Information Processing: RNA Processing

So far, we have looked at the mechanism by which the information in genes (DNA) is transcribed into RNA. The newly made RNA, also known as the primary transcript is further processed before it is functional. Both prokaryotes and eukaryotes process their ribosomal and transfer RNAs.

mRNA processing

The major difference in RNA processing, however, between prokaryotes and eukaryotes, is in the processing of messenger RNAs. We will focus on the processing of mRNAs in this section. You will recall that in bacterial cells, the mRNA is translated directly as it comes off the DNA template. In eukaryotic cells, RNA synthesis, which occurs in the nucleus, is separated from the protein synthesis machinery, which is in the cytoplasm. The initial product of transcription of an mRNA is sometimes referred to as the pre-mRNA. After it has been processed and is ready to be exported from the nucleus, it is called the mature

YouTube Lectures by Kevin
HERE & HERE

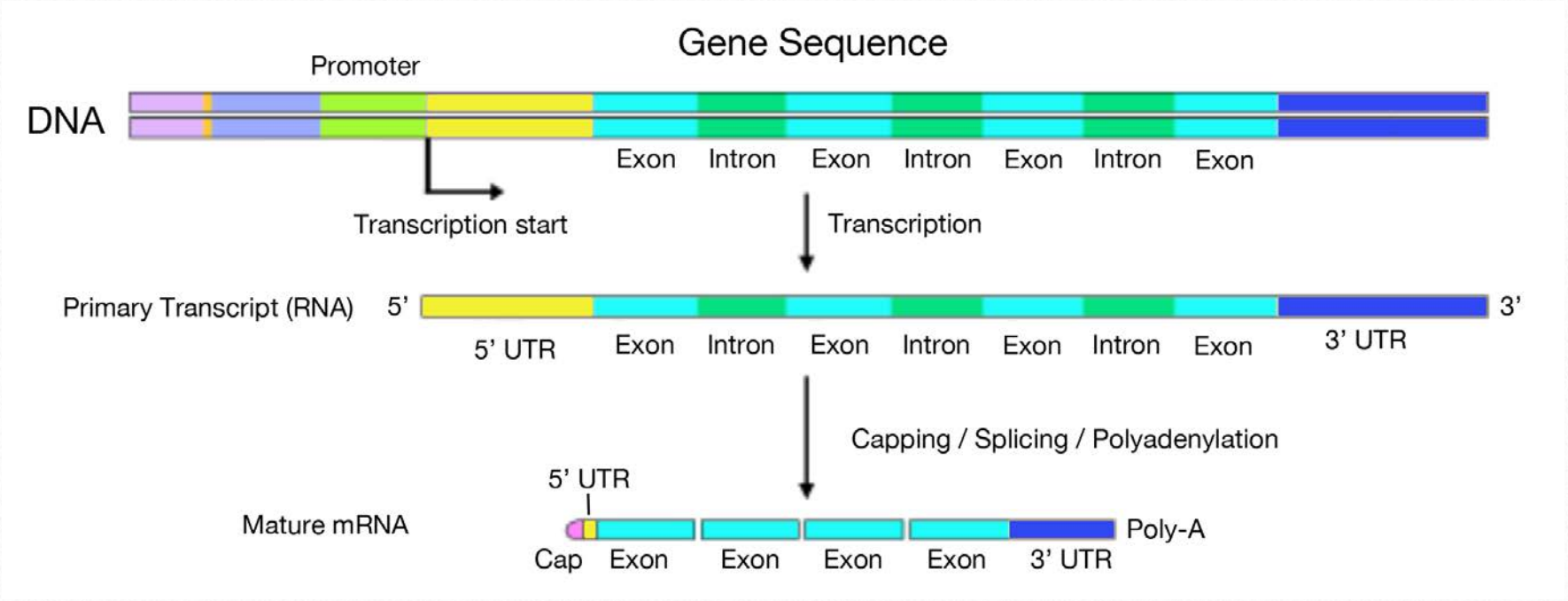

Figure 7.67 - Steps in processing of pre-mRNA

mRNA. The three main processing steps for mRNAs are (Figure 7.67):

- Capping at the 5' end
- Splicing to remove introns
- Addition of a polyA tail at the 3' end.

Although this description suggests that these processing steps occur post-transcriptionally, after the entire gene has been transcribed, there is evidence that processing occurs co-transcriptionally. That is, the steps of processing are occurring as the mRNA is being made. Proteins involved in mRNA processing have been shown to be associated with the phosphorylated C-terminal domain (CTD) of RNA polymerase II.

Capping

As might be expected, the addition of an mRNA cap at the 5' end is the first step in mRNA processing, since the 5'end of the RNA is the first to be made. Capping occurs once the first 20-30 nucleotides of the RNA

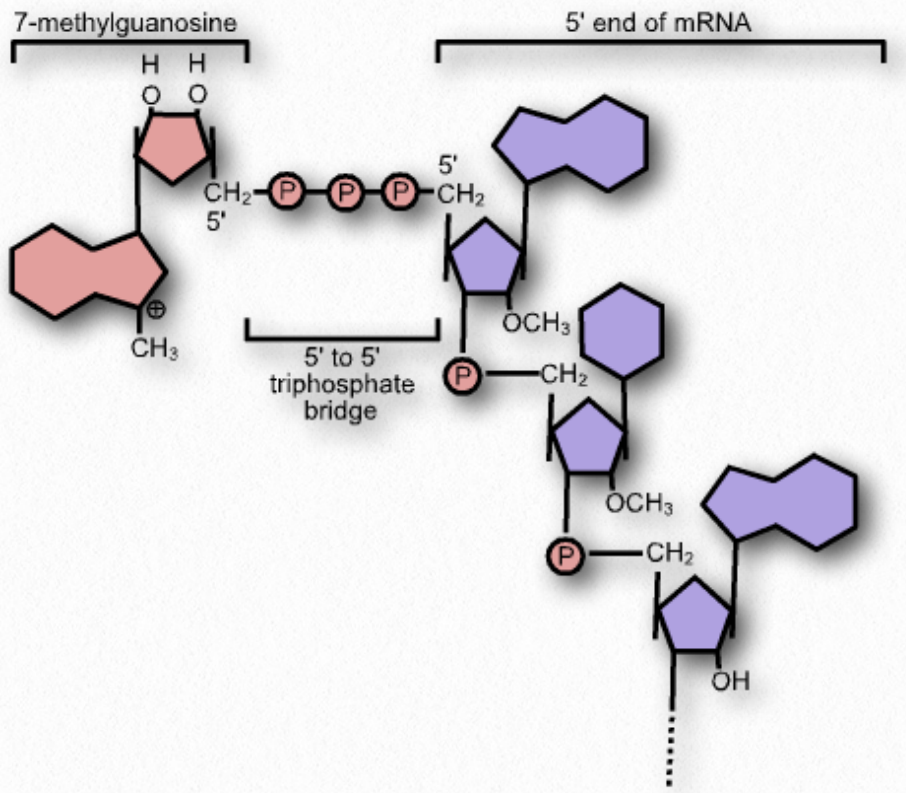

Figure 7.68 - 5' capping of eukaryotic mRNAs
Wikipedia

have been synthesized. The addition of the cap involves removal of a phosphate from the first nucleotide in the RNA to generate a diphosphate. This is then joined to a guanosine monophosphate which is subsequently methylated at N^7 of the guanine to form the 7mG cap structure (Figure 7.68). This cap is recognized and bound by a complex of proteins that remain associated with the cap till the mRNA has been transported into the cytoplasm. The cap protects the 5' end of the mRNA from degradation by nucleases and also helps to position the mRNA correctly on the ribosomes during protein synthesis.

mRNA is sent out of the nucleus to be used to direct protein synthesis.

Intron removal

Introns are removed from the pre-mRNA by the activity of a complex called the spliceosome. The spliceosome is made up of proteins and small RNAs that are associated to form protein-RNA enzymes called small nuclear ribonucleoproteins or snRNPs (pronounced snurps).

Splice junctions

The splicing machinery must be able to recognize splice junctions (i.e., where each exon ends and its associated intron begins) in order to correctly cut out the introns and join the exons to make the mature, spliced mRNA. What signals indicate exon-intron boundaries? The junctions between exons and introns are indicated by specific base sequences. The consensus sequence at the 5' exon-intron junction (also called the 5' splice site) is AG-

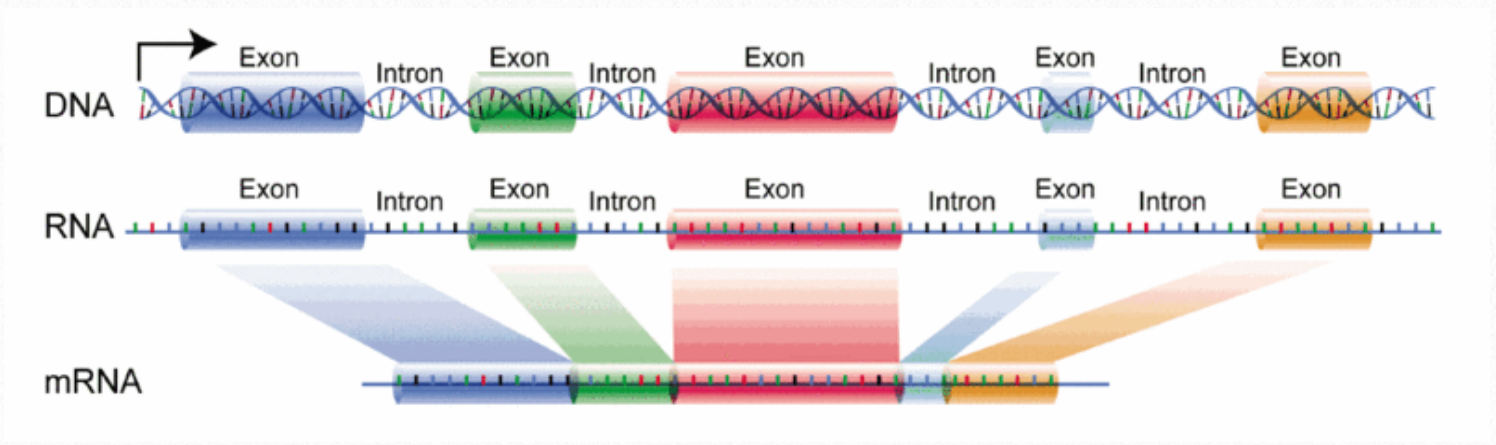

Figure 7.69 - Removal of introns from the primary transcript

Splicing

Eukaryotic genes have introns, noncoding regions that interrupt the gene. The mRNA copied from genes containing introns will also therefore have noncoding regions that interrupt the information in the gene. These noncoding regions must be removed (Figure 7.69) before the

Interactive Learning Module HERE

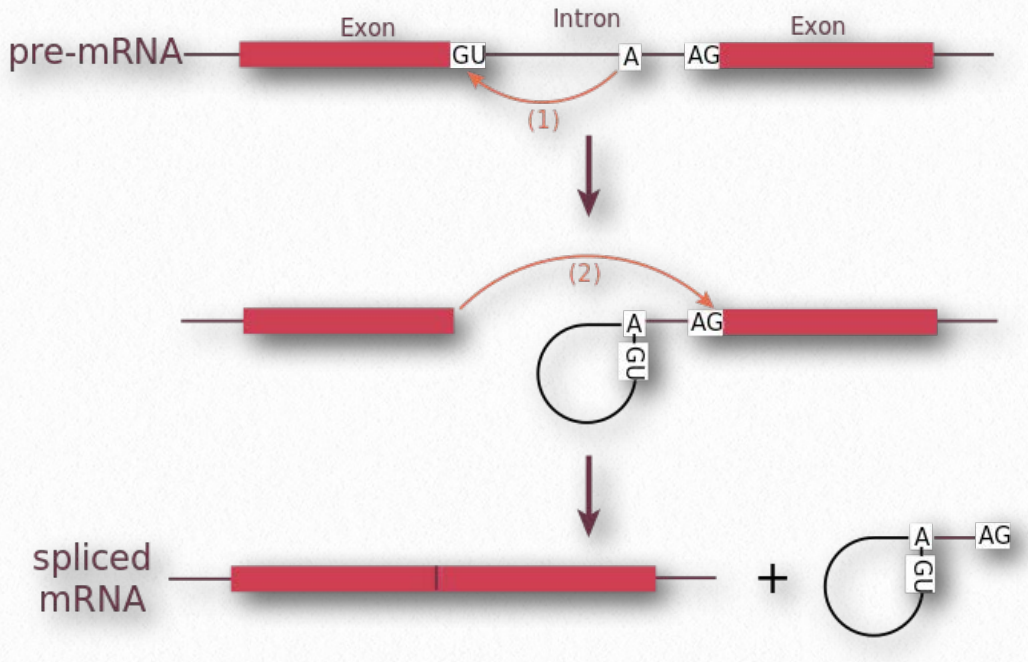

Figure 7.70 - Splicing of introns

Wikipedia

GURAGU. In this sequence, the intron starts with the second G (R stands for any purine). The 3' splice junction has the consensus sequence YAGRNNN, where YAG is within the intron, and RNNN is part of the exon (Y stands for any pyrimidine, and N for any nucleotide).

There is also a third important sequence within the intron, about a hundred nucleotides from the 3' splice site, called a branch point or branch site, that is important for splicing. This site is defined by the presence of an A followed by a string of pyrimidines. The importance of this site will be seen when we consider the steps of splicing.

Splicing mechanism

There are two main steps in splicing. The first step is the nucleophilic attack by the 2'OH of the branch point A on the 5' splice site (the junction of the 5' exon and the intron). As a result of a *trans*-esterification reaction, the 5' exon is released, and a lariat-shaped molecule composed of the 3' exon and the intron sequence is generated (Figure 7.70). In the second step, the 3' OH of the 5' exon attacks the 3' splice site, and the two exons are joined together, and the lariat-shaped intron is released.

Spliceosome

As mentioned earlier, splicing is carried out by a complex consisting of small RNAs and proteins. The five small RNAs crucial to this complex, U1, U2, U4, U5 and U6 are found associated with proteins, as snRNPs. These and many other proteins work together to facilitate splicing. Although many details remain to be worked out, it appears that components of the splicing machinery associate with the CTD of the RNA polymerase and that this association is important for efficient splicing. The assembly of the spliceosome requires the stepwise interaction of the various snRNPs and other splicing factors (Figure

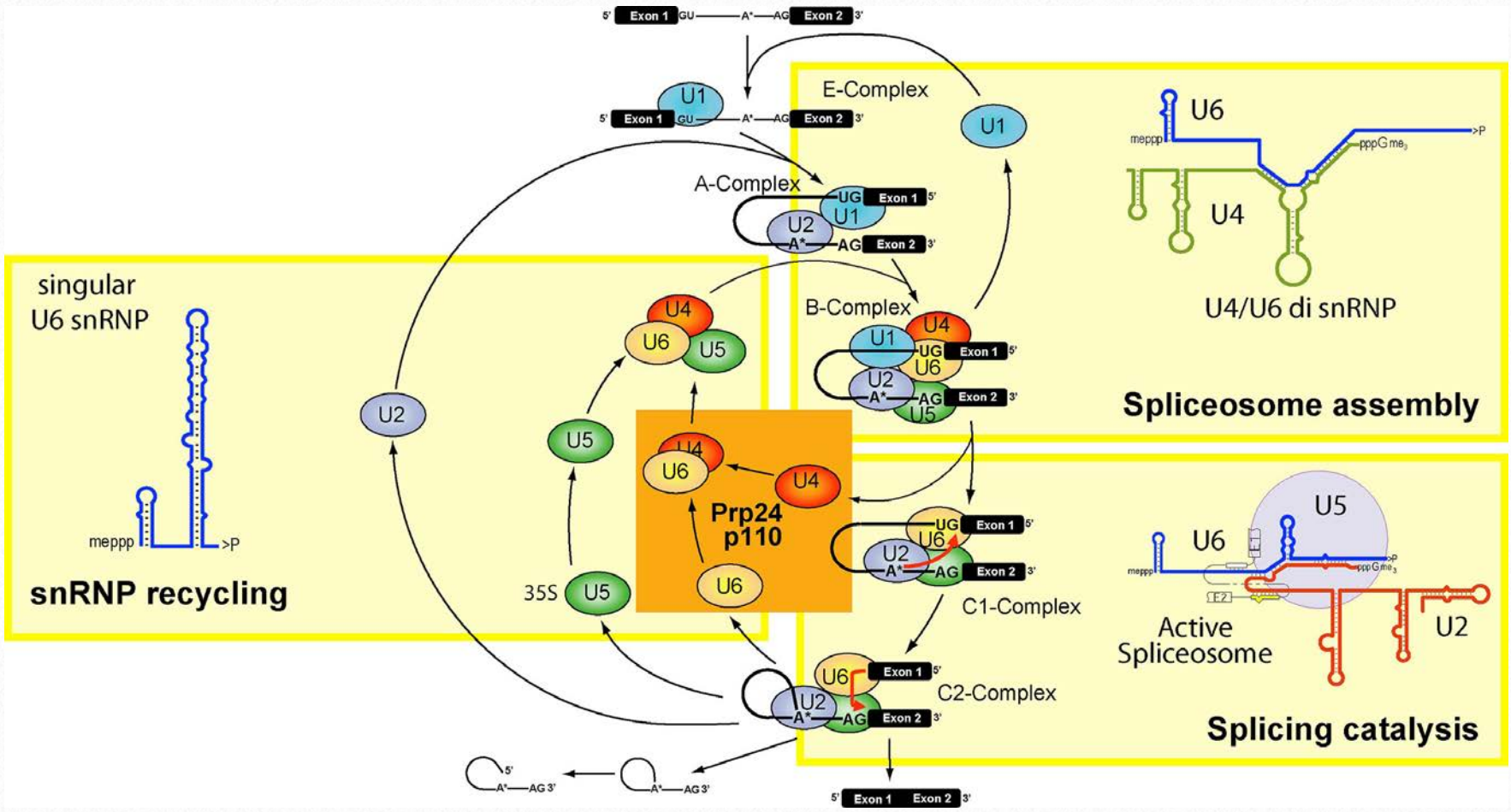

Figure 7.71 - Assembly of the spliceosome complex

Wikipedia

7.71). The initial step in this process is the interaction of the U1 snRNP with the 5' splice site. Additional proteins such as U2AF (AF = associated factor) are also loaded onto the pre-mRNA near the branch site. This is followed by the binding of the U2 snRNA to the branch site.

Next, a complex of the U4/U6 and U5 snRNPs is recruited to the spliceosome to generate a pre-catalytic complex. This complex undergoes rearrangements that alter RNA-RNA and protein-RNA interactions, resulting in displacement of the U4 and U1 snRNPs and the formation of the catalytically active spliceosome. This complex then carries out the two splicing steps described earlier.

YouTube Lectures by Kevin HERE & HERE

Alternative splicing

On average, human genes have about 9 exons each. However, the mature mRNAs from a gene containing nine exons may not include all of them. This is because the exons in a pre-mRNA can be spliced together in different combinations to generate different mature mRNAs. This is called alternative splicing, and allows the production of many different proteins using relatively few genes, since a single RNA with many exons can, by combining different exons dur-

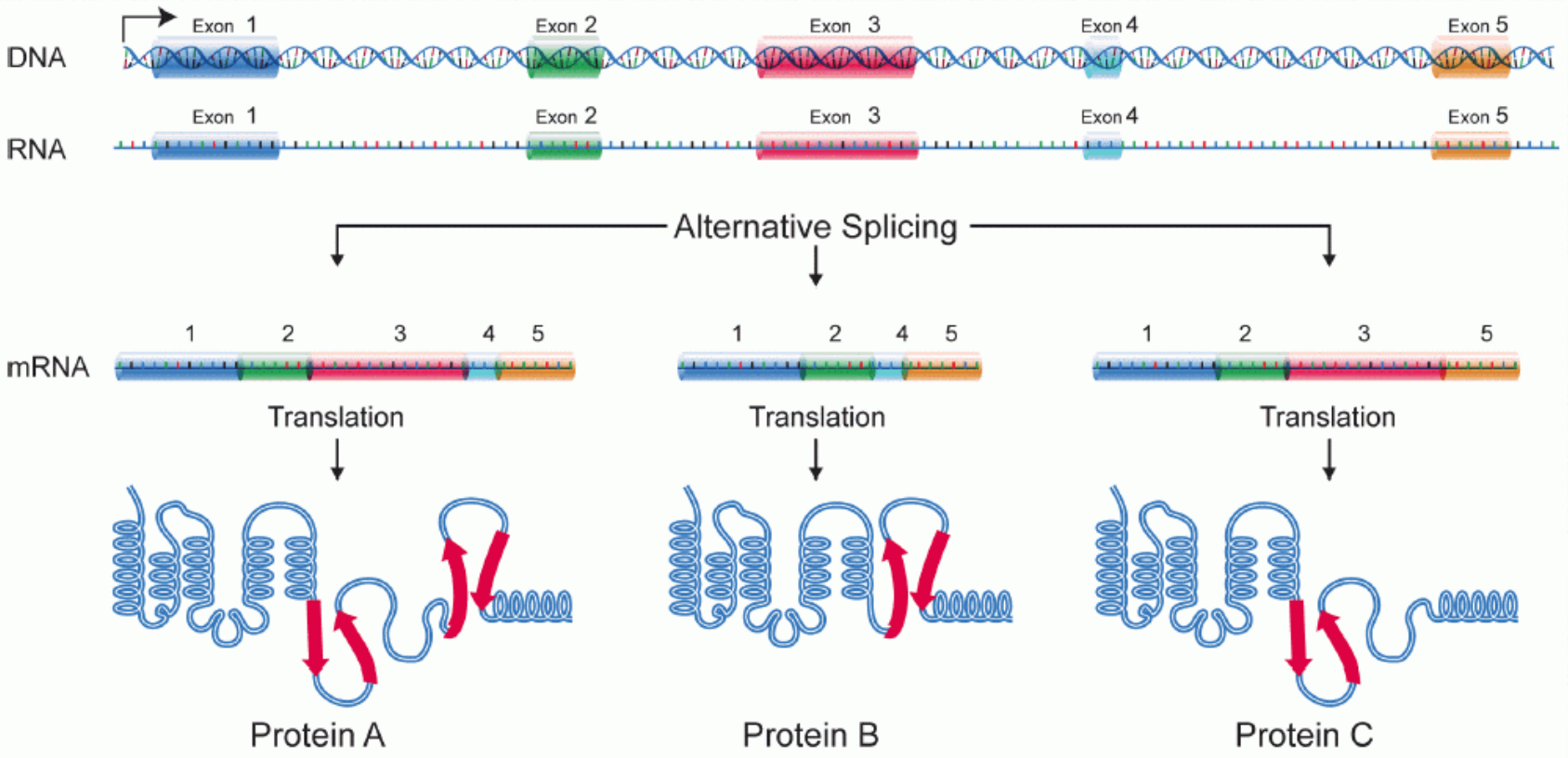

Figure 7.72 - Alternative splicing leads to different forms of a protein from the same gene sequence

ing splicing, create many different protein coding messages. Because of alternative splicing, each gene in our DNA gives rise, on average, to three different proteins. Alternative splicing allows the information in a single gene to be used to specify different proteins in different cell types or at different developmental stages (Figure 7.72).

Polyadenylation

The 3' end of a processed eukaryotic mRNA typically has a "poly(A) tail" consisting of about 200 adenine-containing nucleotides. These residues are added by a template-independent enzyme, poly(A)polymerase, following cleavage of

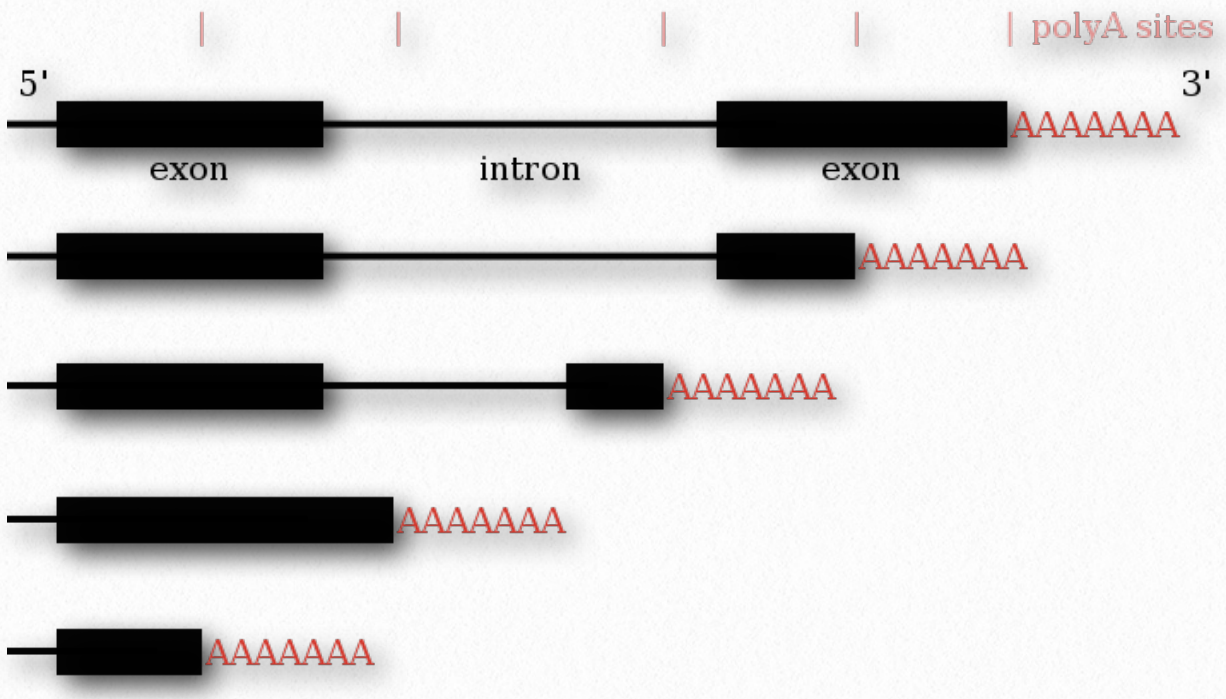

Figure 7.73 - Alternative poly-adenylation sites for a gene

the RNA at a site near the 3' end of the new transcript. Components of the polyadenylation machinery have been shown to be associated with the CTD of the RNA polymerase, showing that all three steps of pre-mRNA processing are tightly linked to transcription. There is evidence that the polyA tail plays a role in efficient translation of the mRNA, as well as in the stability of the mRNA. Like alternative splice sites, genes can have alternative polyA sites as well (Figure 7.73).

The cap and the polyA tail on an mRNA are also indications that the mRNA is complete (i.e., not defective). Once protein-coding messages have been processed by capping, splicing and addition of a poly A tail, they are transported out of the nucleus to be translated in the cytoplasm. Mature mRNAs are sent into the cytoplasm bound to export proteins that interact with the nuclear pore complexes in the nuclear envelope (Figure 7.74). Once the mature mRNA has been translocated to the cytoplasm, it is ready to be translated.

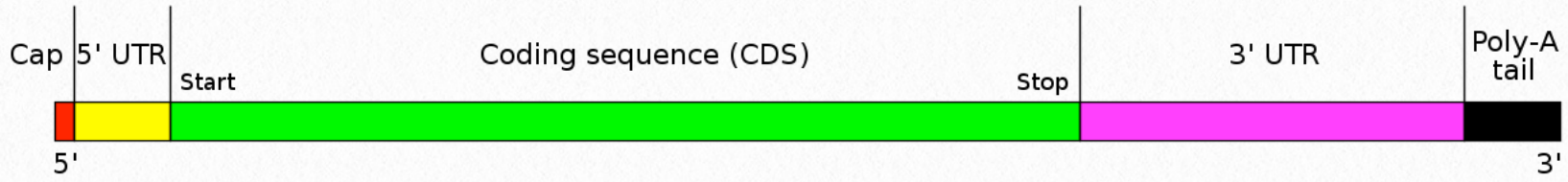

Figure 7.74 - Structure of a mature eukaryotic mRNA

RNA editing

In addition to undergoing the three processing steps outlined above, many RNAs undergo further modification called RNA editing. Editing has been observed in not only mRNAs but also in transfer RNAs and ribosomal RNAs. As the name suggests, RNA editing is a process during which the sequence of the transcript is altered post-transcriptionally. A well-studied example of RNA editing is the alteration of the sequence of the mRNA for apolipoprotein B (see also HERE). The editing results in the deamination of a cytosine in the transcript to form a uracil, at a specific location in the mRNA. This change converts the codon at this position, CAA, which encodes a glutamine, into UAA, a stop codon. The consequence of this is that a shorter version of the protein is made, when the edited transcript is translated. It is interesting that the editing of this transcript occurs

Interactive Learning Module HERE

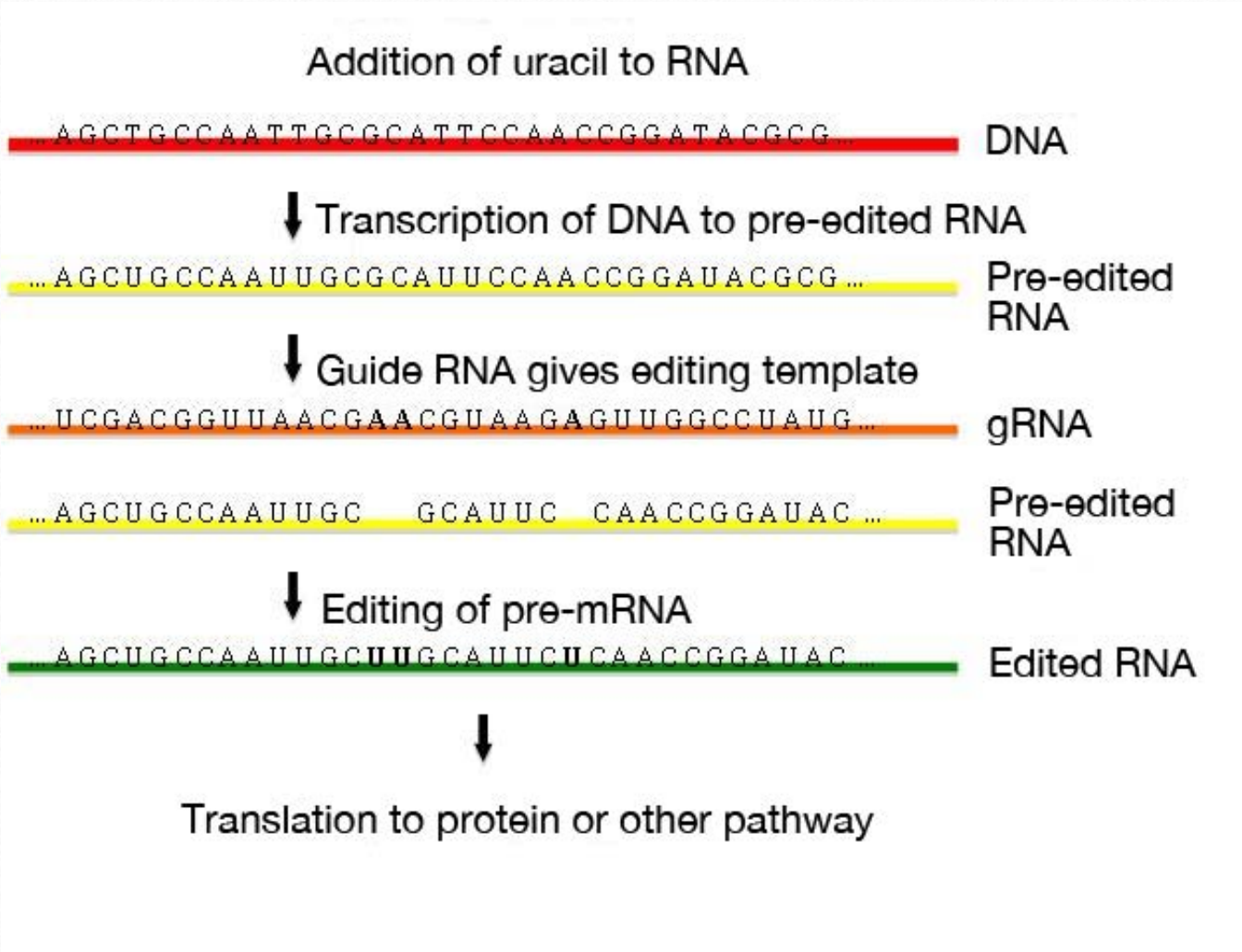

Figure 7.75 - Template guided - one mechanism of RNA editing

in intestinal cells but not in liver cells. Thus, the protein product of the apolipoprotein B gene is longer in the liver than it is in the intestine.

Insertion/deletion

Another kind of RNA editing involves the insertion or deletion of one or more nucleotides. One example of this sort of editing is seen in the mitochondrial RNAs of trypanosomes. Small guide RNAs indicate the sites at which nucleotides are inserted or deleted to produce the mRNA that is eventually translated (Figure 7.75).

The effect of either of these kinds of editing on the mRNA is that the encoded protein product is different, providing another point at which the product of expression of a gene can be controlled.

tRNA synthesis & processing

tRNAs are synthesized by RNA polymerase III, which makes precursor molecules called pre-tRNA that then undergo proc-

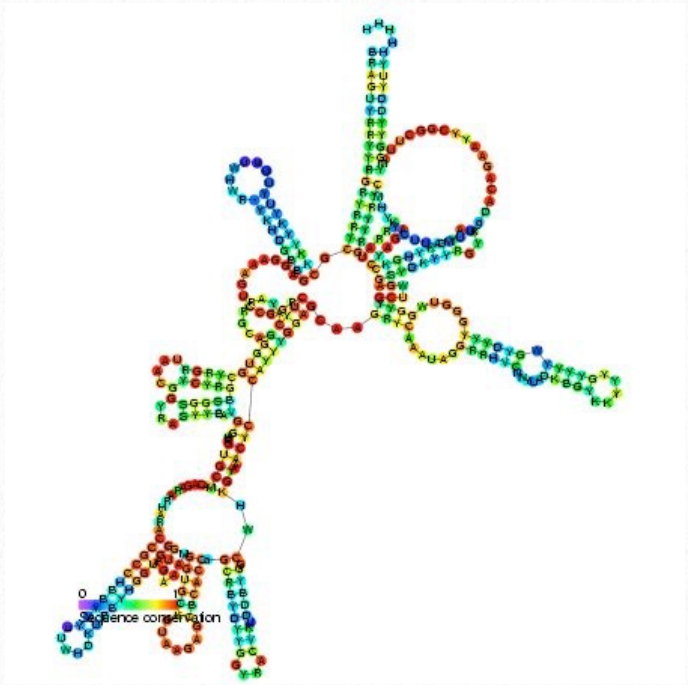

Figure 7.76 - Structure of the RNA component of ribonuclease P

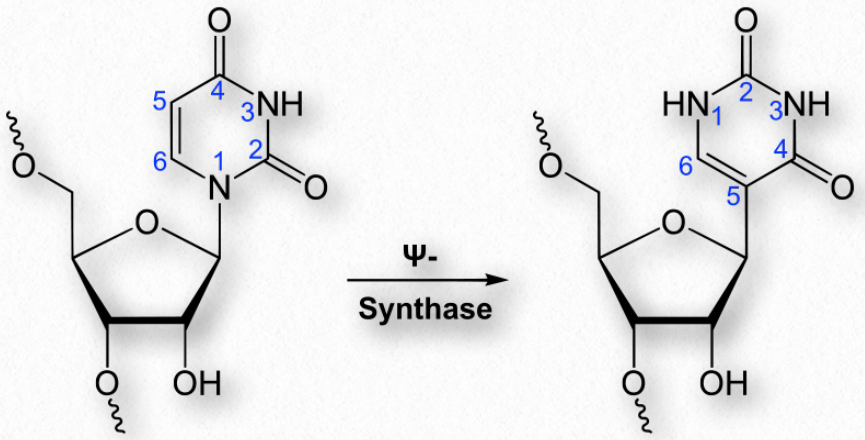

Figure 7.77 - Synthesis of pseudouridine from uridine

Wikipedia

essing to generate mature tRNAs. The initial transcripts contain additional RNA sequences at both the 5' and 3' ends. Some pre-tRNAs also contain introns. These additional sequences are removed from the transcript during processing.

The 5' leader sequence of the pre-tRNA (the additional nucleotides at the 5'-end) is removed by an unusual endonuclease called ribonuclease P (RNase P - Figure 7.76). RNase is a ribonucleoprotein complex composed of a catalytic RNA and numerous proteins. The 3' trailer sequence (extra nucleotides at the 3' end of the pre-tRNA) is later removed by different nucleases. All tRNAs must have a 3' CCA sequence that is necessary for the charging of the tRNAs with amino acids. In bacteria, this CCA sequence is encoded in the tRNA gene, but in eukaryotes, the CCA sequence is added post-transcriptionally by an enzyme called tRNA nucleotidyl transferase (tRNT).

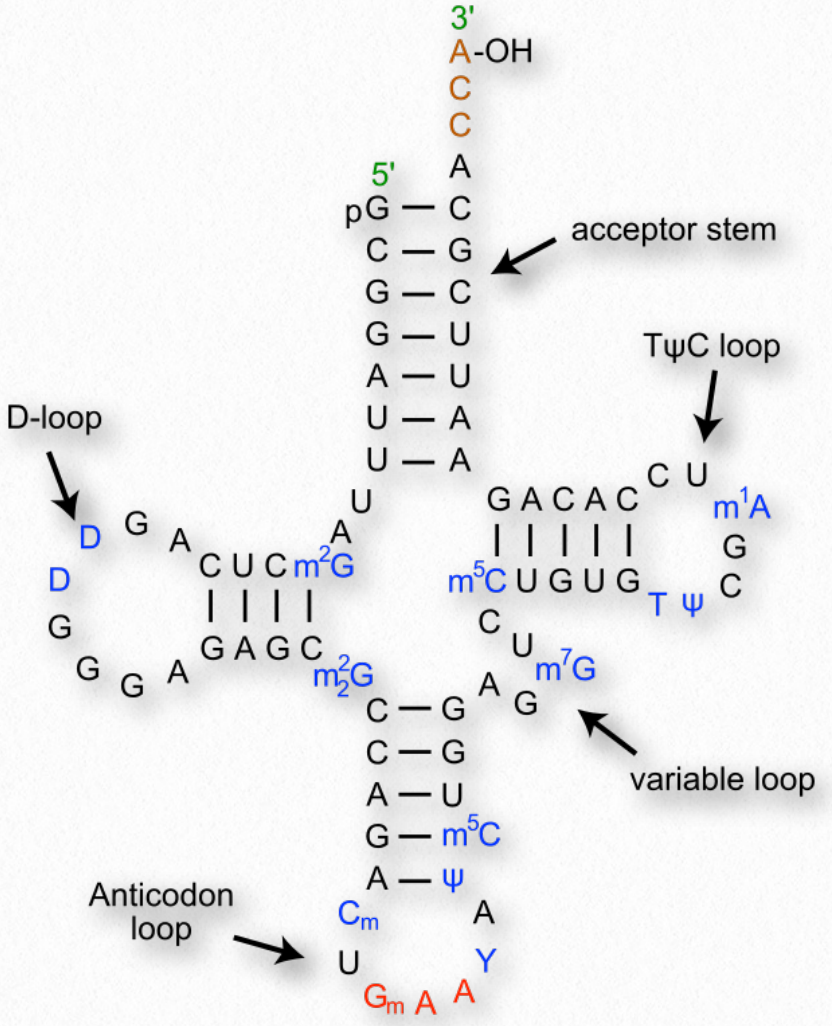

Figure 7.78 - Sequence of a mature tRNA

Wikipedia

Introns

As mentioned earlier, some tRNA precursors contain an intron located in the anticodon arm. In eukaryotes, this intron is typically found immediately 3' to the anticodon. The introns is spliced out with the help of a tRNA splicing endonuclease and a ligase.

Base modifications

Mature tRNAs contain a high proportion of bases other than the usual adenine (A), guanine (G), cytidine (C) and uracil (U). These unusual bases are produced by modifying the bases in the tRNA to form variants,

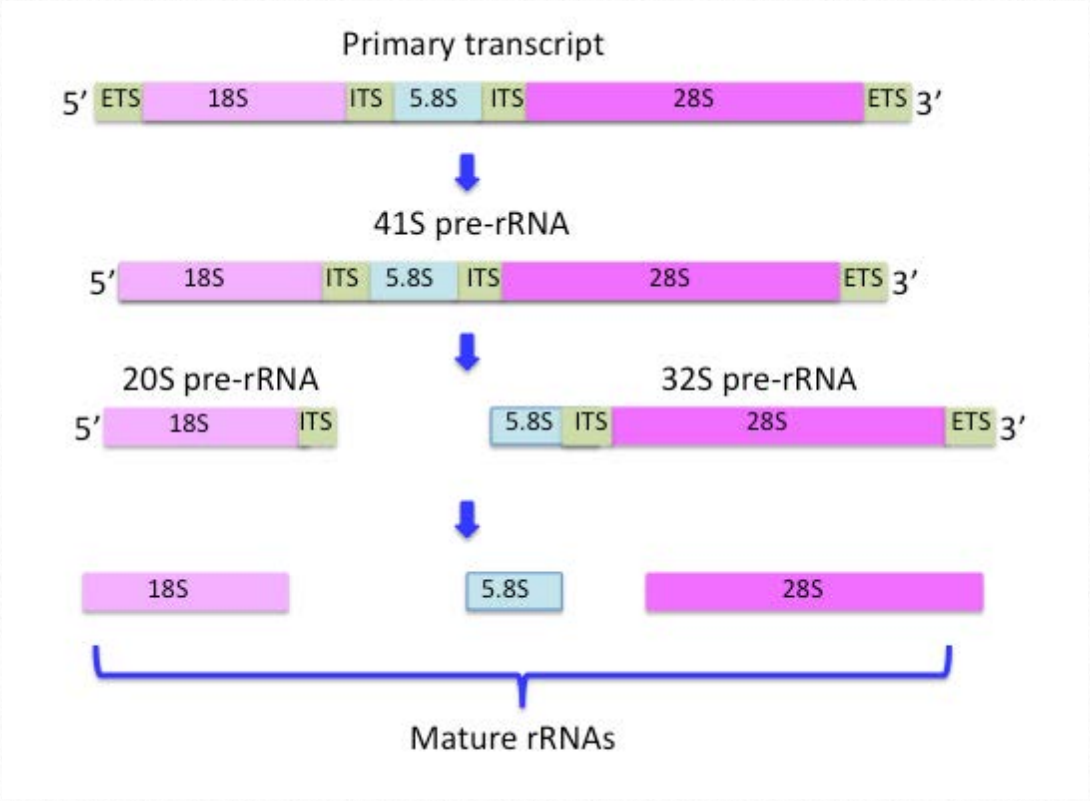

Figure 7.79 - Processing of ribosomal RNA

such as pseudouridine (Figure 7.77) or dihyrouridine. Modifications to the bases are introduced into the tRNA at the final processing step by a variety of specialized enzymes. Different tRNAs have different subsets of modifications at specific locations, often the first base of the anti-codon (the wobble position).

YouTube Lectures by Kevin HERE & HERE

rRNA synthesis and processing

Cells contain many copies of rRNA genes (between 100 and 2000 copies are seen in mammalian cells). These genes are organized in transcription units separated by non-transcribed spacers. Each transcription unit contains sequences coding for 18S, 5.8S and 28S rRNA, and is transcribed by RNA polymerase I into a single long transcript (47S). The 5S rRNA is separately transcribed.

The sizes of ribosomal RNAs are, by convention, indicated by their sedimentation coefficients, which is a measure of their rate of sedimentation during centrifugation. Sedimentation is expressed in Svedberg units (hence the S at the end of the number) with larger numbers indicating greater mass.

The initial transcript contains 5' and 3' external transcribed spacers (ETS) as well as internal transcribed sequences (ITS). The primary transcript is first trimmed at both ends by nucleases to give a 45S pre-rRNA. Further processing of the pre-rRNA through cleavages guided by RNA-protein complexes containing snoRNAs (small nucleolar RNAs), gives rise to the mature 18S, 5.8S and 28S rRNAs (Figure 7.79). Ribosomal RNAs are also modified both on the ribose sugars and on the bases. Interestingly, methylation of ribose sugars is the major modification in rRNA. The modified base pseudouridine is also common in rRNA. Other modifications include base methylation, and acetylation. These modifications are thought to be important in modulating ribosome function.

Graphic images in this book were products of the work of several talented students. Links to their Web pages are below

Click HERE for Aleia Kim's Web Page

Click HERE for Pehr Jacobson's Web Page

Click HERE for Penelope Irving's Web Page

Click HERE for Martha Baker's Web Page

Problem set related to this section **HERE**

Point by Point summary of this section **HERE**

To get a certificate for mastering this section of the book, click **HERE**

Kevin Ahern's free iTunes U Courses - Basic / Med School / Advanced

Biochemistry Free & Easy (our other book) HERE / Facebook Page

Kevin and Indira's Guide to Getting into Medical School - iTunes U Course / Book

To see Kevin Ahern's OSU ecampus courses - BB 350 / BB 450 / BB 451

To register for Kevin Ahern's OSU ecampus courses - BB 350 / BB 450 / BB 451

Biochemistry Free For All Facebook Page (please like us)

Kevin Ahern's Web Page / Facebook Page / Taralyn Tan's Web Page

Kevin Ahern's free downloads HERE

OSU's Biochemistry/Biophysics program HERE

OSU's College of Science HERE

Oregon State University HERE

Email **Kevin Ahern** / **Indira Rajagopal** / **Taralyn Tan**

The Codon Song

To the tune of "*When I'm Sixty Four*"

Metabolic Melodies Website HERE

Building of proteins, you oughta know
Needs amino A's
Peptide bond catalysis in ribosomes
Triplet bases, three letter codes

Mixing and matching nucleotides
Who is keeping score?
Here is the low down
If you count codons
You'll get sixty four

Got - to - line - up - right
16-S R-N-A and
Shine Dalgarno site

You can make peptides, every size
With the proper code
Start codons positioned
In the P site place
Initiator t-RNAs

UGA stops and AUGs go
Who could ask for more?
You know the low down
Count up the codons
There are sixty four

Recording by Tim Karplus
Lyrics by Kevin Ahern