

## Review

# The agents of natural genome editing

Guenther Witzany\*

Telos–Philosophische Praxis, Vogelsangstraße 18c, A-5111-Buermoos, Austria

\* Correspondence to: Guenther Witzany, Tel/Fax: +43-6274-6805; E-mail: witzany@sbg.at

**The DNA serves as a stable information storage medium and every protein which is needed by the cell is produced from this blueprint via an RNA intermediate code. More recently it was found that an abundance of various RNA elements cooperate in a variety of steps and substeps as regulatory and catalytic units with multiple competencies to act on RNA transcripts. Natural genome editing on one side is the competent agent-driven generation and integration of meaningful DNA nucleotide sequences into pre-existing genomic content arrangements, and the ability to (re-)combine and (re-)regulate them according to context-dependent (i.e. adaptational) purposes of the host organism. Natural genome editing on the other side designates the integration of all RNA activities acting on RNA transcripts without altering DNA-encoded genes. If we take the genetic code seriously as a natural code, there must be agents that are competent to act on this code because no natural code codes itself as no natural language speaks itself. As code editing agents, viral and subviral agents have been suggested because there are several indicators that demonstrate viruses competent in both RNA and DNA natural genome editing.**

**Keywords:** persistent virus, repetitive element, transposable element, cooped adaptation, riboagent, non-coding RNA

### Introduction

Cellular life is the main subject of biology and the history of evolution starts with the emergence of the first living cell. For decades, there was no doubt that to speak about life meant to speak about living cells and cellular assemblies. Virology elucidated the fact that viruses code for typical features that are not part of cellular life and are not found in any cell, which seems to date viruses as older than cellular life (Domingo et al., 2008; Koonin, 2009). Additionally there are some indicators that viruses and viral-derived elements are the agents that edit the genome in host organisms, such as the non-lytic but persistent colonization of prokaryotes and eukaryotes by viral agents (Villarreal, 2005).

This view is supported by the hypothesis of a pre-cellular RNA world with an abundance of competing and cooperating ribozymes. Research on the agents that act on transcripts from the DNA information storage medium demonstrated a present RNA world with high diversity and an abundance of ribozymatic and regulatory functions in all key steps and even substeps of cellular replication such as expression, transcription, translation, and repair (Gesteland et al., 2006).

At the present stage we can identify two complementary kinds of natural genome editing: (i) long lasting and stable inherited alterations of the cellular DNA genomes by persistent viral infections that alter genetic host identity and (ii) co-opted adaptations of ribozymatic or ribozyme-like parts as remnants of former viral colonizers that now act

as a great variety of regulatory elements in the present highly dynamic RNA world shortly after transcription out of stable DNA storage medium. There are several indicators that both kinds of natural genome editing agents are epigenetically regulated (Zuckerkindl and Cavalli, 2007; Maksakova et al., 2008).

### Remnants of persistent viral infection events

Persistent viral infection events most probably determine gene word order in both prokaryotes and eukaryotes (see Supplementary data). Viral colonizers therefore play major roles in evolution and diversity of organisms. A variety of examples demonstrate that these viral colonizers are still active in specific developmental processes in that they are expressed in a rather limited developmental window until the process is finished. Then they are regulatory silenced again.

Also well documented is the co-opted function of a great variety of former retroviral parts such as env, gag, pol, that now play important roles in gene regulation of host organisms. These 'defectives' (Villarreal, 2005) that now function as effective regulatory elements share similar features as their viral relatives and in most cases act as ribozymatic structures or in ribozyme-like functions. They do not alter inheritable DNA content but dynamically act on RNA transcripts, although there are some examples in which RNA genome editing may also become conserved status, especially by being reverse transcribed into DNA.

### Repetitive and transposable elements

Due to most prominent persistent remnants of former viral infection events we have to look at transposable elements. Since the

research of Barbara McClintock, it has been proven that the genome as an information storage medium is not a fixed molecular structure, but bears highly dynamic elements which change their position within the genetic sequence order, known as transposable elements or mobile genetic elements. Formerly, they were also known as ‘jumping genes’. Some are active as RNA intermediate (retrotransposons), while some are active as mobile DNA (DNA transposons). Mobile genetic elements are interacting genetic agents (Bapteste and Burian, 2010) that replicate by either cut and paste (DNA transposons; class II elements) or copy and paste (retrotransposons; class I elements) processes. Mobile genetic elements are flanked by repeat sequences.

Together with non-coding DNAs which encode a variety of non-coding RNAs (see below), mobile genetic elements share repeat sequences as essential parts of their identity. This is an important feature, because non-repeat sequences are the most relevant part of protein coding sequences of translational mRNAs, a coherent protein coding line-up of exons in which all intronic sequences are spliced out. The intronic sequences are known as regulatory elements of great diversity, and all of them possess repeat sequences. In contrast to former assumptions on the dominance of protein coding DNA in the (human) genome, we now know that only 1.2% of the DNA codes for proteins, 43% consists of repeat and mobile elements, up to 18% consists of dispersed elements. Repetitive sequences are found either as interspersed repeats up to 20–30 kb or tandem repeats, rather short DNA fragments adjacent to each other (Jurka et al., 2007).

Contrary to the one gene–one protein consensus of the last century, we now know that there are many overlapping genes and dispersed genetic elements which together constitute genetic elements. The largest part of complex genomes is represented by repetitive intronic mobile elements, which essentially contribute to evolution and diversification of host genomes and are not selfish or parasitic, but more or less symbiotic. Long terminal repeats (LTRs) constitute compact nuclear structures such as centromeres and the related telomeres (Witzany, 2008). Repetitive genetic elements delineate centromeres and form telomeres by non-LTR-retrotransposons (Shapiro and Sternberg, 2005).

A large part of genomic repetitive DNA is reverse transcribed, and plays a major role in the physical structured order of the genome as well as formatting functions for expression, replication, transmission, repair, restructuring, cell division, and differentiation (Sternberg and Shapiro, 2005). These genomic repetitive DNA sequences, that are reverse transcribed into retroelements, have major roles in genome formatting and regulation and are active/competent in promoter/enhancer processes, transcript elongation, tissue-specific mRNA targeting, mRNA translation, identifying ‘recognition’ sequences for origins of replication, and chromatin functions (Sternberg and Shapiro, 2005). If we look at the essential roles that repeat elements play in (i) transcription (promoters, enhancers, silencers, transcription attenuation, terminators, and regulatory RNAs), (ii) post-transcriptional RNA processing (mRNA targeting, RNA editing), (iii) translation (enhancement of SINE mRNA translation), (iv) DNA replication (origins, centromeres, telomeres, meiotic pairing, and recombination), (v) localization and movement, chromatin organization (heterochromatin, nucleosome positioning

elements, epigenetic memory, methylation, epigenetic imprinting, and modification), (vi) error correction and repair (double-strand break repair by homologous recombination, methyl-directed mismatch repair) and (vii) DNA restructuring (antigenic variation, phase variation, genome plasticity, uptake and integration of laterally transferred DNA, chromatin diminution, VDJ recombination, and immunoglobulin class switching), it is apparent that these essential agents are all retroelements such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), LTR retrotransposons, non-LTR retrotransposons, and ALUs (Sternberg and Shapiro, 2005). Retrotransposons are stress-inducible elements that may become active during maternal stress, which act during early foetal life and may become non-Mendelian inherited epigenetic traits (Wagner et al., 2008; Huda and Jordan, 2009; Sciamanna et al., 2009). As demonstrated by a variety of experiments, environmental influences then may alter epigenetic marking which can be inherited for several generations (Jablonka and Lamb, 1989, 2002; Sternberg, 2002; Jaenisch and Bird, 2003; Slotkin and Martienssen, 2007). Especially in the brain, environmental information can be conveyed into RNA-based regulatory networks via RNA editing that leads to inheritable RNA-directed epigenetic changes (Mattick, 2009).

Reverse transcribed repetitive elements regulate genome organization. Interestingly, similar to viruses, these elements are widely dispersed and fractioned, although they share a coordinated behaviour. Changes in repetitive elements establish new genome architectures. Whereas in mammals, genome formatting occurs mainly by SINEs, LINEs, and ALUs, in plant genomes, LTR retrotransposons play major roles in this respect. Genome size in both is determined by repetitive DNA abundance. Interestingly, the distance between coding sequences and regulatory repetitive sequences is an important parameter (Zuckermandl, 2002).

Microbial cells are protected from viral infections by clustered regularly interspaced short palindromic repeats that act together with proteins to cleave single-stranded RNA (preferentially within U-rich regions). These repeats are therefore evolutionarily very old (Beloglazova et al., 2008; Haurwitz et al., 2010).

All of these features have important consequences on our genetic perspective because they indicate that the organization of proteins can change without modifications in the coding sequences just by the alteration of splicing processes or even the rearrangements of coded sequences, i.e. the exons in the mRNA. Interestingly, most A to I RNA editing occurs within ALU elements (Levanon et al., 2005) and retroelements can be exapted (co-opted) into neogenes (see Supplemental data), which means that retroelements are essential parts of the *de novo* generation of gene inventions (Brosius, 1999, 2003, 2009; Blackburn, 2000; Mallet et al., 2004).

Additionally, it must be mentioned that transposable elements may contribute to major evolutionary processes in altering and expanding host genomes by changes of gene regulation or co-opted adaptation. But transposable elements must also be suppressed sufficiently so as not to cause disease (Schumann, 2007), and are therefore the subject of epigenetic regulation (Lisch, 2008). Such suppression is stabilized by epigenetic regulation such as RNA interference, DNA methylation, and histone

modifications (Slotkin and Martienssen, 2007). The suppression states may become unstable by environmental changes that cause stress to populations. Mobilized transposable elements can restructure the genome and displace populations from adaptive peaks (Zeh et al., 2009).

### RNA agents at the roots of the tree of life: the RNA world hypothesis

With good reason, it is assumed that before the emergence of DNA, there was an ancient RNA world with a great variety and high density of self-replicating ribozymes. From the early RNA world perspective, the whole diversity of processes within and between evolutionarily later derived cells depends on various RNAs. Therefore, the characteristics of modern RNAs suggest a pre-cellular RNA world which must have been dominated by consortia-based evolution similar to what is currently seen in RNA viruses (Domingo et al., 2008). A variety of RNAs can be identified at the roots of the tree of life, such as genetic polymers inside membrane vesicles of a hypothesized protocell (Chen et al., 2006) and riboswitches (Breaker, 2006), a variety of catalytic strategies of self-cleaving ribozymes that act as sequence-specific RNA cleavage agents (Tang and Breaker, 2000; Ke and Doudna, 2006), the structure and function of group I introns (Houglund et al., 2006), and the roles of RNA in the synthesis of proteins (Moore and Steitz, 2006).

Small nucleolytic ribozymes resemble hammerhead, hairpin, hepatitis delta virus and *Neurospora grassa* Varkud satellite ribozymes. Ribosomal RNA components are catalytically active in polypeptide synthesis and are of major importance in every living cell (Hamann and Westhof, 2007). These enzymes have important features such as (i) the role of ribosomes in the translation from RNA into protein (Noller, 2006), (ii) a great diversity of actions in the modern DNA world necessary for every cellular process seen with ribonucleoproteins (Cech et al., 2006), (iii) small nuclear ribonucleoproteins (snRNPs) (Tycowski et al., 2006), (iv) small nucleolar RNPs (Matera et al., 2007), (v) the assemblies which build spliceosomes, the insertion/deletion competence for site-specific modifications of RNA molecules (Simpson, 2006) and (vi) the unique feature of all retroelements, i.e. reverse transcriptase and other telomerases (Blackburn, 2006). Last but not least, we can find crucial parts of the present RNA world such as group II introns with self-splicing competencies (Pyle and Lambowitz, 2006), the important roles of SINEs and LINEs (Weiner, 2006), as well as the whole range and abundance of non-coding RNAs (Witzany, 2009). Also, the RNA ligase of T4 is found in all three domains of life (Ho and Shuman, 2002). All of these agents can be identified as descendants of an early RNA world, which evolved prior to DNA-encoded cellular life and are the predecessors of the functions of cellular life present since the last universal common ancestor in all three domains of life.

Group II introns are self-splicing RNAs capable of removing themselves from their primary transcripts (Valles et al., 2008), which most likely are forerunners of the splicesomal introns because a few members of this intron class reassemble on the RNA level from independent transcripts (*trans*-splicing) (Knoop and Brennicke, 1994). Some of them encode reverse

transcriptases and are active retroelements. The reverse transcriptases of group II introns are related to the reverse transcriptases of non-LTRs and both are mobile genetic elements, which use target-primed reverse transcription (Toor et al., 2001).

Based on these facts, some authors suggest an RNA-continuity hypothesis, which assumes that all RNA-processing events are relics (Demongeot et al., 2009) that can be traced back to the pre-cellular early RNA world and that lead later on to the origin of the ribosomes, spliceosomes, snoRNAs and mRNAs, and via exaptation (see above), are the origins of introns most relevant to the eukaryotic superkingdom. Introns therefore represent genetic invasions (viral-like infections that do not harm host) with consequences on host genome identity and regulation (Villarreal, 2005; Penny et al., 2009).

### Important riboagents act as module-like consortia

Three RNA assemblages are currently known to play vital roles in editing genetic text as a read-and-write medium. This indicates sequence-specific identification competence for insertion and deletion activities, which alter semantic content (the function which leads to altered regulation or altered protein production) of primary transcripts out of the DNA storage medium. Editosomes, spliceosomes, and ribosomes constituted of a variety of components that counter-regulate themselves during assembly. Slight changes in the composition structure delete their functionality, so it must be assumed that the assembly is a complementarity, that all parts play an important function. These subcomponents are ribomodules that constitute highly active riboagents with a variety of competencies to act on genetic transcripts.

#### Editosome

RNA editing is a co- or post-transcriptional process, which alters the RNA sequence derived complementarily to the DNA from which it was transcribed. RNA editing changes gene sequences at the RNA level. The edited mRNA specifies an amino acid sequence that is different from the protein that could be expected and is encoded by the genomic DNA of the primary transcript (Takenaka et al., 2008). RNA editing alterations of such transcribed RNA sequences occur by modification, substitution, and insertion/deletion processes (Smith, 2008). According to the suggestion of Grosjean and Björk (2004), any sequence alteration that changes the genetic meaning of a transcript is termed editing, whereas structural changes solely are called modifications.

We know of U insertion/deletion in kinetoplastids, *Trypanosoma*, *Leishmania*, *Crithidia* and *Bodonis* through guided RNA targeting, followed by U insertion or deletion (and ligation) in mRNAs. Interacting guide RNAs and pre-mRNAs integrate Watson-Crick base pairing as well as G:U base pairing to determine the cleavage site followed by U insertion or U deletion (Stuart et al., 2005; Carnes and Stuart, 2008), C insertion in pharum through co-transcriptional C insertion in mRNAs, rRNAs, and tRNAs. G insertion in paramyxoviruses occurs through co-transcriptional G insertion in mRNAs, A insertion in Ebola viruses and GA deletion in rats. C to U conversion in plants occurs through C-deamination in mRNAs and similar in mammals. In plant mRNAs, hundreds of editing sites within mRNAs of chloroplasts and mitochondria have been found (Odintsova and Yurina, 2000, 2005).

RNA editing is most important in the central nervous system of mammals. U to C conversion is found in land plants through U amination in mRNAs. A to I conversion through A deamination occurs in several species (Smith, 2008). RNA editing also occurs in tRNAs which affect codon sense by adenosine deaminases active on RNA (ADAR), especially in yeast and mammals by adenine deaminases active on tRNA (ADAT), which are important for translation processes. This includes A to I editing as well as C to U editing of tRNAs. tRNA editing shows some features of quality control, especially the precise recognition of substrates and substrate discrimination capability in a pool of nearly identical substrates (Alfonzo, 2008).

Editing sites have to be identified individually to differentiate a C to be edited from a C which should not be edited. The discriminating information can be found in the nucleotide sequence surrounding a given site. This means that syntactic context is relevant for recognition. Thus, each editing site carries its own recognition context (Takenaka et al., 2008). Interestingly, most of the 12000 currently identified A to I editing sites in about 1200 different genes in the human genome have been found not in coding regions but in Alu-repeat sequences.

RNA editing plays a major role in the intracellular RNA world and represents the natural habitat of most remnants of the early RNA world, such as the great variety of RNA transcript processing with its coordinated and associated protein molecules up to all steps and substeps of translational processes. It is a complementary feature of RNA processing, such as RNA modification, RNA splicing, RNA catalysis by group I intron RNases and other small ribozymes and RNA interference (Homann, 2008).

To get a mature and functional RNA for translation, it must be determined which transcript has to be processed. In this respect, RNA editing alters the sequence information of transcripts which means that the end products such as RNAs or proteins can be different, although the DNA is the same. RNA editing must recognize the site which has to be edited. This is not a random process. In a second step, the editing site has to be processed by RNAs or proteins. In a further step, enzymatic agents have to be directed to the editing site to process the editing. All steps are catalysed by ribonucleoprotein complexes that have been termed the editosome, which plays similar important roles as related agents such as the ribosome or spliceosome. Additionally, the editosome is constituted of a variety of subcomponents, which counterbalance each other (Homann, 2008).

Interestingly, a number of protein motifs have been identified, which are relevant to RNA binding such as the RNA recognition motif (RNA binding domain), the ribonucleoprotein domain, the K-homology domain and the double-stranded RNA-binding domain. Most of these serve as high-affinity recognition motifs of any RNA sequence (Homann, 2008).

RNA editing changes the molecular syntax of RNA transcripts and therefore changes the information content by substitution or an insertion/deletion competence, which has similar consequences of sign-using agents which change character combinations out of the repertoire of signs available to species-specific code-using populations. A limited number of signs in evolutionary protocols such as the DNA genetic code can be used to produce an abundance of messages, which

means, in this case, the production of regulatory actions such as recombination, repair, start, stop, and in parallel protein products. These are processed by RNAs and associated proteins, which act in a coordinated manner and in some cases not only on mRNAs but also additionally on rRNAs and tRNAs (Gott, 2003).

This process can be found in viruses (Hausmann et al., 1999), mitochondria of single cell flagellates (Gott and Rhee, 2008) in plant organelles and mitochondria (Homann, 2008). The APOPEC family are prominent RNA editing agents in mammals, which are involved in C and U deamination at the DNA level and result in immune functions against retroviral infections (Holmes et al., 2003). Today, it is assumed that all eukaryotic post-transcriptional RNAs are edited i.e. rRNA, RNAs, microRNAs (miRNAs) and tRNAs, so that we can conclude that any pre-translational RNA has to be edited to become a mature template for translation (Ochsenreiter and Hajduk, 2008). All of these are strictly coordinated steps such as (i) the formation of a specific and flexible RNA structure, (ii) site-specific binding of a protein factor, (iii) creation of a nucleation site to get a certain stem loop structure and (iv) restricting the whole process to a specific target site. RNA editing is the opposite of a randomly occurring process.

Interestingly, the large majority of plant RNA editing changes transcripts back into those forms, which are conserved over distantly related plant species, so it seems to function as a kind of repair mechanism (Speijer, 2008). In this respect, a number of editosome proteins of the 20S editosome have shown high sequence similarity to DNA repair enzymes (Panigrahi et al., 2003). C to U editing frequently occurs in the chloroplasts of angiosperms, but no consensus motif has been identified as of yet in the case of splice sites. Interestingly, phylogenomic analyses have demonstrated that editing frequencies and editing patterns do not correlate with the phylogenetic trees of plants (Sugiura, 2008). In flowering plant mitochondria, the 400 RNA editing sites in the coding regions of mRNAs involve mainly C to U changes, whereas in non-flowering plants, the reverse reaction can be seen in the amination of U to C. The evolutionary forerunners of plants, green algae, do not show any RNA editing (Turmel et al., 2003; Takenaka et al., 2008), which may indicate that the evolution of plants occurred by genetic acquisition of the spliceosome via plant precursors (Knoop and Rudinger, 2010).

Some data indicate that ADARs recognize adenines specifically in different sequence contexts (Homann, 2008). RNA editing occurs mainly in non-coding regions, while some A to I editing occurs in coding regions. Since adenosines are interpreted as guanosines during translation, this nucleotide alteration can trigger a codon exchange in the coding region (Bass, 2002). ADARs are involved in miRNA-mediated genetic regulation, epigenetic heterochromatin formation, and RNA interference (Jantsch and Öhman, 2008).

RNA editing and RNA interference are in competition, which indicates an addiction module (Jantsch and Öhman, 2008). In certain tissues, ADAR expression is suppressed by RNA interference. All ADARs share a conserved deaminase domain at their C-terminal ends and a varying number of double-stranded RNA-binding domains. Also, the amino termini are highly variable and, additionally, a variety of nuclear export signals and nuclear



import signals have been identified (Jantsch and Öhman, 2008). Interestingly, the phenomenon of hyper-editing, a kind of non-selective editing, is mainly found in viral RNAs and seems to be a kind of antiviral defence reaction. Hyper-editing in non-viral organisms can be found in non-coding RNA regions, e.g. in regions with Alu elements in the human genome, and ADAR-mediated RNA editing of repetitive elements is poorly conserved. ADARs do not have unlimited access to RNA transcripts and compete with other obligate RNA binding and modifying proteins working on any primary RNA transcript, such as snRNPs or Drosha and Dicer dsRNA-binding proteins, so that ADAR enzymes can act on the transcript at specific target sites.

The 20S editosome components, which act as insertion/deletion editors, seem to be a perfect example of an addiction module in that both 'parties' act as counterbalanced subcomplexes (Carnes and Stuart, 2008). Especially, if we look at the subcomplexes involved in kinetoplastid RNA editing, this is strikingly interesting. Here, we find kinetoplastid RNA editing involving proteins, endonucleases, exoUases, TUTases, ligases, and helicases, all of which are with different functions as several subsets (Carnes and Stuart, 2008).

#### Spliceosome

In complex organisms, especially in mammals, before the transcript is processed to the final mRNA for translation into protein, natural genetic engineering occurs which we call alternative splicing. As in ribosome and editosome assembly and also in spliceosome construction, a ribonucleoprotein complex is assembled in various steps that cuts out introns and splices exons together. RNA editing predates splicing and is heavily interconnected, so that the editosome and spliceosome are important co-players. The spliceosomal ribonucleoproteins are mainly small nuclear RNAs that are interconnected with at least 300 different proteins, which are involved in mammalian pre-RNA splicing. The subunits are U1, U2, U4/U6, and U5 snRNPs, the PRP19 complex protein subunit, and some smaller protein components. In addition to the major player (the U2-dependent spliceosome), metazoan cells contain a minor spliceosome (U1 dependent). If they act within the same transcript, they complementarily interact in intron excision (Matlin and Moore, 2007), but the minor variant works on a rather limited number (1%) of all introns. The U6 subunit shows similarities with the structure of a self-splicing group II intron catalytic effector domain 5 (Seetharamann et al., 2006).

Interestingly, the variety of steps in which the subunits of the final spliceosome are produced are counterbalanced within competing parts that regulate the stepwise processing of subcomplexes of the spliceosome. Beneath the first product, the H complex, the E complex which competes the H complex, the SR proteins within the E complex, the A complex, the B complex and the C complex are steps until the final catalytic processing occurs, i.e. exon ligation.

After this final splicing procedure of the mature spliceosome, the remaining RNA products are actively discharged from the spliceosome and the snRNPs and are recycled for further catalytic processes. Dependent on these regulations, the end product may vary concerning the context dependency of the regulation process, which is highly sensitive to various needs and

circumstances. In consequence, spliceosomal regulation differentiates the inclusion (splicing enhancers) or exclusion (splicing silencers) of exons in the final mRNA (House and Lynch, 2008). Splicing regulation occurs by competing *cis*-acting elements that precisely balance regulatory proteins. This means that the competition of exon ligation and ATP-dependent exon rejection may change spliceosomal assembly from the beginning until the end. This means alternative splicing is regulated up until the latest stages of splicing reactions. The highly dynamic character of spliceosomal actions is a further example of the importance of concrete *in vivo* entanglement in the organism, i.e. the pragmatics of genome editing. Along a row of keystones in spliceosomal assembly, spliceosomal functions may lead to different choices and decisions according to local needs and cellular (organismal) requirements.

The crucial competence of the spliceosome is the appropriate and precise splice site recognition of consensus sequences that define exon–intron boundaries. According to the flexibility of splicing in the various time windows of spliceosomal assembly, the correct identification of sequence recognition is a multiple competence, which may change quickly (Matlin and Moore, 2007). As happens in the editosome, the spliceosome can also vary the end products of a given genetic sequence (gene).

#### Ribosome

The detections of the early RNA world prior to cellular life, the whole range of ribozymatic activity in the present biological world and the complementary properties of an abundance of RNA agents have opened a perspective in which RNAs are agents which act on genetic sequences, stabilize complex forms by helper proteins, and regulate/modulate proteins in all cellular life without altering DNA sequence order. One of the most abundant key agents in this respect is undoubtedly the ribosome, which is the key player in any process we term as translation, i.e. in any protein synthesis from of a mature mRNA transcript after editosome and spliceosome activities. Because the ribosome is at least the most important player in cellular life replication and is absent in viral genomes, it must have been a crucial evolutionary step from the capsid-encoding organisms of the pre-cellular RNA world and the ribosome-encoding organisms in the cellular life world (Forterre and Prangishvili, 2009, 2010).

Ribosomes are composed of two-thirds RNA and one-third protein. Ribosomes are assembled into a functional complex. As it is understood today, ribosomal proteins are vital for structural stabilization. Around the catalytic site of the ribosome, there are only RNAs and no ribosomal proteins (Belousoff et al., 2010). This means that the ribosome was originally a ribozyme and the proteins are not involved in the catalytic activity (Moore and Steitz, 2006; Hamann and Westhof, 2007). RNA agents invented DNA as a more stable information storage medium and proteins as more functional components interconnected with RNA structures.

A similar functional division of labour is found in the H/ACA ribonucleoproteins, which mediate post-transcriptional introduction of pseudouridine into ribosomal RNAs and spliceosomal small nuclear RNAs. The RNAs dictate site specificity, while the protein components mediate stability (Karjolic and Yu, 2008).

Interestingly, tRNAs did not evolve first to serve in protein synthesis. As demonstrated by Maizel et al. (1993, 1999), they

represent a composition of formerly different components, with one-half serving to mark single-stranded RNA for replication in the RNA world, whereas the lower half of the tRNA is a later acquisition. As demonstrated in nanoarchaeota (Randau et al., 2005), the various tRNA species are encoded as two half genes, one encoding the conserved T-loops and 3' acceptor stem, the other encoding the D-stem and the 5' acceptor stem subunit. In nanoarchaeota, the CCA sequence (which is important in tRNAs for protein synthesis in nearly all cellular life) is not encoded in the tRNA genes but is added posttranscriptionally by an enzyme (Xiong et al., 2003). It seems that the evolution of protein synthesis has coupled with a variety of older genetic agents and seems to be another example of co-opted adaptation. Coherent with these findings are investigations, which have demonstrated that pre-tRNAs act in self-cleavage, which is clearly a ribozymatic reaction independent of translation (Phizicky, 2005; Wegrzyn and Wegrzyn, 2008).

#### *Non-coding RNAs*

Non-coding RNAs that function in gene regulation coordinate and organize various actions, such as chromatin modification and epigenetic memory, transcriptional regulation, control of alternative splicing, RNA modification and RNA editing, control of mRNA turnover, and control of translation and signal transduction (Amaral et al., 2008). In contrast to former opinions on the expression levels of genomes, it is now increasingly clear that most eukaryotic genomes are highly expressed and a great abundance of non-coding RNAs with regulatory functions are transcribed.

Most of these non-coding RNAs are alternatively spliced and divided into smaller RNAs that are integral parts of RNP complexes. They regulate nearly all aspects of gene regulation. Small RNA species include miRNAs, siRNAs, small nuclear RNAs and small nucleolar RNAs, and transfer RNAs (Yazgan and Krebs, 2007). Although recent research has tried to evaluate the enormous regulatory networks of small RNAs, the role of thousands of longer transcripts is not yet clear. We know that they play important roles in histone modification and methylation, that is, epigenetic control of developmental processes such as the mammalian HOX clusters (Amaral et al., 2008), and also transcriptional interference, promoter inactivation and effects on enzymatic pathways. Interestingly, these large non-coding RNAs are found as interlacing and overlapping sense and antisense transcripts derived from introns or intergenic regions. Similar to their smaller relatives, they are involved in the formation of RNP complexes.

miRNAs and their associated proteins are RNPs. Small non-coding RNAs also share a special competence for epigenetic regulation of gene expression and are derived from repetitive genomic sequences (Farazi et al., 2008). The capacity for epigenetic regulation of gene expression includes the 'recognition' (identification) of specific sequences in other nucleic acids and is common to RNAs (Filipowicz, 2000), especially small nuclear RNAs and tRNAs that (i) identify splice junctions in both pre-mRNAs and codons and (ii) process both the subunits of the spliceosome and the ribosome (Mattick, 2003). Interestingly, this indicates that tRNAs are not only translational factors but are also involved in the regulation of various processes including replication of various replicons such as

plasmids. Their ribozymatic capability has also been shown in that pre-tRNA can self-excite intronic sequences (Wegrzyn and Wegrzyn, 2008). There is also evidence that tRNAs can flow back to the nucleus from their site of action in the cytoplasm (Phizicky, 2005).

This implicates non-coding RNA capacity for self/non-self distinction as well as for identifying the molecular syntax. If one of these does not function, that is, error or damage occurs, the regulation or structural features of small non-coding RNAs do not function. A great variety of non-coding RNAs are produced in transcriptional processes, such as small RNAs, long non-coding RNAs, miRNAs, piwi RNAs, small nuclear RNAs and small nucleolar RNAs (relatives of guide RNAs) (Witzany, 2009).

Interestingly, small nuclear RNAs and small nucleolar RNAs counter-regulate their activities that indicate addiction modules. Small nucleolar RNAs guide pseudouridylation and 2'-O-ribose methylation of rRNAs in forming duplexes with their target. Small Cajal body-specific RNAs guide modifications of spliceosomal RNAs. Small nucleolar RNAs have been identified as a family of mobile genetic elements and other small RNAs derived from snoRNAs (Taft et al., 2009), which indicates that they are the most ancient and numerous family of non-coding RNAs (Dieci et al., 2009).

RNA species such as miRNAs, siRNAs, and piwi RNAs in animals and plants mediate processes in that they guide the binding of protein complexes to specific nucleic acid sequences (Ambros and Chen, 2007). They act both in regulatory processes at the transcriptional level (e.g. by endogenous siRNAs) (Watanabe et al., 2008) when the action potential is activated out of the evolutionary protocols (Vetsigian et al., 2006) fixed in the DNA storage medium, and at the post-transcriptional level in that they stabilize mRNA and translation into proteins (Chen and Rajewsky, 2007). This means that small non-coding RNAs do not solely mediate the transfer of genetic information from DNA to protein, but also act as sequence-specific regulators in the expression of other RNA transcripts and, interestingly, in silencing specific transposons (Bartel, 2004, 2009; Chu and Rana, 2007). Control patterns include mRNA degradation (siRNAs), translational repression (miRNAs), heterochromatin formation and transposon control (piwi RNAs) (Chu and Rana, 2007).

Other small RNAs such as endogenous miRNAs and siRNAs share biogenesis and can perform interchangeable functions (Doench et al., 2003). They cannot be distinguished by their chemical composition or their action, but they differ in their production pathways: (i) miRNAs derive from genetic sequences that are different from known genes, while siRNAs derive from mRNAs, transposons, and viruses, (ii) miRNAs are processed out of transcripts that can form RNA hairpins, while siRNAs are processed from long bimolecular RNA duplexes, (iii) miRNAs are always conserved in related organisms, while endogenous siRNAs are rarely conserved, (iv) miRNAs are produced from genes that are specialists in the silencing of different genes, while siRNAs are typically auto-silencing, such as viruses, transposons and repeats of centromeres (Bartel, 2004, 2009). siRNAs are expressed by extended double-stranded regions of long inverted repeats that can inhibit the expression of nearly any target gene in response to double-stranded RNA and have a very efficient and ancient immune

functions against genetic parasites (Fire, 2005; Sontheimer and Carthew, 2005; Siomi and Siomi, 2009). They function by identifying foreign RNA sequences and inhibiting their replication. The ancient RNAi immune function is based on self/non-self identification competence (Fire, 2005; Obbard et al., 2009). Many of these elements are retrotransposons or transposons and are encoded in the repetitive sequences of the genome (Sontheimer and Carthew, 2005).

miRNAs are single-stranded RNAs with 19–25 nucleotides in length and are generated from endogenous hairpin transcripts of precursor miRNAs with 70 nucleotides (Kim, 2005). The transcription of this pre-miRNA is processed by RNA polymerases pol II and pol III, whereas pol II produces mRNAs, small nucleolar and small nuclear RNAs of the spliceosome. Pol III produces shorter non-coding RNAs, such as tRNAs, some rRNAs, and a nuclear RNA that is part of the spliceosome (Bartel, 2004, 2009). They control not only developmental timing, haematopoiesis, organogenesis, apoptosis, and cell proliferation, but also fat metabolism in flies, neuronal patterning in nematodes, and control of leaf and flower development in plants (Bartel, 2004, 2009). Most of them are processed out of introns. It has been predicted that every metazoan cell type at each developmental stage has a distinct miRNA expression profile (Bartel, 2004, 2009). The most characteristic differences of miRNAs acting in plants and animals are found in the stem loop.

miRNAs and siRNAs seem to have descended from transposable elements with an inherent regulatory ratio to gene regulation that is fulfilled by a variety of siRNAs or miRNAs that act in a coordinated manner in that they share a division of labour in hierarchical steps of suppression and amplification. This is indicated in transposable elements that encode both siRNAs and miRNAs (Piriyaopongsa and Jordan, 2008). They can be found in intronic regions and build stem loop structures (hairpins) as a common feature of active RNA species, such as ribozymes. The defence mechanism of host genomes against transposable element invaders is through siRNA evolved into miRNAs with a new regulatory complexity and a new phenotype. First evolving as an immune function, they were later co-opted as a tool for complex regulatory pathways in host gene expression (Piriyaopongsa and Jordan, 2008). This co-option of identical competencies for different purposes seems to be a common evolutionary pattern for regulatory controls that can be flexibly altered and rearranged to cause phenotypic variation without altering basic components (Mattick and Gagen, 2001).

### Conclusions

Persistent viral lifestyles are counterbalanced viral properties (addiction modules) that transfer complete genetic data sets into host genomes and alter the DNA genetic identity of the host and, additionally the formerly competing viral agents without damage to the host genome content.

RNA consortia edit genetic code without altering inherited DNA content. Transcribed out of DNA cellular sequences, the RNA activated inhabitants from former viral infection events act as modular tools for cellular needs in nearly all cellular processes. That they act not as lytic agents but as part of the host genetic identity is the result of inhabitation by counterbalanced

competing genetic parasites. Addiction modules do not only change genetic identity but also enrich immune functions of host organisms against related parasites through T/A modules. Addiction modules are clearly the result of stable consortial interactions.

Insertion, deletion, rearrangement, recombination, replication, and repair of nucleotide sequences in all detailed steps are coordinated processes in a timely manner. An abundance of small RNAs act as competent agents in most cases combined within a network of other RNAs that bind proteins which together build a fine-tuned network in the division of labour. Interestingly, formerly viral RNA parts now act as modules of cellular gene regulation. Every cellular regulation pattern is counterbalanced by at least two antagonists, which are to be identified as former viral settlers.

Some RNAs such as ribozymes are able to self-replicate. Reverse transcriptase is the only known enzyme that synthesizes DNA out of RNA templates. It is the essential feature of reverse transcribing viruses and retroelements such as retrons, retrotransposons, and retroplasmids.

Essential modules of RNA viruses are ribozymes and ribozymatic structures, which autocatalyse and build ribomodules. Conserved structures such as tRNAs and the complex editosomes, spliceosomes, and ribosomes consist of ensembles of such consortia of riboagents. As endogenized modules, they regulate genetic expression of host organisms.

### Supplementary material

Supplementary material is available at *Journal of Molecular Cell Biology* online.

**Conflict of interest:** none declared.

### References

- Alfonzo, J.D. (2008). Editing of tRNA for structure and function. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 33–50.
- Amaral, P.P., Dinger, M.E., Mercer, T.R., et al. (2008). The eukaryotic genome as an RNA machine. *Science* 319, 1787–1789.
- Ambros, V., and Chen, X. (2007). The regulation of genes and genomes by small RNAs. *Development* 134, 1635–1641.
- Bapteste, E., and Burian, R.M. (2010). On the need for integrative phylogenomics and some steps toward its creation. *Biol. Phil.* 25, 711–736.
- Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* 116, 281–297.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Bass, B.L. (2002). RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846.
- Beloglazova, N., Brown, G., Zimmerman, M.D., et al. (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* 283, 20361–20371.
- Belousoff, M.J., Davidovich, C., Zimmerman, E., et al. (2010). Ancient machinery embedded in the contemporary ribosome. *Biochem. Soc. Transact.* 38, 422–427.
- Blackburn, E.H. (2000). The end of the (DNA) line. *Nat. Struct. Biol.* 7, 847–850.
- Blackburn, E.H. (2006). Telomerase RNA. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 419–436.

- Breaker, R.R. (2006). Riboswitches and the RNA world. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 89–107.
- Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238, 115–134.
- Brosius, J. (2003). The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118, 99–115.
- Brosius, J. (2009). The fragmented gene. *Ann. NY Acad. Sci.* 1178, 186–193.
- Carnes, J., and Stuart, K. (2008). Working together: the RNA editing machinery in *Trypanosoma brucei*. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 143–164.
- Cech, T.R., Moras, D., Nagai, K., et al. (2006). The RNP world. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 309–326.
- Chen, K., and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8, 93–103.
- Chen, I.A., Hanczyc, M.M., Sazani, P.L., et al. (2006). Protocells: genetic polymers inside membrane vesicles. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 58–88.
- Chu, C.Y., and Rana, T.M. (2007). Small RNAs: regulators and guardians of the genome. *J. Cell. Physiol.* 213, 412–419.
- Demongeot, J., Glade, N., Moreira, A., et al. (2009). RNA relics and origin of life. *Int. J. Mol. Sci.* 10, 3420–3441.
- Dieci, G., Preti, M., and Montanini, B. (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 94, 83–88.
- Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. *Genes Dev.* 17, 438–442.
- Domingo, E., Parrish, C.R., and Holland, J.J. (2008). *Origin and Evolution of Viruses*, 2nd edn. San Diego: Academic Press.
- Farazi, T.A., Juranek, S.A., and Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135, 1201–1214.
- Filipowicz, W. (2000). Imprinted expression of small nucleolar RNAs in brain: time for RNomics. *Proc. Natl. Acad. Sci. USA* 97, 14035–14037.
- Fire, A. (2005). Nucleic acid structure and intracellular immunity: some recent ideas from the world of RNAi. *Q. Rev. Biophys.* 38, 303–309.
- Forterre, P. (2010) Manipulation of cellular syntheses and the nature of viruses: the virocell concept. *C. R. Chim. doi:10.1016/j.crci.2010.06.007*.
- Forterre, P., and Prangishvili, D. (2009). The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann. NY Acad. Sci.* 1178, 65–77.
- Gesteland, R.F., Cech, T.R., and Atkins, J.F. (2006). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press.
- Gott, J.M. (2003). Expanding genome capacity via RNA editing. *C. R. Biol.* 326, 901–908.
- Gott, J.M., and Rhee, A.C. (2008). Insertion/deletion editing in *Physarum polycephalum*. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 85–104.
- Grosjean, H., and Bjork, G.R. (2004). Enzymatic conversion of cytidine to lysidine in anticodon of bacterial isoleucyl-tRNA—an alternative way of RNA editing. *Trends Biochem. Sci.* 29, 165–168.
- Hamann, C., and Westhof, E. (2007). Searching genomes for ribozymes and riboswitches. *Genome Biol.* 8, 210.
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., et al. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329, 1355–1358.
- Hausmann, S., Garcin, D., Delenda, C., et al. (1999). The versatility of paramyxovirus RNA polymerase stuttering. *J. Virol.* 73, 5568–5576.
- Ho, C.K., and Shuman, S. (2002). Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl. Acad. Sci. USA* 99, 12709–12714.
- Holmes, A.J., Gillings, M.R., Nield, B.S., et al. (2003). The gene cassette meta-genome is a basic resource for bacterial genome evolution. *Environ. Microbiol.* 5, 383–394.
- Homann, M. (2008). Editing reactions from the perspective of RNA structure. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 1–32.
- Houglund, J.L., Piccirilli, J.A., Forconi, M., et al. (2006). How the group I introns works: a case study of RNA structure and function. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 133–205.
- House, A.E., and Lynch, K.W. (2008). Regulation of alternative splicing: more than just the ABCs. *J. Biol. Chem.* 283, 1217–1221.
- Huda, A., and Jordan, K.I. (2009). Epigenetic regulation of mammalian genomes by transposable elements. *Ann. NY Acad. Sci.* 1178, 276–284.
- Jablonka, E., and Lamb, M.J. (1989). The inheritance of acquired epigenetic variations. *J. Theor. Biol.* 139, 69–83.
- Jablonka, E., and Lamb, M.J. (2002). The changing concept of epigenetics. *Ann. NY Acad. Sci.* 981, 82–96.
- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33, 245–254.
- Jantsch, M.F., and Öhmann, M. (2008). RNA editing by adenosine deaminases that act on RNA (ADARs). In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 51–84.
- Jurka, J., Kapitonov, V.V., Kohany, O., et al. (2007). Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genom. Human Genet.* 8, 241–259.
- Karijolic, J., and Yu, Y.T. (2008). Insights into the protein components of the box H/ACA RNP. *Curr. Proteomics* 5, 129–137.
- Ke, A., and Doudna, J.A. (2006). Catalytic strategies of self-cleaving ribozymes: relics of an RNA world? In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 109–131.
- Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* 6, 376–385.
- Knoop, V., and Brennicke, A. (1994). Evidence for a group II intron in *Escherichia coli* inserted into a highly conserved reading frame associated with mobile DNA sequences. *Nucleic Acids Res.* 22, 1167–1171.
- Knoop, V., and Rudinger, M. (2010). DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett.* 584, 4287–4291.
- Koonin, E.V. (2009). On the origin of cells and viruses: primordial virus world scenario. *Ann. NY Acad. Sci.* 1178, 47–64.
- Levanon, K., Eisenberg, E., Rechavi, G., et al. (2005). Letter from the editor: adenosine-to-inosine RNA editing in Alu repeats in the human genome. *EMBO Rep.* 6, 831–835.
- Lisch, D. (2008). Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* 60, 43–66.
- Maizels, A., and Weiner, A.M. (1993). The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 577–602.
- Maizels, N., Weiner, A.M., Yue, D., et al. (1999). New evidence for the genomic tag hypothesis: archaeal CCA-adding enzymes and tRNA substrates. *Biol. Bull.* 196, 331–334.
- Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* 65, 3329–3347.
- Mallet, F., Bouton, O., Prudhomme, S., et al. (2004). The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *Proc. Natl. Acad. Sci. USA* 101, 1731–1736.
- Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* 8, 209–220.
- Mattick, J.S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25, 930–939.
- Mattick, J.S. (2009). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann. NY Acad. Sci.* 1178, 29–46.
- Mattick, J., and Gagen, M.J. (2001). The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* 18, 1611–1630.
- Matlin, A.J., and Moore, M.J. (2007). Spliceosome assembly and composition. *Adv. Exp. Med. Biol.* 623, 14–35.



- Moore, P.B., and Steitz, T.A. (2006). The roles of RNA in the synthesis of protein. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 257–285.
- Noller, H.F. (2006). Evolution of ribosomes and translation. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 287–307.
- Obbard, D.J., Gordon, K.H., Buck, A.H., et al. (2009). The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 99–115.
- Ochsenreiter, T., and Hajduk, S. (2008). The function of RNA editing in trypanosomes. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 181–197.
- Odintsova, M.S., and Yurina, N.P. (2000). RNA editing in plant chloroplasts and mitochondria. *Russ. J. Plant Physiol.* 47, 274–284.
- Odintsova, M.S., and Yurina, N.P. (2005). Genomics and evolution of cellular organelles. *Genetica* 41, 1170–1182.
- Panigrahi, A.K., Schnauffer, A., Ernst, N.L., et al. (2003). Identification of novel components of *Trypanosoma brucei* editosomes. *RNA* 9, 484–492.
- Penny, D., Hoepfner, M.R., Poole, A.M., et al. (2009). An overview of the introns-first theory. *J. Mol. Evol.* 69, 527–540.
- Phizicky, E.M. (2005). Have tRNA, will travel. *Proc. Natl. Acad. Sci. USA* 102, 11127–11128.
- Piriapongsa, J., and King Jordan, I. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 14, 814–821.
- Pyle, A.M., and Lambowitz, A.M. (2006). Group II introns: ribozymes that splice RNA and invade DNA. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 469–506.
- Randau, L., Calvin, K., Hall, M., et al. (2005). The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge–helix–bulge motifs of joined tRNA halves. *Proc. Natl. Acad. Sci. USA* 102, 17934–17939.
- Schumann, G.G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1 mediated retrotransposition. *Biochem. Soc. Transact.* 35, 637–642.
- Sciamanna, I., Vitullo, P., Curatolo, A., et al. (2009). Retrotransposons, reverse transcriptase and the genesis of new genetic information. *Gene* 448, 180–186.
- Seetharaman, M., Eldho, N.V., Padgett, R.A., et al. (2006). Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. *RNA* 12, 235–247.
- Shapiro, J.A., and Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 1–24.
- Simpson, L. (2006). Uridine insertion/deletion RNA editing as a paradigm for sitespecific modifications of RNA molecules. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 401–417.
- Siomi, H., and Siomi, M.C. (2009). RISC hitchhikes onto endosome trafficking. *Nat. Cell Biol.* 11, 1049–1051.
- Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285.
- Smith, H.C. (2008). Editing informational content of expressed DNA sequences and their transcripts. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 249–265.
- Sontheimer, E.J., and Carthew, R.W. (2005). Silence from within: endogenous siRNAs and miRNAs. *Cell* 122, 9–12.
- Speijer, D. (2008). Evolutionary aspects of RNA editing. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 199–227.
- Sternberg, R. (2002). On the roles of repetitive DNA elements in the context of a unified genomic-epigenetic system. *Ann. NY Acad. Sci.* 981, 154–188.
- Sternberg, R., and Shapiro, J.A. (2005). How repeated retroelements format genome function. *Cytogenet. Genome Res.* 110, 108–116.
- Stuart, K.D., Schnauffer, A., Ernst, N.L., et al. (2005). Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.* 30, 97–105.
- Sugiura, M. (2008). RNA editing in chloroplasts. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 123–142.
- Taft, R.J., Glazov, E.A., Lassmann, T., et al. (2009). Small RNAs derived from snoRNAs. *RNA* 15, 1233–1240.
- Takenaka, M., Van Der Merwe, J.A., Verbitskiy, D., et al. (2008). RNA editing in plant mitochondria. In: Göringer, H.U. (ed). *RNA Editing*. Berlin: Springer Verlag, 105–122.
- Tang, J., and Breaker, R.R. (2000). Structural diversity of self cleaving ribozymes. *Proc. Natl. Acad. Sci. USA* 97, 5784–5789.
- Toor, N., Hausner, G., and Zimmerly, S. (2001). Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* 7, 1142–1152.
- Turmel, M., Otis, C., and Lemieux, C. (2003). The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *Plant Cell* 15, 1888–1903.
- Tycowski, K.T., Kolev, N.G., Conrad, N.K., et al. (2006). The ever-growing world of small nuclear ribonucleoproteins. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 327–368.
- Valles, Y., Halanych, K.M., and Boore, J.L. (2008). Group II introns break new boundaries: presence in a bilaterian's genome. *PLoS One* 3, e1488.
- Vetsigian, K., Woese, C., and Goldenfeld, N. (2006). Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. USA* 103, 10696–10701.
- Villarreal, L.P. (2005). *Viruses and the Evolution of Life*. Washington: American Society for Microbiology Press.
- Wagner, K.D., Wagner, N., Ghanbarian, H., et al. (2008). RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev. Cell* 14, 962–969.
- Watanabe, T., Totoki, Y., Toyoda, A., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543.
- Weiner, A.M. (2006). SINEs and LINEs: troublemakers, saboteurs, benefactors, ancestors. In: Gesteland, R.F., Cech, T.R., and Atkins, J.F. (eds). *The RNA World*, 3rd edn. New York: Cold Spring Harbor Laboratory Press, 507–534.
- Wegrzyn, G., and Wegrzyn, A. (2008). Is tRNA only a translation factor or also a regulator of other processes? *J. Appl. Genet.* 49, 115–122.
- Witzany, G. (2008). The viral origins of telomeres and telomerases and their important role in eukaryogenesis and genome maintenance. *Biosemiotics* 1, 191–206.
- Witzany, G. (2009). Non-coding RNAs: persistent viral agents as modular tools for cellular needs. *Ann. NY Acad. Sci.* 1178, 244–267.
- Xiong, Y., Li, F., Wang, J., et al. (2003). Crystal structures of an archaeal class I CCA-adding enzyme and its nucleotide complexes. *Mol. Cell* 12, 1165–1172.
- Yazgan, O., and Krebs, J.E. (2007). Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes. *Biochem. Cell Biol.* 85, 484–496.
- Zeh, D.W., Zeh, J.A., and Ishida, Y. (2009). Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays* 31, 715–726.
- Zuckerandl, E. (2002). Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. *Genetica* 115, 105–129.
- Zuckerandl, E., and Cavalli, G. (2007). Combinatorial epigenetics, 'junk DNA', and the evolution of complex organisms. *Gene* 390, 232–242.