ISSN 0006-2979, Biochemistry (Moscow), 2019, Vol. 84, No. 8, pp. 870-883. © Pleiades Publishing, Ltd., 2019.
Russian Text © The Author(s), 2019, published in Biokhimiya, 2019, Vol. 84, No. 8, pp. 1099-1114.

REVIEW

The Role of Reverse Transcriptase in the Origin of Life

R. N. Mustafin[1,a]* and E. K. Khusnutdinova[2]

[1]*Bashkir State Medical University, 450008 Ufa, Russia*
[2]*Institute of Biochemistry and Genetics, Ufa Federal Research Center of the Russian Academy of Sciences, 450054 Ufa, Russia*
[a]*e-mail: ruji79@mail.ru*

Received March 8, 2019
Revised May 14, 2019
Accepted May 14, 2019

Abstract—It has been suggested that RNA polymerase ribozyme displaying reverse transcriptase and integrase activities has played a vital role in the origin of life on Earth. Here, we present a hypothesis that formation of universal ancestral units of all living organisms – retroelements – in the evolution was mediated by reverse transcriptase. The propensity of retroelements to mutations and their insertion capacity have formed a basis for the origin of complex DNA structures – primary genomes – that have given rise to archaea, eukaryotes, bacteria, and viruses. Conserved properties of retroelements have been preserved throughout the evolution; their modifications have facilitated the emergence of mechanisms for the interactions between proteins and nucleic acids. Life has evolved due to insertional mutagenesis and competition of autonomously replicating polynucleotides that allowed to preserve structures with adaptive properties. We hypothesize that natural selection of mechanisms for the defense against insertions based on the ribonuclease activity of reverse transcriptase ribozyme has led to the emergence of all universal enzymatic systems for the processing of RNA molecules. These systems have been and still remain the key sources of structural and functional transformations of genomes in the course of evolution. The data presented in this review suggest that the process of translation, which unifies the nucleic acid and protein worlds, has developed as a modification of the defense mechanisms against insertions. Polypeptides formed by this defense system have potentiated the activity of ribozymes in the composition of ribonucleoproteins (RNPs) and even functionally replaced them as more efficient catalysts of biological reactions. Here, we analyze the mechanisms of retroelement involvement in the structural and regulatory transformations of eukaryotic genomes supposedly reflecting the adaptive principles that had originated during the beginning of life on Earth. Simultaneously with the evolution of existing proteins, retroelements have served as sources of new ribozymes, such as long non-coding RNAs. These ribozymes can function in complexes with proteins in the composition of RNPs, as well as display independent catalytic and translational activities; their genes have a potential for the transformation into protein-coding genes. Hence, the conserved principles of RNA, DNA, and proteins interregulation formed at the time of life origin on Earth have been used throughout the evolution.

DOI: 10.1134/S0006297919080030

Keywords: reverse transcriptase, polymerase, processing, ribozymes, retroelements, transposable elements, evolution

The existence of the RNA world at the time of life origin on Earth is now a prevailing concept in modern science. Numerous data have been accumulated on the existence of ribozymes with inherent catalytic activities. Although nowadays all preserved life forms on Earth depend on proteinaceous enzymes, replication of genetic information and its transformation into functional sequences in the RNA world had been mediated only by polyribonucleotide molecules. Such mechanisms can be reconstructed under laboratory conditions in a complete absence of proteins. Moreover, it was experimentally confirmed that ribozymes are capable of self-replication with the exponential growth [1]. In the RNA world, organisms required an RNA polymerase ribozyme for the RNA-based inheritance and expression of "RNA genes". Although this ancestral enzyme has been lost during the course of evolution, the key functional aspects of RNA-catalyzed RNA replication can be investigated *by proxy* using ribozymes generated by *in vitro* selection, such as RNA polymerase R18 [2].

Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeat; LINE, long interspersed nuclear element; LTR, long terminal repeat; ncRNA, non-coding RNA; non-LTR-RE, retroelement not containing long terminal repeats; RE, retroelement; RISC, RNA-induced silencing complex; RNAi, RNA interference; RNP, ribonucleoprotein; RT, reverse transcriptase; SINE, short interspersed nuclear element; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; TE, transposable element.

* To whom correspondence should be addressed.

In 2016, Horning and Joyce [3] reported successful *in vitro* application of the RNA polymerase ribozyme that was capable of replicating short RNA sequences and synthesis of various functional molecules, including aptamers and ribozymes, in the absence of proteins, which proved complete self-reliance of RNA replication and expression. Modulation of the ribozyme activity *in vitro* allowed to reproduce the self-sustained RNA evolution. It is possible that under non-equilibrium boundary conditions allowing to overcome the effects of thermodynamic equilibrium (e.g., dilution and degradation of oligonucleotides) [4], the observed processes were similar to those that had occurred during the origin of life on Earth.

It was suggested that the flux of heat energy through the open pores in submerged rock has facilitated oligonucleotide replication with selection for longer strands. In experiments utilizing the interplay of molecular thermophoresis and laminar convection, strands of 75 nucleotides survived, while the shorter ones ceased to exist. The interplay between the two mechanisms is based on the balancing of the entropy of mixing. Thermophoresis directs molecules horizontally from the warm side to the colder one. At the same time, the liquid moves vertically via laminar convection and takes the molecules with it. Convection diverts horizontal thermophoretic exhaustion and enhances vertical accumulation of the molecules. This interaction between molecular motion and liquid flow results in the effective net transport of oligonucleotides to the lower part of the compartment. As a result, accumulation balances diffusional dilution and solves the problem of molecule concentration related to the process of life origin [4].

Protein-free RNA had been functional already prior to the formation of cellular life forms – the last universal common ancestor of cells existed ~3.9 billion years ago [5]. Despite this fact, ribozymes remain successful and somewhat prevalent functional elements that include conserved RNAs [tRNAs, rRNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), spliceosome ribozymes] and variable polyribonucleotides [group I and II introns, non-coding RNAs (ncRNAs)]. The first ribozyme – group I intron – was discovered in the ciliate *Tetrahymena* already in 1982 [6]. The ribonuclease P ribozyme involved in the processing of tRNA molecules was found in 1983 [7]. The hammerhead-type ribozymes from plant virus satellite RNAs were described in 1986 [8]. Their numerous variants have been found in all existing Kingdoms of life [9]. Over the recent years, the role of long ncRNA ribozymes in the formation of new protein-encoding genes [10-13], genome regulation in ontogenesis, and interaction with the mobile genetic elements (transposable elements, TEs) [14, 15] have been reported. These data could reflect the ancient principles of evolution that had formed during the life origin, when the RNA–DNA world had transformed into the world involving proteins.

The self-replication of DNA biopolymers, which are more stable than RNA and can preserve adaptive sequences vital for their survival, has become an essential requirement for the exponential growth of the amount and diversity of polynucleotides on Earth. The key event in this process was emergence of reverse transcriptase (RT), which facilitates transfer of the genetic information encoded in the RNA polymer to DNA. Conservation of RT in the genomes of all organisms indicates its universal importance throughout the evolution. During the origin of life, RT could be a ribozyme that had provided a new genetic material to DNA, even from a single source, due to the accumulation of mutations in RNA molecules. The frequency of mutations in RNA (10^{-3}-10^{-4}) is much higher in comparison with DNA (10^{-5}-10^{-8}) [16]. Indeed, Samanta, and Joyce were able to obtain the RNA polymerase ribozyme that exhibited RT properties as a secondary activity [17]. Furthermore, correlation between the structures of protein domains in RT and ribozymes was revealed [18], which indicated their continuity in evolution and alternative use in identical biological processes, as well as the ancient origin of the RT protein. In this connection, it can be suggested that the properties of protein analogues of ribozymes could reflect the characteristics of their ancestral RNA precursors.

The RT protein possesses the activity of ribonuclease H [19]. Other analogues of this enzyme are Argonaute, Dicer, Cas9, transposases, integrases, and splicing systems [20]. Hence, it is likely that ribozymes with the properties of these enzymes have originated from RT through its modifications via mutations. This means that the RT ribozyme had the potential not only to perform reverse transcription, but also to catalyze integration of the formed polynucleotides molecule and processing of RNA molecules, which has facilitated the evolutionary success associated with the lengthening of the complex DNA structures.

High RNA mutability was the cause of variability of early genomes but, at the same time, presented the threat for maintaining adaptive properties of polynucleotides required for their stable self-replication. Only those complex DNA molecules and their transcripts survived that have developed the defense mechanisms against insertions into the most important adaptive DNA regions. These defense mechanisms have become the basis for the emergence of clustered regularly interspaced short palindromic repeats (CRISPRs), RNA interference (RNAi), spliceosomal splicing, and translation. The conditions for their formation have been provided by the RNase activity of RT ribozyme. It is most likely that the original function of rRNAs, tRNAs, and processing ribozymes involved in their maturation was the defense against insertions (Fig. 1), as suggested by the conservation of relations of rRNAs, tRNAs, snRNAs, snoRNAs with retroelements (REs) in the course of evolution [21-25].

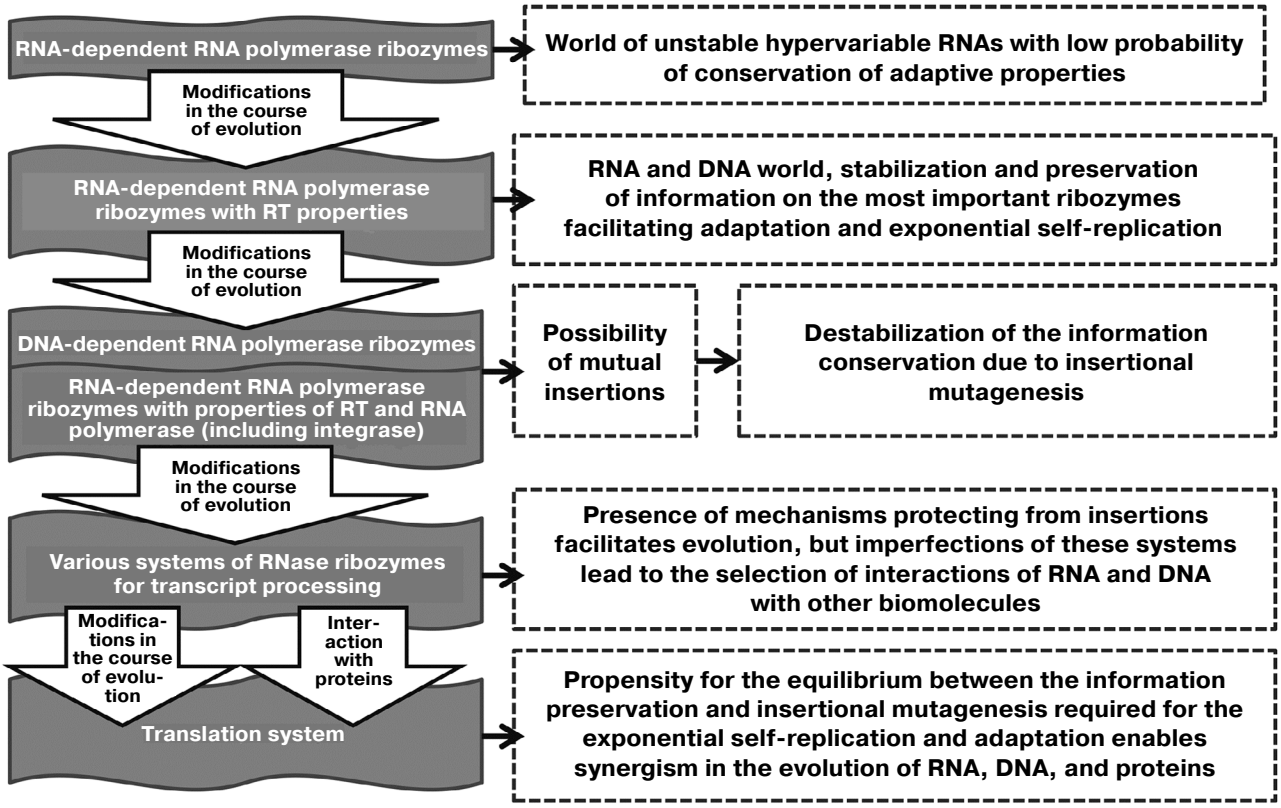

Fig. 1. Evolution of ribozymes during the origin of life on Earth.

It is commonly accepted that RT plays a key role in the evolution of all forms of life. REs are abundant in eukaryotic genomes and comprise from 3 to 85% of all nucleotide sequences in DNA [26]. According to the modern classification (http://www.girinst.org/repbase/), REs belong to the class I TEs [27] and can be autonomous or non-autonomous. The majority of the latter are short interspersed nuclear elements (SINEs) [28] that are transferred from one genome locus to another via the "copy and paste" mechanism with the help of long interspersed nuclear elements (LINEs). At the same time, SINEs have common nucleotide sequences and properties with REs lacking long terminal repeats (non-LTR-REs) [29]. SINEs are derivatives of tRNAs, snRNAs, and 5S rRNA [21, 24, 29-31], the processing products of which participate in RNAi for transposons [23, 25]. This implies the association of REs with the emergence of the translation system at the early stages of life evolution. The data on the use of tRNA molecules as primers for reverse transcription of LTR-REs could serve as a corroboration of the evolutionary relationship between the REs containing long terminal repeats (LTR-REs) and non-LTR-REs [28].

DNA transposons (DNA-TEs) translocating via the rolling circle (*Helitron*, *Maverick*) and cut-and-paste (*TIR*, *Crypton*) mechanisms belong to the class II TEs [32]. They encode the enzyme transposase that catalyzes DNA-TE excision and integration into a new genome region [28]. Phylogenetic analysis revealed that DNA-TE proteins are of the ancient origin and contain specific structural motifs (folds) that had already existed during the period when the first cellular organisms appeared [33]. The enzymes required for the amplification and extension of biopolymers have been of the utmost importance when the life was emerging. The cut-and-paste mechanism has become relevant at the later stages of evolution; hence, it can be suggested that DNA-TEs have originated from REs. An argument in favor of this hypothesis is the emergence in RT of ribonuclease H domains [19] (characteristic for structurally similar transposase and integrase [34, 35]) as the secondary activity. Ribonuclease H belongs to the superfamily of nucleotidyltransferases that also includes transposase, retroviral integrase, resolvase, and nuclease of the RNA-induced silencing complex (RISC) [36]. Enzymes similar to ribonuclease H participate in the universal conserved defense systems against viruses and TEs, including RISC (Argonaute, Dicer) and CRISPR (Cas9) [20]. Considering the possibility of RNA succession by protein domains in the course of evolution [18], one can assume that RT ribozymes could simultaneously possess the properties of replicases and ribonucleases; modification of these prop-

erties has resulted in the formation of new processing systems.

It has been suggested that emergence of RNA molecules capable of replication of primitive RNA genome was the pivotal event in the origin of life. Moreover, the earliest and simplest biological processes depended on RNA for both inheritance and metabolism. Current data on the central catalytic role of RNA in splicing, gene expression, and translation, as well as multiple RNA functions in the formation of specific receptors and catalysts [2], prove the existence of the RNA world at the early stages of evolution. It is possible that the TE-like elements have served as universal units during the origin of life. Their interaction and competition in the formation of primary genomes have served as a basis for all following evolutionary processes. However, due to the high level of mutability of transposon nucleotide sequences [16], it is difficult to evaluate the role of these elements in the formation of genomes of modern organisms. Analysis of human DNA with the help of specific oligonucleotides revealed that TEs comprise >67% of all nucleotide sequences, rather than 45%, as it was assumed before [37]. Modern analytical methods have allowed identification of "hidden" TEs in the genomes of various animal species in order to determine their role in the formation of regulatory elements and protein-coding genes in the course of evolution. As a result, significantly higher number of nucleotide sequences originating from TEs have been identified [38].

It is likely that all Earth biomass, starting from moment of life appearance, has been formed due to the existence of TEs as universal units. The study by Horning and Joyce [3], in which artificial *in vitro* RNA evolution has been reconstructed by decreasing the accuracy of ribozyme functioning in the protein-free medium (resulting in 10% of errors during elongation of new RNA chains) can serve as a corroboration of this statement. Random mutations occurring at each tenth nucleotide in all RNA molecules can give rise to 10^{14} different RNA variants, which proves the efficiency of evolution occurring with the participation of polymerases capable of making a large number of errors [3]. As a result, the molecular variability has been maintained during transition from the RNA world to the world including DNA, RNA, and proteins. This is in agreement with the fact that most RTs introduce errors due to the absence of proofreading (3′-5′ exonuclease) domain. RNA-dependent RNA polymerases that also lack the 3′-5′ exonuclease domain responsible for proofreading activity are considered evolutionary precursors of RTs [39]. It is plausible that preservation of the ability to introduce errors indicates conservation of the ancient properties of RTs that have played a universal role in the evolution of all living organisms. The similarity between the ribozyme domains and their protein analogues [18] suggests that RTs of the RNA world also possessed this property.

According to the alternative hypothesis suggested by Freeland et al. in 1999 [40], DNA was the last to appear in the RNA–DNA–protein triad. As an argument in favor of this statement, the authors noted that the biochemical reduction of ribonucleotides to deoxyribonucleotides is out of the scope of RNA catalytic abilities [40]. However, the authors did not consider the possibility of abiogenesis of DNA monomers in the absence of intermediate step (ribonucleotides). Moreover, it was proven that translation requires larger amount of hereditary information than it can be supported by the RNA genomes [17], which speaks in favor of our hypothesis that proteins appeared the last in the course of evolution. Most likely, the translation system has emerged as one of numerous defense mechanisms of early DNA genomes against insertions that used catalytic polypeptide domains in the complex with ribozymes. This suggestion is supported by the data on the role of products of tRNA, snoRNA, and rRNA processing [23, 25] in RNAi for transposons.

FEATURES OF REVERSE TRANSCRIPTASE

Proteins of the RT family have the same evolutionary origin, which has been proven based on the homology of their amino acid sequences and presence of RTs in all domains of life [39]. The RT enzyme is responsible for the conversion of RNA into cDNA and is required for the invasion of REs and their proliferation in the genomes. RT was first identified in 1970 by two research groups in the RNA-containing mouse leukemia virus and Rous sarcoma virions [41, 42]. In 1989, bacterial RTs were discovered in REs known as retrons in *Myxococcus xanthus* [43] and *Escherichia coli* B [44]. More than 50% of currently identified bacterial RTs are encoded by the group II introns, which are mobile REs with the ribozyme properties [45]. RTs have served not only as a key element in the RNA world transition into the RNA–DNA–protein world, but also preserved their universal ability to accelerate evolution at all the later stages. It is believed that the precursors of RTs were RNA-dependent RNA polymerases [39] that at the time of the origin of life on Earth were represented by ribozymes [2, 3]. Hence, ancient RTs could also be ribozymes capable of synthesizing DNA on the RNA template (as it has been proven in [17]).

The use of nucleoside triphosphates for elongation of either RNA or DNA molecules is a fundamental property of all living organisms. The reaction catalyzed by RNA ribozyme is similar to the activity of telomerase. Telomerase is a specialized DNA polymerase that contains a short RNA in its composition which serves as a template for the synthesis of telomeric DNA repeats. It was suggested that ribozymes with the activities similar to the activity of nucleotidyltransferase had played the role of telomerases before the emergence of protein-mediated

catalysis. Although tRNA nucleotidyltransferase is mostly responsible for the addition and maintenance of the CCA-3′ tail at the acceptor stem of tRNA, this cell enzyme can also function as a "telomerase" in the viral RNA genomes that accept the tRNA-like structures at their 3′-ends [46]. In this regard, the data on close association between tRNAs and REs in eukaryotes [21, 25, 28] may reflect the evolutionary relationship of these molecules. As universal mediators, tRNAs could have been originally used as a required link in RNA and DNA polymerization (likely, even before the emergence of the integrase ribozyme and DNA elongation via insertions). Further selection of certain tRNAs capable of interaction with amino acids mediated by various ribozymes has connected the polynucleotide world to the protein world via the triplet genetic code universal for all living organisms. Close relationships between tRNAs and REs containing complementary nucleotide sequence and using tRNAs as templates confirm this suggestion [19]. The natural ribozyme that catalyzes RNA polymerization via elongation of the RNA primer was identified in the living ciliate *Tetrahymena thermophila*, which corroborates the theory of self-replication of pre-biotic RNAs [47]. It was experimentally demonstrated that unstable non-functional complexes consisting of shorter 3′-truncated oligonucleotides could be stabilized and become functional through the non-enzymatic elongation of the primer [48].

It has been suggested that unlike RTs, other DNA polymerase families have developed the proofreading mechanisms which provide higher fidelity of DNA synthesis during genome replication. In order to determine if the lack of proofreading activity is a historical coincidence or functional limitation of reverse transcription, high-fidelity thermostable DNA polymerase for efficient use of RNA templates was designed, which proved that the combination of proofreading activity with reverse transcription is possible [39]. Apparently, this result indicates conservation of the ancient type of RT, because amplification mediated by this RT is characterized by high mutability, required for the variability and acquiring of new adaptive features. This universal property ensuring DNA amplification and its variability could have been a key element in the origin of life.

The similarity between the structures of ribozymes and RT protein domains [18] suggests the existence of evolutionary pathways for the transformation of catalytic RNA molecules into the protein ones, as well close relationship between these molecules and their interchangeability (Fig. 2). RNA-dependent RNases are considered as precursors of RTs [39], with RNA ribozymes as their ancient forms. Hence, it is only logical to assume that RT ribozymes should be available in living organisms. Indeed, the fact that the highly organized RNA polymerase ribozyme can function as RT has been confirmed. This RT activity has been of vital importance for the transition from the RNA to DNA genomes during the origin of life on Earth. It was suggested that this activity has emerged as a secondary function of RNA-dependent RNA polymerase ribozymes [17]. Although REs also encode other proteins enhancing the efficiency of reverse transcription, their RT possesses dual enzymatic activity: it can copy both RNA and DNA templates as a DNA polymerase and cleaves RNA in the composition of the RNA/DNA hybrids as ribonuclease H [19].

The presence of ribonuclease H activity in RT suggests the universal role of this enzyme in the evolution of key systems for transcript processing in live organisms. This is related to the fact that ribonuclease H belongs to the nucleotidyltransferase superfamily, which also includes transposase, retroviral integrase, resolvase, and

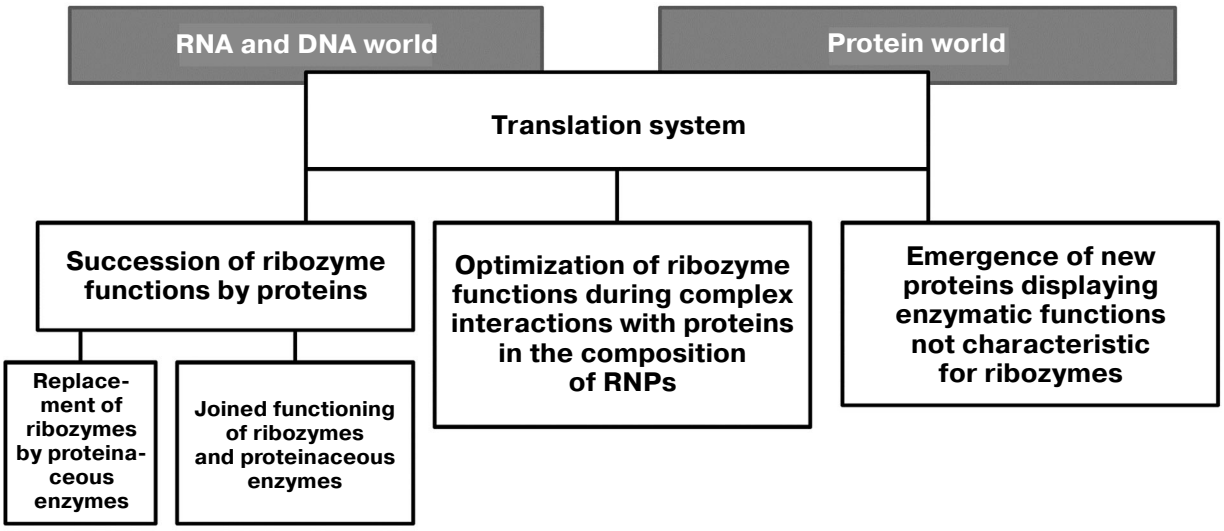

Fig. 2. Relationship between ribozymes and proteinaceous enzymes in the course of evolution.

RISC nuclease Argonaute [36]. There is an evolutionary succession in the 3D structures of domains of RNA ribozymes and proteinaceous enzymes with similar functions [18]. Hence, it is very likely that during the origin of life on Earth, some modifications of RT ribozymes resulted in the appearance of ribonuclease activity, thus becoming the basis for the evolution of processing systems.

Comparison of structures of retroviral integrases, transposases, and their binding sites in DNA revealed that both types of enzymes could have common biochemical and genetic properties, which indicates the possibility that they have originated from a common ancient sequence [34]. Both transposase and retroviral integrase have the same domain similar to ribonuclease H, with the DDE/D catalytic triad that coordinates divalent cations required for DNA cleavage and integration [35]. Chimeric ribozymes with the properties of different enzymes have been described, which confirms the possibility of a multifunctional nature of ancient RTs. In particular, the ribozyme from the *Clostridium difficile* bacterium displays features characteristic for group I introns and insertion elements [49]. The RE encodes integrase and RT as a single polypeptide that is processed by the protease encoded in the same RE [50]. These data suggest that some versions of RT ribozymes could have additional ribonuclease and integrase activities. The mechanistic similarity between the DNA transposase and integrase enzymes from REs emphasizes their close evolutionary relation. It is assumed that retroviruses have originated from the LTR-RE Ty3/Gypsy. The copies of their DNA can be found in the genomes of the majority of living organisms [51]. The evolutionary kinship and interrelationship between different enzymes responsible for the transposition indicated a possible scenario of life origin from REs as structural and functional units that have given rise to other TEs. Their competition and interactions, as well as ability for amplification via reverse transcription, could have been the basis for the evolution.

Mutual parasitism of TEs has resulted in the proliferation of non-autonomous REs that initially non-coding enzymes required for the transposition. These elements contain nucleotide sequences that play an important role in the regulatory processes. For example, *Alu* REs are associated with the activation of gene expression and demonstrate high regulatory potential depending on the chromatin state [52]. *Alu* elements belong to SINEs and have originated from the reversely transcribed 7SL RNA. The latter exists in the composition of RNPs as a ribozyme or as a component of the signal-recognizing particle (SRP) required for the ribosome transport through the transmembrane pores. Interestingly enough, the secondary structures of SRPs from animals, plants, fungi, bacteria, and archaea are very similar [22], which implies ancient evolutionary origin of 7SL RNA. This provides the basis for suggestion that SRPs have originated from TEs – the very first universal source of all genes in the evolution. *Alu* and other SINEs, as well as 7SL RNA, are processed by the Dicer enzyme with the formation of small ncRNAs [53].

ORIGIN OF TRANSLATION SYSTEM AND SUCCESSION OF THE RIBOZYME FUNCTIONS BY PROTEINS

Emergence of RT ribozymes from RNA polymerases and their selection have been essential steps in the maintenance and proliferation of systems for the autonomous RNA replication. This is due to the high stability of DNA and conservation of adaptive genetic material. Transcription of this genetic material back to RNA by DNA-dependent RNA polymerases resulted in the formation of polynucleotide molecules suitable for successful amplification. Similar to the emergence of integrase function by modification of RT ribozymes in the RNA world, multiple insertions have facilitated evolution and proliferation of systems of autonomously replicating polynucleotides. However, this process has posed danger for stable DNA genomes because of the accumulation of mutations, when newly formed adaptive properties were disrupted by new insertions. As a result, new mechanisms of genome competition in a form of RNA-processing systems have developed in the course of selection. The basis for the development of these defense mechanisms was modification of the RT ribonuclease domain structure. It is most likely that RNA polymerases and RT ribozymes have become a source of various processing systems providing the defense against insertions through their modification via mutations. This ability has its reflection in the protein world, e.g., the RT protein possesses the properties of ribonuclease H [19]. The members of its superfamily are major participants in the processing systems in all living organisms (RNAi, CRISPR system, splicing [20, 36]).

It is possible that the translation system has become one of the mechanisms of processing and defense in the RNA–DNA world, which can explain the succession of the ribozyme domains by proteins in the course of evolution and their relationship in the regulation of the same biological processes. The possibility of protein control of biological reactions without involvement of ribozymes also speaks in favor of this suggestion, because selection and conservation of polypeptides aiming to the improvement of defense mechanisms implies advantage of their use over the use of ribozymes. Initially, rRNAs, tRNAs, and ribozymes processing them could have originated from TEs and performed other functions required for the competition of early genomes, including as components of other defense systems. In particular, tRNAs are used as primers by REs and retroviruses [19], which may reflect an ancient evolutionary property of tRNA. Furthermore, rRNAs [23, 54], snRNAs [23], tRNAs [55], and

snoRNAs [56, 57] undergo non-random processing. The obtained products of processing are used for RNAi of TE [23, 25, 55, 58]. Ribozymes participate in various processes and perform numerous functions. For example, tRNAs are used not only as primers, but also in RNAi and transport of amino acids. Proteinaceous enzymes participate in various processes and perform numerous functions. For example, RT telomerases (TERTs) are used for the regulation of gene transcription in addition to telomere elongation [59]. Evolutionary relationships between the translation system ribozymes and transposons are manifested as site-specific features of TE integration. For example, the R-element, miniature inverted repeat transposable elements (MITEs), and *Pokey* insert into multigenic rDNA [60], while the R2 retroelement inserts into the gene regions of the 28S rDNA [61].

The same close relationship and evolutionary kinship have been revealed between all components of other RNA-processing systems and TEs. For example, spliceosomal introns have originated from group II introns. The latter are TEs with the ribozyme properties and RT activity [62]. At the same time, TEs in eukaryotes serve as a source of spliceosomal introns [63, 64], splicing signals [65, 66], splicing enhancers and silencers [67, 68], and structural and functional components of the spliceosome itself [69]. Moreover, similarly to TEs, spliceosomal introns are processed by specific ribonucleases with the production of ncRNAs [70].

Due to the emergence of translation system, living organisms have moved to a new level of structural and functional organization. As a result, proteinaceous enzymes were able to replace ribozymes in competing genomic systems, to function in the content of RNP complexes, and to form domains with new activities not characteristic for ribozymes. Despite these developments, the properties of ribozymes exclusive for these molecules have been conserved in the course of evolution, which proves their role in the origin of translation system. In particular, rRNA molecules in the content of ribosome directly control protein synthesis. It was shown that the ribosomal catalytic sites are composed solely from RNA. Unlike rRNAs, ribosomal proteins are located at a significantly greater distance from the active sites of the ribosomes [71]. It cannot be ruled out that ribosomal ribozymes have originated from the modified TEs in the course of formation of systems competing with other egoistic elements. Participation of ncRNAs formed by rRNA processing in RNAi confirms this statement [23]. Furthermore, it is most likely that TEs have also served as evolutionary ancestors of tRNAs. Both the use of processed tRNA transcripts in TE silencing (which implies complementarity of their nucleotide sequences) [25, 55] and the use of tRNAs [21] and rRNAs [21, 29-31] as a basis of non-autonomous TEs in genomes validate this notion.

Hence, the relations between the RNA–DNA world and proteins have emerged due to RTs and universal conserved RT-containing structures, such as REs and their derivatives (DNA-TEs). Successful proliferation and conservation during evolution of the reversely transcribed rRNAs and tRNAs as autonomous TEs in various eukaryotic genomes indicate their evolutionary kinship with other TEs and are ensured by their common properties with autonomous TEs, as well as the presence of homologous nucleotide sequence in their composition [23]. Interestingly enough, the 3′-end of SINE3, which has originated from the 5S rRNA, displays significant similarity with the CR1-like non-LTR RE. Similarly to the CR1-like REs, the copies of SINE3 are not flanked by the duplicates of target sites, and their 3′-ends consist of the $(ACATT)_n$ and $(ATT)_n$ microsatellites [29]. Transcriptionally active SINE28 that have originated from the 3′-end of the large ribosomal subunit (28S) was identified in mammalian genomes [31]. Some insect genomes contain HaSE3 chimeric SINEs consisting of tRNAs and 5S rRNAs [30]. SINEU elements originating from the U1 and U2 snRNAs were identified in the genomes of crocodiles, alligators, and gavials [24], which suggests evolutionary relationship between translation and splicing. This is in agreement with the fact that snRNAs also undergo non-random processing [23] for RNAi of TE sequences in the genome. Hence, components that associated the RNA–DNA world with proteins during the origin of life have formed due to RT and TEs. The main principles of these processes have been preserved in the course of evolution (Fig. 3).

ROLE OR REVERSE TRANSCRIPTASE IN THE EVOLUTION OF PROKARYOTES

Although RTs are considered as eukaryotic enzymes, they are also abundant in bacteria [72]. RTs mediate gene exchange between the plasmids and the nucleoid, thereby ensuring transmission of adaptive features between cells and their descendants [73]. REs (retrons) were described already in 1989 in the prokaryotic genomes of *E. coli* [44] and *M. xanthus* [43]. In 2002, a prokaryotic RE termed diversity-generating retroelement (DGR) was identified [74]. Phylogenetic analysis revealed that bacterial RTs could be classified into 17 groups: RTs from group II introns, retrons, and retron-like elements, DGR RTs, Abi-like RTs, RTs from the CRISPR-Cas system, G2L (group II-like RTs), and 11 other RT groups with unknown properties [75]. Bacterial genomes contain numerous non-characterized RTs and associated sequences with a wide variety of domain structures. With the exception of group II introns, RT-containing elements do not display any features of active retromobility but are involved in the defense of genomes against phages [72]. These genomic structures have retained their ancient evolutionary property aiming to ensure the competition of egoistic genomes, in the course of which new

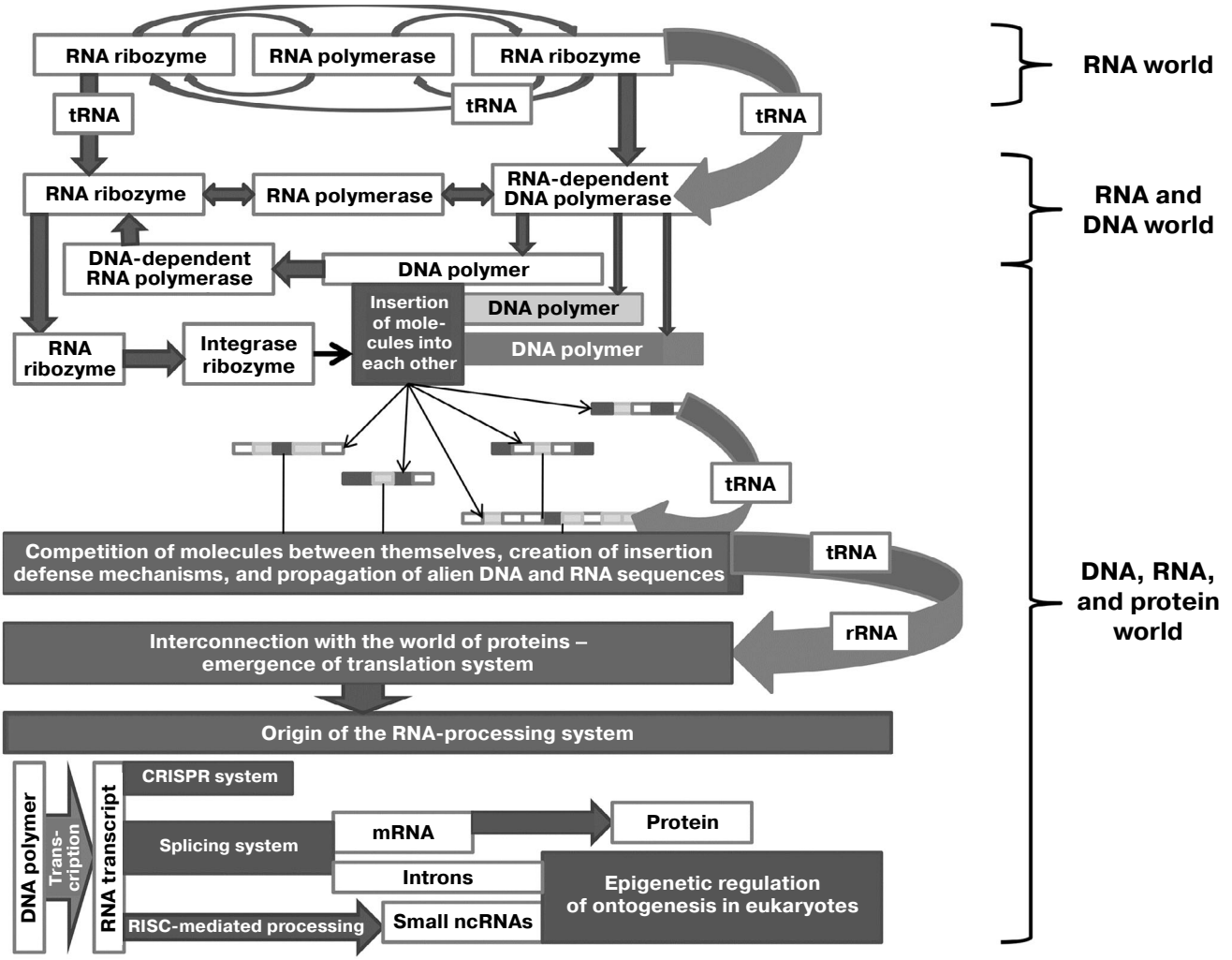

Fig. 3. Schematic representation of the evolution of processing systems as defense mechanisms.

defense systems have been formed. Group II introns encode more than 50% bacterial RTs that can generate complementary DNA on the RNA template. These introns act as ribozymes and mobile REs and are classified into A, B, C, D, E, F, G, CL1/2 (chloroplast-like), and ML (mitochondrion-like) groups. All of them have developed via association of ancient RTs with the structural catalytic RNAs [75].

Group II introns encode their own RTs that stabilize catalytically active RNA for the forward and reverse splicing and convert integrated intron RNA into DNA [76]. Group II introns were first identified in 1986 in the mitochondrial transcript from *Saccharomyces cerevisiae* [77]. It was suggested that these elements were evolutionary predecessors not only of spliceosomal introns, but also of spliceosomes themselves, REs, and eukaryotic telomeres [76]. The structure of the catalytic domain of these RTs is very similar to telomerase, and the active site for splicing – to the spliceosome Prp8 protein, which suggests evolutionary relationship between splicing and retromobility [18]. Moreover, it was shown that the Prp8 protein has originated from RT transposons in the course of evolution [69].

It was demonstrated that RTs of the type III CRISPR-Cas system are related to the RTs encoded by the group II introns. Association of RTs with the CRISPR-Cas system has occurred multiple times in the course of evolution. Phylogenetic analysis of RTs from the CRISPR system allowed to classify them into 12 groups and revealed co-evolution of RTs with the Cas1 proteins, as well as their horizontal transfer to the archaeal genomes [45]. RTs contain sequences with the properties of ribonuclease H, which belongs to the nucleotidyltransferase superfamily together with Cas9 [20]. Hence, it can be suggested that the CRISPR system has formed in the course of evolution from the defense mechanisms exhibited by one RE group against other TEs. Prokaryotes fight viruses with the help of innate

immune systems of receptor modification, restriction-enabled modification, and abortive infection. RNA-mediated CRISPR-Cas adaptive immune system has been identified in the majority of archaea and ~40% bacteria [45]. RT plays an important role in this system by catalyzing reverse transcription of foreign RNA [78].

Integration of new nucleotide sequences (spacers) in prokaryotes with the help of CRISPR system is accompanied by the removal of other parts of genome (repeats), which preserves the size of the nucleoid DNA. As a result, the maximum genome length of *M. xanthus* is only 16-times greater than the minimal length of the *Micoplasma genitalium* genome. The evolutionary success and an extremely wide variety of phenotypes and genome sizes in eukaryotes could be explained by the universal role of insertions in their evolution and functioning, which, in turn, could reflect the processes that had occurred during the life origin on Earth. The sizes of eukaryotic genomes can differ 70,000 times (from 2.3 Mbp in *S. cerevisiae* to 148,852 Mbp in *Paris japonica*) [79]. However, a significant homology between the CRISPR and RNAi systems with participation of small interfering RNAs was reported [20]. It can be suggested that the common ancestor of all domains of life had a defense mechanism modification that provided the basis for the development of CRISPR system in prokaryotes and RISC in eukaryotes. The specifics of their functioning affected the structural features of genomes from different Kingdoms of life. Despite the lengthy period of evolution from the moment of life emergence, modern organisms still possess the mechanisms reflecting succession of the ribozyme functions by proteins at the early stages of evolution, long ncRNAs being the most prominent example.

ROLE OF REVERSE TRANSCRIPTASE IN THE EVOLUTION OF EUKARYOTES

Eukaryotic REs are classified into 2 main groups according to the mechanism of their transposition and DNA organization. The first class consists of LTR-REs characterized by the presence of direct repeats of several hundred base pairs at their ends and *pol* and *gag* genes. The product of the *pol* gene contains several enzyme domains (RT, ribonuclease H, integrase, and proteinase). The *gag* gene encodes GAG (Group-specific AntiGen) proteins. REs belonging to the second class – non-LTR-REs – contain one open reading frame (ORF) encoding a protein with the RT and EN (endonuclease) domains or two ORFs (the first one is similar to the retroviral *gag* gene and the second one encodes RT and EN domains). It was suggested that LTR-REs have evolved from non-LTR REs via development of integrase, while retroviruses originated from LTR-REs via inclusion of envelope genes (env domains) from other viruses [45]. Genome sequencing has revealed that eukaryotes can contain more RT genes than genes encoding any other protein. Around 40% of nucleotide sequences in the mammalian genomes are REs, mainly LINEs and SINEs. The content of TEs in plant DNA can reach 91% (*Asparagus officinalis*), mainly due to LTR-REs [80].

Eukaryotic RTs are involved in addition of telomeres, replication of mitochondrial plasmids, transcription and proliferation of retroviral genomes [39], and gene duplication [80]. RTs comprise a basis of REs, which serve in eukaryotes as key sources for the formation of both regulatory [65] and protein-coding sequences via either domestication of REs themselves [81, 82] or exonization of their nucleotide sequences [66, 83, 84]. REs were found to be the source of the most conserved properties in eukaryotic genomes that distinguish them from prokaryotes. For example, centromeric satellites [85] and centromere-binding proteins CENP/CENH3 [86] interacting with the centromeric REs [87], telomeres [88] and telomerase [89], transcription factors [65] and their binding sites [90] have originated from REs. Many microRNAs [91] and their target nucleotide sequences in the protein-coding genes have originated from REs and other TEs [65, 66, 81-84]. One cannot rule out that ncRNA-processing proteins, as well as the main components of the RNAi system, have also originated from TEs in the course of evolution, similar to spliceosome. Based on the mechanisms of evolution of eukaryotic genomes, it can be suggested that REs and RTs could be the sources of the origin of life itself, which is associated with the use of REs as universal units conserved through all the stages of evolution. Functioning of long ncRNAs is typical for eukaryotes. They represent the best example reflecting the principles of TE participation in the emergence of self-regulating systems and succession of ribozymes by proteinaceous enzymes.

Long ncRNAs are ribozymes capable of functioning either on their own or in the content of RNPs. They participate in the transcription regulation by affecting histone modification and DNA-binding complexes, including transcription factors [92]. Furthermore, nucleotide sequences of many processed transcripts of long ncRNAs completely match the sequences of TEs [15], and >80% of active domains of all long ncRNAs have originated from transposons [14]. REs can function as genes of long ncRNAs [93, 94], i.e., processed RE transcripts display ribozyme properties, which may reflect their ancient conserved feature that could have been a key factor in the origin of life. Investigating the role of long ncRNAs in the formation of new protein-coding genes could reveal similar processes at the early stages of evolution, when transition from the RNA–DNA world into the protein world had occurred by the succession of the ribozyme functions by polypeptide enzymes. The origin of evolutionary new eukaryotic enzymes from the translation products of long ncRNA transcripts has been demonstrated in independent studies [10-13] and was possible due to the ability of

long ncRNAs to bind ribosomes and be translated into functional peptides [95].

Due to the RT function, active proliferation of REs in eukaryotic genomes has become the source of biochemically active, strictly regulated non-coding elements, such as transcription factor binding sites and ncRNA. In the course of evolution, this process facilitates rearrangement of genetic regulatory networks, thus providing foundation for genetic variability and emergence of new species [96]. In addition, REs are important sources of hybrid dysgenesis during formation of new species. The outbreaks of transposition events that facilitated genetic instability have been observed following hybridization. The association between hybrid transpositions and demethylation has been observed in both mammals and plants [97].

Similar to the long ncRNAs, pre-mRNA molecules of protein-coding genes can form various 3D configurations. Moreover, their secondary and tertiary structures play an important role in the processing and stability of pre-mRNAs [98], which suggests functional significance of the evolutionary precursors of these RNAs and their evolutionary origin from TEs. For example, the abundance of TEs sequences in long ncRNA domains could be associated with their ability to form conformations participating in the essential biological reactions [14]. Various algorithms using minimization of free energy and maximum expected accuracy, as well as comparative evolutionary methods, are being developed to predict the secondary structures of RNAs. However, these tools are not ideal. The secondary structure of pre-mRNA could either promote or inhibit the splicing of the molecule itself, depending on the features of nucleotide sequences in its introns. Following the splicing, the RNA secondary structure could also affect the molecule stability and regulation of RNAi. Splicing of pre-mRNA molecules is also influenced by their tertiary structure. The G-quadruplex, which can enhance or inhibit splicing via creating or obscuring the sites of RNA binding with the protein, is an example [98]. REs possess similar properties – they can control recognition of the splicing sites through the formation of specific 3D hairpin structures [99].

Processing of introns of protein-coding genes into functional ncRNAs [70] make them similar to TEs [91]. Conservation of RNAi involving TE transcripts in the course of evolution indicates a universal nature of this mechanism. Active use and conservation of genetic code (encoding of amino acids with nucleotides) could be associated with the possibility of peptide interactions with the primary and secondary DNA structures [100] in order to ensure the defense mechanisms. Both ribozymes or proteins are efficient in different types of biological reactions. If ribozymes provide unique advantages in a certain process, they preserve their functions. However, newly formed proteinaceous enzymes are capable of replacing these RNA molecules by ensuring better adaptation. This non-equilibrium state is a source of evolutionary transformations and generation of new protein-coding genes in the DNA regions containing genes of ncRNAs [10-13].

It is generally assumed that the majority of RT-like enzymes belong to REs or viruses and do not have any defined function in the host cell with the exception of telomerase. However, a unique class of RT-associated genes termed *rvt* was discovered. The *rvt* genes are components of the host cell genome (and not RE components); they exist as single copies and can contain introns at the evolutionary conserved positions. The *rvt* genes can change under selection pressure; they have been found in all major taxonomic groups including Protista, fungi, animals, plants, and bacteria, although their distribution is not phylogenetically uniform [101]. RTs are used in the generation of high-copy genes, such as *MADS-box* and cytochrome P450 genes [102]. Histone genes are also characterized by cluster organization and do not contain introns, which suggests their origin from TEs of the common ancestor of all eukaryotes. This hypothesis is supported by the fact that the *CENP/CENH* gene encoding a variant of the H3 histone, has originated from TE [86]. Small ncRNAs are capable of exerting their epigenetic effect via histone modification [103], while TEs themselves serve as the sources of ncRNAs [91]. Different types of TEs are associated with specific modifications of histones in the regions of their location (LTRs and LINEs – with H3K9me3 mark; other TEs – with H3K27me3 marks) [104]. The size distribution of the conserved and non-conserved introns has the maxima close to the lengths of dinucleosomal and nucleosomal DNAs [105]. The obtained data indicate that TEs possess universal properties and represent conserved units of all living organisms capable of forming self-regulatory complex structures that can efficiently evolve. Participation of histones in the regulation of genome function reflects the fact that during the life appearance, proteins have been used not only for the continuity of ribozyme functions but also for the formation of new optimal adaptive properties. Moreover, TEs have given rise to genes directly involved in the chromatin remodeling [106]. It was found that TEs have been the source for the acetyltransferase complex HDP1/2 [107], chromatin modification factors BEAF-32 and HIM-17 [108], centromeric proteins Abp1 [65], and insulators [109].

A considerable body of data has been accumulated on the universal nature of RT as a key link connecting the RNA and DNA worlds in the origin of life on Earth. Conservation of the RT involvement in the formation of new principles of self-regulation and self-reproduction is manifested in the evolution of bacteria, archaea, and eukaryotes, which could reflect similar mechanisms in the origin of ancestral pre-cellular forms of life. Modified versions of the RNA polymerase ribozyme could have

possessed the properties of RT, ribonuclease, and integrase to support self-reproduction of RNA and DNA biopolymers. However, preservation of information on the adaptive properties essential for the survival of DNA-based biomolecules required the balance between insertions (providing variability and lengthening of early genomes) and protection of important genetic material. In this connection, selection of mechanisms protecting from the insertions has occurred together with the formation of RNA-processing systems (RISC, CRISPR, splicing, and translation). The sources for these systems might have been modified RT ribozymes possessing various ribonuclease activities. Amino acids and peptides could have been used for the optimization of protective mechanisms in competing egoistic polynucleotide molecules. Later, the improvement of the translation system has become the basis for the interconnection of the RNA–DNA world with polypeptides. In these processes, succession of the functions of ribozymes by protein enzymes occurred. This principle is conserved in eukaryotes, because ribozymes of long ncRNA can be translated with the formation of functional molecules. In the course of selection, genes of long ncRNAs could have been transformed into the protein-coding ones, with the processed transcripts preserving their enzymatic activity. As a result, REs in eukaryotes have become the sources of transcription factors and their binding sites, centromeres and centromere-binding protein, telomeres and telomerase, protein-coding genes and ncRNAs interacting with them, introns, and components of spliceosomes. This represents a universal principle of self-regulation of REs and their derivatives required for the life origin and its evolution.

Conflict of interest. The authors declare no conflict of interest in financial or any other sphere.

Ethical approval. This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

1. Lincoln, T. A., and Joyce, G. F. (2009) Self-sustained replication of an RNA enzyme, *Science*, **323**, 1229-1232; doi: 10.1126/science.1167856.
2. Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme, *Science,* **332**, 209-212; doi: 10.1126/science.1200752.
3. Horning, D. P., and Joyce, G. F. (2016) Amplification of RNA by an RNA polymerase ribozyme, *Proc. Natl Acad. Sci. USA*, **113**, 9786-9791; doi: 10.1073/pnas.1610103113.
4. Kreysing, M., Keil, L., Lanzmich, S., and Braun, D. (2015) Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length, *Nat. Chem.*, **7**, 203-208; doi: 10.1038/nchem.2155.
5. Betts, H. C., Puttick, M. N., Clark, J. W., Williams T. A., Donoghue, P. C. J., and Pisani, D. (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin, *Nat. Ecol. Evol.*, **2**, 1556-1562; doi: 10.1038/s41559-018-0644-x.
6. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*, *Cell*, **31**, 147-157.
7. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell*, **35**, 849-857.
8. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R., and Bruening, G. (1986) Autolytic processing of dimeric plant virus satellite RNA, *Science*, **231**, 1577-1580.
9. De la Pena, M., Garcia-Robles, I., and Cervera, A. (2017) The hammerhead ribozyme: a long history for a short RNA, *Molecules*, **22**, E78; doi: 10.3390/molecules22010078.
10. Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A., and Begun, D. J. (2006) Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression, *Proc. Natl. Acad. Sci. USA*, **103**, 9935-9939.
11. Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008) *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*, *Genetics*, **179**, 487-496; doi: 10.1534/genetics.107.084491.
12. Xie, C., Zhang, Y. E., Chen, J. Y., Liu, C. J., Zhou, W. Z., Li, Y., Zhang, M., Zhang, R., Wei, L., and Li, C. Y. (2012) Hominoid-specific *de novo* protein-coding genes originating from long non-coding RNAs, *PLoS Genet.*, **8**, e1002942; doi: 10.1371/journal.pgen.1002942.
13. Ruiz-Orera, J., Messeguer, X., Subirana, J. A., and Alba, M. M. (2014) Long non-coding RNAs as a source of new peptide, *Elife*, **3**, e03523; doi: 10.7554/eLife.03523.
14. Johnson, R., and Guigo, R. (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs, *RNA*, **20**, 959-976; doi: 10.1261/rna.044560.114.
15. Kapusta, A., and Feschotte, C. (2014) Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications, *Trends Genet.*, **30**, 439-452; doi: 10.1016/j.tig.2014.08.004.
16. Lukash, L. L. (2007) Mutagenesis induced by integration processes and evolution of nuclear genome, *Biopolym. Cell*, **23**, 172-187.
17. Samanta, B., and Joyce, G. F. (2017) A reverse transcriptase ribozyme, *Elife*, **6**, e31153; doi: 10.7554/eLife.31153.
18. Qu, G., Kaushal, P. S., Wang, J., Shigematsu, H., Piazza, C. L., Agrawal, R. K., Belfort, M., and Wang, H. W. (2016) Structure of a group II intron in complex with its reverse transcriptase, *Nat. Struct. Mol. Biol.*, **23**, 549-557; doi: 10.1038/nsmb.3220.
19. Hughes, S. H. (2015) Reverse transcription of retroviruses and LTR retrotransposons, *Microbiol. Spectr.*, **3**, MDNA3-0027-2014; doi: 10.1128/microbiolspec.MDNA3-0027-2014.
20. Moelling, K., and Broecker, F. (2015) The reverse transcriptase-RNase H: from viruses to antiviral defense, *Ann. N. Y. Acad. Sci.*, **1341**, 126-135; doi: 10.1111/nyas.12668.

21. Gogolevsky, K. P., Vassetzky, N. S., and Kramerov, D. A. (2009) 5S rRNA-derived and tRNA-derived SINEs in fruit bats, *Genomics*, **93**, 494-500; doi: 10.1016/j.ygeno.2009.02.001.
22. Rosenbland, M. A., Larsen, N., Samuelsson, T., and Zwieb, C. (2009) Kinship in the SRP RNA family, *RNA Biol.*, **6**, 508-516.
23. Li, Z., Ender, C., Meister, G., Moore, P. S., Chang, Y., and 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.
24. Kojima, K. K. (2015) A new class of SINEs with snRNA gene-derived heads, *Genome Biol. Evol.*, **7**, 1702-1712; doi: 10.1093/gbe/evv100.
25. Martinez, G., Choudury, S. G., and Slotkin, R. K. (2017) tRNA-derived small RNAs target transposable element transcripts, *Nucleic Acids Res.*, **45**, 5142-5152; doi: 10.1093/nar/gkx103.
26. Startek, M. P., Nogly, J., Gromadka, A., Grzebelus, D., and Gambin, A. (2017) Inferring transposons activity chronology by TRANAcendence-TEs database and *de novo* mining tool, *BMC Bioinformatics*, **18**, 422; doi: 10.1186/s12859-017-1824-4.
27. Alzohairy, A. M., Gyulai, G., Jansen, R. K., and Bahieldin, A. (2013) Transposable elements domesticated and neofunctionalized by eukaryotic genomes, *Plasmid*, **69**, 1-15; doi: 10.1016/j.plasmid.2012.08.001.
28. Kramerov, D. A., and Vassetzky, N. S. (2011) Origin and evolution of SINEs in eukaryotic genomes, *Heredity (Edinb.)*, **107**, 487-495; doi: 10.1038/hdy.2011.43.
29. Kapitonov, V. V., and Jurka, J. (2003) A novel class of SINE elements derived from 5S rRNA, *Mol. Biol. Evol.*, **20**, 694-702.
30. Wang, J., Wang, A., Han, Z., Zhang, Z., Li, F., and Li, X. (2012) Characterization of three novel SINE families with unusual features in *Helicoverpa armigera*, *PLoS One*, **7**, e31355; doi: 10.1371/journal.pone.0031355.
31. Longo, M. S., Brown, J. D., Zhang, C., O'Neill, M. J., and O'Neill, R. J. (2015) Identification of a recently active mammalian SINE derived from ribosomal RNA, *Genome Biol. Evol.*, **7**, 775-788; doi: 10.1093/gbe/evv015.
32. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A. H. (2007) A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.*, **8**, 973-982.
33. Abrusan, G., Zhang, Y., and Szilagyi, A. (2013) Structure prediction and analysis of DNA transposon and LINE retrotransposons proteins, *J. Biol. Chem.*, **288**, 16127-16138; doi: 10.1074/jbc.M113.451500.
34. Rice, P. A., and Baker, T. A. (2001) Comparative architecture of transposase and integrase complexes, *Nat. Struct. Biol.*, **8**, 302-307.
35. Wolkowicz, U. M., Morris, E. R., Robson, M., Trubitsyna, M., and Richardson, J. M. (2014) Structural basis of Mos1 transposase inhibition by the anti-retroviral drug Raltegravir, *ACS Chem. Biol.*, **9**, 743-751; doi: 10.1021/cb400791u.
36. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis, *Cell*, **121**, 1005-1016.
37. De Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Polock, D. D. (2011) Repetitive elements may comprise over two-thirds of the human genome, *PloS Genet.*, **7**, e1002384; doi: 10.1371/journal.pgen.1002384.
38. Goerner-Potvin, P., and Bourque, G. (2018) Computational tools to unmask transposable elements, *Nat. Rev. Genet.*, **19**, 688-704; doi: 10.1038/s41576-018-0050-x.
39. Ellefson, J. W., Gollihar, J., Shoroff, R., Shivram, H., Lyer, V. R., and Ellington, A. D. (2016) Synthetic evolutionary origin of a proofreading reverse transcriptase, *Science*, **352**, 1590-1593; doi: 10.1126/science.aaf5409.
40. Freeland, S. J., Knight, R. D., and Landweber, L. F. (1999) Do proteins predate DNA, *Science*, **286**, 690-692.
41. Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour, *Nature*, **226**, 1209-1211.
42. Temin, H. M., and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma, *Nature*, **226**, 1211-1213.
43. Lampson, B. C., Inouye, M., and Inouye, S. (1989) Reverse transcriptase with concomitant ribonuclease H activity in the cell-free synthesis of branched RNA-linked msDNA of *Myxococcus xanthus*, *Cell*, **56**, 701-707.
44. Lim, D., and Maas, W. K. (1989) Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA–RNA compound in *E. coli B*, *Cell*, **56**, 891-904.
45. Toro, N., Martinez-Abarca, F., and Gonzalez-Delgado, A. (2017) The reverse transcriptases associated with CRISPR-Cas systems, *Sci. Rep.*, **7**, 7089; doi: 10.1038/s41598-017-07828-y.
46. Ekland, E. H., and Bartel, D. P. (1996) RNA-catalyzed RNA polymerization using nucleoside triphosphates, *Nature*, **382**, 373-376.
47. Been, M. D., and Cech, T. R. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the *Tetrahymena* ribozyme, *Science*, **239**, 1412-1416.
48. Adamala, K., Engelhart, A. E., and Szostak, J. W. (2015) Generation of functional RNAs from inactive oligonucleotide complexes by non-enzymatic primer extension, *J. Am. Chem. Soc.*, **137**, 483-489.
49. Braun, V., Mehlig, M., Moos, M., Rupnik, M., Kalt, B., Mahony, D. E., and von Eichel-Streiber, C. (2000) A chimeric ribozyme in clostridium difficile combines features of group I introns and insertion elements, *Mol. Microbiol.*, **36**, 1447-1459.
50. Gao, X., and Voytas, D. F. (2005) A eukaryotic gene family related to retroelements integrases, *Trends Genet.*, **21**, 133-137.
51. Skala, A. M. (2014) Retroviral DNA transposition: themes and variations, *Microbiol. Spectr.*, **2**; doi: 10.1128/microbiolspec.MDNA3-0005-2014.
52. Zeng, L., Pederson, S. M., Cao, D., Qu, Z., Hu, Z., Adelson, D. L., and Wei, C. (2018) Genome-wide analysis of the association of transposable elements with gene regulation suggests that *alu* elements have the largest overall regulatory impact, *J. Comput. Biol.*, **25**, 551-562; doi: 10.1089/cmb.2017.0228.
53. Ren, Y. F., Li, G., Wu, J., Xue, Y. F., Song, Y. J., Lv, L., Zhang, X. J., and Tang, K. F. (2012) Dicer-dependent biogenesis of small RNAs derived from 7SL RNA, *PLoS One*, **7**, e40705; doi: 10.1371/journal.pone.0040705.
54. Jacob, M. D., Audas, T. E., Mullineux, S. T., and Lee, S. (2012) Where no RNA polymerase has gone before: novel

functional transcripts derived from the ribosomal intergenic spacer, *Nucleus*, **3**, 315-319.
55. Kumar, P., Anaya, J., Mudunuri, S. B., and Dutta, A. (2014) Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets, *BMC Biol.*, **12**, 78; doi: 10.1186/s12915-014-0078-0.
56. Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., Pferffer, S., Rajewsky, N., and Meister, G. (2008) A human snoRNA with microRNA-like functions, *Mol. Cell.*, **32**, 519-528; doi: 10.1016/j.molcel.2008.10.017.
57. Taft, R. J., Glazov, E. A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J. S. (2009) Small RNAs derived from snoRNAs, *RNA*, **15**, 1233-1240; doi: 10.1261/rna.1528909.
58. Venkatesh, T., Suresh, P. S., and Tsutsumi, R. (2016) tRFs: miRNAs in disguise, *Gene*, **579**, 133-138; doi: 10.1016/j.gene.2015.12.058.
59. Zhou, J., Ding, D., Wang, M., and Cong, Y. S. (2014) Telomerase reverse transcriptase in the regulation of gene expression, *BMB Rep.*, **47**, 8-14.
60. Elliott, T. A., Stage, D. E., Crease, T. J., and Eickbush, T. H. (2013) In and out of the rRNA genes: characterization of Pokey elements in the sequenced *Daphnia* genome, *Mob. DNA*, **4**, 20; doi: 10.1186/1759-9753-4-20.
61. Jamburuthugoda, V. K., and Eickbush, T. H. (2014) Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase, *Nucleic Acids Res.*, **42**, 8405-8415; doi: 10.1093/nar/gku514.
62. Novikova, O., and Belfort, M. (2017) Mobile group II introns as ancestral eukaryotic elements, *Trends Genet.*, **33**, 773-783; doi: 10.1016/j.tig.2017.07.009.
63. Wang, D., Su, Y., Wang, X., Lei, H., and Yu, J. (2012) Transposon-derived and satellite-derived repetitive sequences play distinct functional roles in mammalian intron size expansion, *Evol. Bioinform. Online*, **8**, 301-319; doi: 10.4137/EBO.S9758.
64. Yenerall, P., and Zhou, L. (2012) Identifying the mechanisms of intron gain: progress and trends, *Biol. Direct.*, **7**, 29; doi: 10.1186/1745-6150-7-29.
65. Feschotte, C. (2008) The contribution of transposable elements to the evolution of regulatory networks, *Nat. Rev. Genet.*, **9**, 397-405.
66. Tajnik, M., Vigilante, A., Braun, S., Hanel, H., Luscombe, N. M., Ule, J., Zarnack, K., and Koning, J. (2015) Intergenic Alu exonization facilitates the evolution of tissue-specific transcript ends, *Nucleic Acids Res.*, **43**, 10492-10505; doi: 10.1093/nar/gkv956.
67. Lei, H., and Vorechovsky, I. (2005) Identification of splicing silencers and enhancers in sense Alus: a role for pseudoacceptors in splice site repression, *Mol. Cell. Biol.*, **25**, 6912-6920.
68. Pastor, T., Talotti, G., Lewandowska, M. A., and Pagani, F. (2009) An Alu-derived intronic splicing enhancer facilitates intronic processing and modulates aberrant splicing in ATM, *Nucleic Acids Res.*, **37**, 7258-7267; doi: 10.1093/nar/gkp778.
69. Galej, W. P., Oubridge, C., Newman, A. J., and Nagai, K. (2013) Crystal structure of Prp8 reveals active site cavity of the spliceosome, *Nature*, **493**, 638-643; doi: 10.1038/nature11843.
70. Rearick, D., Prakash, A., McSweeny, A., Shepard, S. S., Fedorova, L., and Fedorov, A. (2011) Critical association of ncRNA with introns, *Nucleic Acids Res.*, **39**, 2357-2366; doi: 10.1093/nar/gkq1080.
71. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science*, **289**, 920-930.
72. Zimmerly, S., and Wu, L. (2015) An unexplored diversity of reverse transcriptases in bacteria, *Microbiol. Spectr.*, **3**, MDNA3-0058-2014; doi: 10.1128/microbiolspec.MDNA3-0058-2014.
73. Ravin, N. V., and Shestakov, S. V. (2013) Genome of prokaryotes, *Vavilov Zh. Genet. Selekts.*, **17**, 972-984.
74. Liu, M., Deora, R., Doulatov, S. R., Gingery, M., Eiserling, F. A., Preston, A., Maskell, D. J., Simons, R. W., Cotter, P. A., Parkhill, J., and Miller, J. F. (2002) Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage, *Science*, **295**, 2091-2094.
75. Toro, N., Martinez-Abarca, F., Gonzalez-Delgado, A., and Mestre, M. R. (2018) On the origin and evolutionary relationships of the reverse transcriptases associated with type III CRISPR-Cas systems, *Front. Mircrobiol.*, **9**, 1317; doi: 10.3389/fmicb.2018.01317.
76. Lambowitz, A. M., and Zimmerly, S. (2011) Group II introns: mobile ribozymes that invade DNA, *Cold Spring Harb. Perspect. Biol.*, **3**, a003616; doi: 10.1101/cshperspect.a003616.
77. Peebles, C. L., Perlman, P. C., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., and Cheng, H. L. (1986) A self-splicing RNA excises an intron lariat, *Cell*, **44**, 213-223.
78. Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A. M., and Fire, A. Z. (2016) Direct CRISPR spacer acquisition from RNA by natural reverse transcriptase–Cas1 fusion protein, *Science*, **351**, aad4234; doi: 10.1126/science.aad4234.
79. Elliott, T. A., and Greqory, T. R. (2015) Do larger genomes contain more diverse transposable elements? *BMC Evol. Biol.*, **15**, 69-81; doi: 10.1186/s12862-015-0339-8.
80. Kubiak, M. R., and Makalowska, I. (2017) Protein-coding genes' retrocopies and their functions, *Viruses*, **9**, E80; doi: 10.3390/v9040080.
81. Zdobnov, E. M., Campillos, M., Harrington, E. D., Torrents, D., and Bork, P. (2005) Protein coding potential of retroviruses and other transposable elements in vertebrate genomes, *Nucleic Acids Res.*, **33**, 946-954.
82. Campillos, M., Doerks, T., Shah, P. K., and Bork, P. (2006) Computational characterization of multiple Gag-like human protein, *Trends Genet.*, **22**, 585-589.
83. Sela, N., Kim, E., and Ast, G. (2010) The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates, *Genome Biol.*, **11**, R59; doi: 10.1186/gb-2010-11-6-r59.
84. Schmitz, J., and Brosius, J. (2011) Exonization of transposed elements: a challenge and opportunity for evolution, *Biochimie*, **93**, 1928-1934; doi: 10.1016/j.biochi.2011.07.014.
85. Cheng, Z. J., and Murata, M. (2003) A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives, *Genetics*, **164**, 665-672.
86. Kipling, D., and Warburton, P. E. (1997) Centromeres, CENP-B and Tigger too, *Trends Genet.*, **13**, 141-145.

87. Mestrovic, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Satovic, E., and Plohl, M. (2015) Structural and functional liaisons between transposable elements and satellite DNAs, *Chromosome Res.*, **23**, 583-596; doi: 10.1007/s10577-015-9483-7.
88. Arkhipova, I. R. (2018) Neutral theory, transposable elements, and eukaryotic genome evolution, *Mol. Biol. Evol.*, **35**, 1332-1337; doi: 10.1093/molbev/msy083.
89. Garavis, M., Gonzalez, C., and Villasante, A. (2013) On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution, *Genome Biol. Evol.*, **5**, 1142-1150; doi: 10.1093/gbe/evt079.
90. De Souza, F. S., Franchini, L. F., and Rubinstein, M. (2013) Exaptation of transposable elements into novel *cis*-regulatory elements: is the evidence always strong, *Mol. Biol. Evol.*, **30**, 1239-1251; doi: 10.1093/molbev/mst045.
91. Gim, J., Ha, H., Ahn, K., Kim, D. S., and Kim, H. S. (2014) Genome-wide identification and classification of microRNAs derived from repetitive elements, *Genomic Inform.*, **12**, 261-267; doi: 10.5808/GI.2014.12.4.261.
92. Long, Y., Wang, X., Youmans, D. T., and Cech, T. R. (2017) How do lncRNAs regulate transcription? *Sci. Adv.*, **3**, eaao2110; doi: 10.1126/sciadv.aao2110.
93. Lu, X., Sachs, F., Ramsay, L., Jacques, P. E., Goke, J., Bourque, G., and Ng, H. H. (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity, *Nat. Struct. Mol. Biol.*, **21**, 423-425; doi: 10.1038/nsmb.2799.
94. Honson, D. D., and Macfarlan, T. S. (2018) A lncRNA-like role for LINE1s in development, *Dev. Cell*, **46**, 132-134; doi: 10.1016/j.devcel.2018.06.022.
95. Anderson, D. M., Anderson, K. M., Cang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., and Olson, E. N. (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance, *Cell*, **160**, 595-606; doi: 10.1016/j.cell.2015.01.009.
96. Chuong, E. B., Elde, N. C., and Feschotte, C. (2017) Regulatory activities of transposable elements: from conflicts to benefits, *Nat. Rev. Genet.*, **18**, 71-86; doi: 10.1038/nrg.2016.139.
97. Fontdevila, A. (2005) Hybrid genome evolution by transposition, *Cytogenet. Genome Res.*, **110**, 49-55.
98. Soemedi, R., Cygan, K. J., Rhine, C. L., Glidden, D. T., Taggart, A. J., Lin, C. L., Fredericks, A. M., and Fairbrother, W. G. (2017) The effects of structure on pre-mRNA processing and stability, *Methods*, **125**, 36-44; doi: 10.1016/j.ymeth.2017.06.001.
99. Kralovicova, J., Patel, A., Searle, M., and Vorechovsky, I. (2015) The role of short RNA loops in recognition of a single-hairpin exon derived from a mammalian-wide interspersed repeat, *RNA Biol.*, **12**, 54-69, doi: 10.1080/15476286.2015.1017207.
100. Khavinson, V. Kh., Solovyov, A. Yu., and Shataeva, L. K. (2006) Molecular mechanism of interaction between oligopeptides and double-stranded DNA, *Bul. Exp. Biol. Med.*, **141**, 457-461; doi: 10.1007/s10517-006-0198-9.
101. Gladyshev, E. A., and Arkhipova, I. R. (2011) A widespread class of reverse transcriptase-related cellular genes, *Proc. Natl. Acad. Sci. USA*, **108**, 20311-20316; doi: 10.1073/pnas.1100266108.
102. Kim, S., and Choi, D. (2018) New role of LTR-retrotransposons for emergence and expansion of disease-resistance genes and high-copy gene families in plants, *BMB Rep.*, **51**, 55-56.
103. Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., Xiao, J., and Wang, S. (2016) Transposon-derived small RNA is responsible for modified function of WRKY45 locus, *Nat. Plants*, **2**, 16016-16023; doi: 10.1038/nplants.2016.16.
104. Trizzino, M., Kapusta, A., and Brown, C. D. (2018) Transposable elements generate regulatory novelty in a tissue-specific fashion, *BMC Genomics*, **19**, 468; doi: 10.1186/s12864-018-4850-3.
105. Vinogradov, A. E. (2011) *Base Function of Genome Structure of Eukaryotes*: Doctoral (Biol.) dissertation [in Russian], St. Petersburg.
106. Joly-Lopez, Z., and Bureau, T. E. (2018) Exaptation of transposable element coding sequences, *Curr. Opin. Genet. Dev.*, **49**, 34-42; doi: 10.1016/j.gde.2018.02.011.
107. Duan, C. G., Wang, X., Pan, L., Miki, D., Tang, K., Hsu, C. C., Lei, M., Zhong, Y., Hou, Y. J., Wang, Z., Zhang, Z., Mangrauthia, S. K., Xu, H., Zhang, H., Dilkes, B., Tao, W. A., and Zhu, J. K. (2017) A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation, *Cell Res.*, **27**, 226-240; doi: 10.1038/cr.2016.147.
108. Sinzelle, L., Izsvak, Z., and Ivics, Z. (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes, *Cell. Mol. Life Sci.*, **66**, 1073-1093; doi: 10.1007/s00018-009-8376-3.
109. Wang, J., Vicente-Garcia, C., Seruggia, D., Molto, E., Fernandez-Minan, A., Neto, A., Lee, E., Gomez-Skarmeta, J. L., Montoliu, L., Lunyak, V. V., and Jordan, I. K. (2015) MIR retrotransposons sequences provide insulators to the human genome, *Proc. Natl. Acad. Sci. USA*, **112**, 4428-4437; doi: 10.1073/pnas.1507253112.