

Protein Synthesis¹

I. Introduction: In the last section we learned the basics of DNA and RNA synthesis. Today will be devoted to the synthesis of proteins and in the next class we will consider how this synthesis is controlled. It sounds like we will once again be talking about mechanisms and in fact we will spend much time on the "nuts and bolts" of the process. But beyond the process, what we will be looking at are the mechanisms that make one cell different from another and the means whereby sequences of nucleotides interact with the environment to produce the structures and processes of life.

II. About Genes: You certainly know that genes are units of heredity and that they are composed of DNA found in a cell's chromosomes. However, for the purposes of today's discussion we will need to know considerably more about the action and structure of genes. So, some definitions:

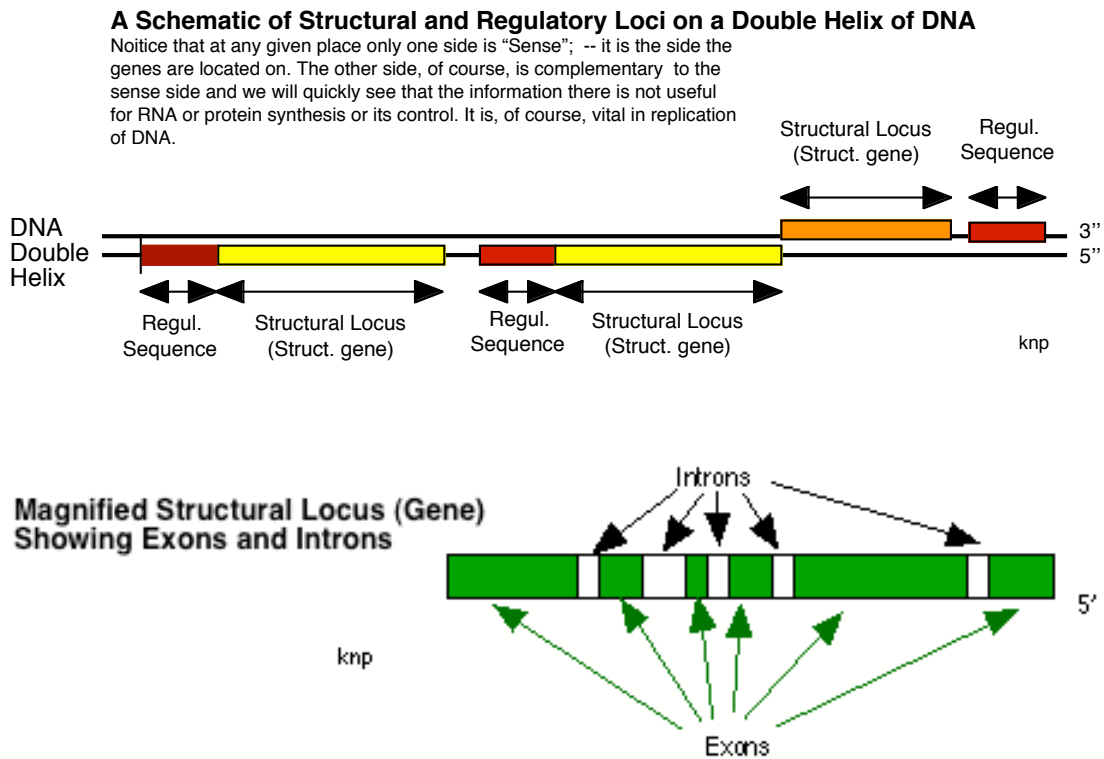
- **Gene:** this term can be used broadly or very specifically. As originally (broadly) construed, it means some inherited difference between individuals that increases the likelihood of showing a certain trait. As generally construed by molecular biologists, a gene is a specific section of DNA that contains information that is used somehow to construct the organism. Using this type of definition, there are two general categories of genes.
- **Structural gene:** a gene that contains information that ultimately specifies the primary structure of a protein (in which case it also specifies the structure of a molecule of messenger RNA, see below) or it contains the information needed to produce a particular type of transfer or ribosomal RNA (see below) or ribozyme. Structural genes in eukaryotic organisms can be subdivided into two types of region:
 - **Exon:** the sequence of nucleotides within the structural gene that carries information eventually used to synthesize a protein
 - **Intron:** "intervening sequences" -- sequences within a structural gene, often quite long, that do not contain information used to specify the primary structure of a protein (or non-mRNA structure). No thoroughly convincing hypothesis has yet been put forth for their role; we'll see more about them later
- **(b) Regulatory genes (regulatory sequences):** sections of DNA, usually near to a particular structural gene locus. Their purpose is to help to control the access to information in structural genes. Regulatory sequences are far more common than structural genes although they tend to be relatively short in length when compared to structural genes. There are typically many regulatory sequences associated with each structural gene.
- **Locus:** a synonym for the term gene but it implies a particular location or place where certain instructions are located. The term traces from the fact

¹ © 2006 by K.N. Prestwich, Dept. Biology, College of the Holy Cross, Worcester, MA 01610

genes are in fact located in predictable places (that is, on certain chromosomes, next to certain other genes, and generally in a certain exact physical location). This said, a locus might be an exact position or a general location on a chromosome that contains a number of genes (including the one of interest). We will use the term in its more exact sense.

- **Alleles:** slightly different nucleotide sequences for a given locus (gene); they may be found in different individuals or on similar chromosomes in the same individual. Don't fret too much about this one right now, as it will not be a major part of our discussions. It will come into more use later when we do a unit on transmission (Mendelian) genetics.

So, here are a couple of pictures of what we need to know about the organization of genes (for the moment).



III. Types of RNA: RNA is all the same in that it is always:

- Composed of ribonucleotides,
- It contains uracil instead of thymine,
- Consists of a single strand (there are a few viral exceptions to this),
- it is synthesized using a structural gene (DNA) as a template (once again, there are a few viral exceptions to this)

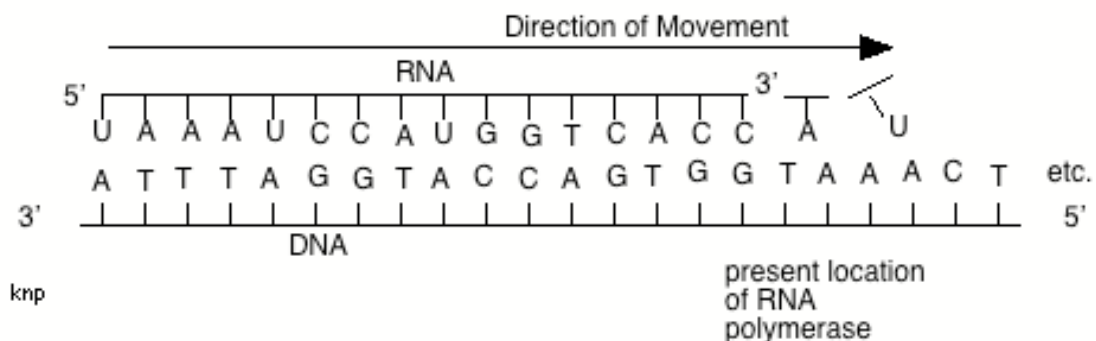
However, there are functionally three general categories of RNA. As with our usual rules of form and function (see the second notes for the course), they

have slight differences in structure that go along with their functions. Note also that within each of these categories, there are many versions, each with its own unique structure and properties.

Messenger RNA (mRNA): mRNA has a central role in the production of proteins. mRNA carries the information found in a structural gene out of the nucleus (in the case of eukaryotic organisms) into the cytoplasm and to the ribosomes where other types of RNA (tRNA), the ribosomes themselves, and various enzymes cause appropriate amino acids to be lined up and polymerized into a protein. The mRNA molecule contains information that actually specifies the order of these amino acids. Since most proteins are large molecules consisting of several hundred or more amino acids, mRNAs also tend to be large molecules. More about this below.

Since the mRNA contains a copy of the information found in the DNA, we often term it a "**transcript**" and the process of synthesizing mRNA as **transcription**. Transcription is no different than the RNA synthesis process we learned in the last set of notes -- it is exactly the same process used in the synthesis of all functional types of RNA. **The only thing that makes one molecule mRNA and another tRNA is the function of the molecule** (which will be reflected in its size, the nucleotide sequence, and the shape it assumes when it is completed – form and function again). In any case, recall what we learned in the last set of notes:

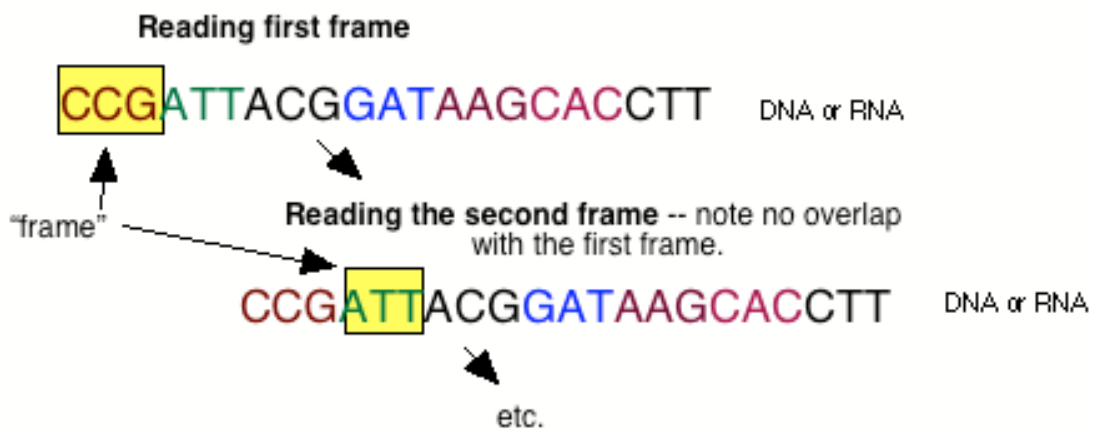
- The RNA polymerase reads a particular gene in the 3 to 5' direction
- The strand of RNA it synthesizes is made as the 5' side of the incoming NTP adds to the 3' end of the growing nucleic acid ("5' to 3' addition).
- Thus, the resulting RNA molecules differ from each other according to differences in the DNA template used to direct their synthesis. In the drawing below, the gene template is shown at the bottom; the antiparallel strand is not shown:



As you know, mRNA contains information about the order of amino acids for a particular type of protein. That is, the gene and then the mRNA molecule both contain a map of the primary structure of a protein. Recall that proteins are all made of amino acids of which there are 20 types. In the 1950s one of the most important pursuits in molecular biology was to figure out how amino acids were specified – that is to figure out the basic syntax of the nucleotide language.

It seemed to most of those who studied the problem that the most likely rules for the "genetic code" would be somewhat akin to aspects of written language. That is:

- A certain number of shapes (like letters) would be joined together to stand for a certain "concept" – an amino acid. These sequences could be thought of as genetic words.
- They knew the structure of DNA and realized that, unlike human languages, different words were not written with physical spaces between them. The reason for this expectation was that it was known that in both DNA and RNA, sequential nitrogenous bases are the same distance apart. Yet it still made sense that the words would most likely **not overlap** each other anymore than they do in our written languages.
- If that was the case and if the information did not overlap, then exact rules would be needed to read the information. The most logical conclusion was that a certain number of letters would be read as word and then whatever it was that was doing the reading would shift ahead an entire frame to the next word. Let's say the words were 3 letters long (which they turned out to be). Furthermore, let's assume that the language consists of exactly 4 letters: A, C, G, and T (or U). The hypothesis as to how the information would be displayed and read is:



Now, why should the frames be three long? The answer is that if they are any shorter it is not possible to specify all the information that we know genetic systems specify. And, if the words are any longer, they will waste space, material, and energy. Here's the logic. Recall that since we are writing instructions to build proteins that we need 20 words -- one for each amino acid. Now, each nucleotide can be either A, C, G or T (U). How many unique words (shapes) can be produced using this 4 letter language if the words are always 3 letters long?

- If the words were only 1 nucleotide long, there would be four possible words (A, C, G, or U).

- If they were two nucleotides long, there are a possible of 16 unique words (e.g., AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, and UU).

Notice that the total number was equal to $4^{\text{length of the word}}$ which is $4^2 = 16$ in this case.

- If the "words" are actually 3 places long, then there are a total of $4^3 = 64$ unique combinations. If you are feeling ambitious, write them all out.

Notice that now we have 64 "words" that we can use to say 20 things.

It ends up that the language of heredity, like most languages has more than one word for the same thing. Many amino acids are specified by two or three different "words". **Three letter sequences of nucleotides** on an mRNA molecule are termed **codons**. There is no commonly used name for the complement of these three letter sequences that was found in the DNA used to direct the synthesis of RNA. Just think of them as DNA codons. Of the 64 codons,

- 61 combinations specify amino acids (many amino acids have more than one codon)
- 3 codons do not correspond to an amino acid and instead act as "stop" signals.

I hate this term but we say that the **triplet code** is "**degenerate**" because it has more than one codon for many amino acids. A better term is to call it **redundant**.

Other points: I can't emphasize too much the fact that there are no "gaps" in RNA molecules between codons; regardless of how they are often illustrated in books. You should be clear on this and clear on the reason why.

Suppose that a form of life from another planet was found to use a hereditary molecule and mRNA like molecule that contained 6 nucleotides (three groups of complementary types). Suppose also that it used 33 different kinds of amino acids. What would be the minimum length of a codon?

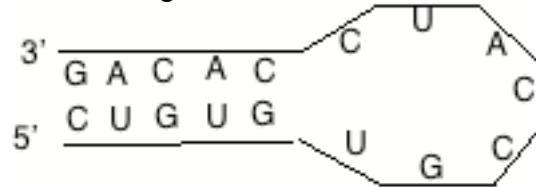
Back to earth. If an mRNA molecule is 300 nucleotides long, what is the maximum number of codons it contains?

One last observation. Recall from earlier in these notes that structural genes in eukaryotes have regions called "exons" and "introns". Both of these types of areas end up in the initial or "**primary RNA transcript**". We'll see a bit later in these notes that the introns must all be excised before the primary transcript becomes a functional mRNA molecule. However, for the moment, let's take a side trip and consider the production of other types of RNA.

Transfer RNA (tRNA): These are very short molecules that consist of only about 75 nucleotides. They are not straight, string-like molecules. As soon as they are synthesized, they tend to fold on themselves. This folding is caused by the fact that certain sequences that are some distance apart from each other are complementary. For example, the sequence:

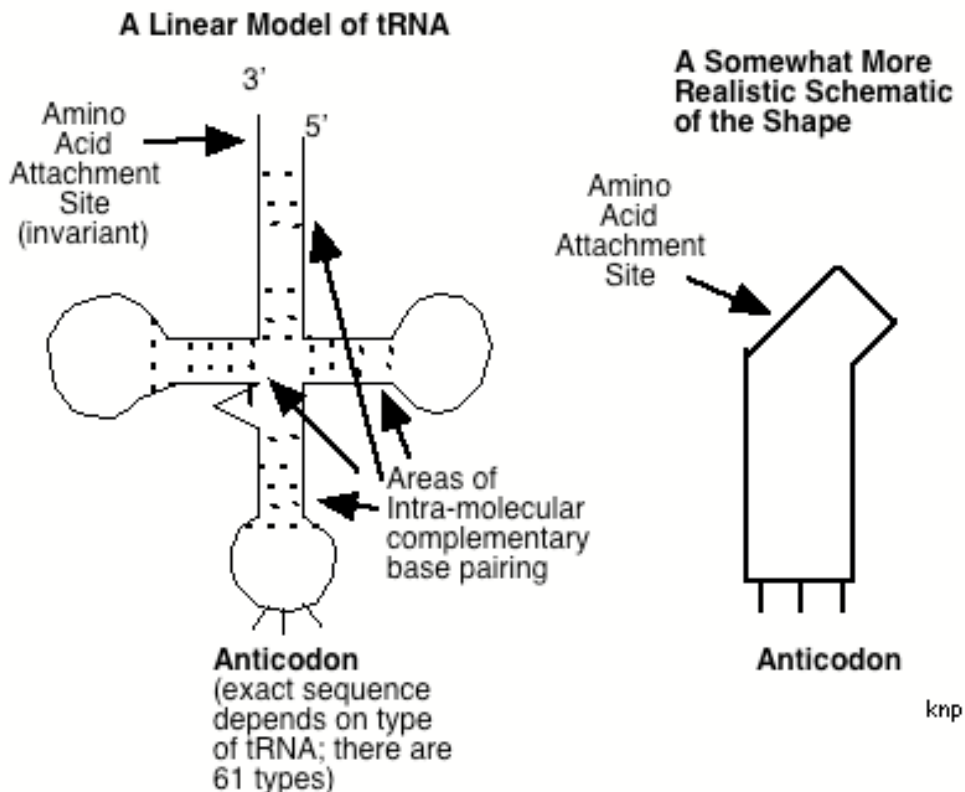
3' -GACACCUACCGUCUGUC -5'

can fold on itself to give the following structure:



This sort of thing is very common in RNA, especially in tRNA and ribosomal RNA.

In transfer RNA the result of this self-pairing is a molecule that has 4-paired areas and three loops. This general configuration, which is shared by all types of RNA molecules, is shown below. Don't get the idea however that the resulting molecules are all flat structures with three loops. In fact they fold in such a way that the result is a curved structure (see next page) that is sometimes (misleadingly) termed the "**L**" structure. Nevertheless, we will use the idea of an L structure because it greatly simplifies what we need to know about tRNA.



About this "L" shape. Notice from the models above that the "top" is where the 3' and 5' ends are near together. This is **where an amino acid may be attached -- at the 3' end**. At the bottom of the tRNA molecule is a loop with something called an "**Anticodon**". This is a place where three nitrogenous bases stick more or less outwards from the molecule. This region is the part of the tRNA molecule that will interact with mRNA during protein synthesis. Obviously, **a particular anticodon will only interact with its specific complementary mRNA codon (see above)**.

Moreover, the **anticodon is the central place for identifying the type of tRNA molecule**; in the "L" model, it is the bottom of the "L". So, the anticodon is used to identify:

- A particular type of tRNA
- and therefore the type of amino acid that it carries.

Ribosomal RNA (rRNA): Last but not least, these are very large RNA molecules with numerous folded and paired areas. They assume very complex shapes reminiscent to those found in proteins. And as with proteins, their shape determines their function. Most of the genes for the many types of rRNA are found near the part of the nucleus we call the **nucleolus**. These structures (there can be many in one nucleus) are most visible in cells that are actively synthesizing proteins and therefore that need large number of ribosomes.

This leads us to ribosomes. **Ribosomes** are, by cellular standards, small structures composed of a number of different types of structural proteins, enzymes and rRNA. Most people consider them organelles even though they do not have membranes; we will also call them organelles. Functionally, they are the sites of protein synthesis --period. They are either found floating around more or less free in the cytoplasm or they are bound to the endoplasmic reticulum ("rough endoplasmic reticulum or rough ER). It is also worth noting once again that mitochondria produce their own ribosomes that although generally similar to those in the rest of the cell, resemble those of bacteria more. This is considered to be further evidence that mitochondria originated from bacteria that joined with early eukaryotic cells. The mitochondrial ribosomes are used to synthesize proteins that are coded for by the mitochondrial DNA mentioned previously.

Finally, a word about ribosomal structure. All ribosomes are composed of two subunits called the **large** and **small** subunits. Guess what -- one is bigger than the other.

- The small subunit has a binding site for mRNA and
- The large subunit has three regions called A, P, and E sites that are used to organize tRNA and the growing peptide chain.
- The large and small subunits normally only associate with each other during protein synthesis -- when it is complete, they dissociate.

Recycling: The "players" of protein synthesis (nucleotides, nucleic acids, proteins and amino acids) will have many lives. Keep in mind that not only are there many examples of each specific type of these chemicals but also that any individual molecule or structure will be used over and over again. Eventually, for some reason, each complex molecule (RNA or protein) will be broken down to its constituent parts. The most common fate of these parts (nucleotides or amino acids) will be that they will be recycled; it is far less likely that they will be metabolized. This is exactly like the situation we saw earlier with ATP, coenzymes, and enzymes.

An example. A molecule of a particular type of tRNA will be joined with the particular type of amino acid that it carries, and participate in protein synthesis where it loses its amino acid. It can then pick up another amino acid of the same type and once again participate in protein synthesis. This cycle will continue many times before this particular tRNA is broken down into NMP and these are then converted back to NTPs and used to synthesize new molecules of RNA.

Let's Set the Stage: In any cell that is capable of synthesizing proteins, there will already be plenty of tRNA of each type present, along with ribosome parts (see below), all 20 types of amino acids, and plenty of NTPs for RNA synthesis. Moreover, at any time most of the tRNA molecules will already be carrying the type of amino acid for which they are specific. Let's look at how amino acids come to be associated with a given type of tRNA:

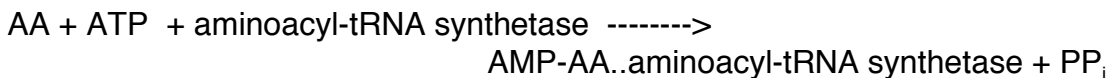
Production of Aminoacyl tRNAs: When a given tRNA has the appropriate amino acid attached to it, it is called an **aminoacyl tRNA** or simply a "charged tRNA".

Here we go with the term "charged" again. In this case, it simply means that the tRNA is carrying an amino acid. It has nothing to do with issues of electrical polarity. Kind of like the fact that phosphagen buffers have nothing to do with pH buffers; in biology, the same word frequently has many meanings!

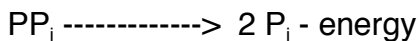
This charging process is very important. For protein synthesis to work correctly, it is absolutely vital that the correct amino acid be attached to the correct tRNA -- in other words, the correct amino acid with the correct anticodon. This vital task is accomplished by a class of more than 20 related enzymes called **aminoacyl-tRNA synthetases**. Each species of aminoacyl-tRNA synthetase is specific for a particular type of amino acid and tRNA. Thus, the wrong amino acid cannot be put on a given type of tRNA.

Like any synthesis, the production of amino-acyl-tRNA requires an energy source. As usual, the energy source is ATP. Just as with nucleic acid synthesis, we will see that 2 ~P are used to attach each amino acid. The result is that the process is very difficult to reverse and it requires lots of energy.

Here is an overview of the process of attaching one amino acid to its tRNA. Several steps are involved. The first involves "activating the amino acid (AA):

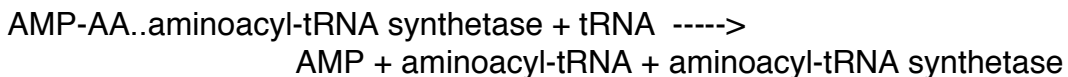


the pyrophosphate immediately is hydrolyzed (in the cytosol):



which of course makes the previous step essentially irreversible.

When the stages above are complete, the appropriate tRNA attaches to the enzyme's active site next to the AMP-AA (already present from the previous step). Again, recall that the shape of the active site prevents anything but the correct amino acid and tRNA from coming together:

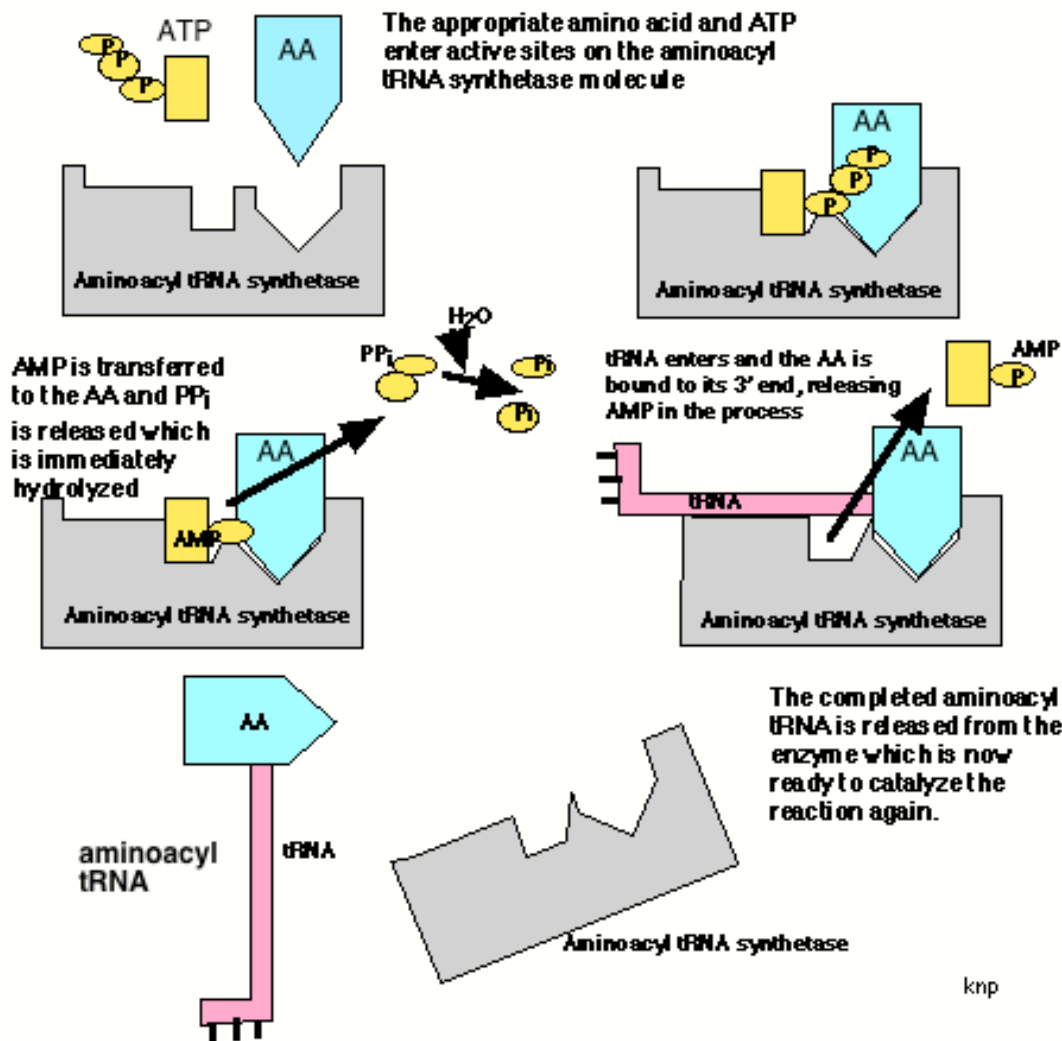


In case you are interested, the amino acid attaches to the 3' end of the tRNA at its carboxylic acid end. Thus, the amine end is left free.

There are a couple of additional things to realize about this reaction:

1. Notice how directly the enzyme (aminoacyl-tRNA synthetase) participates – in this case it covalently binds to the AMP-aminoacyl. Actually this sort of combination is common mechanism in catalyzed reactions. Here the attachment is covalent and in other cases it is looser.
2. Notice that as usual the enzyme is not lost in the process -- by the end of the process, the enzyme is ready to be used for another round. On the other hand, the substrates have definitely changed -- ATP was degraded to AMP and 2 Pi and a tRNA and amino acid have been joined.

A summary diagram is shown on the top of the next page:



The Central Dogma and Protein Synthesis: The stage is set and players are ready. We are now ready to overview the entire process of protein synthesis. In the process, you will see how information moves from the hereditary molecule (DNA or sometimes RNA) to the cell and how it helps to determine the structure

of a protein and eventually metabolic pathways and the large-scale features of an organism.

Processes in the Nucleus – RNA synthesis and Processing

Primary Transcripts: Under appropriate conditions (that we will cover in the next set of notes that deal with genetic regulation) RNA molecules are transcribed from genes that ultimately code for proteins. In eukaryotes these **primary transcripts** contain both introns and exons -- they contain a transcript of the entire gene. Introns correspond to nucleotide sequences that do not code for anything. Introns need to be removed before protein synthesis. So, the next step:

mRNA Processing: Before the primary transcript mRNA leaves the nucleus it is intercepted by a series of enzymes and RNA molecules. There are three general processes that occur:

- **Capping**
- Addition of a **poly-A tail**
- **Intron removal**

Let's look at each of these in a bit more detail:

Capping involves attaching a **modified version of the nucleotide guanine to the 5' end of the mRNA**. This process does two things.

- Enzymes that degrade RNA have a hard time getting started on molecules with this unusual version of guanine. Thus, capped mRNAs are harder to break down.
- This adds an additional ID tag to the 5' end of the molecule. This ID is important for allowing the mRNA to attach correctly to the ribosome during the translation process (see below).

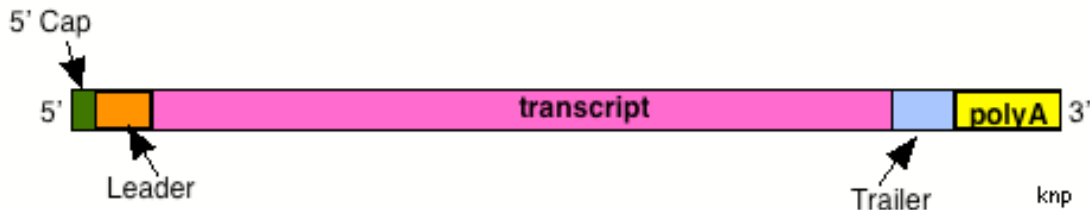
Poly A tails are just what they say they are -- long runs of the nucleotide adenine (30 to 200).

- This tail also makes it harder for **RNAases** to act on the mRNA.
- In addition, **it helps the mRNA to get out of the nucleus** -- apparently it serves as an ID tag to some proteins in the nuclear membrane that

Intron Removal and Splicing: Finally, there is **the problem of introns**. Recall that they do not code for any amino acids in proteins and therefore they must be removed. After capping and adding a tail, a series of proteins and ribonucleic acids assemble at various spots along the mRNA transcript -- wherever there are introns. They snip out the introns and join the remaining exons on either side of a removed intron. This is called **SPLICING** (the term comes from splicing a rope, as all of you sailors and ex girl and boy scouts know). The result is a piece of mRNA that is ready to direct protein synthesis. By the way, the excised introns

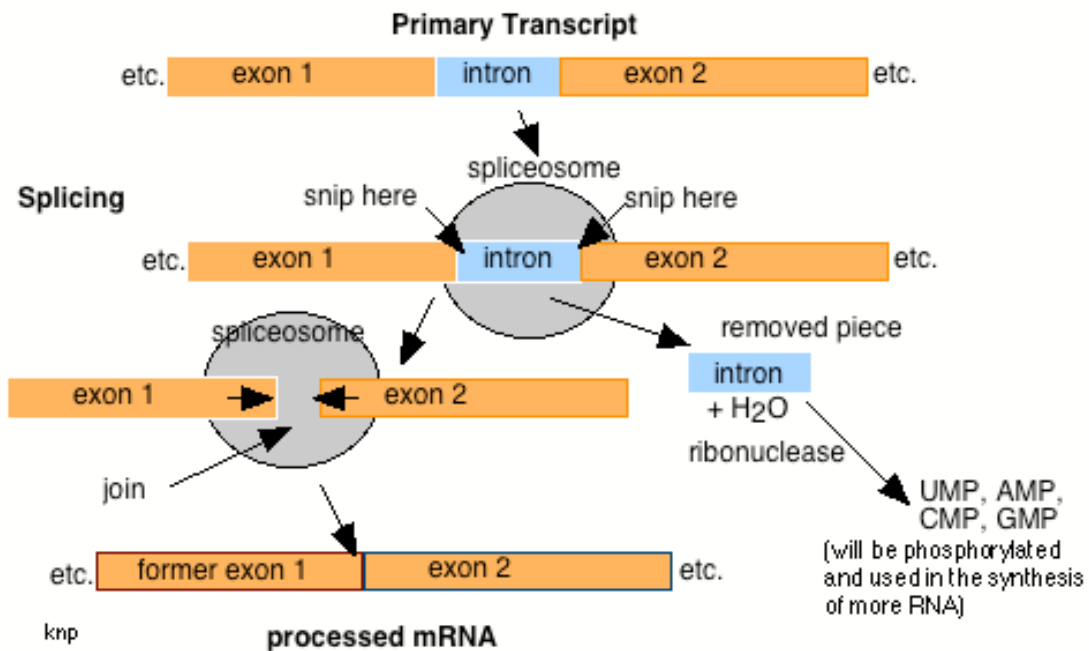
are broken down by RNAase into NMPs that are then converted to NTPs and used in the synthesis of other molecules of RNA.

By way of overview then, a fully processed mRNA molecule in a eukaryote looks like this:



A fully processed mRNA molecule. The transcript has been stripped of introns and has a "cap" of modified guanine on the 5' end and a tail of adenines on the 3' end. The "Leader" and "trailer" refer to sections of the original transcript that do not code for parts of the protein but that are important in the translation process.

and here is an over view of the splicing process:



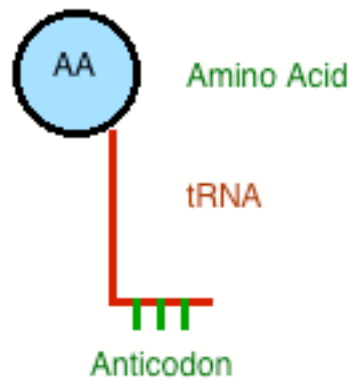
One of the most interesting features of the splicing process is that some of the catalysts used are made entirely of RNA and are generally called **ribozymes** (from ribonucleic acid "enzymes"). While ribozymes are certainly not nearly as numerous as protein catalysts, more and more are being discovered.

Translation: This is the process whereby the information contained in the mRNA transcript is used to direct the synthesis of proteins. The term "translation" refers

to the fact that instructions in the form of a linear series of codons (linear sequence of nucleotides) are used to create a specific linear sequence of amino acids. Thus, information, written in the language of DNA and RNA (a sequence of nucleotides), becomes instead represented as a series of amino acids. The entire process occurs in the cytoplasm (or the matrix of a mitochondrion). The players in the translation process include:

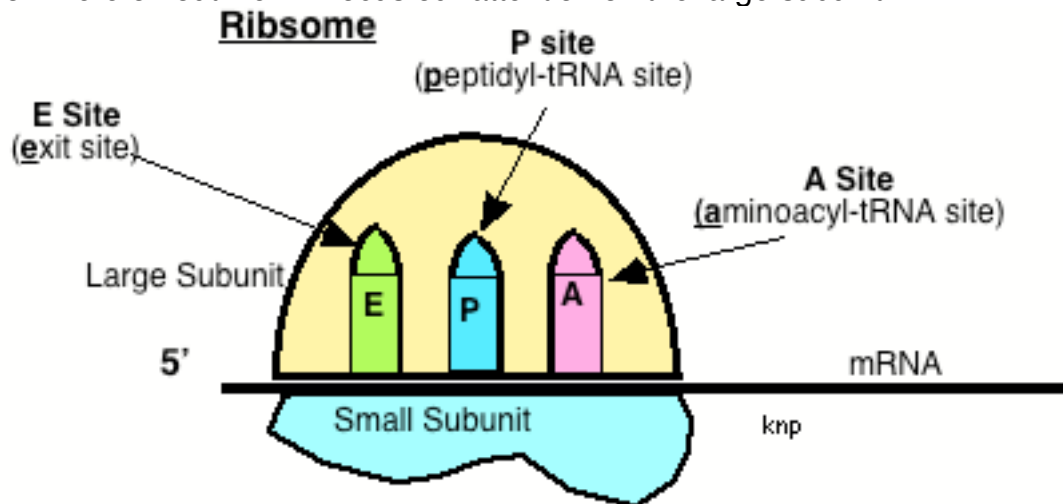
- a fully processed **mRNA** molecule (see above)
- **ribosomes**
- **protein "helpers"** -- recognition, elongation, and termination factors. All of these have to do with specific stages in the translation process.
- **aminoacyl-tRNA** (*i.e.*, "charged" tRNA molecules -- see earlier in these notes). We will illustrate these using the "L" notation discussed earlier (see top of next page).

A Schematic of an Aminoacyl tRNA



The Steps of Translation:

Recall that ribosomes consist of a small and large subunit and that these two are not normally bound together. Before they can bind together, mRNA must bind to the small subunit. Then it joins with a large subunit to form a ribosome. From here on out we will focus our attention on the large subunit.

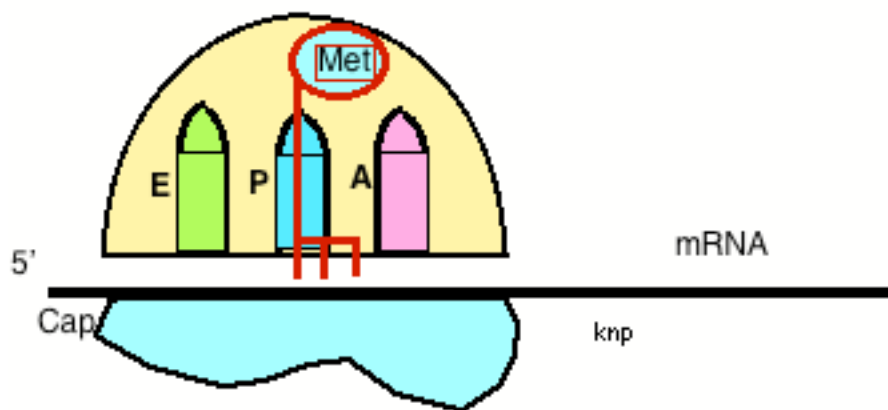


In particular, note the three tRNA binding areas on the large subunit:

- The **E** or **exit site** is the last place where a tRNA molecule is bound to the complex before it is released so that it can go and pick up another amino acid.
- The **P** or **peptidyl-tRNA** site will normally hold a tRNA that contains the entire growing peptide. We will see that the tRNA molecule in this site changes with the reading of each codon.
- The **A** or **aminoacyl-tRNA** site is where each new incoming AA-tRNA (aminoacyl tRNA) first binds.

Step 1: Initiation: The small subunit attaches to the 5' end of the mRNA (by identifying the cap). Using some energy carried in GTP and a series of proteins called **initiation factors** this complex joins with the large subunit. The mRNA lines up with the tRNA sites in the large subunit such that a sequence of nucleotides that is located a short distance downstream from the cap lines up under the P site. This particular codon (the one now under the P site) is called the **start codon** (AUG). Thus, it will pair with an aa-tRNA with the anticodon UAC; this particular tRNA **carries the amino acid methionine**:

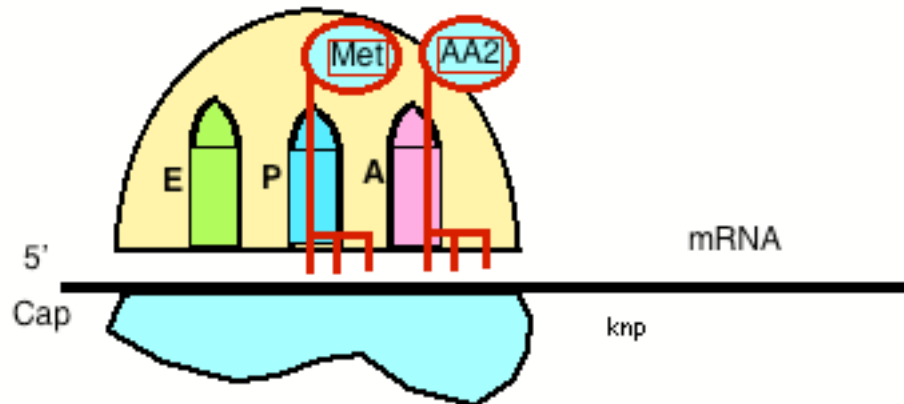
The Initiator aa-tRNA in the P-site complexed to the start codon of the mRNA



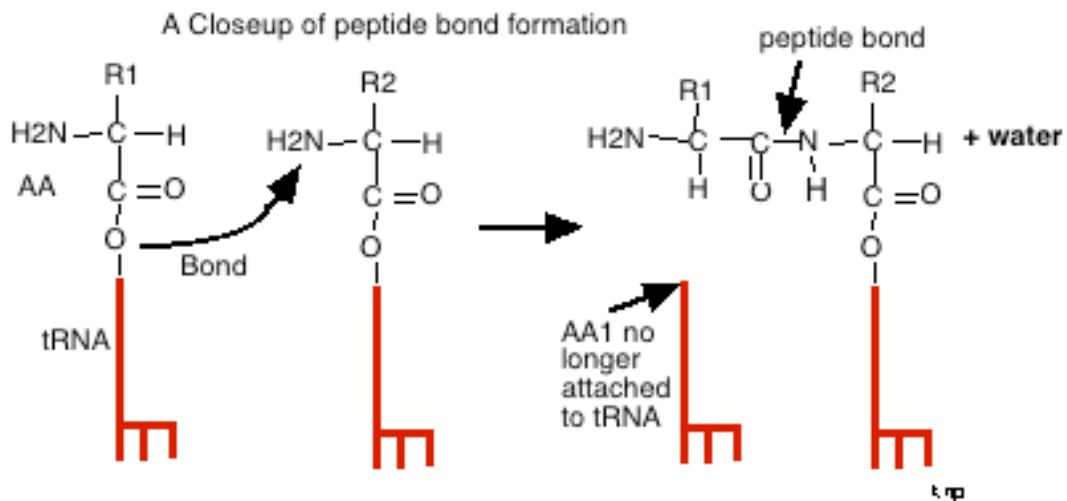
Don't learn the start codon and anticodons sequences (unless you want to)!

Step 2: Elongation: Now we will start building the polypeptide chain. Recall that aa-tRNAs are everywhere, having been synthesized in preparation for protein synthesis. At random they pop in and out of the **A site**. When one comes in that has the anticodon for the codon exposed under the A site, it will hydrogen bond. This is called **codon recognition**:

The Initiator aa-tRNA remains in the P-site while codon recognition occurs in the A site and a second amino acid is lined up along the mRNA

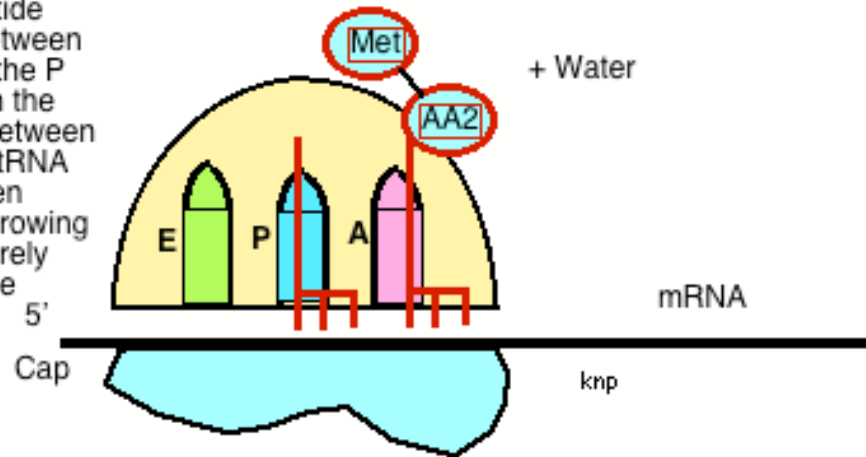


Step 3: Peptide Bond Formation: the amino acid attached to the tRNA in the P-site is bonded to the amine (free) end of the amino acid in the A site. Since amino acids are bound to the tRNA via their carboxyl end, this means that the first amino acid has now been transferred to the tRNA in the A site:



One item of interest here is that the formation of the peptide bond is catalyzed by a ribozyme found in the large subunit of the ribosome.

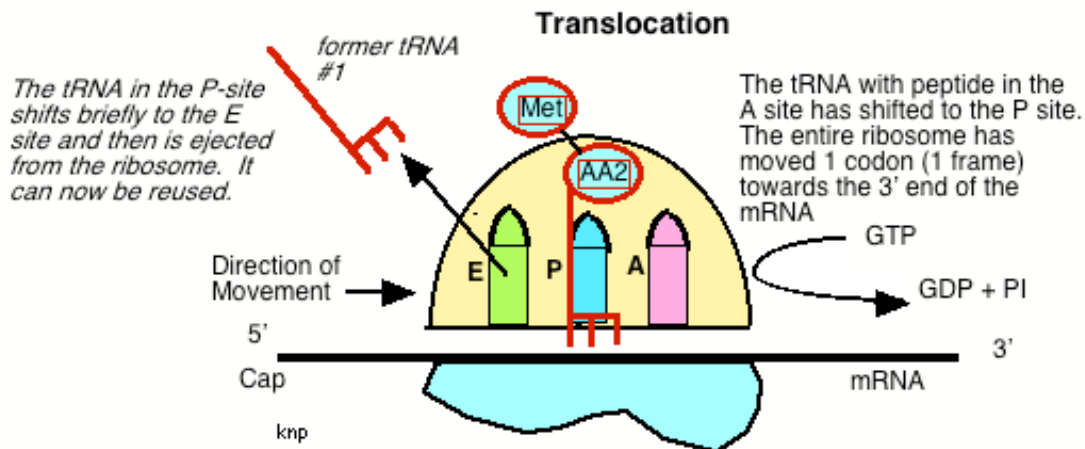
Elongation: A peptide bond has formed between the amino acids on the P and A site tRNAs. In the process, the bond between the amino acid and tRNA in the P site has been broken and so the growing peptide chain is entirely attached to the A site tRNA.



Where did the energy to drive the formation of a peptide bond (a decidedly non-spontaneous reaction) come from? (hint – look earlier in the notes).
 What general type of reaction occurred when the peptide bond was formed?

Step 4: Translocation: in this step, **the whole ribosome does the equivalent of walking one codon in the 5' to 3' direction.** The result is that:

- The "amino acid-less" tRNA in the P site ends up in the E site as a result of the entire ribosome moving one codon towards the 3' end of the mRNA. **This tRNA is ejected upon arrival in the E site;** it leaves the ribosome and is re-charged and reused.
- the **tRNA holding the growing chain ends up in the P site (from A)**, again as a result of the movement of the ribosome in the 3' direction along the mRNA. It remains bound to its codon on the mRNA
- This process requires energy in the form of GTP
- Notice again that **the translocation moves the reading exactly 1 codon or one reading frame.**



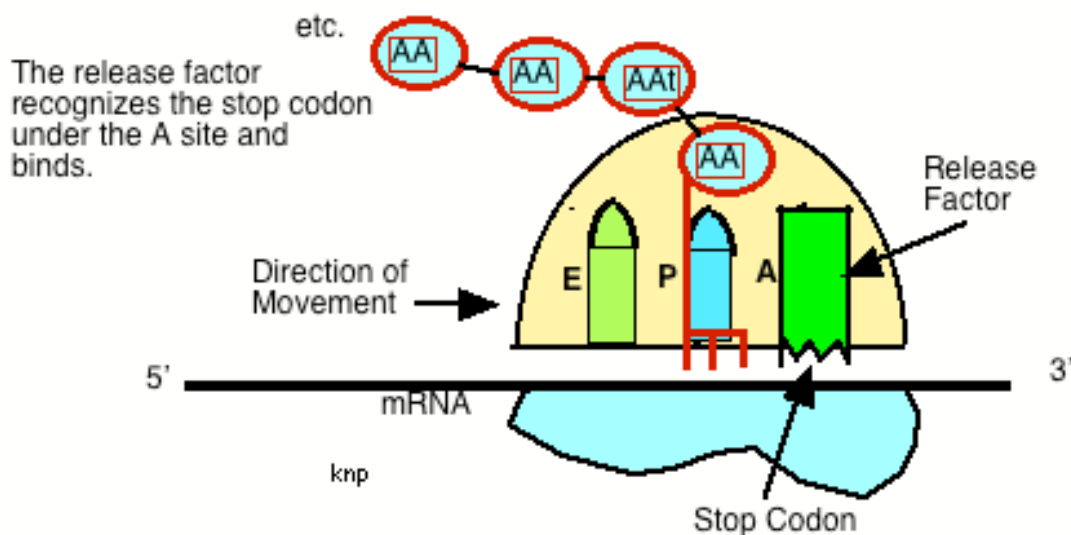
Step 5: Do Steps 1 to 4 Over and Over. The process (codon recognition, elongation, and translocation) will continue over and over as the ribosome moves

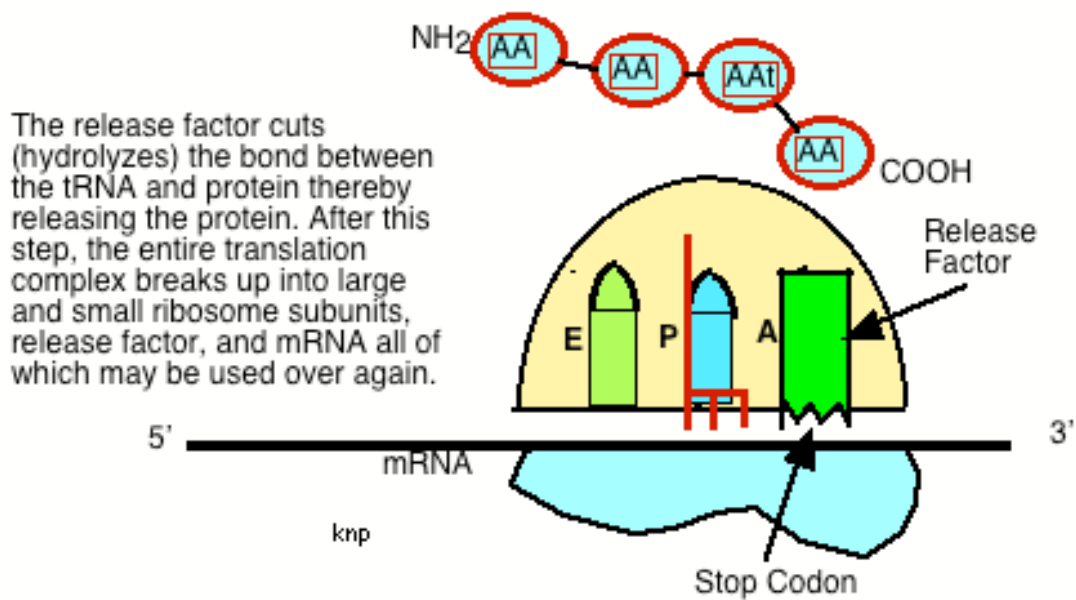
down the mRNA one frame (codon) at a time. Notice that this mechanism reads the information on the mRNA in a non-overlapping matter -- the reading frame (the exposed 3 nucleotide sequence) always moves exactly 3 nucleotides in each cycle. Thus, as we have mentioned before, codons are distinct.

Note also the use of more $\sim P$, this time for the translocation step. As usual, in animals this comes from oxidative phosphorylation.

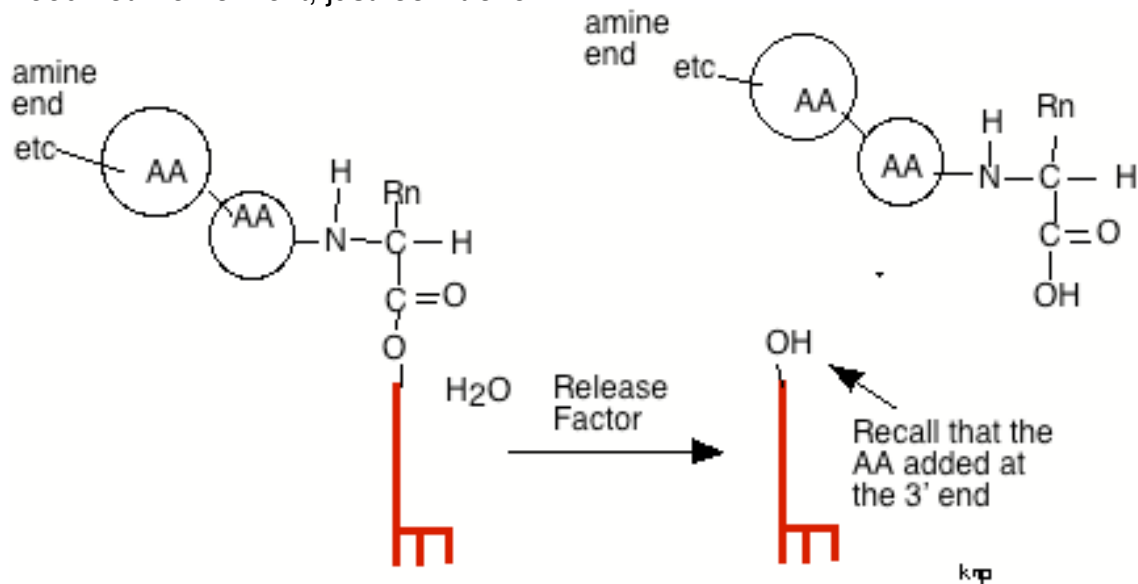
Step 6: Termination: Eventually we get to near the end of the mRNA. Just before the end of the mRNA is a section called the trailer and at its start is what is called a **stop codon**. What are stop codons? Recall that of the 64 codons, only 61 specified amino acids. Thus, three sequences (**UAA, UAG, UGA**) do not code for amino acids (NO NEED TO LEARN THESE SEQUENCES!). When they are encountered the polypeptide chain cannot elongate further. What actually happens is that:

- When a stop codon appears under the A site after the translocation step, **there is no tRNA that can attach**.
- However, a protein called a **release factor** binds to the mRNA at the stop codon.
- This factor will catalyze the hydrolysis of the bond between the polypeptide and the tRNA in the P site. As a result, the polypeptide is released.
- Once this happens, the entire ribosome breaks down into its small and large subunits, the mRNA is released so that it can potentially be translated again, and the release factor is also set free to be used by another ribosome (as usual, remember that there will be many release factor proteins milling about the cell:





What follows is a more detailed view of the hydrolysis and its consequences; you need not memorize it, just look it over:



Notice that proteins can recognize particular amino acid sequences. How is this possible? Note: we have already seen examples of proteins recognizing specific nucleotide sequences -- recall the replication origin sites on DNA. We will see more examples of protein DNA interaction shortly.

Time and protein synthesis: **at the body temperature of a mammal or bird, typically a polypeptide chain can grow at the rate of about 10 amino acids per second.** Thus, if a protein is at a typical length of say 150 amino acids, it will take 15 seconds to construct (exclusive of initiation and termination).

POST-TRANSLATIONAL MODIFICATION: THE ROLE OF THE "ENVIRONMENT"

A structural gene directly specifies only the primary structure of a protein. The exact types of higher order structure that will develop depend very much on the sort of environment a protein finds itself in and the types of interactions a protein has with other compounds, especially other proteins.

In fact all of these things could be lumped together as the "environment". In this case, the environment is the chemical and physical conditions that a protein encounters in a cell over its lifetime (yes proteins eventually "die" -- they are broken back down into amino acids). Let's look at these factors.

In all cells, there is an expected normal environment with respect to pH , temperature, the concentrations of various ions, and osmolarity. All of these factors clearly can affect the higher order structure of proteins. This expected environment is achieved partly as a result of regulation of these variables and partly as a matter of the conditions normally found in the external environment of the organism. For instance, in animals that cannot regulate their body temperature, their cells will be at whatever temperature is found in the general environment. The animal will only be able to exist in environments where the range of temperatures is such that proteins fold and perform correctly.

Where does regulation occur in cells and in organisms? List a couple of ways that a cell could attempt to regulate things such as osmolarity, pH and ion concentrations.

Explain how changes in ionic concentration and pH would affect the higher order (2° and above) structure of proteins but not their primary structures.

Proteins assume their higher order structures in part during the translation process. As the protein is created and lengthens, various side groups begin to interact and cause various types of bending and folding. Amino acids with non-polar side groups are twisted into the center of the protein away from water and from polar side groups. Obviously, any thing that affects the polar properties of the cytoplasm will affect the folding process -- either more or less of any polar solute such as H^+ or any other ions can obviously result in a protein with a different shape.

Beyond this there are often specific factors that work only on certain types or parts of proteins and that can have very significant effects. Here are some examples of what are generally called **post-translational modifications**.

Disulfide bridges: one type of amino acid, cysteine, contains sulfur in its side chain. This side chain can react with other cysteines -- the sulfur atoms in each side chain will bind to each other (C -- S -- S -- C) in what is called a disulfide bridge. Wherever these form, a relatively strong connection is made between different areas of a protein. These covalent bonds tend to hold a molecule in

particular shape with much more force than H bonds. It takes specific enzymes in most cases to catalyze the formation of disulfide bridges.

Change or modification of the primary sequence: Once again, an enzyme removes particular amino acids from end or the other or chemically modifies the side groups of particular types of amino acids. It should be obvious that such alterations may cause tremendous changes in the higher order structure. Try to visualize what would happen to a large folded molecule if a portion of it were removed. In other cases, enzymes cut the "backbone" of the protein in a particular place and this induces important structural changes.

An important example is a functional condition for enzymes, where they are inactive and waiting to be activated. Such enzymes are usually called **zymogens**. Zymogens are enzymes that are produced so that they fold up into an inactive configuration. However, when needed, some factor acts on them to cause them to rapidly change shape into an active form. Good examples are a number of digestive enzymes and proteins involved in blood clotting. Obviously, you don't want digestive enzymes to be produced in their active state -- if you did, they might well digest the contents of the cell that makes them. Instead, you want to be able to turn them on only under the appropriate conditions. Likewise with blood clotting. You want the proteins needed for clotting to be present in sufficient numbers but you do not want them to do anything until you are bleeding. So specific substances are released when bleeding begins that interact with these clotting factors and cause them to become active.

What cellular structure would normally contain zymogens?

Questions:

1. Terms to know:

aminoacyl-tRNA synthetases

PP_i

aminoacyl-tRNA

primary transcript

capping

poly-A tail

splicing

spliceosome

mRNA processing

ribozyme

translation

A, P, and E sites

large and small subunits

initiation factors

start codon

translocation

reading frame

Termination

stop codon

release factor

disulfide bridge

zymogen

post-translational modification

2. What events in protein synthesis require energy? (include all steps starting with a completely edited (processed) mRNA transcript)

3. Describe the process of translation.

4. In general terms discuss how information flows from DNA to the cell and how information that is ultimately stored in genes results in both the structure of proteins and in overall biological process. Is information in genes all that is needed to construct a functional protein? Is the environment within a cell that acts on a particular protein independent of the action of other genes?