RNA Structure

Ribonucleic acid (RNA) is a polymer of ribonucleotides (figure 13.4) that contains the sugar ribose and the bases adenine, guanine, cytosine, and uracil (instead of thymine). The nucleotides are joined by a phosphodiester bond, just as they are in DNA. Most RNA molecules are single stranded. However, an RNA strand can coil back on itself to form secondary structures such as hairpins with complementary base pairing and helical organization (p. 304). The formation of double-stranded regions in RNA is often critical to its function.

⏭ *Regulation of transcription elongation (section 34.3)*

Protein Structure

Proteins are polymers of amino acids linked by peptide bonds; thus they are also called polypeptides. An amino acid is defined by the presence of a central carbon (the α carbon) to which are attached a carboxyl group, an amino group, and a side chain (figure 13.6). Twenty amino acids are normally used to form proteins. However, two unusual amino acids have recently been discovered in some proteins (section 13.6). Amino acids differ in terms of their side chains. Depending on the structure of the side chain, the amino acid can be described as nonpolar, polar, or

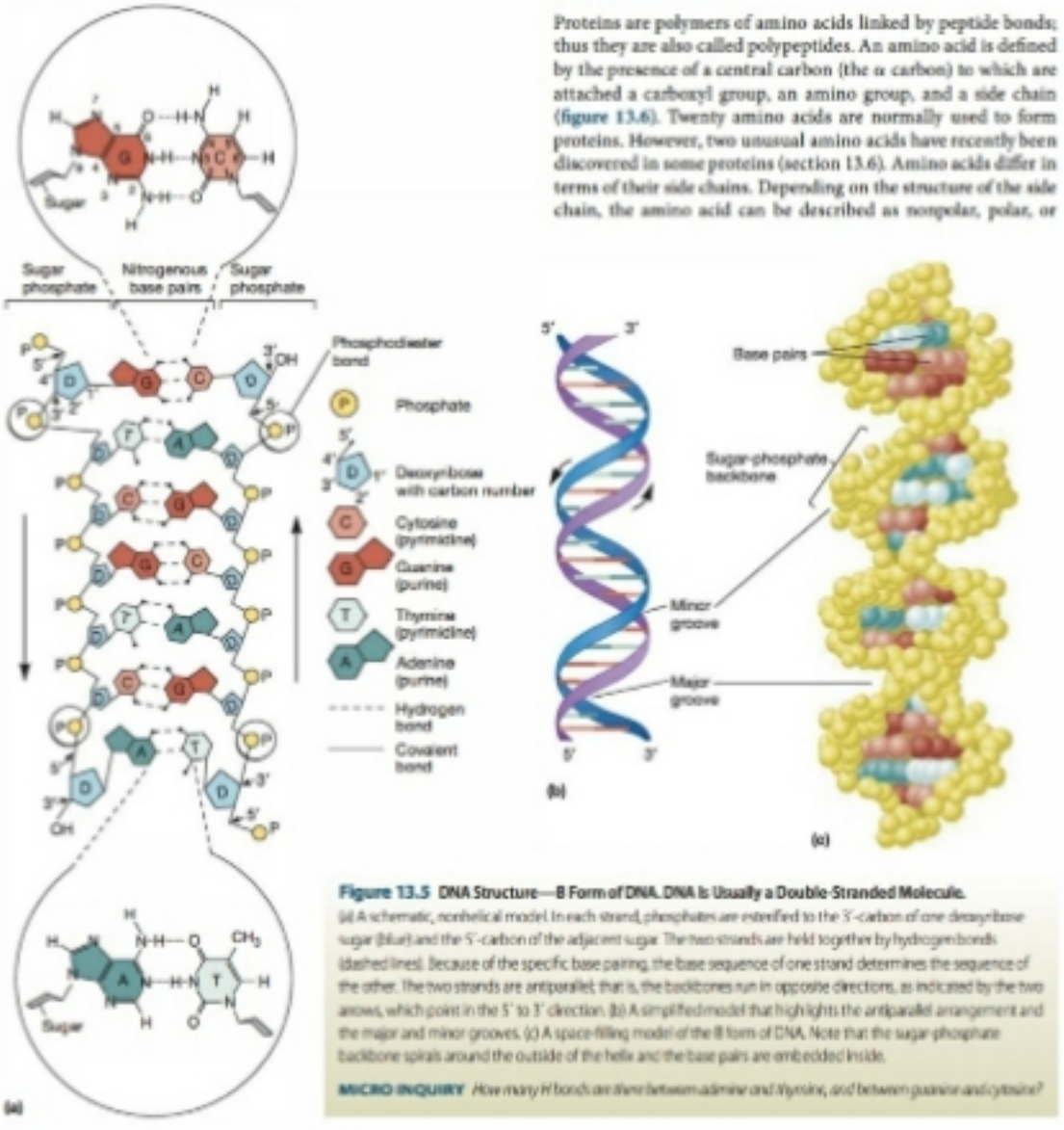


Figure 13.5 DNA Structure—B Form of DNA. DNA Is Usually a Double-Stranded Molecule.
(a) A schematic, nonhelical model. In each strand, phosphates are esterified to the 5′-carbon of one deoxyribose sugar (blue) and the 5′-carbon of the adjacent sugar. The two strands are held together by hydrogen bonds (dashed lines). Because of the specific base pairing, the base sequence of one strand determines the sequence of the other. The two strands are antiparallel; that is, the backbones run in opposite directions, as indicated by the two arrows, which point in the 5′ to 3′ direction. (b) A simplified model that highlights the antiparallel arrangement and the major and minor grooves. (c) A space-filling model of the B form of DNA. Note that the sugar-phosphate backbone spirals around the outside of the helix and the base pairs are embedded inside.

MICRO INQUIRY *How many H bonds are there between adenine and thymine, and between guanine and cytosine?*

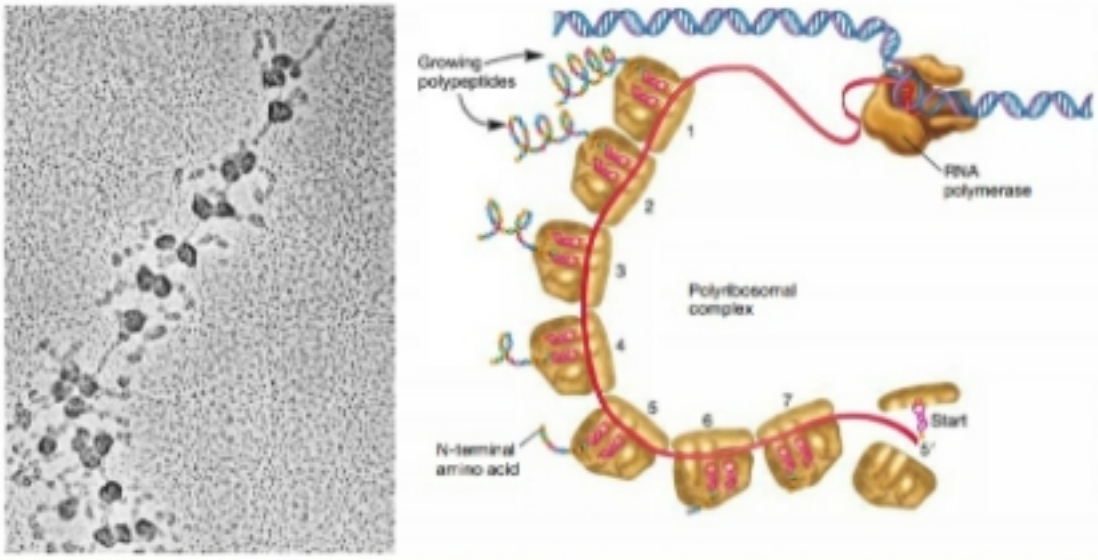


Figure 13.33 Coupled Transcription and Translation in Bacteria. (a) A transmission electron micrograph showing a polyribosome. (b) A schematic representation of coupled transcription and translation. As the DNA is transcribed, ribosomes bind the free 5′ end of the mRNA. Thus translation is started before transcription is completed. Note that there are multiple ribosomes bound to the mRNA, forming a polyribosome. The ribosomes are shown at different points in the translation process. Ribosomes 1, 2, 5, and 7 have completed the transpeptidation reaction, but translocation has not yet occurred. Ribosomes 3 and 4 have an A site containing an incoming aminoacyl-tRNA. Transpeptidation has not occurred. Ribosome 6 shows elongation upon completion of both transpeptidation and translocation. The tRNA bearing the growing polypeptide is in the P site and the empty tRNA is in the E site.

MICRO INQUIRY *Why is simultaneous transcription and translation impossible in eukaryotes?*

tRNA to assume a cloverleaf conformation (figure 13.34a). However, the three-dimensional structure looks like the letter L (figure 13.34b). One important feature of tRNAs is the acceptor stem, which holds the activated amino acid. The 3′ end of all tRNAs has the same —C—C—A sequence, and in all cases, the amino acid is attached to the terminal adenylic acid (A). Another important feature of the tRNA is the **anticodon.** The anticodon is complementary to the mRNA codon and is located on the anticodon arm (figure 13.34a).

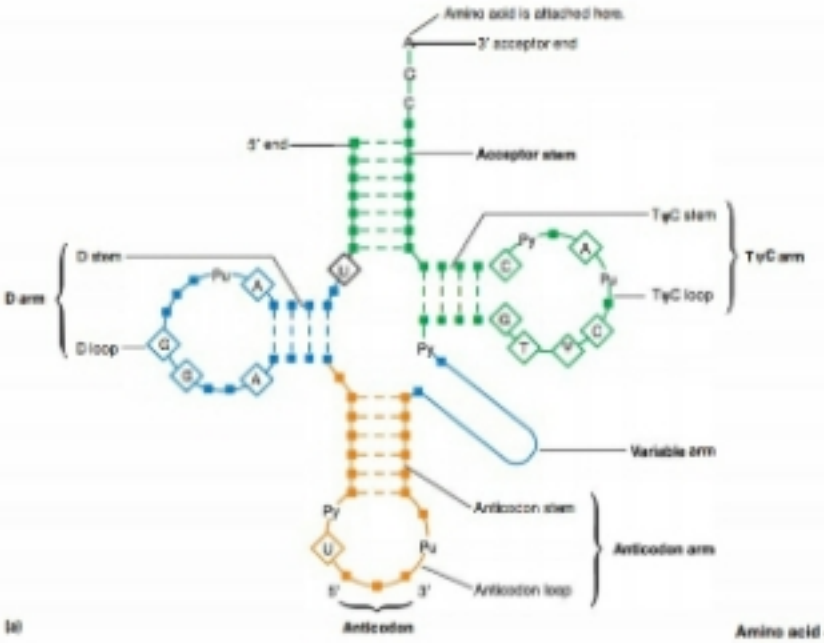
Enzymes called **aminoacyl-tRNA synthetases** catalyze amino acid activation (figure 13.35). As is true of DNA and RNA synthesis, the reaction is driven to completion when ATP is hydrolyzed to release pyrophosphate. The amino acid is attached to the 3′-hydroxyl of the terminal adenylic acid on the tRNA by a high-energy bond. The storage of energy in this bond provides the fuel needed to generate the peptide bond when the amino acid is added to the growing peptide chain.

There are at least 20 aminoacyl-tRNA synthetases, each specific for a single amino acid and its tRNAs (cognate tRNAs). It is critical that each tRNA attach the corresponding amino acid because if an incorrect amino acid is attached to a tRNA, it will be incorporated into a polypeptide in place of the correct amino acid. The protein synthetic machinery recognizes only the anticodon of the aminoacyl-tRNA and cannot tell whether the correct amino acid is attached. Some aminoacyl-tRNA synthetases proofread just like DNA polymerases do. If the wrong amino acid is attached to tRNA, the enzyme hydrolyzes the amino acid from the tRNA, rather than release the incorrect product.

● *Aminoacyl-tRNA Structure*

Ribosome Structure

Protein synthesis takes place on ribosomes that serve as workbenches, with mRNA acting as the blueprint. Recall that ribosomes are formed from two subunits, the large subunit and the small subunit, and each contains one or more rRNA molecules and numerous polypeptide chains. A bacterial ribosome and its components are shown in figure 13.36. The



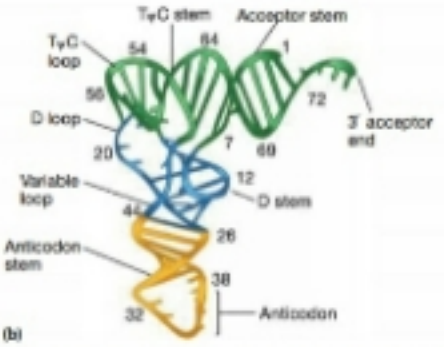


Figure 13.34 tRNA Structure. (a) The two-dimensional cloverleaf structure for tRNA. In addition to the anticodon arm, three other arms are readily observed: the D or DHU arm, the T or TψC arm, and the variable arm. The D and T arms are named because of the presence of unusual nucleosides. The variable arm is of different lengths depending on the tRNA; the other arms are fairly constant in size. Bases found in all tRNAs are in diamonds; purine and pyrimidine positions in all tRNAs are labeled Pu and Py, respectively. (b) The three-dimensional structure of tRNA. The various regions are distinguished with different colors.

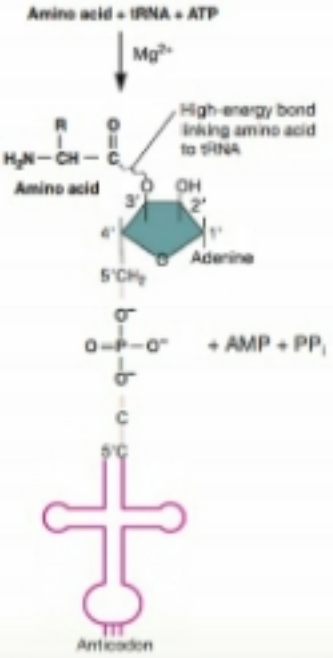
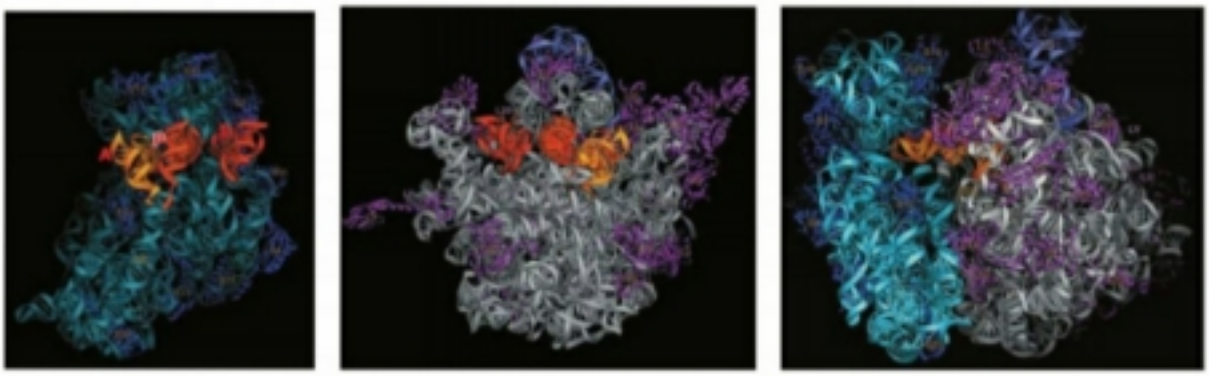


Figure 13.35 Aminoacyl-tRNA Synthetase Reaction. The amino acid is attached by the appropriate aminoacyl-tRNA synthetase to the 3'-hydroxyl of adenylic acid by a high-energy bond (red).

MICRO INQUIRY *What would be the outcome if an aminoacyl-tRNA synthetase added the wrong amino acid to a tRNA (i.e., the anticodon specified a different amino acid than that added to the 3' end of the tRNA)?*

region of the ribosome directly responsible for translation is called the translational domain. Both subunits contribute to this domain. The growing peptide chain emerges from the large subunit at the exit domain (figure 13.36d).

Ribosomal RNA is thought to have three roles. (1) It contributes to ribosome structure. (2) The 16S rRNA of the 30S subunit is needed for the initiation of protein synthesis because its



(a) 30S subunit contains:
16S rRNA
21 polypeptides
(0.9×10^6 daltons)

(b) 50S subunit contains:
5S rRNA
23S rRNA
34 polypeptides
(1.8×10^6 daltons)

(c) 70S ribosome (2.8×10^6 daltons)

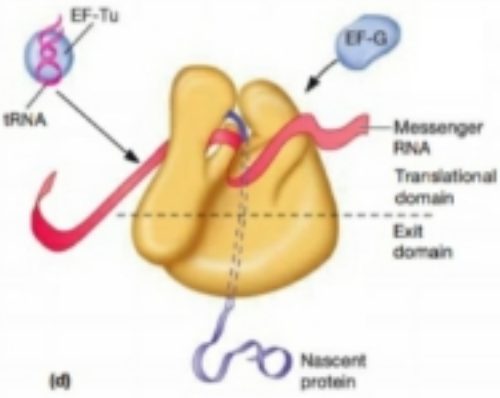


Figure 13.36 Bacterial Ribosome Structure. (a) Interior interface view of the 30S subunit of the *Thermus thermophilus* 70S ribosome showing the positions of the A, P, and E site tRNAs. (b) Interior interface view of the *T. thermophilus* 50S subunit and portions of its three tRNAs. (c) The complete *T. thermophilus* 70S ribosome viewed from the right-hand side with the 30S subunit on the left and the 50S subunit on the right. The anticodon arm of the A site tRNA is visible in the interface cavity. (d) A diagram of ribosomal structure. The components in figures (a)–(c) are colored as follows: 16S rRNA, cyan; 23S rRNA, gray; 5S rRNA, light blue; 30S proteins, dark blue; 50S proteins, magenta; and A, P, and E site tRNAs (gold, orange, and red, respectively).

3′ end binds to a site on the leader of the mRNA called the Shine-Dalgarno sequence; thus the Shine-Dalgarno sequence is part of the **ribosome-binding site** (**RBS**). This helps position the mRNA on the ribosome. The 16S rRNA also binds a protein needed to initiate translation (initiation factor 3) and the 3′ CCA end of amino-acyl-tRNA. (3) The 23S rRNA is a ribozyme that catalyzes peptide bond formation.

Initiation of Protein Synthesis

Like transcription and DNA replication, protein synthesis is divided into three stages: initiation, elongation, and termination. The initiation of protein synthesis is very elaborate. Apparently the complexity is necessary to ensure that the ribosome does not start synthesizing a polypeptide chain in the middle of a gene—a disastrous error.

Bacteria begin protein synthesis with a modified aminoacyl-tRNA, *N*-formylmethionyl-tRNA^{fMet} (fMet-tRNA), which is coded for by the start codon AUG (**figure 13.37**). The amino acid of the initiator tRNA has a formyl group covalently

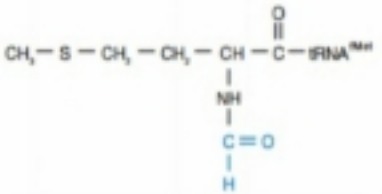


Figure 13.37 ***N*-formylmethionyl-tRNA^{fMet} Is the Initiator tRNA Used by Bacteria.** The formyl group is in color. Archaea and eukaryotes use methionyl-tRNA for initiation.

MICRO INQUIRY *Why would it be impossible for fMet-tRNA to initiate peptide bond formation with another amino acid? (Hint: Examine figure 13.40 closely.)*

acids.

1. Amino acids involved in protein synthesis.

About 150 amino acids are found in nature, of which only 20 are specified by the genetic code. Only these 20 amino acids take part in protein synthesis. Among the other amino acids found in proteins are *cystine* and *hydroxyproline*. *Cystine* is a double amino acid consisting of two *cysteine* units. It is formed by oxidation which produces a disulphide linkage between two cysteine molecules. This disulphide linkage

Fig. 13.3.The genetic code.

1st base in the codon	Second base in the codon: U		C		A		G		3rd base in the codon
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine	C
	UUA	Leucine	UCA	Serine	UAA	TERMINATION (3) (ochre)	UGA	TERMINATION (3) (opal)	A
	UUG	Leucine (2)	UCG	Serine	UAG	TERMINATION (3) (amber)	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 (1) (INITIATION)	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 (4)	A
	GUG	Valine (2)	GCG	Alanine	GAG	Glutamic acid	GGG	Glycine	G

(The numbers in brackets refer to the notes for Table 13.3).

plays an important part in maintaining the tertiary structure of proteins. It is cysteine which is incorporated in the polypeptide chain during protein synthesis. Cystine is then formed as a result of interaction between two cysteine molecules. *Hydroxyproline*, which is particularly found in collagen, is formed by the addition of hydroxyl (OH) group to *proline* after the latter has been incorporated into the polypeptide chain.

The genes of a cell contain coded information for the maintenance and reproduction of the cell. They direct the arrangement of the 20 types of amino acids into the polypeptide chains of the protein molecules. A polypeptide chain typically contains about 100-300 amino acids and is formed by specific arrangement of the 20 types of amino acids.

2. The genetic code is a triplet code.

DNA contains four kinds of nucleotides (of A, T, G and C), and proteins are synthesized from 20 different types of amino acids. A basic problem regarding the genetic code was : how many bases of DNA specify one amino acid ? In a *singlet code* each base or letter would specify one amino acid. Only 4 of the 20 types of amino acids would be coded unambiguously by a singlet code (Table 13.1). In a two-letter or *doublet* code two bases would specify one amino acid. Here 16 (4 x 4) of the 20 amino acids can be specified, but there would be ambiguous deteremination of a number of amino acids. A *triplet* or three-letter code was first suggested by the physicist Gamow in 1954. According to the triplet code three letters or bases specify one amino acid. Thus 64 (4 x 4 x 4) distinct triplets of purine and/or pyrimidine bases determine the 20 amino acids. These triplets have been called *codons*. Since there are 64 codons and only 20 amino acids it is obvious that there are many more codons than there are amino acids, i.e. the code is *degenerate*. Experimental evidence shows that the code is a triplet one and that 61 of the 64 codons code for individual amino acids during protein synthesis.

Table 13.1. The maximum possible number of codons in the singlet, doublet and triplet codes.

Type of code	*Number of bases in codon*	*Number of codons*	*Ambiguous/ degenerate*
Singlet code	1	4	Ambiguous
Doublet code	2	4x4=16	Ambiguous
Triplet code	3	4x4x4=64	Degenerate

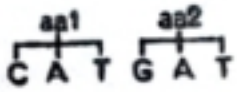
A *quadruplet code* would have 4x4x4x4 = 256 codons, and would show even more degeneracy than the triplet code.

3. The code is non-overlapping.

Since the DNA molecule is a long chain of nucleotides, it could be read either in an *overlapping* or *non-overlapping* manner. The genetic code could thus be overlapping or non-overlapping. The reading of the code by these two different ways would yield different results. In the *non-overlapping code* six nucleotides would code for *two* amino acids, while in the *overlapping* code up to *four* could be coded (Fig. 13.1). In the non-overlapping code each letter is read only once while in the overlapping code it would be read three times, each time as a part of a different word. Mutational changes in one letter would affect only one word in the non-overlapping code, while it would affect three words in the overlapping code.

Non-overlapping code.

(C, A, T & G are bases.
aa1 and aa2 are amino acids).



Overlapping code.

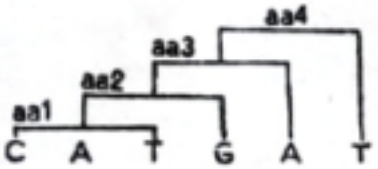


Fig. 13.1. Non-overlapping and overlapping code.

Studies on gene mutations show that *the code is of the non-overlapping type*. In the tobacco mosaic virus (TMV) mutation of one base of the nucleic acid into another results in the alteration of only a single amino acid. Similarly, studies on normal and sickle cell haemoglobins show that a single mutational change results in the substitution of only one amino acid.

Recently it has been shown that in the bacterial virus $\phi X174$ there is a possibility of overlapping of genes (Barrel and coworkers 1976, Sanger *et al*, 1977). Two genes contain the codes for second proteins with different amino acid sequences. Each of the two genes codes for different sequences of amino acids by a frame shift *(overlapping code)* and produces two totally different proteins (see 'Genes within genes', in chapter 9).

4. The code is commaless.

Is the genetic code read in an uninterrupted manner from one end of the nucleic acid chain to the other ? Or are there bases (commas) bet-

ween successive codons? A code with commas could be represented as follows (the X represents a base acting as a comma).

UUU *X* CUC *X* GUA *X* UCC *X* ACC ..... Bases

Phe Leu Val Ser Thr ..... Amino acids

A mutation resulting in an addition or deletion of a base would affect only one amino acid of the polypeptide chain. The total genetic message would be only slightly changed.

UUU *X*—UC *X* GUA *X* UCC *X* ACC ..... Bases

Phe *Changed aa* Val Ser Thr ..... Amino acids

A commaless code would not have the comma bases and can be represented thus :

UUU CUC GUA UCC ACC ..... Bases

Phe Leu Val Ser Thr ..... Amino acids

In such a code any mutation involving a deletion of a base (—C) would result in a drastic change in the genetic message.

UUU UCG UAU CCA CC ..... Bases

Phe *Ser Tyr Pro* ..... Amino acids

The entire series of amino acids following the deletion would change.

All the available evidence indicates that *the code is commaless*, i.e. there are no demarcating signals between codons. The work of Khorana and his associates cited below gives clear evidence of a commaless code. Long synthetic polynucleotides with specific repeating sequences were used for translation of protein chains. Thus the repeating sequence CUCUCU.....contains the codons CUC (for *leucine*) and UCU (for *serine*). When this sequence is used for translation of proteins, neither amino acid is incorporated into the protein unless the other is also present. This result can only be explained by a commaless triplet code where there would have to be alternate translation of CUC and UCU codons.

5. The code has polarity.

If a gene is to specify the same protein repeatedly it is essential that the code must be read between fixed start and end points. These points are the *initiation* and the *termination codons*, respectively. It is also essential that the code must be read in a fixed *direction*. In other words the code must have *polarity*. It is obvious that if the code is read in opposite directions it would specify two different proteins, since the codons would have reversed base sequences. Thus if the message given below is read

from left to right the first codon, UUG, would specify *leucine*. If read

Codons	UUG	AUC	GUC	UCG	CCA	ACA	AGG
⟶	Leu	Ile	Val	Ser	Pro	Thr	Arg
	Val	Leu	Leu	Ala	Thr	Thr	Gly ⟵

from right to left the codon would become GUU and would specify *valine*. It is thus seen that the sequence of amino acids constituting the protein would undergo a drastic change if the code is read in the opposite direction. The available evidence indicates that the message in mRNA is read in the 5'→3' direction. The polypeptide chain is synthesized in the N→C direction, i.e. from the *amino* (NH_2) terminal to the *carboxyl* (COOH) terminal.

6. Codons and anticodons.

During translation the codons of mRNA pair with complementary anticodons of tRNA. Since mRNA is read in a polar manner in the 5'→3' direction, the codons are also written in the 5'→3' direction. Thus the codon AUG is written as 5'AUG3'. The corresponding anticodon on tRNA should therefore be written as 5'CAU3', In such a configuration the first bases of both codon and anticodon would be the ones at the 5' end and third bases at the 3' end.

Base number		1	2	3	
Codon (mRNA)	5'	A	U	G	3'
Anticodon (tRNA)	3'	U	A	C	5'
Base number		3	2	1	

Often, however, the anticodon is written in the 3'→5' direction so as to bring about an easier correlation between the bases of the codon and anticodon. Thus the anticodon for AUG is written as 3' UAC 5' or, more simply, UAC. Here the first letter in the codon is at the 5' end and the first letter of the anticodon at the 3' end.

7. Initiation codons.

The starting amino acid in the synthesis of most protein chains is *methionine* (eukaryotes) or *N-formylmethionine* (prokaryotes). Methionyl or N-formyl methionyl-tRNA specifically binds to *initiation sites* containing the AUG codon. This codon is therefore called the *initiation codon*. Less often, GUG also serves as the initiation codon in bacterial protein synthesis. Normally GUG is the codon for *valine*. In the phage MS2, GUG is the initiation codon for the A protein. GUG has been found to initiate protein synthesis when the normal AUG codon is lost by deletion. However, initiation by GUG is less efficient, since it has a lower affinity for fMet-tRNA.

Both AUG and GUG codons show ambiguity in one sense, since each of them codes for two different amino acids. When these two codons are at *initiation positions* of mRNA they code for *N-formyl methionine*. In *internal positions* AUG codes for *methionine* and GUG for *valine*.

8. Termination codons.

Three of the 64 codons do not specify any tRNA, and were hence called *nonsense codons*. These codons are UAG (*amber*), UAA (*ochre*) and UGA (*opal* or *umber*). Since they bring about termination of polypeptide chain synthesis they are also called *termination codons*. UAG was the first termination codon to be discovered. It was named 'amber' after a graduate student named Bernstein (the German for 'amber') who helped in the discovery of a class of mutations. Apparently to give uniformity the other two termination codons were also named after colours.

Termination codons do not code for any amino acids and hence cause termination and release of polypeptide chains. Apparently no tRNA species has anticodons complementary to the termination codons. There are mRNAs with single termination codons and also mRNAs with two successive termination codons (e.g. MS2 coat protein mRNA). Termination codons are not read by any tRNA molecules but by proteins called *release factors*. In *prokaryotes* there are three release factors RF-1, RF-2 and RF-3. RF-1 recognizes UAA and UAG, while RF-2 recognizes UAA and UGA. RF-3 stimulates RF-1 and RF-2. In *eukaryotes* a single release factor (RF) recognizes all three termination codons.

9. The code is degenerate.

As mentioned previously, there are 64 possible codons in a triplet code, of which 61 have been shown to code amino acids. Since only 20 amino acids take part in protein synthesis, it is obvious that there are many more codons than amino acid types. Except for *tryptophan* and *methionine*, which have a single codon each, all other amino acids involved in protein synthesis have more than one codon. *Phenylalanine*, *tyrosine*, *histidine*, *glutamine*, *asparagine*, *lysine*, *aspartic acid*, *glutamic acid* and *cysteine* have two codons each. *Isoleucine* has three codons. *Valine*, *proline*, *threonine*, *alanine* and *glycine* have four codons each. *Leucine*, *arginine* and *serine* have six codons each (Table 13.2). This variability in the number of codons for different amino acids may at least partially account for the unequal distribution of the different amino acids in protein. In general, the frequency of appearance of amino acids in proteins roughly corresponds to the number of available codons.

Table 13.2 Number of codons coding for different amino acids. Amino acids in categories 2-5 are coded by more than one codon. Such codons are called degenerate.

Amino Acids	Number of codons
1. Tryptophan, methionine	1
2. Phenylalanine, tyrosine, histidine, glutamine, asparagine lysine, aspartic acid, glutamic acid, cysteine	2 2
3. Isoleucine	3
4. Valine, proline, threonine, alanine, glycine	4
5. Leucine, arginine, serine	6

10. The wobble hypothesis.

The triplet code is a degenerate one, with many more codons than the number of amino acid types coded. An explanation for this degeneracy is provided by the 'wobble hypothesis' proposed by Crick (1966). Since there are 61 codons specifying amino acids, the cell should contain 61 different tRNA molecules, each with a different anticodon. Actually, however, the number of tRNA molecule types discovered is much less than 61. This implies that the anticodons of some tRNAs read more than one codon on mRNA.

According to the wobble hypothesis only the first two positions of a triplet codon on mRNA have a precise pairing with the bases of the tRNA anticodon. The pairing of the third position bases of the codon may be ambiguous, and varies according to the nucleotide present in this position. Thus a single tRNA type is able to recognize two or more codons differing only in the third base. The anticodon UCG of *serine tRNA* recognizes two codons, AGC and AGU. The bonding between UCG and AGC follows the usual Watson-Crick pairing pattern. In UCG-AGU pairing, however, hydrogen bonding takes place between G and U. This is a departure from the usual Watson-Crick pairing mechanism where G pairs with C and A with U. Such interaction between the third bases is referred to as '*wobble pairing*'.

mRNA codons (serine)	5′	AG*C*	3′		5′	AG*U*	3′
tRNA anticodon	3′	UC*G*	5′		3′	UC*G*	5′

The degeneracy of the code is not random. Mostly, the different codons for a particular amino acid have the same first two letters (*leucine*, *serine* and *arginine* are exceptions). Thus the first two letters of all the four codons for *valine* are GU and for *alanine* GC.

Notes for Table 13.3

1. *Methionine* is the starting amino acid of eukaryote polypeptides, and *N-formyl methionine* of bacterial polypeptides. The codon for these amino acids is AUG, which is called the *initiation codon* because it starts protein synthesis.
2. *N-formyl methionine* is probably also coded by UUG and GUG, which then also function as *initiation codons*.
3. UAA, UAG and UGA, called **ochre, amber and opal**, respectively, were called *'nonsense' codons* since they do not code for any aimno acid. They are responsible for signalling the termination of the polypeptide chain, and are therefore called *termination codons or stop codons*. They are read by specific proteins called *release factors*, and not by any tRNAs.
4. GGA probably also codes for *glutamic acid*. When a codon codes for more than one amino acid the code is said to be *ambiguous*.

When only two codons specify an amino acid the third letters of the codons are either *both purines* or *both pyrimidines*: never one purine and one pyrimidine.

THIRD BASE

	Pyrimidine	Purine	Pyrimidine	Purine	
Codons	UU*U*	UU*A*	CA*U*	CA*A*	Watson-Crick pairing
	UU*C*	UU*G*	CA*C*	CA*G*	Wobble pairing
Amino acids	Phe	Leu	His	Glu	
Anticodons	AA*A*	AA*U*	GU*A*	GU*U*	

It is possible to predict the minimum number of tRNAs required to translate the different codons specifying a particular amino acid. The amino acid *leucine* is specified by six codons: UUA, UUG, CUU, CUC, CUA and CUG. The first two letters of two codons are UU and of four codons CU. Hence at least two different tRNAs are required, since the first two letters of a codon do not have wobble pairing with the anticodon.

Of the four codons having CU two (CUU and CUC) have *pyrimidines* as their third bases and two (CUA and CUG) have *purines*. Hence they cannot be read by the same anticodon, because the purine of the anticodon can only pair with a pyrimidine and *vice versa*. The CU-codons must therefore be read by at least two different anticodons. Thus *at least three codons* are required to read the anticodons for leucine.

Presumably the anticodons AUU reads UUA and UAG, GAA reads CUU and CUC, and GAU reads CUA and CUG. It will be seen that each anticodon has Watson-Crick pairing with third base of one codon, and wobble pairing with the third base of the other codon.

THIRD BASE

	Purine		Pyrimidine		Purine	
Codons	UU*A*	UU*G*	CU*U*	CU*C*	CU*A*	CU*G*
Anticodons	AA*U*	AA*U*	GA*A*	GA*A*	GA*U*	GA*U*
Type of pairing	WCP	WP	WCP	WP	WCP	WP

WCP = Watson-Crick pairing with anticodon.

WP = Wobble pairing with anticodon.

The six codons of leucine and their pairing with three types of anticodons.

The anticodon of certain tRNAs contains *inosine* (I), a deamination product of adenosine. Inosine closely resembles G and would thus be expected to pair with C. It is found to bond with C, A or U. Thus a tRNA with its anticodon having I in the wobble position would be able to pair with three codons having C, A or U in the third position. *Isoleucine* is the only amino acid to have three codons, these being AU*U* AU*C* and AU*A*. A tRNA molecule having the anticodon UAI would be able to pair with all these three codons (Fig. 13.3).

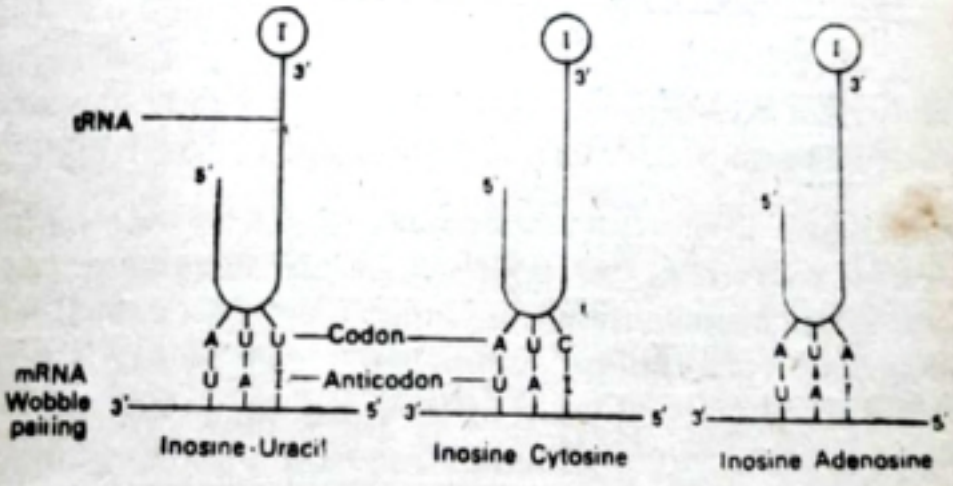


Fig. 13.3. Wobble pairing of inosine with uracil cytosine and adenine.

Codons for isoleucine	AU*U*	AU*C*	AU*A*
Anticodon	UA*I*	UA*I*	UA*I*

I cannot pair with G. Had it been able to do so the anticodon UAI would have also paired with AUG, the initiation codon specifying *methionine*.

Wobble pairing takes place in only certain combinations. Three types have been proposed : (i) U in the wobble position of the tRNA anticodon can pair with A or G of the mRNA codon, (ii) G can pair with U or C and (iii) I can pair with A, U or C.

11. Deciphering the code.

Prior to the breakthrough Nirenberg-Matthaei experiments of 1961, work on the genetic code was in the area of theoretical speculations and development of concepts. This period also has its importance because it laid the theoretical foundations for subsequent experimental work. The physicist George Gamow proposed the *diamond code* (1954) and the *triangle code* (1955) and gave a comprehensive theoretical framework to the different facets of the genetic code. Gamow's proposals dealt with the general features of the genetic code as recognized to-day. The main features of his proposals were :

i) A *triplet codon* corresponding to one amino acid of the polypeptide chain.

ii) *Direct template translation* by codon-amino acid pairing.

iii) Translation of the code in an *overlapping* manner.

iv) *Degeneracy* of the code, i.e. an amino acid being coded by more than one codon.

v) *Colinearity* of nucleic acid and the primary protein synthesized.

vi) *Universality* of the code, i.e. the code being essentially the same for different organisms.

In 1957 Brenner showed that the overlapping triplet code is an impossibility, and subsequent work has shown that the code is a *non-overlapping* one (with the exceptions mentioned previously).

Gamow's idea of direct template relationship between nucleic acid was challenged when Crick proposed his *adaptor hypothesis*. According to this hypothesis *adaptor molecules* intervene between nucleic acid and amino acids during translation. It is now known that tRNA molecules act as adaptors between codons of mRNA and amino acids of the resulting polypeptide chain. An offshoot of the adaptor hypothesis is the *comma-free* code.

The in vitro approach. In principle the most direct way to determine the code would be to determine the sequence of amino acids in a protein and the sequence of nucleotides in the mRNA specifying the proteins. Although this can be done today, no adequate methods were available in the early 1960s. Therefore, more indirect methods had to be employed.

The discovery of the enzyme *polynucleotide phosphorylase* in 1955 by Grunberg-Manago and Ochoa made possible the synthesis of polynucleotides containing only a single type of nucleotide repeated several times. Thus it became possible to synthesize *polynucleotides* containing only U, A, C or G nucleotides.

Polynucleotide	Configuration	Codes for :
Polyuridylic acid or Poly(U)	UUUUUU	Polyphenylalanine
Polyadenylic acid or Poly(A)	AAAAAA	Polylysine
Polycytidilic acid or Poly(C)	CCCCCC	Polyproline
Polyguanidylic acid or Poly(G)	GGGGGG	Inactive

The action of polynucleotide phosphorylase may be represented thus:

$$(RNA)_n + \text{Ribonucleoside diphosphate} \rightleftharpoons (RNA)_{n+1} + Pi$$

Polynucleotide phosphorylase differs from *RNA polymerase* used to transcribe mRNA from DNA in that :

(i) it does not require a *template* or *primer*

(ii) the activated substrates are *ribonucleotide diphosphates* and not *triphosphates*, and

(iii) *orthophosphate* (Pi) is produced instead of *pyrophosphate* (PPi).

The breaking of the genetic code was made possible by the use of synthetic polynucleotides and trinucleotides. The different types of techniques used include the use of polymers containing a single type of nucleotide (*homopolymers*), the use of mixed polymers containing more than one type of nucleotide (*heteropolymers*) in random or defined sequences and the use of *trinucleotides* ('*minimessengers*') in filter binding.

i) *The use of polymers containing a single type of nucleotide* (*homopolymers*). The landmark experiment which led to the decipbering of the genetic code was performed by Nirenberg and Matthaei (1961). Cells of the bacterium *E.coli* were broken open by grinding with finely powdered alumina to yield cell sap. The DNA of the cells was broken down by adding the enzyme *deoxyribonuclease*. The template for synthesizing new mRNA was thus destroyed. The cell sap was centrifuged to remove the heavier fragments of the cell wall and cell membranes. The slower-sedimenting cell-free extract contained ribosomes, enzymes, DNA, mRNA and tRNA.

To the cell-free system were added energy sources (ATP, GTP) and a mixture of all 20 amino acids, at least one of which was labelled with radioactive material (^{14}C). Protein synthesis took place in the cell-free system and as the mRNA is unstable, came to an end within a few

minutes. On addition of a crude fraction of mRNA and incubation of the mixture at 37°C for about an hour protein synthesis was resumed.

Addition of trichloroacetic acid stopped the reaction and precipitated the proteins. The free amino acids remained in solution. The protein precipitate was washed and its radioactivity measured by placing it in a radiation counting instrument. The amount of radio-activity indicated the amount of labelled amino acids incorporated into the protein. It was found that the ribosomes synthesized protein in response to addition of mRNA.

Another crucial component of this experiment was the use of the synthetic polynucleotide *poly(U)* in the cell-free system. Such polymers containing one type of monomer are called *homopolymers*. It was found that the ribosomes read the code in poly(U) and synthesized *polyphenylalenine*, a polypeptide containing only *phenyialanine* repeated over and over again. Since the synthetic messenger–RNA *poly(U)* coded for *polyphenylalanine*, the codon UUU was identified for *phenylalanine*. Thus the first code word to be deciphered was UUU.

This discovery was extended in the laboratories of Nirenberg and Ochoa. The experiment was repeated using synthetic *poly(A)* and *poly (C)* chains, which gave *polylysine* and *polyproline* chains, respectively. Thus AAA was indentified as the code for *lysine* and CCC as the code for *proline*. *Poly (G)* was found to be inactive, because it formed a triple-stranded helical structure.

ii) *The use of mixed polymers (heteropolymers) with random sequences.* Further elucidation of the code took place by using synthetic messengers containing two kinds of bases. This technique was employed in the laboratories of Ochoa and Nirenberg and led the deduction of the composition of the codons for the 20 amino acids. The synthetic messengers contained bases distributed at random (*random copolymers*). For example in a random polymer using U and A nucleotides eight triplets are possible:

UUU, UUA, UAA, UAU, AAA, AAU, AUU and AUA. Theoretically eight amino acids could be coded by these eight codons. Actual experiments, however, yielded only six, *phenylalanine, lysine, tyrosine, leucine, isoleucine* and *asparagine*. By varying the relative compositions of U and A in the synthetic messenger, and determining the percentage of the different amino acids in the proteins formed, it was possible to deduce the composition of the code for different amino acids.

iii) *The use of mixed polymers (heteropolymers) with defined sequences.* A few years later Khorana combinined organic synthetic and enzymatic techniques to synthesize RNA polymers having defined sequ-

ences. Combination of C and U nucleotides led to the synthesis of the polynucleotide CUCUCUCUC..... This contains alternating CUC and UCU codons. Ribosomes code this message to form a polypeptide of alternating *leucine* and *serine* amino acids.

Completely synthetic systems has been produced in the laboratory. Combining U and A results in the formation of a *dinucoletide* UA. Combining two such dinucleotides produces a *tetranucleotide* UAUA. Adding two such tetranucleotides results in UAUAUAUA. By further addition of such chains it is possible to obtain a *polynucleotide* which contains alternating U and A nucleotides indefinitely : UAUAUAUA-UAUAUA..... This polynucleotide contains the codons UAU and AUA repeated several times. These codons code for *tyrosine* and *isoleucine*, respectively. The polypeptide chain produced contains alternating tyrosine and isoleucine units.

It is possible to combine a UA doublet with a UC doublet to yield UAUC. By adding such tetranucleotides it is possible to get a short chain with message UAUCUAUCUAUC..... This can be resolved into the codons UAU, CUA, UCU and AUC, which code for *tyrosine*, *leucine*, *serine* and *isoleucine*, respectively. It is thus possible that in the future, polynucleotides coding for any desired protein may be synthesized. This in effect would be the artificial production of genes by synthetic methods.

iv) *The use of trinucleotides* (*minimessengers*) *in filter binding*. In 1964 Leder and Nirenberg developed a more direct technique for determining codons of amino acids. This technique employs *cellulose nitrate filters* and has been called the filter binding technique. Cellulose nitrate filters were originally used to isolate ribosomes from microorganisms. The ribosomes are left behind on the filter, while the tRNAs wash through the filter when mRNA is absent. In the presence of mRNA and ribosomes the tRNAs stick to the filter. mRNA causes binding of amino acid charged tRNA to the ribosomes. This technique was developed in the laboratories of Nirenberg and Khorana for analysing the genetic code.

A mixture of the synthetic messenger poly(U) and ribosomes was prepared on the filter. The various tRNAs, each carrying a specific amino acid labelled by ^{14}C, were individually passed through the filter. Only phenylalanine-tRNAs became attached to the poly(U) messenger and were retained on the filter. The other amino acid-tRNAs passed through. This showed that UUU coded for *phenylalanine*.

It was found that when the long messengers were substituted by different *trinucleotides* (triplets) the same results were obtained.

has been verified. (For exceptions to the colinearily principle see 'Split genes' in chapter 9).

13. The code is universal.

The genetic code is valid for all organisms ranging from bacteria to man. It is essentially the same for all organisms and is therefore said to be universal. The universality of the code was demonstrated by Marshall, Caskey and Nirenberg (1967), who found that *E. coli* (bacterium), *Xenopus laevis* (amphibian) and guinea pig (mammal) amino acyl tRNAs use almost the same code. This showed that the code is essentially universal.

Other evidence for the universality of the code comes from a study of gene mutations. Such mutations result in amino acid substitutions. Amino acid substitution resulting from gene mutations are known for coat protein in tobacco mosaic virus (TMV), α chain of tryptophan synthetase in *E. coli* and haemoglobin in man. A change in a single base can account for nearly all amino acid substitutions. This proves the universality of the code.

The code has remained constant since the time it was fixed when complex bacteria evolved (about three billion years ago). Any mutation altering the code reading would change the reading of mRNA. This in turn may change the amino acid sequence of the proteins synthesized by the organism. As many of these changes could be lethal, there would be a strong selection pressure against such a mutation. Hence the constancy of the code over a long period of time. Changes in proteins take place only with respect to the positions of particular amino acids. Moreover, only a few such changes take place at a time. Most mutations result in only a single amino acid substitution

In translation, the order of nucleotides of mRNA is read as triplet codons and determines the order of amino acids in the growing polypeptide chain.

RNA polymerase binds to a particular binding site at the beginning of a gene. Starting from there, it moves along the strand and encounters with each DNA nucleotide and adds the corresponding complementary RNA nucleotide to the growing mRNA strand. When ***RNA polymerase*** arrives at the stop signal at the opposite end of the gene, it gets disengaged from the DNA and a newly assembled mRNA molecule is released. The transcribed mRNA moves out of the nucleus, undergoes processing and attaches to the ribosomes to direct protein synthesis. Transcription occurs only during G_0 phase when the cell is matabolically active but not preparing for cell division or during G_1 and G_2 substages of interphase at the time of protein synthesis when the cell is getting ready for next cell division.

Translation begins when an tRNA molecule within the ribosome recognises and binds to the initiation codon on mRNA. The ribosome then moves along mRNA molecule, three nucleotides (= one codon) at a time. Codon specific tRNA with its amino acid comes to join the codon on mRNA. This is then transferred to the growing polypeptide chain. The ribosome continues to move on mRNA and its polypeptide chain keeps elongating, until ribosome reaches the stop/termination codon which does not code for any amino acid. The translation stopes, and the ribosomes, the polypeptide chain and mRNA are all set free.

13.5 Mechanism of Translation

The process of translation involves following steps:

- Activation of amino acids.
- Attachment of activated amino acids with tRNA (Amino-acetylation of tRNA).
- Stages of translation: Initiation, elongation and termination of polypeptide chain.
- Modification of released polypeptide chain.

1. Activation of Amino Acids

Amino acids in the cytoplasm occur in inactive form and they cannot take part in protein synthesis. Hence these are activated by giving them energy. The activation is provided by ATP, whose molecules unite with the amino acids forming highly reactive **amino acid phosphate-adenyl complexes** which are known as

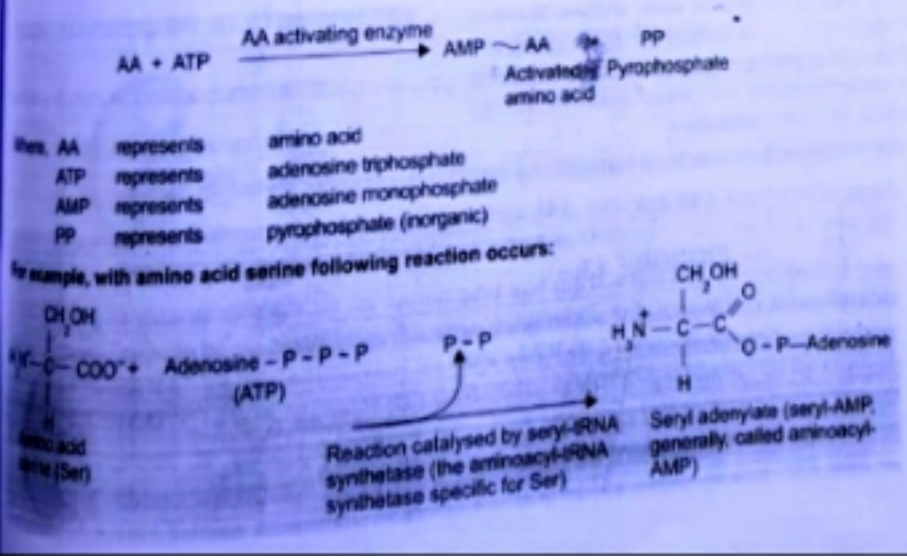


FIGURE 13.4
Reaction that occurs in the activation of amino acid serine with the help of enzyme *amino-acyl tRNA synthetase*

aminoacyl adenylates. The process of activation is governed by specific enzy... ***aminoacyl tRNA synthetase***. Usually each amino acid has its own specific ***amino... tRNA synthetase enzyme***. Hence, there are as many enzymes as the number... amino acids (*i.e.* 20). Each enzyme has double specificity. The specificity is pro... by **recognition region**. It recognises its own amino acid and finds out its own tR... The mechanism of activation is represented in Fig 13.4.

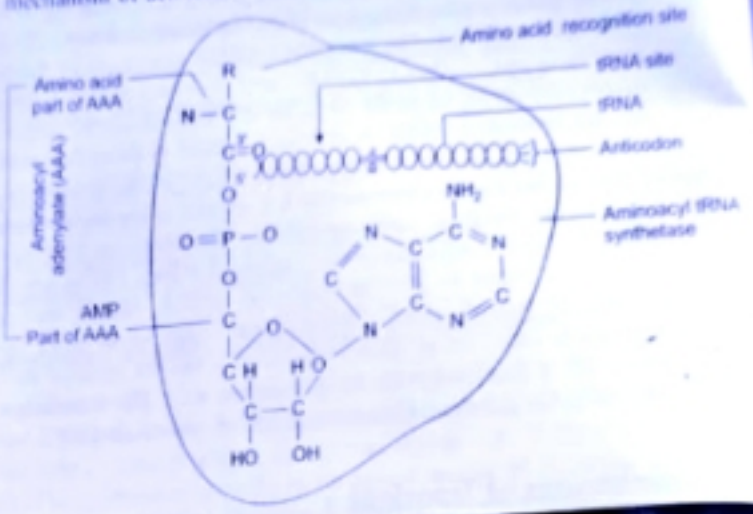


FIGURE 13.5

Aminoacyl adenylate (AAA) consisting of enzyme amino acyl tRNA synthetase, amino acid and tRNA

Extension: Aminoacyl tRNA Synthetases

The enzymes **aminoacyl tRNA synthetases** are required to link amino acids to their cognate tRNA molecules. There are twenty different ***aminoacyl tRNA synthetases*** in each cell, one for each of the 20 amino acids that take part in protein synthesis. But as many as 60 different tRNA molecules are found in the cell. It means some amino acids have more than one tRNA and only one aminoacyl tRNA synthetase for all tRNAs of that amino acid. It means:

1. The same aminoacyl tRNA synthetase enzyme can recognise all the tRNAs for the same amino acid, in case there are more than one tRNAs for one amino acid.
2. Different or more than one synthetases are present in the cell for each of those amino acids which are specified by more than one codon.

The specificity of ***aminoacyl tRNA synthetases*** ensures the joining of correct amino acid with each tRNA with a high degree of accuracy. However, wrong pairing of amino acid and tRNA does occur because each of these enzymes have to exercise two kinds of specificity, *i.e.* an ***aminoacyl tRNA synthetase*** has to recognise a proper tRNA and a proper amino acid. Since all tRNA molecules have a very similar primary, secondary and tertiary structures, pairing of enzyme with specific tRNA does go wrong sometimes.

The linkage of amino acid with tRNA is through an ester bond and needs energy from ATP.

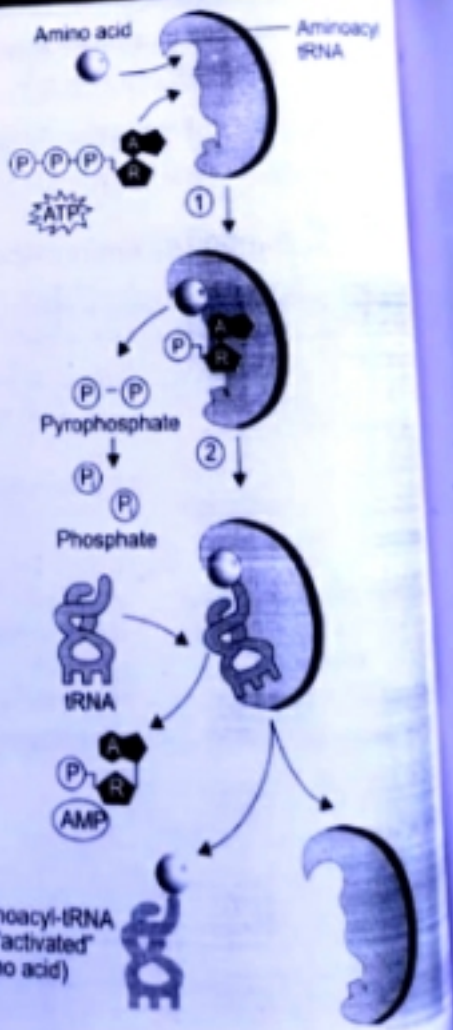


FIGURE 13.6

Process of amino acid activation by enzyme aminoacyl tRNA-synthetase and its attachment with tRNA

2. Attachment of Activated Amino Acid with tRNA or Formation of Aminoacyl-tRNA (Charging of tRNA)

The enzyme bound activated amino acids, **aminoacyl adenylates (AAA)** become attached to the 3′ end of their respective tRNA molecules. The attachment is catalysed by the same enzymes, ***aminoacyl transfer RNA synthetases***, that catalyse activation of their amino acid. The product thus formed is known as **aminoacyl transfer RNA complex** (aminoacyl tRNA). The reaction is as under:

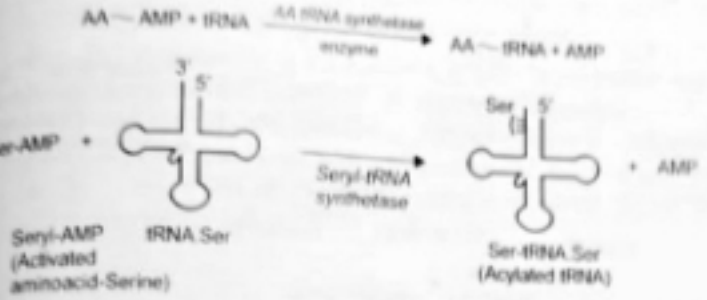


FIGURE 13.7
Reaction that occurs in the binding of activated amino acid serine with tRNA with the help of enzyme aminoacyl tRNA synthetase

This is important to note that particular varieties of amino acids join specific tRNA molecules. This means that for 20 amino acids at least, there are 20 different tRNA molecules and 20 different enzymes. The tRNA charged with its cognate amino acid serves as an **adaptor molecule** for decoding the information on mRNA. Therefore, tRNA is also called **adaptor RNA**. The tRNA with attached amino acid is said to be **acylated** or **charged tRNA.** The tRNA molecule without an amino acid is **uncharged tRNA** while with an incorrect amino acid, it is called **mischarged tRNA.** There are no tRNA for termination codons. Hence, synthesis of polypeptide chain terminates at a codon for which there is no tRNA.

3. Stages during Translation

Process of translation can be separated into the following steps: (1) Initiation (2) Elongation, and (3) Termination.

Though the process of translation includes the same three steps in both prokaryotes and eukaryotes, these are more complex in eukaryotic cells. Therefore, translation in prokaryotes and eukaryotes are discussed separately.

6 Translation in Prokaryotes

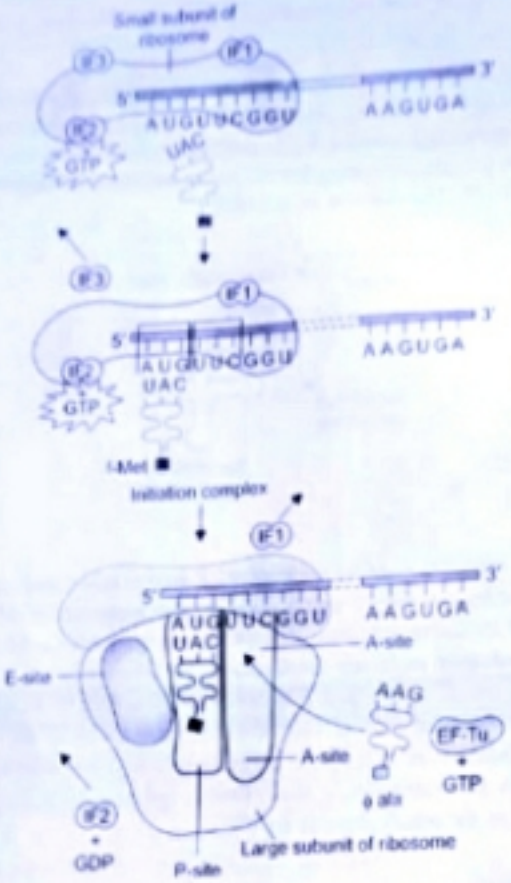
1. Initiation of Polypeptide Chain in Prokaryotes (Formation of Initiation Complex)

In prokaryotes initiation involves following four steps:

1. Three **initiation factors (IFs)**, called **IF1, IF2** and **IF3** bind to the small subunit of ribosome with GTP attached to IF2.
2. **Initiator tRNA** with its amino acid and mRNA binds to the ribosomal subunit. In *E. coli* and other bacteria the initiator tRNA is called **formylatable tRNA**, and is represented as $\mathbf{tRNA^{f\text{-}Met}}$. It carries **N-formyl-methionine (f-Met),** which is a modified methionine with a formyl group on its nitrogen atom. In prokaryotes and in eukaryotic organelles (mitochondria and chloroplasts), each polypeptide chain starts with N-formyl-methionine amino acid (amino acid methionine formylated on its amino group). IF2 with its

FIGURE 13.8

Steps in the formation of initiation complex for the initiation of polypeptide chain in *E. coli*



Role of Initiation Factors

Requirements for Initiation

The initiation of polypeptide chain synthesis requires:

- The two ribosomal subunits 30S and 50S.
- mRNA to be translated.
- Protein initiation factors (IFs).
- GTP.
- The initiator tRNA with initiation amino acid, i.e., formylatable tRNA with N-formylmethionine. The complex is represented as **f-Met-tRNA$^{f\text{-}Met}$**.

Role of Prokaryotic IF1

- IF1 associates with 30S ribosomal subunit in the A site.
- It prevents aminoacyl-tRNA from entering.
- It modulates IF2 binding to the ribosome by increasing its affinity.
- It prevents 50S subunit from binding to 30S and formation of 70S subunit.

Role of Prokaryotic IF2

- It has a ribosome dependent GTPase activity.
- It binds to initiator tRNA and also to P site on 30S subunit.
- It controls the entry of tRNA onto the ribosome.
- It transfers f-Met-tRNA$^{f\text{-}Met}$ to P site of 50S ribosome.
- It causes hydrolysis of GTP releasing energy when 50S subunit joins initiation complex to form complete ribosome.

Role of Prokaryotic IF3

- Stabilises free 30S subunit which is released by the dissociation of 70S ribosome.
- Prevents 30S subunit from reassociating with 50S subunit.
- Enables the initiation complex to bind quickly to mRNA through codon-anticodon pairing.
- Checks the accuracy of recognition of first aminoacyl-tRNA.

GTP helps to identify the location of N-formylmethionine by $tRNA^{fMet}$ in the P-site of 30S subunit of ribosome. It is the only aminoacyl tRNA that can bind to the P-site of small ribosomal subunit.

The initiation factors, GTP, mRNA and 30S subunit of ribosome all collectively constitute **30S preinitiation complex**.

3. Anticodon of $tRNA^{fMet}$ base pairs with the initiation codon, **AUG**, the first codon on the 5′ end of mRNA. AUG is brought into correct position when mRNA binds to 30S ribosomal subunit by its ribosome binding site. This completes the formation of **30S preinitiation complex**.

The ribosomal binding site on mRNA is also called **leader sequence** or **Shine-Dalgarno Sequence** (after its discoverer). It consists of 3–9 purine purine sequence base pairs with a pyrimidine rich complementary sequence on the 3′ end of 16S rRNA of 30S subunit of ribosome. The 3′ end of 16S rRNA was earlier called mRNA binding site.

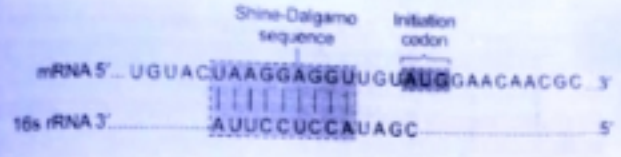


FIGURE 13.9
Base pairing between Shine-Dalgarno sequence or ribosomal binding site (leader sequence) in mRNA and complementary region near 3′ terminus of 16S rRNA

4. **30S preinitiation complex** formed this way joins with 50S ribosomal subunit, generating **70S initiation complex**. The energy needed for the binding of 50S subunit is provided by the hydrolysis of GTP. Mg^{2+} ions are required during this process. This association releases IF2 from 30S subunit. The entire process of polypeptide initiation in prokaryotes can be represented as follows:

- 30S subunit + IF1 + IF2 + IF3 + GTP ⟶ 30S - IF1 - IF2 - IF3 - GTP complex
- 30S – IF1 – IF2 – IF3 – GTP complex + f-Met-$tRNA^{fMet}$ + mRNA ⟶ 30S preinitiation complex + IF1 + IF3
- 30S preinitiation complex + 50S subunit ⟶ 70S initiation complex + IF2 + GDP + P_i

i.e. 30S subunit + f-Met-$tRNA^{fMet}$ + mRNA + 50S subunit + GTP ⟶ 70S initiation complex + GDP + P_i

When 50S subunit of ribosome joins with the 30S preinitiation complex, the f-Met-$tRNA^{fMet}$ occupies its P site. This enables the pairing of f-Met-tRNA anticodon with AUG initiation codon in mRNA. Thus, the reading frame is defined.

The junction of 30S to 50S subunits involves (i) contacts between 16S rRNA and 23S rRNA; (ii) interactions between rRNA of each subunit with proteins in the other and (iii) protein-protein interactions. Some of these interactions are:

- 3′ terminus of 16S rRNA interacts directly with mRNA at initiation by pairing with Shine–Dalgarno sequence.
- Specific regions of 16S rRNA interact directly with the anticodon region of tRNA in both A-site and P-site.
- Subunit interaction involves interactions between 16S and 23S rRNAs.

2. Elongation of Polypeptide Chain

After the formation of 70S mRNA-f-Met-tRNA$^{f\text{Met}}$ complex, the elongation of polypeptide chain begins by the regular addition of amino acids in following steps:

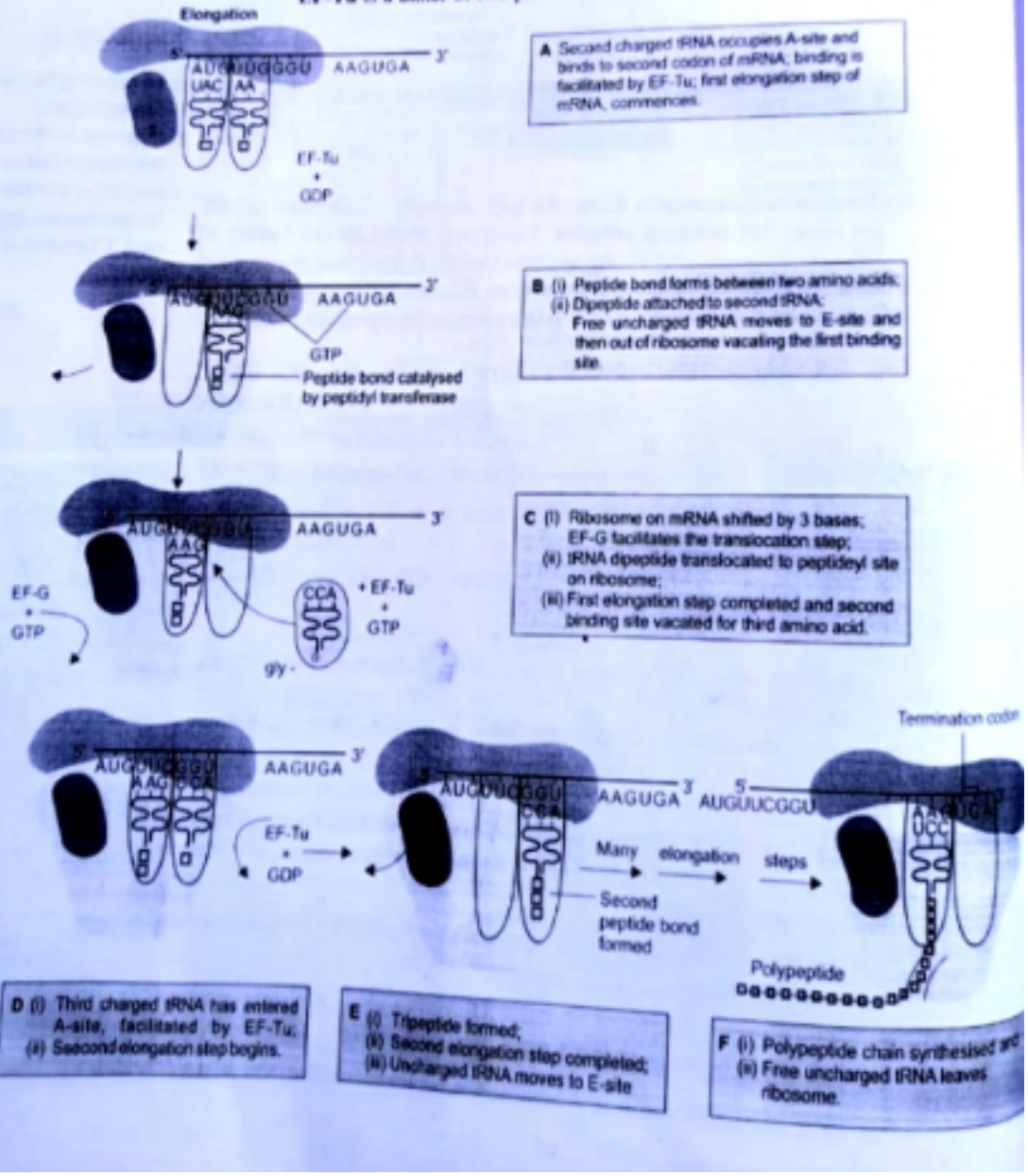
(i) Binding of AA-tRNA at Site-A of larger subunit of ribosome

The large subunit of each ribosome has two slots for the attachment of two molecules of tRNA. These are called **P-site (peptidyl or donor site)** and **A-site (amino acyl or acceptor site)**. The incoming aminoacyl-tRNA complex (AA-tRNA) attaches to the acceptor site and base pairs with mRNA codon present at the ribosome's A-site. In the first round of elongation, the codon in A-site is the one immediately next to the start codon. It receives the second aminoacyl-tRNA. The tRNA carrying peptide chain shifts to the peptidyl or donor site, *i.e.* f-Met-tRNA$^{f\text{Met}}$ attaches to the P-site.

The binding of new aminoacyl tRNA needs two **elongation factors (EFs)**, namely **EF-Tu** and **EF-Ts**. The required energy is provided by the hydrolysis of 2GTP. **EF-Tu** is a dimer of two proteins, each of which binds to one GTP. **EF-Tu** com-

FIGURE 13.10

Formation of peptide bond and elongation of growing polypeptide chain in prokaryotes



aminoacyl tRNA to A-site and is released after the transfer of aminoacyl tRNA. **EF-Ts** regenerates EF-Tu-2GTP for the next aminoacyl tRNA.

The aminoacyl tRNA binding sequence can be summarised as follows:

(1) EF-Tu + 2GTP + AA-tRNA ⟶ EF-Tu-2GTP-AA tRNA

(2) EF-Tu-2GTP-AA-tRNA + 70S complex ⟶ 70S complex-AA tRNA + EF-Tu-2GDP + 2Pi

(3) EF-Tu-2GDP $\xrightarrow{\text{EF-Ts}}$ EF-Tu + 2GTP

70S complex + AA-tRNA + 2GTP ⟶ 70S complex-AA-tRNA + 2GDP + 2Pi

(ii) Formation of peptide bond

With f-Met-tRNA$^{f\text{-}Met}$ now at the **P**-site and second aminoacyl tRNA at **A**-site, a peptide bond is formed between carboxyl group of f-Met amino acid from peptidyl tRNA and amino group (NH_2^+) of amino acid from aminoacyl-tRNA of site A. As a result, two amino acids are now attached to the second tRNA, present at A-site (Fig. 13.10B).

Peptide bond formation is catalysed by ***peptidyl transferase***. It is an enzymatic activity inherent in 23S rRNA of large ribosomal subunit (50S). Thus, 23S rRNA is a **ribozyme**. The energy for peptide bond formation is provided by the hydrolysis of "high energy" ester bond by which f-Met was attached to its tRNA.

(iii) Translocation (Movement of peptidyl-tRNA from A-site to P-site)

After the formation of peptide bond, tRNA$^{f\text{-}Met}$ on **P**-site is without amino acid and tRNA at A-site has a dipeptide. Three movements occur at this point:

- The uncharged tRNA moves to leave **E**-site on ribosome and is finally released in the cytoplasm to start a new polypeptide chain.
- The **dipeptidyl tRNA** (with two amino acids) moves from **A**-site to **P**-site leaving the A-site vacant (Fig. 13.10C). This is called **translocation**. This requires enzyme ***translocase***, and an elongation factor **EF-G** with a bounded **GTP** which attaches to ribosome. The GTP provides energy for translocation as EF-G leaves the ribosome.
- As the peptidyl tRNA translocates, it remains hydrogen-bonded to mRNA and pulls mRNA along with it. As a result, ribosome moves along the length of mRNA in 5′→3′ direction so that the next codon (*i.e.* third codon) on mRNA is available at **A**-site to receive next aminoacyl tRNA having correct anticodon (Fig. 13.10D).

As one ribosome moves along the length of mRNA, the initiation point of mRNA becomes free. It can form an initiation complex with 30S subunit of another ribosome. In this way a number of ribosomes get attached to a single mRNA molecule.

Each successive amino acid is added in this way to polypeptide chain as mRNA is read in 5′→3′ direction. The amino end of the growing polypeptide chain is thought to pass out of the ribosome through a tunnel in the 50S subunit of ribosome. Polypeptide synthesis is very rapid. In *E. coli*, a polypeptide chain of 400 amino acids is formed in 10 seconds.

During the process of protein synthesis a number of ribosomes could be seen attached to a single mRNA molecule, each with a polypeptide chain under formation, the size of polypeptide chains on different ribosomes being different. This complex is known as **polyribosome complex.**

TABLE 13.1: Various factors involved during translation

Process	Factors	Role
Initiation	IF-1	Stabilises 30S subunit.
	IF-2	Binds f-Met-tRNA (the initiator tRNA) to 30S mRNA complex; stimulates GTP hydrolysis
	IF-3	Binds 30S subunit to initiation site of mRNA; dissociates ribosomes (70S) into subunits (30S + 50S) following termination.
Elongation	EF-Tu	Brings aminoacyl-tRNA to the A-site.
	EF-Ts	Generates active EF-Tu.
	EF-G/EF2	Catalyses translocation of polypeptidyl tRNA (GTP-dependent) from site A to P site.
Termination	RF1	Catalyses release of the polypeptide chain from tRNA and dissociation of the translation complexes; specific for UAA and UAG termination codons.
	RF2	Behaves like RF1, specific for UGA and UAA codons.
	RF3	Stimulates RF1 and RF2.

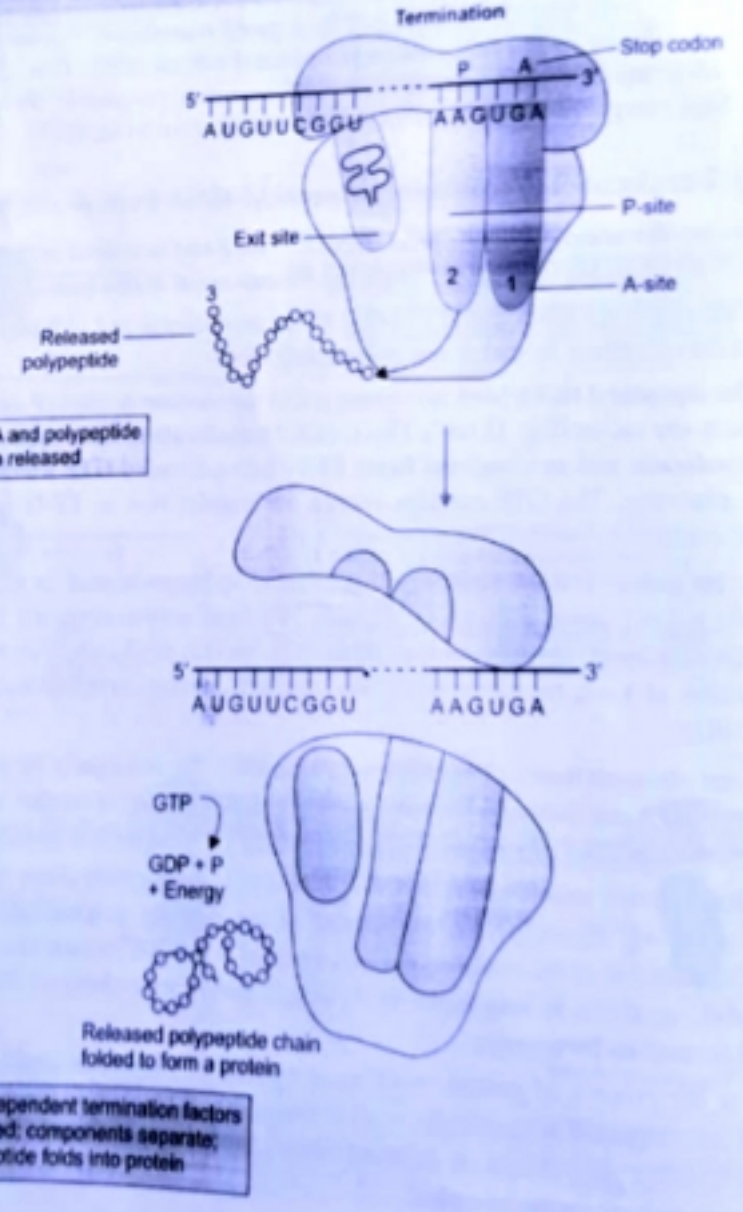


FIGURE 13.11

Diagram showing termination of translation and release of polypeptide chain

3. Termination of Polypeptide Chain

Termination of polypeptide chain occurs when any one of the three termination codons (UAA, UAG or UGA) present on mRNA at the 3' end of each cistron arrives at the ribosome's **A**-site. There are no tRNA molecules that recognise these termination codons. Releasing factors **RF1** and **RF2** are needed for chain termination because these recognise termination codons and terminate translation. **RF1** is specific for UAG and **RF2** is specific for UGA. The releasing factor with the termination codon forms a complex which induces the enzyme ***peptidyl transferase*** to catalyse the termination and release of polypeptide chain (**Lipmann**, 1973). Factor **RF3** stimulates **RF1** and **RF2**. The free ribosome now dissociates into two subunits with the help of dissociation factor IF3.

Polysomes or Polyribosomes

During protein synthesis, when polypeptide chain has grown to 25 amino acid residues, the **AUG** initiation codon of encoding mRNA is completely free of the ribosome to start a new polypeptide chain. Formation of a new polypeptide chain is then initiated. Now two ribosomes are attached to one mRNA molecule. When second ribosome has also moved along mRNA and has a 25 amino acids long polypeptide chain, a third ribosome also attaches to initiation codon of mRNA. This process of movement and reinitiation continues until mRNA is covered with a series of ribosomes aligned at a distance of one ribosome per 80 nucleotides. Such a large translation unit with several ribosomes attached to one mRNA is called **polyribosome** or simply a **polysome.**

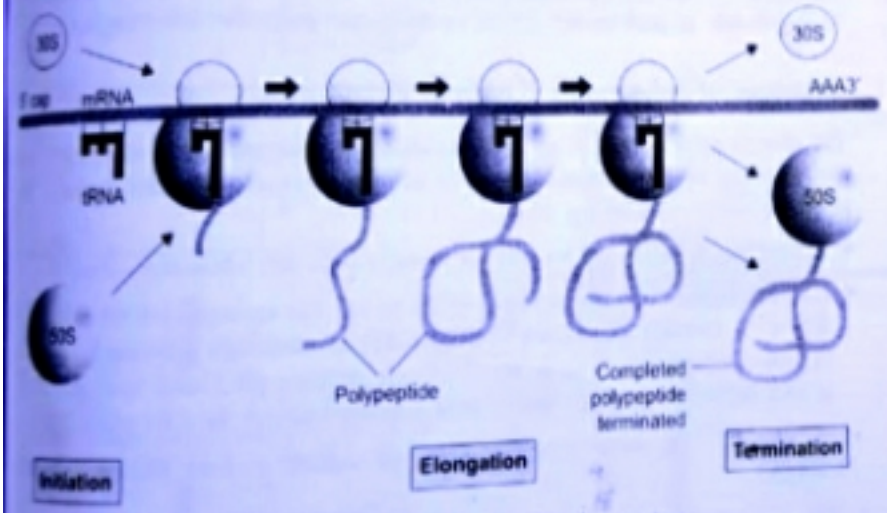


FIGURE 13.12
Polysome showing a mRNA molecule with several ribosomes, each with individual polypeptide chain of different length. At initiation, the two subunits of ribosome are separate and join, and at termination the two subunits of ribosome get disassembled.

Coupled Transcription-Translation in Prokaryotes

During transcription, the mRNA molecule being synthesised has a free 5' terminus. Its ribosome binding site is transcribed first, followed by codon (AUG) that initiates synthesis of polypeptide chain, then region of mRNA associated with elongation of polypeptide chain and finally the stop codon for termination of polypeptide chain. Because, in bacteria no nuclear membrane separates DNA from cytoplasm and its ribosomes, the process of translation or polypeptide synthesis begins even before mRNA synthesis is completed and mRNA is released from DNA. Electron micrographs have been obtained where mRNA is still attached to DNA and process of synthesis of polypeptide chain is seen. This process is called **coupled transcription-translation.** It does not occur in eukaryotes because transcription occurs inside the nucleus and translation in the cytoplasm.

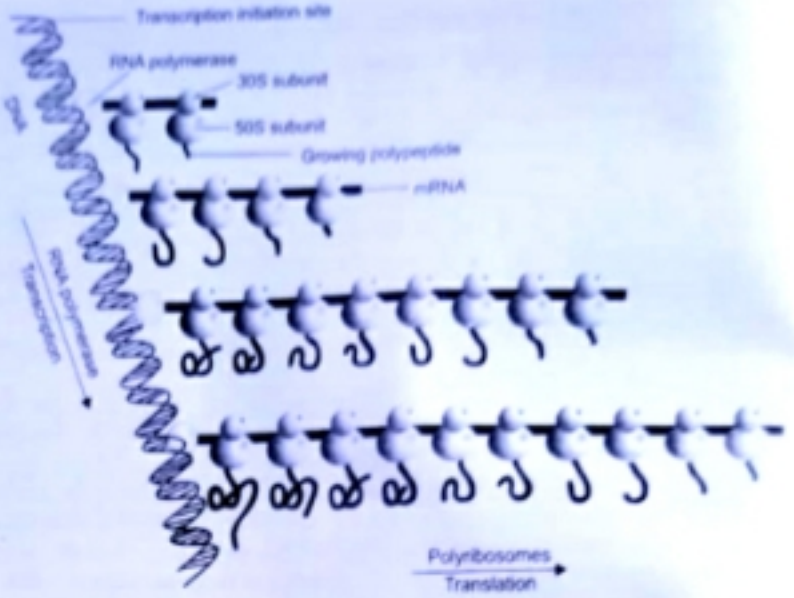


FIGURE 13.13
Coupled transcription and translation in Bacteria

13.7 Translation in Eukaryotes

The process of polypeptide synthesis in eukaryotes follows the same general pattern as in the case of prokaryotes. It can be discussed under the following heads:

1. Initiation of Polypeptide Chain in Eukaryotes

The process of initiation of polypeptide chain in eukaryotes is more complex and needs at least twelve proteins, marked as eIFs (**eukaryotic initiation factors**). The initiation involves following steps:

- The initiation factors eIF1A and eIF3 bind to the 40S ribosomal subunit.
- GTP binds to eIF2 and this complex binds to initiator aminoacyl tRNA (methionyl $tRNA^{Met}$), forming **Met-$tRNA^{Met}$ eIF2-GTP**. In mammals initiation factor eIF2 has three subunits α, β and γ. The α-subunit of factor eIF2 binds to eIF2. γ binds to Met-$tRNA^{Met}$ and eIF2, and subunit β is supposed to be a recycling factor.
- Met-$tRNA^{Met}$-eIF2-GTP associates with 40S subunit to form **40S preinitiation complex.**
- The 40S preinitiation complex (40S-Met $tRNA^{Met}$-eIF2-GTP) binds to 5′ end of mRNA.
- mRNA is brought to the 40S subunit of ribosome by eIF4E.
- The eIF4E recognises and binds to the 5′ cap of mRNA.
- The e1F4G binds to both e1F4E at the 5′ cap and to PABP (the poly A binding protein) at the 3′ poly A tail and also to eIF4A and eIF4B.
- The small subunit (40S) of ribosome with initiation Met-$tRNA^{Met}$ + eIF2α + GTP slides along mRNA to reach initiation codon AUG and gets associated with 60S subunit of ribosome. Factor eIF5 releases eIF2 and eIF3 factors and brings about the association of 60S subunit with the initiation complex. GTP hydrolyses to provide energy for this binding.

13.8 Protein Maturation and Secretion

After reading this section, you should be able to:

- Discuss the role of molecular chaperones in protein folding, and list some important examples of chaperones
- Describe the role of protein splicing in protein maturation
- Distinguish translocation of proteins from protein secretion
- List bacterial translocation systems, and indicate whether they function in Gram-positive, Gram-negative, or both types of bacteria

As a polypeptide emerges from a ribosome, it is not yet ready to assume its cellular functions. Protein function depends on its three-dimensional shape. Proteins must be properly folded and in some cases associated with other protein subunits to generate a functional enzyme (e.g., DNA and RNA polymerases are multimeric proteins). In addition, proteins must be delivered to the proper subcellular or extracellular site. We now discuss these posttranslational events.

Protein Folding and Molecular Chaperones

Although the amino acid sequence of a polypeptide determines its final conformation, helper proteins aid the newly formed or nascent polypeptide in folding to its proper functional shape. These proteins, called **molecular chaperones** or simply chaperones, recognize only unfolded polypeptides or partly denatured proteins and do not bind to normal, functional proteins. Their role is essential because the cytoplasm is filled with new polypeptide chains. Under such conditions, it is possible for polypeptides to fold improperly and aggregate to form nonfunctional complexes. Molecular chaperones suppress incorrect folding and may reverse any incorrect folding that has already taken place. They are so important that chaperones are present in all cells.

Several chaperones and cooperating proteins aid proper protein folding in *E. coli*: chaperones DnaK, DnaJ, GroEL, and GroES; and the stress protein GrpE. After a sufficient length of nascent polypeptide extends from the ribosome, a series of reactions involving DnaJ and DnaK fold the protein into its native conformation. This requires the expenditure of ATP (**figure 13.42**). Sometimes the polypeptide does not reach its native conformation

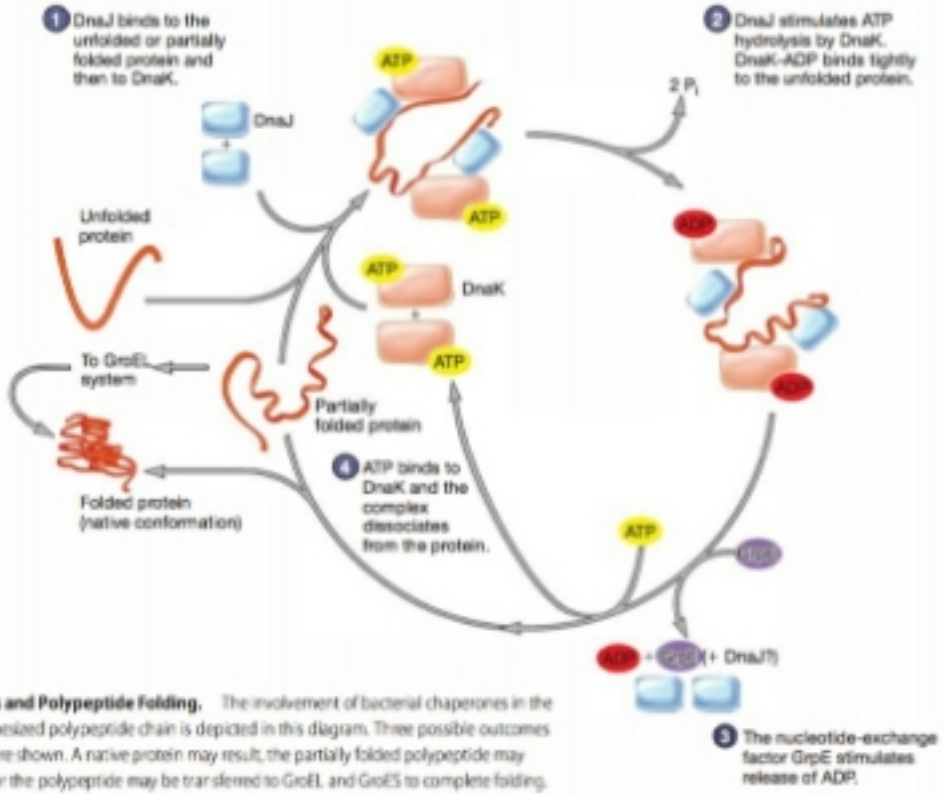


Figure 13.42 Chaperones and Polypeptide Folding. The involvement of bacterial chaperones in the proper folding of a newly synthesized polypeptide chain is depicted in this diagram. Three possible outcomes of a chaperone reaction cycle are shown. A native protein may result, the partially folded polypeptide may bind again to DnaK and DnaJ, or the polypeptide may be transferred to GroEL and GroES to complete folding.

in one series of reactions, and the folding process may be repeated. Alternatively, the partially folded protein may be transferred to chaperones GroEL and GroES, which complete the folding. This chaperone system also expends ATP as it folds the protein into its proper conformation.

Chaperones were first discovered because they dramatically increase in concentration when cells are exposed to high temperatures, metabolic poisons, and other stressful conditions that cause protein denaturation. Thus many chaperones are called **heat shock proteins**. When an *E. coli* culture is switched from 30 to 42°C, the concentrations of some 20 different heat-shock proteins increase greatly within about 5 minutes. If the cells are exposed to a lethal temperature, the heat shock proteins are still synthesized but most other proteins are not. Thus chaperones protect the cell from thermal damage and other stresses as well as promote the proper folding of new polypeptides. For example, DnaK protects *E. coli* RNA polymerase from thermal inactivation in vitro. In addition, DnaK reactivates thermally inactivated RNA polymerase, especially if ATP, DnaJ, and GrpE are present. GroEL and GroES also protect intracellular proteins from aggregation.

Protein Splicing

A further level of complexity in the formation of proteins has been discovered in microbes belonging to all three domains of life. Some microbial proteins are spliced after translation. In protein splicing, a part of the polypeptide is removed before the polypeptide folds into its final shape. Self-splicing proteins begin as larger precursor proteins composed of an internal intervening sequence called an **intein** (about 130 to 600 amino acids in length) flanked by external sequences called **exteins** (figure 13.43). Inteins remove themselves from the precursor protein. When the splicing is completed, two proteins have been formed: the intein protein and the protein formed by splicing the two exteins together.

Protein Translocation and Secretion in Bacteria

It has been estimated that almost one third of the proteins synthesized by cells leave the cytoplasm to reside in membranes, the periplasmic space of bacterial and archaeal cells, or the external environment. It is not surprising then that over 15 different systems for moving proteins out of the cytoplasm have evolved. Some of these systems are found in all domains of life. Others are unique to bacterial cells, and others are observed only in Gram-negative bacteria. When proteins are moved from the cytoplasm to the membrane or to the periplasmic space, the movement is called translocation. Protein secretion refers to the movement of proteins from the cytoplasm to the external environment. Many of the secretion pathways are designated with numbers (e.g., type I secretion system, type II secretion system, etc.).

Why are so many proteins moved out of the cytoplasm? Many important proteins are located in membranes. These include transport proteins that bring needed materials into the cell and take wastes out of the cell. They also include proteins

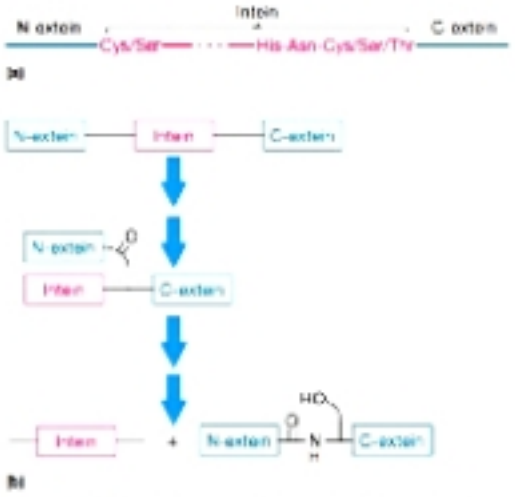


Figure 13.43 Protein Splicing. (a) A generalized illustration of intein structure. The amino acids that are commonly conserved (enhanced) in inteins are shown. Note that many are rich in hydroxyl-containing amino acids. (b) A simplified overview of splicing.

involved in electron transport. In Gram-negative bacteria, the periplasmic space is loaded with proteins such as chemotaxis proteins, enzymes involved in cell wall synthesis, and periplasmic components of nutrient uptake systems. Many organisms secrete hydrolytic enzymes into the external environment. These enzymes break down macromolecules into monomers that are more easily brought into the cell. The protein subunits of external structures such as flagella and fimbriae must also be moved out of the cell and assembled on its external surface. Pathogenic microbes often release toxins that are important in the infection process.

Protein secretion poses different difficulties, depending on the structure of the cell envelope. For Gram-positive bacteria to secrete proteins, the proteins must be translocated across the plasma membrane. Once across the plasma membrane, the protein either passes through the relatively porous peptidoglycan into the external environment or becomes embedded in or attached to the peptidoglycan. Gram-negative bacteria have more hurdles to jump when they secrete proteins. They, too, must transport the proteins across the plasma membrane, but to complete secretion, the proteins must be transported across the outer membrane.

Common Translocation and Secretion Systems

The major pathway for translocating proteins across the plasma membrane is the Sec (secretion) pathway (figure 13.44). In Gram-negative bacteria, proteins can be transported across the outer membrane by several different mechanisms, some of which bypass the Sec system, moving proteins directly from the cytoplasm

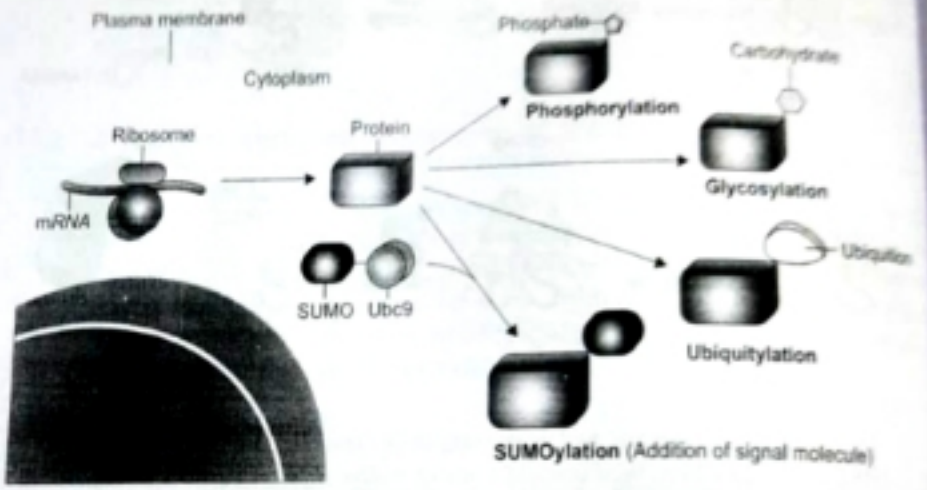
2. **Biochemical Modifications**

The biological changes in proteins to achieve their native functional form are:

1. Proteolytic Cleavage: The amino acid '**N-formyl-methionine**' in prokaryotes and methionine in eukaryotes attached to the N-terminus of polypeptide chain are removed by hydrolysis. This is called **proteolytic cleavage** and is an essential step. In approximately 50 percent of eukaryotic proteins an acetyl group is added to their amino end.
2. Amino Acid Modification: The activity of enzymes and some other proteins is altered by **phosphorylation, methylation** or **hydroxylation** of some amino acids. For example, hydroxyl groups (OH) of serine, threonine and tyrosine are modified by phosphoryl group from ATP. The hydroxylation of proline and lysine changes the protein to collagen (a structural protein).
3. Attachment of Carbohydrates (Glycosylation): In glycoproteins, carbohydrate molecules are covalently bonded to the proteins. These provide immunological protection, cell-cell recognition and blood clotting.

FIGURE 13.22

Biochemical modification of protein by phosphorylation, glycoslyation or methylation of amino acids



4. Addition of Prosthetic Groups: Many enzymes can function in association with a covalently bound **cofactor** or **prosthetic group**. The prosthetic groups are heme, FAD, biotin and pantothenic acid.
5. Trimming of Polypeptide Chain: In some cases long polypeptide chains are trimmed by enzyme ***exopeptidase.*** The amino acid residues are removed one by one either from N- or C-terminus. For example, long polypeptide chain of preinsulin is cut to a small functional insulin of only 51 amino acids.
6. Signal Molecules: These are added or removed from the N-terminal end of some proteins (Sumoylation). For transportation to their specific destination proteins have a sequence of up to 30 amino acids at their N-terminal. This is called **signal sequence.** Once a protein has reached its destination the signal sequence is removed before it assumes the functional status.
7. Polypeptide Chains: Group are also complexed with metal ions or prosthetic group. The tertiary and quarternary level of protein structure is attained only when some metal ion is complexed with it. Iron in case of haemoglobin and Mg in case of chlorophyll is essential for their quarternary structure.