Study material

B.Sc. (H) Botany II

Molecular Biology

Unit 6- Processing and modification of RNA

**Exon shuffling**

<https://www.nature.com/articles/nbt0501_423>

Evolution of eukaryotes is mediated by sexual recombination of parental genomes. Crossovers occur in random, but homologous, positions at a frequency that depends on DNA length. As exons occupy only 1% of the human genome and introns about 24%, by far most of the crossovers occur between exons, rather than inside. The natural process of creating new combinations of exons by intronic recombination is called exon shuffling.

<https://www.ncbi.nlm.nih.gov/pubmed/22948334>

Exon shuffling has been characterized as one of the major evolutionary forces shaping both the genome and the proteome of eukaryotes. This mechanism was particularly important in the creation of multidomain proteins during animal evolution, bringing a number of functional genetic novelties.

<https://en.wikipedia.org/wiki/Exon_shuffling>

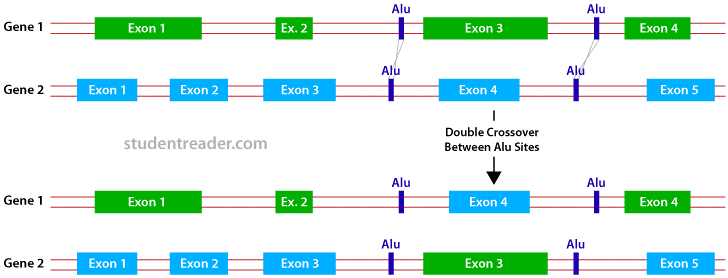
In 1978 Walter Gilbert discovered that the existence of introns could play a major role in the evolution of proteins. It was noted that recombination within introns could help assort exons independently and that repetitive segments in the middle of introns could create hotspots for recombination to shuffle the exonic sequences. However, the presence of these introns in eukaryotes and absence in prokaryotes created a debate about the time in which these introns appeared. Two theories arose: the "introns early" theory and the "introns late" theory. Supporters of the "introns early theory" believed that introns and RNA splicing were the relics of the RNA world and therefore both prokaryotes and eukaryotes had introns in the beginning. However, prokaryotes eliminated their introns in order to obtain a higher efficiency, while eukaryotes retained the introns and the genetic plasticity of the ancestors. On the other hand, supporters of the "introns late" theory believe that prokaryotic genes resemble the ancestral genes and introns were inserted later in the genes of eukaryotes. What is clear now is that the eukaryotic exon-intron structure is not static, introns are continually inserted and removed from genes and the evolution of introns evolves parallel to exon shuffling. In order for exon shuffling to start to play a major role in protein evolution the appearance of spliceosomal introns had to take place. This was due to the fact that the self-splicing introns of the RNA world were unsuitable for exon-shuffling by intronic recombination. These introns had an essential function and therefore could not be recombined. Additionally there is strong evidence that spliceosomal introns evolved fairly recently and are restricted in their evolutionary distribution. Therefore, exon shuffling became a major role in the construction of younger proteins.

Moreover, to define more precisely when exon shuffling became significant in eukaryotes, the evolutionary distribution of modular proteins that evolved through this mechanism were examined in different organisms (i.e., Escherichia coli, Saccharomyces cerevisiae, Arabidopsis thaliana, etc.) These studies suggested that there was an inverse relationship between the genome compactness and the proportion of intronic and repetitive sequences. As well as the fact that exon shuffling became significant after metazoan radiation.

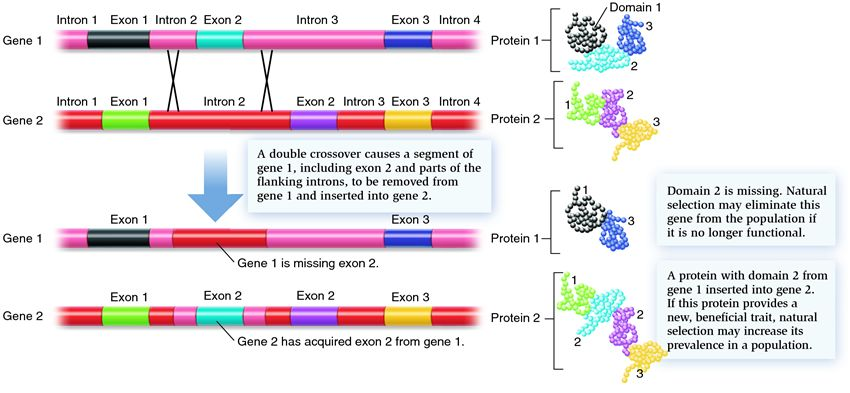
Mechanisms

Crossover during sexual recombination of parental genomes

Evolution of eukaryotes is mediated by sexual recombination of parental genomes and since introns are longer than exons most of the crossovers occur in noncoding regions. In these introns there are large numbers of transposable elements and repeated sequences which promote recombination of nonhomologous genes. In addition it has also been shown that mosaic proteins are composed of mobile domains which have spread to different genes during evolution and which are capable of folding themselves. There is a mechanism for the formation and shuffling of said domains, this is the modularization hypothesis. This mechanism is divided into three stages. The first stage is the insertion of introns at positions that correspond to the boundaries of a protein domain. The second stage is when the "protomodule" undergoes tandem duplications by recombination within the inserted introns. The third stage is when one or more protomodules are transferred to a different nonhomologous gene by intronic recombination. All states of modularization have been observed in different domains such as those of hemostatic proteins.



<https://images.app.goo.gl/rxTpusE7XrynNHtB8>

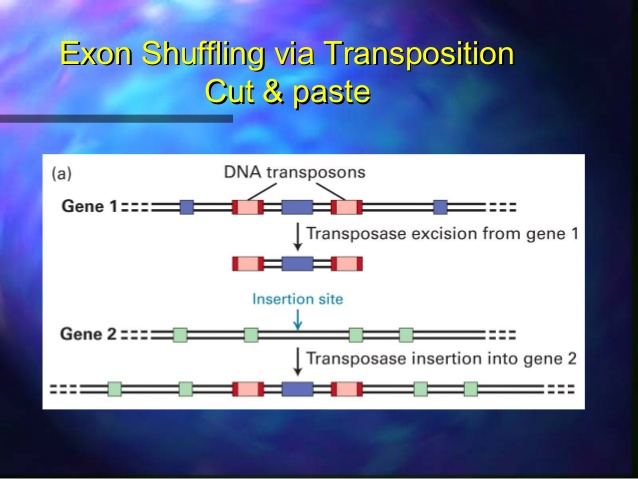


<https://images.app.goo.gl/VxWdVay8zPS4GuVz6>

Transposon mediated

Long interspersed element (LINE)-1

A potential mechanism for exon shuffling is the long interspersed element (LINE) -1 mediated 3' transduction. However it is important first to understand what LINEs are. LINEs are a group of genetic elements that are found in abundant quantities in eukaryotic genomes. LINE-1 is the most common LINE found in humans. It is transcribed by RNA polymerase II to give an mRNA that codes for two proteins: ORF1 and ORF2, which are necessary for transposition. Upon transposition, L1 associates with 3' flanking DNA and carries the non-L1 sequence to a new genomic location. This new location does not have to be in a homologous sequence or in close proximity to the donor DNA sequence. The donor DNA sequence remains unchanged throughout this process because it functions in a copy-paste manner via RNA intermediates; however, only those regions located in the 3' region of the L1 have been proven to be targeted for duplication. Nevertheless, there is reason to believe that this may not hold true, as shown by the following example. The human ATM gene is responsible for the human autosomal-recessive disorder ataxia-telangiectasia and is located on chromosome 11. However, a partial ATM sequence is found in chromosome 7. Molecular features suggest that this duplication was mediated by L1 retrotransposition: the derived sequence was flanked by 15bp target side duplications (TSD), the sequence around the 5' end matched with the consensus sequence for L1 endonuclease cleavage site and a poly(A) tail preceded the 3' TSD. But since the L1 element was present in neither the retrotransposed segment nor the original sequence the mobilization of the segment cannot be explained by 3' transduction. Additional information has led to the belief that trans-mobilization of the DNA sequence is another mechanism of L1 to shuffle exons, but more research on the subject must be done.



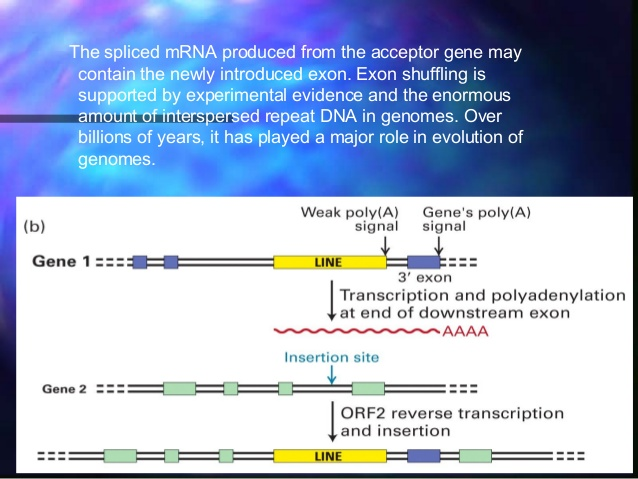
<https://images.app.goo.gl/WLa6WxyqrXE4pjYSA>

Helitron

Another mechanism through which exon shuffling occurs is by the usage of helitrons. Helitron transposons were first discovered during studies of repetitive DNA segments of rice, worm and the thale crest genomes. Helitrons have been identified in all eukaryotic kingdoms, but the number of copies varies from species to species. Helitron encoded proteins are composed of a rolling-circle (RC) replication initiator (Rep) and a DNA helicase (Hel) domain. The Rep domain is involved in the catalytic reactions for endonuclelytic cleavage, DNA transfer and ligation. In addition this domain contains three motifs. The first motif is necessary for DNA binding. The second motif has two histidines and is involved in metal ion binding. Lastly the third motif has two tyrosines and catalyzes DNA cleavage and ligation. There are three models of gene capture by helitrons: the 'read-through" model 1 (RTM1), the 'read-through" model 2 (RTM2) and a filler DNA model (FDNA). According to the RTM1 model an accidental "malfunction" of the replication terminator at the 3' end of the Helitron leads to transposition of genomic DNA. It is composed of the read-through Helitron element and its downstream genomic regions, flanked by a random DNA site, serving as a "de novo" RC terminator. According to the RTM2 model the 3' terminus of another Helitron serves as an RC terminator of transposition. This occurs after a malfunction of the RC terminator. Lastly in the FDNA model portions of genes or non-coding regions can accidentally serve as templates during repair of ds DNA breaks occurring in helitrons. Even though helitrons have been proven to be a very important evolutionary tool, the specific details for their mechanisms of transposition are yet to be defined. An example of evolution by using helitrons is the diversity commonly found in maize. Helitrons in maize cause a constant change of genic and nongenic regions by using transposable elements, leading to diversity among different maize lines.

Long-terminal repeat (LTR) retrotransposons

Long-terminal repeat (LTR) retrotransposons are part of another mechanism through which exon shuffling takes place. They usually encode two open reading frames (ORF). The first ORF named gag is related to viral structural proteins. The second ORF named pol is a polyprotein composed of an aspartic protease (AP)which cleaves the polyprotein, an Rnase H (RH) which splits the DNR-RNA hybrid, a reverse transcriptase (RT) which produces a cDNA copy of the transposons RNA and a DDE integrase which inserts cDNA into the host's genome. Additionally LTR retrotransponsons are classified into five subfamilies: Ty1/copia, Ty3/gypsy, Bel/Pao, retroviruses and endogenous retroviruses. The LTR retrotransponsons require an RNA intermediate in their transposition cycle mechanism. Retrotransponsons synthesize a cDNA copy based on the RNA strand using a reverse transcriptase related to retroviral RT. The cDNA copy is then inserted into new genomic positions to form a retrogene. This mechanism has been proven to be important in gene evolution of rice and other grass species through exon shuffling.



<https://images.app.goo.gl/c2T7uo22EvLKzE6s6>

Illegitimate recombination- Lastly, illegitimate recombination (IR) is another of the mechanisms through which exon shuffling occurs. IR is the recombination between short homologous sequences or nonhomologous sequences. There are two classes of IR: one corresponds to errors of enzymes which cut and join DNA (i.e., DNases.) This process is initiated by a replication protein which helps generate a primer for DNA synthesis. While one DNA strand is being synthesized the other is being displaced. This process ends when the displaced strand is joined by its ends by the same replication protein. The second class of IR corresponds to the recombination of short homologous sequences which are not recognized by the previously mentioned enzymes. However, they can be recognized by non-specific enzymes which introduce cuts between the repeats. The ends are then removed by exonuclease to expose the repeats. Then the repeats anneal and the resulting molecule is repaired using polymerase and ligase.

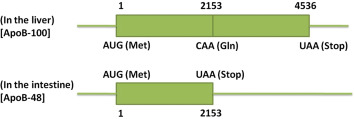
**RNA editing**

<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/rna-editing>

RNA editing is a post-transcriptional process in which nucleotide changes are introduced into a RNA sequence, many of which can thus contribute to proteomic sequence variation.

RNA editing is a process through which the nucleotide sequence specified in the genomic template is modified to produce a different nucleotide sequence in the transcript. RNA editing is an important mechanism of genetic regulation that amplifies genetic plasticity by allowing the production of alternative protein products from a single gene. There are two generic classes of RNA editing in nuclei, involving enzymatic deamination of either C-to-U or A-to-I nucleotides. The best characterized example of C-to-U RNA editing is that of apolipoprotein B (apoB), which is mediated by a holoenzyme that contains a minimal core composed of an RNA-specific cytidine deaminase apobec-1, and its cofactor apobec-1 complementation factor (ACF). C-to-U editing of apoB RNA generates two different isoforms—apoB100 and apoB48—from a single transcript. Both are important regulators of lipid transport and metabolism, and are functionally distinct. C-to-U apoB RNA editing is regulated by a range of factors including developmental, nutritional, environmental, and metabolic stimuli.

RNA editing is a molecular process by which protein-coding gene change its message. One of the most common types of RNA editing is C-to-U conversion. This conversion may occur partially or completely in some tissues but not in others, leading to differential gene expression. Occasionally, it can produce a new protein with a different function from the unedited transcript, foe example, apolipoprotein B gene, one of the lipid carriers in the blood (Cooper, 1999). There are two types – Apo B-100 and apo B-48. Despite differences in length, amino acid sequences of the gigantic protein apo B-100 (4536 amino acid) with that of apo B-48 (2152 amino acid), the result of alignment for the alignable part is 100% identity. It was found that apo B-48 is translated from a very long mRNA that is identical to that of apo B-100 with the execution of an in-frame stop codon resulting from the RNA editing of codon 2153 from CAA (Gln) to UAA (stop). Thus, by using RNA editing, two quite different proteins are produced from the same gene



RNA editing in trypanosomatids occurs by the insertion and deletion of uridine nucleotides (Us), and this editing transforms nonfunctional RNA transcripts into mature mRNAs that can be translated into proteins. Editing occurs by a coordinated series of catalytic steps: cleavage of the mRNA editing site, U addition or removal, and ligation. A multiprotein complex called the editosome catalyzes RNA editing, and uses partially complementary guide RNAs (gRNAs) to direct the process. The sequences of gRNAs specify the sites and numbers of Us that are inserted or deleted. Each gRNA specifies editing at multiple sites, and a single gRNA can specify both U insertion and deletion at different sites. The editosome sediments at ∼20S in glycerol gradients, and contains at least 20 different proteins. Its catalytic activities include endoribonuclease, U‐specific exoribonuclease (exoUase), 3′ terminal uridine transferase (TUTase), RNA ligase, and RNA helicase. Known editosome substrates are transcribed from the mitochondrial maxicircle DNA, and primarily encode proteins of the oxidative phosphorylation pathway. Twelve of the 20 identified Trypanosoma brucei maxicircle gene transcripts undergo posttranscriptional RNA editing. Editing can be extensive, as in the case of COIII mRNA, where 547 uridine nucleotides are inserted and 41 are deleted to generate translatable mRNA. Such extensively edited mRNAs require numerous gRNAs to complete editing.

<https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1574-6976.1999.tb00401.x>

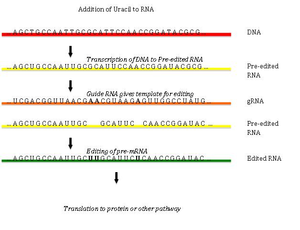
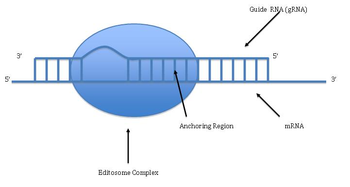
The term RNA editing describes those molecular processes in which the information content is altered in an RNA molecule. To date such changes have been observed in tRNA, rRNA and mRNA molecules of eukaryotes, but not prokaryotes. The demonstration of RNA editing in prokaryotes may only be a matter of time, considering the range of species in which the various RNA editing processes have been found. RNA editing occurs in the nucleus, as well as in mitochondria and plastids, which are thought to have evolved from prokaryotic‐like endosymbionts. Most of the RNA editing processes, however, appear to be evolutionarily recent acquisitions that arose independently. The diversity of RNA editing mechanisms includes nucleoside modifications such as C to U and A to I deaminations, as well as non‐templated nucleotide additions and insertions. RNA editing in mRNAs effectively alters the amino acid sequence of the encoded protein so that it differs from that predicted by the genomic DNA sequence.

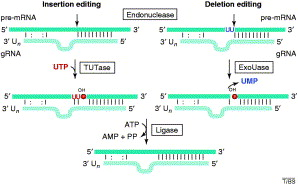
<https://en.wikipedia.org/wiki/RNA_editing>

RNA editing is a molecular process through which some cells can make discrete changes to specific nucleotide sequences within an RNA molecule after it has been generated by RNA polymerase. RNA editing may include the insertion, deletion, and base substitution of nucleotides within the RNA molecule. RNA editing is relatively rare, with common forms of RNA processing (e.g. splicing, 5'-capping, and 3'-polyadenylation) not usually considered as editing. RNA editing has been observed in some tRNA, rRNA, mRNA, or miRNA molecules of eukaryotes and their viruses, archaea, and prokaryotes. RNA editing occurs in the cell nucleus and cytosol, as well as within mitochondria and plastids. In vertebrates, editing is rare and usually consists of a small number of changes to the sequence of the affected molecules. In other organisms, extensive editing (pan-editing) can occur; in some cases the majority of nucleotides in an mRNA sequence may result from editing. RNA-editing processes show great molecular diversity, and some appear to be evolutionarily recent acquisitions that arose independently. The diversity of RNA editing phenomena includes nucleobase modifications such as cytidine (C) to uridine (U) and adenosine (A) to inosine (I) deaminations, as well as non-template nucleotide additions and insertions. RNA editing in mRNAs effectively alters the amino acid sequence of the encoded protein so that it differs from that predicted by the genomic DNA sequence.

Editing by insertion or deletion

RNA editing through the addition and deletion of uracil has been found in kinetoplasts[A kinetoplast is a network of circular DNA (called kDNA) inside a large mitochondrion] from the mitochondria of Trypanosoma brucei. Because this may involve a large fraction of the sites in a gene, it is sometimes called "pan-editing" to distinguish it from topical editing of one or a few sites. Pan-editing starts with the base-pairing of the unedited primary transcript with a guide RNA (gRNA), which contains complementary sequences to the regions around the insertion/deletion points. The newly formed double-stranded region is then enveloped by an editosome, a large multi-protein complex that catalyzes the editing. The editosome opens the transcript at the first mismatched nucleotide and starts inserting uridines. The inserted uridines will base-pair with the guide RNA, and insertion will continue as long as A or G is present in the guide RNA and will stop when a C or U is encountered. The inserted nucleotides cause a frameshift, and result in a translated protein that differs from its gene. The mechanism of the editosome involves an endonucleolytic cut at the mismatch point between the guide RNA and the unedited transcript. The next step is catalyzed by one of the enzymes in the complex, a terminal U-transferase, which adds Us from UTP at the 3' end of the mRNA. The opened ends are held in place by other proteins in the complex. Another enzyme, a U-specific exoribonuclease, removes the unpaired Us. After editing has made mRNA complementary to gRNA, an RNA ligase rejoins the ends of the edited mRNA transcript.[9][10] As a consequence, the editosome can edit only in a 3' to 5' direction along the primary RNA transcript. The complex can act on only a single guide RNA at a time. Therefore, a RNA transcript requiring extensive editing will need more than one guide RNA and editosome complex.

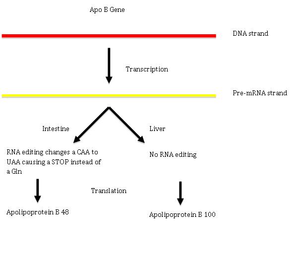
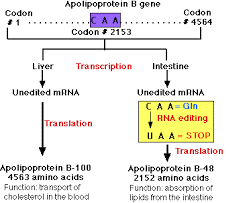


<https://images.app.goo.gl/pi97VdmxtdGQsHyS8>

Editing by deamination

C-to-U editing

The editing involves cytidine deaminase that deaminates a cytidine base into a uridine base. An example of C-to-U editing is with the apolipoprotein B gene in humans. Apo B100 is expressed in the liver and apo B48 is expressed in the intestines. In the intestines, the mRNA has a CAA sequence edited to be UAA, a stop codon, thus producing the shorter B48 form. C-to-U editing often occurs in the mitochondrial RNA of flowering plants. Different plants have different degrees of C-to-U editing; for example, eight (8) editing events occur in mitochondria of the moss Funaria hygrometrica, whereas over 1,700 editing events occur in the lycophytes Isoetes engelmanii. C-to-U editing is performed by members of the pentatricopeptide repeat (PPR) protein family. Angiosperms have large PPR families, acting as trans -factors for cis -elements lacking a consensus sequence; Arabidopsis has around 450 members in its PPR family. There have been a number of discoveries of PPR proteins in both plastids and mitochondria.

<https://images.app.goo.gl/PbtKGh8XK5wpfXAY7>

A-to-I editing

Adenosine-to-inosine (A-to-I) modifications contribute to nearly 90% of all editing events in RNA. The deamination of adenosine is catalyzed by the double-stranded RNA-specific adenosine deaminase (ADAR), which typically acts on pre-mRNAs. The deamination of adenosine to inosine disrupts and destabilizes the dsRNA base pairing, therefore rendering that particular dsRNA less able to produce siRNA, which interferes with the RNAi pathway. The wobble base pairing causes deaminated RNA to have a unique but different structure, which may be related to the inhibition of the initiation step of RNA translation. Studies have shown that I-RNA (RNA with many repeats of the I-U base pair) recruits methylases that are involved in the formation of heterochromatin and that this chemical modification heavily interferes with miRNA target sites. There is active research into the importance of A-to-I modifications and their purpose in the novel concept of epitranscriptomics, in which modifications are made to RNA that alter their function. A long established consequence of A-to-I in mRNA is the interpretation of I as a G, therefore leading to functional A-to-G substitution, e.g. in the interpretation of the genetic code by ribosomes. Newer studies however, have weakened this correlation by showing that I's can also be decoded by the ribosome (although in a lesser extent) as A's and U's. Furthermore it was shown that I's lead to the stalling of ribosomes on the I-rich mRNA. The development of high-throughput sequencing in recent years has allowed for the development of extensive databases for different modifications and edits of RNA. RADAR (Rigorously Annotated Database of A-to-I RNA editing) was developed in 2013 to catalog the vast variety of A-to-I sites and tissue-specific levels present in humans, mice, and flies. The addition of novel sites and overall edits to the database are ongoing. The level of editing for specific editing sites, e.g. in the filamin A transcript, is tissue-specific. The efficiency of mRNA-splicing is a major factor controlling the level of A-to-I RNA editing.

Alternative mRNA editing

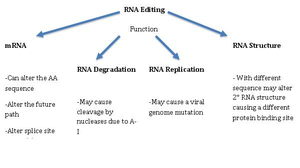
Alternative U-to-C mRNA editing was first reported in WT1 (Wilms Tumor-1) transcripts, and non-classic G-A mRNA changes were observed in HNRNPK (heterogeneous nuclear ribonucleoprotein K) transcripts in both malignant and normal colorectal samples. The latter changes were also later seen alongside non-classic U-to-C alterations in brain cell TPH2 (tryptophan hydroxylase 2) transcripts. Although the reverse amination might be the simplest explanation for U-to-C changes, transamination and transglycosylation mechanisms have been proposed for plant U-to-C editing events in mitochondrial transcripts. A recent study reported novel G-to-A mRNA changes in WT1 transcripts at two hotspots, proposing the APOBEC3A (apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3A) as the enzyme implicated in this class of alternative mRNA editing. It was also shown that alternative mRNA changes were associated with canonical WT1 splicing variants, indicating their functional significance.

RNA editing in plant mitochondria and plastids

It has been shown in previous studies that the only types of RNA editing seen in the plants' mitochondria and plastids are conversion of C-to-U and U-to-C (very rare). RNA-editing sites are found mainly in the coding regions of mRNA, introns, and other non-translated regions. In fact, RNA editing can restore the functionality of tRNA molecules. The editing sites are found primarily upstream of mitochondrial or plastid RNAs. While the specific positions for C to U RNA editing events have been fairly well studied in both the mitochondrion and plastid,[39] the identity and organization of all proteins comprising the editosome have yet to be established. Members of the expansive PPR protein family have been shown to function as trans-acting factors for RNA sequence recognition. Specific members of the MORF (Multiple Organellar RNA editing Factor) family are also required for proper editing at several sites. As some of these MORF proteins have been shown to interact with members of the PPR family, it is possible MORF proteins are components of the editosome complex. An enzyme responsible for the trans- or deamination of the RNA transcript remains elusive, though it has been proposed that the PPR proteins may serve this function as well. RNA editing is essential for the normal functioning of the plant's translation and respiration activity. Editing can restore the essential base-pairing sequences of tRNAs, restoring functionality. It has also been linked to the production of RNA-edited proteins that are incorporated into the polypeptide complexes of the respiration pathway. Therefore, it is highly probable that polypeptides synthesized from unedited RNAs would not function properly and hinder the activity of both mitochondria and plastids. C-to-U RNA editing can create start and stop codons, but it cannot destroy existing start and stop codons. A cryptic start codon is created when the codon ACG is edited to be AUG.

RNA editing in viruses

RNA editing in viruses (i.e., measles, mumps, or parainfluenza) is used for stability and generation of protein variants. Viral RNAs are transcribed by a virus-encoded RNA-dependent RNA polymerase, which is prone to pausing and "stuttering" at certain nucleotide combinations. In addition, up to several hundred non-templated A's are added by the polymerase at the 3' end of nascent mRNA. These As help stabilize the mRNA. Furthermore, the pausing and stuttering of the RNA polymerase allows the incorporation of one or two Gs or As upstream of the translational codon. The addition of the non-templated nucleotides shifts the reading frame, which generates a different protein.



Origin and evolution of RNA editing

The RNA-editing system seen in the animal may have evolved from mononucleotide deaminases, which have led to larger gene families that include the apobec-1 and adar genes. These genes share close identity with the bacterial deaminases involved in nucleotide metabolism. The adenosine deaminase of E. coli cannot deaminate a nucleoside in the RNA; the enzyme's reaction pocket is too small for the RNA strand to bind to. However, this active site is widened by amino acid changes in the corresponding human analog genes, APOBEC1 and ADAR, allowing deamination. The gRNA-mediated pan-editing in trypanosome mitochondria, involving templated insertion of U residues, is an entirely different biochemical reaction. The enzymes involved have been shown in other studies to be recruited and adapted from different sources. But the specificity of nucleotide insertion via the interaction between the gRNA and mRNA is similar to the tRNA editing processes in the animal and Acanthamoeba mithochondria. Eukaryotic ribose methylation of rRNAs by guide RNA molecules is a similar form of modification. Thus, RNA editing evolved more than once. Several adaptive rationales for editing have been suggested. Editing is often described as a mechanism of correction or repair to compensate for defects in gene sequences. However, in the case of gRNA-mediated editing, this explanation does not seem possible because if a defect happens first, there is no way to generate an error-free gRNA-encoding region, which presumably arises by duplication of the original gene region. This thinking leads to an evolutionary proposal called "constructive neutral evolution" in which the order of steps is reversed, with the gratuitous capacity for editing preceding the "defect".

RNA editing may be involved in RNA degradation

A study looked at the involvement of RNA editing in RNA degradation. The researchers specifically looked at the interaction between ADAR and UPF1, an enzyme involved in the nonsense-mediated mRNA decay pathway (NMD). They found that ADAR and UPF1 are found within the suprasliceosome and they form a complex that leads to the down-regulation of specific genes. The exact mechanism or the exact pathways that these two are involved in are unknown at this time. The only fact that this research has shown is that they form a complex and down-regulate specific genes.

mRNA transport

<https://www.nature.com/articles/nrm2255>

The transport of RNA molecules from the nucleus to the cytoplasm is fundamental for gene expression. The different RNA species that are produced in the nucleus are exported through the nuclear pore complexes via mobile export receptors. Small RNAs (such as tRNAs and microRNAs) follow relatively simple export routes by binding directly to export receptors. Large RNAs (such as ribosomal RNAs and mRNAs) assemble into complicated ribonucleoprotein (RNP) particles and recruit their exporters via class-specific adaptor proteins. Export of mRNAs is unique as it is extensively coupled to transcription (in yeast) and splicing (in metazoa).

<https://www.pnas.org/content/102/47/17008.long>

After mRNAs are synthesized, processed, and become associated with a number of different proteins at the transcription site, they are released into the nucleoplasm. The mechanism by which these large mRNA-protein (mRNP) complexes then move through dense nucleoplasm to reach the nuclear pores has been the subject of intense study and speculation. Early workers proposed that mRNP complexes are transferred along a chain of receptors until they reach a nuclear pore, expending metabolic energy in the process. This solid-state transport model is supported by observations made in fixed nuclei that show some transcripts distributed along tracks that originate from the locus of the parent gene. A second theory, called the “gene-gating” hypothesis, proposes that active genes are situated near the nuclear periphery and that mRNAs exit the nucleus through the nearest pores. This idea is supported by observations that certain mRNAs exit from one side of the nucleus and that, in yeast, many transcriptionally active gene loci are located near the nuclear periphery. By contrast, a number of other studies have found that mRNP complexes move quite freely within the nucleus. This view is supported by studies of the distribution of newly synthesized Balbiani ring RNA in the salivary gland cells of insects, fluorescence recovery after photobleaching and fluorescence correlation spectroscopy studies of probes that bind to the poly(A) tails of mRNAs, and from single-particle analysis of mRNP complexes bound to GFP-linked proteins.

<https://www.sciencedirect.com/science/article/abs/pii/S0955067406000573?via%3Dihub>

All movement of molecules and macromolecules between the cytoplasm and the nucleus takes place through nuclear pore complexes (NPCs), very large macromolecular complexes that are the only channels connecting these compartments. mRNA export is mediated by multiple, highly conserved protein factors that couple steps of nuclear pre-mRNA biogenesis to mRNA transport. Mature messenger ribonucleoproteins (mRNPs) diffuse from sites of transcription to NPCs, although some active genes are positioned at the nuclear periphery where they interact physically with components of NPCs. As properly processed mRNPs translocate through the pore, certain mRNP proteins are removed.

<https://sciencing.com/mrna-leave-nucleus-10050146.html>

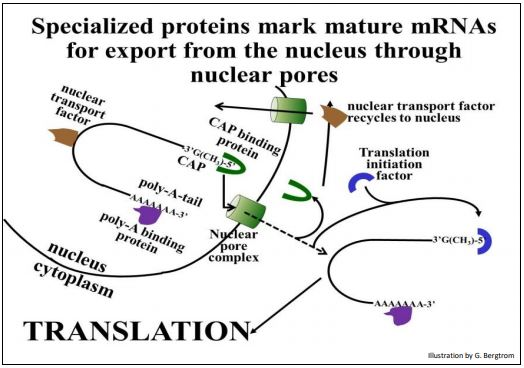
After mRNA molecules are synthesized at the transcription site, they must make their journey to the sites of translation, the ribosomes. Ribosomes appear both free in the cell cytoplasm and attached to a membranous organelle called the endoplasmic reticulum, both of which lie outside the nucleus. Before the mRNA can pass through the double plasma membrane that makes up the nuclear envelope (or nuclear membrane), it must reach the membrane somehow. This occurs by the binding of the new mRNA molecules to transport proteins. Before the resulting mRNA-protein (mRNP) complexes can move to the edge, they become thoroughly mixed inside the substance of the nucleus, so that those mRNP complexes that happen to form near the edge of the nucleus have no better a chance at exiting the nucleus at a given time after formation than do mRNP processes close to the interior. When mRNP complexes encounter regions of the nucleus heavy in DNA, which in this environment exists as chromatin (i.e., DNA bound to structural proteins), it can become stalled, just like a pickup truck being bogged down in heavy mud. This stalling can be overcome by the input of energy in the form of ATP, which prods the bogged-down mRNP in the direction of the edge of the nucleus.

The nucleus needs to protect the all-important genetic material of the cell, yet it also must have a means of exchanging proteins and nucleic acids with the cell cytoplasm. This is accomplished via "gates" consisting of proteins and known as nuclear pore complexes (NPC). These complexes have a pore running through the double membrane of the nuclear envelope and a number of different structures on either side of this "gate." The NPC is enormous by molecular standards. In human beings, it has a molecular mass of 125 million Daltons. In contrast, a molecule of glucose has a molecular mass of 180 Daltons, making it about 700,000 times smaller than the NPC complex. Both nucleic acid and protein transport into the nucleus and the movement of these molecules out of the nucleus occur via the NPC. On the cytoplasmic side, the NPC has what is called a cytoplasmic ring as well as cytoplasmic filaments, both of which serve to help anchor the NPC in place in the nuclear membrane. On the nuclear side of the NPC is a nuclear ring, analogous to the cytoplasmic ring on the opposite side, as well as a nuclear basket.

A variety of individual proteins participate in the movement of mRNA and a diverse variety of other molecular cargoes out of the nucleus, with the same applying to movement of substances into the nucleus.

<https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Basic_Cell_and_Molecular_Biology_(Bergtrom)/10%3A_Transcription_and_RNA_Processing/10.7%3A_RNA_and_Ribosome_Export_from_the_Nucleus>

The 5’ methyl guanosine cap and the poly(A) tail collaborate to facilitate exit of mRNAs from the nucleus into the cytoplasm. We now understand that proteins in the nucleus participate in the export process. A nuclear transport receptor binds along the mature (or maturing) mRNA, a poly-A-binding protein binds along the poly-A tail of the message, and another protein binds at or near the methyl guanosine CAP itself. These interactions enable transport of the mRNA through nuclear pores. After the mRNA is in the cytoplasm, the nuclear transport receptor re-cycles back into the nucleus while a translation initiation factor replaces the protein bound to the CAP. The nuclear transport process is summarized in the illustration below.



<https://images.app.goo.gl/HRsGZsRkaC7odS22A>

The mature mRNA, now in the cytoplasm, is ready for translation. Translation is the process of protein synthesis mediated by ribosomes and a host of translation factors (including the initiation factor in the illustration above.

Ribozymes

<https://en.wikipedia.org/wiki/Ribozyme>

Ribozymes (ribonucleic acid enzymes) are RNA molecules that have the ability to catalyze specific biochemical reactions, including RNA splicing in gene expression, similar to the action of protein enzymes. The 1982 discovery of ribozymes demonstrated that RNA can be both genetic material (like DNA) and a biological catalyst (like protein enzymes), and contributed to the RNA world hypothesis, which suggests that RNA may have been important in the evolution of prebiotic self-replicating systems. The most common activities of natural or in vitro-evolved ribozymes are the cleavage or ligation of RNA and DNA and peptide bond formation. Within the ribosome, ribozymes function as part of the large subunit ribosomal RNA to link amino acids during protein synthesis. They also participate in a variety of RNA processing reactions, including RNA splicing, viral replication, and transfer RNA biosynthesis. Examples of ribozymes include the hammerhead ribozyme, the VS ribozyme, Leadzyme and the hairpin ribozyme.

Before the discovery of ribozymes, enzymes, which are defined as catalytic proteins, were the only known biological catalysts. In 1967, Carl Woese, Francis Crick, and Leslie Orgel were the earliest to suggest that RNA could act as a catalyst. This idea was based upon the discovery that RNA can form complex secondary structures. These ribozymes were found in the intron of an RNA transcript, which removed itself from the transcript, as well as in the RNA component of the RNase P complex, which is involved in the maturation of pre-tRNAs. In 1989, T.R. Cech and Sidney Altman shared the Nobel Prize in chemistry for their "discovery of catalytic properties of RNA." The term ribozyme was introduced by Kelly Kruger et al. in 1982 in a paper published in Cell. It had been a firmly established belief in biology that catalysis was reserved for proteins. However, the idea of RNA catalysis is motivated in part by the old question regarding the origin of life: Which comes before, enzymes that do the work of the cell or nucleic acids that carry the information required to produce the enzymes? The concept of "ribonucleic acids as catalysts" circumvents this problem.

In the 1980s T. Cech, at the University of Colorado at Boulder, was studying the excision of introns in a ribosomal RNA gene in *Tetrahymena thermophila*. While trying to purify the enzyme responsible for the splicing reaction, he found that the intron could be spliced out in the absence of any added cell extract. As much as they tried, Cech and his colleagues could not identify any protein associated with the splicing reaction. After much work, Cech proposed that the intron sequence portion of the RNA could break and reform phosphodiester bonds. At about the same, Sidney Altman, a professor at Yale University, was studying the way tRNA molecules are processed in the cell when he and his colleagues isolated an enzyme called RNase-P, which is responsible for conversion of a precursor tRNA into the active tRNA. Much to their surprise, they found that RNase-P contained RNA in addition to protein and that RNA was an essential component of the active enzyme. This was such a foreign idea that they had difficulty publishing their findings. The following year, Altman demonstrated that RNA can act as a catalyst by showing that the RNase-P RNA subunit could catalyze the cleavage of precursor tRNA into active tRNA in the absence of any protein component.

Since Cech's and Altman's discovery, other investigators have discovered other examples of self-cleaving RNA or catalytic RNA molecules. Many ribozymes have either a hairpin – or hammerhead – shaped active center and a unique secondary structure that allows them to cleave other RNA molecules at specific sequences. It is now possible to make ribozymes that will specifically cleave any RNA molecule. These RNA catalysts may have pharmaceutical applications. For example, a ribozyme has been designed to cleave the RNA of HIV. If such a ribozyme were made by a cell, all incoming virus particles would have their RNA genome cleaved by the ribozyme, which would prevent infection.

Structure and mechanism

Despite having only four choices for each monomer unit (nucleotides), compared to 20 amino acid side chains found in proteins, ribozymes have diverse structures and mechanisms. In many cases they are able to mimic the mechanism used by their protein counterparts. For example, in self cleaving ribozyme RNAs, an in-line SN2 reaction is carried out using the 2’ hydroxyl group as a nucleophile attacking the bridging phosphate and causing 5’ oxygen of the N+1 base to act as a leaving group . In comparison, RNase A, a protein that catalyzes the same reaction, uses a coordinating histidine and lysine to act as a base to attack the phosphate backbone.

As protein enzymes metal binding is also critical to the function of many ribozymes. Often these interactions use both the phosphate backbone and the base of the nucleotide, causing drastic conformational changes. There are two mechanism classes for the cleavage of phosphodiester backbone in the presence of metal. In the first mechanism, the internal 2’- OH group attacks phosphorus center in a SN2 mechanism. Metal ions promote this reaction by first coordinating the phosphate oxygen and later stabling the oxyanion. The second mechanism also follows a SN2 displacement, but the nucleophile comes from water or exogenous hydroxyl groups rather than RNA itself. The smallest ribozyme is UUU, which can promote the cleavage between G and A of the GAAA tetranucleotide via the first mechanism in the presence of Mn2+. The reason why this trinucleotide rather than the complementary tetramer catalyze this reaction may be because the UUU-AAA pairing is the weakest and most flexible trinucleotides among the 64 conformations, which provides the binding site for Mn2+.

Phosphoryl transfer can also be catalyzed without metal ions. For example, pancreatic ribonuclease A and hepatitis delta virus(HDV) ribozymes can catalyze the cleavage of RNA backbone through acid-base catalysis without metal ions. Hairpin ribozyme can also catalyze the self-cleavage of RNA without metal ions but the mechanism is still unclear.

Ribozyme can also catalyze the formation of peptide bond between adjacent amino acid by lowering the activation entropy.

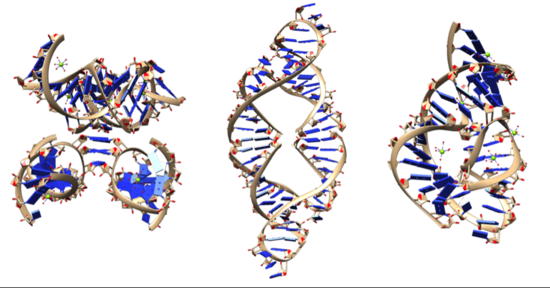


Image showing the diversity of ribozyme structures. From left to right: leadzyme, hammerhead ribozyme, twister ribozyme

Although ribozymes are quite rare in most cells, their roles are sometimes essential to life. For example, the functional part of the ribosome, the biological machine that translates RNA into proteins, is fundamentally a ribozyme, composed of RNA tertiary structural motifs that are often coordinated to metal ions such as Mg2+ as cofactors. In a model system, there is no requirement for divalent cations in a five-nucleotide RNA catalyzing trans-phenylalanation of a four-nucleotide substrate with 3 base pairs complementary with the catalyst, where the catalyst/substrate were devised by truncation of the C3 ribozyme. The best-studied ribozymes are probably those that cut themselves or other RNAs, as in the original discovery by Cech and Altman. However, ribozymes can be designed to catalyze a range of reactions, many of which may occur in life but have not been discovered in cells. RNA may catalyze folding of the pathological protein conformation of a prion in a manner similar to that of a chaperonin.

Ribozymes and the origin of life

RNA can also act as a hereditary molecule, which encouraged Walter Gilbert to propose that in the distant past, the cell used RNA as both the genetic material and the structural and catalytic molecule rather than dividing these functions between DNA and protein as they are today; this hypothesis is known as the "RNA world hypothesis" of the origin of life. Since nucleotides and RNA and thus ribozymes can arise by inorganic chemicals, these are candidates for the earliest enzymes, and "replicators", i.e. information-containing macro-molecules that replicate themselves. An example of a self-replicating ribozyme that ligates two substrates to generate an exact copy of itself was described in 2002.

Applications

Ribozymes have been proposed and developed for the treatment of disease through gene therapy. One major challenge of using RNA based enzymes as a therapeutic is the short half-life of the catalytic RNA molecules in the body. To combat this, the 2’ position on the ribose is modified to improve RNA stability. One area of ribozyme gene therapy has been the inhibition of RNA-based viruses. A type of synthetic ribozyme directed against HIV RNA called gene shears has been developed and has entered clinical testing for HIV infection. Similarly, ribozymes have been designed to target the hepatitis C virus RNA, SARS coronavirus (SARS-CoV), Adenovirus and influenza A and B virus RNA. The ribozyme is able to cleave the conserved regions of the virus’s genome which has been shown to reduce the virus in mammalian cell culture. Despite these efforts by researchers, these projects have remained in the preclinical stage.

Known ribozymes

Well validated naturally occurring ribozyme classes:

GIR1 branching ribozyme

glmS ribozyme

Group I self-splicing intron

Group II self-splicing intron - Spliceosome is likely derived from Group II self-splicing ribozymes.

Hairpin ribozyme

Hammerhead ribozyme

HDV ribozyme

rRNA - Found in all living cells and links amino acids to form proteins.

RNase P

Twister ribozyme

Twister sister ribozyme

VS ribozyme

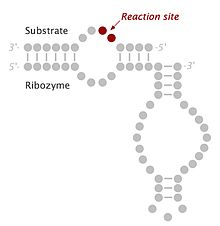
Pistol ribozyme

Hatchet ribozyme

Viroids

<https://en.wikipedia.org/wiki/Hairpin_ribozyme>

The hairpin ribozyme is a small section of RNA that can act as a ribozyme. Like the hammerhead ribozyme it is found in RNA satellites of plant viruses. It was first identified in the minus strand of the tobacco ringspot virus (TRSV) satellite RNA where it catalyzes self-cleavage and joining (ligation) reactions to process the products of rolling circle virus replication into linear and circular satellite RNA molecules. The hairpin ribozyme is similar to the hammerhead ribozyme in that it does not require a metal ion for the reaction.

 Secondary structure of a minimal hairpin ribozyme with substrate RNA bound. Circles represent individual nucleotides and lines indicate canonical (Watson-Crick) base pairs

The minimal hairpin ribozyme-substrate complex folds into a secondary structure that includes two domains, each consisting of two short base paired helices separated by an internal loop. Domain A (helix 1 – loop A – helix 2) contains the substrate and the primary substrate-recognition region of the ribozyme. Domain B (helix 3 – loop B – helix 4) is larger and contains the primary catalytic determinants of the ribozyme. The two domains are covalently joined via a phosphodiester linkage that connects helix 2 to helix 3. These domains must interact with one another in order for catalysis to occur. When the minimal ribozyme-substrate complex is allowed to fold under conditions of low ionic strength, the two domains stack one atop the other, forming an inactive, extended structure that resembles a hairpin. In order for catalysis to occur, the two domains lie parallel to one another in a fold that resembles a paperclip. In various publications, this RNA has been termed either the "paperclip" or "hairpin" ribozyme. Despite the fact that the former name has proven to be more accurate, the latter has become the commonly accepted nomenclature. In the laboratory, a functional interaction between the two domains is promoted by the addition of cations, whose positive charge suffices to overcome the electrostatic repulsion of the negatively charged RNA backbone. In nature, association of the two domains is assisted through a combination of metal ions (including Mg2+) and the presence of two additional helical domains that are not present in the minimal ribozyme-substrate complex but serve to promote proper three-dimensional folding. These additional domains stack upon helices 2 and 3, thereby promoting the association of the two functional domains through what is termed a four-way helical junction.

The hairpin ribozyme is an RNA motif that catalyzes RNA processing reactions essential for replication of the satellite RNA molecules in which it is embedded. These reactions are self-processing, i.e. a molecule rearranging its own structure. Both cleavage and end joining reactions are mediated by the ribozyme motif, leading to a mixture of interconvertible linear and circular satellite RNA molecules. These reactions are important for processing the large multimeric RNA molecules that are generated by rolling circle replication. At the end of the replication cycle, these large intermediates of satellite RNA replication are processed down to unit length molecules (circular or linear) before they can be packaged by viruses and carried to other cells for further rounds of replication.

The hairpin ribozyme has been identified in only 3 naturally occurring sequences:

Satellite RNA of tobacco ringspot virus (sTRSV), satellite RNA of chicory yellow mottle virus (sCYMV), satellite RNA of arabis mosaic virus (sARMV).

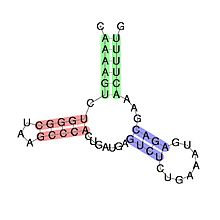
In common with several other ribozymes and protein ribonucleases, the cleavage reaction of the hairpin ribozyme generates RNA fragments with termini consisting of a 2',3'-cyclic phosphate and a 5'-hydroxyl group. The ligation reaction appears to be a simple reversal of cleavage, i.e. covalent joining of RNA fragments ending with a 2',3'-cyclic phosphate and a 5'-hydroxyl group to generate the ordinary 3'-5' phosphodiester linkage used in both RNA and DNA. Studies of this reaction in multiple ribozymes have served to establish that the reaction chemistry (catalytic mechanism) is an endogenous property of the RNA molecule itself and is not mediated by metal ions, as is true for some protein enzymes and some other ribozymes. Moreover, cleavage activity is still observed when Mg2+ is replaced by [Co(NH3)6]3+. Co3+ binds NH3 so tightly in solution that NH3 does not dissociate to any appreciable extent, and therefore does not become protonated. This suggests there is no metal-catalyzed proton transfer or direct coordination to the RNA, but instead metals are only required for folding. Furthermore, in crystal structures of a ribozyme-inhibitor complex and a transition state mimic, it was shown that the three-dimensional architecture splays apart A-1 and G+1, positioning the 2'-OH of A-1 for an in-line nucleophilic attack on the scissile phosphate linkage. Additionally, G8, A38, and A9 have been suggested to play roles in the catalysis by deprotonating the 2'-OH of A-1, stabilizing the developing negative charge of the pentacoordinate phosphate oxygens, and protonating the 5'-O leaving group of G+1.

<https://en.wikipedia.org/wiki/Hammerhead_ribozyme>

The hammerhead ribozyme is an RNA motif that catalyzes reversible cleavage and ligation reactions at a specific site within an RNA molecule. It is one of several catalytic RNAs (ribozymes) known to occur in nature. It serves as a model system for research on the structure and properties of RNA, and is used for targeted RNA cleavage experiments, some with proposed therapeutic applications. Named for the resemblance of early secondary structure diagrams to a hammerhead shark, hammerhead ribozymes were originally discovered in two classes of plant virus-like RNAs: satellite RNAs and viroids. Hammerhead ribozymes were found in RNA plant pathogens like viroids and viral satellites. A hammerhead ribozyme was also reported in the satellite DNA of newt genomes. New examples of this ribozyme were then found in the genomes of unrelated organisms like schistosomes, cave crickets, Arabidopsis thaliana and a few mammals like rodents and the platypus. It was found that the hammerhead ribozyme occurs in a wide variety of bacterial and eukaryal genomes, and even in humans. Similar reports confirmed and extended these observations, unveiling the hammerhead ribozyme as a ubiquitous catalytic RNA in all life kingdoms. In eukaryotic genomes, many of the detected hammerhead ribozymes seem to be related to short interspersed retroelements (SINEs), with the exception of a family of strikingly conserved hammerheads found in the genomes of all amniotes. These hammerhead ribozymes (the so-called HH9 and HH10) occur in the introns of a few specific genes and point to a preserved biological role during pre-mRNA biosynthesis.

The hammerhead ribozyme carries out a very simple chemical reaction that results in the breakage of the substrate strand of RNA, specifically at C17, the cleavage-site nucleotide. Although RNA cleavage is often referred to as hydrolysis, the mechanism employed does not in fact involve the addition of water. Rather, the cleavage reaction is simply an isomerization that consists of rearrangement of the linking phosphodiester bond. It is the same reaction, chemically, that occurs with random base-mediated RNA degradation, except that it is highly site-specific and the rate is accelerated 10,000-fold or more.

The minimal hammerhead ribozyme is composed of three base paired helices, separated by short linkers of conserved sequence as shown in the crystal structure. These helices are called I, II and III. The conserved uridine-turn links helix I to helix II and usually contains the sequence CUGA. Helix II and III are linked by a sequence GAAA. The cleavage reaction occurs between helix III and I, and is usually a C. The structure of a full length ribozyme shows that there are extensive interactions between the loop of stem II and stem I.

 secondary structure and sequence conservation of Hammerhead ribozyme (type III)

Cleavage by phosphodiester isomerization

The cleavage reaction is a phosphodiester isomerization reaction that is initiated by abstraction of the cleavage-site ribose 2’-hydroxyl proton from the 2’-oxygen, which then becomes the attacking nucleophile in an “in-line” or SN2(P)-like reaction, although it is not known whether this proton is removed prior to or during the chemical step of the hammerhead cleavage reaction. (The cleavage reaction is technically not bimolecular, but behaves in the same way a genuine SN2(P) reaction does; it undergoes inversion of configuration subsequent to forming an associative transition-state consisting of a pentacoordinated oxyphosphrane.) The attacking and leaving group oxygens will both occupy the two axial positions in the trigonal bipyramidal transition-state structure as is required for an SN2-like reaction mechanism.

The 5’-product, as a result of this cleavage reaction mechanism, possesses a 2’,3’-cyclic phosphate terminus, and the 3’-product possesses a 5’-OH terminus, as with nonenzymatic alkaline cleavage of RNA. The reaction is therefore reversible, as the scissile phosphate remains a phosphodiester, and may thus act as a substrate for hammerhead RNA-mediated ligation without a requirement for ATP or a similar exogenous energy source. The hammerhead ribozyme-catalyzed reaction, unlike the formally identical non-enzymatic alkaline cleavage of RNA, is a highly sequence-specific cleavage reaction with a typical turnover rate of approximately 1 molecule of substrate per molecule of enzyme per minute at pH 7.5 in 10 mM Mg2+ (so-called “standard reaction conditions” for the minimal hammerhead RNA sequence), depending upon the sequence of the particular hammerhead ribozyme construct measured. This represents an approximately 10,000-fold rate enhancement over the nonezymatic cleavage of RNA.

Requirement for divalent metal ions

All ribozymes were originally thought to be metallo-enzymes. It was assumed that divalent metal ions like Mg2+ were thought to have two roles: Promote proper folding of RNA and to form the catalytic core.[20] Since RNA itself did not contain enough variation in the functional groups, metal ions were thought to play a role at the active site, as was known about proteins. The proposed mechanism for the Mg2+ ion was: the deprotonation of the 2'-OH group by a Magnesium.aqua.hydroxy complex bound by the pro-R oxygen at the phosphate-cleavage site, followed by nucleophilic attack of the resultant 2'-alkaoxide on the scissile phosphate forming a pentacoordinate phosphate intermediate. The last step is the departing of the 5' leaving group, yielding a 2',3'-cyclic phosphate with an inverted configuration. It was presumed that hexahydrated magnesium ions, which exist in equilibrium with magnesium hydroxide, could play the roles of general acid and general base, in a way analogous to those played by two histidines in RNase A. An additional role for divalent metal ions has also been proposed in the form of electrostatic stabilization of the transition-state.

Not a metallo-enzyme

In 1998 it was observed that the hammerhead ribozyme, and hairpin ribozyme, do not require the presence of metal ions for catalysis, provided a sufficiently high concentration of monovalent cation is present to permit the RNA to fold. This suggested that the RNA itself, rather than serving as an inert, passive scaffold for the binding of chemically active divalent metal ions, is instead itself intimately involved in the chemistry of catalysis. The latest structural results, described below, indeed confirm that two invariant nucleotides, G12 and G8, are positioned consistent with roles as the general base and general acid in the hammerhead cleavage reaction.