutionize RNA research in the near future, as it happened in


**Dawid Głów**[1]

**Martyna Nowacka**[1]

**Krzysztof J. Skowronek**[1,✉]

**Janusz M. Bujnicki**[1,2,✉]

[1]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
[2]Department of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland

✉Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 4 Ks. Trojdena St., 02-109 Warsaw, Poland; e-mail: kskowronek@iimcb.gov.pl (Krzysztof Skowronek); e-mail: iamb@genesilico.pl (Janusz Bujnicki)



Received: June 13, 2016
Accepted: June 21, 2016

**Key words:** protein engineering, ribonucleases, RNA cleavage, RNA-protein complexes


**Abbreviations:** dsRBD – double-stranded RNA Binding Domain; RBM – RNA binding motif; dsRNA – double-stranded RNA; nt – nucleotide; RNase – ribonuclease; ssRNA – single-stranded RNA.


**Acknowledgement:** Research on RNases in the authors' laboratory was supported by Polish National Centre for Research and Development [NCBR, grant 245550 to J.M.B.]; European Research Council [ERC, StG grant RNA+P=123D grant to J.M.B, PoC grant eRNases to J.M.B.] and Polish National Science Centre [NCN, grant 2011/01/B/NZ1/00209 to K.J.S]; D.G. was additionally supported by scholarships from the European Social Fund, Human Capital Operational Programme [UDA-POKL.08.02.01-14-041/09]. J.M.B was also supported by the "Ideas for Poland" fellowship from the Foundation for Polish Science.

**Table 1.** RNases covered in this review

| RNase | Occurrence | Superfamily | Fold type | Cleaved sequence | Comments |
|---|---|---|---|---|---|
| Natural sequence-specific RNases | | | | | |
| RNase A | Mammals, some amphibia | α/β | RNase A | Y^N, | Y = C or U |
| RNase V (angiogenin) | Vertebrates | α/β | RNase A | C^N | U^N cleaveage is 12-times less efficient than C^N cleavage |
| RNase T1 | Fungus | α/β | RNase T1 | G^N | |
| RNase U2 | Fungus | α/β | RNase T1 | R^N | R = G or A |
| α-sarcin | Fungus | α/β | RNaseT1 | GAG^A | Loop sequence |
| RegB | T4 bacteriophage | α/β | RelE/YoeB | GG^AG | 3′ G of GGAG sequence must be base-paired; ribosomal protein S1 activates the cleavage |
| MazF | Bacteria | α/β | CccB/MazF/Kid | Depends on the MazF | |
| Cas9 | Bacteria, bacteriophages | α/β | HNH & RNase H-like | CCNNNNN^N | Specifically recognizes 20 nt sequences complementary to guide RNA (PAM) and the HNH domain cleaves the ssRNA substrate at 4 resiudes from PAM |
| RNase III | Bacteria, bacteriophages, some fungi | α | RNase III | U^NNANA/ U^NNANA | In dsRNA |
| MiniIII | Gram-positive bacteria, plants' plastids | α | RNase III | SWWS | Cleaved sequence depends of organism of origin; L3 ribosomal protein activates the cleavage of natural substrate *in vitro* |
| Sequence specific fusion RNases | | | | | |
| RNase A + oligonucleotides | – | – | – | Depends on oligonucleotide sequence | |
| RNase H + oligonucleotides | – | – | – | Depends on oligonucleotide sequence | Cleaves the RNA strand in a DNA/RNA hybrid |
| RNase H + Zinc finger domain | – | – | – | ^NNNNNUUCUUCCCC | Cleaves the RNA strand in a DNA/RNA hybrid |
| PIN + PUF domain | – | – | – | UU^GAUAUAUUCG^ | - |
| RNase T1 + RNA binding motifs | – | – | – | Depends on the sequence recognized by the RBM | |

the case of restriction enzymes used for DNA manipulation. Such a revolution in the techniques of molecular biology of RNA could have a significant influence on both basic and applied research. Specific RNases able to catalyse cleavage of specific RNA sequence could in future find application in diagnostics or therapy of some genetic diseases as well as perhaps be used in antiviral therapies. As a result, research on RNase specificity and on molecular mechanisms of RNase-RNA interactions are very important in the context of searching for natural molecular tools and creating new ones.

The best way to uncover the mechanisms of enzyme specificity is obtaining the structures of enzyme-substrate complexes at the atomic resolution. This goal is attainable with X-ray crystallography or NMR spectroscopy. The availability of a high-resolution structure makes it possible to establish the functional role of individual amino acid residues in the substrate selection process. In a further perspective, such knowledge enables protein engineering which is aimed at, for example, increasing substrate specificity or its alteration by manipulating non-covalent interactions between an enzyme and its substrate. The structures of numerous RNases, including sequence-specific ones, have been determined experimentally. Unfortunately, so far there are not many structures available of RNase complexes with RNA. Additionally, in many structures of RNase complexes, RNases are bound only with fragments of the substrate, such as nucleotides or their analogs (Tab. 2).

In this article, we focus on describing endoribonucleases acting on various types of RNA molecules and having

**Table 2.** List of the structures mentioned in the text

| RNase | Structure determination method | Ligands | Organism | PDB accession number |
|---|---|---|---|---|
| RNase A | Crystallography | DNA (5′-ATAAG-3′) | *Bos taurus* | 1RCN (complex with homolog of substrate) |
| RNase V (angiogenin) | Crystallography | – | *Homo sapiens* | 1ANG |
| RegB | NMR | – | Bacteriophage T4 | 2HX6 |
| αSarcin | Crystallography | – | *Aspergillus giganteus* | 1DE3 |
| RNase T1 | Crystallography | Guanosine-2′,3′-cyclic phosphorothioate and 3′-GMP | *Aspergillus oryzae* | (2-6)GSP |
| | | 2′,3′-CGPS | | 1GSP, 7GSP |
| RNase U2 | Crystallography | 3′AMP | *Ustilago spharogena* | 3AGN, 3AGO |
| MazF | Crystallography | RNA | *Bacillus subtilis* | 4MDX |
| | | DNA homologue of substrate | *Escherichia coli* | 5CR2 |
| RNase III | Crystallography | 2 × 28 nt RNA | *Aquifex aeolicus* | 2EZ6 (complex with products) 1YZ9, 1YYW, 1YYO, 1YYK (complex with intermediates) |
| MiniIII | Crystallography | – | *Bacillus subtilis* | 4OUN |

preferences for cleaving RNAs with particular ribonucleotide sequences. We also describe molecular mechanisms that govern these preferences. It is worth mentioning that substrate specificity of a large number of RNases is limited to preferences for very short sequences, often in the context of a specific secondary RNA structure. Therefore, we subjectively selected enzymes to describe examples of the best studied and arguably the most interesting RNases which specifically recognize and cut their substrates (Tab. 1). In the second part of the article, we present examples of specificity engineering of these enzymes and the potential of such manipulations for creating tools useful in biotechnology.

## NATURALLY OCCURRING, SPECIFIC ENDORIBONUCLEASES ACTING ON SINGLE-STRANDED RNA

### RNase A

An overview of RNases exhibiting substrate specificity towards the sequence, we start with pancreatic Ribonuclease A, a model representative of this class of enzymes which has been intensively investigated for over 50 years. RNase A is one of the best characterized enzymes [4]. It is worth emphasizing that it was one of the first proteins in history, and also the first enzyme, for which the amino acid sequence was determined, and also one of the first proteins and enzymes whose three-dimensional structures were determined by X-ray crystallography (after myoglobin, haemoglobin and lysozyme). RNase A was first crystallized over 50 years ago and the resolution of diffraction data, which was obtained at that time, was 2 Å [5]. Christian B. Anfinsen, Stanford Moore and William H. Stein received the Nobel Prize in chemistry in 1972 for their research devoted to this protein.

RNase A catalyses the hydrolysis of phosphodiester bonds of RNA molecules on the 3′ side of nucleotide residues which contain pyrimidine bases (the cleaved sequence is Y^N, where Y is C or U, N is any residue and ^ indicates the cleavage site), however the CN dinucleotide is cut twice as fast as the UN dinucleotide (Tab. 1).

Structural data show that the specificity of RNase A results from interactions between the pyrimidine and side chains of Thr45 and Phe120 (1RCN, Fig. 1, Tab. 2). Thr45 forms hydrogen bond only with the pyrimidine base, which results in spatial exclusion of the purine base (A, G and I). The side chain of the Phe120 takes part in van der Waals interactions with the pyrimidine and with the Thr45 side chain. Apart from these two residues, which are the most important for RNase A specificity, Asp83 and Ser123 also participate in binding of the pyrimidine residue forming

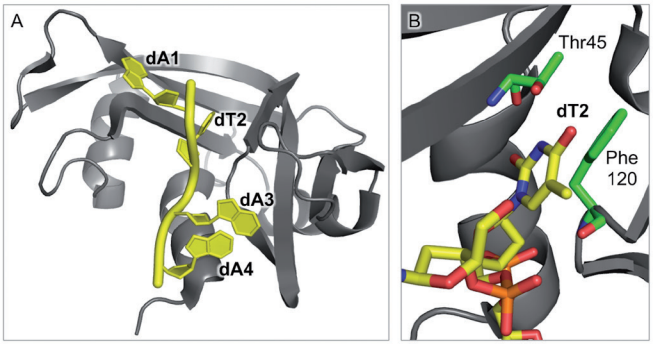


**Figure 1.** Crystal structure of RNase A in complex with d(ApTpApG) (1RCN) (A). RNA is represented in yellow; protein residues involved in base-specific contacts are marked in green. Detailed representation of sequence-dependent interactions (B).

bonds with the phosphate group [6]. RNase A, upon binding its substrate, interacts with four RNA residues around the cleavage site, two on the 5′ side and two on the 3′ side. These bonds do not depend on the RNA sequence. Interactions between RNase A and RNA were presented and discussed by Fontecilla-Camps *et al*. [7].

RNase A is an eponymous member of the RNase A superfamily that groups many various proteins with ribonucleolytic activity. These enzymes have similar structures and, to some degree, amino acid sequences, especially within the evolutionarily conserved active centre.

RNase V (ANGIOGENIN)

RNase V is an RNase A superfamily member. It is also called angiogenin as it plays an important role in the process of angiogenesis, the development of new blood vessels. This process occurs during the tissue development of higher organisms, but it is also of key importance in cancer progression.

Human RNase V catalyses the cleavage of 18S, 28S rRNA [8] and 5S rRNA *in vitro* [9] as well as tRNA after injection into *Xenopus* oocytes [10]. In human cells, angiogenin is also activated during oxidative stress, heat shock or ultraviolet irradiation, and cleaves tRNA leading to translational repression [11]. It seems that the substrate specificity of angiogenin is determined not by the RNA sequence itself but rather its secondary structure (the hairpin or pseudoknot), sometimes formed as a result of RNA interactions with other proteins [12].

The amino acid sequence of RNase V shows considerable similarity (33% identity) to the amino acid sequence of RNase A [13]. Available crystallographic structures of both nucleases also show high similarity (1RCN, 1ANG, Tab. 2). The region of angiogenin that corresponds to the pyrimidine residue binding pocket in RNase A contains many identical or very similar elements. In angiogenin, Thr44 and Ile115 correspond to Thr45 and Phe120 of RNase A, and Thr80 replaces Asp83 [14,15]. In this case, the role of Thr44 is as important as that of Thr45 in RNase A, and substitution of any of these residues influences the nucleolytic activity of angiogenin [16]. Despite numerous similarities, differences in the structure of the active centres of the two nucleases can also be noticed. The pocket that binds the pyrimidine residue within the active centre of angiogenin, contains an additional side chain of Glu117. This change may be responsible for the significant reduction in endonucleolytic activity of RNase V (by approx. $10^5$–$10^6$) as compared to RNase A [8]. The presence of an additional chain in the active centre may also restrict RNase V specificity. It cuts sequences containing a C residue (within the C^N site) 12 times more efficiently than sequences that contain U (in the context of the U^N site) (Tab. 1) [17], which means a six-fold increase in selectivity as compared to RNase A.

On several occasions, the specificity of angiogenin was engineered. An example is the substitution of Thr44 by Asp, which changed the preference of the enzyme to U^N [16], or obtaining an angiogenin chimera containing corresponding RNase A residues (38-42) in positions 38-41. The chimeric RNase had activity that was similar to the activity of RNase A [18]. Crystallographic structures of both engineered proteins are available [19], however, the structure of RNase V in the complex with RNA is so far unknown.

RNase T1

RNase T1 was discovered in 1957 in a cell extract from *Aspergillus oryzae* [20]. Since then it was a subject of numerous investigations; nonetheless its physiological substrate remains unknown. The enzymatic activity of RNase T1 is restricted to the cleavage of the ssRNA after guanine residue (within G^N sites) (Tab. 1) [21]. Structures of RNase T1 and its mutants were determined at high resolution by X-ray crystallography [22] and NMR spectroscopy [23]. The structural core of this enzyme consists of four antiparallel β strands and one α helix [24]. Several structures of RNase T1 complexed with single and di- or tri-nucleotides (or nucleotide analogs) are available as well. The most interesting are the structures analysed by time-resolved crystallography [25], which represent different stages of this reaction at high resolution (1-7GSP, Fig. 2, Tab. 2) and therefore allow precise dissection of the hydrolysis of guanosine 2′,3′-cyclophosphorothioate by RNase T1. RNA hydrolysis conducted by RNase T1 can be divided into two steps: formation of intermediates by transphosphorylation of substrate and hydrolysis of these intermediates to a product with 3′ phosphate. The guanine base specificity of RNase T1 is also well understood. Guanine is mainly bound by the segment Tyr42-Glu46. Hydrogen bonds are formed between the nucleobase and the N-H groups of the peptide bonds of Asn43, Asn44, Tyr45, and the side chain of Glu46. Furthermore, the guanine is sandwiched between two tyrosine residues (Tyr42 and Tyr45). It forms a single hydrogen bond with the carbonyl group of Asn98, which is situated outside of the Tyr42-Glu46 segment. Conservation of this segment among the RNase T1 family members is very high [26] and these enzymes are specific for G or at least prefer G-containing RNA as a substrate. However, most of the contacts are formed between the nucleobase and main-chain peptide bonds. Therefore the reason for the conservation of this segment is unclear.

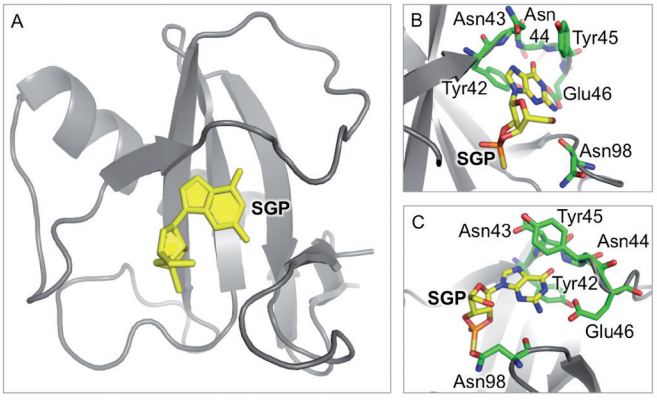


**Figure 2.** Crystal structure of RNase T1 in complex with 2′,3′-CGPS and 3′-GMP (2GSP) (A). Substrates are represented in yellow; protein residues involved in base-specific contacts are marked in green. Detailed representation of sequence-dependent interactions (B and C).

RNase U2

Although RNase U2 belongs to the RNase T1 family, it constitutes a clear exception in substrate specificity. This enzyme, apart from the guanine residue, also recognizes A-containing substrates (Tab. 1). A decrease in the specificity of this RNase is the result of differences in the structure of the pocket, which binds nucleobases. The Tyr42 – Glu46 region of RNase T1 corresponds to the Tyr44 – Asp50 region in RNase U2 [27]. Base-specific contacts are also different than those in RNase T1. Adenine forms four hydrogen bonds to the main-chain atoms of Tyr44 and the side-chain carboxyl group of Glu49 (3AGN, 3AGO, Tab. 2) [27].

α-SARCIN

α-Sarcin RNase is produced by a mould, *Aspergillus giganteus* [28] and it represents a group of ribonucleases that belong to the group of ribotoxins related to the RNase T1 family. α-Sarcin blocks the activity of the ribosome by cleaving the phosphodiester bond on the 3′ side of the G4325 residue in 28S rRNA. This nucleotide residue is located in the so-called α-Sarcin/ricin loop, whose cleavage prevents the binding of the EF2 and EF1 transcription factors by the ribosome and leads to ribosome inactivation. There is a range of structural and biochemical similarities between α-Sarcin and RNase T1. Just like in RNase T1, the core of the α-Sarcin structure consists of five β strands and a single α helix [29]. The active centres of both nucleases consist of identical amino acid residues – His50, Glu96, Arg121 and His137 [29]. As a result, the chemical reaction catalysed by α-Sarcin proceeds exactly in the same way as the one catalysed by RNase T1 [30]. Biochemical tests have shown that α-Sarcin recognizes and cuts most efficiently the GAG^A sequence situated in the α-Sarcin/ricin loop of 28S rRNA. It has also been shown that this sequence must be located in the context of an appropriate secondary structure (Tab. 1), in this case, a double-stranded RNA helix consisting of at least three base pairs [31,32]. An analysis of the α-Sarcin complex model with RNA (1DE3, Tab. 2) revealed two structural elements that are probably responsible for the substrate specificity of this protein. These are: a fragment of loop 2 (Trp51-Gly55), which could recognize the GAGA sequence, and a lysine-rich fragment of loop 3, which could be responsible for recognizing and binding of an appropriate secondary RNA structure [29].

RegB RNase

RegB is one of the most "mysterious" nucleases. It is encoded by the T4 bacteriophage and the majority of T4-like bacteriophages [33]. RegB is an important factor involved in controlling bacteriophage multiplication. The expression of the RegB coding gene takes place in bacterial cells at an early stage of bacteriophage infection. This nuclease participates in regulation of the transition between the early and late infection phases by inactivation of early-phase mRNA [34]. Under *in vivo* conditions, RegB specifically cuts RNA in the middle of the GGAG sequence located at the ribosome-binding site in the majority of prokaryotic mRNAs (Shine-Dalgarno sequences, AGGAGG). Another RNA sequence cleaved by RegB, although much less efficiently, is GGAU. It should be emphasized that RegB is a transphosphorylase that generates a cutting product with a phosphate group at the 3′ end. Interestingly, RegB retains its catalytic activity throughout the entire lytic cycle of the T4 bacteriophage, but it does not cleave transcripts of the intermediate and late phases [35]. Moreover, GGAG motifs, located in the coding sequences (and not at the ribosome-binding site) of early mRNA, are not cut either. The *in vitro* activity of RegB increases over 100-fold (depending on the substrate) if the ribosomal S1 protein is present in the reaction [35]. The guanine residue from the 3′ end of the GGAG motif in the preferred RegB substrate *in vitro* is in the first base pair of a double-stranded structure (Tab. 1). Therefore, it seems very likely that RegB recognizes not only the nucleotide sequence but also the specific secondary RNA structure. The formation of such a structure may depend on the presence of the S1 protein [36]. Despite the fact that the RegB structure was solved (2HX6, Tab. 2), it is not known how exactly this RNase selects its RNA substrate [37].

MazF RNase

MazF RNase is a ribotoxin, one of two elements of the so-called toxin-antitoxin system. Under physiological conditions, MazF is combined with a protein that constitutes the so-called antitoxin (MazE), which blocks the RNase active centre. In such a system, MazF is inactive and it cannot catalyse the RNA cleavage reaction. Under stress conditions, ClpAPP proteinase degrades MazE [38] and activates MazF. MazF activation leads to a change in the gene expression level and, consequently, it may lead to programmed cell death [39, 40]. It was previously thought that only mRNA molecules are the substrate of this nuclease; however, recent data show that also tRNA molecules are recognized and cut by MazF [41]. It is worth emphasizing that MazF nucleases from various bacteria cleave different RNA sequences. MazF from *Escherichia coli* (EcMazF) recognizes U^AC [42] or A^CA sequences [43]. MazFs from other bacteria recognize sequences of various lengths that often contain the ACA sub-sequence. For example: MazF from *Bacillus subtilis* (BsMazF) recognizes and cuts the U^ACAU motif [44], and ChBK, a second MazF-like nuclease encoded in the *E. coli* chromosome, cuts the D^ACD sequence (where D stands for G, A, or U) [45]. On the other hand, *Mycobacterium tuberculosis* has as many as seven different genes encoding MazF-like toxins (MazF-mc1-7). Some of them recognize and cut the same sequence as MazF from *E. coli*, while others show different sequence preferences. MazF-mt3 cuts the U^C-CUU sequence [46], and MazF-mt7 the U^CGCU sequence [47] (Tab. 1). It should be emphasized that not all sequences containing motifs recognized by particular MazF proteins are cleaved by these enzymes, which implies that secondary structures (or their absence) may also play a role in the substrate recognition process [48,49].

The structures of complexes of representatives of the MazF group of RNases with RNA are known. The BsMazF structure was described first (4MDX, Tab. 2) [50]. It helped to elucidate the mechanism of specific interaction of BsMazF with all five residues of its RNA substrate motif (U^ACAU). Uracil in position 1 is bound in the pocket formed by Pro26, Ala49, Glu70, Arg71 and Ser73 residues; however, hydro-

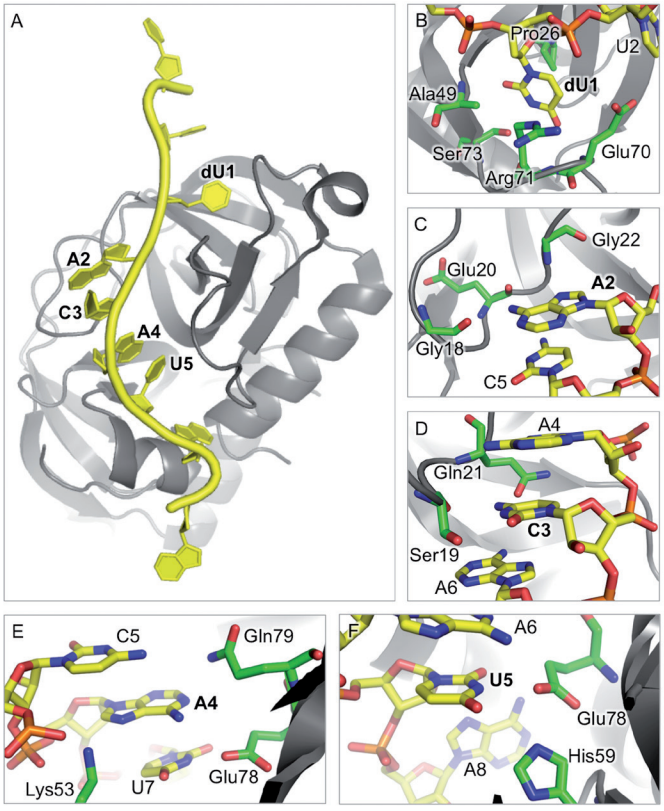


**Figure 3.** Crystal structure of RNase MazF in complex with ssRNA (4MDX) (A). RNA is represented in yellow; protein residues involved in base-specific contacts are marked in green. Detailed representation of sequence-dependent interactions (B-F).

gen bonds with the Glu70 and Ser73 residues involve Watson-Crick edges. Adenine in the second position of the motif forms hydrogen bonds through its Hoogsteen edge with atoms of carbonyl and amide N-H groups of Gly18, Glu20 and Gly22 residues. Cytosine in position 3 forms hydrogen bonds with Gln21 and Ser19. Adenine in position 4 forms hydrogen bonds through Watson-Crick and Hoogsteen edges with Glu78, Gly79 and Lys53, while uracil in position 5 forms a hydrogen bond with Glu78 and a weak polar interaction with His59 (Fig. 3 and Fig. 4) [50].

Another available structure of an MazF complex with RNA is the EcMazF structure (5CR2, Tab. 2). The amino acid sequence of this protein, which recognizes and cuts the N^ACA motif, shows only 23% identity with BsMazF. Differences in the amino acid sequences of these proteins are responsible for differences in the structures of the segments responsible for substrate binding. The EcMazF pocket, which binds the first nucleotide residue of the NACA sequence, is much deeper than in BsMazF. Therefore, EcMazF may probably also bind purine residues. This, in turn, may account for the reduced specificity of EcMazF as compared to BsMazF. Apart from the differences in the binding of the first residue, the majority of interactions between the EcMazF protein and its substrate are the same as in the case of BsMazF. An exception is the hydrogen bond formed by Glu78 and Gln79 residues, which are not retained in the EcMazF-RNA complex. These amino acid residues in EcMazF

U^ACAU
1 2 3 4 5

| Ribonucleotide residue | Amino acid residues |
|---|---|
| 1U | Pro26, Ala49, Glu70, Arg71, Ser73 |
| 2A | Gly18, Glu20, Gly22 |
| 3C | Gln21, Ser19 |
| 4A | Glu78, Gly79, Lys53 |
| 5U | Glu78, His59 |

**Figure 4.** Schematic representation of sequence-specific interactions between BsMazF and RNA. The sequence of RNA recognized and cleaved by BsMazF is shown, where ^ indicates the cleavage site. Specific interactions between BsMazF and each nucleotide of the RNA sequence are presented in the table.

are replaced by Asp76 and Gln77. A shorter side chain of Asp76 prevents the formation of the hydrogen bond, while in the case of Gln77, the side chain is located too far from the RNA to form a contact. In the case of BsMazF, the amino acid residue corresponding to Asp76 binds uracil in position 5 (Fig. 3). It could explain the lack of specificity of EcMazF for this position of its substrate. In summary, EcMazF recognizes the ACA sequence; therefore, it does not recognize the first and last nucleotide residue in the BsMazF substrate, namely UACAU. Additionally, the absence of hydrogen bonds allows the "sliding" of the recognized sequence in the binding pocket. As a result EcMazF can cut RNA at both the U^AC site and the A^CA site [51].

RNase Cas9

Cas9 (CRISPR associated protein 9) is a family of endonucleases that cleave nucleic acid molecules (usually DNA, but in some cases also RNA) base-paired to a targeting RNA molecule. To describe the specificity of the Cas9 nuclease, it is necessary to present the mechanism of the CRISPR system (Clustered Regularly Interspaced Short Palindromic Repeats), an "immune" mechanism of prokaryotic organisms. In many bacteria and the majority of archaea, CRISPR forms special loci that provide an acquired immunity to bacteriophages and plasmids. This immunity involves sequence-dependent recognition and fragmentation of foreign genetic material. Host immunity is acquired through prior contact with a given foreign DNA, which is "remembered" by the insertion of a fragment of this DNA between repetitions of short sequences situated in the CRISPR locus. Due to the diversity of known classes and types of this mechanism, we will present only class 2, type II of the CRISPR. In this case, the process of degradation of foreign DNA requires the presence of CRISPR RNA molecules (crRNA) and trans-activating crRNA molecules (tracrRNA), which are encoded in the CRISPR loci. It is the crRNA sequence that determines the specificity of CRISPR. crRNA binds to tracrRNA and forms a double-stranded structure bound by the Cas9 nuclease, which is an executive part of this system. After forming the complex, Cas9 cuts sequences that are complementary to crRNA. The Cas9 nuclease most frequently used in biotechnology is that from *Streptococcus pyogenes*, which recognizes sequences 20 nucleotides long, adjacent to the PAM sequence (protospacer-adjacent motif), NGG

[52]. The cleavage site is situated between the 3rd and 4th base pair from the PAM sequence (CCNNNN^ in the target DNA sequence and ^NNNNGG in the non-complementary sequence). Cas9 contains two nuclease domains that belong to the HNH and RNase H-like nuclease superfamily, respectively. The HNH domain is responsible for cleaving the complementary sequence (with CCN) and the RNase H-like domain is responsible for cleaving the non-complementary sequence (with NGG). [53].

The CRISPR mechanism has became the basis for producing tools for very specific recognition and cleavage of various DNA sequences. tracrRNA and crRNA were replaced with one molecule, the so-called sgRNA (single guide RNA) or gRNA (guide RNA), whose structure resembles the structure of the crRNA-tracrRNA complex. The sequence of the 5′ end of this molecule determines the cleavage site in the complementary DNA strand, while the sequence of the 3′ end allows the production of a double-stranded RNA structure bound by Cas9.

The first Cas9 nuclease which specifically cleaves RNA but not DNA was discovered in 2013. It was Cas9 from *Francisella novicida*, which forms a complex with a unique scaRNAs (small, CRISPR/Cas-associated RNA) [54] and hydrolyses RNA transcripts encoding bacterial lipoproteins. In subsequent studies, the same nuclease was used for sequence-dependent cleavage of viral RNA in eukaryotic cells [55]. Independently, another research group showed the possibility of programmed cleavage of ssRNA molecules by Cas9 from *Streptococcus pyogenes*. The use of the so-called PAMmers, i.e. DNA oligomers with a length of at least 15 nt complementary to the ssRNA sequences whose 5′ end began from the PAM sequence (NGG), considerably increased the efficiency of the cleavage reaction. In contrast to DNA cleavage by Cas9, ssRNA is cut between the 4th and 5th nucleotide before the PAM sequence (Fig. 5, Tab. 1) [56]. In practice, this means that it is possible to design a nucleolytic specificity of Cas9 for any RNA sequence of 20 nucleotide residues, which contains a sequence complementary to PAM, i.e. CCN, at the 5′ end. This makes Cas9 one of the most specific ribonucleases available. Adjustment of the Cas9 specificity is possible by changing a fragment of the gRNA sequence and designing an appropriate PAMmer.

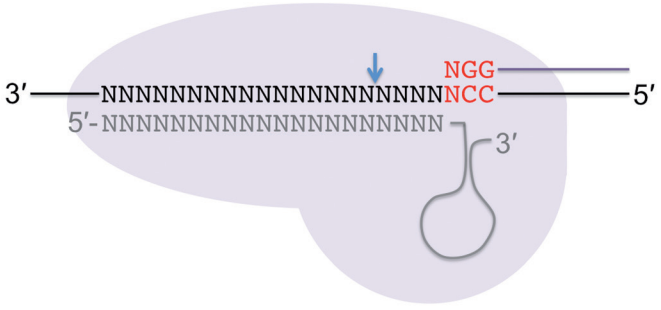


**Figure 5.** Schematic representation of Cas9 Nuclease, *S. pyogenes* sequence recognition and RNA cleavage. The cleaved RNA molecule is presented in black. gRNA is presented in grey. PAM sequence is marked in red. PAM together with the purple line stand for the DNA PAMmer, which is required for sequence-specific cleavage of RNA molecule. The arrow indicates the cleavage site between nucleotides 4 and 5 downstream of the NCC sequence complementary to the PAM sequence.

The Cas9-based technology is currently one of the fastest-developing technologies based on molecular biology tools. The only limitation in designing RNA-specific Cas9 is the requirement to retain the PAM sequence next to the target sequence. Recent results demonstrate that it is possible to change the specific recognition of the PAM sequence by Cas9 nucleases specific for DNA. Therefore, it seems that similar engineering of RNA-cutting Cas9 nucleases is possible and, as a result, development of Cas9 RNases that cleave any desired ssRNA sequences should be possible.

## NATURALLY OCCURRING, SPECIFIC ENDORIBONUCLEASES ACTING ON DOUBLE-STRANDED RNA

### RNase III

A superfamily of enzymes related to RNase III can simultaneously cut both strands of dsRNA. They are characterized by the presence of a unique α-helical domain that is capable of dimerization to form a structure with two active centres, each responsible for cutting one of the RNA strands in dsRNA. Amino acid residues that are evolutionarily conserved among all RNase III superfamily members are responsible for binding divalent metal cations and phosphodiester bond cleavage. The products of the reaction catalysed by RNase III enzymes usually contain 3′ sticky ends that are two residue long. In addition to the RNase III domain, enzymes of this superfamily may contain other structural elements. These additional elements are mostly responsible for the selection of the RNA substrate, interactions with it or with other proteins, and regulation of the enzymatic activity [57-59]. Consequently, the presence of these additional elements is used for classification of RNase III family enzymes [60].

Eponymous representatives of the RNase III family (e.g. RNase III from *E. coli*) belong to class I. Enzymes from this class consist of a catalytic domain and a dsRNA binding domain (dsRBD) and they form homodimers. Crystallographic (2EZ6, 1YYK, 1YYO, 1YYW, 1YZ9, Tab. 2) and biochemical tests have shown that the selection of the substrate sequence requires contact of the protein with at least two regions in dsRNA. These regions, which are called the proximal box and distal box, are sequences with the length of 4 and 2 base pairs, respectively [61]. The third dsRNA region, which can also be important during the substrate selection process, is the so-called middle box [62,63]. Interactions between RNase III and the RNA are mediated by four RBM1-4 regions (RNA binding motifs), which are situated both in the dsRNA binding domain and in the catalytic domain (Fig. 6). In the structure of the RNase III complex from *Aquifex aeolicus* with dsRNA, it has been observed that the majority of the interactions take place between side chains of amino acid residues and riboses or phosphates of the RNA, which means that these interactions are sequence non-specific, but allow the selection of dsRNA molecules with a defined length and appropriate arrangement of the enzyme and substrate. An exception is the hydrogen bond formed between the conserved Gln157, and the O2

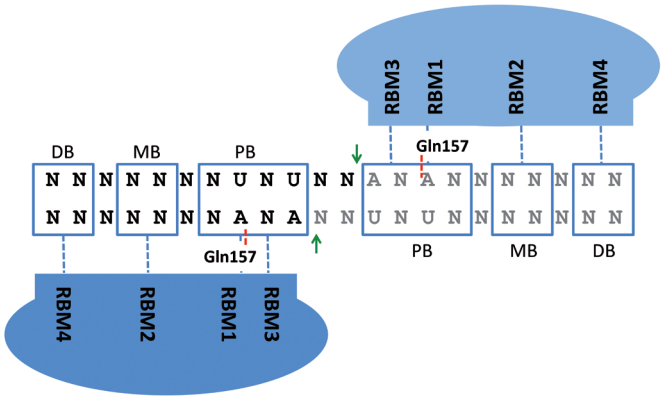


**Figure 6.** Schematic representation of the interactions between RNaseIII dimer and its dsRNA target sequence. The boxed fragments of sequence indicate distal box (DB) middle box (MB) and proximal box (PB). Interactions between RNA binding motifs and sequence boxes are indicated by dashed lines. Sequence-specific contacts are indicated by red dashed lines. Arrows indicate the cleavage sites.

atom of a base-paired adenine [63]. This glutamine residue is situated in RBM1, the N-terminal helix of the dsRNA-binding domain, and it interacts with the A/U base pair situated in the proximal part of the dsRNA. Substitution of this base pair by G/C inhibits the binding of dsRNA by RNase III [64]. Apart from this, despite the lack of specific interactions between RNase III and dsRNA in position 1 of the proximal box, the substitution of the A/U base-pair by G/C in this place also disrupts the binding of the substrate (Fig. 6, Tab. 1).

Enzymes belonging to class II of RNases III (with Drosha being their best-known representative), have two catalytic domains that interact with each other, a dsRBD domain and a polyproline domain. The structure of human Drosha has been recently determined [65], but in the absence of structural data concerning enzyme-substrate complex, the atomic details of the substrate selection mechanism for this class of enzymes remains unknown.

Class III of RNases III is represented by Dicer. This enzyme is composed of two slightly different catalytic domains, dsRBD domain, a PAZ domain, a domain with unknown function and an N-terminal domain with helicase activity. It is known that Dicer acts as a "molecular ruler" that measures the distance from one of the pre-miRNA ends (3′ or 5′), which is bound by the PAZ domain, to the cleavage site [66]. So far, approximately 2000 substrates of this RNase have been identified. They are very diverse, both, in terms of sequence and secondary structure [67]. In the case of perfect double-strand substrates, Dicer shows a detectable sequence preference. Dicer generates miRNA with the guanine residue at the 5′-end much less efficiently and it prefers an uracil residue in this position [68].

### MiniIII RNase

Class IV of RNases III consists of the smallest enzymes belonging to this family (hence their name, MiniIII RNases), because, as opposed to canonical RNase III, they do not contain a double-stranded RNA binding domain and comprise only the catalytic domain. Like classical RNase III, these enzymes are homo-dimeric. MiniIII-coding sequences are found in the genomes of Gram-positive bacteria and plant plastids. The MiniIII RNase from *Bacillus subtilis* (BsMiniIII) has been so far the most studied enzyme of class IV of RNases III. The physiological function of BsMiniIII is the cleavage of a double-stranded fragment formed by both ends of pre-23S rRNA in the last stage of its maturation [69]. The presence of the ribosomal L3 protein significantly increases the efficiency of this reaction *in vitro*, most likely by facilitating the recognition of the target site in pre-rRNA [70]. In plant plastids, MiniIII also participates in the degradation of introns that have been spliced out [71].

The main pre-rRNA cleavage site of BsMIniIII is situated at the end of a short double-stranded structure, in the ACC sequence. The cleavage specificity of BsMiniIII RNase has been also studied *in vitro* on long dsRNAS where it was found that this enzyme strongly prefers cleaving within the ACC^U sequence [72]. Despite the fact that the mechanism of substrate recognition by MiniIII has not yet been established, due to differences in the RNase III and MiniIII structures, it must be different from the mechanism identified in classic RNase III. Apart from the lack of double-stranded RNA binding domain, MiniIII distict due to the presence of an additional α-helix (α5b) and a short loop between the α5b and α6 helices in the catalytic domain, which replace a long loop present in classic RNase III. A model of the BsMiniIII complex (4OUN, Tab. 2) with dsRNA implies that this loop can play a significant role in the MiniIII substrate selection by interactions with the major groove of dsRNA. This prediction has been confirmed experimentally [72].

## SEQUENCE SPECIFIC RNASES CONSTRUCTED BY PROTEIN ENGINEERING

### RNase A FUSED WITH OLIGONUCLEOTIDES

The conjugation of a synthetic oligonucleotide to RNase A was probably the first attempt to obtain RNase targeting specific sequences in RNA. As a result of limited digestion of RNase A by subtilisin, active RNase S is formed. It can be separated into two components: S-peptide and S-protein (amino acid residues 1-20 and 21-124, respectively), which do not show any activity on their own. After mixing both proteins at a 1:1 stoichiometric ratio, they form a complex with activity comparable to native RNase A. S-peptide, which contains the cysteine residue in position 1, was synthesized and then covalently joined with a deoxyribooligonucleotide containing an S-tiopyridic residue at the 3′ end. The resulting fusion enzyme specifically cut the RNA sequence complementary to the sequence of the oligonucleotide used (Tab. 1) [73].

### RNase H FUSED WITH OLIGONUCLEOTIDES

RNase HI degrades the RNA strand of the double stranded DNA-RNA hybrid [74] of at least four base pairs [75]. In the structure of an RNase HI complex with a DNA/RNA hybrid, hydrogen bonds are formed between amino acid residues and the major and minor DNA/RNA grooves, leading to a weak sequence preference [76]. For this reason, this enzyme is basically sequence non-specific. However, RNase H from *E. coli* has a very weak preference for se-

quences containing the WNAGWGNNCWUNNN^NAWGNNCWCUNW motif.

The first attempt at RNase HI fusion aiming at creating a sequence-specific enzyme took place in 1992. The RNase H from *E. coli* used for this purpose, was first subjected to targeted mutagenesis. All cysteine residues were replaced with Ala and Glu135 was replaced with Cys. Then the GTCATCTCC DNA oligonucleotide with a maleimid group at the 5′ end was covalently attached to Cys135. Such a fusion of RNase H cleaves the RNA sequence complementary to the conjugated oligonucleotide between the 5th and the 6th ribonucleotide residue (GGAGA^UGAC) (Tab. 1) [77]. Another example is the fusion of RNase H with DNA oligonucleotides complementary to viral RNA. Specific conjugates were obtained which cut RNA of the hepatitis B virus *in vitro* [77] and RNA of hepatitis C RNA virus *in vitro* and *in vivo* [78].

### RNase H IN FUSION WITH A ZINC FINGER

Fusions of proteins with oligonucleotides require a complicated chemical procedure. It is much easier to produce fusion proteins with new domain combinations by fusing parts of various genes.

The zinc finger domain of the Cys2His2 type is a small protein domain that selectively binds a tri-nucleotide sequence in double-stranded nucleic acids. There are many classes of zinc fingers with various functions, however, the most important function from the point of view of this article is binding nucleic acids in a specific way and the possibility of engineering this specificity [79].

An appropriately designed and optimized fusion of the RNase H catalytic domain from *Bacillus halodurans* (originally without any particular sequence specificity) and the ZfQQR zinc finger resulted in an enzyme that exhibited sequence-specific cleavage of the RNA strand in a DNA/RNA hybrid. The zinc finger used for the development of this fusion protein specifically binds GGGGAAGAA DNA sequence, while RNase H cleaves the RNA strand between the 5th and the 6th base pair of the hybrid upstream the sequence bound by the zinc finger. The target sequence can be therefore presented as the following motif: N^NNNNNUUCUUCCCC [80].

### PIN NUCLEASE IN FUSION WITH PUF DOMAIN

The PIN domain from a non-specific nuclease SMC6 and the sequence-specific PUF domain (Pumilio/fem-3-binding domain) of the PUM1 protein were used to develop a recombinant sequence-specific ssRNA endoRNase. The catalytic function in this protein is fulfilled by the PIN domain, whose molecular structure has been quite well studied [81]. The PUF domain consists of modules, which specifically recognize and bind the consecutive ribonucleotide residues of the ssRNA sequence. Three amino acid residues in conserved positions of the PUF module and the Watson-Crick edge of the RNA molecule participate in these interactions. Substitutions of the amino acid residues in these conserved positions lead to altered specificity of the individual modules, which makes construction of PUF domains with any sequential specificities relatively easy [82]. For optimal activity of the PIN+PUF fusion, the linker between the domains must be of the right length. The use of 8 modules allowed the development of a nuclease that specifically binds the UUGAUAUA sequence and cleaves the RNA between the 4th and 5th nucleotide downstream of this sequence (UUGAUAUANNNN^N) (Tab. 1). This nuclease also cuts the RNA sequence inside the bound motif, but with lower efficiency [83,84]. These results indicate a great potential of engineering of such fusion proteins and their numerous potential applications.

### RNase T1 IN FUSION WITH RNA-BINDING MOTIFS

The fusion Tα RNase consists of RNase T1, in which part of the L3 loop has been substituted by the corresponding domain from α-sarcin. This modification was the first attempt to engineer specificity in RNase T1. The chimeric protein was designed in a way that allows the chimera to cut after a guanine residue not only in ssRNA but also in dsRNA [85]. The replacement of the L3 loop was also performed in order to develop an RNase T-tat fusion. The fragment of the L3 loop of RNase T1 was replaced by the TAT protein peptide of the HIV-1 virus. This peptide is responsible for binding to an early transcript of Tar RNA (the region responsible for transcription activation). The constructed chimera selectively recognized and hydrolysed HIV-1 TAR RNA both *in vitro* and *in vivo* (Tab. 1) [86].

## SUMMARY AND PERSPECTIVES

Ribonucleases are a highly diverse group of proteins. Many researchers have shown interest in their substrate specificity. There is no doubt, however, that the substrate specificities of endoribonucleases are much less studied than the specificities of enzymes which cleave DNA molecules, such as restriction enzymes, commonly used in molecular biology. The reasons for this lie, amongst other things, in the nature of the RNases. Many of them recognize not only a nucleotide sequence but also the RNA structure, which often makes it difficult to interpret the experimental results that compare the effectiveness of digestion of RNA substrates sequences. This is due to the fact that, as in the case of RNA a change of the nucleotide sequence may also cause a change in its structure. In addition, a change in the substrate structure may depend on many external and cellular factors, which are difficult to identify and even more difficult to control under *in vivo* conditions. Taken together, these make the determination of the properties of a given enzyme very difficult. While some RNases cut RNA non-specifically, the specificity of others is sometimes limited to a very short sequence or, in extreme cases, to the presence of a single nucleotide. RNases often recognize and cut several similar motifs in the RNA sequence with variable efficiency. The recognition of more than one sequence motif by a given enzyme and the functional redundancy of RNases in a cell result in a situation when a mutation leading to dysfunction of one of RNases is compensated by another enzyme from a given class without blocking RNA synthesis and/or degradation [2].

Structural biology is very helpful in understanding the molecular mechanisms determining the specificity of a given enzyme. The availability of the structure of an enzyme complex with its substrate is also important when engineering of a given protein is planned. RNase engineering has been successfully conducted for many years. It utilizes, amongst other things, the creation of chimeric RNases and fusions with other proteins. In this way, it was possible to combine the non-specific RNase H with a zinc finger domain and thus to obtain a fusion protein capable of sequence-specific cleavage. However, determination of more high-resolution structures of RNase complexes with RNA is essential to foster RNase engineering with the aim of creating new molecular tools.

Recent years have brought many fundamental discoveries, revealing completely new roles fulfilled by non-coding RNA molecules. The rapidly growing interest in RNA research exposes limitations resulting from the absence of enzymatic tools, analogous to those available in DNA research for years. One of basic causes of revolutionary changes in the 20th century biology, leading to the establishment of molecular biology, was the capability of planned and reproducible DNA fragmentation using restriction enzymes. It seems that a similar breakthrough in the 21st century might involve research on non-coding RNA and its biological functions. From a methodological point of view, this will certainly require new methods of RNA modification analysis using third-generation high-throughput sequencing of RNA molecules and mass spectrometry, but most of all, new enzymatic tools enabling planned *in vitro* RNA manipulations. It is possible that the results of sequence specificity studies of known endoribonucleases and first attempts at engineering their specificity presented in this article are an announcement of such breakthrough discoveries.

**REFERENCES**


1. Aravind L, Koonin EV (2001) A natural classification of ribonucleases. Methods Enzymol 341: 3-28
2. Houseley J, Tollervey D (2009) The many pathways of RNA degradation. Cell 136: 763-776
3. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709-1712
4. Beintema JJ, Kleineidam RG (1998) The ribonuclease A superfamily: general discussion. Cell Mol Life Sci 54: 825-832
5. Fankuchen I (1941) An X-Ray and Crystallographic Study of Ribonuclease. J Gen Physiol 24: 315-316
6. Wlodawer A, Sjolin L (1983) Structure of ribonuclease A: results of joint neutron and X-ray refinement at 2.0-A resolution. Biochemistry 22: 2720-2728
7. Fontecilla-Camps JC, de Llorens R, le Du MH, Cuchillo CM (1994) Crystal structure of ribonuclease A.d(ApTpApApG) complex. Direct evidence for extended substrate recognition. J Biol Chem 269: 21526-21531
8. Shapiro R, Riordan JF, Vallee BL (1986) Characteristic ribonucleolytic activity of human angiogenin. Biochemistry 25: 3527-3532
9. Rybak SM, Vallee BL (1988) Base cleavage specificity of angiogenin with *Saccharomyces cerevisiae* and *Escherichia coli* 5S RNAs. Biochemistry 27: 2288-2294
10. Saxena SK, Rybak SM, Davey RT, Jr., Youle RJ, Ackerman EJ (1992) Angiogenin is a cytotoxic, tRNA-specific ribonuclease in the RNase A superfamily. J Biol Chem 267: 21982-21986
11. Yamasaki S, Ivanov P, Hu GF, Anderson P (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. J Cell Biol 185: 35-42
12. Gao X, Xu Z (2008) Mechanisms of action of angiogenin. Acta Biochim Biophys Sin (Shanghai) 40: 619-624
13. Strydom DJ, Fett JW, Lobb RR, Alderman EM, Bethune JL, Riordan JF, Vallee BL (1985) Amino acid sequence of human tumor derived angiogenin. Biochemistry 24: 5486-5494
14. Acharya KR, Shapiro R, Allen SC, Riordan JF, Vallee BL (1994) Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. Proc Natl Acad Sci USA 91: 2915-2919
15. Leonidas DD, Shapiro R, Allen SC, Subbarao GV, Veluraja K, Acharya KR (1999) Refined crystal structures of native human angiogenin and two active site variants: implications for the unique functional properties of an enzyme involved in neovascularisation during tumour growth. J Mol Biol 285: 1209-1233
16. Curran TP, Shapiro R, Riordan JF (1993) Alteration of the enzymatic specificity of human angiogenin by site-directed mutagenesis. Biochemistry 32: 2307-2313
17. Shapiro R (1998) Structural features that determine the enzymatic potency and specificity of human angiogenin: threonine-80 and residues 58-70 and 116-123. Biochemistry 37: 6847-6856
18. Harper JW, Fox EA, Shapiro R, Vallee BL (1990) Mutagenesis of residues flanking Lys-40 enhances the enzymatic activity and reduces the angiogenic potency of angiogenin. Biochemistry 29: 7297-302
19. Holloway DE, Chavali GB, Hares MC, Baker MD, Subbarao GV, Shapiro R, Acharya KR (2004) Crystallographic studies on structural features that determine the enzymatic specificity and potency of human angiogenin: Thr44, Thr80, and residues 38-41. Biochemistry 43: 1230-1241
20. Sato K, Egami F (1957) The specificity of T1 ribonuclease. C R Seances Soc Biol Fil 151: 1792-1796
21. Egami F, Oshima T, Uchida T (1980) Specific interaction of base-specific nucleases with nucleosides and nucleotides. Mol Biol Biochem Biophys 32: 250-77
22. Koepke J, Maslowska M, Heinemann U, Saenger W (1989) Three-dimensional structure of ribonuclease T1 complexed with guanylyl-2',5'-guanosine at 1.8 A resolution. J Mol Biol 206: 475-4788
23. Pfeiffer S, Karimi-Nejad Y, Ruterjans H (1997) Limits of NMR structure determination using variable target function calculations: ribonuclease T1, a case study. J Mol Biol 266: 400-423
24. Arni R, Heinemann U, Tokuoka R, Saenger W (1988) Three-dimensional structure of the ribonuclease T1 2'-GMP complex at 1.9-A resolution. J Biol Chem 263: 15358-15368
25. Zegers I, Loris R, Dehollander G, Fattah Haikal A, Poortmans F, Steyaert J, Wyns L (1998) Hydrolysis of a slow cyclic thiophosphate substrate of RNase T1 analyzed by time-resolved crystallography. Nat Struct Biol 5: 280-283
26. Yoshida H (2001) The ribonuclease T1 family. Methods Enzymol 341: 28-41
27. Noguchi S (2010) Isomerization mechanism of aspartate to isoaspartate implied by structures of Ustilago sphaerogena ribonuclease U2 complexed with adenosine 3'-monophosphate. Acta Crystallogr D Biol Crystallogr 66: 843-849
28. Olson BH, Goerner GL (1965) Alpha Sarcin, a new antitumor agent. I. Isolation, purification, chemical composition, and the identity of a new amino acid. Appl Microbiol 13: 314-321
29. Perez-Canadillas JM, Santoro J, Campos-Olivas R, Lacadena J, Martinez del Pozo A, Gavilanes JG, Rico M, Bruix M (2000) The highly refined solution structure of the cytotoxic ribonuclease alpha-sarcin reveals the structural requirements for substrate recognition and ribonucleolytic activity. J Mol Biol 299: 1061-1073
30. Lacadena J, Martinez del Pozo A, Lacadena V, Martinez-Ruiz A, Mancheno JM, Onaderra M, Gavilanes JG (1998) The cytotoxin alpha-sarcin behaves as a cyclizing ribonuclease. FEBS Lett 424: 46-48

31. Endo Y, Gluck A, Chan YL, Tsurugi K, Wool IG (1990) RNA-protein interaction. An analysis with RNA oligonucleotides of the recognition by alpha-sarcin of a ribosomal domain critical for function. J Biol Chem 265: 2216-2222
32. Gluck A, Endo Y, Wool IG (1994) The ribosomal RNA identity elements for ricin and for alpha-sarcin: mutations in the putative CG pair that closes a GAGA tetraloop. Nucleic Acids Res 22: 321-324
33. Piesiniene L, Truncaite L, Zajanckauskaite A, Nivinskas R (2004) The sequences and activities of RegB endoribonucleases of T4-related bacteriophages. Nucleic Acids Res 32: 5582-5595
34. Uzan M, Favre R, Brody E (1988) A nuclease that cuts specifically in the ribosome binding site of some T4 mRNAs. Proc Natl Acad Sci USA 85: 8895-8899
35. Sanson B, Hu RM, Troitskayadagger E, Mathy N, Uzan M (2000) Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. J Mol Biol 297: 1063-1074
36. Lebars I, Hu RM, Lallemand JY, Uzan M, Bontems F (2001) Role of the substrate conformation and of the S1 protein in the cleavage efficiency of the T4 endoribonuclease RegB. J Biol Chem 276: 13264-13272
37. Odaert B, Saida F, Aliprandi P, Durand S, Crechet JB, Guerois R, Laalami S, Uzan M, Bontems F (2007) Structural and functional studies of RegB, a new member of a family of sequence-specific ribonucleases involved in mRNA inactivation on the ribosome. J Biol Chem 282: 2019-2028
38. Aizenman E, Engelberg-Kulka H, Glaser G (1996) An Escherichia coli chromosomal "addiction module" regulated by guanosine 3',5'-bispyrophosphate: a model for programmed bacterial cell death, Proc Natl Acad Sci USA 93: 6059-6063
39. Amitai S, Kolodkin-Gal I, Hananya-Meltabashi M, Sacher A, Engelberg-Kulka H (2009) *Escherichia coli* MazF leads to the simultaneous selective synthesis of both "death proteins" and "survival proteins". PLoS Genet 5: e1000390
40. Vesper O, Amitai S, Belitsky M, Byrgazov K, Kaberdina AC, Engelberg-Kulka H, Moll I (2011) Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. Cell 147: 147-157
41. Schifano JM, Cruz JW, Vvedenskaya IO, Edifor R, Ouyang M, Husson RN, Nickels BE, Woychik NA (2016) tRNA is a new target for cleavage by a MazF toxin. Nucleic Acids Res 44: 1256-1270
42. Munoz-Gomez AJ, Santos-Sierra S, Berzal-Herranz A, Lemonnier M, Diaz-Orejas R (2004) Insights into the specificity of RNA cleavage by the *Escherichia coli* MazF toxin. FEBS Lett 567: 316-320
43. Zhang Y, Zhang J, Hoeflich KP, Ikura M, Qing G, Inouye M (2003) MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. Mol Cell 12: 913-23.
44. Park JH, Yamaguchi Y, Inouye M (2011) Bacillus subtilis MazF-bs (EndoA) is a UACAU-specific mRNA interferase. FEBS Lett 585: 2526-2532
45. Zhang Y, Zhu L, Zhang J, Inouye M (2005) Characterization of ChpBK, an mRNA interferase from *Escherichia coli*. J Biol Chem 280: 26080-26088
46. Schifano JM, Vvedenskaya IO, Knoblauch JG, Ouyang M, Nickels BE, Woychik NA (2014) An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. Nat Commun 5: 3538
47. Zhu L, Phadtare S, Nariya H, Ouyang M, Husson RN, Inouye M (2008) The mRNA interferases, MazF-mt3 and MazF-mt7 from Mycobacterium tuberculosis target unique pentad sequences in single-stranded RNA. Mol Microbiol 69: 559-569
48. Christensen SK, Pedersen K, Hansen FG, Gerdes K (2003) Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. J Mol Biol 332: 809-819
49. Zhang Y, Zhang J, Hara H, Kato IInouye M (2005) Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. J Biol Chem 280: 3143-3150
50. Simanshu DK, Yamaguchi Y, Park JH, Inouye M, Patel DJ (2013) Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in Bacillus subtilis. Mol Cell 52: 447-458
51. Zorzini V, Mernik A, Lah J, Sterckx YG, De Jonge N, Garcia-Pino A, De Greve H, Versees W, Loris R (2016) Substrate recognition and activity regulation of the *Escherichia coli* mRNA endonuclease MazF. J Biol Chem 291: 10950-10960
52. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816-821
53. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA 109: E2579-2586
54. Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 497: 254-257
55. Price AA, Sampson TR, Ratner HK, Grakoui A, Weiss DS (2015) Cas9-mediated targeting of viral RNA in eukaryotic cells. Proc Natl Acad Sci USA 112: 6164-6169
56. O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature 516: 263-266
57. Nicholson AW (2003) The ribonuclease III superfamily: forms and functions in RNA maturation, decay, and gene silencing. NY: Cold Spring Harbor Press
58. Ji X (2008) The mechanism of RNase III action: how dicer dices. Curr Top Microbiol Immunol 320: 99-116
59. MacRae IJ, Doudna JA (2007) Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr Opin Struct Biol 17: 138-145
60. Olmedo G, Guzman P (2008) Mini-III, a fourth class of RNase III catalyses maturation of the Bacillus subtilis 23S ribosomal RNA. Mol Microbiol 68: 1073-1076
61. Pertzev AV, Nicholson AW (2006) Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of Escherichia coli ribonuclease III. Nucleic Acids Res 34: 3708-3721
62. Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X (2005) Intermediate states of ribonuclease III in complex with double-stranded RNA. Structure 13: 1435-1442
63. Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. Cell 124: 355-366
64. Zhang K, Nicholson AW (1997) Regulation of ribonuclease III processing by double-helical sequence antideterminants. Proc Natl Acad Sci USA 94: 13437-13441
65. Kwon SC, Nguyen TA, Choi YG, Jo MH, Hohng S, Kim VN, Woo JS (2016) Structure of Human DROSHA. Cell 164: 81-90
66. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. Science 311: 195-198
67. Kozlowski PS-RJ, Legacz M, Magnus M, Krzyzosiak WJ (2008) Structures of microRNA precursors. Springer Netherlands
68. Starega-Roslan J, Galka-Marciniak P, Krzyzosiak WJ (2015) Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer. Nucleic Acids Res 43: 10939-10951
69. Redko Y, Bechhofer DH, Condon C (2008) Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. Mol Microbiol 68: 1096-1106
70. Redko Y, Condon C (2009) Ribosomal protein L3 bound to 23S precursor rRNA stimulates its maturation by Mini-III ribonuclease. Mol Microbiol 71: 1145-1154
71. Hotto AM, Castandet B, Gilet L, Higdon A, Condon C, Stern DB (2015) Arabidopsis chloroplast mini-ribonuclease III participates in rRNA maturation and intron recycling. Plant Cell 27: 724-740

72. Glow D, Pianka D, Sulej AA, Kozlowski LP, Czarnecka J, Chojnowski G, Skowronek KJ, Bujnicki JM (2015) Sequence-specific cleavage of dsRNA by Mini-III RNase. Nucleic Acids Res 43: 2864-2873
73. Zuckermann RN, Schultz PG (1988) A hybrid sequence-selective ribonuclease S. J Am Chem Soc 110: 6592–6594
74. Stein H, Hausen P (1969) Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. Science 166: 393-395
75. Donis-Keller H (1979) Site specific enzymatic cleavage of RNA. Nucleic Acids Res 7: 179-192
76. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell 121: 1005-1016
77. Kanaya S, Nakai C, Konishi A, Inoue H, Ohtsuka EIkehara M (1992) A hybrid ribonuclease H. A novel RNA cleaving enzyme with sequence-specific recognition. J Biol Chem 267: 8492-8498
78. Fukuma T, Walton CM, Wu CH, Wu GY (2003) Conjugation of an antisense oligodeoxynucleotide to ribonuclease h results in sequence-specific cleavage and intracellular inhibition of HCV gene expression. Bioconjug Chem 14: 295-301
79. Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science 263: 671-673
80. Sulej AA, Tuszynska I, Skowronek KJ, Nowotny M, Bujnicki JM (2012) Sequence-specific cleavage of the RNA strand in DNA-RNA hybrids by the fusion of ribonuclease H with a zinc finger. Nucleic Acids Res 40: 11563-11570
81. Takeshita D, Zenno S, Lee WC, Saigo K, Tanokura M (2007) Crystal structure of the PIN domain of human telomerase-associated protein EST1A. Proteins 68: 980-989
82. Wang X, McLachlan J, Zamore PD, Hall TM (2002) Modular recognition of RNA by a human pumilio-homology domain. Cell 110: 501-512
83. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z (2012) Engineering RNA endonucleases with customized sequence specificities. Nat Commun 3: 1147
84. Zhang W, Wang Y, Dong S, Choudhury R, Jin Y, Wang Z (2014) Treatment of type 1 myotonic dystrophy by engineering site-specific RNA endonucleases that target (CUG)(n) repeats. Mol Ther 22: 312-320
85. Chen DT, Lin A (2002) Domain swapping in ribonuclease T1 allows the acquisition of double-stranded activity. Protein Eng 15: 997-1003
86. Dow-Tien C, Yuan-Jhih T, Alan L (2008) Creating a ribonuclease T-tat that preferentially recognizes and hydrolyzes HIV-1 TAR RNA *in vitro* and *in vivo*. Nucleic Acids Res 36: 963-969

# Specyficzności sekwencyjne endorybonukleaz

**Dawid Głów[1], Martyna Nowacka[1], Krzysztof J. Skowronek[1,✉], Janusz M. Bujnicki[1,2,✉]**

[1]Laboratorium Bioinformatyki i Inżynierii Białka, Międzynarodowy Instytut Biologii Molekularnej i Komórkowej w Warszawie, ul. Ks. Trojdena 4, 02-109 Warszawa, Polska
[2]Zakład Bioinformatyki, Instytut Biologii Molekularnej i Biotechnologii, Wydział Biologii, Uniwersytet im. Adama Mickiewicza, ul. Umultowska 89, 61-614 Poznań, Polska

✉e-mail: kskowronek@iimcb.gov.pl; iamb@genesilico.pl

**Słowa kluczowe:** inżynieria białek, rybonukleazy, cięcie RNA, kompleksy białek z RNA

**STRESZCZENIE**

**Rybonukleazy są to enzymy nukleolityczne rozkładające cząsteczki RNA powszechnie występujące w organizmach żywych. Enzymy te są zaangażowane w podstawowe procesy komórkowe, między innymi w dojrzewanie RNA towarzyszące powstaniu funkcjonalnych cząsteczek, jak i degradację RNA umożliwiającą usuwanie cząsteczek wadliwych, obcych lub takich, które spełniły już swoje funkcje komórkowe. Degradacja RNA jest również jednym z głównych procesów determinujących poziom akumulacji transkryptów w komórce, a co za tym idzie ważnym elementem systemu regulacji ekspresji genów. Rybonukleazy mogą katalizować reakcje z udziałem cząsteczek RNA zawierających swoiste sekwencje, struktury lub sekwencje w obrębie konkretnej struktury, mogą też przecinać cząsteczki RNA niespecyficznie. W niniejszym artykule przedstawiamy wybrane endorybonukleazy przecinające wiązanie fosfodiestrowe wewnątrz cząsteczek RNA w specyficznie rozpoznawanych sekwencjach lub ich pobliżu. Przedstawiamy również przykłady konstruowania rybonukleaz z myślą o zapotrzebowaniu na specyficzne narzędzia molekularne, które mogłyby zostać wykorzystane między innymi w pracy badawczej nad cząsteczkami RNA.**