CATALYTIC RNA

5.26.1 Introduction

Key Terms

A **ribozyme** is an RNA that has catalytic activity.

RNA editing describes a change of sequence at the level of RNA following transcription.

The idea that only proteins have enzymatic activity was deeply rooted in biochemistry. (Yet devotées of protein function once thought that only proteins could have the versatility to be the genetic material!) A rationale for the identification of enzymes with proteins lies in the view that only proteins, with their varied three-dimensional structures and variety of side-groups, have the flexibility to create the active sites that catalyze biochemical reactions. But the characterization of systems involved in RNA processing has shown this view to be an over-simplification.

Several types of catalytic reactions are now known to reside in RNA. **Ribozyme** has become a general term used to describe an RNA with catalytic activity, and it is possible to characterize the enzymatic activity in the same way as a more conventional enzyme. Some RNA catalytic activities are directed against separate substrates, while others are intramolecular (which limits the catalytic action to a single cycle).

Introns of the group I and group II classes possess the ability to splice themselves out of the pre-mRNA that contains them. Engineering of group I introns has generated RNA molecules that have several other catalytic activities related to the original activity.

The enzyme ribonuclease P is a ribonucleoprotein that contains a single RNA molecule bound to a protein. The RNA possesses the ability to catalyze cleavage in a tRNA substrate, while the protein component plays an indirect role, probably to maintain the structure of the catalytic RNA.

The common theme of these reactions is that the RNA can perform an intramolecular or intermolecular reaction that involves cleavage or joining of phosphodiester bonds *in vitro.* Although the specificity of the reaction and the basic catalytic activity is provided by RNA, proteins associated with the RNA may be needed for the reaction to occur efficiently *in vivo.*

RNA splicing is not the only means by which changes can be introduced in the informational content of RNA. In the process of **RNA editing**, changes are introduced at individual bases, or bases are added at particular positions within an mRNA. The insertion of bases (most commonly uridine residues) occurs for several genes in the mitochondria of certain lower eukaryotes; like splicing, it involves the breakage and reunion of bonds between nucleotides, but also requires a template for

coding the information of the new sequence.

CATALYTIC RNA

5.26.2 Group I introns undertake self-splicing by transesterification

Key Terms

Autosplicing (Self-splicing) describes the ability of an intron to excise itself from an RNA by a catalytic action that depends only on the sequence of RNA in the intron.

Key Concepts

- The only factors required for autosplicing *in vitro* by group I introns are a monovalent cation, a divalent cation, and a guanine nucleotide.
- Splicing occurs by two transesterifications, without requiring input of energy.
- The 3′–OH end of the guanine cofactor attacks the 5′ end of the intron in the first transesterification.
- The 3′–OH end generated at the end of the first exon attacks the junction between the intron and second exon in the second transesterification.
- The intron is released as a linear molecule that circularizes when its 3′–OH terminus attacks a bond at one of two internal positions.
- The G^{414}–A^{16} internal bond of the intron can also be attacked by other nucleotides in a *trans*-splicing reaction.

Group I introns are found in diverse locations. They occur in the genes coding for rRNA in the nuclei of the lower eukaryotes *Tetrahymena thermophila* (a ciliate) and *Physarum polycephalum* (a slime mold). They are common in the genes of fungal mitochondria. They are present in three genes of phage T4 and also are found in bacteria. Group I introns have an intrinsic ability to splice themselves. This is called **self-splicing** or **autosplicing**. (This property is found also in the group II introns discussed in *Molecular Biology 5.24.11 Group II introns autosplice via lariat formation*. (749))

Self-splicing was discovered as a property of the transcripts of the rRNA genes in *T. thermophila* (for an account of the discovery see *Great Experiments 4.4 RNA catalysis*). The genes for the two major rRNAs follow the usual organization, in which both are expressed as part of a common transcription unit. The product is a 35S precursor RNA with the sequence of the small rRNA in the 5′ part, and the sequence of the larger (26S) rRNA toward the 3′ end.

In some strains of *T. thermophila,* the sequence coding for 26S rRNA is interrupted by a single, short intron. When the 35S precursor RNA is incubated *in vitro,* splicing occurs as an autonomous reaction. The intron is excised from the precursor and accumulates as a linear fragment of 400 bases, which is subsequently converted to a

circular RNA. These events are summarized in **Figure 26.1**.

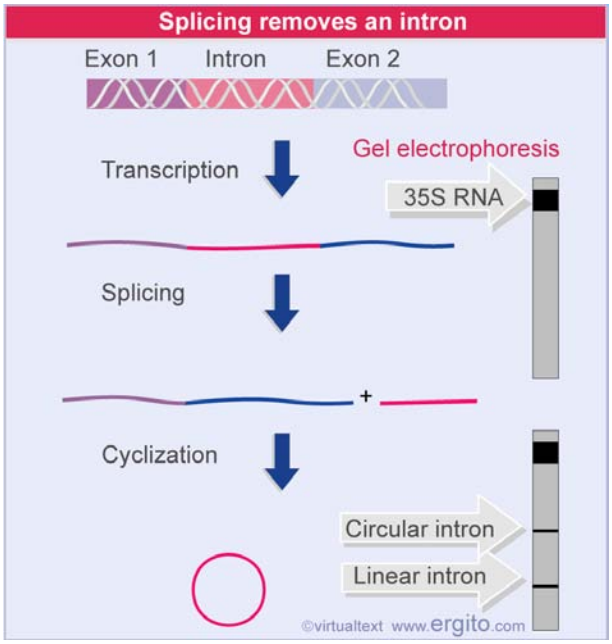

Figure 26.1 Splicing of the *Tetrahymena* 35S rRNA precursor can be followed by gel electrophoresis. The removal of the intron is revealed by the appearance of a rapidly moving small band. When the intron becomes circular, it electrophoreses more slowly, as seen by a higher band.

The reaction requires only a monovalent cation, a divalent cation, and a guanine nucleotide cofactor. No other base can be substituted for G; but a triphosphate is not needed; GTP, GDP, GMP, and guanosine itself all can be used, so there is no net energy requirement. The guanine nucleotide must have a 3′–OH group (747).

The fate of the guanine nucleotide can be followed by using a radioactive label. The radioactivity initially enters the excised linear intron fragment. The G residue becomes linked to the 5′ end of the intron by a normal phosphodiester bond.

Figure 26.2 shows that three transfer reactions occur. In the first transfer, the guanine nucleotide behaves as a cofactor that provides a free 3′–OH group that attacks the 5′ end of the intron. This reaction creates the G–intron link and generates a 3′–OH group at the end of the exon. The second transfer involves a similar chemical reaction, in which this 3′–OH then attacks the second exon. The two transfers are connected; no free exons have been observed, so their ligation may occur as part of the same reaction that releases the intron. The intron is released as a linear molecule, but the third transfer reaction converts it to a circle.

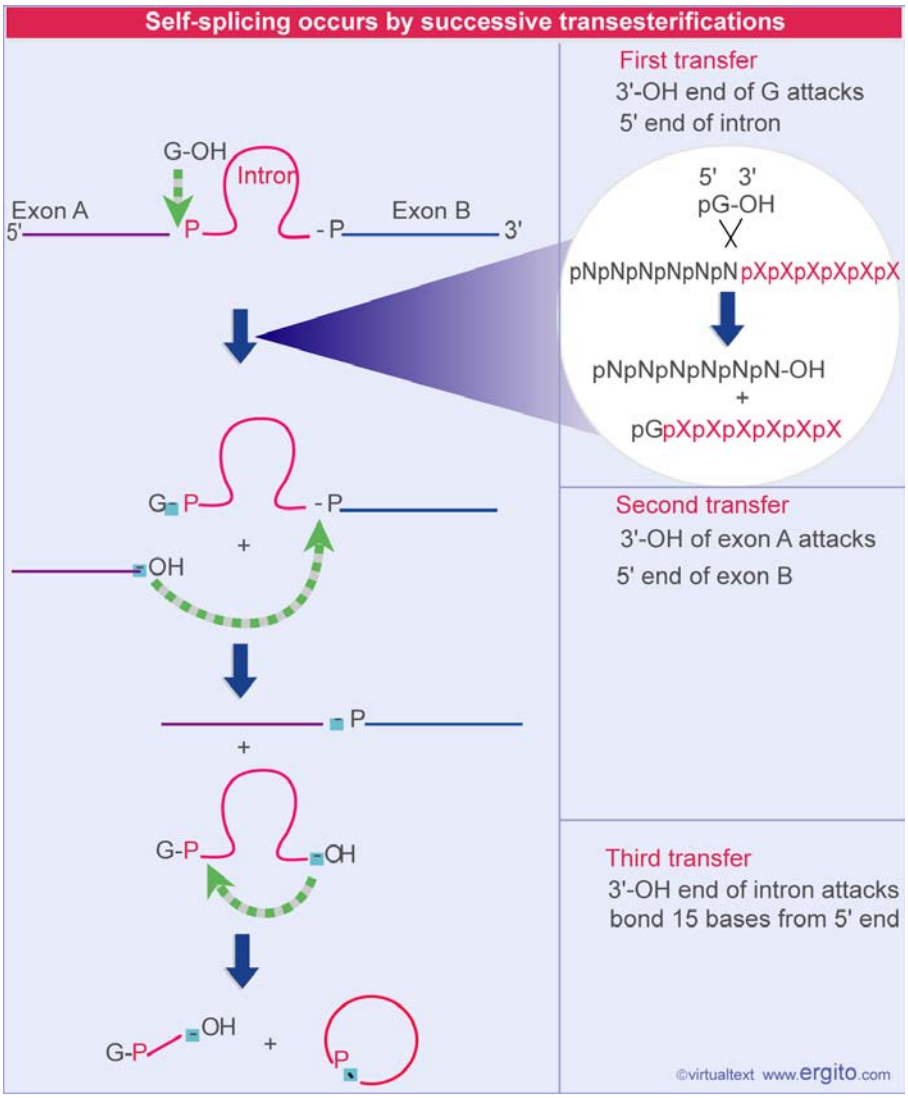

Figure 26.2 Self-splicing occurs by transesterification reactions in which bonds are exchanged directly. The bonds that have been generated at each stage are indicated by the shaded boxes.

Each stage of the self-splicing reaction occurs by a transesterification, in which one phosphate ester is converted directly into another, without any intermediary hydrolysis. Bonds are exchanged directly, and energy is conserved, so the reaction does not require input of energy from hydrolysis of ATP or GTP. (There is a parallel for the transfer of bonds without net input of energy in the DNA nicking-closing enzymes discussed in *Molecular Biology 4.15 Recombination and repair*.)

If each of the consecutive transesterification reactions involves no net change of energy, why does the splicing reaction proceed to completion instead of coming to equilibrium between spliced product and nonspliced precursor? The concentration of GTP is high relative to that of RNA, and therefore drives the reaction forward; and a change in secondary structure in the RNA prevents the reverse reaction.

The in vitro *system includes no protein so the ability to splice is intrinsic to the RNA.*

The RNA forms a specific secondary/tertiary structure in which the relevant groups are brought into juxtaposition so that a guanine nucleotide can be bound to a specific site and then the bond breakage and reunion reactions shown in **Figure 26.2** can occur. Although a property of the RNA itself, the reaction is assisted *in vivo* by proteins, which stabilize the RNA structure (748).

The ability to engage in these transfer reactions resides with the sequence of the intron, which continues to be reactive after its excision as a linear molecule. **Figure 26.3** summarizes its activities:

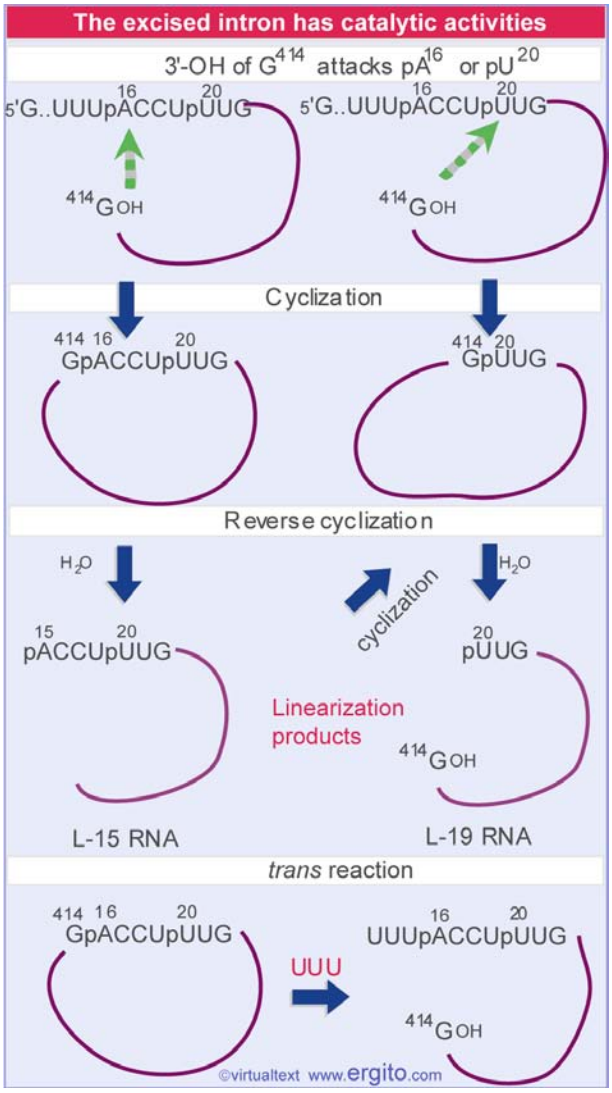

Figure 26.3 The excised intron can form circles by using either of two internal sites for reaction with the 5′ end, and can reopen the circles by reaction with water or oligonucleotides.

The intron can circularize when the 3′ terminal G attacks either of two positions near the 5′ end. The internal bond is broken and the new 5′ end is transferred to the 3′ –OH end of the intron. The *primary cyclization* usually involves reaction between the terminal G^{414} and the A^{16}. This is the most common reaction (shown as the third

transfer in **Figure 26.2**). Less frequently, the G^{414} reacts with U^{20}. Each reaction generates a circular intron and a linear fragment that represents the original 5′ region (15 bases long for attack on A^{16}, 19 bases long for attack on U^{20}). The released 5′ fragment contains the original added guanine nucleotide.

Either type of circle can regenerate a linear molecule *in vitro* by specifically hydrolyzing the bond (G^{414}–A^{16} or G^{414}–U^{20}) that had closed the circle. This is called a *reverse cyclization.* The linear molecule generated by reversing the primary cyclization at A^{16} remains reactive, and can perform a secondary cyclization by attacking U^{20}.

The final product of the spontaneous reactions following release of the intron is the L-19 RNA, a linear molecule generated by reversing the shorter circular form. This molecule has an enzymatic activity that allows it to catalyze the extension of short oligonucleotides (not shown in the figure, but see **Figure 26.8**).

The reactivity of the released intron extends beyond merely reversing the cyclization reaction. Addition of the oligonucleotide UUU reopens the primary circle by reacting with the G^{414}–A^{16} bond. The UUU (which resembles the 3′ end of the 15-mer released by the primary cyclization) becomes the 5′ end of the linear molecule that is formed. This is an *intermolecular* reaction, and thus demonstrates the ability to connect together two different RNA molecules.

This series of reactions demonstrates vividly that the autocatalytic activity reflects a generalized ability of the RNA molecule to form an active center that can bind guanine cofactors, recognize oligonucleotides, and bring together the reacting groups in a conformation that allows bonds to be broken and rejoined. Other group I introns have not been investigated in as much detail as the *Tetrahymena* intron, but their properties are generally similar (750; for review see 255; 256).

The autosplicing reaction is an intrinsic property of RNA *in vitro,* but to what degree are proteins involved *in vivo?* Some indications for the involvement of proteins are provided by mitochondrial systems, where splicing of group I introns requires the *trans*-acting products of other genes. One striking case is presented by the *cyt18* mutant of *N. crassa,* which is defective in splicing several mitochondrial group I introns. The product of this gene turns out to be the mitochondrial tyrosyl-tRNA synthetase! This is explained by the fact that the intron can take up a tRNA-like tertiary structure that is stabilized by the synthetase and which promotes the catalytic reaction (2512).

This relationship between the synthetase and splicing is consistent with the idea that splicing originated as an RNA-mediated reaction, subsequently assisted by RNA-binding proteins that originally had other functions. The *in vitro* self-splicing ability may represent the basic biochemical interaction. The RNA structure creates the active site, but is able to function efficiently *in vivo* only when assisted by a protein complex.

Last updated on 6-20-2002

Reviews

255. Cech, T. R. (1985). *Self-splicing RNA: implications for evolution*. Int. Rev. Cytol. 93, 3-22.

256. Cech, T. R. (1987). *The chemistry of self-splicing RNA and RNA enzymes*. Science 236, 1532-1539.

References

747. Cech, T. R. et al. (1981). *In vitro splicing of the rRNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence*. Cell 27, 487-496.

748. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982). *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell 31, 147-157.

749. Belfort, M., Pedersen-Lane, J., West, D., Ehrenman, K., Maley, G., Chu, F., and Maley, F. (1985). *Processing of the intron-containing thymidylate synthase (td) gene of phage T4 is at the RNA level*. Cell 41, 375-382.

750. Been, M. D. and Cech, T. R. (1986). *One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity*. Cell 47, 207-216.

2512. Myers, C. A., Kuhla, B., Cusack, S., and Lambowitz, A. M. (2002). *tRNA-like recognition of group I introns by a tyrosyl-tRNA synthetase.* Proc. Natl. Acad. Sci. USA 99, 2630-2635.

CATALYTIC RNA

5.26.3 Group I introns form a characteristic secondary structure

Key Concepts

- Group I introns form a secondary structure with 9 duplex regions.
- The core of regions P3, P4, P6, P7 has catalytic activity.
- Regions P4 and P7 are both formed by pairing between conserved consensus sequences.
- A sequence adjacent to P7 base pairs with the sequence that contains the reactive G.

All group I introns can be organized into a characteristic secondary structure, with 9 helices (P1-P9). **Figure 26.4** shows a model for the secondary structure of the *Tetrahymena* intron. Two of the base-paired regions are generated by pairing between conserved sequence elements that are common to group I introns. P4 is constructed from the sequences *P* and *Q;*, P7 is formed from sequences *R* and *S*. The other base-paired regions vary in sequence in individual introns. Mutational analysis identifies an intron "core," containing P3, P4, P6, and P7, which provides the minimal region that can undertake a catalytic reaction. The lengths of group I introns vary widely, and the consensus sequences are located a considerable distance from the actual splice junctions.

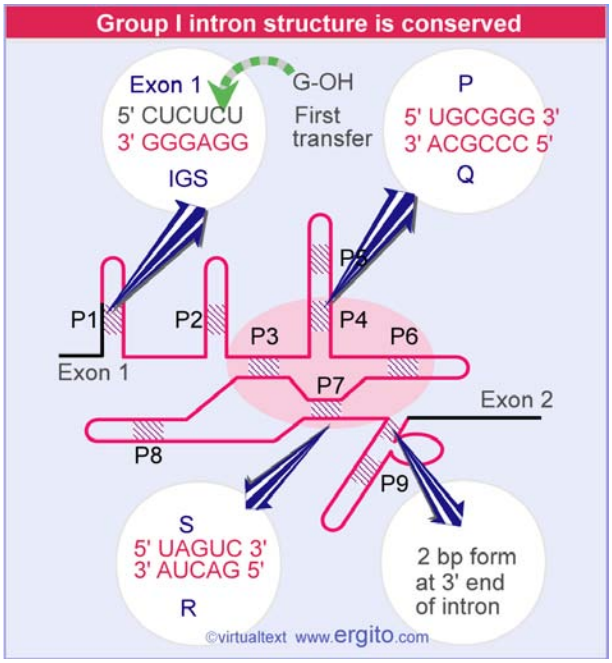

Figure 26.4 Group I introns have a common secondary structure that is formed by 9 base paired regions. The sequences of regions P4 and P7 are conserved, and identify the individual sequence elements P, Q, R, and S. P1 is created by pairing between the end of the left exon and the IGS of the intron; a region between P7 and P9 pairs with the 3′ end of the intron.

Some of the pairing reactions are directly involved in bringing the splice junctions into a conformation that supports the enzymatic reaction. P1 includes the 3′ end of the left exon. The sequence within the intron that pairs with the exon is called the IGS, or internal guide sequence. (Its name reflects the fact that originally the region immediately 3′ to the IGS sequence shown in the figure was thought to pair with the 3′ splice junction, thus bringing the two junctions together. This interaction may occur, but does not seem to be essential.) A very short sequence, sometimes as short as 2 bases, between P7 and P9, base pairs with the sequence that immediately precedes the reactive G (position 414 in *Tetrahymena*) at the 3′ end of the intron (752).

The importance of base pairing in creating the necessary core structure in the RNA is emphasized by the properties of *cis*-acting mutations that prevent splicing of group I introns. Such mutations have been isolated for the mitochondrial introns through mutants that cannot remove an intron *in vivo*, and they have been isolated for the *Tetrahymena* intron by transferring the splicing reaction into a bacterial environment. The construct shown in **Figure 26.5** allows the splicing reaction to be followed in *E. coli*. The self-splicing intron is placed at a location that interrupts the tenth codon of the β-galactosidase coding sequence. The protein can therefore be successfully translated from an RNA only after the intron has been removed.

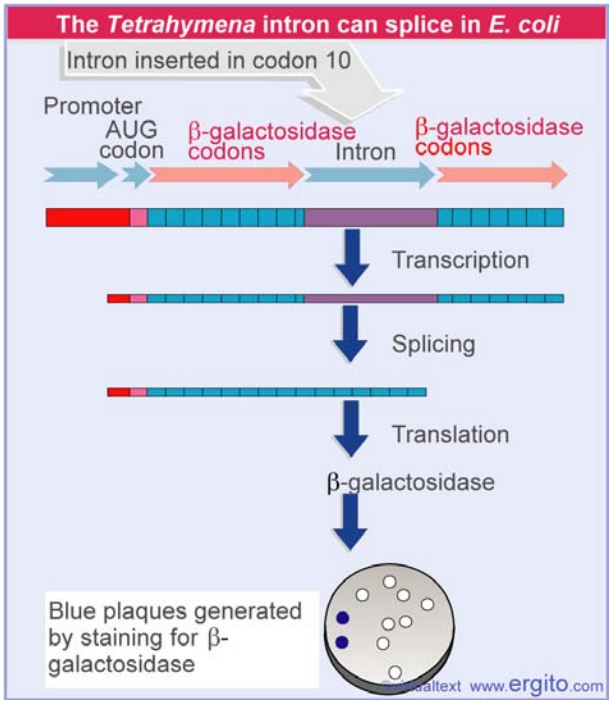

Figure 26.5 Placing the *Tetrahymena* intron within the β -galactosidase coding sequence creates an assay for self-splicing in *E. coli*. Synthesis of β -galactosidase can be tested by adding a compound that is turned blue by the enzyme. The sequence is carried by a bacteriophage, so the presence of blue plaques (containing infected bacteria) indicates successful splicing.

The synthesis of β-galactosidase in this system indicates that splicing can occur in conditions quite distant from those prevailing in *Tetrahymena* or even *in vitro*. One interpretation of this result is that self-splicing can occur in the bacterial cell. Another possibility is that there are bacterial proteins that assist the reaction.

Using this assay, we can introduce mutations into the intron to see whether they prevent the reaction. Mutations in the group I consensus sequences that disrupt their base pairing stop splicing. The mutations can be reverted by making compensating changes that restore base pairing.

Mutations in the corresponding consensus sequences in mitochondrial group I introns have similar effects. A mutation in one consensus sequence may be reverted by a mutation in the complementary consensus sequence to restore pairing; for example, mutations in the R consensus can be compensated by mutations in the S consensus.

Together these results suggest that the group I splicing reaction depends on the formation of secondary structure between pairs of consensus sequences within the intron. The principle established by this work is that *sequences distant from the splice junctions themselves are required to form the active site that makes self-splicing possible* (751).

References

751. Burke, J. M. et al. (1986). *Role of conserved sequence elements 9L and 2 in self-splicing of the Tetrahymena ribosomal RNA precursor*. Cell 45, 167-176.

752. Michel, F. and Wetshof, E. (1990). *Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis*. J. Mol. Biol. 216, 585-610.

CATALYTIC RNA

5.26.4 Ribozymes have various catalytic activities

Key Concepts

- By changing the substrate binding-site of a group I intron, it is possible to introduce alternative sequences that interact with the reactive G.
- The reactions follow classical enzyme kinetics with a low catalytic rate.
- Reactions using 2′–OH bonds could have been the basis for evolving the original catalytic activities in RNA.

The catalytic activity of group I introns was discovered by virtue of their ability to autosplice, but they are able to undertake other catalytic reactions *in vitro*. All of these reactions are based on transesterifications. We analyze these reactions in terms of their relationship to the splicing reaction itself.

The catalytic activity of a group I intron is conferred by its ability to generate a particular secondary and tertiary structure that creates active sites, equivalent to the active sites of a conventional (proteinaceous) enzyme. **Figure 26.6** illustrates the splicing reaction in terms of these sites (this is the same series of reactions shown previously in **Figure 26.2**).

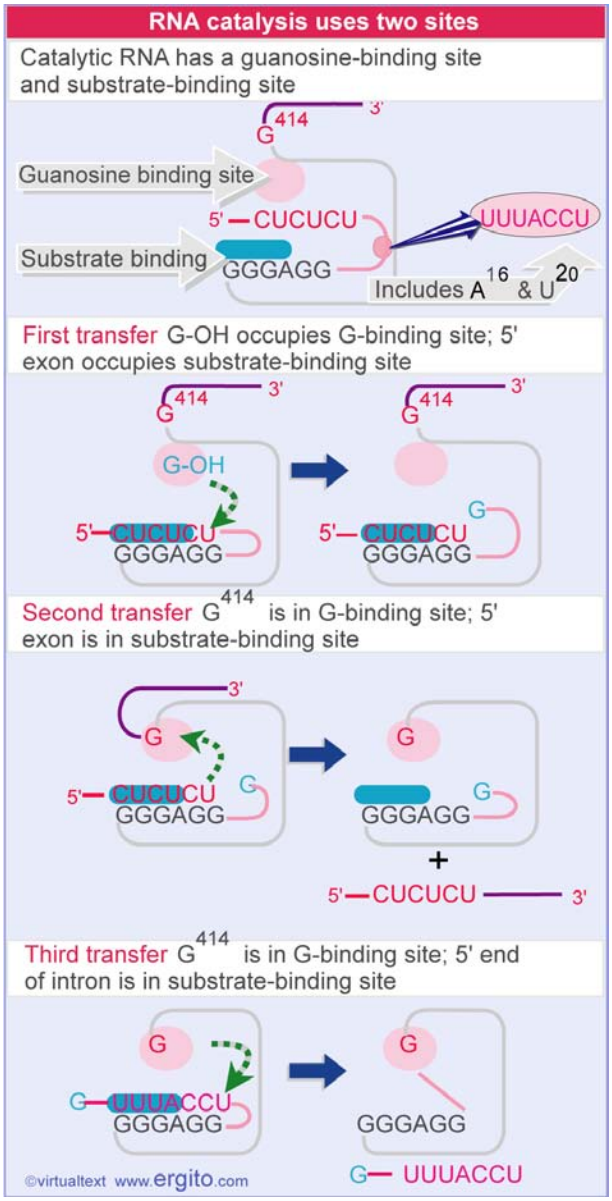

Figure 26.6 Excision of the group I intron in *Tetrahymena* rRNA occurs by successive reactions between the occupants of the guanosine-binding site and substrate-binding site. The left exon is red, and the right exon is purple.

The substrate-binding site is formed from the P1 helix, in which the 3′ end of the first intron base pairs with the IGS in an intermolecular reaction. A guanosine-binding site is formed by sequences in P7. This site may be occupied either by a free guanosine nucleotide or by the G residue in position 414. In the first transfer reaction, it is used by free guanosine nucleotide; but it is subsequently occupied by G^{414}. The second transfer releases the joined exons. The third transfer creates the circular intron.

Binding to the substrate involves a change of conformation; before substrate binding, the 5′ end of the IGS is close to P2 and P8, but after binding, when it forms the P1

helix, it is close to conserved bases that lie between P4 and P5. The reaction is visualized by contacts that are detected in the secondary structure in **Figure 26.7**. In the tertiary structure, the two sites alternatively contacted by P1 are 37 Å apart, which implies a substantial movement in the position of P1.

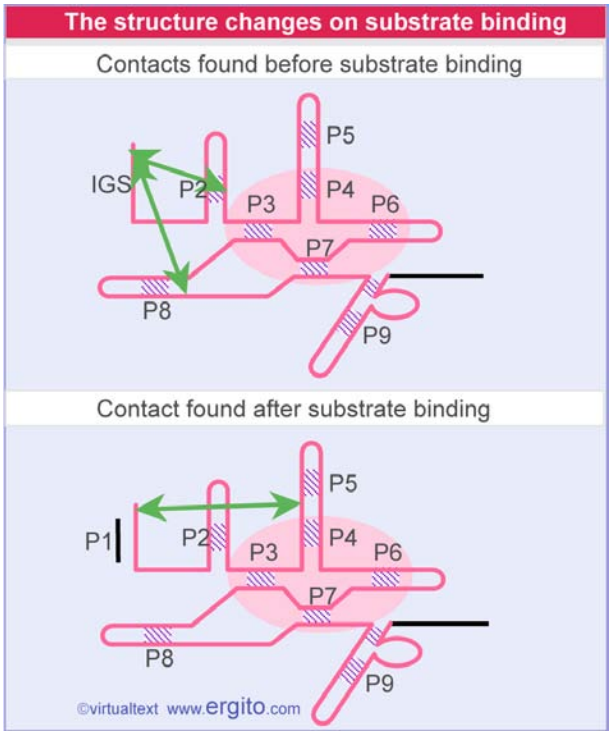

Figure 26.7 The position of the IGS in the tertiary structure changes when P1 is formed by substrate binding.

The L-19 RNA is generated by opening the circular intron (shown as the last stage of the intramolecular rearrangements shown in **Figure 26.3**). It still retains enzymatic abilities. These resemble the activities involved in the original splicing reaction, and we may consider ribozyme function in terms of the ability to bind an intramolecular sequence complementary to the IGS in the substrate-binding site, while binding either the terminal G^{414} or a free G-nucleotide in the G-binding site.

Figure 26.8 illustrates the mechanism by which the oligonucleotide C_5 is extended to generate a C_6 chain. The C_5 oligonucleotide binds in the substrate-binding site, while G^{414} occupies the G-binding site. By transesterification reactions, a C is transferred from C_5 to the 3′–terminal G, and then back to a new C_5 molecule. Further transfer reactions lead to the accumulation of longer cytosine oligonucleotides. The reaction is a true catalysis, because the L-19 RNA remains unchanged, and is available to catalyze multiple cycles. The ribozyme is behaving as a nucleotidyl transferase.

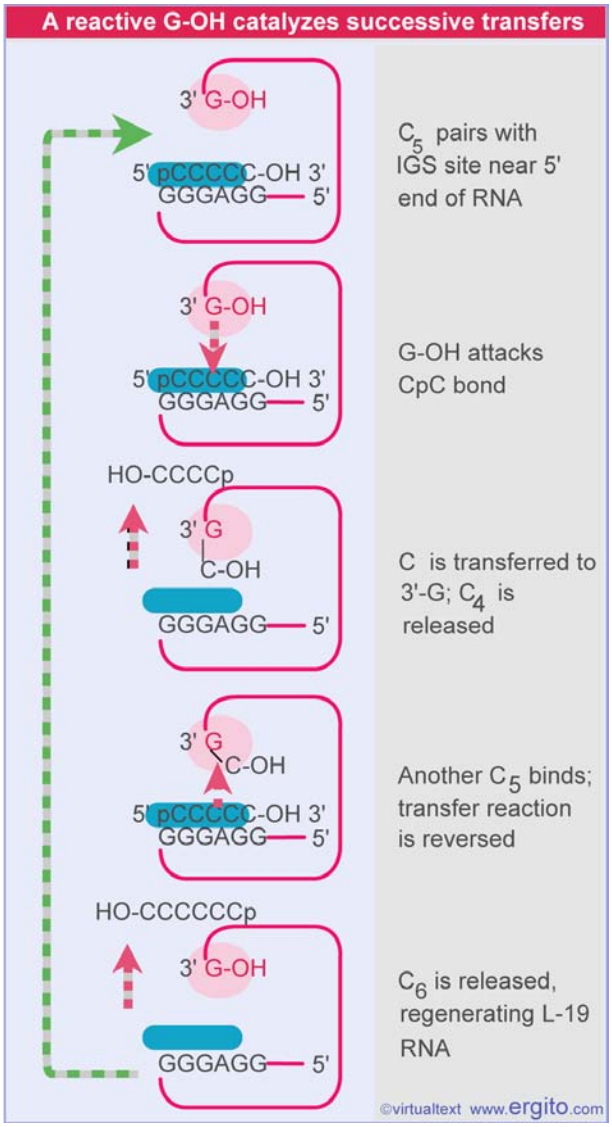

Figure 26.8 The L-19 linear RNA can bind C in the substrate-binding site; the reactive G-OH 3′ end is located in the G-binding site, and catalyzes transfer reactions that convert 2 C5 oligonucleotides into a C4 and a C6 oligonucleotide.

Some further enzymatic reactions are characterized in **Figure 26.9**. The ribozyme can function as a sequence-specific endoribonuclease by utilizing the ability of the IGS to bind complementary sequences. In this example, it binds an external substrate containing the sequence CUCU, instead of binding the analogous sequence that is usually contained at the end of the left exon. A guanine-containing nucleotide is present in the G-binding site, and attacks the CUCU sequence in precisely the same way that the exon is usually attacked in the first transfer reaction. This cleaves the target sequence into a 5′ molecule that resembles the left exon, and a 3′ molecule that bears a terminal G residue. By mutating the IGS element, it is possible to change the specificity of the ribozyme, so that it recognizes sequences complementary to the new sequence at the IGS region (for review see 257).

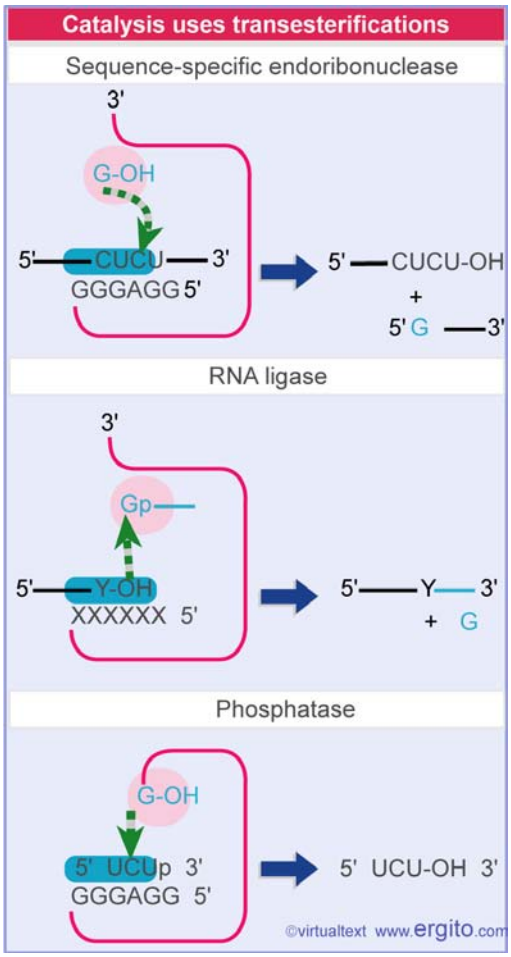

Figure 26.9 Catalytic reactions of the ribozyme involve transesterifications between a group in the substrate-binding site and a group in the G-binding site.

Altering the IGS, so that the specificity of the substrate-binding site is changed to enable other RNA targets to enter, can be used to generate a ligase activity. An RNA terminating in a 3′–OH is bound in the substrate site, and an RNA terminating in a 5′–G residue is bound in the G-binding site. An attack by the hydroxyl on the phosphate bond connects the two RNA molecules, with the loss of the G residue.

The phosphatase reaction is not directly related to the splicing transfer reactions. An oligonucleotide sequence that is complementary to the IGS and terminates in a 3′–phosphate can be attacked by the G^{414}. The phosphate is transferred to the G^{414}, and an oligonucleotide with a free 3′–OH end is then released. The phosphate can then be transferred either to an oligonucleotide terminating in 3′–OH (effectively reversing the reaction) or indeed to water (releasing inorganic phosphate and completing an authentic phosphatase reaction).

The reactions catalyzed by RNA can be characterized in the same way as classical enzymatic reactions in terms of Michaelis-Menten kinetics. **Figure 26.10** analyzes the reactions catalyzed by RNA. The K_M values for RNA-catalyzed reactions are

low, and therefore imply that the RNA can bind its substrate with high specificity. The turnover numbers are low, which reflects a low catalytic rate. In effect, the RNA molecules behave in the same general manner as traditionally defined for enzymes, although they are relatively slow compared to protein catalysts (where a typical range of turnover numbers is 10^3–10^6).

RNA catalysis is enzymatic

Enzyme	Substrate	K_M (mM)	Turnover (/min)
19 base virusoid	24 base RNA	0.0006	0.5
L-19 Intron	CCCCCC	0.04	1.7
RNAase P RNA	pre-tRNA	0.00003	0.4
RNAase P complete	pre-tRNA	0.00003	29
RNAase T1	GpA	0.05	5,700
β-galactosidase	lactose	4.0	12,500

Figure 26.10 Reactions catalyzed by RNA have the same features as those catalyzed by proteins, although the rate is slower. The K_M gives the concentration of substrate required for half-maximum velocity; this is an inverse measure of the affinity of the enzyme for substrate. The turnover number gives the number of substrate molecules transformed in unit time by a single catalytic site.

How does RNA provide a catalytic center? Its ability seems reasonable if we think of an active center as a surface that exposes a series of active groups in a fixed relationship. In a protein, the active groups are provided by the side chains of the amino acids, which have appreciable variety, including positive and negative ionic groups and hydrophobic groups. In an RNA, the available moieties are more restricted, consisting primarily of the exposed groups of bases. Short regions are held in a particular structure by the secondary/tertiary conformation of the molecule, providing a surface of active groups able to maintain an environment in which bonds can be broken and made in another molecule. It seems inevitable that the interaction between the RNA catalyst and the RNA substrate will rely on base pairing to create the environment. Divalent cations (typically Mg^{2+}) play an important role in structure, typically being present at the active site where they coordinate the positions of the various groups. They play a direct role in the endonucleolytic activity of virusoid ribozymes (see *Molecular Biology 5.26.9 Viroids have catalytic activity*).

The evolutionary implications of these discoveries are intriguing. The split personality of the genetic apparatus, in which RNA is present in all components, but proteins undertake catalytic reactions, has always been puzzling. It seems unlikely that the very first replicating systems could have contained both nucleic acid and protein.

But suppose that the first systems contained only a self-replicating nucleic acid with primitive catalytic activities, just those needed to make and break phosphodiester bonds. If we suppose that the involvement of 2′–OH bonds in current splicing reactions is derived from these primitive catalytic activities, we may argue that the original nucleic acid was RNA, since DNA lacks the 2′–OH group and therefore

could not undertake such reactions. Proteins could have been added for their ability to stabilize the RNA structure. Then the greater versatility of proteins could have allowed them to take over catalytic reactions, leading eventually to the complex and sophisticated apparatus of modern gene expression.

Reviews

257. Cech, T. R. (1990). *Self-splicing of group I introns*. Annu. Rev. Biochem. 59, 543-568.

CATALYTIC RNA

5.26.5 Some group I introns code for endonucleases that sponsor mobility

Key Terms

Intron homing describes the ability of certain introns to insert themselves into a target DNA. The reaction is specific for a single target sequence.

Key Concepts

- Mobile introns are able to insert themselves into new sites.
- Mobile group I introns code for an endonuclease that makes a double-strand break at a target site.
- The intron transposes into the site of the double-strand break by a DNA-mediated replicative mechanism.

Certain introns of both the group I and group II classes contain open reading frames that are translated into proteins. Expression of the proteins allows the intron (either in its original DNA form or as a DNA copy of the RNA) to be *mobile*: it is able to insert itself into a new genomic site. Introns of both groups I and II are extremely widespread, being found in both prokaryotes and eukaryotes. Group I introns migrate by DNA-mediated mechanisms, whereas group II introns migrate by RNA-mediated mechanisms.

Intron mobility was first detected by crosses in which the alleles for the relevant gene differ with regard to their possession of the intron. Polymorphisms for the presence or absence of introns are common in fungal mitochondria. This is consistent with the view that these introns originated by insertion into the gene. Some light on the process that could be involved is cast by an analysis of recombination in crosses involving the large rRNA gene of the yeast mitochondrion.

This gene has a group I intron that contains a coding sequence. The intron is present in some strains of yeast (called ω^+) but absent in others (ω^-). Genetic crosses between ω^+ and ω^- are *polar:* the progeny are usually ω^+.

If we think of the ω^+ strain as a donor and the ω^- strain as a recipient, we form the view that in $\omega^+ \times \omega^-$ crosses a new copy of the intron is generated in the ω^- genome. As a result, the progeny are all ω^+.

Mutations can occur in either parent to abolish the polarity. Mutants show normal segregation, with equal numbers of ω^+ and ω^- progeny. The mutations indicate the nature of the process. Mutations in the ω^- strain occur close to the site where the intron would be inserted. Mutations in the ω^+ strain lie in the reading frame of the intron and prevent production of the protein. This suggests the model of **Figure 26.11**, in which the protein coded by the intron in an ω^+ strain recognizes the site

where the intron should be inserted in an ω^- strain and causes it to be preferentially inherited.

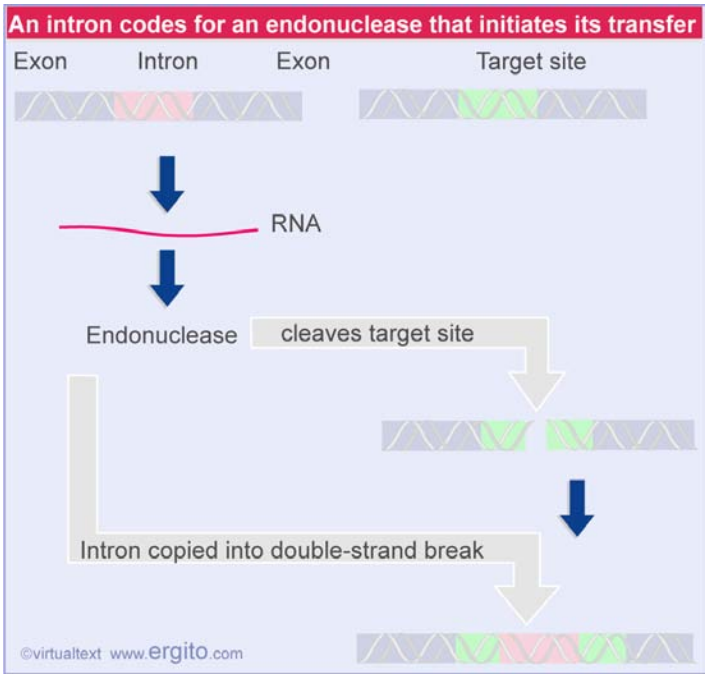

Figure 26.11 An intron codes for an endonuclease that makes a double-strand break in DNA. The sequence of the intron is duplicated and then inserted at the break.

What is the action of the protein? The product of the ω intron is an endonuclease *that recognizes the* ω^- *gene as a target for a double-strand break.* The endonuclease recognizes an 18 bp target sequence that contains the site where the intron is inserted. The target sequence is cleaved on each strand of DNA 2 bases to the 3′ side of the insertion site. So the cleavage sites are 4 bp apart, and generate overhanging single strands.

This type of cleavage is related to the cleavage characteristic of transposons when they migrate to new sites (see *Molecular Biology 4.16 Transposons*). The double-strand break probably initiates a gene conversion process in which the sequence of the ω^+ gene is copied to replace the sequence of the ω^- gene. The reaction involves transposition by a duplicative mechanism, and occurs solely at the level of DNA. Insertion of the intron interrupts the sequence recognized by the endonuclease, thus ensuring stability.

Many group I introns code for endonucleases that make them mobile. Several different families of endonucleases are found; one common feature is the presence of the amino acid sequence LAGLIDADG near the active site (for review see 4515). Similar introns often carry quite different endonucleases. There are differences in the details of insertion; for example, the endonuclease coded by the phage T4 *td* intron cleaves a target site that is 24 bp upstream of the site at which the intron is itself inserted. The dissociation between the intron sequence and the endonuclease sequence is emphasized by the fact that the same endonuclease sequences are found in inteins (sequences that code for self-splicing proteins; see *Molecular*

Biology 5.26.12 Protein splicing is autocatalytic).

The variation in the endonucleases means that there is no homology between the sequences of their target sites. The target sites are among the longest and therefore the most specific known for any endonucleases (with a range of 14-40 bp). The specificity ensures that the intron perpetuates itself only by insertion into a single target site and not elsewhere in the genome. This is called **intron homing**.

Introns carrying sequences that code for endonucleases are found in a variety of bacteria and lower eukaryotes. These results strengthen the view that introns carrying coding sequences originated as independent elements.

Last updated on December 24, 2003

Reviews

4515. Belfort, M. and Roberts, R. J. (1997). *Homing endonucleases: keeping the house in order*. Nucleic Acids Res. 25, 3379-3388.

CATALYTIC RNA

5.26.6 Some group II introns code for reverse transcriptases

Key Concepts

- Some group II introns code for a reverse transcriptase that generates a DNA copy of the RNA sequence that transposes by a retroposon-like mechanism.

Most of the open reading frames contained in group II introns have regions that are related to reverse transcriptases. Introns of this type are found in organelles of lower eukaryotes and also in some bacteria. The reverse transcriptase activity is specific for the intron, and is involved in homing. The reverse transcriptase generates a DNA copy of the intron from the pre-mRNA, and thus allows the intron to become mobile by a mechanism resembling that of retroviruses (see *Molecular Biology 4.17.2 The retrovirus life cycle involves transposition-like events*). The type of retrotransposition involved in this case resembles that of a group of retroposons that lack LTRs, and which generate the 3′–OH needed for priming by making a nick in the target (see **Figure 17.20** in *Molecular Biology 4.17.12 LINES use an endonuclease to generate a priming end*).

The best characterized mobile group II introns code for a single protein in a region of the intron beyond its catalytic core. The typical protein contains an N-terminal reverse transcriptase activity, a central domain associated with maturase activity, and a C-terminal endonuclease domain. The endonuclease initiates the transposition reaction, and thus plays the same role in homing as its counterpart in a group I intron. The reverse transcriptase generates a DNA copy of the intron that is inserted at the homing site. The endonuclease also cleaves target sites that resemble, but are not identical to the homing site, at much lower frequency, leading to insertion of the intron at new locations (3330).

Figure 26.12 illustrates the transposition reaction for a typical group II intron. The endonuclease makes a double-strand break at the target site. A 3′ end is generated at the site of the break, and provides a primer for the reverse transcriptase. The intron RNA provides the template for the synthesis of cDNA. Because the RNA includes exon sequences on either side of the intron, the cDNA product is longer than the region of the intron itself, so that it can span the double-strand break, allowing the cDNA to repair the break. The result is the insertion of the intron (754; for review see 259).

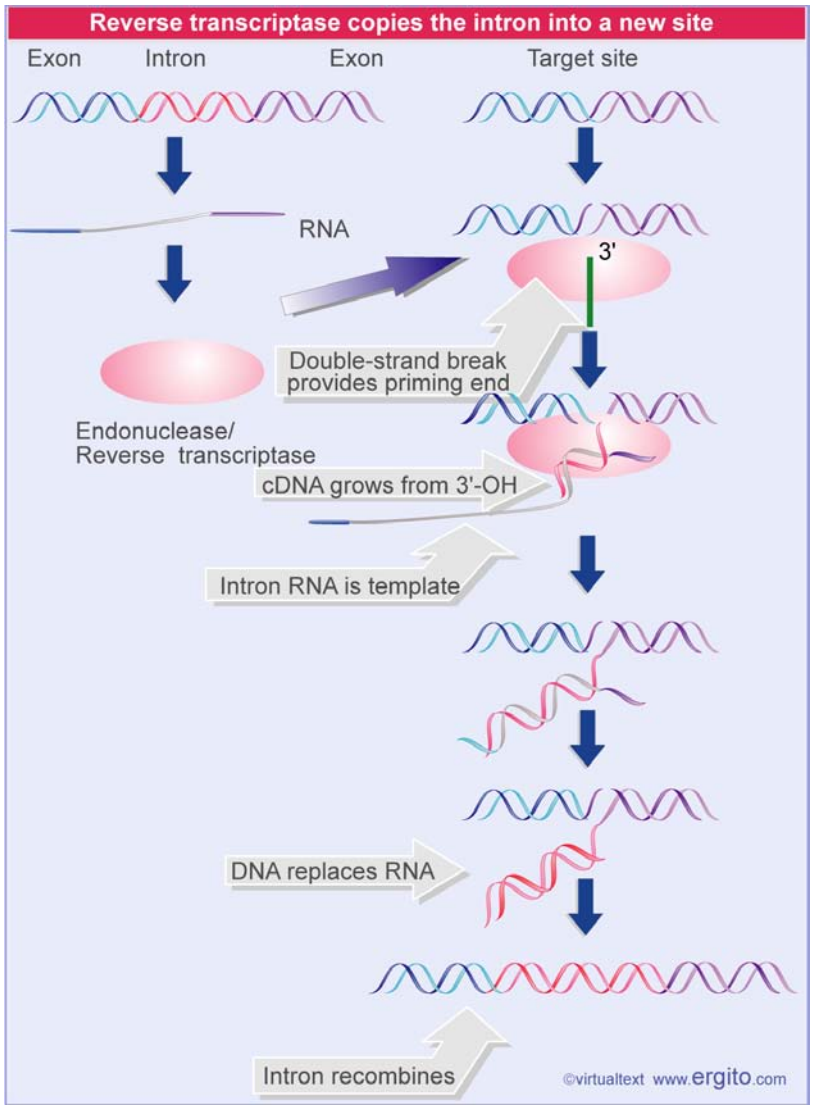

Figure 26.12 Reverse transcriptase coded by an intron allows a copy of the RNA to be inserted at a target site generated by a double-strand break.

Last updated on 1-22-2003

Reviews

259. Lambowitz, A. M. and Belfort, M. (1993). *Introns as mobile genetic elements*. Annu. Rev. Biochem. 62, 587-622.

References

754. Zimmerly, S. et al. (1995). *Group II intron mobility occurs by target DNA-primed reverse transcription*. Cell 82, 545-554.

3330. Dickson, L., Huang, H. R., Liu, L., Matsuura, M., Lambowitz, A. M., and Perlman, P. S. (2001). *Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites.* Proc. Natl. Acad. Sci. USA 98, 13207-13212.

CATALYTIC RNA

5.26.7 Some autosplicing introns require maturases

Key Terms

A **maturase** is a protein coded by a group I or group II intron that is needed to assist the RNA to form the active conformation that is required for self-splicing.

Both types of autosplicing intron may code for proteins invovled in their perpetuation and homing, such as endonucelases in group I and reverse transcriptases in group II. In addition, both tyeps of intron may code for **maturase** activities that are required to assist the splicing reaction (753).

The maturase activity is part of the single open reading frame coded by the intron. In the example of introns that code for homing endonucleases, the single protein product has both endonuclease and maturase activity. Mutational analysis shows that the two activities are independent (4516).

Structural analysis shows that the endonuclease and maturase activities are provided by different active sites in the protein, each coded by a separate domain (4517). The endonuclease site binds to DNA, but the maturase site binds to the intron RNA. **Figure 26.13** shows the structure of one such protein bound to DNA. A characteristic feature of the endonuclease is the presence of parallel α helices, containing the hallmark LAGLIDADG sequences, leading to the two catalytic amino acids. The maturase activity is located some distance away on the surface of the protein.

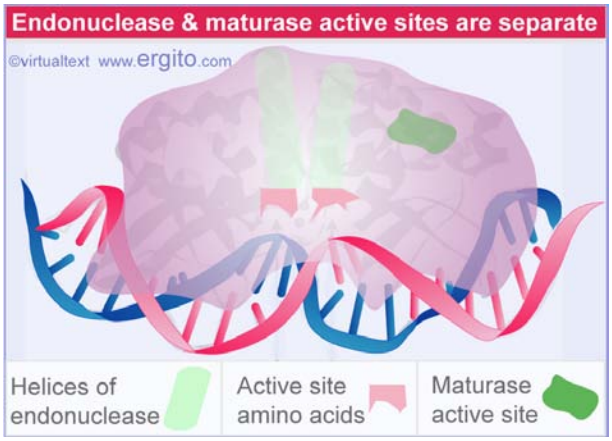

Figure 26.13 A homing intron codes for and endonuclease of the LAGLIDADG family that also has maturase activity. The LAGLIDADG sequences are part of the two α helices that terminate in the catalytic amino acids close to the DNA duplex. The maturase active site is identified by an arginine residue elsewhere on the surface of the protein.

Introns that code for maturases may be unable to splice themselves effectively in the absence of the protein activity. The maturase is in effect a splicing factor that is

required specifically for splicing of the sequence that codes for it. It functions to assist the folding of the catalytic core to form an active site (3329).

The coexistence of endonuclease and maturase activities in the same protein suggests a route for the evolution of the intron. **Figure 26.14** suggests that the intron originated in an independent autosplicing element. The insertion into this element of a sequence coding for an endonuclease gave it mobility. However, the insertion might well disruupt the ability of the RNA sequence to fold into the active structure. This would create pressure for assistance from proteins that could restore folding ability. The incorporation of such a sequence into the intron would maintain its independence.

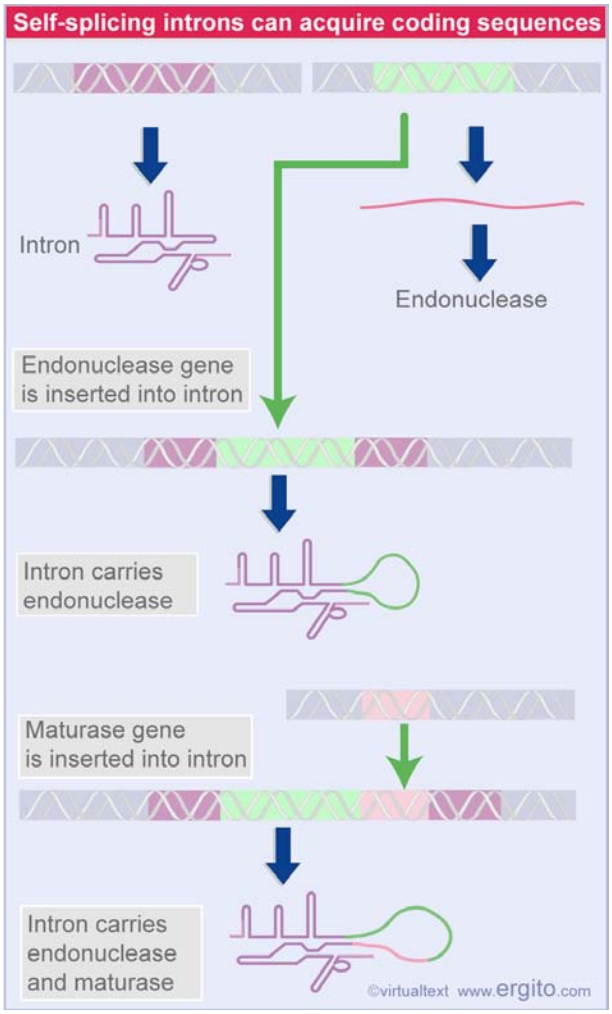

Figure 26.14 The intron originated as independent sequence coding for a self-splicing RNA. The insertion of the endonuclease sequence created a homing intron that was mobile. Then the insertion of the maturase sequence enhanced the ability of the itnron sequewnces to fold into the active structure for splicing.

Some group II introns that do not code for maturase activities may use comparable proteins that are coded by sequences in the host genome. This suggests a possible

route for the evolution of general splicing factors. The factor may have originated as a maturase that specifically assisted the splicing of a particular intron. The coding sequence became isolated from the intron in the host genome, and then it evolved to function with a wider range of substrates that the original intron sequence. The catalytic core of the intron could have evolved into an snRNA.

Last updated on January 6, 2004

References

753. Carignani, G. et al. (1983). *An RNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in S. cerevisiae.* Cell 35, 733-742.

3329. Matsuura, M., Noah, J. W., and Lambowitz, A. M. (2001). *Mechanism of maturase-promoted group II intron splicing.* EMBO J. 20, 7259-7270.

4516. Henke, R. M., Butow, R. A., and Perlman, P. S. (1995). *Maturase and endonuclease functions depend on separate conserved domains of the bifunctional protein encoded by the group I intron aI4 alpha of yeast mitochondrial DNA.* EMBO J. 14, 5094-5099.

4517. Bolduc et al. (2003). *Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endopnuclease and group I splicing factor.* Genes. Dev. 17, 2875-2888.

CATALYTIC RNA

5.26.8 The catalytic activity of RNAase P is due to RNA

Key Concepts

- Ribonuclease P is a ribonucleoprotein in which the RNA has catalytic activity.

One of the first demonstrations of the capabilities of RNA was provided by the dissection of ribonuclease P, an *E. coli* tRNA-processing endonuclease. Ribonuclease P can be dissociated into its two components, the 375 base RNA and the 20 kD polypeptide. Under the conditions initially used to characterize the enzyme activity *in vitro,* both components were necessary to cleave the tRNA substrate.

But a change in ionic conditions, an increase in the concentration of Mg^{2+}, renders the protein component superfluous. *The RNA alone can catalyze the reaction!* Analyzing the results as though the RNA were an enzyme, each "enzyme" catalyzes the cleavage of multiple substrates. Although the catalytic activity resides in the RNA, the protein component greatly increases the speed of the reaction, as seen in the increase in turnover number (see **Figure 26.10**).

Because mutations in either the gene for the RNA or the gene for protein can inactivate RNAase P *in vivo,* we know that both components are necessary for natural enzyme activity. Originally it had been assumed that the protein provided the catalytic activity, while the RNA filled some subsidiary role, for example, assisting in the binding of substrate (it has some short sequences complementary to exposed regions of tRNA). But these roles are reversed!

CATALYTIC RNA

5.26.9 Viroids have catalytic activity

Key Terms

A **viroid** is a small infectious nucleic acid that does not have a protein coat.

A **virusoid (Satellite RNA)** is a small infectious nucleic acid that is encapsidated by a plant virus together with its own genome.

Key Concepts

- Viroids and virusoids form a hammerhead structure that has an self-cleaving activity.
- Similar structures can be generated by pairing a substrate strand that is cleaved by an enzyme strand.
- When an enzyme strand is introduced into a cell, it can pair with a substrate strand target that is then cleaved.

Another example of the ability of RNA to function as an endonuclease is provided by some small plant RNAs (~350 bases) that undertake a self-cleavage reaction. As with the case of the *Tetrahymena* group I intron, however, it is possible to engineer constructs that can function on external substrates (756).

These small plant RNAs fall into two general groups: viroids and virusoids. The **viroids** are infectious RNA molecules that function independently, without enacapsidation by any protein coat. The **virusoids** are similar in organization, but are encapsidated by plant viruses, being packaged together with a viral genome. The virusoids cannot replicate independently, but require assistance from the virus. The virusoids are sometimes called **satellite RNAs**.

Viroids and virusoids both replicate via rolling circles (see **Figure 13.16**). The strand of RNA that is packaged into the virus is called the plus strand. The complementary strand, generated during replication of the RNA, is called the minus strand. Multimers of both plus and minus strands are found. Both types of monomer are generated by cleaving the tail of a rolling circle; circular plus strand monomers are generated by ligating the ends of the linear monomer.

Both plus and minus strands of viroids and virusoids undergo self-cleavage *in vitro*. The cleavage reaction is promoted by divalent metal cations; it generates 5′–OH and 2′–3′–cyclic phosphodiester termini. Some of the RNAs cleave *in vitro* under physiological conditions. Others do so only after a cycle of heating and cooling; this suggests that the isolated RNA has an inappropriate conformation, but can generate an active conformation when it is denatured and renatured.

The viroids and virusoids that undergo self-cleavage form a "hammerhead" secondary structure at the cleavage site, as drawn in the upper part of **Figure 26.15**.

The sequence of this structure is sufficient for cleavage. When the surrounding sequences are deleted, the need for a heating-cooling cycle is obviated, and the small RNA self-cleaves spontaneously. This suggests that the sequences beyond the hammerhead usually interfere with its formation.

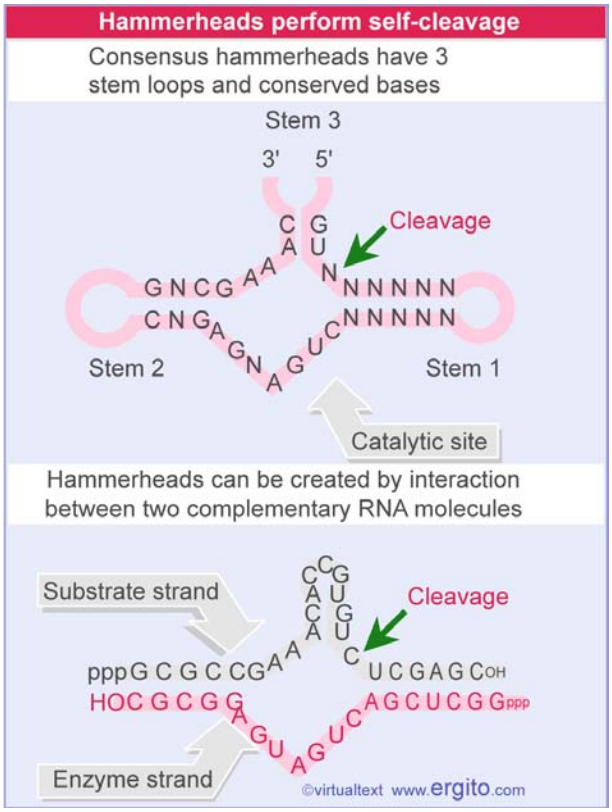

Figure 26.15 Self-cleavage sites of viroids and virusoids have a consensus sequence and form a hammerhead secondary structure by intramolecular pairing. Hammerheads can also be generated by pairing between a substrate strand and an "enzyme" strand.

The active site is a sequence of only 58 nucleotides. The hammerhead contains three stem-loop regions whose position and size are constant, and 13 conserved nucleotides, mostly in the regions connecting the center of the structure. The conserved bases and duplex stems generate an RNA with the intrinsic ability to cleave (757).

An active hammerhead can also be generated by pairing an RNA representing one side of the structure with an RNA representing the other side. The lower part of **Figure 26.15** shows an example of a hammerhead generated by hybridizing a 19 base molecule with a 24 base molecule. The hybrid mimics the hammerhead structure, with the omission of loops I and III. When the 19 base RNA is added to the 24 base RNA, cleavage occurs at the appropriate position in the hammerhead.

We may regard the top (24 base) strand of this hybrid as comprising the "substrate," and the bottom (19 base) strand as comprising the "enzyme." When the 19 base RNA is mixed with an excess of the 24 base RNA, multiple copies of the 24 base RNA are

cleaved. This suggests that there is a cycle of 19 base–24 base pairing, cleavage, dissociation of the cleaved fragments from the 19 base RNA, and pairing of the 19 base RNA with a new 24 base substrate. The 19 base RNA is therefore a ribozyme with endonuclease activity. The parameters of the reaction are similar to those of other RNA-catalyzed reactions (see **Figure 26.10**; for review see 258; 2270).

The crystal structure of a hammerhead shows that it forms a compact V-shape, in which the catalytic center lies in a turn, as indicated diagrammatically in **Figure 26.16**. An Mg^{2+} ion located in the catalytic site plays a crucial role in the reaction. It is positioned by the target cytidine and by the cytidine at the base of stem 1; it may also be connected to the adjacent uridine. It extracts a proton from the 2′–OH of the target cytidine, and then directly attacks the labile phosphodiester bond. Mutations in the hammerhead sequence that affect the transition state of the cleavage reaction occur in both the active site and other locations, suggesting that there may be a substantial rearrangement of structure prior to cleavage (758).

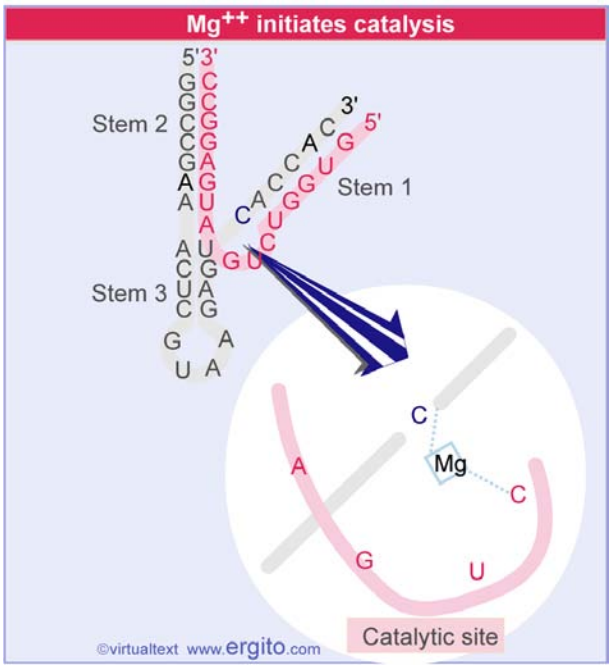

Figure 26.16 A hammerhead ribozyme forms a V-shaped tertiary structure in which stem 2 is stacked upon stem 3. The catalytic center lies between stem 2/3 and stem 1. It contains a magnesium ion that initiates the hydrolytic reaction.

It is possible to design enzyme-substrate combinations that can form hammerhead structures, and these have been used to demonstrate that introduction of the appropriate RNA molecules into a cell can allow the enzymatic reaction to occur *in vivo*. A ribozyme designed in this way essentially provides a highly specific restriction-like activity directed against an RNA target. By placing the ribozyme under control of a regulated promoter, it can be used in the same way as (for example) antisense constructs specifically to turn off expression of a target gene under defined circumstances.

Reviews

258. Symons, R. H. (1992). *Small catalytic RNAs*. Annu. Rev. Biochem. 61, 641-71.

2270. Doherty, E. A. and Doudna, J. A. (2000). *Ribozyme structures and mechanisms*. Annu. Rev. Biochem. 69, 597-615.

References

756. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme*. Cell 35, 849-857.

757. Forster, A. C. and Symons, R. H. (1987). *Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site*. Cell 50, 9-16.

758. Scott, W. G., Finch, J. T., and Klug, A. (1995). *The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage*. Cell 81, 991-1002.

CATALYTIC RNA

5.26.10 RNA editing occurs at individual bases

Key Terms

RNA editing describes a change of sequence at the level of RNA following transcription.

Key Concepts

- Apolipoprotein-B and glutamate receptors have site specific deaminations catalyzed by cytidine and adenosine deaminases that change the coding sequence.

A prime axiom of molecular biology is that the sequence of an mRNA can only represent what is coded in the DNA. The central dogma envisaged a linear relationship in which a continuous sequence of DNA is transcribed into a sequence of mRNA that is in turn directly translated into protein. The occurrence of interrupted genes and the removal of introns by RNA splicing introduces an additional step into the process of gene expression: the coding sequences (exons) in DNA must be reconnected in RNA. But the process remains one of information transfer, in which the actual coding sequence in DNA remains unchanged.

Changes in the information coded by DNA occur in some exceptional circumstances, most notably in the generation of new sequences coding for immunoglobulins in mammals and birds. These changes occur specifically in the somatic cells (B lymphocytes) in which immunoglobulins are synthesized (see *Molecular Biology 5.25 Immune diversity*). New information is generated in the DNA of an individual during the process of reconstructing an immunoglobulin gene; and information coded in the DNA is changed by somatic mutation. The information in DNA continues to be faithfully transcribed into RNA.

RNA editing is a process in which *information changes at the level of mRNA.* It is revealed by situations in which the coding sequence in an RNA differs from the sequence of DNA from which it was transcribed. RNA editing occurs in two different situations, with different causes. In mammalian cells there are cases in which a substitution occurs in an individual base in mRNA, causing a change in the sequence of the protein that is coded. In trypanosome mitochondria, more widespread changes occur in transcripts of several genes, when bases are systematically added or deleted.

Figure 26.17 summarizes the sequences of the apolipoprotein-B gene and mRNA in mammalian intestine and liver. The genome contains a single (interrupted) gene whose sequence is identical in all tissues, with a coding region of 4563 codons. This gene is transcribed into an mRNA that is translated into a protein of 512 kD representing the full coding sequence in the liver.

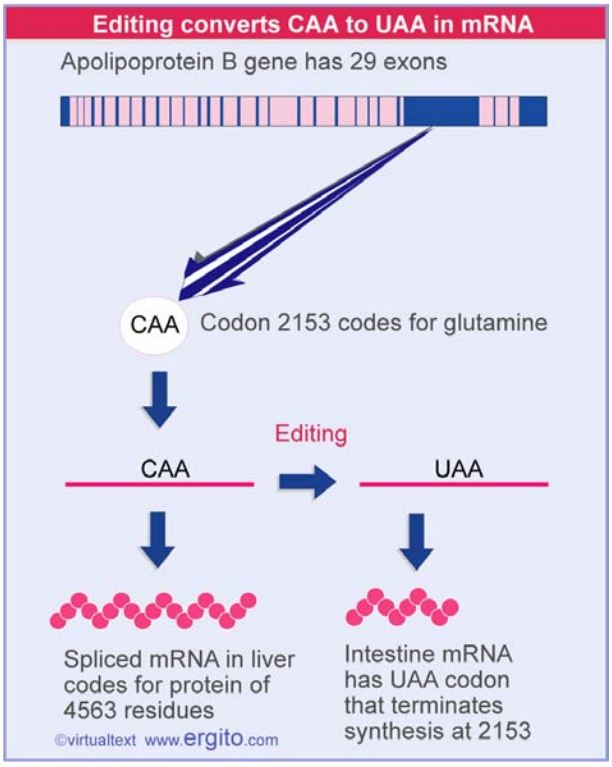

Figure 26.17 The sequence of the apo-B gene is the same in intestine and liver, but the sequence of the mRNA is modified by a base change that creates a termination codon in intestine.

A shorter form of the protein, ~250 kD, is synthesized in intestine. This protein consists of the N-terminal half of the full-length protein. It is translated from an mRNA whose sequence is identical with that of liver except for a change from C to U at codon 2153. This substitution changes the codon CAA for glutamine into the ochre codon UAA for termination (759).

What is responsible for this substitution? No alternative gene or exon is available in the genome to code for the new sequence, and no change in the pattern of splicing can be discovered. We are forced to conclude that a change has been made directly in the sequence of the transcript.

Another example is provided by glutamate receptors in rat brain. Editing at one position changes a glutamine codon in DNA into a codon for arginine in RNA; the change affects the conductivity of the channel and therefore has an important effect on controlling ion flow through the neurotransmitter. At another position in the receptor, an arginine codon is converted to a glycine codon (760).

The editing event in apo-B causes C_{2153} to be changed to U; both changes in the glutamate receptor are from A to I (inosine). These events are *deaminations* in which the amino group on the nucleotide ring is removed. Such events are catalyzed by enzymes called cytidine and adenosine deaminases, respectively. This type of editing appears to occur largely in the nervous system. There are 16 (potential) targets for cytosine deaminase in *D. melanogaster*, and all are genes involved in neurotransmission. In many cases, the editing event changes an amino acid at a

functionally important position in the protein.

What controls the specificity of an editing reaction? Enzymes that undertake deamination as such often have broad specificity – for example, the best characterized adenosine deaminase acts on any A residue in a duplex RNA region. Editing enzymes are related to the general deaminases, but have other regions or additional subunits that control their specificity. In the case of apoB editing, the catalytic subunit of an editing complex is related to bacterial cytidine deaminase, but has an additional RNA-binding region that helps to recognize the specific target site for editing. A special adenosine deaminase enzyme recognizes the target sites in the glutamate receptor RNA, and similar events occur in a serotonin receptor RNA (762).

The complex may recognize a particular region of secondary structure in a manner analogous to tRNA-modifying enzymes or could directly recognize a nucleotide sequence. The development of an *in vitro* system for the apo-B editing event suggests that a relatively small sequence (~26 bases) surrounding the editing site provides a sufficient target. **Figure 26.18** shows that in the case of the GluR-B RNA, a base-paired region that is necessary for recognition of the target site is formed between the edited region in the exon and a complementary sequence in the downstream intron. A pattern of mispairing within the duplex region is necessary for specific recognition. So different editing systems may have different types of requirement for sequence specificity in their substrates (761).

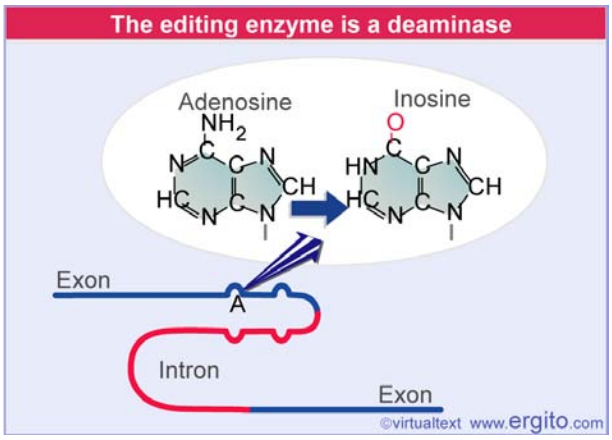

Figure 26.18 Editing of mRNA occurs when a deaminase acts on an adenine in an imperfectly paired RNA duplex region.

Last updated on 9-10-2003

References

759. Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., and Scott, J. (1987). *A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine*. Cell 50, 831-840.

760. Sommer, B. et al. (1991). *RNA editing in brain controls a determinant of ion flow in glutamate-gated channels*. Cell 67, 11-19.

761. Higuchi, M. et al. (1993). *RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency*. Cell 75, 1361-1370.

762. Navaratnam, N. et al. (1995). *Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site*. Cell 81, 187-195.

CATALYTIC RNA

5.26.11 RNA editing can be directed by guide RNAs

Key Terms

A **guide RNA** is a small RNA whose sequence is complementary to the sequence of an RNA that has been edited. It is used as a template for changing the sequence of the pre-edited RNA by inserting or deleting nucleotides.

Key Concepts

- Extensive RNA editing in trypanosome mitochondria occurs by insertions or deletions of uridine.
- The substrate RNA base pairs with a guide RNA on both sides of the region to be edited.
- The guide RNA provides the template for addition (or less often deletion) of uridines.
- Editing is catalyzed by a complex of endonuclease, terminal uridyltransferase activity, and RNA ligase.

Another type of editing is revealed by dramatic changes in sequence in the products of several genes of trypanosome mitochondria. In the first case to be discovered, the sequence of the cytochrome oxidase subunit II protein has a frameshift relative to the sequence of the *coxII* gene. The sequences of the gene and protein given in **Figure 26.19** are conserved in several trypanosome species. How does this gene function?

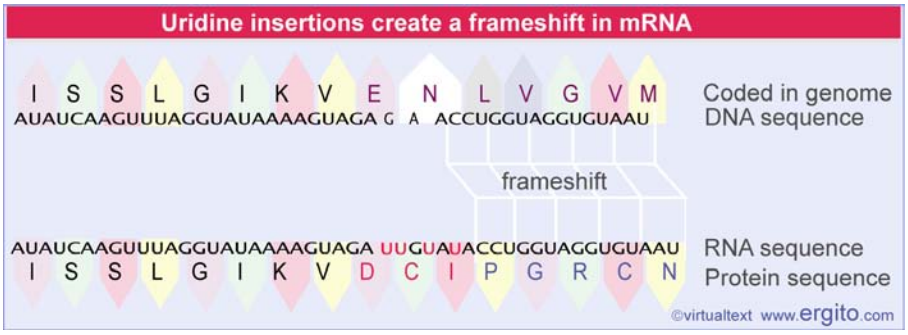

Figure 26.19 The mRNA for the trypanosome *coxII* gene has a frameshift relative to the DNA; the correct reading frame is created by the insertion of 4 uridines.

The *coxII* mRNA has an insert of an additional four nucleotides (all uridines) around the site of frameshift (see *Great Experiments 4.7 RNA editing*). The insertion restores the proper reading frame; it inserts an extra amino acid and changes the amino acids on either side. No second gene with this sequence can be discovered, and we are forced to conclude that the extra bases are inserted during or after transcription (764). A similar discrepancy between mRNA and genomic sequences is found in genes of the SV5 and measles paramyxoviruses, in these cases involving the addition of G

residues in the mRNA.

Similar editing of RNA sequences occurs for other genes, and includes deletions as well as additions of uridine. The extraordinary case of the *coxIII* gene of *T. brucei* is summarized in **Figure 26.20**.

Figure 26.20 Part of the mRNA sequence of *T. brucei coxIII* shows many uridines that are not coded in the DNA (shown in red) or that are removed from the RNA (shown as T).

More than half of the residues in the mRNA consist of uridines that are not coded in the gene. Comparison between the genomic DNA and the mRNA shows that no stretch longer than 7 nucleotides is represented in the mRNA without alteration; and runs of uridine up to 7 bases long are inserted (1277).

What provides the information for the specific insertion of uridines? A **guide RNA** contains a sequence that is complementary to the correctly edited mRNA. **Figure 26.21** shows a model for its action in the cytochrome *b* gene of *Leishmania* (1280).

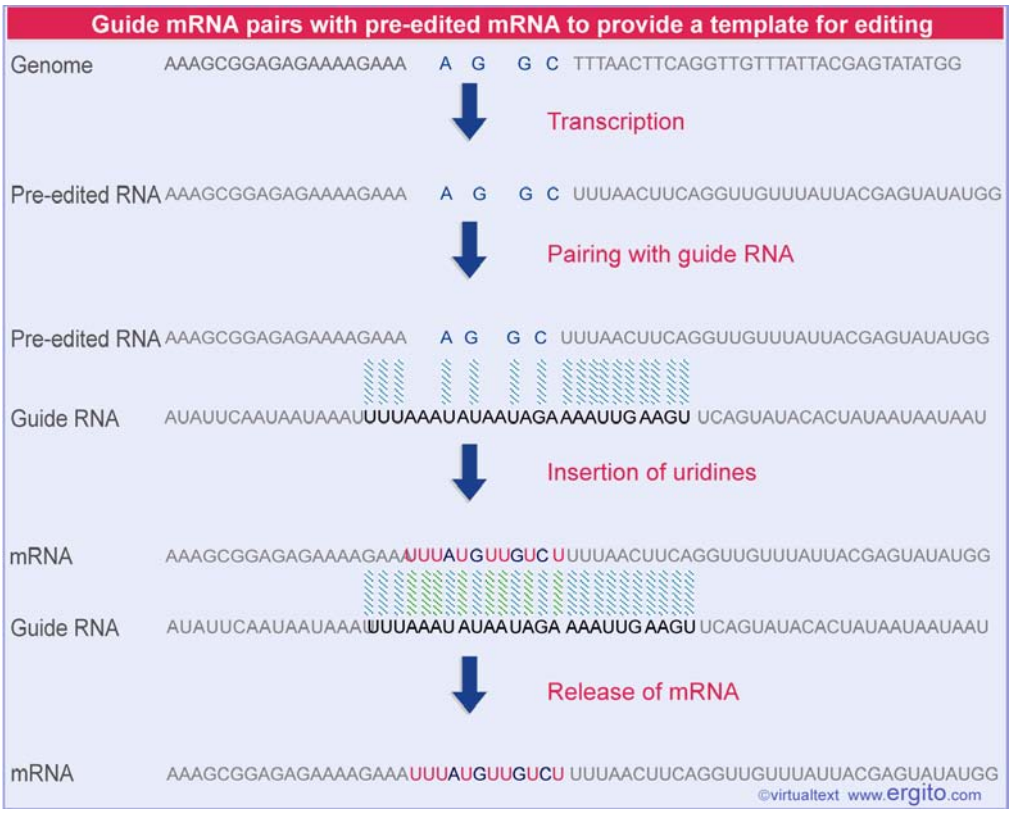

Figure 26.21 Pre-edited RNA base pairs with a guide RNA on both sides of the region to be edited. The guide RNA provides a template for the insertion of uridines. The mRNA produced by the insertions is complementary to the guide RNA.

The sequence at the top shows the original transcript, or pre-edited RNA. Gaps show where bases will be inserted in the editing process. 8 uridines must be inserted into this region to create the valid mRNA sequence.

The guide RNA is complementary to the mRNA for a significant distance including and surrounding the edited region. Typically the complementarity is more extensive on the 3′ side of the edited region and is rather short on the 5′ side. Pairing between the guide RNA and the pre-edited RNA leaves gaps where unpaired A residues in the guide RNA do not find complements in the pre-edited RNA. The guide RNA provides a template that allows the missing U residues to be inserted at these positions. When the reaction is completed, the guide RNA separates from the mRNA, which becomes available for translation.

Specification of the final edited sequence can be quite complex; in this example, a lengthy stretch of the transcript is edited by the insertion altogether of 39 U residues, and this appears to require two guide RNAs that act at adjacent sites. The first guide RNA pairs at the 3′–most site, and the edited sequence then becomes a substrate for further editing by the next guide RNA.

The guide RNAs are encoded as independent transcription units. **Figure 26.22** shows a map of the relevant region of the *Leishmania* mitochondrial DNA. It includes the "gene" for cytochrome *b*, which codes for the pre-edited sequence, and two regions

that specify guide RNAs. Genes for the major coding regions and for their guide RNAs are interspersed.

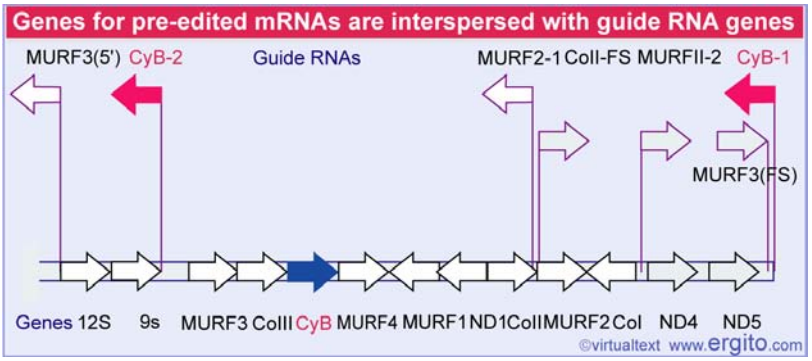

Figure 26.22 The *Leishmania* genome contains genes coding for pre-edited RNAs interspersed with units that code for the guide RNAs required to generate the correct mRNA sequences. Some genes have multiple guide RNAs.

In principle, a mutation in either the "gene" or one of its guide RNAs could change the primary sequence of the mRNA, and thus of the protein. By genetic criteria, each of these units could be considered to comprise part of the "gene." Since the units are independently expressed, they should of course complement in *trans.* If mutations were available, we should therefore find that 3 complementation groups were needed to code for the primary sequence of a single protein.

The characterization of intermediates that are partially edited suggests that the reaction proceeds along the pre-edited RNA in the 3′–5′ direction. The guide RNA determines the specificity of uridine insertions by its pairing with the pre-edited RNA.

Editing of uridines is catalyzed by a 20S enzyme complex that contains an endonuclease, a terminal uridyltransferase (TUTase), and an RNA ligase, as illustrated in **Figure 26.23** (2399). It binds the guide RNA and uses it to pair with the pre-edited mRNA. The substrate RNA is cleaved at a site that is (presumably) identified by the absence of pairing with the guide RNA, a uridine is inserted or deleted to base pair with the guide RNA, and then the substrate RNA is ligated. UTP provides the source for the uridyl residue. It is added by the TUTase activity; it is not clear whether this activity, or a separate exonuclease, is responsible for deletion. [At one time it was thought that a stretch of U residues at the end of guide RNA might provide the source for added U residues or a sink for deleted residues, but transfer of U residues to guide RNAs appears to be an aberrant reaction that is not responsible for editing (767).]

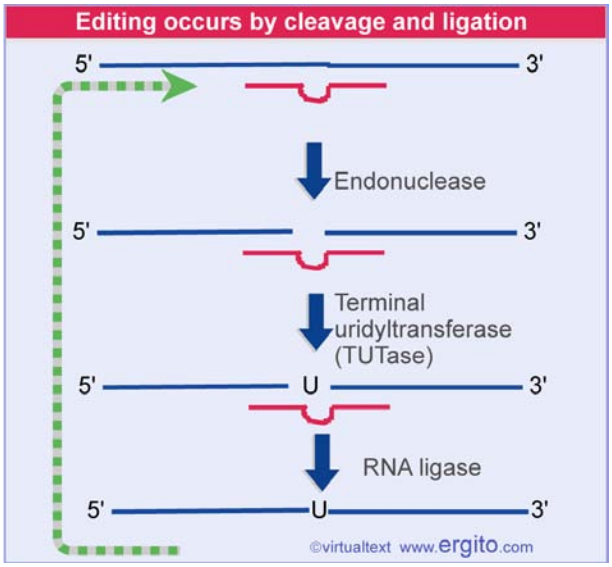

Figure 26.23 Addition or deletion of U residues occurs by cleavage of the RNA, removal or addition of the U, and ligation of the ends. The reactions are catalyzed by a complex of enzymes under the direction of guide RNA.

The structures of partially edited molecules suggest that the U residues are added one a time, and not in groups. It is possible that the reaction proceeds through successive cycles in which U residues are added, tested for complementarity with the guide RNA, retained if acceptable and removed if not, so that the construction of the correct edited sequence occurs gradually. We do not know whether the same types of reaction are involved in editing reactions that add C residues.

Last updated on 4-17-2002

References

764. Benne, R., Van den Burg J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H., and Tromp, M. C. (1986). *Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA.* Cell 46, 819-826.

767. Seiwert, S. D., Heidmann, S. and Stuart, K. (1996). *Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid editing*. Cell 84, 831-841.

1277. Feagin, J. E., Abraham, J. M., and Stuart, K. (1988). *Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei.* Cell 53, 413-422.

1280. Blum, B., Bakalara, N., and Simpson, L. (1990). *A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information.* Cell 60, 189-198.

2399. Aphasizhev, R., Sbicego, S., Peris, M., Jang, S. H., Aphasizheva, I., Simpson, A. M., Rivlin, A., and Simpson, L. (2002). *Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing.* Cell 108, 637-648.

CATALYTIC RNA

5.26.12 Protein splicing is autocatalytic

Key Terms

Protein splicing is the autocatalytic process by which an intein is removed from a protein and the exteins on either side become connected by a standard peptide bond.

Extein sequences remain in the mature protein that is produced by processing a precursor via protein splicing.

An **intein** is the part that is removed from a protein that is processed by protein splicing.

Key Concepts

- An intein has the ability to catalyze its own removal from a protein in such a way that the flanking exteins are connected.
- Protein splicing is catalyzed by the intein.
- Most inteins have two independent activities: protein splicing and a homing endonuclease.

Protein splicing has the same effect as RNA splicing: a sequence that is represented within the gene fails to be represented in the protein. The parts of the protein are named by analogy with RNA splicing: **exteins** are the sequences that are represented in the mature protein, and **inteins** are the sequences that are removed. The mechanism of removing the intein is completely different from RNA splicing. **Figure 26.24** shows that the gene is translated into a protein precursor that contains the intein, and then the intein is excised from the protein. About 100 examples of protein splicing are known, spread through all classes of organisms. The typical gene whose product undergoes protein splicing has a single intein.

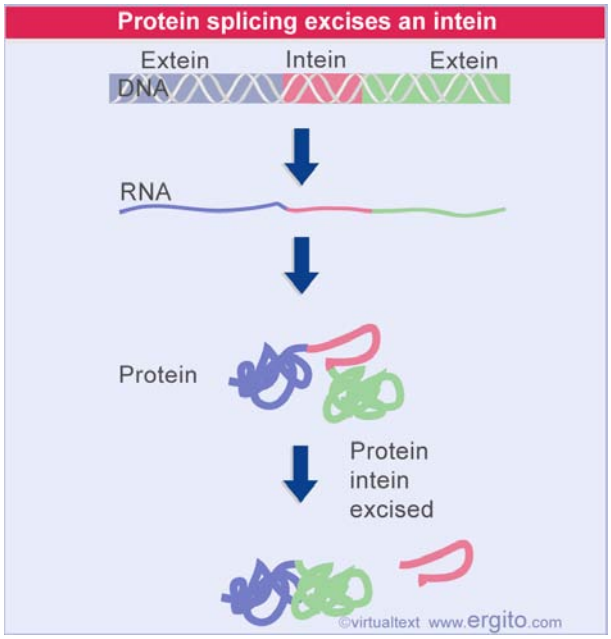

Figure 26.24 In protein splicing the exteins are connected by removing the intein from the protein.

The first intein was discovered in an archaeal DNA polymerase gene in the form of an intervening sequence in the gene that does not conform to the rules for introns (2327). Then it was demonstrated that the purified protein can splice this sequence out of itself in an autocatalytic reaction (2328). The reaction does not require input of energy and occurs through the series of bond rearrangements shown in **Figure 26.25**. It is a function of the intein, although its efficiency can be influenced by the exteins.

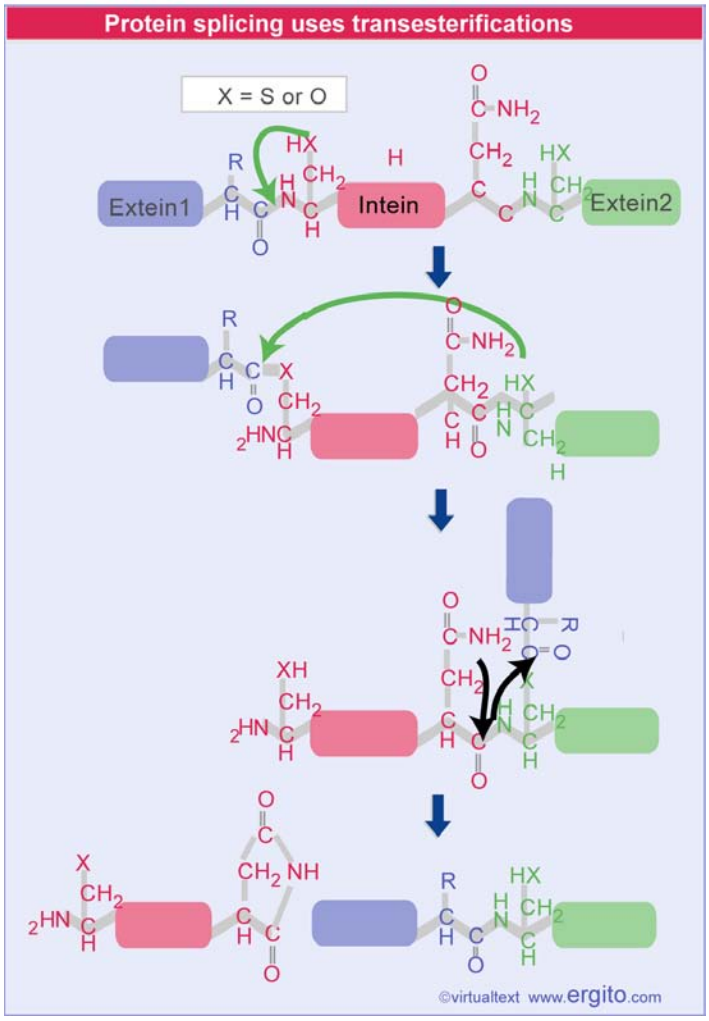

Figure 26.25 Bonds are rearranged through a series of transesterifications involving the -OH groups of serine or proline or the -SH group of cysteine until finally the exteins are connected by a peptide bond and the intein is released with a circularized C-terminus.

The first reaction is an attack by an -OH or -SH side chain of the first amino acid in the intein on the peptide bond that connects it to the first extein. This transfers the extein from the amino-terminal group of the intein to an N-O or N-S acyl connection. Then this bond is attacked by the -OH or -SH side chain of the first amino acid in the second extein. The result is to transfer extein1 to the side chain of the amino-terminal acid of extein2. Finally, the C-terminal asparagine of the intein cyclizes, and the terminal NH of extein2 attacks the acyl bond to replace it with a conventional peptide bond. Each of these reactions can occur spontaneously at very low rates, but their occurrence in a coordinate manner rapidly enough to achieve protein splicing requires catalysis by the intein (for review see 2267).

Inteins have characteristic features. They are found as in-frame insertions into coding sequences. They can be recognized as such because of the existence of homologous genes that lack the insertion. They have an N-terminal serine or cysteine (to provide the -XH side chain) and a C-terminal asparagine. A typical intein has a sequence of

~150 amino acids at the N-terminal end and ~50 amino acids at the C-terminal end that are involved in catalyzing the protein splicing reaction. The sequence in the center of the intein can have other functions.

An extraordinary feature of many inteins is that they have homing endonuclease activity. A homing endonuclease cleaves a target DNA to create a site into which the DNA sequence coding for the intein can be inserted (see **Figure 26.11** in *Molecular Biology 5.26.5 Some group I introns code for endonucleases that sponsor mobility*). The protein splicing and homing endonuclease activities of an intein are independent (2329).

We do not really understand the connection between the presence of both these activities in an intein, but two types of model have been suggested. One is to suppose that there was originally some sort of connection between the activities, but that they have since become independent and some inteins have lost the homing endonuclease. The other is to suppose that inteins may have originated as protein splicing units, most of which (for unknown reasons) were subsequently invaded by homing endonucleases. This is consistent with the fact that homing endonucleases appear to have invaded other types of units also, including most notably group I introns.

Last updated on 2-6-2002

Reviews

2267. Paulus, H. (2000). *Protein splicing and related forms of protein autoprocessing.* Annu. Rev. Biochem. 69, 447-496.

References

2327. Perler, F. B. et al. (1992). *Intervening sequences in an Archaea DNA polymerase gene.* Proc. Natl. Acad. Sci. USA 89, 5577-5581.

2328. Xu, M. Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993). *in vitro protein splicing of purified precursor and the identification of a branched intermediate.* Cell 75, 1371-1377.

2329. Derbyshire, V., Wood, D. W., Wu, W., Dansereau, J. T., Dalgaard, J. Z., and Belfort, M. (1997). *Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution.* Proc. Natl. Acad. Sci. USA 94, 11466-11471.

CATALYTIC RNA

5.26.13 Summary

Self-splicing is a property of two groups of introns, which are widely dispersed in lower eukaryotes, prokaryotic systems, and mitochondria. The information necessary for the reaction resides in the intron sequence (although the reaction is actually assisted by proteins *in vivo*). For both group I and group II introns, the reaction requires formation of a specific secondary/tertiary structure involving short consensus sequences. Group I intron RNA creates a structure in which the substrate sequence is held by the IGS region of the intron, and other conserved sequences generate a guanine nucleotide binding site. It occurs by a transesterification involving a guanosine residue as cofactor. No input of energy is required. The guanosine breaks the bond at the 5′ exon-intron junction and becomes linked to the intron; the hydroxyl at the free end of the exon then attacks the 3′ exon-intron junction. The intron cyclizes and loses the guanosine and the terminal 15 bases. A series of related reactions can be catalyzed via attacks by the terminal G-OH residue of the intron on internal phosphodiester bonds. By providing appropriate substrates, it has been possible to engineer ribozymes that perform a variety of catalytic reactions, including nucleotidyl transferase activities.

Some group I and some group II mitochondrial introns have open reading frames. The proteins coded by group I introns are endonucleases that make double-stranded cleavages in target sites in DNA; the cleavage initiates a gene conversion process in which the sequence of the intron itself is copied into the target site. The proteins coded by group II introns include an endonuclease activity that initiates the transposition process, and a reverse transcriptase that enables an RNA copy of the intron to be copied into the target site. These types of introns probably originated by insertion events. The proteins coded by both groups of introns may include maturase activities that assist splicing of the intron by stabilizing the formation of the secondary/tertiary structure of the active site.

Catalytic reactions are undertaken by the RNA component of the RNAase P ribonucleoprotein. Virusoid RNAs can undertake self-cleavage at a "hammerhead" structure. Hammerhead structures can form between a substrate RNA and a ribozyme RNA, allowing cleavage to be directed at highly specific sequences. These reactions support the view that RNA can form specific active sites that have catalytic activity.

RNA editing changes the sequence of an RNA after or during its transcription. The changes are required to create a meaningful coding sequence. Substitutions of individual bases occur in mammalian systems; they take the form of deaminations in which C is converted to U, or A is converted to I. A catalytic subunit related to cytidine or adenosine deaminase functions as part of a larger complex that has specificity for a particular target sequence.

Additions and deletions (most usually of uridine) occur in trypanosome mitochondria and in paramyxoviruses. Extensive editing reactions occur in trypanosomes in which as many as half of the bases in an mRNA are derived from editing. The editing reaction uses a template consisting of a guide RNA that is complementary to the mRNA sequence. The reaction is catalyzed by an enzyme complex that includes an

endonuclease, terminal uridyltransferase, and RNA ligase, using free nucleotides as the source for additions, or releasing cleaved nucleotides following deletion.

Protein splicing is an autocatalytic reaction that occurs by bond transfer reactions and input of energy is not required. The intein catalyzes its own splicing out of the flanking exteins. Many inteins have a homing endonuclease activity that is independent of the protein splicing activity.