RNA and Protein Synthesis

Considerable evidence suggests that RNA molecules evolved prior to DNA molecules and proteins, and that many processes now involving DNA and proteins were previously accomplished by RNA alone. Many RNA molecules have enzymatic activity, and like DNA contain genetic information stored as nucleotide sequences. In living cells encountered today, RNA molecules are formed from DNA templates through transcription, but this was not always so. It is no wonder then that RNA molecules play such important roles within living cells.

Types of RNA molecules:

Although cellular RNA molecules are typically made via transcription using DNA as a template, they are coded for by different genes and have a variety of different functions. RNA molecules typically found in both prokaryotic and eukaryotic cells include:

1) **Messenger-RNA** (**m-RNA**) – Messenger-RNA molecules are coded for by regions of DNA called **structural genes**, and carry the nucleotide sequences determining what types of proteins the cell will make. They carry this "message" from the nucleus or nucleoid to the ribosomes where proteins are made.

2) **Ribosomal-RNA** (**r-RNA**) – Ribosomal-RNA molecules bind with proteins to form **ribosomes**, the sites of protein synthesis. In prokaryotic cells r-RNA molecules occur as 23S, 16S and 5S segments. The 23S and 5S segments occur within the 50S ribosomal subunits, while the 16S segments occur within the 30S subunits. Ribosomal-RNA is largely responsible for ribosome function (except for the formation of peptide bonds) and coordinates the attachment of t-RNA **anti-codons** with m-RNA **codons** during translation. Most of the RNA within a prokaryotic cell is ribosomal.

3) **Transfer-RNA** (**t-RNA**) – Transfer-RNA molecules are relatively small, being only 74 to 93 nucleotides in length. Their function is to carry individual amino acid molecules to the ribosomes for protein synthesis. Each t-RNA molecule is folded upon itself to form a shape roughly similar to a cloverleaf. The top or "stem" forms the amino acid binding site and typically has the base sequence CCA at the 3' end. The side loops are called T and D loops, and the bottom loop contains the **anti-codon region** (as set of three bases). **Aminoacyl-t-RNA-synthetase** enzymes catalyze the formation of covalent bonds binding amino acids to t-RNA molecules. There are typically many different aminoacyl-t-RNA-synthetase enzymes within each cell because each one catalyzes only the reaction binding a specific amino acid to the CCA sequence of each t-RNA molecule. The resulting complex is called **aminoacyl-t-RNA**.

Eukaryotic cells typically contain additional types of RNA not found in prokaryotic cells. Though these are not directly involved in protein synthesis, they are of considerable importance. Some additional types of RNA include:

1) **Small or short-RNA** (**s-RNA**) – Small-RNA molecules bind with proteins to form structures called **spliceosomes** and are involved in a process called **post-transcriptional modification**. In eukaryotic cells, nearly all RNA molecules undergo some modification following transcription, so these small-RNA molecules play a vital role. In the case of m-RNA molecules, post-transcriptional modification involves the removal of multiple sections called **introns** (intervening regions) from the original RNA transcript. The remaining sections, called **exons** (expressed regions) are then bond together to form a shorter m-RNA molecule (in some cases over half the original m-RNA transcript is composed of introns). Post-transcriptional modification also involves the application of a methylated-guanine "cap" (which is attached to the 5' end before transcription is completed), and a poly-adenine "tail". The splicing of exons following the removal of introns can occur in a variety of ways in some m-RNA molecules, and this increases the variety of proteins that can be generated.

2) **Micro-RNA** (**mi-RNA**) – Micro-RNA molecules are very small, usually only 21 bases in length. They regulate the expression of genes by binding with m-RNA molecules. At this time, researchers believe there are probably hundreds of genes coding for different mi-RNA molecules, and that the potential for regulating cellular processes by means of mi-RNA is enormous.

3) **Small interfering-RNA** (**si-RNA**) – Small interfering-RNA molecules (also called short interfering or silencing-RNA) are typically 20-25 nucleotides in length, and double-stranded. Like mi-RNA they play a significant role in regulating the expression of eukaryotic genes.

Translation – Protein synthesis:

The term "**translate"** means to give an equivalent in another language, therefore, when not applied to biological activities, **translation** refers to the process of reproducing a written or spoken text in a different language while retaining the original meaning. When applied to biological systems, it has essentially the same meaning.

Translation is protein synthesis. During translation, information contained in **m-RNA** molecules in the form of nucleotide sequences (transcribed from small sections of DNA; also nucleotide sequences), is **translated** into information contained in **polypeptides** in the form of amino acid sequences. The language change is from nucleotide sequence to amino acid sequence. Some polypeptides are fully functional proteins, while others undergo modification to become functional units. Recall that proteins with quaternary structure contain multiple polypeptide chains.

Some background information:

When the structure of DNA was first determined (through the efforts of Rosalind Franklin, Maurice Wilkins, James Watson and Francis Crick – see Wikipedia DNA) in the early 1950s, researchers were uncertain how this relatively simple molecule could contain the information necessary to control the physiological processes within living cells. It became obvious that protein structure was dependent on the nucleotide sequences of DNA molecules, but it was not immediately apparent how this could be. Proteins typically contain 20 different amino acids, and the **primary structure** of each polypeptide is

determined by the sequence or arrangement of these monomers (just as the sequence or arrangement of letters determines the informational content of words on a page). Each DNA molecule (and each m-RNA molecule) contains only four different nitrogenous bases, so researchers knew individual bases (translated separately) could not determine amino acid sequences. When "read" in pairs (groups of 2), the bases could be arranged to yield 16 different sets, but this was still not enough. Then researchers considered translating base sequences arranged in sets of three. In this configuration, they form 64 possible sets (3-letter "words"), which is more than enough to code for 20 different amino acids. By synthesizing nucleotide sequences containing only one type of base, i.e., polyadenine, polycytosine, polyguanine, etc. researchers were able to determine portions of the "**genetic code**", the sets of bases coding for different types of amino acid; polyadenine sequences yielded polypeptides containing only lysine, polycytosine sequences yielded polypeptides containing only proline, polyguanine sequences yielded polypeptides containing only glycine, etc. Eventually the entire genetic code was revealed.

The Genetic Code:

The genetic code typically presented in biology textbooks (and in the lecture syllabus) is made up of nucleotides (bases) arranged in groups of three. These groups are called **codons**, and occur within messenger-RNA molecules and DNA molecules. Since nucleotides arranged in sets of three form 64 different combinations (64, 3-letter "words"), and because proteins typically contain only 20 different amino acids, there is considerable overlap. Most of the amino acids incorporated into proteins are "coded for" by multiple different codons, so the genetic code contains **redundancy**. Some references refer to the code as being degenerate, but the overlap has benefit because it reduces the potentially lethal effects of certain types of mutation as will be explained later.

In addition to the codons translated into amino acids, the genetic code contains four codons with alternate functions. The codon "**AUG**" codes for the amino acid **methionine**, but also serves as a start or **initiator codon**. The codons "**UAA**", "**UAG**", and "**UGA**" do not code for any amino acids, and were initially referred to as non-sense codons; these code for "stop" or termination of the growing polypeptide chain. Named "Ocher", "Amber" and "Umber" (color names), these three **terminator codons** play an essential role in protein synthesis. Although similar in all living organisms, the genetic code as presented in most texts is **not** universal. Variations are common within mitochondria. For example, mitochondria found within animals and eukaryotic microbes use "UGA" as the code for tryptophan instead of stop, most animal mitochondria use "AUA" as the code for methionine rather than isoleucine, all mitochondria associated with vertebrate animals use "AGA" and "AGG" as chain terminators, and yeast mitochondria use "CUA, CUG, CUC and CUU" to code for threonine instead of leucine. Only plant mitochondria use the universal code. Beyond that, researchers have now developed strains of *E. coli* and *Saccharomyces cerevisiae* that use specific redundant codons as the codes for amino acids not normally incorporated into proteins. The code apparently has considerable versatility.

Translation:

The translation process occurs in association with **ribosomes** and involves the three types of RNA common to all types of cells, i.e., m-RNA, r-RNA and t-RNA. Since m-RNA molecules carry the genetic message, they must interact directly with the ribosomes.

Translation can be broken into a series of stages as follows:

1) **Initiation** - Initially, the **small ribosomal subunit** (30S in prokaryotes, 40S in eukaryotes) binds to m-RNA at a site "upstream" of the initiator codon. Some bacterial m-RNA molecules carry the nucleotide sequence "5'-AGGAGG" (the Shine-Dalgarno sequence) which binds to the complimentary sequence "3'-UCCUCCA" (the anti-Shine-Dalgarno sequence) on the 16S ribosomal-RNA portion of the ribosome. In bacteria lacking a Shine-Dalgarno sequence, the binding of m-RNA to the small ribosomal subunit involves interaction between ribosomal proteins and purine rich regions of RNA upstream of the initiator. In eukaryotes a sequence beginning with "A" or "G" followed by "CCACC" (the Kozak concensus sequence) serves as the ribosome binding site.

   Once the binding of m-RNA to the small ribosomal subunit has occurred, a number of **translation initiation** factors (IF) bind to the ribosome along with a t-RNA carrying the amino acid methionine. In prokaryotes, this is **N-formylmethionine** (fmet) rather than the methionine commonly incorporated into proteins. Following the binding of this aminoacyl-t-RNA molecule, the second ribosomal subunit (50S in prokaryotes or 60S in eukaryotes) binds to the m-RNA.

2) **Elongation**
   The large ribosomal subunit has three t-RNA binding sites designated as the A-site, the P-site and the E-site. The A-site is the **amino acid binding site** and binds with incoming t-RNA molecules (**aminoacyl-t-RNA**), i.e., those carrying individual amino acids to the ribosome. The P-site is the **polypeptide binding site**, and once translation has begun, will bind with the **peptidyl-t-RNA** molecule, i.e., the one holding the growing polypeptide chain. The E-site binds with the **exiting-t-RNA**, i.e., the free t-RNA molecule remaining after the amino acid it was carrying has been added to the growing peptide chain.

   The binding of **aminoacyl-t-RNA molecules** involves interaction between **t-RNA anti-codons** and **m-RNA codons**. Hydrogen bonds forming between the complimentary bases insure that each amino acid will be incorporated in the correct location. A ribosomal enzyme called **peptidyl transferase** catalyzes chemical reactions resulting in the formation of **peptide bonds** between adjacent amino acids. Without this enzyme activity, proteins cannot be made. **Peptidyl transferase** activity apparently involves RNA molecules, not ribosomal proteins.

   Translation requires that the ribosomes travel along m-RNA molecules from the 5' ends toward the 3'ends. As they do, the codons associated with the A-site change and new aminoacyl-t-RNA molecules can bind. As peptide bonds are formed between adjacent amino acids, the growing peptide chain is transferred from the t-RNA in the P-site to the one occupying the A-site (explaining the name peptidyl transferase). The t-RNA released by the transfer of the peptide chain, occupies the E-site until the ribosome moves, and is then released. These "exiting" t-RNA molecules can again interact with **aminoacyl-t-RNA synthase** enzymes, pick up new amino acids, and return to the ribosome, i.e., they can be used over and over again.

3) **Termination**
   Translation is terminated when the ribosome reaches a stop or terminator codon. In both prokaryotic and eukaryotic cells, the process involves the interaction of two or more proteins called **chain release factors**. When a stop codon is recognized, the finished polypeptide chain is released from the last t-RNA, i.e., the covalent bond attaching the last amino acid added to its corresponding t-RNA molecule is hydrolyzed. Following this reaction, the t-RNA is released from the ribosome, and the two ribosomal subunits separate, releasing the m-RNA.

Since messenger-RNA molecules have a relatively short life expectancy (about 2 minutes in some eukaryotic cells), bacteria often increase translation efficiency by binding multiple ribosomes to the same m-RNA during translation. A group of ribosomes bound to and translating the same m-RNA is called a **polyribosome** or **polysome**. 0

In eukaryotic cells, m-RNA molecules typically carry the information copied from individual genes (transcription is **monocistronic**), but in prokaryotic cells, m-RNA molecules often carry the information copied from multiple genes (transcription is **polycistronic**). **Polycistronic m-RNA molecules** carry information that is translated into multiple polypeptides. In *E.coli* the m-RNA copied from the tryptophan biosynthesis operon contains about 7000bp and is translated into five different enzymes. Each protein is coded for by a section of m-RNA having its own start and stop signals, i.e., initiation and termination sequences.

For a fun and simple review of the translation process take a look at this site.
http://www.phschool.com/science/biology_place/biocoach/translation/term.html