Translation :

From RNA to Protein

The RNA polymerase has done its job (or in the case of prokaryotes, may still be in the process of doing its job), so now what happens to the RNA? For RNA that is destined to provide instructions for making a protein, then it needs to be translated, which is a job for Superman^{tm}! Oops, actually it's a job for ribosomes.

Ribosomes are a complex of RNA and protein that bind to and processively move down (from 5' to 3' end) a strand of mRNA, picking up aminoacyl-tRNAs, checking to see if they are complementary to the RNA tri-nucleotide being "read" at the moment, and adding them to the new polypeptide chain if they are. The RNA part of the ribosomes are generated by the organism's general purpose RNA polymerase in prokaryotes, and generated by the RNA polymerases I and III in eukaryotes. Recall that RNA Pol II is used by eukaryotes to generate protein-coding mRNA's. Although the numbers of RNA strands and protein subunits differ between the prokaryote and eukaryote, the mechanism for translation is remarkably well conserved.

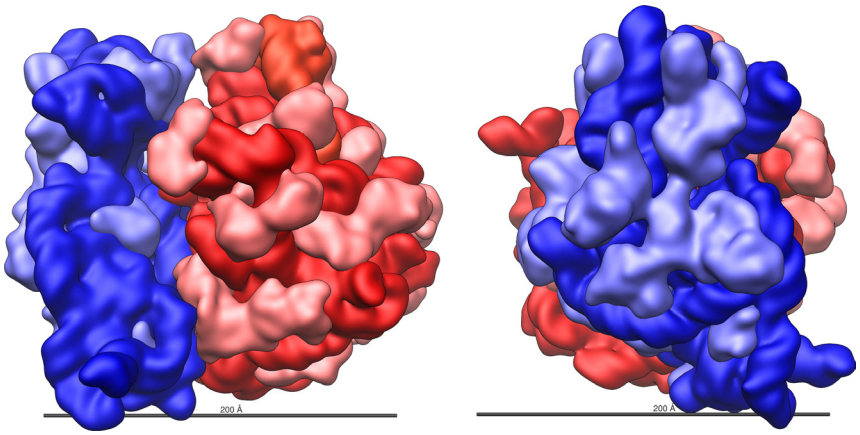


Figure 1. Two views of a prokaryotic ribosome. The large ribosomal subunit (50S) is shown in red, while the small ribosomal subunit (30S) is shown in blue. 3D images generated from data in the RCSB Protein Data Bank.

Using this book: This book is designed to be used in both introductory and advanced cell biology courses. The primary text is generally on the left side of the vertical divider, and printed in black. Details that are usually left to an advanced course are printed in blue and found on the right side of the divider. Finally, additional biomedically relevant information can be found in red print on either side of the divider.

Prokaryotic ribosomes

The prokaryotic ribosomes contain 3 RNA strands and 52 protein subunits which can be divided into 1 RNA and 21 proteins in the small ribosomal subunit (aka the 30S subunit) and 2 RNA and 31 proteins in the large ribosomal subunit (50S subunit). The small subunit locates the start site and moves along the RNA. The large ribosomal subunit contains the aminoacyl transferase enzyme activity that connects amino acids to make a protein. Neither subunit is sufficient to carry out translation by itself. They must come together to form the full 70S ribosome for translation to occur. If you aren't already familiar with the nomenclature, you're probably thinking that it's obvious why I went into biology rather than math. My competence at basic computation aside, there is a method to the madness. The "S" in 30S or 50S indicates Svedberg units, or a measurement of the sedimentation rate when the molecules in question are centrifuged under standard conditions. Because the rate of sedimentation depends on both the mass *and the shape* of a molecule, numbers do not always add up.

The genes for the prokaryotic rRNA molecules are arranged in an operon and thus come from a single transcript. Depending on the organism, there may be several such operons in the genome to ensure steady production of this crucial enzymatic complex. However, these RNAs are not translated, so instead of having multiple translation start codons to signal the beginning of each gene, the single transcript is cleaved post-transcriptionally by Ribonuclease III (RNase III) into 25S, 18S, and 5S segments, and these are then further trimmed by RNase III and RNase M into the final 23S, 16S, and 5S rRNAs found in the ribosomes. Although the 5S does not appreciably differ in sedimentation rate, it is in fact slightly "shaved" in post-transcriptional editing.

Eukaryotic ribosomes

Like the RNA molecules in prokaryotic ribosomes, the eukaryotic rRNA molecules are also post-transcriptionally cleaved from larger transcripts. This processing, and the subsequent assembly of the large and small ribosomal subunits are carried out in the nucleolus, a region of the nucleus specialized for ribosome production, and containing not only high concentrations of rRNA and ribsomal proteins, but also RNA polymerase I and RNA polymerase III. In contrast, RNA polymerase II, as befits its broader purpose, is found throughout the nucleus. The 40S small ribosomal subunit in eukaryotes also has just 1 rRNA, and has 33 proteins. The 60S, or large ribosomal subunit in eukaryotes has three rRNA molecules, two of which are roughly analogous to the prokaryote (28S and 5S eukaryotic, 23S and 5S prokaryotic), and one, the 5.8S, that binds with complemen-

Because of the density of material in the nucleolus needed for constant ribosome production, it is often readily visible under various types of microscopy despite not being bounded by a membrane.

tary sequence on part of the 28S rRNA. It also contains 50 proteins. These ribosomal subunits have roughly the same function as the prokaryotic versions: the small subunit in conjunction with various initiation factors is responsible for finding the start site and positioning the ribosome on the mRNA, while the large subunit houses the docking sites for incoming and spent aminoacyl-tRNAs and contains the catalytic component to attach amino acids via peptide bonds.

The ribosomal RNA precursors (pre-rRNA) are remarkably conserved in eukaryotes, with the 28S, 5.8S, and 18S rRNAs encoded within a single transcript. This transcript, synthesized by RNA polymerase I, always has the same 5' to 3' order: 18S, 5.8S, 28S. After the 45S pre-rRNA is transcribed, it is immediately bound by nucleolar proteins in preparation for cleavage and base modification. However, it is primarily the small nuclelolar RNAs (snoRNAs), not the proteins, that determine the position of the modifications. The primary modifications are 2'-hydroxylmethylation and transformation of some uridines into pseudouridines. These snoRNAs are sometimes transcribed independently by RNA polymerase III or II, but often are formed from the introns of pre-mRNA transcripts. Oddly, some of these introns come from pre-mRNAs that form unused mRNAs!

For the nucleolus to be the site of ribosome assembly, the ribosomal proteins must be available to interact with the rRNAs. By mechanisms discussed in the next chapter, the mRNAs for the ribosomal proteins are translated (as are all proteins) in the cytoplasm, but the resulting proteins are then imported into the nucleus for assembly into either the large or small ribosomal subunit. The subunits are then exported back out to the cytoplasm, where they can carry out their function.

The Genetic Code

We have blithely described the purpose of the DNA chromosomes as carrying the information for building the proteins of the cell, and the RNA as the intermediary for doing so. Exactly how is it, though, that a molecule made up of just four different nucleotides joined together (albeit thousands and even thousands of thousands of them), can tell the cell which of twenty-odd amino acids to string together to form a functional protein? The obvious solution was that since there are not enough individual unique nucleotides to code for each amino acid, there must be combinations of nucleotides that designate particular amino acids. A doublet code, would allow for only 16 different combinations (4 possible nucleotides in the first position x 4 possible nucleotides in the 2nd position = 16 combinations) and would not be enough to encode the 20 amino acids. However, a triplet code would yield 64 combinations, easily enough to encode

20 amino acids. So would a quadruplet or quintuplet code, for that matter, but those would be wasteful of resources, and thus less likely. Further investigation proved the existence of a triplet code as described in the table below.

1st base

2nd base	U		C		A		G		3rd base
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine	C
	UUA	Leucine	UCA	Serine	UAA	Stop	UGA	Stop	A
	UUG	Leucine	UCG	Serine	UAG	Stop	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine	C
	CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine	A
	CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine	G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine	C
	AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine	A
	AUG	Methionine (Start)	ACG	Threonine	AAG	Lysine	AGG	Arginine	G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic Acid	GGU	Glycine	U
	GUC	Valine	GCC	Alanine	GAC	Aspartic Acid	GGC	Glycine	C
	GUA	Valine	GCA	Alanine	GAA	Glutamic Acid	GGA	Glycine	A
	GUG	Valine	GCG	Alanine	GAG	Glutamic Acid	GGG	Glycine	G

Nonpolar, aliphatic Polar, uncharged Aromatic Positively charged Negatively charged

With so many combinations and only 20 amino acids, what does the cell do with the other possibilities? The genetic code is a *degenerate code*, which means that there is redundancy so that most amino acids are encoded by more than one triplet combination (codon). Although it is a redundant code, it is not an ambiguous code: under normal circumstances, a given codon encodes one and only one amino acid. In addition to the 20 amino acids, there are also three "stop codons" dedicated to ending translation. The three stop codons also have colloquial names: UAA (ochre), UAG (amber), UGA (opal), with UAA being the most common in prokaryotic genes. Note that there are no dedicated start codons: instead, AUG codes for both methionine and the start of translation, depending on the circumstance, as explained forthwith. The initial Met is a methionine, but in prokaryotes, it is a specially modified formyl-methionine (f-Met). The tRNA is also specialized and is different from the tRNA that carries methionine to the ribosome for addition to a growing polypeptide. Therefore, in referring to a the loaded initiator tRNA, the usual nomenclature is fMet-$tRNA_i$ or fMet-$tRNA_f$. There also seems to be a little more leeway in defining the start site in prokaryotes than in eukaryotes, as some bacteria use GUG or UUG. Though these codons normally encode valine and leucine, respectively, when they are used as start codons, the initiator tRNA brings in f-Met.

The colloquial names were started when the discoverers of UAG decided to name the codon after a friend whose last name translated into "amber". Opal and ochre were named to continue the idea of giving stop codons color names.

The stop codons are sometimes also used to encode what are now considered the 21st and 22nd amino acids, selenocysteine (UGA) and pyrrolysine (UAG). These amino acids have been discovered to be consistently encoded in some species of prokarya and archaea.

Although the genetic code as described is nearly universal, there are some situations in which it has been modified, and the modifications retained in evolutionarily stable environments. The mitochondria in a broad range of organisms demonstrate stable changes to the genetic code including converting the AGA from encoding arginine into a stop codon and changing AAA from encoding lysine to encoding asparagine. Rarely, a change is found in translation of an organismic (nuclear) genome, but most of those rare alterations are conversions to or from stop codons.

Other minor alterations to the genetic code exist as well, but the universality of the code in general remains. Some mitochondrial DNAs can use different start codons: human mitochondrial ribosomes can use AUA and AUU. In some yeast species, the CGA and CGC codons for arginine are unused. Many of these changes have been cataloged by the National Center for Biotechnology Information (NCBI) based on work by Jukes and Osawa at the University of California at Berkeley (USA) and the University of Nagoya (Japan), respectively.

tRNAs are rather odd ducks

In prokaryotes, tRNA can be found either as single genes or as parts of operons that can also contain combinations of mRNAs or rRNAs. In any case, whether from a single gene, or after the initial cleavage to separate the tRNA transcript from the rest of the transcript, the resulting pre-tRNA has an N-terminal leader (41 nt in E. coli) that is excised by RNase P. That cleavage is universal for any prokaryotic tRNA. After that, there are variations in the minor excisions carried out by a variety of nucleases that produce the tRNA in its final length though not its mature sequence, as we will see in a few paragraphs.

Eukaryotic pre-tRNA (transcribed by RNA polymerase III) similarly has an N-terminal leader removed by RNase P. Unlike the prokaryote though, the length can vary between different tRNAs of the same species. Some eukaryotic pre-tRNA transcripts also contain introns, especially in the anticodon loop, that must be spliced out for the tRNA to function normally. These introns are different from the self-splicing or spliceosome-spliced transcripts discussed in the transcription chapter. Here, the splicing function is carried out not by ribozymes, but by conventional (protein) enzymes. Interestingly, RNaseP also removes a 3' sequence from the pre-tRNA, but then another 3' sequence is added back on. This new 3' end is always CCA, and is added by three successive rounds with tRNA nucleotidyl transferase.

Earlier in this textbook, when RNA was introduced, it was noted that although extremely similar to DNA on many counts, it is normally single stranded, and that property, combined with the opportunity for complementary base pairing within a strand, allows it to do something far different than double-stranded DNA: it can form highly complex secondary structures. One of the simplest and clearest examples of this is tRNA, which depends on its conformation to accomplish its cellular function. The prototypical cloverleaf-form tRNA diagram is shown in fig. 2 on the left, with a 3D model derived from x-ray crystallographic data on the right. As you can see, the fully-splayed-out shape

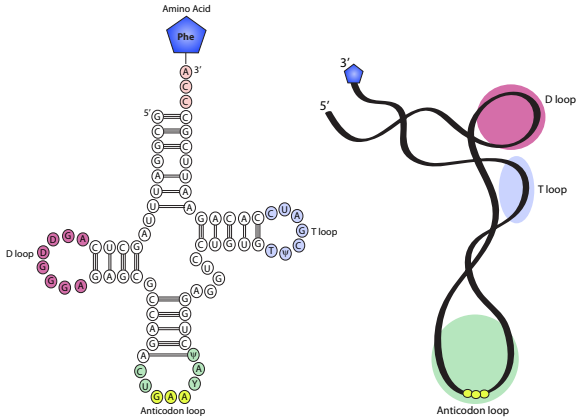


Figure 2. tRNA structure. At left, the classic clover-leaf, splayed out for simplicity. At right, a more accurate representation of the tRNA in pseudo-3D.

has four stem-loop "arms" with the amino acid attached to the acceptor arm, which is on the opposite side of the tRNA from the anticodon arm, which is where the tRNA must match with the mRNA codon during translation. Roughly perpendicular to the acceptor-anticodon axis are the D arm and the TψC arm. In some tRNAs, there are actually five total arms with a very short loop between the TψC and anticodon arms. The arm-like stem and loop structures are formed by two areas of strong complementarity (the stems, base-paired together) interrupted by a short non-complementary sequence (the loop). In general terms, the arms are used to properly position the tRNA within the ribosome as well as recognizing the mRNA codon and bringing in the correct amino acid.

When it comes time for the tRNA to match its anticodon with the codon on the mRNA, the code is not followed "to the letter" if you will pardon the pun. There is a phenomenon called "wobble" in which a codon-anticodon match is allowed and stabilized for translation even if the nucleotide in the third position is not complementary. Wobble can occur because the conformation of the tRNA allows a little flexibility to that position of the anticodon, permitting H-bonds to form where they normally would not. This is not a universal phenomenon though: it only applies to situations where a U or a G is in the first position of the anticodon (matching the third position of codon). Following the convention of nucleic acid sequences, the sequence is always written 5' to 3', even though in the case of codon-anticodon matching, the strands of mRNA and tRNA are antiparallel:

```
tRNA  3' U A A 5'
mRNA  5' A U U 3'
```

In addition to being allowed a bit of wobble in complementary base-pairing, tRNA molecules have another peculiarity. After being initially incorporated into a tRNA through conventional transcription, there is extensive modification of some of the bases of the tRNA. This affects both purines and pyrimidines, and can range from simple additions such as methylation or extensive restructing of the sugar skeleton itself, as in the conversion of guanosine to wyosine (W). Over 50 different modifications have been catalogued to date. These modifications can be nearly universal, such as the dihydrouridine (D) found in the tRNA D loop, or more specific, such as the G to W conversion found primarily in $tRNA^{Phe}$ of certain species (examples have been identified in both prokaryotic and eukaryotic species). As many as 10% of the bases in a tRNA may be modified. Naturally, alterations to the sugar base of the nucleotides can also alter the base-pairing characteristics. For example, one common modified base, inosine, can complement U, C, or A. This aberrant complementary base pairing can be equal among the suitor bases, or it may be biased, as in the case of 5-methoxyuridine, which can recognize A, G, or U, but the recognition of U is poor.

The knowledge of the genetic code begs the question: how is the correct amino acid attached to any given tRNA? A class of enzymes called the *aminoacyl tRNA synthetases* are responsible for recognizing both a specific tRNA and a specific amino acid, binding an ATP for energy and then joining them together (sometimes called charging the tRNA) with hydrolysis of the ATP. Specificity is a difficult task for the synthetase since amino acids are built from the same backbone and are so similar in mass. Distinguishing between tRNA molecules is easier, since they are larger and their secondary structures also allow for greater variation and therefore greater ease of discrimination. There is also a built-in pre-attachment proofreading mechanism in that tRNA molecules that fit the synthetase well (i.e. the correct ones) maintain contact longer and allow the reaction to proceed whereas ill-fitting and incorrect tRNA molecules are likely to disassociate from the synthetase before it tries to attach the amino acid.

Charging an aminoacyl tRNA synthetase with its amino acid requires energy. The synthetase first binds a molecule of ATP and the appropriate amino acid, which react resulting in the formation of aminoacyl-adenylate and pyrophosphate. The PPi is released and the synthetase now binds to the proper tRNA. Finally, the amino acid is transferred to the tRNA. Depending on the class of synthetase, the amino acid may attach to the 2'-OH of the terminal A (class I) or to the 3'-OH of the terminal A (class II) of the tRNA. Phe-tRNA synthetase is the exception: it is structurally a class II enzyme but transfers the Phe onto the 2'-OH. Note that amino acids transferred onto the 2'-OH are soon moved to the 3'-OH anyway due to a transesterification reaction.

As soon as the RNA has emerged from the RNAP and there is sufficient space to accomodate a ribosome, translation can begin in prokaryotes. In fact, for highly expressed genes, it would not be unusual to see multiple RNA polymerases transcribing the DNA and multiple ribosomes on each of the transcripts translating the mRNA to protein! The process begins with the small ribosomal subunit (and only the small subunit - if it is attached to the large subunit, it is unable to bind the mRNA), which binds to the mRNA loosely and starts to scan it for a recognition sequence called the Shine-Dalgarno sequence, after its discoverers. Once this is recognized by the small ribosomal subunit rRNA, the small subunit is positioned around the start codon (AUG). This process is facilitated by initiation factors as follows. The 30S ribosomal subunit dissociates from the 50S ribosomal subunit if it was associated with one, and binds to intiation factors IF-1 and IF-3. IF-1 binds to the A site, where it prevents new aminoacyl-tRNA molecules from entering before the full ribosome is assembled. It also facilitates the assembly and stabilization of the initiation complex. IF-3 is required to allow the 30S subunit to bind to mRNA. Once this has occurred, IF-2-GTP arrives on scene, carrying with it the initiator aminoacyl-tRNA. This settles into the P site, which is positioned so that the anticodon of the tRNA settles over the AUG start codon of the mRNA. Hydrolysis of the GTP attached to IF-2 and release of all the initiation factors is needed to allow the 50S subunit to bind to the 30S subunit to form the full and fully functional ribosome. Because GTP hydrolysis was required, the joining of the subunits is irreversible spontaneously, and requires expenditure of energy upon termination of translation. Once the 50S subunit joins with the 30S subunit, the A site is ready to accept the next aminoacyl-tRNA.

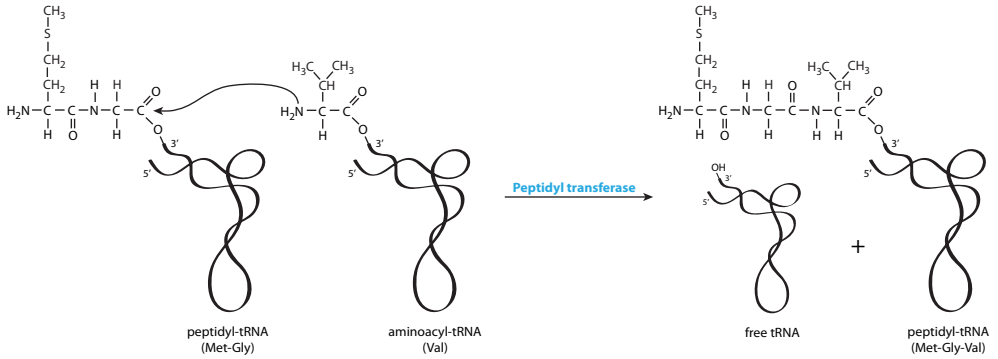


Figure 4. Peptide bond formation at the addition of the third amino acid. The previous two amino acids are peptide bonded together as well as attached to the tRNA from the second amino acid. The aminoacyl-tRNA bond is broken and transferred/transformed to the peptide bond connecting the initial dipeptide to the third amino acid.

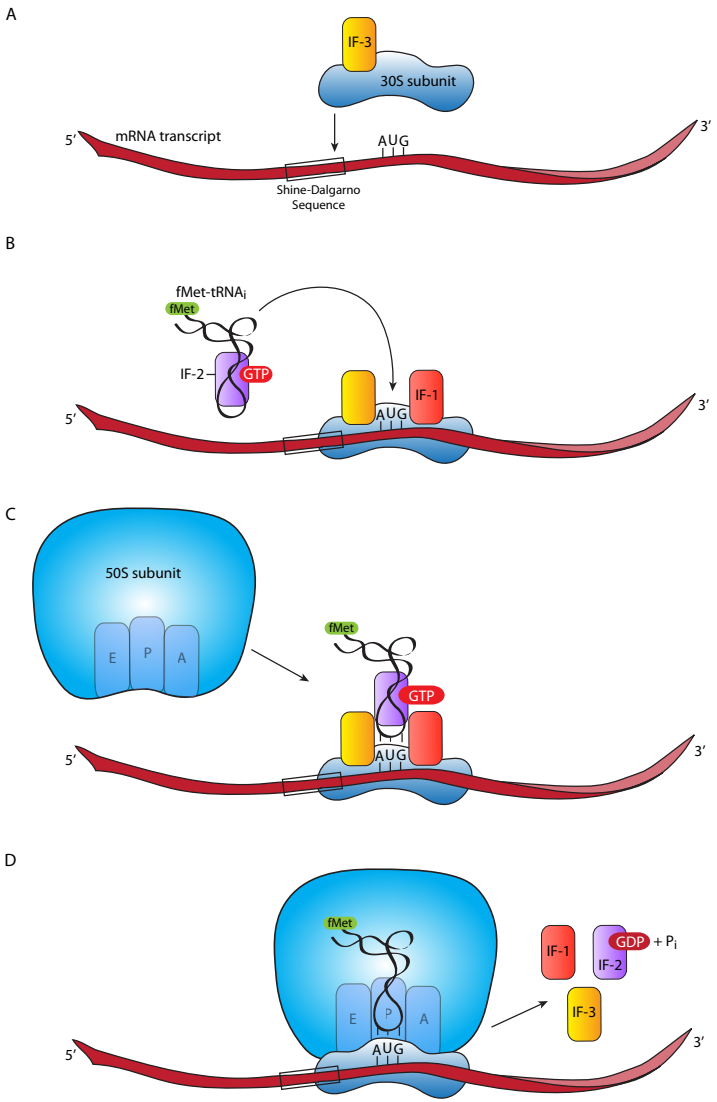


Figure 3. Initiation of Translation in Prokaryotes. (A) 30S subunit binds to Shine-Dalgarno sequence. (B) fMet-tRNAi is loaded into the middle slot of the small ribosomal subunit. Initiation factors occupyh the other two slots. (C) The large ribosomal subunit docks with the small subunit. (D) The initiation factors are released and the ribosome is ready to start translation.

A common and understandable misconception is that the new amino acid brought to the ribosome is added onto the growing polypeptide chain. In fact, the mechanism is exactly the opposite: the polypeptide is added onto the new amino acid (fig. 4). This begins from the second amino acid to be added to a new protein (fig. 5). The first amino acid, a methionine, you should recall, came in along with IF-2 and the initiator tRNA. The new aminoacyl-tRNA is escorted by EF-Tu, an elongation factor that carries a GTP. Once the aa-tRNA is in place, EF-Tu hydrolyzes the GTP and dissociates from the aminoacyl-tRNA and ribosome.

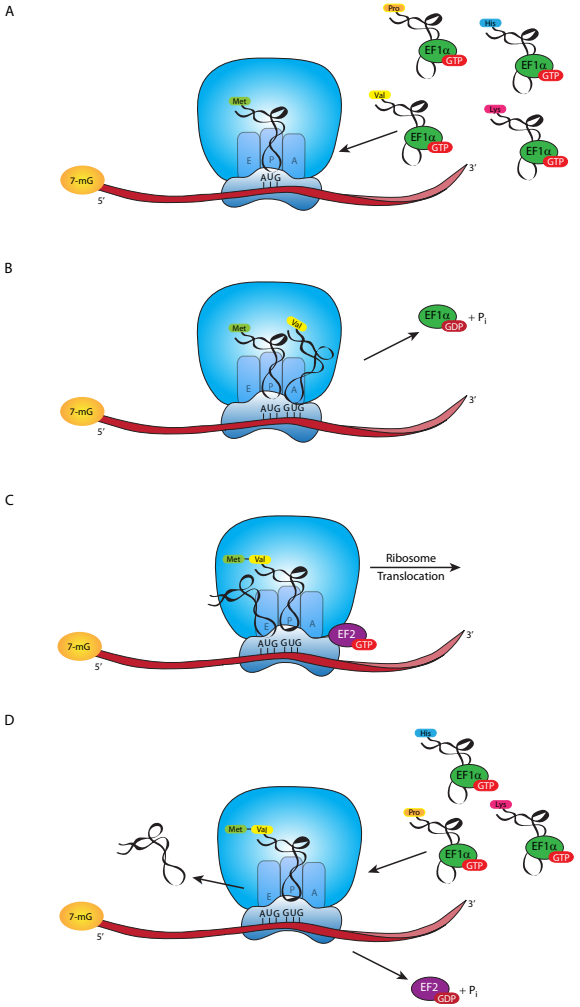


Figure 5. Elongation of the polypeptide chain. (A) a new aminoacyl-tRNA drops into the A slot of the ribosome. (B) the fMet is moved off its $tRNA_i$ and peptide-bonded to the new amino acid, which is still attached to its tRNA. (C) Ribosome shifts over to the right. (D) empty $tRNA_i$ is ejected.

For a long time, there was a bit of mystery surrounding the simultaneous docking of two tRNA molecules on immediately adjacent codons of mRNA. Under normal conditions, there should not be enough room, since the tRNAs are fairly bulky and one should obstruct the other from reaching the mRNA to make a codon-anticodon match. The matter was finally cleared up in 2001 with x-ray crystallographic examinations showing a bend in the mRNA between the codon in the P slot and the codon in the A slot. The bend puts the two associated tRNAs at slightly different angles and thus creates just enough room for both to maintain basepairing hydrogen bonds with the mRNA. See Yusupov et al, *Science* 292 (5518): 883-896, 2001.

When a new aminoacyl-tRNA drops into the A slot of the ribosome, the anticodon is lined up with the codon of the mRNA. If there is no complementarity, the aminoacyl-tRNA soon floats back out of the slot to be replaced by another candidate. However, if there is complementarity (or something close enough, recalling the idea of wobble) then H-bonds form between the codon and anti-codon, the tRNA changes conformation, which shifts the conformation of EF-Tu, causing hydrolysis of GTP to GDP + Pi, and release from the aa-tRNA. The codon-anticodon interaction is stable long enough for the catalytic activity of the ribosome to hydrolyze the bond between fMet and the tRNAf in the P slot, and attach the fMet to the new amino acid with a peptide bond in the A slot. The new amino acid is still attached to its tRNA, and as this process occurs, the ribosome shifts position with respect to the mRNA and tRNAs. This puts the now-empty (no amino acid attached) tRNAf in the E slot, the tRNAaa in the P slot, attached to that aa which is bonded to Met, and the A slot is again open for a new tRNA to come in. The elongation factor EF-G binds near the A slot as soon as EF-Tu leaves, and is required for ribosomal translocation, providing energy for the process by hydrolyzing a GTP that it carries with it to the ribosome. From my students' experiences, the best way to learn this seems to be to study the diagrams and see the movements of the molecules, filling in the mechanistic details in your mind. This process continues until the ribosome brings the A slot in line with a stop codon.

There is no tRNA with an anticodon for the stop codon. Instead, there is a set of release factors that fit into the A site of the ribosome, bind to the stop codon, and actvate the ribosome to cut the bond between the polypeptide chain and the last tRNA (fig. 6). Depending on which stop codon is present either RF1 (recognizes UAA or UAG) or RF2 (for UAA or UGA) first enters the A slot. The RF1 or RF2 is complexed with RF3, which is involved in subsequent releasing of the RF complex from the A slot. This is necessary because once the polypeptide has been released from the ribosome, the mRNA must be released. Ribosome releasing factor (RRF) also binds in the A slot, which causes a conformational change in the ribosome releasing the previous and now empty tRNA. Finally, EF-G binds to RRF, and with an accompanying hydrolysis of GTP, causes dissociation of the ribosome into separate large and small subunits. Note that it is the combination of EF-G/RRF that causes dissociation; EF-G alone plays a different role in ribosome movement when it is not at the stop codon.

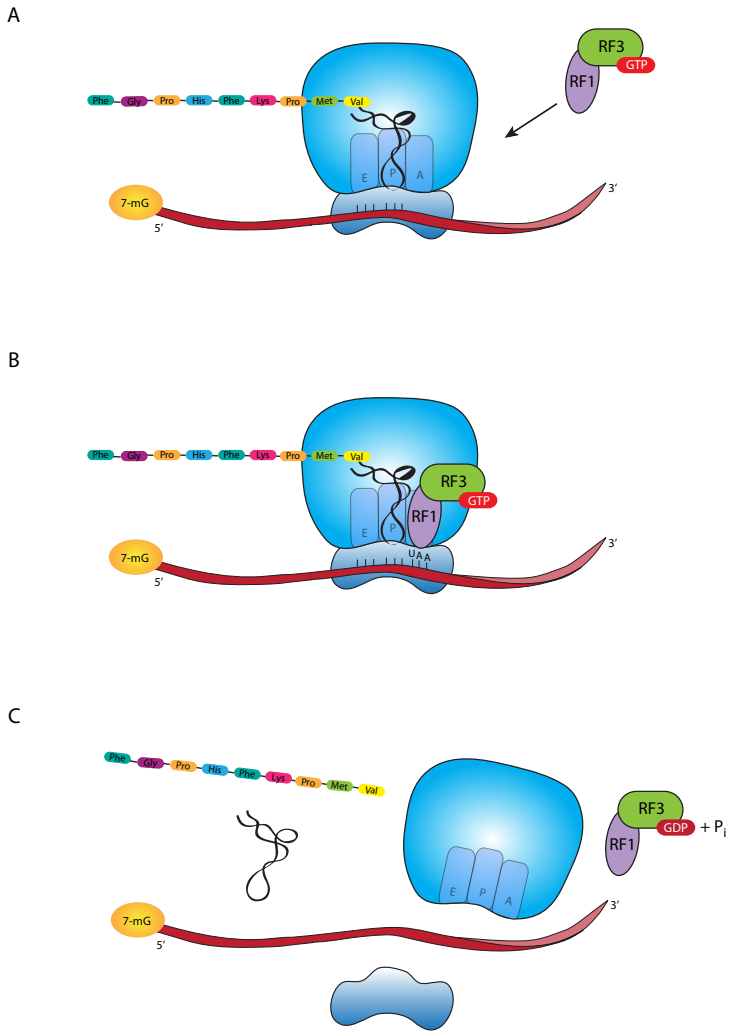


Figure 6. Termination of translation.

Eukaryotic Translation

Eukaryotic translation, as with transcription, is satisfyingly similar (from a student studying point of view, or from an evolutionary conservation one) to the prokaryotic case. The initiation process is slightly more complicated, but the elongation and termination processes are the same, but with eukaryotic homologues of the appropriate elongation and release factors.

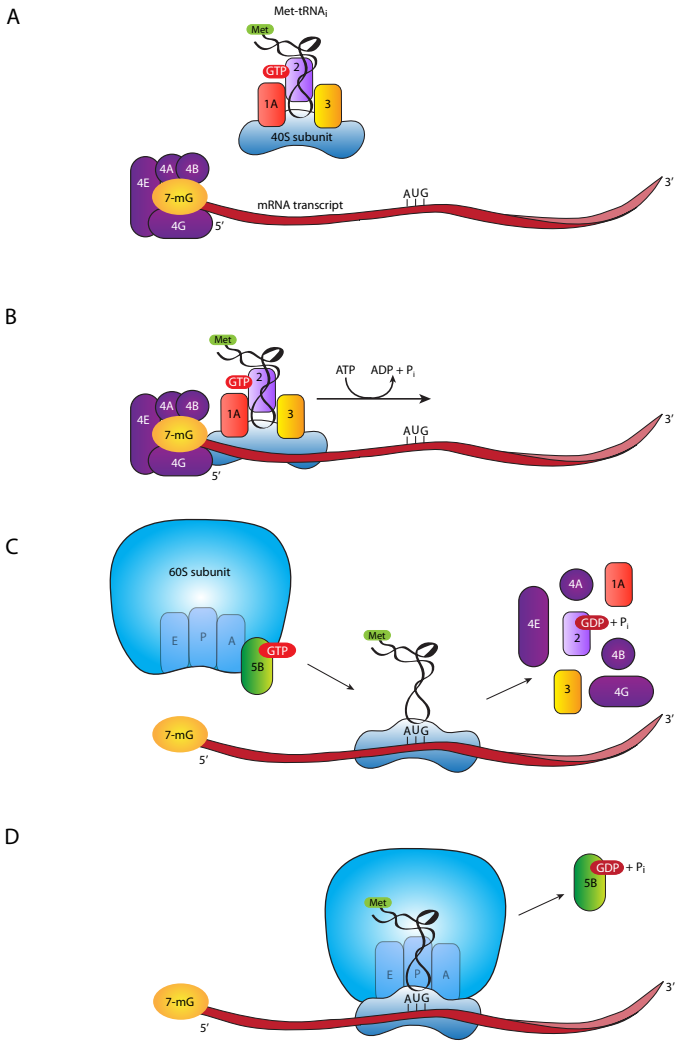


Figure 7. Initiation of Translation in Eukaryotes.

For eukaryotes, each mRNA encodes one and only one gene (as opposed to multi-gene transcripts such as operons), so there isn't much question of which AUG is a start codon, and which are just regular methionines. Therefore, there is no requirement for a Shine-Dalgarno sequence in eukaryotes. The small ribosomal subunit, accompanied by eukaryotic initiation factors eIF-3, eIF-2, and Met-tRNAi, together known as the terneary complex, binds to eIF-1A. Meanwhile, eIF-4A, -4B, -4E, and -4G bind to the 5' (7-methyguanosine) cap of the mRNA (fig. 7A). The small subunit complex and the eIF4/mRNA cap-binding complex interact to form the 43S complex, which then begins scanning the mRNA from 5' to 3' looking for the first AUG.

Once the 43S scanning complex has found the start codon, the initiation factors drop off, and the large ribosomal subunit arrives. The large ribosomal subunit has bound eIF-6, which prevents it from reassociating with small subunits, and its removal is required first. Another factor, eIF-5 enters the scene during the coupling process between the large and small ribosomal subunits, and hydrolysis of an eIF-5-attached GTP is required to complete the docking of the subunits and the formation of a complete functional ribosome on the mRNA.

Elongation is functionally the same as in prokaryotes except that the functions of EF-Tu is taken care of by EF-1α, also with hydrolysis of GTP. EF-2 is the eukaryotic analog of EF-G, and utilizes GTP hydrolysis for translocation of the ribosome. Termination uses eukaryotic homologues of the release factors, though eRF-1 takes the place of both prokaryotic RF-1 and RF-2.

Although polyribosomes (aka polysomes) can form on both prokaryotic and eukaryotic mRNAs, eukaryotic polysomes have an additional twist. Technically, a polysome is simply an mRNA with multiple ribosomes translating it simultaneously, but in eukaryotes, the polysome also has a unique morphology because it utilizes PABPI, or poly-A binding protein. This protein not only binds to the 3' poly-A tail of an mRNA, it also interacts with the eIF-4 initiation factors, which thus loops the mRNA into a circular shape. That way, once the ribosome reaches the end of the gene and releases from the mRNA, it is physically near the beginning of the mRNA to start translating again.

Usually, but not always, the first AUG is the start codon for eukaryotic genes. However, the context of the AUG matters, and it is a much stronger (i.e. more frequently recognized and used) start codon if there is a purine residue (A or G) at -3 and a G at +4. See Kozak, M., *Biochimie* **76**: 815-821, 1994.

Regulation of Translation

Gene expression is primarily regulated at the pre-transcriptional level, but there are a number of mechanisms for regulation of translation as well. One well-studied animal system is the iron-sensitive RNA-binding protein, which regulates the expression of genes involved in regulating intracellular levels of iron ions. Two of these genes, ferritin, which safely sequesters iron ions inside cells, and transferrin, which transports iron from the blood into the cell, both utilize this translational regulation system in a feedback loop to respond to intracellular iron concentration, but they react in opposite ways. The key interaction is between the iron response elements (IRE), which are sequences of mRNA that form short stem-loop structures, and IRE-BP, the protein that recognizes and binds to the IREs. In the case of the ferritin gene, the IRE sequences are situated upstream of the start codon. When there is high iron, the IRE-BP is inactive, and the stem-loop structures are melted and overrun by the ribosome, allowing translation of ferritin, which is an iron-binding protein. As the iron concentration drops, the IRE-BP is activated and binds around the IRE stem-loop structures, stabilizing them and preventing the ribosome from proceeding. This prevents the production of ferritin when there is little iron to bind.

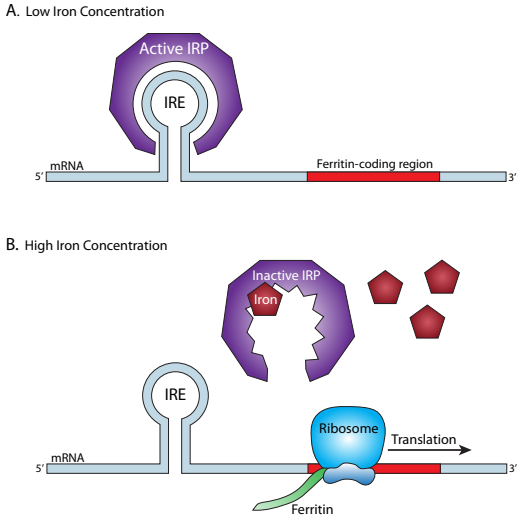


Figure 8. Pre-translational control of gene expression by iron-response protein (IRP), which binds to either the iron-response element (IRE), unless it has bound iron.

Transferrin also uses iron response elements and IRE-binding proteins, but in a very different mechanism. The IRE sequences of the transferrin gene are located downstream of the *stop* codon, and play no direct role in allowing or preventing translation.

However, when there is low intracellular iron and there is a need for more transferrin to bring iron into the cell, the IRE-BP is activated as in the previous case, and it binds to the IREs to stabilize the stem-loop structures. In this case; however, it prevents the 3' poly-A tail degradation that would normally occur over time. Once the poly-A tail is degraded, the rest of the mRNA is destroyed soon thereafter. As mentioned in the transcription chapter, the longer poly-A tails are associated with greater persistence in the cytoplasm, allowing more translation before they are destroyed. The IRE-BP system in this case externally prolongs the lifetime of the mRNA when that gene product is needed in higher amounts.

Since mRNA is a single-stranded nucleic acid and thus able to bind complementary sequence, it is not too surprising to find that one of the ways that a cell can regulate translation is using another piece of RNA. Micro RNAs (miRNAs) were discovered as very short (~20 nucleotides) non-protein-coding genes in the nematode, C. elegans. Since their initial discovery (Lee et al, *Cell* **75**: 843-54, 1993), hundreds have been found in various eukaryotes, including humans. The expression pattern of the miRNA genes is highly specific to tissue and developmental stage. Many are predicted to form stem-loop structures, and appear to hybridize to 3'-untranslated sequences of mRNA thus blocking initiation of translation on those mRNA molecules. They may also work through a mechanism similar to the siRNA discussed below, but there is clear evidence that mRNA levels are not necessarily altered by miRNA-directed translational control.

MicroRNAs are currently under investigation for their roles as either oncogenes or tumor suppressors (reviewed in Garzon et al, *Ann. Rev. Med.* **60**: 167-79, 2009). Approximately half of known human miRNAs are located at fragile sites, breakpoints, and other regions associated with cancers (Calin et al, *Proc. Nat. Acad. Sci. (USA)* **101**: 2999-3004, 2004). For example, miR-21 is not only up-regulated in a number of tumors, its overexpression blocks apoptosis - a necessary step to allow abnormal cells to contine to live and divide rather than die out. Conversely, miR-15a is significantly depressed in some tumor cells, and overexpression can slow or stop the cell cycle, even inducing apoptosis.

Another mechanism for translational control that uses small RNA molecules is RNA interference (RNAi). This was first discovered as an experimentally induced repression of translation when short *double-stranded* RNA molecules, a few hundred nucleotides in length and containing the same sequence as a target mRNA, were introduced into cells. The effect was dramatic: most of the mRNA with the target sequence was quickly destroyed. The current mechanistic model of RNAi repression is that first, the double-stranded molecules are cleaved by an endonuclease called *Dicer*, which cleaves with over-hanging single-stranded 3' ends. This allows the short fragments (siRNA, ~20nt long) to form a complex with several proteins (RISC, RNA-induced silencing complex). The RISC splits the double-stranded fragments into single strands, one of which is an exact complement to the mRNA. Because of the complementarity, this is a stable interaction, and the double-stranded region appears to signal an endonuclease to destroy the mRNA/siRNA hybrid.

The final method of controlling levels of gene expression is control after the fact, i.e., by targeted destruction of the gene product protein. While some proteins keep working until they fall apart, others are only meant for short-term use (e.g. to signal a short phase in the cell cycle) and need to be removed for the cell to function prop-

erly. Removal, in this sense, would be a euphemism for chopped up and recycled. The ubiquitin-proteasome system is a tag-and-destroy mechanism in which proteins that have outlived their usefulness are polyubiquitinated. Ubiquitin is a small (76 amino acids, ~5.6 kDa), highly conserved (96% between human and yeast sequences) eukaryotic protein (fig. 9) that can be attached to other proteins through the action of three sequential enzymatic steps, each catalyzed by a different enzyme.

E1 activates the ubiquitin by combining it with ATP to make ubiquitin-adenylate, and then transfers the ubiquitin to itself via a cysteine thioester bond. Through a trans(thio) esterification reaction, the ubiquitin is then transferred to a cysteine in the E2 enzyme, also known as ubiquitin-conjugating enzyme. Finally, E3, or ubiquitin ligase, interacts with both E2-ubiquitin and the protein designated for destruction, transferring the ubiquitin to the target protein. After several rounds, the polyubiquitinated protein is send to the proteasome for destruction.

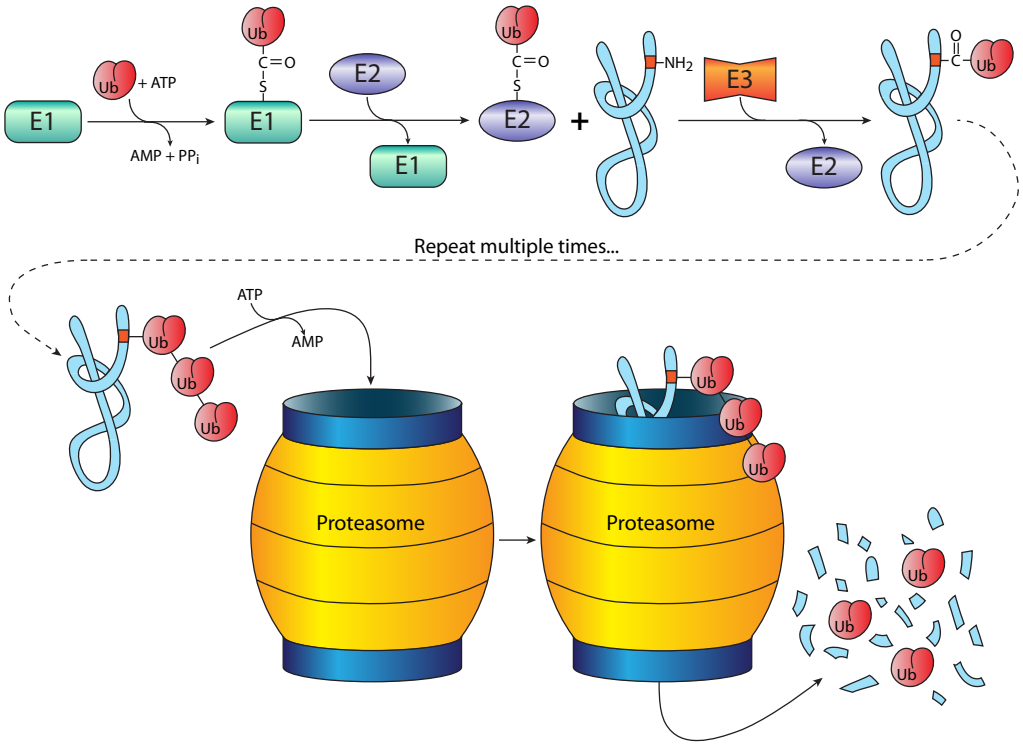
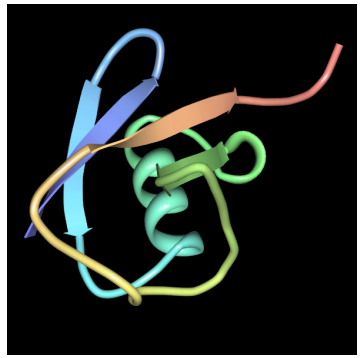


Figure 10. Polyubiquitination of a targeted protein (blue) requires three ubiquitinating enzymes, E1, E2, and E3. Once tagged, the protein is positioned in the proteasome by binding of the polyubiquitin tail to the outer surface of the proteasome. The proteasome then cleaves the protein into small polypeptides.

Figure 9. Ubiquitin. This 3D representation was generated from the file 1ubi (synthetic human ubiquitin) in the RCSB Protein Data Bank



Mutations in E3 genes can cause a variety of human medical disorders such as the neurodevelopmental disorders Angelman syndrome, Hippel-Lindau syndrome, or the general growth disorder known as 3-M syndrome. Mechanisms linking malfunction in ubiquitination pathways and symptoms of these disorders are not currently known.

Proteasomes are very large protein complexes arranged as a four-layered barrel (the 20S subunit) capped by a regulatory subunit (19S) on each end. The two outer rings are each composed of 7 α subunits that function as entry gates to the central rings, each of which is composed of 7 β subunits, and which contain along the interior surface, 6 proteolytic sites. The 19S regulatory units control the opening and closing of the gates into the 20S catalytic barrel. The entire proteasome is sometimes referred to as a 26S particle.

A polyubiquitinated protein is first bound to the 19S regulatory unit in an ATP-dependent reaction (the 19S contains ATPase activity). 19S unit opens the gates of the 20S unit, possibly involving ATP hydrolysis, and guides the protein into the central proteolytic chamber. The protease activity of proteasomes is unique in that it is a threonine protease, and it cuts most proteins into regular 8-9 residue polypeptides, although this can vary.

As we will see in the cell cycle chapter, proteasomes are a crucial component to precise regulation of protein functions.

ER, Golgi and Vesicles :

Post-translational Processing and Vesicular Transport

Using this book: This book is designed to be used in both introductory and advanced cell biology courses. The primary text is generally on the left side of the vertical divider, and printed in black. Details that are usually left to an advanced course are printed in blue and found on the right side of the divider. Finally, additional biomedically relevant information can be found in red print on either side of the divider.

Once a polypeptide has been translated and released from the ribosome, it may be ready for use, but often it must undergo post-translational processing in order to become fully functional. While many of these processes are carried out in both prokaryotes and eukaryotes, the presence of organelles provides the need as well as some of the mechanisms for eukaryote-specific modifications such as glycosylation and targeting.

Proteolytic Cleavage

The most common modification is proteolytic cleavage. Some of the pre-cleavage polypeptides are immediately cleaved, while others are stored as inactive precursors to form a pool of enzymes (or other kinds of proteins) that can be activated very quickly, on a timescale of seconds to minutes, as compared to having to go through transcription and translation, or even just translation. Interestingly, though methionine (Met) is universally the first amino acid of a newly synthesized polypeptide, many proteins have that methionine cleaved off (also true for some prokaryotic f-Met).

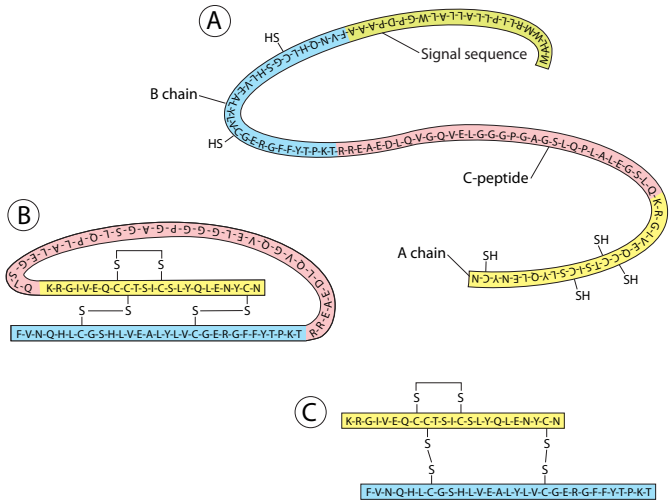


Figure 1. Proteolytic processing is necessary to make biologically active insulin. (A) The linear protein contains a signal sequence, which is cleaved after the protein enters the ER, an A chain, a B chain, and a C-peptide. (B) Inside the ER, the proinsulin (insulin precursor) folds and disulfide bonds form between cysteines. (C) Finally, two cleavages release the C peptide, which leaves the A and B chains attached by the disulfide bonds. This is now active insulin.

Activation of proteins by cleavage of precursors is a common theme: the precursor protein is termed a proprotein, and the peptide that is cleaved off of it to activate the protein is called the propeptide. Among the better known examples of proteins that are derived from proproteins are the hormone insulin, the cell death protein family of caspases, and the Alzheimer-associated neural protein β-amyloid. Insulin is an interesting example (fig. 1) in mammals: preproinsulin (inactive as a hormone) is first translated from the insulin mRNA. After a cleavage that removes an N-terminal sequence, proinsulin (still inactive) is generated. The proinsulin forms some internal disulfide bonds, and when the final proteolytic action occurs, a substantial chunk (called the C-peptide) is taken out of the middle of the proinsulin. Since the protein was internally disulfide bonded though, the two end pieces remain connected to become the active insulin hormone.

Another interesting protein processing example is that of collagen assembly (fig. 2). As you will read in chapter 13, collagen is a very large secreted protein that provides structure and shock absorbance for the extracellular matrix in animals. You can find it in skin, hooves, cartilage, and various connective tissues. An individual collagen protein is actually a twisted triple-helix of three subunits. The collagen subunits are made as procollagen, and propeptides are lopped off of both N- and C- termini to generate the final protein. However, they are not cleaved off until after the three subunits assemble around one another. In fact, collagen subunits that have already been processed do not assemble into triple-helical proteins. The propeptide sequences are clearly necessary for efficient assembly of the final protein complex.

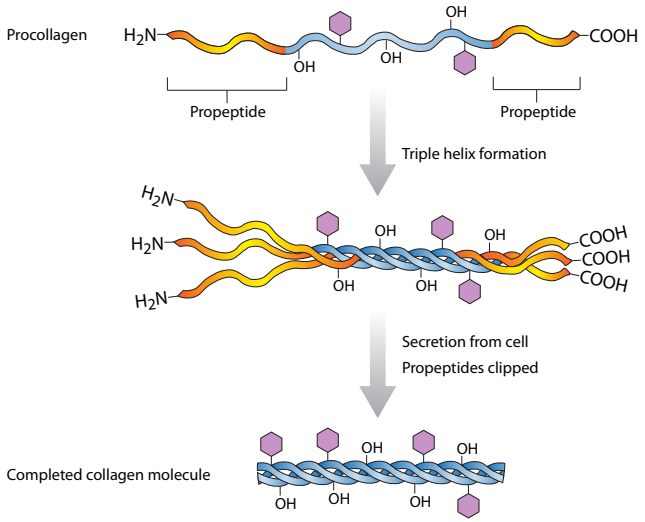


Figure 2. Processing and assembly of procollagen into collagen.

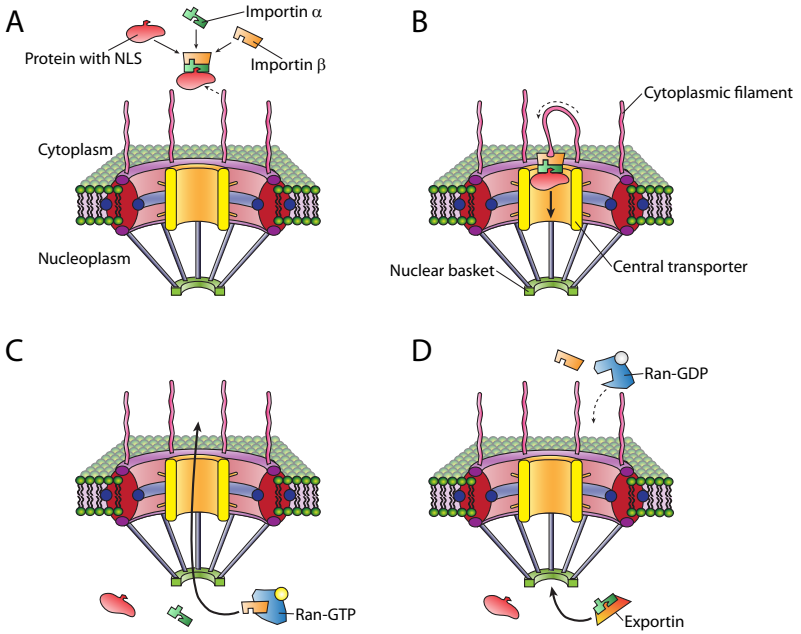
Protein Trafficking

The idea that propeptide sequences have important functions in protein maturation beyond just keeping them from being active is not exclusive to assembly. A major class of cleaved peptide sequences is signal peptides. Signal peptides direct the protein from the cytoplasm into a particular cellular compartment. In the case of prokaryotes, this essentially means the cell membrane, but for eukaryotes, there are specific signal peptides that can direct the protein to the nucleus, to the mitochondria, to the endoplasmic reticulum, and other intracellular organelles. The peptides are specifically recognized by receptors on the membranes of particular compartments, which then help to guide the insertion of the protein into or through the membrane. Almost all protein synthesis in eukaryotes is carried out in the cytoplasm (with the exception of a few proteins in the chloroplasts and mitochondria), so proteins found in any other compartment or embedded in any membrane must have been targeted and transported into that compartment by its signal sequence.

The nucleus is one such compartment, and examples of the proteins found within include DNA and RNA polymerases, transcription factors, and histones. These and other nuclear proteins have an N-terminal signal sequence known as the NLS, or nuclear localization signal. This is a well-studied pathway that involves a set of importin adapter proteins and the nuclear pore complex (fig. 3). Transport into the nucleus is particularly challenging because it has a double membrane (remember that it is contiguous with the endoplasmic reticulum membrane. Although there are other mechanisms for making proteins that are embedded in the nuclear membrane, the primary mechanism for import and export of large molecules into and out of the nucleus itself is the *nuclear pore complex*. The complex is very large and can be made of over 50 different proteins (nucleoporins, sometimes called nups). The nucleoporins are assembled into a large open octagonal pore through the nuclear membranes. As figure 3 indicates, there are antenna-like fibrils on the cytoplasmic face, and these help to guide proteins from their origin in the cytoplasm to the nuclear pore, and on the nuclear side there is a basket structure. Of course, not all proteins are allowed into the nucleus, and the mechanism for distinguishing appropriate targets is straightforward. The protein must bear a nuclear localization signal (NLS). While in the cytoplasm, an importin-α protein binds to the NLS of a nuclear protein, and also binds to an importin-β. The importin-β is recognized and bound by the nuclear pore complex. The details of the transport mechanism are murky, but phenylalanine-glycine repeats in the nucleoporin subunits (FG-nups) are thought to be involved.

Although this is primarily considered a eukaryotic process given that there are so many potential targets, prokaryotes do have membrane proteins (in fact, some 800 different ones in E. coli comprising ~20% of total protein), and they are positioned there with the aid of insertase enzymes such as YidC and complexes such as Sec translocase. The Sec translocase uses a signal recognition particle (SRP) much like that in eukaryotes, and will be discussed later in this chapter when the SRP is introduced. YidC, which has eukaryotic homologues (e.g. Oxa1 in mitochondria), is a 61 kDa transmembrane protein that is placed in the membrane through an SRP-Sec translocase mechanism. Once there, YidC interacts with nascent polypeptides (once they reach ~70 amino acids long) that have begun to interact with the lipids of the cell membrane, and pushes the protein into/through the membrane.

Figure 3. Transport through nuclear pore.



Once the nucleoprotein-importin aggregate is moved into the nucleus, Ran-GTP, a small GTPase, causes the aggregate to dissociate (fig. 3c). The imported protein is released in the nucleus. The importins are also released in the nucleus, but they are exported back out again to be reused with another protein targeted for the nucleus.

Export from the nucleus to the cytoplasm also occurs through the nuclear pore. The Ran-GTP is also a part of the export complex (fig. 3d), and in conjunction with an *exportin* protein and whatever is to be exported, is moved out of the nucleus via the nuclear pore. Once in the cytoplasm, the hydrolysis of GTP to GDP by Ran (activated by Ran-GAP, a cytoplasmic protein) provides the energy to dissociate the cargo (e.g. mRNA) from the exporting transport molecules. The Ran-GDP then binds to importins, re-enters the nucleus, and the GDP is exchanged for GTP.

The nuclear pore is the only transport complex that spans dual membrane layers, although there are coordinated pairs of transport complexes in double-membraned organelles such as mitochondria. The transport proteins in the outer mitochondrial membrane link with transport proteins in the inner mitochondrial membrane to move matrix-bound proteins (e.g. those involved in the TCA cycle) in from the cytoplasm. The complexes that move proteins across the outer membrane are made up of Tom (translocator outer membrane) family of proteins. Some of the proteins will stay em-

The mechanisms of small GTPase activation of other processes will be discussed again in more detail in later chapters (cytoskeleton, signaling). The key to understanding the mechanism is to remember that the GTPase hydrolyzes GTP to GDP, but still holds onto the GDP. Although the GTPase will hydrolyze GTP spontaneously, the GTPase-activating protein, GAP (or Ran-GAP in this case) greatly speeds the rate of hydrolysis. In order to cycle the system back to GTP, the GDP is not re-phosphorylated: it is *exchanged* for a new GTP. The exchange is greatly facilitated by the action of an accessory protein, the guanine nucleotide exchange factor (GEF), in this particular case, a Ran-GEF.

bedded in the outer membrane: they are processed by a SAM (sorting and assembly machinery) complex also embedded in the outer membrane). Meanwhile, others continue to the Tim (translocator inner membrane) proteins that move them across the inner membrane. As with the nuclear proteins, there is a consensus signal sequence on mitochondrial proteins that is bound by cytosolic chaperones that bring them to the Tom transporters. As shown in the table below, there are signal sequences/propeptides that target proteins to several other compartments.

Nucleus	–PPKKKRKV–
ER entry	MMSFVSLLLVGILFWATEAEQLTKCEVFQ–
ER retention	–KDEL
Mito. Matrix	MLSLRQSIRFFKPATRTLCSSRYLL–
Peroxisome	–SKL
Peroxisome	–RLXXXXXHL

Of particular importance for the rest of this chapter, is the sequence targeting proteins to the endoplasmic reticulum, and by extension, any proteins destined for the ER, the Golgi apparatus, the cell membrane, vesicles and vesicularly-derived compartments, and secretion out of the cell. Here, in addition to an N-terminal signal sequence, the position of secondary internal signal sequences (sometimes called signal patches) helps to determine the disposition of the protein as it enters the ER.

The initial insertion requires recognition of the signal sequence by *SRP*, the signal recognition protein. The SRP is a G-protein and exchanges its bound GDP for a GTP upon binding to a protein's signal sequence. The SRP with its attached protein then docks to a receptor (called the SRP receptor, astoundingly enough) embedded in the ER membrane and extending into the cytoplasm. The SRP usually binds as soon as the signal sequence is available, and when it does so, it arrests translation until it is docked to the ER membrane. Incidentally, this is the origin of the "rough" endoplasmic reticulum: the ribosomes studding the ER are attached to the ER cytoplasmic surface by the nascent polypeptide it is producing and an SRP. The SRP receptor can exist on its own or in association with a *translocon*, which is a bipartite translocation channel. The SRP receptor (SR) is also a GTPase, and is usually carrying a GDP molecule when unassociated. However, upon association with the translocon it exchanges its GDP for a GTP. These GTPs are important because when the SRP binds to the SR, both GTPase activities are activated and the resulting release of energy dissociates both from the translocon and the nascent polypeptide. This relieves the block on translation imposed by the SRP, and the new protein is pushed on through the translocon as it is being synthesized. Once the signal sequence has completely entered the lumen of the ER, it reveals a recognition site for *signal peptidase*, a hydrolytic enzyme that resides in the ER lumen and whose purpose is to snip off the signal peptide.

Prokaryotes also use an SRP homolog. In E. coli, the SRP is simple, made up of one protein subunit (Ffh) and a small 4.5S RNA. By comparison, some higher eukaryotes have an SRP comprised of six different proteins subunits and a 7S RNA. Similarly, there is a simple prokaryotic homologue to the SRP receptor, FtsY. An interesting difference is that FtsY generally does not interact with exported proteins, and appears to be necessary only for membrane-embedded proteins. Otherwise, there are many similarities in mechanism for SRP-based insertion of membrane proteins in eukaryotic and prokaryotic species, including GTP dependence, and completion of the mechanism by a translocase (SecYEG in E. coli).

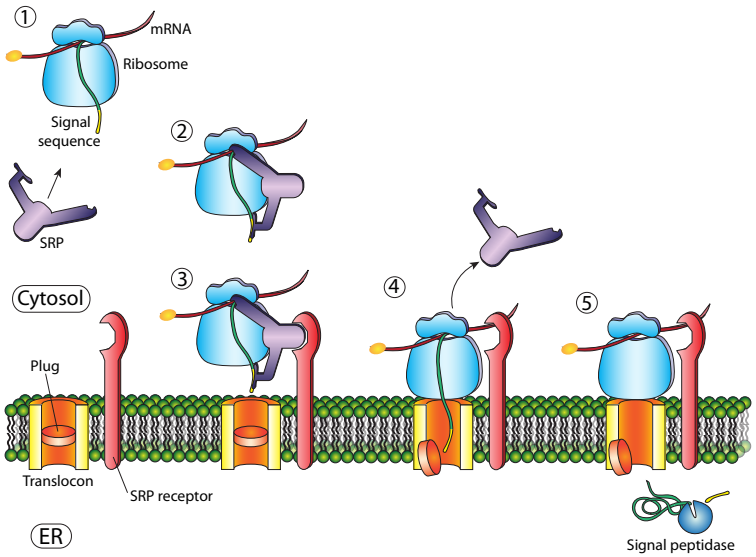


Figure 4. SRP and its receptor SR mediate movement of proteins through the ER membrane. The SRP recognizes the signal sequence and binds to it and the ribosome, temporarily arresting translation. The SRP-polypeptide-ribosome complex is bound by its receptor, SR, which positions the complex on a translocon. Once the ribosome and polypeptide are docked on the translocon, the SRP dissociates, and translation resumes, with the polypeptide moving through the translocon as it is being synthesized.

If that was the only signal sequence in the protein, the remainder of the protein is synthesized and pushed through the translocon and a soluble protein is deposited in the ER lumen, as shown in figure 4. What about proteins that are embedded in a membrane? Transmembrane proteins have internal signal sequences (sometimes called signal patches). Depending on their relative locations, they may be considered either *start*-transfer or *stop*-transfer sequence, where "transfer" refers to translocation of the peptide through the translocon. This is easiest to understand by referring to fig. 5. If

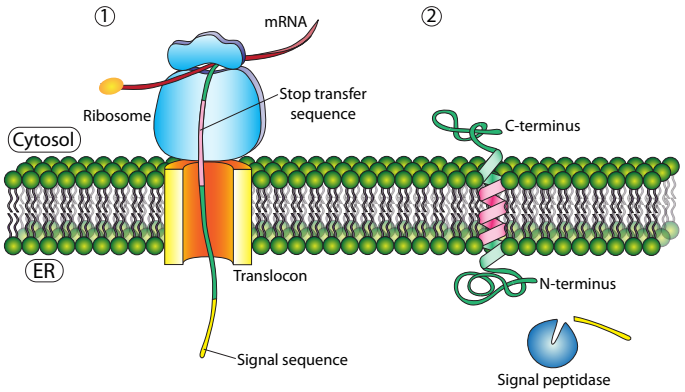


Figure 5. Single-pass transmembrane protein insertion. (1) the signal sequence has allowed the ribosome to dock on a translocon and newly made polypeptide is threaded through until the stop-transfer sequence. (2) The hydrophobic stop transfer sequence gets "stuck" in the membrane, forcing the rest of the polypeptide to stay in the cytoplasm as it is translated.

there is a significant stretch of mostly-uninterrupted hydrophobic residues, it would be considered a stop-transfer signal, as that part of the protein can get stuck in the translocon (and subsequently the ER membrane) forcing the remainder of the protein to remain outside the ER. This would generate a protein that inserts into the membrane once, with its N-terminus in the ER lumen and the C-terminus in the cytoplasm. In a multi-pass transmembrane protein, there could be several start- and stop- transfer hydrophobic signal patches. Building on the single-pass example, if there was another

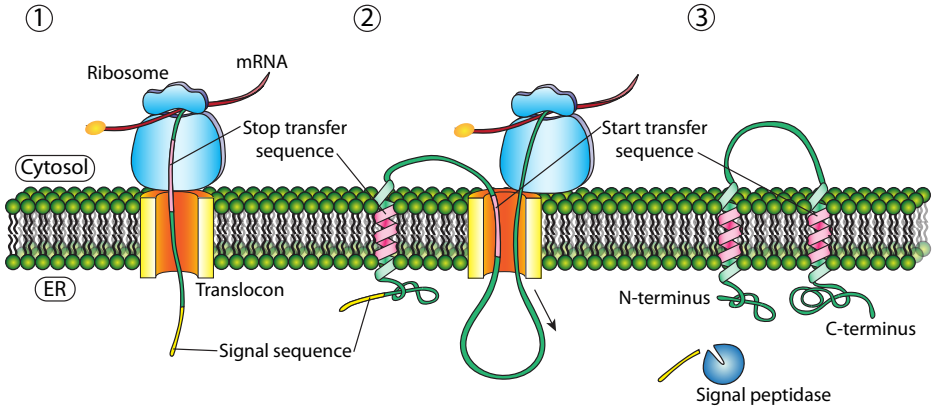


Figure 6. Insertion of 2-pass transmembrane protein.

signal patch after the stop-transfer sequence, it would act as a start-transfer sequence, attaching to a translocon and allowing the remainder of the protein to be moved into the ER. This results in a protein with both N- and C- termini in the ER lumen, passing through the ER membrane twice, and with a cytoplasmic loop sticking out. Of course, the N-terminus could be on the other side. For a cytoplasmic N-terminus, the protein cannot have an N-terminal signal sequence (fig. 7). It has an internal signal patch instead. It plays essentially the same role, but the orientation of the patch means that the N-terminal stays cytoplasmic. The polypeptide translated after the patch is fed into the ER. And just as in the last example, multiple stop- and start-sequences can reinsert the protein in the membrane and change the facing of the next portion.

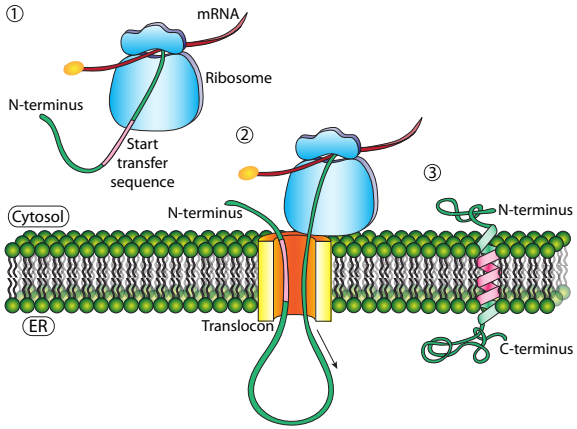


Figure 7. Insertion of a single-pass protein with N-terminus in cytoplasm uses a signal patch but no N-terminal signal.