

Handout #4 pages 51-62

(6 pages Text + Figures 24-29)

For handouts 1-4 you can use Lodish 4th edition chapter 11 as a reference. Access Lodish Textbook via <http://www.ncbi.nlm.nih.gov/books> (search for "Lodish"). You can then search for "RNA processing" which is in chapter 11.

Part #7: Editing of pre-mRNA

The discovery of RNA editing in the mid-1980's was another addition in the expanding list of amazing things that cells do to their pre-mRNA to process it to mature mRNA. RNA editing refers to the observation that the primary sequence of certain pre-mRNAs is changed by the deletion and/or insertion of Uridine nucleotides. Thus RNA editing is an exception to the rule that the sequence of the pre-mRNA and the mature mRNA corresponds to the sequence found in the gene. Figure 24 shows an example of RNA editing found in mammalian cells. The mechanism of RNA editing and the ribonucleoprotein editing complex are currently under investigation and we will not discuss it in this class. To date it appears likely that RNA editing is extremely rare in mammals with only a few examples: one is the serum protein apolipoprotein B (Apo-B) shown in Figure 24 and another example is one of the subunits of the glutamate receptor involved in neurotransmission. You should remember the following features of RNA editing.

1. RNA editing requires small RNAs called guide RNAs which basepair to the pre-mRNA thereby directing the Editosome to add or delete Uridine nucleotides.
2. RNA editing is widespread in the mitochondria and chloroplasts of plants and in protozoan mitochondria. It is so prevalent that some mRNAs have more **than half** of their primary sequence changed to a new sequence. This makes it difficult to map the edited mature mRNA to the gene from which it was transcribed.
3. In contrast to point #2, mammals have very few examples of RNA editing and both involve only single nucleotide changes, which however are critical in the function of the protein product.

Part #8: Export of mRNA: the Rev protein of HIV.

Pre-mRNA processing occurs in the nucleus, whereas mRNAs get translated in the cytoplasm. This simple fact indicates that the mRNAs must be exported through nuclear pores prior to becoming fully functional. Until the late 1980's it was thought that mRNA export was a default, automatic step in the maturation of functional mRNA. The only exception to this idea was in the case of incompletely processed pre-mRNAs, ie. pre-mRNAs where some of the introns had not been removed or a polyA tail had not been added. These type of pre-mRNAs were observed to accumulate in the nucleus and were not found in the cytoplasm, thus, it was concluded that they were defective in nuclear export. The biological rationale for this defect was based on the idea that incompletely processed pre-mRNAs potentially could encode toxic proteins if translated, thus it made sense for the cell to degrade them rather than to export them to the cytoplasm.

In the late 1980's investigations into the functioning of the HIV-1 virus (HIV for short) that causes AIDS revealed that the HIV-encoded Rev protein is critical in expression of the HIV genes by facilitating the export of unspliced HIV pre-mRNA. This regulation has been extensively studied and is shown in Figure 25. Upon infection, HIV integrates a DNA copy of its RNA genome into the host-cell DNA. The HIV genome is transcribed as one long transcription unit of 9000 bases which then undergoes alternative splicing or no splicing to yield 3 types of mRNAs: the 9000 nt unspliced RNA, a 4000 nt incompletely spliced mRNA, and a 2000nt completely spliced mRNA (which codes for Rev protein). As the Rev protein levels build up, it enters the nucleus of the infected cell and facilitates the export of the 9000nt unspliced HIV pre-mRNA that is needed in the cytoplasm for producing new viral particles that need the full-length 90000nt viral DNA. Nuclear Rev also facilitates the nuclear export of the 4000nt partially spliced HIV pre-mRNA (necessary for expression of other essential HIV proteins). In cells lacking the HIV Rev protein, the incompletely spliced HIV pre-mRNAs remain in the nucleus. This Rev-dependent export of the unspliced HIV mRNA requires an RNA sequence in the HIV pre-mRNA called the Rev Response Element or RRE (see Figure 25). The RRE-Rev complex is sufficient for export as it can be attached to unrelated, heterologous pre-mRNAs and facilitate their export from the nucleus to the cytoplasm.

This is the end of the pre-mRNA processing part of the course.

Now we move to rRNA & tRNA processing, catalytic RNA (Ribozymes) & the RNA world Hypothesis

Part #9: rRNA and tRNA processing and genes in introns.

In Parts 1-8 of these lecture handouts, we were concerned with the processing of pre-mRNAs since they encode for all the myriad types of proteins that exist in the cell. However, 95% of the total RNA in a mammalian cell is comprised of ribosomal RNA (rRNA) and transfer RNA (tRNA). These RNAs also undergo processing steps as part of their maturation. Note that my last lecture #6 will cover the functions of rRNA and tRNA in the ribosome and in translation.

9A. Ribosomal RNA Processing, snoRNAs.

Eukaryotes contain four rRNAs which are highly conserved in sequence and structure. 3 of the 4 rRNAs are synthesized as a single pre-rRNA transcript which, as shown in Figure 26A and B, then undergoes an orderly series of endonucleolytic cleavage steps and exonuclease steps resulting in the three types of rRNA found in eukaryotic ribosomes, namely the 5.8S, 18S and 28S rRNAs. The fourth rRNA is produced in a separate transcription-processing pathway. These processing steps occur in the nucleolus, a densely packed structure found in the nucleus that is the ribosome assembly center. Each type of rRNA is also extensively modified by methylation of the 2'-OH group, the same group that distinguishes RNA from DNA, however the functional significance of these modifications remains to be determined. rRNAs also contains other types of modified nucleotides such as pseudouridine.

rRNA and snoRNA processing in eukaryotes: additional things to remember.

1. rRNA processing requires a number of distinct snoRNP complexes (small nucleolar ribonucleoproteins) which are comprised of snoRNAs (small nucleolar RNAs) that stably bind snoRNA proteins.
2. rRNA processing occurs in the nucleolus.
3. As the long pre-rRNA transcript is made it is immediately bound by protein-RNA complexes (snoRNPs) that exonucleolytically and endonucleolytically cleave the transcript to yield 3 of the 4 mature rRNAs. The fourth rRNA is produced from a separate pre-rRNA transcript.
4. All 4 rRNAs contain many modified nucleotide bases.
5. Most snoRNAs are produced from the introns of functional mRNAs (Figure 26C). This means the "junk" sequences of introns are not always to be thrown away. There are also examples of mRNAs which do not code for a protein and undergo splicing only to produce snoRNAs from their introns. This latter example completely reverses our thinking about exons and introns because the intron codes for a functional snoRNA gene whereas the exon is discarded as nonfunctional. Later I will also mention that certain miRNAs are embedded in introns again underscoring that introns can produce important products.

9B. Transfer RNA Processing.

tRNAs are highly conserved, short RNA molecules (~80nts) that function in decoding the genetic information into protein sequence during the process of translation. tRNAs are produced from longer transcripts and some tRNAs contain introns which must be removed by splicing. As shown in Figure 27, all pre-tRNAs (bacterial, archaea and eukaryotes) contain a 5' end which must be removed by a ribonucleoprotein endonuclease called RNase P which contains a single protein and a single RNA called M1RNA. M1 RNA was one of the earliest catalytic RNAs to be discovered because it could bind and cleave pre-tRNAs in vitro in the absence of proteins. tRNAs undergo up to 4 types of processing modifications which are summarized in Figure 27. These modifications include:

1. all tRNAs in all 3 kingdoms have their 5' end removed by Rnase P (which contains a protein and an RNA. In bacteria the RNA is the catalytic molecule, in eukaryotes it is the protein.
2. all tRNAs in all 3 kingdoms have their 3' end nucleotides replaced by a CCA nucleotide triplet sequence by the CCA-adding enzyme. There is now emerging evidence that the 3' terminal "A" in the CCA is involved in peptide bond formation and so is central to the catalytic mechanism of the ribosome.
3. all tRNAs in all 3 kingdoms have modified bases.
4. some tRNAs require that an intron be spliced out. tRNA splicing uses a special set of endonucleases and ligases that are unrelated to the Spliceosome.

Part #10: Catalytic RNAs and Ribozymes (Figure 28).

The first RNA shown to have catalytic activity was found in the rRNA of a protozoa, *Tetrahymena thermophilus*. This Tetrahymena pre-rRNA contains an intron which is spliced out when this pre-rRNA is incubated with itself in the absence of protein. Subsequently, many self-splicing introns were discovered in rRNAs and tRNAs from all 3 kingdoms of life and are now designated group 1 introns to differentiate them from the splicing of pre-mRNA which is designated group 2 intron splicing. A second class of group 2 introns that are self-splicing, ie do not require proteins for splicing has also been discovered. The features that are similar and different between group 1 and group 2 introns are summarized in Figure 28 and are listed below.

1. **Energy requirement.** Group 1 introns use guanosine (no phosphates) as a cofactor so they have no ATP-based energy requirement. Although group 2 introns do not require guanosine you should remember that spliceosomal group 2 introns require ATP for protein and RNA rearrangements (but no ATP is required for the catalysis).
2. **Introns that are excised.** Group 1 splicing results in the excision of a linear intron that also has an extra guanosine nucleotide at the 5' end. Group 2 splicing results in the removal of a lariat RNA and no extra nucleotides are added.
3. **Polypeptide requirement.** Group 1 introns and the Group 2 self-splicing introns do not require protein to carry out their catalytic activity. Group 2 spliceosome introns do require proteins to carry out the reaction.

Similarities between group 1 and 2 intron splicing include:

1. two step reaction mechanism
2. both reaction steps are transesterifications.
- 2) no energy is required for the catalytic reactions (this is like peptide bond formation).

Part #11: RNA World Hypothesis

The "RNA World" Hypothesis has spawned a literature of its own that is known even outside of the field of molecular biology. This hypothesis is based on a fundamental dilemma inherent to the Central Dogma of Molecular Biology (DNA begets RNA begets protein) namely how does one produce proteins without the genetic code of DNA and how does one produce DNA without the catalytic properties of proteins. In its most-simplified form this dilemma is analogous to the age-old question of "which came first, the chicken or the egg?"

The field of Molecular Evolution has demonstrated that simple compounds, CO₂, O₂, ammonia, nitrate, H₂, H₂O, formaldehyde, etc... are able to combine with each other to produce the known building blocks of life, amino acids, nucleotides, lipids etc... Although the presence of these components in a prebiotic soup appears more likely than ever, there are still several major hurdles in the pathway of molecular evolution that would lead to short RNA polymers. However, even assuming the "molecular biologist's dream" that the prebiotic earth contained short RNAs, DNAs and polypeptides it was clear that self-replicating molecules were needed to get processes akin to natural selection in progress. Thus, a breakthrough molecule was needed which had to be self-replicating, meaning it had to have both template-directed information capabilities as well as enzymatic capabilities. Modern life forms have divided these two properties into two molecules, DNA and polypeptides. Yet as with the chicken and the egg, we need both DNA and protein to get molecular

life started. The properties of RNA have brought it to the forefront as the most likely candidate for being the breakthrough molecule. The idea that RNA was the "first" molecule of life was proposed in the 1960's however, little came of this proposal as RNA appeared to be the dullest member of the 3 macromolecules of the cell. Not only were RNAs referred to as messengers, temporary deliverers of the genetic information stored in DNA, but they were also viewed as having a passive structural role in ribonucleoprotein machines including the ribosome, the signal recognition particle and the spliceosome. RNA was viewed as a structural skeleton to which was appended the polypeptides that catalyzed critical cellular reactions. All that came undone with the discovery of catalytic RNAs and the structural diversity of modern RNAs. This coupled with breakthroughs in in vitro selection systems (see SELEX described in Part #12) has spawned a small industry of researchers who are collecting evidence that RNA preceded both DNA and proteins as being the first molecule of life.

The "RNA World" hypothesis means different things to different authors, but all RNA world hypotheses include 3 basic assumptions: 1) at some time in the evolution of life, genetic continuity was assured by the replication of RNA, 2) basepairing was the key to replication, and 3) genetically encoded proteins were not catalytically involved in points 1 and 2. Evidence for these assumptions has been building and has led to a paradigm shift that RNA has properties ideally and uniquely suited for being the first molecule of life. Subsequent experiments have borne out these properties. RNA is now recognized to have a number of properties that are remarkably useful for being the first molecule. I caution that significant assumptions in this hypothesis lack an experimental basis and so it is at best a hypothesis with significant "holes".

1. Catalytic. Catalytic RNA can do ligation reactions and short-template driven polymerase reactions that show it can be self-replicating. RNA has now been shown to catalytically drive a subset of the reactions involved in the Krebs' glycolysis cycle. Evidence is now very strong that the catalytic heart of the ribosome, ie, peptide bond formation, is RNA catalyzed and may be even 100% catalyzed by RNA. Catalytic RNA can also perform homologous recombination, a key step in producing novel structural and enzymatic motifs by bringing together separate RNA domains.

2. Structural diversity. In proportion to work on proteins and DNA, structural work (X-ray crystallography and NMR) on RNA has only taken off since the late 1990's (with the exception of tRNAs and tRNA synthetases). This short period of time has already revealed that RNA adopts secondary and tertiary shapes that make it indistinguishable from proteins. In contrast DNA is overwhelmingly in B-duplex conformation and the existing deviations are fairly minor when compared to what RNA can do.

3. Short RNAs have activity. Although it is indisputable that RNA has limited catalytic and structural properties as compared with polypeptides, it is notable and surprising that short RNAs (<20 nts) are better than short peptides (of a similar number of repeat units) at folding and at performing catalytic functions. Shortness would have been an important attribute in the prebiotic earth because longer macromolecules are more susceptible to damage caused by UV and gamma ray bombardment, drastic changes in temperature, etc., conditions associated with a prebiotic earth. Unlike polypeptides of a similar length, short RNAs have a higher propensity to fold into coherent defined structures that can contribute to catalyzing reactions and nucleating structures.

Part #12: SELEX and RNA evolution in vitro (Figure 29).

SELEX is an acronym for **S**ystematic **E**volution of **L**igands by **E**xponential Enrichment. In its essence SELEX is a method for directed evolution of novel RNA structures and enzymatic activities in an in vitro system. It combines two techniques: 1) the ability to mutagenize DNA with 2) the ability to select or enrich for particular RNA activities. Thus, it is analogous to the Theory of Evolution by Natural Selection which also combined pre-existing variants (ie. mutants) in a population with selection of the most fit individuals being able to pass hereditary information into the next generation. Figure 29 describes the basic SELEX technique. Essentially, a pool of single stranded DNA oligos is produced by chemical synthesis in which flanking bases are a single unique sequence (they are held constant) while a middle region consists of random sequences. The "complexity" of this pool of DNA is determined by the length of the randomized region which in this case is 30 nts long. Therefore this DNA pool contains $4^{30} = 10^{18}$ unique DNA sequences and is also referred to as a "sequence space". Note that this randomized region can be any length we choose, the number 30 here is just an example. Using PCR and in vitro transcription, an identical RNA copy is made from this DNA pool. The RNA pool then undergoes a selection process in which those RNA sequences able to perform a given function, ie the "Winners" are pulled out of this mixture. The remaining RNA sequences are discarded, ie. the "Losers". The RNA Winners are then reverse transcribed back into a DNA sequence pool, which is now enriched for RNAs with the desired properties. Multiple rounds of this selection process results in only a few Winner sequences which then can be analyzed individually. **A mutagenic variation step** can be introduced by using nucleotide analogues during steps 5 and/or 6 or a variant Taq polymerase enzyme (or variant reverse transcriptase) that has a high error rate.

SELEX has been successfully used to produce:

- 1) RNA aptamers that act like antibodies, they can bind any molecule you imagine, from single amino acids to long polypeptides.
- 2) RNAs with novel functions: example an RNA molecule that enters into a lipid bilayer and then behaves as a nuclear pore.
- 3) RNA with novel enzymatic properties. This supports the idea of a prebiotic RNA World. An example is the enzyme glucose 6 dehydrogenase can now be replaced by an RNA molecule that was identified by SELEX. RNA can also do ligation reactions and polymerization reactions, meaning it can replicate itself.

Practice SELEX problem.

You have isolated and cloned a new protein called X. Based on sequence homology with other known proteins you predict that X will bind to RNA having a structure comprised of a 6bp stem followed by an 8nt long, single stranded loop (ie. this is a stemloop RNA). The problem is you do not know the sequence of the loop. Your goal is to use SELEX to identify the sequence that is "best" loop that will bind to X with high affinity and specificity. Describe all the SELEX steps that will help you to accomplish your goal. In your answer include the enzymes that you will need and whether the DNA or RNA is single stranded (ss) or double stranded (ds).