H1 THE GENETIC CODE

Key Notes

The genetic code is a triplet code

The genetic code is the rules that specify how the nucleotide sequence of an mRNA is translated into the amino acid sequence of a polypeptide. The nucleotide sequence is read as triplets called codons. The codons UAG, UGA and UAA do not specify amino acids and are called termination codons or Stop codons. AUG codes for methionine and also acts as an initiation (Start) codon.

The genetic code is degenerate

Most amino acids in proteins are specified by more than one codon (i.e. the genetic code is degenerate). Codons that specify the same amino acid (synonyms) often differ only in the third base, the wobble position, where base pairing with the anticodon may be less stringent than for the first two positions of the codon.

Universality of the genetic code

The genetic code is not universal but is the same in most organisms. Exceptions are found in mitochondrial genomes where some codons specify different amino acids to that normally encoded by nuclear genes. In mitochondria, the UGA codon does not specify termination of translation but instead encodes for tryptophan. Similarly, in certain protozoa UAA and UAG encode glutamic acid instead of acting as termination codons.

Reading frames

The mRNA sequence can be read by the ribosome in three possible reading frames. Usually only one reading frame codes for a functional protein since the other two reading frames contain multiple termination codons. In some bacteriophage, overlapping genes occur which use different reading frames.

Open reading frames

An open reading frame (ORF) is a run of codons that starts with ATG and ends with a termination codon, TGA, TAA or TAG. Coding regions of genes contain relatively long ORFs unlike noncoding DNA where ORFs are comparatively short. The presence of a long open reading frame in a DNA sequence therefore may indicate the presence of a coding region. Computer analysis of the ORF can be used to deduce the sequence of the encoded protein.

Related topics

RNA structure (G1)
Ribosomal RNA (G8)
Transfer RNA (G9)
Translation in prokaryotes (H2)
Translation in eukaryotes (H3)

The genetic code is a triplet code

During translation, the sequence of an mRNA molecule is read from its 5′ end by ribosomes which then synthesize an appropriate polypeptide. Both in prokaryotes and in eukaryotes, the DNA sequence of a single gene is **colinear** with the amino acid sequence of the polypeptide it encodes. In other words, the

nucleotide sequence of the coding DNA strand, 5′ to 3′, specifies in exactly the same order the amino acid sequence of the encoded polypeptide, N-terminal to C-terminal. The relationship between the nucleotide sequence of the mRNA and the amino acid sequence of the polypeptide is called **the genetic code**. The sequence of the mRNA is read in groups of three nucleotides called **codons**, with each codon specifying a particular amino acid (*Fig. 1*). However, three codons, UAG, UGA and UAA, do not encode an amino acid. Whenever one of these codons is encountered by a ribosome, it leads to termination of protein synthesis. Therefore these three codons are called **termination codons** or **stop codons**. The codon AUG codes for methionine. Although methionine is found at internal positions in polypeptide chains, all eukaryotic polypeptides also start with methionine (see Topic H3) and all prokaryotic polypeptides start with a modified methionine (*N*-formyl methionine; see Topic H2). Therefore the first AUG codon that is read by the ribosome in an mRNA is called the **initiation codon** or **start codon**.

The genetic code is degenerate

Since RNA is composed of four types of nucleotides, there are $4^3 = 64$ possible codons, that is 64 possible triplets of nucleotides with different sequences. However, only 20 amino acids are commonly found in proteins (see Topic B1) so that, in most cases, a single amino acid is coded for by several different codons (see *Fig. 1*). The genetic code is therefore said to be **degenerate**. In fact, only methionine and tryptophan are represented by a single codon. As a result of the genetic code's degeneracy, a mutation that changes only a single nucleotide in

Codon sequence

1st base (5′end)	2nd base U	2nd base C	2nd base A	2nd base G	3rd base (3′end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	**Stop**	**Stop**	A
	Leu	Ser	**Stop**	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Fig. 1. The genetic code.

DNA (**point mutation**), and hence changes only a single nucleotide in the corresponding mRNA, often has no effect on the amino acid sequence of the encoded polypeptide.

Codons that specify the same amino acid are called **synonyms**. Most synonyms differ only in the third base of the codon; for example GUU, GUC, GUA and GUG all code for valine. During protein synthesis, each codon is recognized by a triplet of bases, called an **anticodon**, in a specific tRNA molecule (see Topics G9 and H2). Each base in the codon base pairs with its complementary base in the anticodon. However, the pairing of the third base of a codon is less stringent than for the first two bases (i.e. there is some '**wobble base pairing**') so that in some cases a single tRNA may base pair with more than one codon. For example, phenylalanine tRNA, which has the anticodon GAA, recognizes both of the codons UUU and UUC. The third position of the codon is therefore also called the **wobble position**.

Universality of the genetic code

For many years it was thought that the genetic code is 'universal', namely that all living organisms used the same code. Now we know that the genetic code is almost the same in all organisms but there are a few differences. Mitochondria contain DNA, as double-stranded DNA circles, and the mitochondrial genome codes for about 10–20 proteins. Surprisingly, in **mitochondrial mRNAs**, some codons have different meanings from their counterparts in mRNA in the cytosol. A few examples are given below (N denotes any of the four nucleotides A, G, C or U):

mitochondria	AUA = Met not Ile
mitochondria	UGA = Trp not **Stop**
some animal mitochondria	AGA and AGG = **Stop** not Arg
plant mitochondria	CGG = Trp not Arg
yeast mitochondria	CUN = Thr not Leu

Some unicellular organisms are also now known to use a variant genetic code. For example:

some ciliated protozoa	UAA and UAG = Glu not **Stop**.

Reading frames

Since the sequence of an mRNA molecule is read in groups of three nucleotides (codons) from the 5′ end, it can be read in three possible **reading frames**, depending on which nucleotide is used as the first base of the first codon (*Fig. 2*).

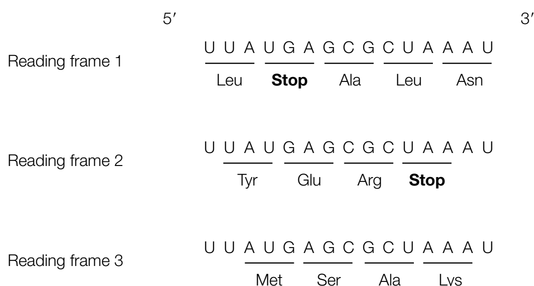


Fig. 2. Three potential reading frames for any given mRNA sequence depending on which nucleotide is 'read' first.

Usually, only one reading frame (reading frame 3 in *Fig. 2*) will produce a functional protein since the other two reading frames will include several termination (**Stop**) codons. The correct reading frame is set *in vivo* by recognition by the ribosome of the initiation codon, AUG, at the start of the coding sequence. Usually one sequence of bases encodes only a single protein. However, in some bacteriophage DNAs, several genes overlap, with each gene being in a different reading frame. This organization of **overlapping genes** generally occurs when the genome size is smaller than can accommodate the genes necessary for phage structure and assembly using only one reading frame.

Open reading frames

In many cases these days, the protein encoded by a particular gene is deduced by cloning (see Section I) and then sequencing the corresponding DNA. The DNA sequence is then scanned using a computer program to identify runs of codons that start with ATG and end with TGA, TAA or TAG. These runs of codons are called **open reading frames** (**ORFs**) and identify potential coding regions. Because genes carry out important cellular functions, the sequence of coding DNA (and of important regulatory sequences) is more strongly conserved in evolution that that of noncoding DNA. In particular, mutations that lead to the creation of termination codons within the coding region, and hence premature termination during translation, are selected against. This means that the coding regions of genes often contain comparatively long ORFs whereas in noncoding DNA, triplets corresponding to termination codons are not selected against and ORFs are comparatively short. Thus, when analyzing the ORFs displayed for a particular cloned DNA, it is usually true that a long ORF is likely to be coding DNA whereas short ORFs may not be. Nevertheless, one must be aware that some exons can be short and so some short ORFs may also be coding DNA. Computer analysis may be able to detect these by screening for the conserved sequences at exon/intron boundaries and the splice branchpoint sequence (see Topic G7). Finally, by referring to the genetic code, computer analysis can predict the protein sequence encoded by each ORF. This is the **deduced protein sequence**.

H2 TRANSLATION IN PROKARYOTES

Key Notes

Overview

During translation the mRNA is read in a 5′ to 3′ direction and protein is made in an N-terminal to C-terminal direction. Translation relies upon aminoacyl-tRNAs that carry specific amino acids and recognize the corresponding codons in mRNA by anticodon–codon base pairing. Translation takes place in three phases; initiation, elongation and termination.

Synthesis of aminoacyl-tRNA

Each tRNA molecule has a cloverleaf secondary structure consisting of three stem loops, one of which bears the anticodon at its end. The amino acid is covalently bound to the 3′ OH group at the 3′ end by aminoacyl synthetase to form aminoacyl-tRNA. The reaction, called amino acid activation, occurs in two steps and requires ATP to form an intermediate, aminoacyl-adenylate.

Initiation of protein synthesis

Each ribosome has three binding sites for tRNAs; an A site where the incoming aminoacyl-tRNA binds, a P site where the tRNA linked to the growing polypeptide chain is bound, and an E site which binds tRNA prior to its release from the ribosome. Translation in prokaryotes begins by the formation of a 30S initiation complex between the 30S ribosomal subunit, mRNA, initiation factors and fMet $tRNA_f^{Met}$. The 30S subunit binds to the Shine–Dalgarno sequence which lies 5′ to the AUG Start codon and is complementary to the 16S rRNA of the small ribosomal subunit. The ribosome then moves in a 3′ direction along the mRNA until it encounters the AUG codon. The 50S ribosomal subunit now binds to the 30S initiation complex to form the 70S initiation complex. In this complex, the anticodon of the fMet $tRNA_f^{Met}$ is base paired to the AUG initiation codon (start codon) in the P site.

Elongation

The elongation cycle consists of three steps: aminoacyl-tRNA binding, peptide bond formation, and translocation. In the first step, the aminoacyl-tRNA corresponding to the second codon binds to the A site on the ribosome as an aminoacyl-tRNA/EF-Tu/GTP complex. After binding, the GTP is hydrolyzed and EF-Tu/GDP is released. The EF-Tu is regenerated via the EF-Tu–EF-Ts exchange cycle. Peptide bond formation is catalyzed by peptidyl transferase between the C-terminus of the amino acyl moiety in the P site and the amino group of the aminoacyl-tRNA in the A site. In the final (translocation) step, EF-G/GTP binds to the ribosome, the deacylated tRNA moves from the P site to the E site, the dipeptidyl-tRNA in the A site moves to the P site, and the ribosome moves along the mRNA to place the next codon in the A site. The GTP is hydrolyzed to GDP and inorganic phosphate. When the next aminoacyl-tRNA binds to the A site in the next round of elongation, the deacylated tRNA is released from the E site.

Termination

The appearance of a UAA or UAG termination (stop) codon in the A site causes release factor RF1 to bind whereas RF2 recognizes UGA. RF3 assists RF1 and RF2. The release factors trigger peptidyl transferase to transfer the polypeptide to a water molecule instead of to aminoacyl-tRNA. The polypeptide, mRNA, and free tRNA leave the ribosome and the ribosome dissociates into its subunits ready to begin a new round of translation.

Related topics

RNA structure (G1)
Transcription in prokaryotes (G2)
Operons (G3)
Transcription in eukaryotes: an overview (G4)
Transcription of protein-coding genes in eukaryotes (G5)
Regulation of transcription by RNA Pol II (G6)
Processing of eukaryotic pre-mRNA (G7)
Ribosomal RNA (G8)
Transfer RNA (G9)
The genetic code (H1)
Translation in eukaryotes (H3)

Overview

Ribosomes exist normally as separate subunits that come together to form a ribosome when they bind to an mRNA near its 5′ end. The ribosome reads the nucleotide sequence from the 5′ to 3′ direction, synthesizing the corresponding protein from amino acids in an N-terminal (amino-terminal) to C-terminal (carboxyl terminal) direction. The amino acids used are covalently bound to tRNA (transfer RNA) molecules to form aminoacyl-tRNAs. Each aminoacyl-tRNA bears a triplet of bases, called an **anticodon**. The ribosome reads each triplet codon of the mRNA in turn and an aminoacyl-tRNA molecule with an anticodon that is complementary to the codon binds to it via hydrogen bonding. A peptide bond is then formed between the incoming amino acid and the growing end of the polypeptide chain.

Overall, protein synthesis (or **translation**) takes place in three stages; **initiation, elongation** and **termination**. During initiation, the mRNA–ribosome complex is formed and the first codon (always AUG) binds the first aminoacyl-tRNA (called **initiator tRNA**). During the elongation phase, the other codons are read sequentially and the polypeptide grows by addition of amino acids to its C-terminal end. This process continues until a termination codon (Stop codon), which does not have a corresponding aminoacyl-tRNA with which to base pair, is reached. At this point, protein synthesis ceases (termination phase) and the finished polypeptide is released from the ribosome. Usually at any one time, many ribosomes are translating an mRNA simultaneously, forming a structure called a **polyribosome** or **polysome**.

Synthesis of aminoacyl-tRNA

Each tRNA molecule has a **cloverleaf secondary structure** with the anticodon accessible at the end of the anticodon stem loop (see *Fig. 1* and Topic G9). During synthesis of the aminoacyl-tRNA, the amino acid is covalently bound to the A residue of the CCA sequence at the 3′ end (*Fig. 1*). Each tRNA molecule carries only a single amino acid. However, because of the redundancy of the genetic code (see Topic H1), several codons may encode the same amino acid and so there will also exist several types of tRNA with corresponding anticodons all bearing the same amino acid. The correct nomenclature is, for example, $tRNA^{Gly}$ for the tRNA that will accept glycine whereas the corre-

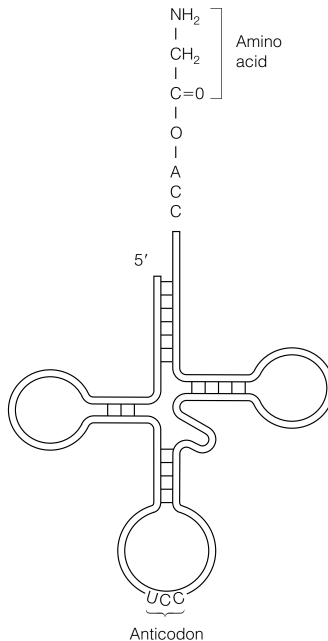
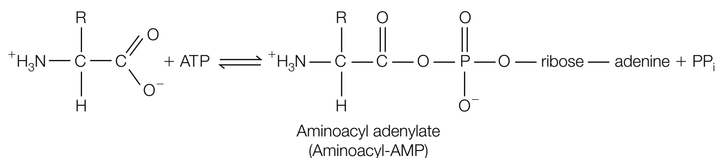


Fig. 1. Structure of an aminoacyl-tRNA.

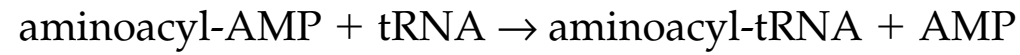
sponding aminoacyl-tRNA is Gly-tRNA^{Gly}, and is the aminoacyl-tRNA shown in *Fig. 1*.

Synthesis of aminoacyl-tRNAs is crucially important for two reasons. First each amino acid must be covalently linked to a tRNA molecule in order to take part in protein synthesis, which depends upon the 'adaptor' function of tRNA to ensure that the correct amino acids are incorporated. Second, the covalent bond that is formed between the amino acid and the tRNA is a high energy bond that enables the amino acid to react with the end of the growing polypeptide chain to form a new peptide bond. For this reason, the synthesis of aminoacyl-tRNA is also referred to as **amino acid activation**. Amino acids that are not linked to tRNAs cannot be added to the growing polypeptide.

The attachment of an amino acid to a tRNA is catalyzed by an enzyme called **aminoacyl-tRNA synthetase**. A separate aminoacyl-tRNA synthetase exists for every amino acid, making 20 synthetases in total. The synthesis reaction occurs in two steps. The first step is the reaction of an amino acid and ATP to form an **aminoacyl-adenylate** (also known as **aminoacyl-AMP**):

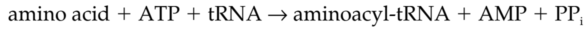


In the second step, without leaving the enzyme, the aminoacyl group of aminoacyl-AMP is transferred to the 3′ end of the tRNA molecule to form aminoacyl-tRNA:



$$\text{aminoacyl-AMP} + \text{tRNA} \rightarrow \text{aminoacyl-tRNA} + \text{AMP}$$

The overall reaction is:



$$\text{amino acid} + \text{ATP} + \text{tRNA} \rightarrow \text{aminoacyl-tRNA} + \text{AMP} + \text{PP}_i$$

and is driven by the subsequent hydrolysis of the pyrophosphate to inorganic phosphate.

Initiation of protein synthesis

Each prokaryotic ribosome, shown schematically in *Fig. 2* (see Topic G8 for details of ribosome structure), has three binding sites for tRNAs. The **aminoacyl-tRNA binding site** (or **A site**) is where, during elongation, the incoming aminoacyl-tRNA binds. The **peptidyl-tRNA binding site** (or **P site**) is where the tRNA linked to the growing polypeptide chain is bound. The **exit site** (or **E site**) is a binding site for tRNA following its role in translation and prior to its release from the ribosome. All three sites (A, P and E) are formed by the rRNA molecules in the ribosome (see Topic G8).

The first codon translated in all mRNAs is AUG which codes for methionine. This AUG is called the **start codon** or **initiation codon**. Naturally, other AUG codons also occur internally in an mRNA where they encode methionine residues internal to the protein. Two different tRNAs are used for these two types of AUG codon; $\text{tRNA}_f^{\text{Met}}$ is used for the initiation codon and is called the **initiator tRNA** whereas $\text{tRNA}_m^{\text{Met}}$ is used for internal AUG codons. In prokaryotes the first amino acid of a new protein is ***N*-formylmethionine** (abbreviated **fMet**). Hence the aminoacyl-tRNA used in initiation is **fMet-$\text{tRNA}_f^{\text{Met}}$** also abbreviated as fMet-tRNA_f. It is essential that the correct AUG is used as the initiation codon since this sets the correct reading frame for translation (see Topic H1). A short sequence rich in purines (5′-AGGAGGU-3′), called the **Shine–Dalgarno sequence**, lies 5′ to the AUG initiation codon (*Fig. 3*) and is complementary to part of the 16S rRNA in the small ribosomal subunit. Therefore this is the binding site for the 30S ribosomal subunit which then migrates in a 3′ direction along the mRNA until it encounters the AUG initiation codon. Thus the Shine–Dalgarno sequence delivers the ribosomal subunit to the correct AUG for initiation for translation.

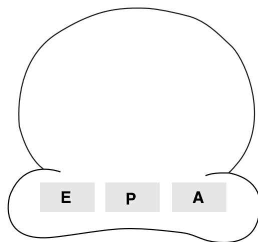


Fig. 2. Schematic of a prokaryotic 70S ribosome showing the peptidyl-tRNA site (P site), aminoacyl-tRNA site (A site) and exit site (E site).

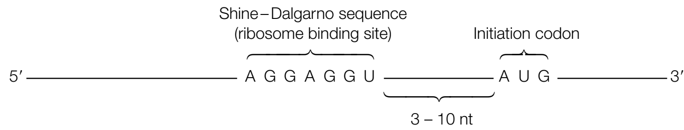


Fig. 3. The Shine–Dalgarno sequence in prokaryotic mRNA.

Initiation of protein synthesis requires proteins called **initiation factors (IFs)**. In prokaryotes, three initiation factors (IF-1, IF-2 and IF-3) are essential. Because of the complexity of the process, the exact order of binding of IF-1, IF-2, IF-3, fMet-tRNA$_f^{Met}$ and mRNA is still unclear. One current model is shown in *Fig. 4* and is described below.

- Initiation begins with the binding of IF-1 and IF-3 to the small (30S) ribosomal subunit. Their role is to stop the 30S subunit binding to the 50S subunit in the absence of mRNA and fMet-tRNA$_f^{Met}$ which would result in a nonfunctional ribosome.

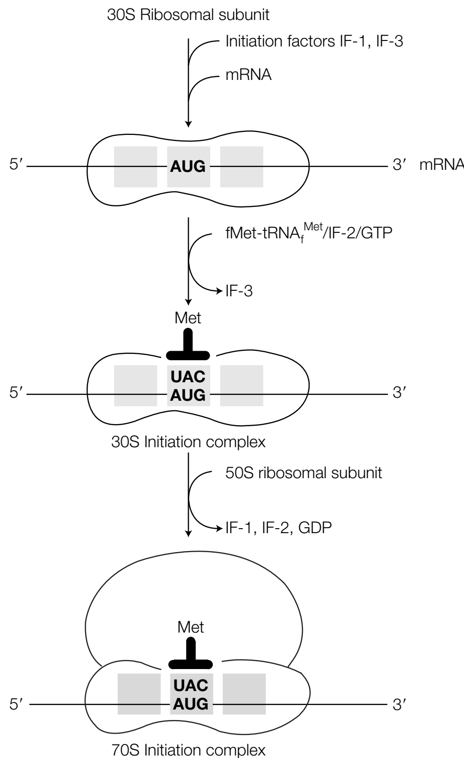


Fig. 4. Initiation of protein synthesis in prokaryotic cells.

- The small subunit then binds to the mRNA via the Shine–Dalgarno sequence and moves 3′ along the mRNA until it locates the AUG initiation codon.
- The initiator tRNA charged with *N*-formylmethionine and in a complex with IF-2 and GTP (fMet-$tRNA_f^{Met}$/IF-2/GTP) now binds.
- IF-3 is released.
- The complex of mRNA, fMet-$tRNA_f^{Met}$, IF-1, IF-2 and the 30S ribosomal subunit is called the **30S initiation complex**.
- The large (50S) ribosomal subunit now binds, with the release of IF-1 and IF-2 and hydrolysis of GTP, to form a **70S initiation complex**.

One important point to note is that, unlike all other aminoacyl-tRNA molecules (which bind to the A site; see below), the binding of fMet-$tRNA_f^{Met}$ occurs directly into the P site.

Elongation

At the start of the first round of elongation (*Fig. 5*), the initiation codon (AUG) is positioned in the P site with fMet-$tRNA_f^{Met}$ bound to it via codon–anticodon base pairing. The next codon in the mRNA is positioned in the A site. Elongation of the polypeptide chain occurs in three steps called the **elongation cycle,** namely **aminoacyl-tRNA binding, peptide bond formation** and **translocation**:

- *Aminoacyl-tRNA binding*: in this first step, the corresponding aminoacyl-tRNA for the second codon binds to the A site via codon–anticodon interaction (*Fig. 5*). Binding of the aminoacyl-tRNA requires **elongation factor EF-Tu** and GTP which bind as an aminoacyl-tRNA/EF-Tu/GTP complex. Following binding, the GTP is hydrolyzed and the EF-Tu is released, now bound to GDP (*Fig. 5*). Before the EF-Tu molecule can catalyze the binding of another charged tRNA to the ribosome, it must be regenerated by a process involving another elongation factor, **EF-Ts**. This regeneration is called the **EF-Tu–EF-Ts exchange cycle** (*Fig. 6*). First, EF-Ts binds to EF-Tu and displaces the GDP. Then GTP binds to the EF-Tu and displaces EF-Ts. The EF-Tu-GTP is now ready to take part in another round of elongation.
- *Peptide bond formation*: the second step, peptide bond formation, is catalyzed by **peptidyl transferase**. In this reaction the carboxyl end of the amino acid bound to the tRNA in the P site is uncoupled from the tRNA and becomes joined by a peptide bond to the amino group of the amino acid linked to the tRNA in the A site (*Fig. 5*). A protein with peptidyl transferase activity has never been isolated. The reason is now clear; in *E. coli* at least, the peptidyl transferase activity is associated with part of the 23S rRNA in the large ribosomal subunit. In other words, peptidyl transferase is a **ribozyme**, a catalytic activity that resides in an RNA molecule (see also Topic G8).
- *Translocation*: in the third step, a complex of **elongation factor EF-G** (also called **translocase**) and GTP (i.e. EF-G/GTP) binds to the ribosome. Three concerted movements now occur, collectively called translocation; the deacylated tRNA moves from the P site to the E site, the dipeptidyl-tRNA in the A site moves to the P site, and the ribosome moves along the mRNA (5′ to 3′) by three nucleotides to place the next codon in the A site. During the translocation events, GTP is hydrolyzed to GDP and inorganic phosphate, and EF-G is released ready to bind more GTP for another round of elongation.

After translocation, the A site is empty and ready to receive the next aminoacyl-tRNA. The A site and the E site cannot be occupied simultaneously. Thus the

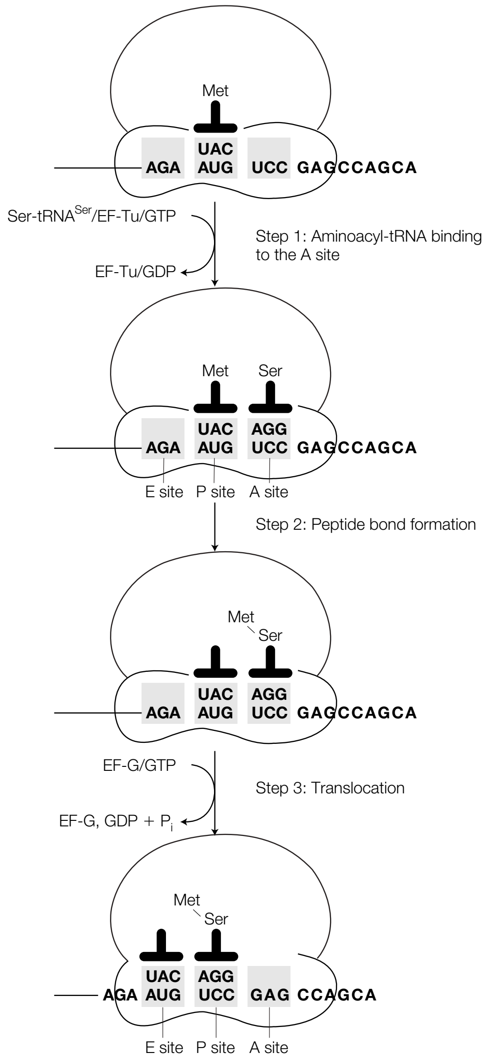


Fig. 5. The elongation phase of protein synthesis in prokaryotes.

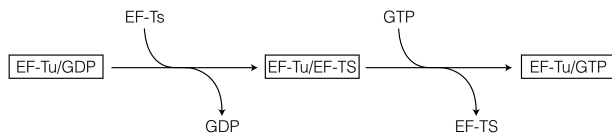


Fig. 6. The EF-Tu–EF-Ts exchange cycle.

deacylated tRNA is released from the E site before the next aminoacyl-tRNA binds to the A site to start a new round of elongation. Elongation continues, adding one amino acid to the C-terminal end of the growing polypeptide for each codon that is read, with the peptidyl-tRNA moving back and forth from the P site to the A site as it grows.

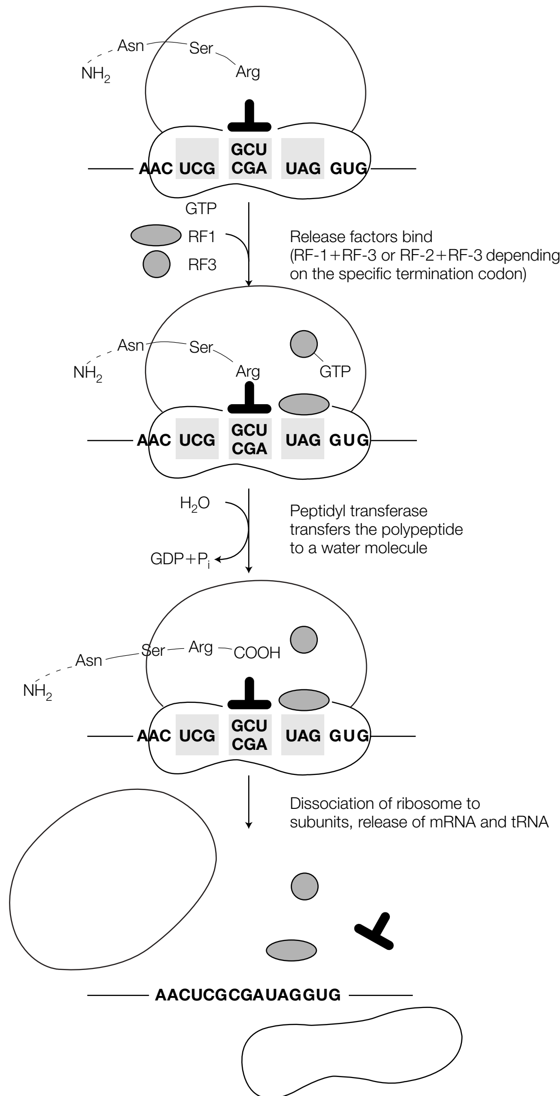


Fig. 7. Termination of protein synthesis in prokaryotic cells.

Termination

Eventually, one of three **termination codons** (also called **Stop codons**) becomes positioned in the A site (*Fig. 7*). These are UAG, UAA and UGA. Unlike other codons, prokaryotic cells do not contain aminoacyl-tRNAs complementary to Stop codons. Instead, one of two **release factors** (**RF-1** and **RF-2**) binds instead. RF-1 recognizes UAA and UAG whereas RF-2 recognizes UAA and UGA. A third release factor, **RF-3**, is also needed to assist RF-1 or RF-2 interaction with the ribosome. Thus either RF-1 + RF-3 or RF-2 + RF-3 bind depending on the exact termination codon in the A site. RF-1 (or RF-2) binds at or near the A site whereas RF-3/GTP binds elsewhere on the ribosome. The release factors cause the peptidyl transferase activity to transfer the polypeptide to a water molecule instead of to aminoacyl-tRNA, effectively cleaving the bond between the polypeptide and tRNA in the P site. To understand this, it is important to realize that the ribosome normally excludes water from the reaction center since this could otherwise hydrolyze the peptidyl-tRNA ester bond and cause premature release of the polypeptide. The release factors appear to work by carrying a water molecule into the peptidyl transferase reaction center so that the hydrolysis now occurs. The free polypeptide now leaves the ribosome, followed by the mRNA and free tRNA, and the ribosome dissociates into 30S and 50S subunits ready to start translation again.

H3 Translation in eukaryotes

Key Notes

Initiation

Eukaryotic ribosomes are larger (80S) and more complex than prokaryotic ribosomes (70S). Initiation is basically similar in prokaryotes and eukaryotes except that in eukaryotes at least nine initiation factors are involved (cf. three factors in prokaryotes), the initiating amino acid is methionine (cf. *N*-formylmethionine in prokaryotes), eukaryotic mRNAs do not contain Shine–Dalgarno sequences (so the AUG initiation codon is detected by the ribosome scanning instead), and eukaryotic mRNA is monocistronic (cf. some polycistronic mRNAs in prokaryotes). Initiation in eukaryotes involves the formation of a 48S preinitiation complex between the 40S ribosomal subunit, mRNA, initiation factors and Met-$tRNA_i^{met}$. The ribosome then scans the mRNA to locate the AUG initiation codon. The 60S ribosomal subunit now binds to form the 80S initation complex.

Elongation

Elongation in eukaryotes requires three eukaryotic initiation factors that have similar functions to the corresponding prokaryotic proteins.

Termination

A single eukaryotic release factor recognizes all three termination codons and requires ATP for activity.

Related topics

RNA structure (G1)
Transcription in prokaryotes (G2)
Operons (G3)
Transcription in eukaryotes: an overview (G4)
Transcription of protein-coding genes in eukaryotes (G5)
Regulation of transcription by RNA Pol II (G6)
Processing of eukaryotic pre-mRNA (G7)
Ribosomal RNA (G8)
Transfer RNA (G9)
The genetic code (H1)
Translation in prokaryotes (H2)

Initiation

The overall mechanism of protein synthesis in eukaryotes is basically the same as in prokaryotes, with three phases defined as initiation, elongation and termination. However, there are some significant differences, particularly during initiation.

- Whereas a prokaryotic ribosome has a sedimentation coefficient (see Topic G8) of 70S and subunits of 30S and 50S, a eukaryotic ribosome has a sedimentation coefficient of 80S with subunits of 40S and 60S (see Topic G8). The composition of eukaryotic ribosomal subunits is also more complex than prokaryotic subunits (see Topic G8) but the function of each subunit is essentially the same as in prokaryotes.

- In eukaryotes, each mRNA is **monocistronic**, that is, discounting any subsequent post-translational cleavage reactions that may occur, the mRNA encodes a single protein. In prokaryotes, many mRNAs are **polycistronic**, that is they encode several proteins. Each coding sequence in a prokaryotic mRNA has its own initiation and termination codons.
- Initiation of protein synthesis in eukaryotes requires at least nine distinct eukaryotic **initiation factors** (**eIFs**) (see *Table 1*) compared with the three initiation factors (IFs) in prokaryotes (see Topic H2).
- In eukaryotes, the initiating amino acid is **methionine**, not *N*-formylmethionine as in prokaryotes.
- As in prokaryotes, a special initiator tRNA is required for initiation and is distinct from the tRNA that recognizes and binds to codons for methionine at internal positions in the mRNA. When charged with methionine ready to begin initiation, this is known as **Met-tRNA$_i^{met}$**.
- The main difference between initiation of translation in prokaryotes and eukaryotes is that in bacteria, a Shine–Dalgarno sequence (see Topic H2) lies 5′ to the AUG initiation codon and is the binding site for the 30S ribosomal subunit, marking this AUG as the one to use for initiation rather than any other AUG internal in the mRNA. The initiation complex is assembled directly over this initiation codon. In contrast, most eukaryotic mRNAs do not contain Shine–Dalgarno sequences. Instead, a 40S ribosomal subunit attaches at the 5′ end of the mRNA and moves downstream (i.e. in a 5′ to 3′ direction) until it finds the AUG initiation codon. This process is called **scanning**.

The full details of initiation in eukaryotes are still not fully known but the process occurs broadly as follows:

- the first step is the formation of a **pre-initiation complex** consisting of the 40S small ribosomal subunit, Met-tRNA$_i^{met}$, eIF-2 and GTP;
- the pre-initiation complex now binds to the 5′ end of the eukaryotic mRNA, a step that requires **eIF-4F** (also called **cap binding complex**) and **eIF-3**. The eIF-4F complex consists of eIF-4A, eIF-4E, and eIF-4G; eIF-4E binds to the 5′ cap on the mRNA (an essential step) whilst eIF-4G interacts with the poly(A)

Table 1. Comparison of protein synthesis factors in prokaryotes and eukaryotes

Prokaryotic	Eukaryotic	Function
Initiation factors		
IF-1, IF-2, IF-3	At least 9 initiation factors, e.g. eIF-1, eIF-2, eIF-2B, eIF-3, eIF-4	Individual factors have functions that differ between prokaryotes and eukaryotes (see the text)
Elongation factors		
EF-Tu	eEF-1A	Aminoacyl tRNA delivery to ribosome
EF-Ts	eEF-1B	Recycling of EF-Tu or eEF-1A
EF-G	eEF-2	Translocation
Termination factors		
RF-1, RF-2, RF-3	eRF-1, eRF-3	Polypeptide chain release

binding protein on the poly(A) tail (which of course implies that the mRNA bends back on itself to allow this interaction to occur). Thus this complex checks that both the 5′ and 3′ ends of the mRNA are intact. The eIF-4A is an ATP-dependent RNA helicase that unwinds any secondary structures in the mRNA, preparing it for translation. This is in contrast to prokaryotic translation where no helicase is needed, presumably because protein synthesis in bacteria can start even as the mRNA is still being synthesized whereas in eukaryotes, transcription in the nucleus and translation in the cytoplasm are separate events which allows time for mRNA secondary structure to form.

- The complex now moves along the mRNA in a 5′ to 3′ direction until it locates the AUG initiation codon (i.e. scanning). The 5′ untranslated regions of eukaryotic mRNAs vary in length but can be several hundred nucleotides long and may contain secondary structures such as hairpin loops. These secondary structures are probably removed by initiation factors of the scanning complex. The initiation codon is usually recognizable because it is often (but not always) contained in a short sequence called the **Kozak consensus** (5′-ACC**AUG**G-3′).
- Once the complex is positioned over the initiation codon, the 60S large ribosomal subunit binds to form an **80S initiation complex**, a step that requires the hydrolysis of GTP and leads to the release of several initiation factors.

Elongation

- The elongation stage of translation in eukaryotes requires three elongation factors, **eEF-1A**, **eEF-IB** and **eEF-2**, which have similar functions to their prokaryotic counterparts EF-Tu, EF-Ts and EF-G (see *Table 1*).
- Although most codons encode the same amino acids in both prokaryotes and eukaryotes, the mRNAs synthesized within the organelles of some eukaryotes use a variant of the genetic code (see Topic H1).
- During elongation in bacteria, the deacylated tRNA in the P site moves to the E site prior to leaving the ribosome (see Topic H2). In contrast, although the situation is still not completely clear, in eukaryotes the deacylated tRNA appears to be ejected directly from the ribosome.

Termination

In eukaryotes, eukaryotic release factor eRF-1 recognizes all three termination codons (UAA, UAG and UGA) and, with the help of protein eRF-3, terminates translation (*Table 1*).

H4 PROTEIN TARGETING

Key Notes

Overview

Both in prokaryotes and eukaryotes, newly synthesized proteins must be delivered to a specific subcellular location or exported from the cell for correct activity. This phenomenon is called protein targeting.

Secretory proteins

Secretory proteins have an N-terminal signal peptide which targets the protein to be synthesized on the rough endoplasmic reticulum (RER). During synthesis it is translocated through the RER membrane into the lumen. Vesicles then bud off from the RER and carry the protein to the Golgi complex, where it becomes glycosylated. Other vesicles then carry it to the plasma membrane. Fusion of these transport vesicles with the plasma membrane then releases the protein to the cell exterior.

Plasma membrane proteins

Plasma membrane proteins are also synthesized on the RER but become inserted into the RER membrane (and hence ultimately the plasma membrane) rather than being released into the RER lumen. The plasma membrane protein may pass once through the plasma membrane (Type I and Type II integral membrane proteins) or may loop back and forth, passing through many times (Type III integral membrane protein). The orientation of the protein in the membrane is determined by topogenic sequences within the polypeptide chain. Type I proteins have a cleaved N-terminal signal sequence and a hydrophobic stop-transfer sequence, Type II have an uncleaved N-terminal signal sequence that doubles as the membrane-anchoring sequence, and Type III have multiple signal sequences and stop-transfer sequences. Proteins destined to be anchored in the membrane by a glycosyl-phosphatidylinositol (GPI) structure have both a cleaved N-terminal signal sequence and a C-terminal hydrophobic sequence that directs addition of the preformed GPI anchor.

Proteins of the endoplasmic reticulum

Proteins destined for the RER have an N-terminal signal peptide, are synthesized on the RER, are translocated into the RER lumen or inserted into the RER membrane. C-terminal amino acid sequences (KDEL in soluble RER lumen proteins, KKXX in type I integral membrane proteins) are recognized by specific receptor proteins and retain the proteins in the ER.

Lysosomal proteins

Lysosomal proteins are targeted to the lysosomes via the addition of a mannose 6-phosphate signal that is added in the *cis*-compartment of the Golgi and is recognized by a receptor protein in the *trans*-compartment of the Golgi. The protein is then transported by specialized vesicles to a late endosome that later matures into a lysosome. The mannose 6-phosphate receptor recycles back to the Golgi for re-use.

Mitochondrial and chloroplast proteins

Most mitochondria and chloroplast proteins are made on free cytosolic ribosomes, released into the cytosol and then taken up into the organelle. Uptake into the mitochondrial matrix requires a matrix-targeting sequence and occurs at sites where the outer and inner mitochondrial membranes come into contact. The process is mediated by hsp70 and hsp60 proteins and requires both ATP hydrolysis and an electrochemical gradient across the inner mitochondrial membrane. Targeting of proteins to other compartments of mitochondria or chloroplasts requires two signals.

Nuclear proteins

Proteins destined for import into the nucleus typically require a nuclear localization signal, four to eight amino acids long, located internally in the protein. Uptake occurs via nuclear pores and requires ATP hydrolysis.

Related topics

Prokaryote cell structure (A1)
Eukaryote cell structure (A2)
Protein structure (B3)
Membrane lipids (E1)
Membrane proteins and carbohydrate (E2)
Transport of macromolecules (E4)
Protein glycosylation (H5)
Electron transport and oxidative phosphorylation (L2)
Photosynthesis (L3)

Overview

Cells must ensure that each newly synthesized protein is sorted to its correct location where it can carry out the appropriate function. This process is called **protein targeting**. In a eukaryotic cell, the protein may be destined to stay in the cytosol, for example an enzyme involved in glycolysis (see Topic J3). Alternatively it may need to be targeted to an organelle (such as a mitochondrion, lysosome, peroxisome, chloroplast or the nucleus) or be inserted into the plasma membrane or exported out of the cell. In bacteria such as *E. coli*, the protein may stay in the cytosol, be inserted into the plasma membrane or the outer membrane, be sent to the space between these two membranes (the periplasmic space) or be exported from the cell. In both prokaryotes and eukaryotes, if a protein is destined for the cytosol, it is made on free ribosomes in the cytosol and released directly into the cytosol. If it is destined for other final locations, specific protein-targeting mechanisms are involved.

Secretory proteins

Proteins destined to be secreted from the eukaryotic cell are synthesized by ribosomes bound to the **rough endoplasmic reticulum (RER)**. As the protein is synthesized, it is translocated across the RER membrane into the lumen of the RER where it folds into its final conformation. The ER then buds off vesicles that carry the protein to the **Golgi apparatus** (see Topic A2) also called the **Golgi complex** (*Fig. 1*). The Golgi has a ***cis*** **face** (where vesicles enter) and a ***trans*** **face** (where vesicles leave). Thus the RER vesicles fuse with the ***cis*** **compartment** of the Golgi, releasing the protein into the Golgi lumen. The protein then moves through the Golgi complex to the ***trans*** **compartment**, being modified *en route* by the addition of carbohydrate residues (glycosylation, see Topic H5). Finally, vesicles bud from the *trans* compartment and carry the glycosylated secretory proteins to the plasma membrane where the vesicles fuse, releasing their contents to the cell exterior. This fusion and extracellular release of protein is also called **exocytosis** (see Topic E4).

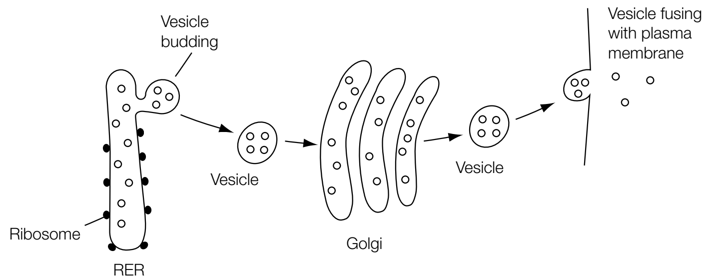


Fig. 1. *Synthesis and exocytosis of secretory proteins; see text for details. The ribosomes attached to the RER are shown as filled-in circles whereas the open circles in the lumen of the RER, vesicles and Golgi complex represent secretory protein molecules.*

Signal hypothesis

A typical secretory protein differs from a cytosolic protein by having a sequence about 13–35 amino acids long at its N-terminal end called a **signal sequence** or **signal peptide**. The signal peptides of different secretory proteins differ in amino acid sequence but there are some common features, for example the center of the sequence usually consists of 10–15 hydrophobic amino acids. The **signal hypothesis** was proposed from early work in this area and predicted that the signal peptide directs the secretory protein to the RER membrane and so targets the protein to cross into the RER lumen and be exported. The signal hypothesis has been shown to apply to protein secretion in animal, plant and bacterial cells. A simplified version of the mechanism is shown in *Fig. 2*.

The mRNA for the secretory protein binds to a free cytoplasmic ribosome and protein synthesis begins. The first part of the protein made is the N-terminal signal peptide. A **signal recognition particle (SRP)**, which is a complex of a 7S RNA and six proteins, binds to the signal peptide and stops further protein synthesis. This stops the secretory protein from being released prematurely into the cytosol. The ribosome–mRNA–SRP complex now binds to an **SRP receptor**, a protein on the surface of the RER. The RER membrane also contains a **ribosome receptor** protein associated with a **protein translocon**. In a concerted series of reactions, the ribosome is held tightly by the ribosome receptor protein, the SRP binds to the SRP receptor and is released from the signal peptide, and translation now continues once more, the nascent polypeptide passing through a pore in the membrane created by the translocon. As it passes through the pore, the signal peptide is cleaved off by a **signal peptidase** on the lumenal face of the RER (*Fig. 2*) and degraded, releasing the rest of the protein into the lumen. The protein is then transported through the Golgi to the cell exterior as described above. Since transport across the RER membrane occurs during protein synthesis, the process is said to be **co-translational**. The released SRP is cycled via its receptor ready for binding to another signal peptide (the **SRP cycle**).

Plasma membrane proteins

Integral plasma membrane proteins are also synthesized by ribosomes on the RER, but become inserted in the RER membrane rather than transported into the lumen. During transport to the Golgi and then to the cell surface, these proteins stay anchored in the membrane, the final vesicles which fuse with the plasma membrane then becoming new plasma membrane (*Fig. 3*). Note that, after inser-

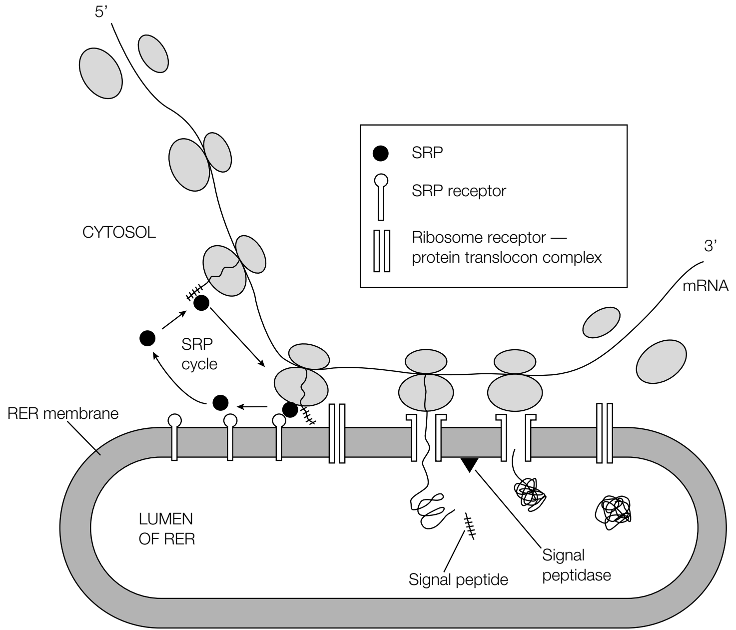


Fig. 2. A simplified version of the signal hypothesis (see the text for details).

tion in the RER membrane, one part of the protein faces in towards the RER lumen but eventually this faces outward on the cell surface. It is this part of the protein that receives the carbohydrate during glycosylation in the RER and Golgi complex so that the carbohydrate is exposed on the cell surface.

Transfer of the plasma membrane protein across the ER membrane occurs during synthesis by a mechanism similar to that for secretory proteins. However, by definition, the protein is destined to remain anchored in the membrane and not enter the RER lumen entirely. There are several ways in which this is achieved, depending on the type of membrane protein. Some integral

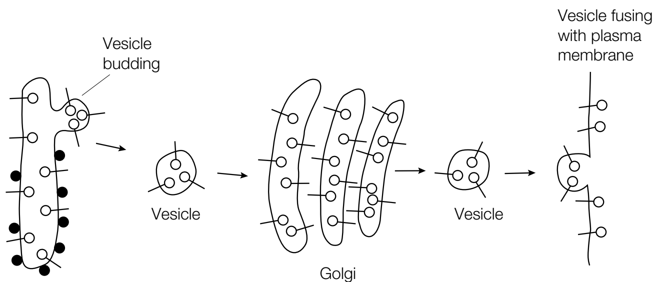


Fig. 3. Synthesis of plasma membrane proteins; see text for details. The ribosomes attached to the RER are shown as filled-in circles whereas the newly synthesized plasma membrane proteins are shown as open circles on stalks.

membrane proteins are **single membrane-spanning proteins**, that is the polypeptide chain crosses the membrane only once, whereas in other cases the protein is a **multiple membrane-spanning protein** (see Topic E2). The orientation of the protein in the membrane and the number of times it spans the lipid bilayer depend on specific **topogenic sequences** within the polypeptide chain. These topogenic sequences are regions of predominantly hydrophobic amino acids, and fall into three types: **N-terminal signal sequences**, **internal signal sequences** and **stop-transfer sequences**.

In the single membrane-spanning **Type I integral membrane proteins** (*Fig. 4a*), in addition to the N-terminal signal sequence which is cleaved from the protein by signal peptidase as in secretory proteins, there is a second hydrophobic sequence located internally in the protein. Thus the protein starts to cross the RER membrane during synthesis just like a secretory protein, but then transfer is stopped before the entire protein is translocated and the protein stays inserted in the membrane through the interaction of the hydrophobic stop-transfer sequence with the hydrophobic interior of the bilayer. In the single membrane-spanning **Type II membrane proteins** (*Fig. 4b*), there is just an N-terminal signal sequence as found in secretory proteins. However, in this case the signal sequence is not cleaved from the membrane protein by signal peptidase and doubles as the membrane anchor. Multiple membrane-spanning **Type III integral membrane proteins** (*Fig. 4c*), which cross the membrane several times, have multiple internal signal peptides and stop-transfer sequences to organize this arrangement during synthesis. The final orientation of the N terminus and the C terminus depends on whether the N-terminal signal sequence is cleaved and whether the final topogenic sequence is an internal signal sequence or a stop-transfer sequence, respectively. Some proteins lack an N-terminal signal sequence and have just an internal signal sequence.

Those proteins that are anchored in the membrane through a covalently attached **glycosyl-phosphatidylinositol (GPI)** structure at the C terminus (*Fig 4d*; also see Topic E2) possess both an N-terminal signal sequence to direct them to the RER membrane and a second hydophobic sequence at the very C terminus. The N-terminal signal sequence is cleaved off by signal peptidase, while the C-terminal sequence directs the addition of a preformed GPI structure to an internal amino acid residue near the C terminus. The GPI structure is built up by sequential addition of sugars (glucosamine and mannose) and ethanoloamine phosphate to phosphatidylinositol (see Topic E1) in the RER membrane. A **transamidase** enzyme then cleaves off the C-terminal signal sequence and concomitantly adds on the complete GPI anchor.

Proteins of the endoplasmic reticulum

The RER contains many proteins that have the role of assisting nascent proteins to fold correctly into their native conformation. Some of these are called **chaperones** (see Topic B3). RER-resident proteins are made on the RER and either pass into the lumen (as do secretory proteins) or are anchored in the membrane (as for Type I integral membrane proteins). However, these proteins contain a **retention signal** at the C terminus that is recognized by specific receptor proteins which retain these proteins in the RER, preventing them from moving along the secretory pathway to the Golgi. In the case of soluble proteins in the lumen of the RER, the retention signal is Lys-Asp-Glu-Leu (or **KDEL** using the one-letter amino acid code) at the C terminus. In the case of Type I integral membrane proteins in the RER membrane, the retention signal is Lys-Lys-Xaa-Xaa (**KKXX**) in the cytosolic C terminus.

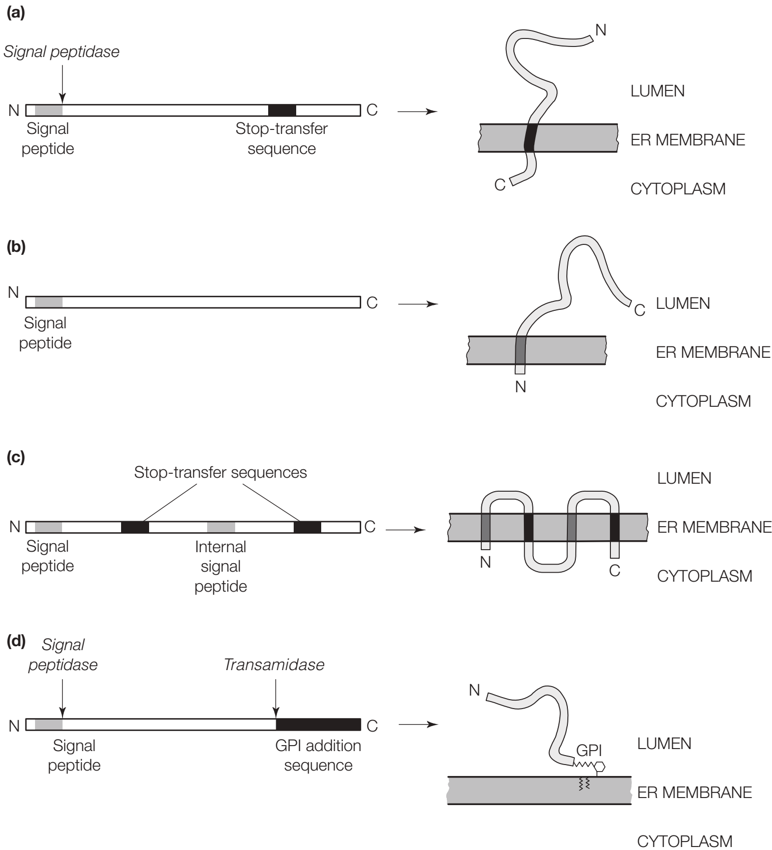


Fig. 4. Insertion of integral membrane proteins into the RER membrane during synthesis. (a) Type I integral membrane protein with a cleavable N-terminal signal sequence and a stop-transfer sequence; (b) Type II integral membrane protein with an uncleaved N-terminal signal sequence; (c) Type III integral membrane protein with multiple signal and stop-transfer sequences; (d) glycosyl-phosphatidylinositol (GPI) anchored membrane protein with a cleavable N-terminal signal sequence and a C-terminal GPI anchor addition sequence.

Lysosomal proteins

Lysosomal enzymes and lysosomal membrane proteins are synthesized on the RER and transported to the *cis* compartment of the Golgi complex. Here they become glycosylated and **mannose 6-phosphate** is added to the protein. The mannose 6-phosphate is the signal that targets the lysosomal protein to its correct destination. It is recognized by **mannose 6-phosphate receptor** proteins in the *trans* compartment of the Golgi which bind to the lysosomal protein and package it in transport vesicles that bud from the Golgi apparatus (*Fig. 5*). The transport vesicles then fuse with sorting vesicles, the contents of which are acidic. The low pH causes dissociation of the lysosomal protein from its receptor and a phosphatase removes the phosphate from the mannose 6-phosphate, preventing it from re-binding to the receptor. Vesicles bud from the sorting vesicle to return the receptor to the Golgi for re-use (**receptor recycling**) and the lysosomal protein is now delivered to the lysosome by vesicle fusion with it (*Fig. 5*).

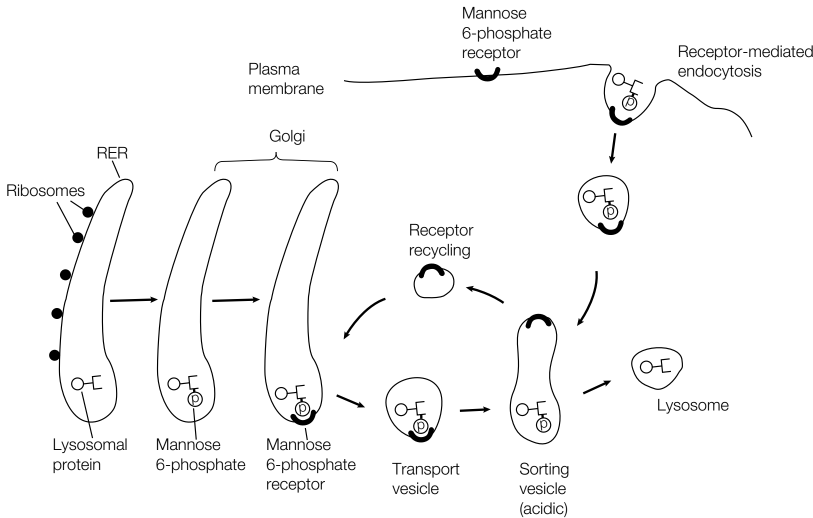


Fig. 5. Synthesis and targeting of lysosomal proteins.

Not all lysosomal proteins take the normal route of protein targeting; some end up being exported by the cell and must be retrieved. This **scavenger pathway** works as follows. The lysosomal glycoprotein binds to mannose 6-phosphate receptors in the plasma membrane and is internalized again by endocytosis (*Fig. 5*). This process, called **receptor-mediated endocytosis**, creates an endocytic vesicle (or **endosome**) that then delivers the lysosomal protein to the lysosome by fusion (see Topic E4).

Mitochondrial and chloroplast proteins

Mitochondria and chloroplasts contain their own DNA, ribosomes, mRNA, etc., and carry out protein synthesis, but very few mitochondrial or chloroplast proteins are made in this way. Rather, the large majority of mitochondrial and chloroplast proteins are encoded by the nuclear genome, are synthesized in the cytosol on free ribosomes, released after synthesis and then imported into the organelle. Thus, the process is **post-translational**. The protein may need to be targeted to any one of several locations; for mitochondria this could be the **outer mitochondrial membrane**, the **inner membrane**, the **intermembrane space** or the mitochondrial **matrix**. Chloroplasts have the same subcompartments plus two other potential destinations, the **thylakoid membrane** and the **thylakoid space** (see Topics A2 and L3). Most is known about mitochondrial protein uptake.

Proteins are targeted to the mitochondrial matrix by an N-terminal **matrix-targeting sequence**. This sequence is typically 15–35 amino acids long and rich in hydrophobic amino acids, the hydroxylated amino acids serine and threonine, and the positively charged amino acids arginine and lysine. These matrix-targeting sequences probably assume an α-helical conformation with the positively charged amino acids on one side of the helix and the hydrophobic amino acids on the other, such that these sequences are amphipathic. After synthesis by cytosolic ribosomes, the protein is released into the cytosol but is kept in an unfolded state by chaperone proteins called the **hsp70** family of proteins which bind to it during synthesis. This is necessary since folded proteins cannot be imported into mitochondria. The hsp70 then transfers the

unfolded protein to an **import receptor** in the outer mitochondrial membrane that is believed to slide along the membrane until it reaches a site where the inner membrane and outer membrane are in contact (a **contact site**). At this point it passes into the matrix via a protein translocator formed from the components of both membranes (*Fig. 6*). As it passes through the pore, the cytoplasmic hsp70 is released, the signal peptide is cleaved off by a signal peptidase, and the protein is bound in the matrix by **mitochondrial hsp70**. The hsp70 is then replaced by **mitochondrial hsp60** which assists the protein to fold correctly into its final active state. Import of proteins into the mitochondrion requires energy from the electrochemical gradient across the inner membrane (see Topic L2) as well as ATP hydrolysis. Protein import into the mitochondrial inner membrane and intermembrane space needs two signals; the protein is first imported into the matrix as described above and then a second signal sequence directs the protein back into the inner membrane or across it into the intermembrane space.

Protein import into chloroplasts follows similar mechanisms to those in mitochondria but the signals used must be different since mitochondria and chloroplasts are present together in some plant cells and yet proteins become targeted to the correct destination.

Nuclear proteins

The nucleus has an inner and an outer membrane (see Topic A2) and is perforated by 3000–4000 **nuclear pores**. Each pore consists of a **nuclear pore complex** of more than 100 different proteins organized in a hexagonal array. Although small molecules can pass through the pore by free diffusion, large proteins entering the nucleus require **a nuclear localization signal**. This is four to eight amino acids long and is rich in the positively charged amino acids lysine and arginine as well as usually containing proline, and is located internally within the polypeptide chain. The protein is taken through the pore in an ATP-requiring step and enters the nucleus without cleavage of the localization signal.

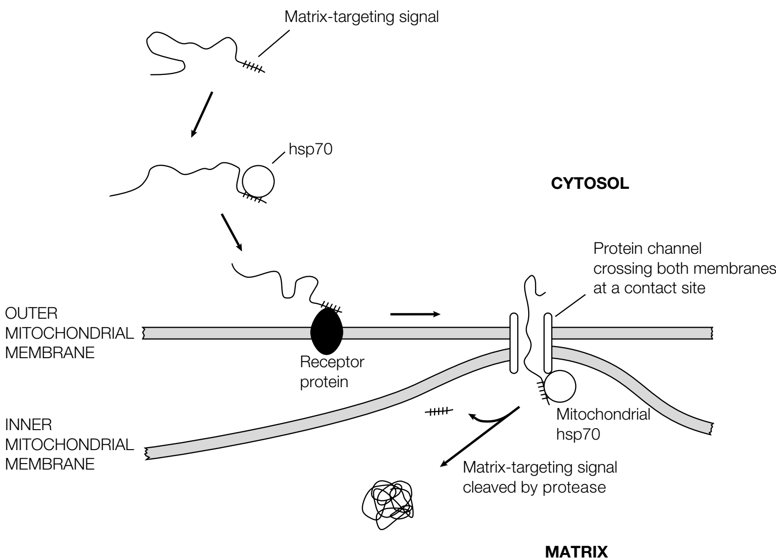


Fig. 6. Uptake of proteins into the mitochondrial matrix; see text for details.

H5 PROTEIN GLYCOSYLATION

Key Notes

Protein glycosylation: overview

Many proteins synthesized by the ribosomes of the RER contain short chains of carbohydrates (oligosaccharides) and are called glycoproteins. The oligosaccharides are of two main types; O-linked (to the OH side chain of Ser or Thr) and N-linked (to the NH_2 side chain of Asn). To be modified by N-linked glycans the Asn residues must be in the consensus sequence Asn-Xaa-Ser/Thr.

Synthesis of O-linked oligosaccharides

O-linked oligosaccharides are synthesized by the sequential addition of monosaccharides to the protein as it passes through the Golgi complex. These oligosaccharides usually consist of only 4–5 sugar residues.

Synthesis of N-linked oligosaccharides

N-linked oligosaccharides are initially synthesized on a dolichol phosphate carrier that is anchored to the RER membrane. The completed precursor structure with the composition $(Glc)_3(Man)_9(GlcNAc)_2$ is then transferred to the protein by the enzyme oligosaccharide transferase. Before leaving the RER the three terminal glucose residues are removed. The resulting high mannose type oligosaccharide may be trimmed down to a pentasaccharide core with the composition $(Man)_3(GlcNAc)_2$ and additional monosaccharides added in the Golgi to produce a complex type of oligosaccharide.

Related topics

Collagen (B5)
Protein targeting (H4)
Monosaccharides and disaccharides (J1)
Cholesterol (K5)

Protein glycosylation: overview

Most proteins made by ribosomes on the rough endoplasmic reticulum (RER) (see Topic H4) are glycoproteins, that is they contain short chains of carbohydrates (oligosaccharides) covalently linked to them during passage through the RER and Golgi complex. Two main types of oligosaccharide linkage exist:

- **O-linked oligosaccharides** are commonly attached to the protein via O-glycosidic bonds to OH groups of Ser or Thr side chains (*Fig. 1a*). The residues that are modified are often in a region of the polypeptide chain that is rich in Ser and/or Thr residues. In plant glycoproteins, the OH groups of hydroxyproline (Hyp) residues can be glycosylated, while the hydroxylysine (Hyl) residues in the mammalian protein collagen can be O-glycosylated (see Topic B5).
- **N-linked oligosaccharides** are linked to the protein via N-glycosidic bonds, to the NH_2 groups of Asn side chains (*Fig. 1b*) where the Asn occurs in the sequence Asn-X-Ser (or Thr) where X is any amino acid except Pro. If the Asn is N-glycosylated the Ser or Thr residue in this motif is not O-glycosylated.

Synthesis of O-linked oligosaccharides

The synthesis of O-linked oligosaccharides occurs by the sequential addition of monosaccharide units to the newly synthesized protein as it passes through the Golgi complex. First, *N*-acetylgalactosamine (GalNAc) is transferred to the

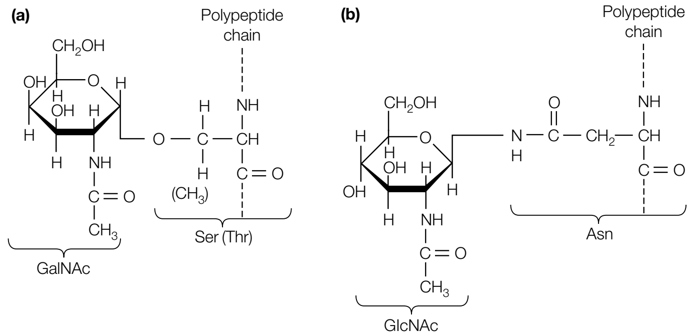


Fig. 1. Structures of oligosaccharide linkages. (a) O-linked glycosidic bond between GalNAc and Ser (Thr) residue. (b) N-linked glycosidic bond between GlcNAc and an Asn residue.

relevant Ser or Thr residue of the protein by GalNAc transferase, an enzyme that uses UDP-GalNAc as the precursor (*Fig. 2*). Other monosaccharides [galactose, *N*-acetylglucosamine (GlcNAc), sialic acid, fucose] (see Topic J1) are then added using the corresponding sugar nucleotides as precursors. The exact type and number (usually only 4–5) of monosaccharides added depends on the protein that is being modified.

Synthesis of N-linked oligosaccharides

In contrast to O-linked oligosaccharides which are built up sequentially on the protein, N-linked oligosaccharides are synthesized as a large, branched precursor structure that is then added en bloc to the acceptor Asn residue. The oligosaccharide is made on a lipid carrier called **dolichol phosphate**. This consists of 22 isoprene (C5) units (see Topic K5) with a terminal phosphate group and it is anchored to the RER membrane.

Synthesis of the oligosaccharide starts by the dolichol phosphate accepting monosaccharides from the cytosolic face of the RER membrane (*Fig. 3*) but when the $(Man)_5(GlcNac)_2$-dolichol phosphate intermediate has formed, this flips orientation and now accepts further monosaccharides from the lumenal side of the RER membrane (*Fig. 3*). All of these subsequent transfers are from dolichol phosphate-linked monosaccharides that are made on the cytoplasmic side of the RER membrane then likewise flipped across to act as donors. The final oligosaccharide, with the composition $(Glc)_3(Man)_9(GlcNAc)_2$ is linked to the dolichol by a high-energy pyrophosphate bond (*Fig. 3*). This provides the energy for transfer of the oligosaccharide to the protein, a reaction catalyzed by a membrane-bound **oligosaccharide transferase** enzyme and which occurs in the RER (*Fig. 4*).

Whilst the protein is still in the RER, the three glucose residues are removed (*Figs. 4 and 5*). Interestingly, glucose residues are added back to the protein if it

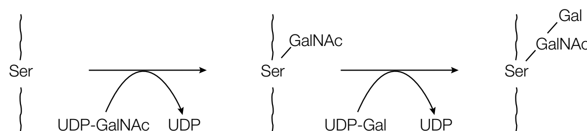


Fig. 2. Synthesis of O-linked oligosaccharide. The example shown is an O-linked oligosaccharide in human immunoglobulin A (IgA).

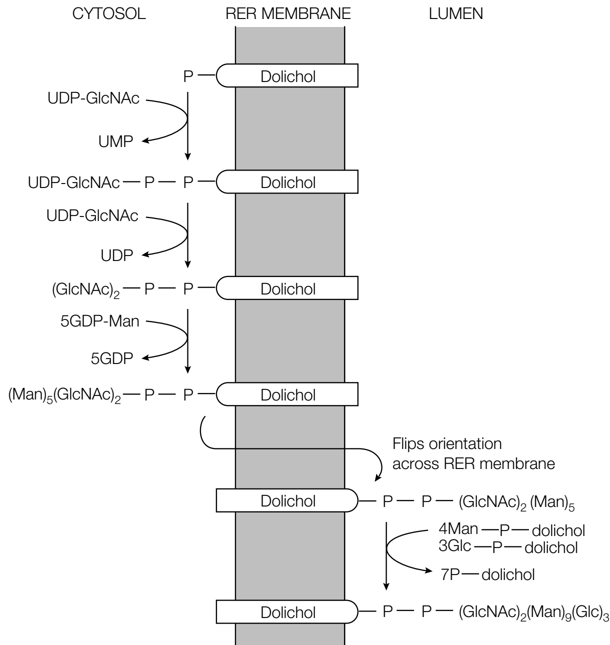


Fig. 3. Synthesis of N-linked oligosaccharides on a dolichol phosphate carrier in the RER membrane.

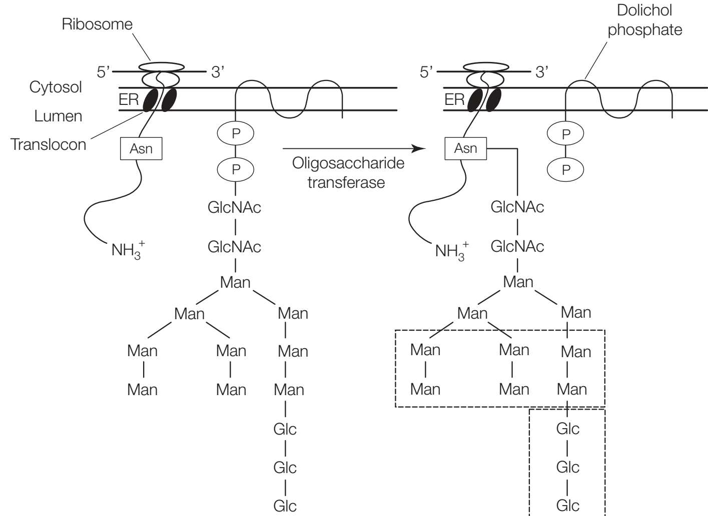


Fig. 4. Transfer of the N-linked oligosaccharide from the dolichol phosphate carrier to an acceptor Asn residue in a growing polypeptide chain by oligosaccharide transferase. The terminal three glucose residues and six mannose residues that are removed during trimming of the oligosaccharide are indicated.

is unfolded or wrongly folded. Thus, only when the protein is correctly folded are all glucose residues finally removed and the protein may continue along the secretory pathway. The folded glycoprotein with its **high-mannose type oligosaccharides** (see *Fig. 3* in Topic J2) is now transported to the Golgi complex via vesicles. As it moves through the Golgi complex, the oligosaccharide is often '**trimmed**' or '**processed**' with the terminal six mannose residues being removed to leave the **pentasaccharide core** (*Figs. 4 and 5*). Further mannose residues and other monosaccharides may then be added to the oligosaccharide in the Golgi to generate the **complex type of oligosaccharide** (*Fig. 5*; also see *Fig. 3* in Topic J2).

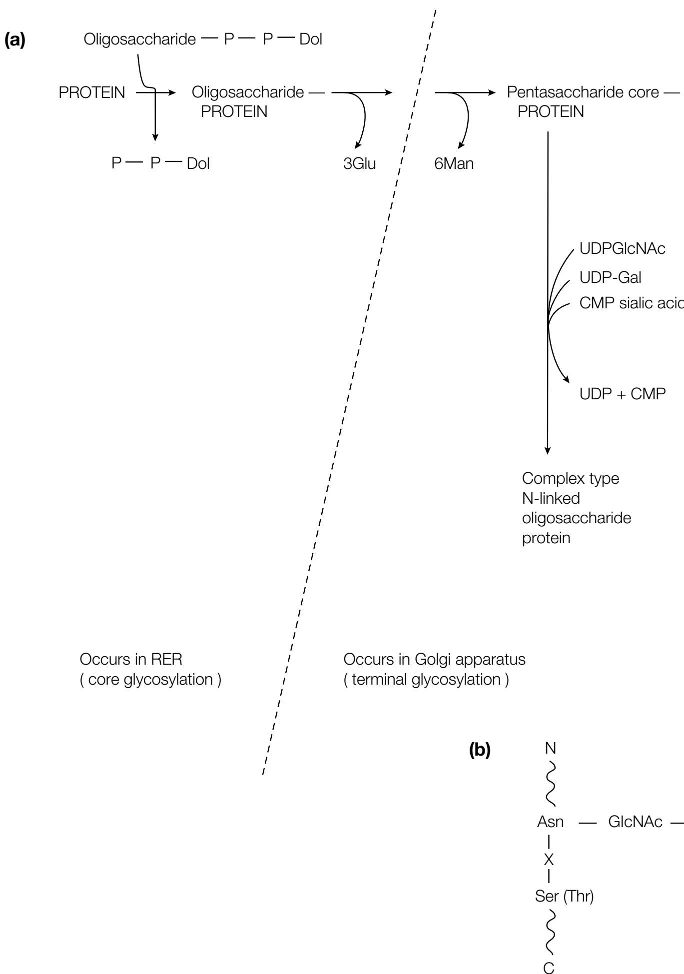


Fig. 5. (a) Transfer of the N-linked oligosaccharide to a protein and its further processing in the RER and Golgi (see text for details). (b) Structure of the pentasaccharide core of N-linked oligosaccharides.