Chapter 11. Synthesis and Processing of the Proteome

Learning outcomes

When you have read Chapter 11, you should be able to:

1. Draw the general structure of a transfer RNA (tRNA) and explain how this structure enables the tRNA to play both a physical and an informational role during protein synthesis
2. Describe how an amino acid becomes attached to a tRNA and outline the processes that ensure that combinations are formed between the correct pairs of amino acids and tRNAs
3. Explain how codons and anticodons interact, and discuss the influence of wobble on this interaction
4. Outline the techniques that have been used to study the structure of the ribosome, and summarize the information that has resulted from these studies
5. Give a detailed description of the process of translation in bacteria and eukaryotes, with emphasis on the roles of the various translation factors, this description including an explanation of how translation is regulated and an outline of the unusual events, such as frameshifting, that can occur during the elongation phase
6. Explain why post-translational processing of proteins is an important component of the genome expression pathway, and describe the key features of protein folding, protein processing by proteolytic cleavage and chemical modification, and intein splicing
7. Describe the major processes responsible for protein degradation in bacteria and eukaryotes

THE END RESULT of genome expression is the proteome, the collection of functioning proteins synthesized by a living cell. The identity and relative abundance of the individual proteins in a proteome represents a balance between the synthesis of new proteins and the degradation of existing ones. The biochemical capabilities of the proteome can also be changed by chemical modification and other processing events. The combination of synthesis, degradation, and modification/processing enables the proteome to meet the changing requirements of the cell and to respond to external stimuli.

In this chapter we will study the synthesis, processing and degradation of the components of the proteome. To understand protein synthesis we will first examine the role of tRNAs in decoding the genetic code and then investigate the events, occurring at the ribosome, that result in polymerization of amino acids into polypeptides. The ribosomal events are sometimes looked upon as the final stage in expression of an individual gene but the polypeptide that is initially synthesized is inactive until it has been folded, and may also have to undergo cutting and chemical modification before it becomes functional. We will study these processing events in Section 11.3. At the end of the chapter we will investigate how the cell degrades proteins that it no longer requires.

11.1. The Role of tRNA in Protein Synthesis

Transfer RNAs play the central role in translation. They are the adaptor molecules, whose existence was predicted by Francis Crick in 1956 (Crick, 1990), which form the link between the mRNA and the polypeptide that is being synthesized. This is both a *physical* link, tRNAs binding to both the mRNA and the growing polypeptide, and an *informational* link, tRNAs ensuring that the polypeptide being synthesized has the amino acid sequence that is denoted, via the genetic code,

by the sequence of nucleotides in the mRNA (*Figure 11.1*). To understand how tRNAs play this dual role we must examine aminoacylation, the process by which the correct amino acid is attached to each tRNA, and codon-anticodon recognition, the interaction between tRNA and mRNA.

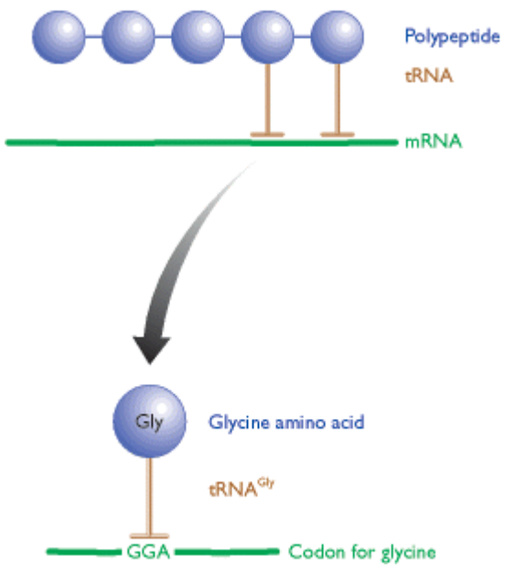

Figure 11.1. **The adaptor role of tRNA in translation**. The top drawing shows the physical role of tRNA, forming an attachment between the polypeptide and the mRNA. The lower drawing shows the informational link, the tRNA carrying the amino acid specified by the codon to which it attaches

11.1.1. Aminoacylation: the attachment of amino acids to tRNAs

Bacteria contain 30–45 different tRNAs and eukaryotes have up to 50. As only 20 amino acids are designated by the genetic code, this means that all organisms have at least some isoaccepting tRNAs, different tRNAs that are specific for the same amino acid. The terminology used when describing tRNAs is to indicate the amino acid specificity with a superscript suffix, using the numbers 1, 2, etc., to distinguish different isoacceptors: for example, two tRNAs specific for glycine would be written as $tRNA^{Gly1}$ and $tRNA^{Gly2}$.

All tRNAs have a similar structure

The smallest tRNAs are only 74 nucleotides in length, and the largest are rarely more than 90 nucleotides. Because of their small size, and because it is possible to purify individual tRNAs, they were among the first nucleic acids to be sequenced, way back in 1965 by Robert Holley's group at Cornell University, New York. The sequences revealed one unexpected feature, that as well as the standard RNA nucleotides (A, C, G and U), tRNAs contain a number of modified nucleotides, 5–10 in any particular tRNA, with over 50 different modifications known altogether (Section 10.3).

Examination of the first tRNA sequence, for $tRNA^{Ala}$ of *Saccharomyces cerevisiae*, showed that the molecule could adopt various base-paired secondary structures. After more tRNAs had been sequenced, it became clear that one particular structure could be taken up by all of them. This is the cloverleaf (*Figure 11.2*), and has the following features:

- The acceptor arm is formed by seven base pairs between the 5′ and 3′ ends of the molecule. The amino acid is attached to the extreme 3′ end of the tRNA, to the adenosine of the invariant CCA terminal sequence (Section 10.2.2).
- The D arm, named after the modified nucleoside dihydrouridine (see *Table 10.5*), which is always present in this structure.
- The anticodon arm contains the triplet of nucleotides called the anticodon which base-pair with the mRNA during translation.
- The V loop contains 3–5 nucleotides in Class 1 tRNAs or 13–21 nucleotides in Class 2 tRNAs.

- The **TΨC arm**, named after the sequence thymidine-pseudouridine-cytosine, which is always present.

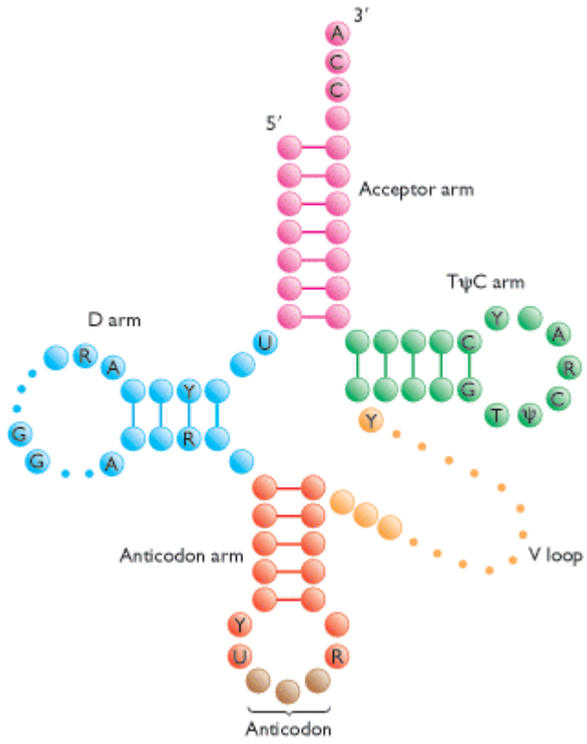

Figure 11.2. **The cloverleaf structure of a tRNA**. The tRNA is drawn in the conventional cloverleaf structure, with the different components labeled. Invariant nucleotides (A, C, G, T, U, Y, where Y = pseudouridine) and semi-invariant nucleotides (abbreviations: R, purine; Y, pyrimidine) are indicated. Optional nucleotides not present in all tRNAs are shown as smaller dots. The standard numbering system places position 1 at the 5′ end and position 76 at the 3′ end; it includes some but not all of the optional nucleotides. The invariant and semi-invariant nucleotides are at positions 8, 11, 14, 15, 18, 19, 21, 24, 32, 33, 37, 48, 53, 54, 55, 56, 57, 58, 60, 61, 74, 75 and 76. The nucleotides of the anticodon are at positions 34, 35 and 36.

The cloverleaf structure can be formed by virtually all tRNAs, the main exceptions being the tRNAs used in vertebrate mitochondria, which are coded by the mitochondrial genome and which sometimes lack parts of the structure. An example is the human mitochondrial $tRNA^{Ser}$, which has no D arm. As well as the conserved secondary structure, the identities of nucleotides at some positions are completely invariant (always the same nucleotide) or semi-invariant (always a purine or always a pyrimidine), and the positions of the modified nucleotides are almost always the same. Many of the invariant nucleotide positions are important in the tertiary structure of tRNA. X-ray crystallography studies have shown that nucleotides in the D and TΨC loops form base pairs that fold the tRNA into a compact L-shaped structure (*Figure 11.3* ; Clark, 2001). Each arm of the L-shape is approximately 7 nm long and 2 nm in diameter, with the amino acid binding site at the end of one arm and the anticodon at the end of the other. The additional base-pairing means that the base-stacking (see page 14) is almost continuous from one end of the tRNA to the other, providing stability to the structure.

Figure 11.3. **The three-dimensional structure of a tRNA**. Additional base pairs, shown in black and mainly between the D and TYC loops, fold the cloverleaf structure shown in *Figure 11.2* into this L-shaped configuration. Depending on its sequence, the V loop might also form interactions with the D arm, as indicated by thin black lines. The color scheme is the same as in *Figure 11.2* . From Freifelder D, *Molecular Biology*, 2nd edition, 1986, Jones and Bartlett Publishers, Sudbury, MA. Reprinted with permission.

Aminoacyl-tRNA synthetases attach amino acids to tRNAs

The attachment of amino acids to tRNAs - 'charging' in molecular biology jargon - is the function of the group of enzymes called aminoacyl-tRNA synthetases. The chemical reaction that results in aminoacylation occurs in two steps. An activated amino acid intermediate is first formed by reaction between the amino acid and ATP, and then the amino acid is transferred to the 3′ end of the tRNA, the link being formed between the -COOH group of the amino acid and the -OH group attached to either the 2′ or 3′ carbon on the sugar of the last nucleotide, which is always an A (*Figure 11.4*).

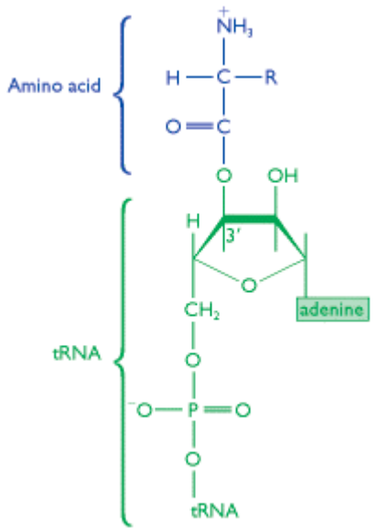

Figure 11.4. **Aminoacylation of a tRNA**. The result of aminoacylation by a Class II aminoacyl-tRNA synthetase is shown, the amino acid being attached via its -COOH group to the 3′-OH of the terminal nucleotide of the tRNA. A Class I aminoacyl-tRNA synthetase attaches the amino acid to the 2′-OH group

With a few exceptions, organisms have 20 aminoacyl-tRNA synthetases, one for each amino acid. This means that groups of isoaccepting tRNAs are aminoacylated by a single enzyme. Although the basic chemical reaction is the same for each amino acid, the 20 aminoacyl-tRNA synthetases fall into two distinct groups, Class I and Class II, with several important differences between them (*Table 11.1*). In particular, Class I enzymes attach the amino acid to the 2′-OH group of the terminal nucleotide of the tRNA, whereas Class II enzymes attach the amino acid to the 3′-OH group (Ibba *et al.*, 2000).

Table 11.1. Features of aminoacyl-tRNA synthetases

Feature	Class I enzymes	Class II enzymes
Structure of the enzyme active site	Parallel β-sheet	Antiparallel β-sheet
Interaction with the tRNA	Minor groove of the acceptor stem	Major groove of the acceptor stem
Orientation of the bound tRNA	V loop faces away from the enzyme	V loop faces the enzyme
Amino acid attachment	To the 2'-OH of the terminal nucleotide of the tRNA	To the 3'-OH of the terminal nucleotide of the tRNA
Enzymes for *	Arg, Cys, Gln, Glu, Ile, Leu, Lys1, Met, Trp, Tyr, Val	Ala, Asn, Asp, Gly, His, LysII, Phe, Pro, Thr, Ser
* The aminoacyl-tRNA synthetase for lysine is a Class I enzyme in some archaea and bacteria and a Class II enzyme in all other organisms. For more details see Arnez and Moras (1997) and Ibba *et al.* (2000).		

Aminoacylation must be carried out accurately: the correct amino acid must be attached to the correct tRNA if the rules of the genetic code are to be followed during protein synthesis. It appears that an aminoacyl-tRNA synthetase has high fidelity for its tRNA, the result of an extensive interaction between the two, covering some 25 nm^2 of surface area and involving the acceptor arm and anticodon loop of the tRNA, as well as individual nucleotides in the D and TΨC arms. The interaction between enzyme and amino acid is, of necessity, less extensive, amino acids being much smaller than tRNAs, and presents greater problems with regard to specificity because

several pairs of amino acids are structurally similar. Errors do therefore occur, at a very low rate for most amino acids but possibly as frequently as one aminoacylation in 80 for difficult pairs such as isoleucine and valine. Most errors are corrected by the aminoacyl-tRNA synthetase itself, by an editing process that is distinct from aminoacylation, involving different contacts with the tRNA (Hale *et al.*, 1997; Silvian *et al.*, 1999).

In most organisms, aminoacylation is carried out by the process just described, but a few unusual events have been documented. These include a number of instances where the aminoacyl-tRNA synthetase attaches the incorrect amino acid to a tRNA, this amino acid subsequently being transformed into the correct one by a second, separate chemical reaction. This was first discovered in the bacterium *Bacillus megaterium* for synthesis of glutamine-tRNA^{Gln} (i.e. glutamine attached to its tRNA). This aminoacylation is carried out by the enzyme responsible for synthesis of glutamic acid-tRNA^{Glu}, and initially results in attachment of a glutamic acid to the tRNA^{Gln} (*Figure 11.5A*). This glutamic acid is then converted to glutamine by transamidation catalyzed by a second enzyme. The same process is used by various other bacteria (although not *Escherichia coli*) and by the archaea. Some archaea also use transamidation to synthesize asparagine-tRNA^{Asn} from aspartic acid-tRNA^{Asn} (Ibba *et al.*, 2000). In both of these cases, the amino acid that is synthesized by the modification process is one of the 20 that are specified by the genetic code. There are also two examples where the modification results in an unusual amino acid. The first example is the conversion of methionine to *N*-formylmethionine (*Figure 11.5B*), producing the special aminoacyl-tRNA used in initiation of bacterial translation (Section 11.2.2). The second example occurs in both prokaryotes and eukaryotes and results in synthesis of selenocysteine, which is specified in a context-dependent manner by some 5′-UGA-3′ codons (Section 3.3.2). These codons are recognized by a special tRNA^{SeCys}, but there is no aminoacyl-tRNA synthetase that is able to attach selenocysteine to this tRNA. Instead, the tRNA is aminoacylated with a serine by the seryl-tRNA synthetase, and then modified by replacement of the -OH group of the serine with an -SeH, to give selenocysteine (*Figure 11.5C* ; Low and Berry, 1996).

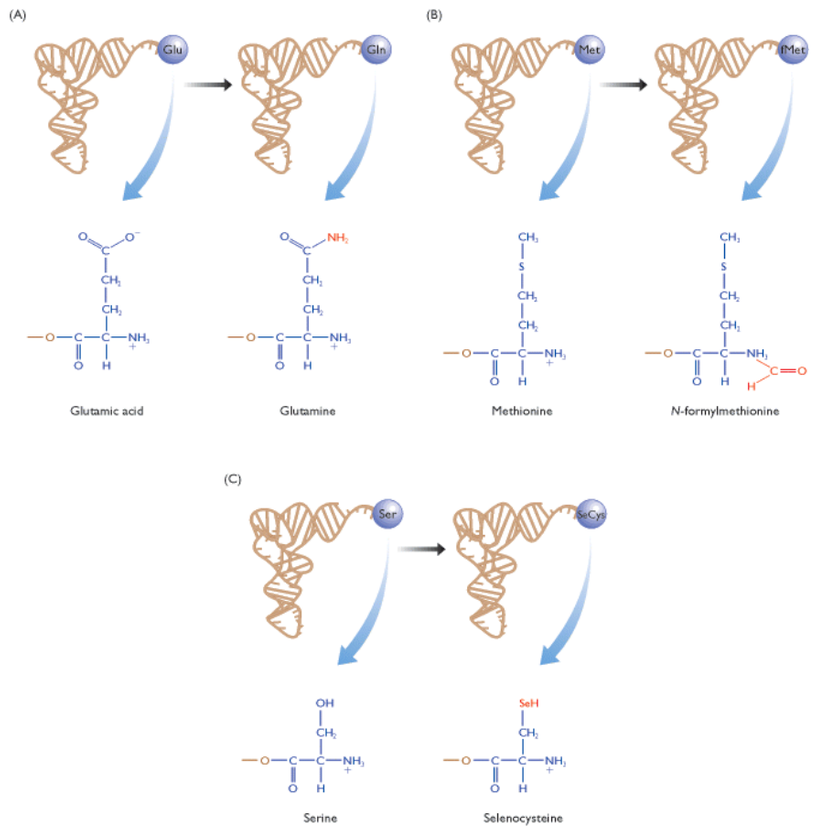

Figure 11.5. Unusual types of aminoacylation. (A) In some bacteria, tRNA^{Gln} is aminoacylated with glutamic acid, which is then converted to glutamine by transamidation. (B) The special tRNA used in initiation of translation in bacteria is aminoacylated with methionine, which is then converted to *N*-formylmethionine. (C) tRNA^{SeCys} in various organisms is initially aminoacylated with serine

11.1.2. Codon-anticodon interactions: the attachment of tRNAs to mRNA

Aminoacylation represents the first level of specificity displayed by a tRNA. The second level is the specificity of the interaction between the anticodon of the tRNA and the mRNA being translated. This specificity ensures that protein synthesis follows the rules of the genetic code (see *Figure 3.20*).

UUU, UUC: phe UUA, UUG: leu	UCU, UCC, UCA, UCG: ser	UAU, UAC: tyr UAA, UAG: stop	UGU, UGC: cys UGA: stop UGG: trp
CUU, CUC, CUA, CUG: leu	CCU, CCC, CCA, CCG: pro	CAU, CAC: his CAA, CAG: gln	CGU, CGC, CGA, CGG: arg
AUU, AUC, AUA: ile AUG: met	ACU, ACC, ACA, ACG: thr	AAU, AAC: asn AAA, AAG: lys	AGU, AGC: ser AGA, AGG: arg
GUU, GUC, GUA, GUG: val	GCU, GCC, GCA, GCG: ala	GAU, GAC: asp GAA, GAG: glu	GGU, GGC, GGA, GGG: gly

Figure 3.20. **The genetic code**. See *Table 3.1* for the three-letter abbreviations of the amino acids.

In principle, codon-anticodon recognition is a straightforward process involving base-pairing between the anticodon of the tRNA and a codon in the mRNA (*Figure 11.6*). The specificity of aminoacylation ensures that the tRNA carries the amino acid denoted by the codon that it pairs with, and the ribosome controls the topology of the interaction in such a way that only a single triplet of nucleotides is available for pairing. Because base-paired polynucleotides are always antiparallel, and because the mRNA is read in the 5′→3′ direction, the first nucleotide of the codon pairs with nucleotide 36 of the tRNA, the second with nucleotide 35, and the third with nucleotide 34.

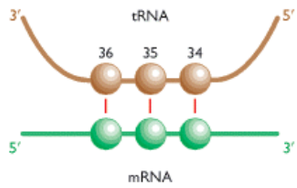

Figure 11.6. **The interaction between a codon and an anticodon**. The numbers indicate the nucleotide positions in the tRNA (see *Figure 11.2*)

In practice, codon recognition is complicated by the possibility of **wobble**. This is another of the principles of gene expression originally proposed by Crick and subsequently shown to be correct. Because the anticodon is in a loop of RNA, the triplet of nucleotides is slightly curved (see *Figures 11.2* and *11.3*) and so cannot make an entirely uniform alignment with the codon. As a result, a non-standard base pair can form between the third nucleotide of the codon and the first nucleotide (number 34) of the anticodon. This is called ‘wobble'. A variety of pairings is possible, especially if the nucleotide at position 34 is modified. In bacteria, the two main features of wobble are (Ikemura, 1981):

- ***G-U base-pairs*** are permitted. This means that an anticodon with the sequence 3′-xxG-5′ can base-pair with both 5′-xxC-3′ and 5′-xxU-3′. Similarly, the anticodon 3′-xxU-5′ can base-pair with both 5′-xxA-3′ and 5′-xxG-3′. The consequence is that, rather than needing a different tRNA for each codon, the four members of a codon family (e.g. 5′-GCN-3′, all coding for alanine) can be decoded by just two tRNAs (*Figure 11.7A*).
- *Inosine* , abbreviated to I, is a modified purine (see *Table 10.5*) that can base-pair with A, C and U. Inosine can only occur in the tRNA because the mRNA is not modified in this way. The triplet 3′-UAI-5′ is sometimes used as the anticodon in a $tRNA^{Ile}$ molecule because it pairs with 5′-AUA-3′, 5′-AUC-3′ and 5′-AUU-3′ (*Figure 11.7B*), which form the three-codon family for this amino acid in the standard genetic code.

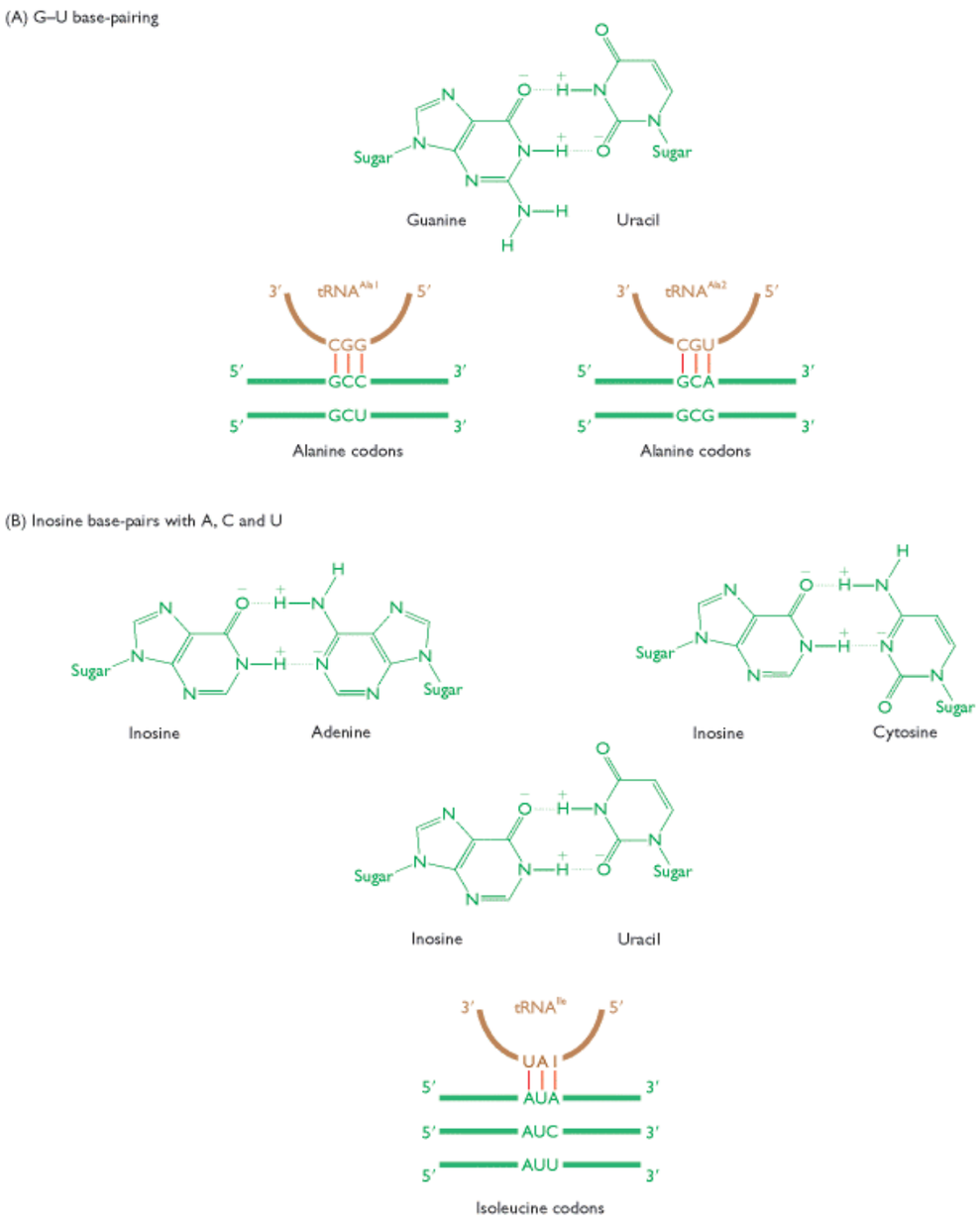

Figure 11.7. **Two examples of wobble in bacteria**. (A) Wobble involving a G-U base pair enables the four-codon family for alanine to be decoded by just two tRNAs. Note that wobble involving G-U also enables accurate decoding of a four-codon family that specifies two amino acids. For example, the anticodon 3′-AAG-5′ can decode 5′-UUC-3′ and 5′-UUU-3′, both coding for phenylalanine (see *Figure 3.20*), and the anticodon 3′-AAU-5′ can decode the other two members of this family, 5′-UUA-3′ and 5′-UUG-3′, which code for leucine. (B) Inosine can base-pair with A, C or U, meaning that a single tRNA can decode all three codons for isoleucine. Dotted lines indicate hydrogen bonds; I, inosine.

Wobble reduces the number of tRNAs needed in a cell by enabling one tRNA to read two or possibly three codons. Hence bacteria can decode their mRNAs with as few as 30 tRNAs. Eukaryotes also make use of wobble but in a restricted way. The human genome, which in this regard is fairly typical of higher eukaryotes, has 48 tRNAs. Of these, 16 are predicted to use wobble to decode two codons each, with the remaining 32 being specific for just a single triplet (*Figure 11.8* ; IHGSC, 2001). The distinctive features compared with wobble in bacteria are:

- G-U wobble is used with eight tRNAs but in every case the wobble involves an anticodon with the sequence 3′-xxG-5′. The alternative version of G-U wobble, where the anticodon sequence is 3′-xxU-5′, appears not to be used in eukaryotes, possibly because this could result in a $tRNA^{ile}$ with the anticodon 3′-UAU-5′ reading the methionine codon 5′-AUG-3′ (

Figure 11.9). Eukaryotes may therefore have a means of preventing this type of wobble from occurring (Percudani, 2001).

- Eight other human tRNAs have anticodons containing inosine (3′-xxI-5′) but these decode only 5′-xxC-3′ and 5′-xxU-3′. The base pairing between I and A is weak, which means that 5′-xxA-3′ codons are only inefficiently recognized by an 3′-xxI-5′ anticodon. To avoid this inefficiency, in every example of wobble involving inosine in the human tRNA set, the 5′-xxA-3′ codon is recognized by a separate tRNA. Note, however, that recognition by a separate tRNA does not preclude the 5′-xxA-3′ codon from also being decoded by the tRNA containing 3′-xxI-5′, albeit inefficiently. This does not compromise the specificity of the genetic code, because wobble involving inosine is limited to those codon families in which all four triplets specify the same amino acid (see *Figure 11.8*).

Other genetic systems use more extreme forms of wobble. Human mitochondria, for example, use only 22 tRNAs. With some of these tRNAs the nucleotide in the wobble position of the anticodon is virtually redundant because it can base-pair with any nucleotide, enabling all four codons of a family to be recognized by the same tRNA. This phenomenon has been called superwobble.

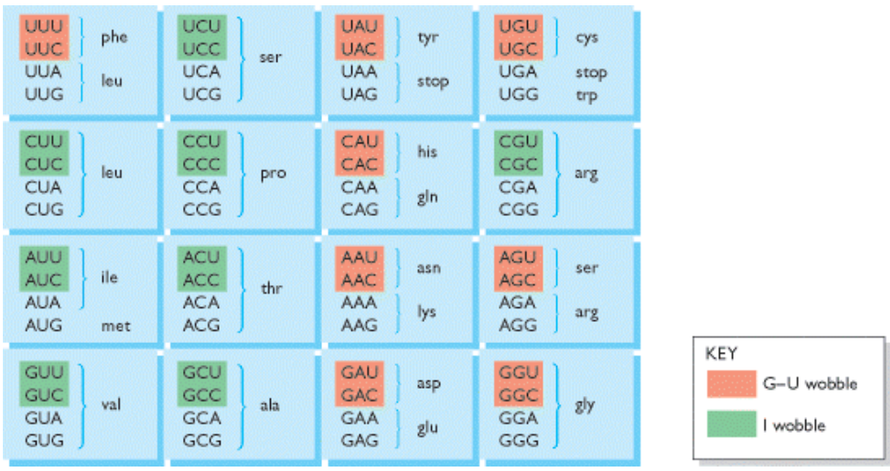

Figure 11.8. **The predicted usage of wobble in decoding the human genome**. Pairs of codons that are predicted to be decoded by a single tRNA using G-U wobble are highlighted in red, and those pairs predicted to be decoded by wobble involving inosine are highlighted in green. Codons that are not highlighted have their own individual tRNAs. The predictions are based largely on examination of the anticodon sequences of the tRNAs that have been located in the draft human genome sequence. The analysis shown here implies that there are 45 tRNAs in human cells - the 16 for the wobble pairs and 29 singletons. In fact there are 48 tRNAs. This is because three codons thought to be decoded as part of a wobble pair (5′-AAU-3′, 5′-AUC-3′ and 5′-UAU-3′) also have their own individual tRNAs, although these are present in low abundance.

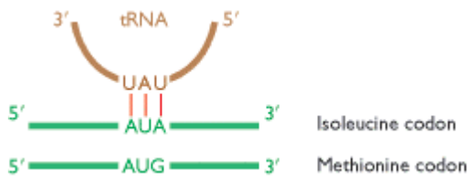

Figure 11.9. **A tRNA with the anticodon 3′-UAU-5′ could read the isoleucine codon 5′-AUA-3′ as well as the methionine codon**

11.2. The Role of the Ribosome in Protein Synthesis

An *E. coli* cell contains approximately 20 000 ribosomes, distributed throughout its cytoplasm. The average human cell contains rather more (nobody has ever counted them all), some free in the cytoplasm and some attached to the outer surface of the endoplasmic reticulum, the membranous network of tubes and vesicles that permeates the cell. Originally, ribosomes were looked on as passive partners in protein synthesis, merely the structures on which translation occurs. This view has changed over the years and ribosomes are now considered to play two active roles in protein synthesis:

- Ribosomes *coordinate* protein synthesis by placing the mRNA, aminoacyl-tRNAs and associated protein factors in their correct positions relative to one another.
- Components of ribosomes, including the rRNAs, *catalyze* at least some of the chemical reactions occurring during translation.

To understand how ribosomes play these roles we will first survey the structural features of ribosomes in bacteria and eukaryotes, and then examine the detailed mechanism for protein synthesis in these two types of organism.

11.2.1. Ribosome structure

Our understanding of ribosome structure has gradually developed over the last 50 years as more and more powerful techniques have been applied to the problem. Originally called ‘microsomes', ribosomes were first observed in the early decades of the 20th century as tiny particles almost beyond the resolving power of light microscopy. In the 1940s and 1950s, the first electron micrographs showed that bacterial ribosomes are oval-shaped, with dimensions of 29 nm × 21 nm, rather smaller than eukaryotic ribosomes, the latter varying a little in size depending on species but averaging about 32 nm × 22 nm. In the mid-1950s, the discovery that ribosomes are the sites of protein synthesis stimulated attempts to define the structures of these particles in greater detail.

Ultracentrifugation was used to measure the sizes of ribosomes and their components

The initial progress in understanding the detailed structure of the ribosome came not from observing them with the electron microscope but by analyzing their components by ultracentrifugation (Technical Note 2.2). Intact ribosomes have sedimentation coefficients of 80S for eukaryotes and 70S for bacteria, and each can be broken down into smaller components (*Figure 11.10*):

- Each ribosome comprises two subunits. In eukaryotes these subunits are 60S and 40S; in bacteria they are 50S and 30S. Note that sedimentation coefficients are not additive because they depend on shape as well as mass; it is perfectly acceptable for the intact ribosome to have an S value less than the sum of its two subunits.
- The large subunit contains three rRNAs in eukaryotes (the 28S, 5.8S and 5S rRNAs) but only two in bacteria (23S and 5S rRNAs). In bacteria the equivalent of the eukaryotic 5.8S rRNA is contained within the 23S rRNA.
- The small subunit contains a single rRNA in both types of organism: an 18S rRNA in eukaryotes and a 16S rRNA in bacteria.
- Both subunits contain a variety of ribosomal proteins, the numbers detailed in *Figure 11.10* . The ribosomal proteins of the small subunit are called S1, S2, etc.; those of the large subunit are L1, L2, etc. There is just one of each protein per ribosome, except for L7 and L12, which are present as dimers.

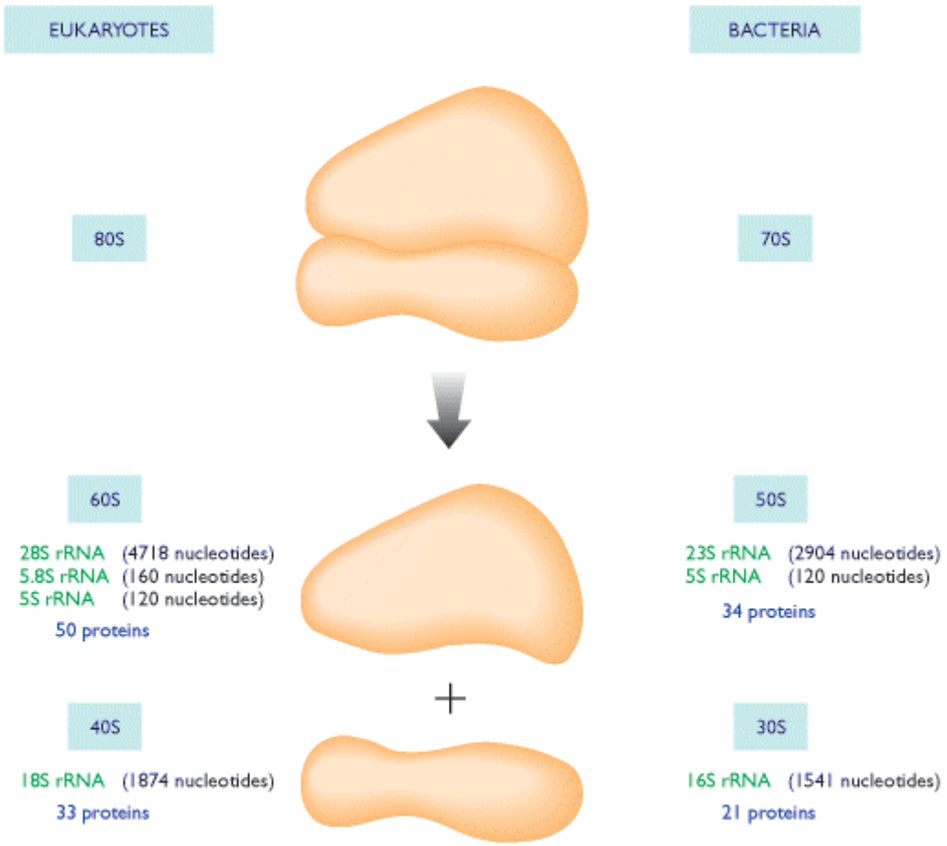

Figure 11.10. **The composition of eukaryotic and bacterial ribosomes**. The details refer to a 'typical' eukaryotic ribosome and the *Escherichia coli* ribosome. Variations between different species mainly concern the numbers of ribosomal proteins.

Probing the fine structure of the ribosome

Once the basic composition of eukaryotic and bacterial ribosomes had been worked out, attention was focused on the way in which the various rRNAs and proteins fit together. Important information was provided by the first rRNA sequences, comparisons between these identifying conserved regions that can base-pair to form complex two-dimensional structures (*Figure 11.11*).

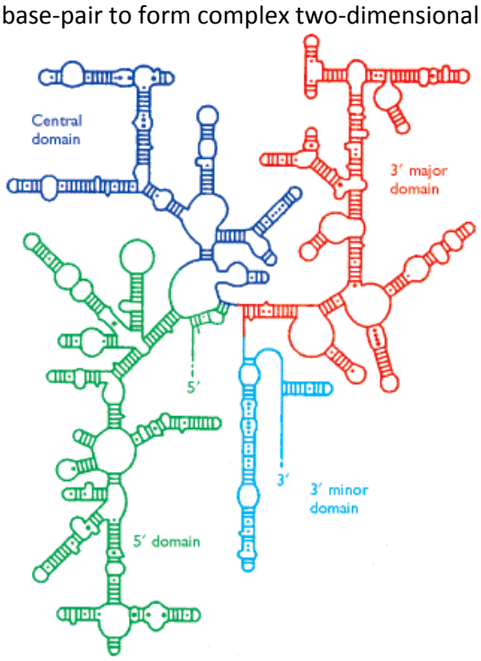

Figure 11.11. **The base-paired structure of the *Escherichia coli* 16S rRNA**. In this representation, standard base pairs (G-C, A-U) are shown as bars; non-standard base pairs (e.g. G-U) are shown as dots

This suggested that the rRNAs provide a scaffolding within the ribosome, to which the proteins are attached, an interpretation that under-emphasizes the active role that rRNAs play in protein synthesis but which nonetheless was a useful foundation on which to base subsequent research. Much of that subsequent research has concentrated on the bacterial ribosome, which is smaller than the eukaryotic version and available in large amounts from extracts of cells grown to high density in liquid cultures. A number of technical approaches have been used to study the bacterial ribosome:

- ***Nuclease protection studies*** (Section 2.2.1) enable contacts between rRNAs and proteins to be identified.
- *Protein-protein crosslinking* identifies pairs or groups of proteins that are located close to one another in the ribosome.
- ***Electron microscopy*** has gradually become more sophisticated, enabling the overall structure of the ribosome to be resolved in greater detail. For example, innovations such as immunoelectron microscopy, in which ribosomes are labeled with antibodies specific for individual ribosomal proteins before examination, have been used to locate the positions of these proteins on the surface of the ribosome.
- *Site-directed hydroxyl radical probing* makes use of the ability of Fe(II) ions to generate hydroxyl radicals that cleave RNA phosphodiester bonds located within 1 nm of the site of radical production. This technique was used to determine the exact positioning of ribosomal protein S5 in the *E.* coli ribosome. Different amino acids within S5 were labeled with Fe(II) and hydroxyl radicals induced in reconstituted ribosomes. The positions at which the 16S rRNA was cleaved were then used to infer the topology of the rRNA in the vicinity of S5 protein (*Figure 11.12* ; Heilek and Noller, 1996).

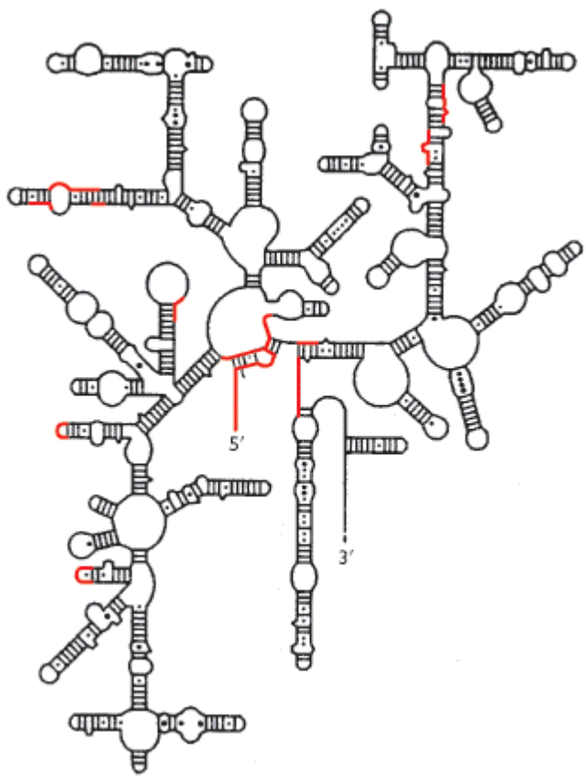

Figure 11.12. **Positions within the *Escherichia coli* 16S rRNA that form contacts with ribosomal protein S5**. The distribution of the contact positions (shown in red) for this single ribosomal protein emphasizes the extent to which the base-paired secondary structure of the rRNA is further folded within the three-dimensional structure of the ribosome. For details of the work that led to these results, see Heilek and Noller (1996).

In recent years these techniques have been increasingly supplemented by X-ray crystallography (Section 9.1.3), which has been responsible for the most exciting insights into ribosome structure. Analyzing the massive amounts of X-ray diffraction data that are produced by crystals of an object as large as a ribosome is a huge task, particularly at the level needed to obtain a structure that is

detailed enough to be informative about the way in which the ribosome works (Pennisi, 1999). This challenge has been met, and structures have been deduced for ribosomal proteins bound to their segments of rRNA (Conn *et al.*, 1999; Agalarov *et al.*, 2000), for the large and small subunits (Ban *et al.*, 2000; Wimberly *et al.*, 2000), and for the entire bacterial ribosome attached to mRNA and tRNAs (Yusupov *et al.*, 2001). As well as revealing the structure of the ribosome (*Figure 11.13*), this recent explosion of information has had an important impact on our understanding of the translation process, as we will see in the next section.

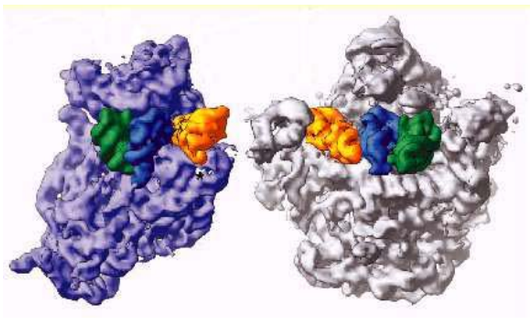

Figure 11.13. **The bacterial ribosome**. The picture shows the ribosome of the bacterium *Thermus thermophilus*. The small subunit is at the top, with the 16S rRNA in light blue and the small subunit ribosomal proteins in dark blue. The large subunit rRNAs are in grey and the proteins in purple. The gold area is the A site (Section 11.2.3) - the point at which aminoacylated tRNAs enter the ribosome during protein synthesis. This site, and most of the region within which protein synthesis actually occurs, is located in the cleft between the two subunits. Reprinted with permission from Mathews and Pe'ery (2001) *Trends Biochem. Sci.*, **26**, 585–587.

11.2.2. Initiation of translation

Although ribosomal architecture is similar in bacteria and eukaryotes, there are distinctions in the way in which translation is carried out in the two types of organism. The most important of these differences occurs during the first stage of translation, when the ribosome is assembled on the mRNA at a position upstream of the initiation codon.

Initiation in bacteria requires an internal ribosome binding site

The main difference between initiation of translation in bacteria and eukaryotes is that in bacteria the translation initiation complex is built up directly at the initiation codon, the point at which protein synthesis will begin, whereas eukaryotes use a more indirect process for locating the initiation point, as we will see in the next section.

When not actively participating in protein synthesis, ribosomes dissociate into their subunits, which remain in the cytoplasm waiting to be used for a new round of translation. In bacteria, the process initiates when a small subunit, in conjunction with the translation initiation factor IF-3 (*Table 11.2*), attaches to the ribosome binding site (also called the Shine-Dalgarno sequence). This is a short target site, consensus 5′-AGGAGGU-3′ in *E.* coli (*Table 11.3*), located about 3–10 nucleotides upstream of the initiation codon, the point at which translation will begin (*Figure 11.14*). The ribosome binding site is complementary to a region at the 3′ end of the 16S rRNA, the one present in the small subunit, and it is thought that base-pairing between the two is involved in the attachment of the small subunit to the mRNA.

Figure 11.14. The ribosome binding site for bacterial translation. In *Escherichia coli*, the ribosome binding site has the consensus sequence 5′-AGGAGGU-3′ and is located between 3 and 10 nucleotides upstream of the initiation codon.

Attachment to the ribosome binding site positions the small subunit of the ribosome over the initiation codon (*Figure 11.15*). This codon is usually 5′-AUG-3′, which codes for methionine, although 5′-GUG-3′ and 5′-UUG-3′ are sometimes used. All three codons can be recognized by the same initiator tRNA, the last two by wobble. This initiator tRNA is the one that was aminoacylated with methionine and subsequently modified by conversion of the methionine to *N*-formylmethionine (see *Figure 11.5B*). The modification attaches a formyl group (-COH) to the amino group, which means that only the carboxyl group of the initiator methionine is free to participate in peptide bond formation. This ensures that polypeptide synthesis can take place only in the N→C direction. The initiator $tRNA_i^{Met}$ is brought to the small subunit of the ribosome by a second initiation factor, IF-2, along with a molecule of GTP, the latter acting as a source of energy for the final step of initiation. Note that the $tRNA_i^{Met}$ is only able to decode the initiation codon; it cannot enter the complete ribosome during the elongation phase of translation during which internal 5′-AUG-3′ codons are recognized by a different $tRNA^{Met}$ carrying an unmodified methionine.

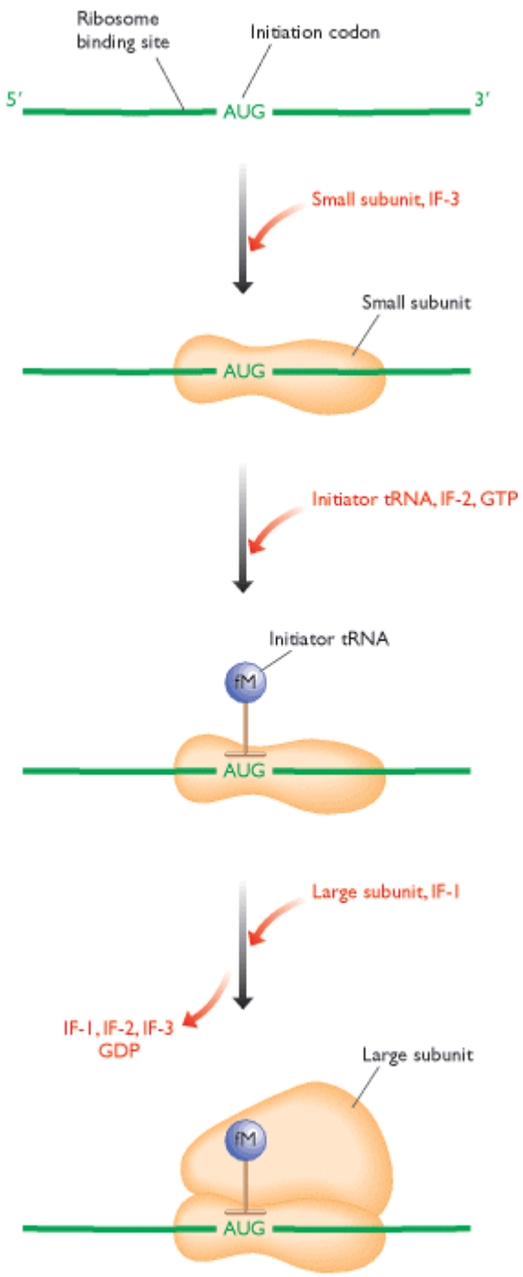

Figure 11.15. **Initiation of translation in *Escherichia coli***. See the text for details. Note that the different components of the initiation complex are not drawn to scale. Abbreviation: fM, *N*-formylmethionine.

Table 11.2. Functions of the bacterial translation factors

Factor	Function
Initiation factors	
IF-1	Unclear; X-ray crystallography studies show that binding of IF-1 blocks the A site (see page 329), so its function may be to prevent premature entry of tRNAs into the A site. Alternatively IF-1 may cause conformational changes that prepare the small subunit for attachment to the large subunit
IF-2	Directs the initiator $tRNA_i^{Met}$ to its correct position in the initiation complex
IF-3	Prevents premature reassociation of the large and small subunits of the ribosome
Elongation factors	
EF-Tu	Directs the next tRNA to its correct position in the ribosome
EF-Ts	Regenerates EF-Tu after the latter has yielded the energy contained in its attached GTP molecule
EF-G	Mediates translocation
Release factors	
RF-1	Recognizes the termination codons 5′-UAA-3′ and 5′-UAG-3′
RF-2	Recognizes 5′-UAA-3′ and 5′-UGA-3′
RF-3	Stimulates dissociation of RF1 and RF2 from the ribosome after termination
Ribosome recycling factor	
RRF	Responsible for disassociating the ribosome subunits after translation has terminated

Table 11.3. Examples of ribosome binding sequences in *Escherichia coli*

Gene	Codes for	Ribosome binding sequence	Nucleotides to the start Codon
E. coli consensus	-	5'-AGGAGGU-3'	10
Lactose operon	Lactose utilization enzymes	5'-A G G A -3'	7
galE	Hexose-1-phosphate uridyltransferase	5'-G G A G -3'	6
rplJ	Ribosomal protein L10	5'-A G G A G -3'	8

Completion of the initiation phase occurs when IF-1 binds to the initiation complex. The precise role of IF-1 is unclear (see *Table 11.2*), but it may induce a conformational change in the initiation complex, enabling the large subunit of the ribosome to attach. Attachment of the large subunit requires energy, which is generated by hydrolysis of the bound GTP, and results in release of the initiation factors.

Initiation in eukaryotes is mediated by the cap structure and poly(A) tail

Only a small number of eukaryotic mRNAs have internal ribosome binding sites (see the next section). Instead, with most mRNAs the small subunit of the ribosome makes its initial attachment at the 5′-end of the molecule and then **scans** along the sequence until it locates the initiation codon. The process requires a plethora of initiation factors and there is still some confusion over the functions of all of these (*Table 11.4*). The details are as follows (*Figure 11.16* ; Dever, 1999).

The first step involves assembly of the pre-initiation complex. This structure comprises the 40S subunit of the ribosome, a ‘ternary complex' made up of the initiation factor eIF-2 bound to the initiator $tRNA^{Met}$ and a molecule of GTP, and three additional initiation factors, eIF-1, eIF-1A, and eIF-3. As in bacteria, the initiator tRNA is distinct from the normal $tRNA^{Met}$ that recognizes internal 5′-AUG-3′ codons but, unlike bacteria, it is aminoacylated with normal methionine, not the formylated version.

Table 11.4. Eukaryotic translation factors

Factor	Function
Initiation factors	
eIF-1	Component of the pre-initiation complex
eIF-1A	Component of the pre-initiation complex
eIF-2	Binds to the initiator $tRNA^{Met}$ within the ternary complex component of the pre-initiation complex; phosphorylation of eIF-2 results in a global repression of translation
eIF-3	Component of the pre-initiation complex; makes direct contact with eIF-4G and so forms the link with the cap binding complex
eIF-4A	Component of the cap binding complex; a helicase that aids scanning by breaking intramolecular base pairs in the mRNA
eIF-4B	Aids scanning, possibly by acting as a helicase that breaks intramolecular base pairs in the mRNA
eIF-4E	Component of the cap binding complex, possibly the component that makes direct contact with the cap structure at the 5′ end of the mRNA
eIF-4F	The cap binding complex, comprising eIF-4A, eIF-4E and eIF-4G, which makes the primary contact with the cap structure at the 5′ end of the mRNA
eIF-4G	Component of the cap binding complex; forms a bridge between the cap binding complex and eIF-3 in the pre-initiation complex; in at least some organisms, eIF-4G also forms an association with the poly(A) tail, via the polyadenylate-binding protein
eIF-5	Aids release of the other initiation factors at the completion of initiation
eIF-6	Associated with the large subunit of the ribosome; prevents large subunits from attaching to small subunits in the cytoplasm
Elongation factors	
eEF-1	Complex of four subunits (eEF-1a, eEF-1b, eEF-1d and eEF-1g); directs the next tRNA to its correct position in the ribosome
eEF-2	Mediates translocation
Release factors	
eRF-1	Recognizes the termination codon
eRF-3	Possibly stimulates dissociation of eRF-1 from the ribosome after termination; possibly causes the ribosome subunits to disassociate after termination of translation

After assembly, the pre-initiation complex associates with the 5′ end of the mRNA. This step requires the cap binding complex (sometimes called eIF-4F), which comprises the initiation factors eIF-4A, eIF-4E and eIF-4G. The contact with the cap might be made by eIF-4E alone (as shown in *Figure 11.16*) or might involve a more general interaction with the cap binding complex (Pestova and Hellen, 1999). The factor eIF-4G acts as a bridge between eIF-4E, bound to the cap, and eIF-3, attached to the pre-initiation complex (Hentze, 1997). The result is that the pre-initiation complex becomes attached to the 5′ region of the mRNA. Attachment of the pre-initiation complex to the mRNA is also influenced by the poly(A) tail, at the distant 3′ end of the mRNA. This interaction is thought to be mediated by the polyadenylate-binding protein (PADP), which is attached to the poly(A) tail (Section 10.1.2). In yeast and plants it has been shown that PADP can form an association with eIF-4G, this association requiring that the mRNA bends back on itself. With artificially uncapped mRNAs, the PADP interaction is sufficient to load the pre-initiation complex onto the 5′ end of the mRNA, but under normal circumstances the cap structure and poly(A) tail probably work together (Preiss and Hentze, 1998). The poly(A) tail could have an important regulatory role, as the length of the tail appears to be correlated with the extent of initiation that occurs with a particular mRNA.

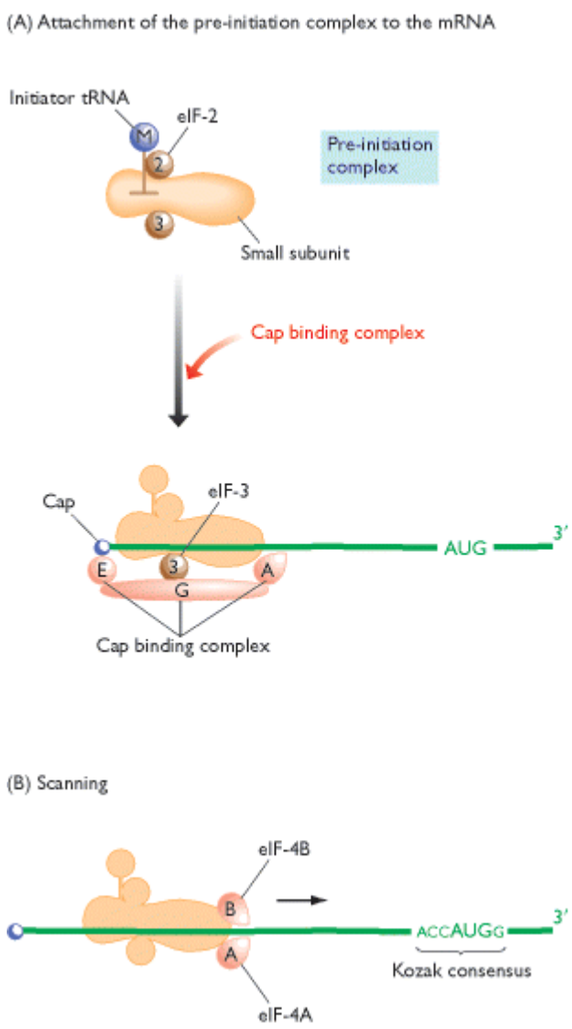

Figure 11.16. **Initiation of translation in eukaryotes**. (A) Assembly of the pre-initiation complex and its attachment to the mRNA. See the text for details. For clarity, several proteins whose precise roles are not understood have been omitted. The overall configuration of the complex is not known: the scheme shown here is based on Hentze (1997). (B) The pre-initiation complex scans along the mRNA until it reaches the initiation codon, which is recognizable because it is located within the Kozak consensus sequence. Scanning is aided by eIF-4A and eIF-4B, which are thought to have helicase activity. It is probable that eIF-3 remains attached to the pre-initiation complex during scanning, as shown here. It is not clear whether eIF-4E and eIF-4G also remain attached at this stage. Note that scanning is an energy-dependent process that requires hydrolysis of ATP. Abbreviation: M, methionine.

After becoming attached to the 5′ end of the mRNA, the initiation complex, as it is now called, has to scan along the molecule and find the initiation codon. The leader regions of eukaryotic mRNAs can be several tens, or even hundreds, of nucleotides in length and often contain regions that form hairpins and other base-paired structures. These are probably removed by a combination of eIF-4A and eIF-4B. eIF-4A, and possibly also eIF-4B, has a helicase activity and so is able to break intramolecular base pairs in the mRNA, freeing the passage for the initiation complex (*Figure 11.16B*). The initiation codon, which is usually 5′-AUG-3′ in eukaryotes, is recognizable because it is contained in a short consensus sequence, 5′-ACCAUGG-3′, referred to as the Kozak consensus.
Once the initiation complex is positioned over the initiation codon, the large subunit of the ribosome attaches. As in bacteria, this requires hydrolysis of GTP and leads to release of the initiation factors. Two final initiation factors are involved at this stage: eIF-5, which aids release of the other factors, and eIF-6, which is associated with the unbound large subunit and prevents it from attaching to a small subunit in the cytoplasm.

Initiation of eukaryotic translation without scanning

The scanning system for initiation of translation does not apply to every eukaryotic mRNA. This was first recognized with the picornaviruses, a group of viruses with RNA genomes which includes the human poliovirus and rhinovirus, the latter being responsible for the common cold. Transcripts from these viruses are not capped but instead have an internal ribosome entry site (**IRES**) which is similar in function to the ribosome binding site of bacteria, although the sequences of IRESs and their positions relative to the initiation codon are more variable than the bacterial versions (Mountford and Smith, 1995). The presence of IRESs on their transcripts means that picornaviruses can block protein synthesis in the host cell by inactivating the cap binding complex, without affecting translation of their own transcripts, although this is not a normal part of the infection strategy of all picornaviruses.

Remarkably, no virus proteins are required for recognition of an IRES by a host ribosome. In other words, the normal eukaryotic cell possesses proteins and/or other factors that enable it to initiate translation by the IRES method (Holcik *et al*., 2000). Because of their variability, IRESs are difficult to identify by inspection of DNA sequences, but it is becoming clear that a few nuclear gene transcripts possess them and that these are translated, at least under some circumstances, via their IRES rather than by scanning. Examples are the mRNAs for the mammalian immunoglobulin heavy-chain binding protein and the *Drosophila* Antennapedia protein (Section 12.3.3). IRESs are also found on several mRNAs whose protein products are translated when the cell is put under stress, for example by exposure to heat, irradiation, or low oxygen conditions. Under these circumstances, cap-dependent translation is globally suppressed (as described in the next section). The presence of IRESs on the 'survival' mRNAs therefore enables these to undergo preferential translation at the time when their products are needed.

Regulation of translation initiation

The initiation of translation is an important control point in protein synthesis, at which two different types of regulation can be exerted. The first of these is global regulation, which involves a general alteration in the amount of protein synthesis occurring, with all mRNAs translated by the cap mechanism being affected to a similar extent. In eukaryotes this is commonly achieved by phosphorylation of eIF-2, which results in repression of translation initiation by preventing eIF-2 from binding the molecule of GTP that it needs before it can transport the initiator tRNA to the small subunit of the ribosome. Phosphorylation of eIF-2 occurs during stresses such as heat shock, when the overall level of protein synthesis is decreased and IRES-mediated translation takes over.
Transcript-specific regulation involves mechanisms that act on a single transcript or a small group of transcripts coding for related proteins. The most frequently cited example of transcript-specific regulation involves the operons for the ribosomal protein genes of *E.* coli (*Figure 11.17A*). The leader region of the mRNA transcribed from each operon contains a sequence that acts as a binding site for one of the proteins coded by the operon. When this protein is synthesized it can either attach to its position on the ribosomal RNA, or bind to the leader region of the mRNA. The rRNA attachment is favored and occurs if there are free rRNAs in the cell. Once all the free rRNAs have been assembled into ribosomes, the ribosomal protein binds to its mRNA, blocking translation initiation and hence switching off further synthesis of the ribosomal proteins coded by that particular mRNA. Similar events involving other mRNAs ensure that synthesis of each ribosomal protein is coordinated with the amount of free rRNA in the cell.

A second example of transcript-specific regulation, one occurring in mammals, involves the mRNA for ferritin, an iron-storage protein (*Figure 11.17B*). In the absence of iron, ferritin synthesis is inhibited by proteins that bind to sequences called iron-response elements located in the leader region of the ferritin mRNA. The bound proteins block the ribosome as it attempts to scan along the mRNA in search of the initiation codon. When iron is present, the binding proteins detach and the mRNA is translated. Interestingly, the mRNA for a related protein - the transferrin receptor involved in the uptake of iron - also has iron-response elements, but in this case detachment of the binding proteins in the presence of iron results not in translation of the mRNA but in its degradation. This is logical because when iron is present in the cell, there is less requirement for transferrin receptor activity because there is less need to import iron from outside.

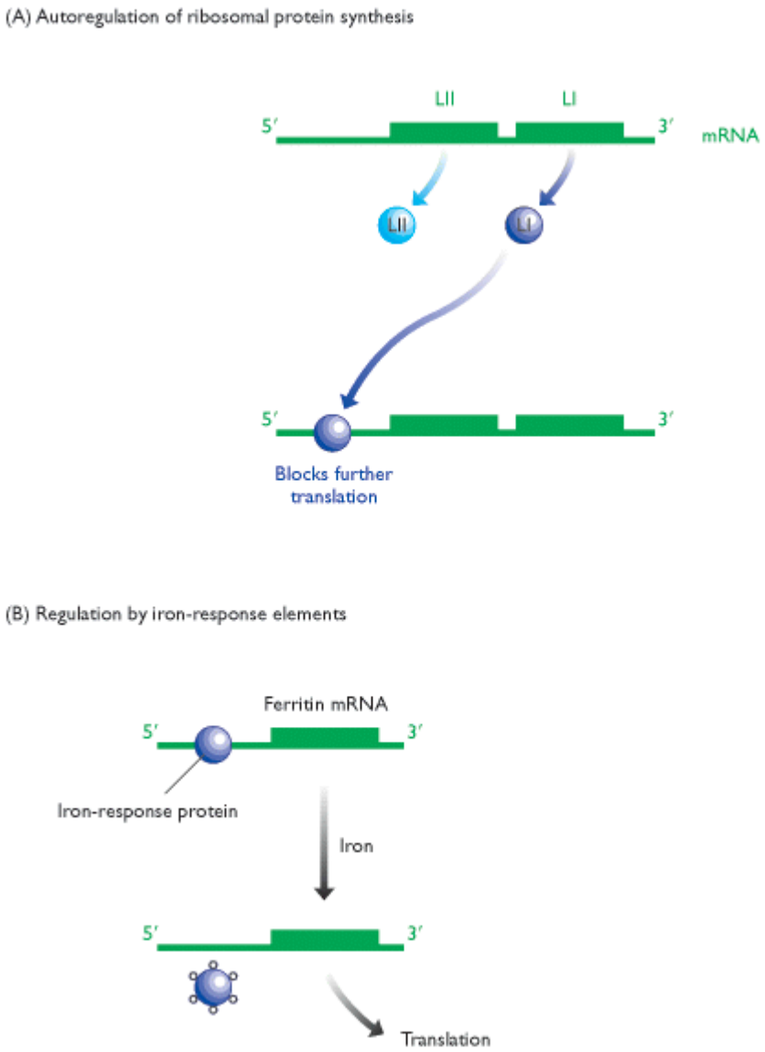

Figure 11.17. **Transcript-specific regulation of translation initiation**. (A) Regulation of ribosomal protein synthesis in bacteria. The L11 operon of *Escherichia coli* is transcribed into an mRNA carrying copies of the genes for the L11 and L1 ribosomal proteins. When the L1 binding sites on the available 23S rRNA molecules have been filled, L1 binds to the 5′ untranslated region of the mRNA, blocking further initiation of translation. (B) Regulation of ferritin protein synthesis in mammals. The iron-response protein binds to the 5′ untranslated region of the ferritin mRNA when iron is absent, preventing the synthesis of ferritin

11.2.3. Elongation of translation

The main differences between translation in bacteria and eukaryotes occur during the initiation phase; the events after the large subunit of the ribosome becomes associated with the initiation complex are similar in both types of organism. We can therefore deal with them together, by looking at what happens in bacteria and referring to the distinctive features of eukaryotic translation where appropriate.

Elongation in bacteria and eukaryotes

Attachment of the large subunit results in two sites at which aminoacyl-tRNAs can bind. The first of these, the **P** or **peptidyl site**, is already occupied by the initiator $tRNA_i^{Met}$, charged with *N*-formylmethionine or methionine, and base-paired with the initiation codon. The second site, the **A** or **aminoacyl site**, covers the second codon in the open reading frame (*Figure 11.18*). The structures revealed by X-ray crystallography show that these sites are located in the cavity between the large and small subunits of the ribosome, the codon-anticodon interaction being associated with the small subunit and the aminoacyl end of the tRNA with the large subunit (*Figure 11.19* ; Yusupov *et al.*, 2001).

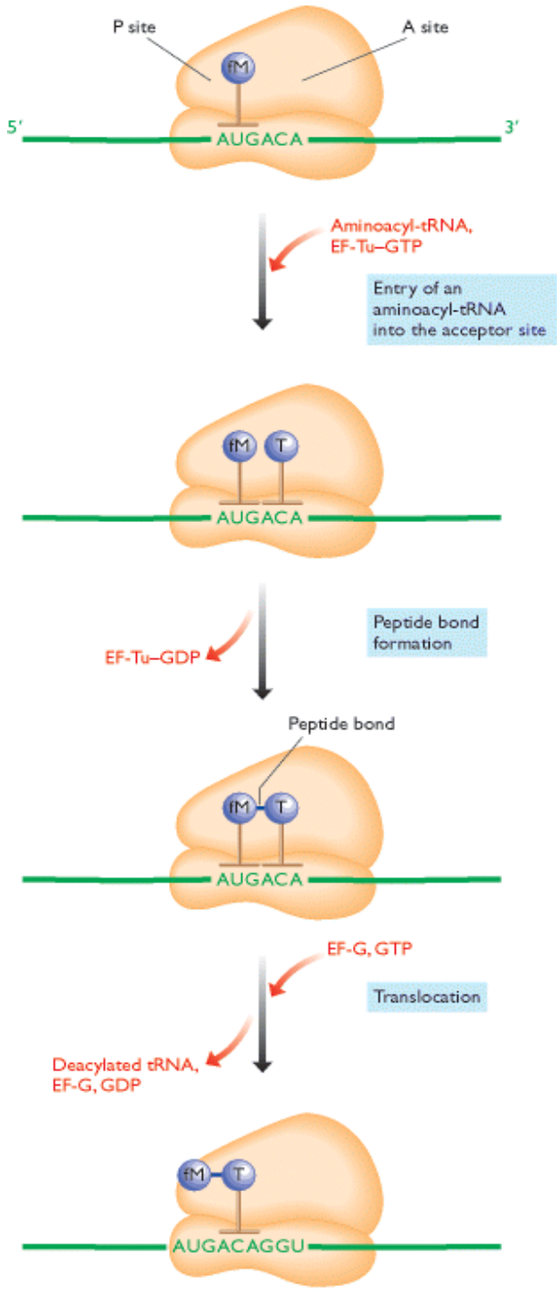

Figure 11.18. **Elongation of translation**. The diagram shows the events occurring during a single elongation cycle in *Escherichia coli*. See the text for details regarding eukaryotic translation. Abbreviations: fM, *N*-formylmethionine; T, threonine

Figure 11.19. **The important sites in the ribosome**. The structure on the left is the large subunit of the *Thermus thermophilus* ribosome; that on the right is the small subunit. The views look down onto the two surfaces that contact one another when the subunits are placed together to make the intact ribosome. The A, P and E sites are labeled, and each one is occupied by a tRNA shown in red or orange. The main part of each tRNA is embedded within the large subunit, with just the anticodon arms and loops associated with the small subunit. Those parts of the ribosome that make the important bridging contacts between the subunits are labeled as B1a, etc. Reprinted with permission from Yusupov *et al.*, *Science*, **292**, 883–896. Copyright 2001 American Association for the Advancement of Science.

The A site becomes filled with the appropriate aminoacyl-tRNA, which in *E.* coli is brought into position by the elongation factor EF-Tu, which ensures that only tRNAs that carry the correct amino acid are able to enter the ribosome, mischarged tRNAs being rejected at this point (Ibba, 2001). EF-Tu is an example of a G protein, meaning that it binds a molecule of GTP which it can hydrolyze to release energy. In eukaryotes the equivalent factor is called eEF-1, which is a complex of four subunits: eEF-1a, eEF-1b, eEF-1d and eEF-1g (see *Table 11.4*). The first of these exists in at least two forms, eEF-1a1 and eEF-1a2, which are highly similar proteins that probably have equivalent functions in different tissues (Hafezparast and Fisher, 1998). Specific contacts between the tRNA, mRNA and the 16S rRNA within the A site ensure that only the correct tRNA is accepted. These contacts are able to discriminate between a codon-anticodon interaction in which all three base pairs have formed, and one in which one or more mis-pairs are present, the latter signaling that the wrong tRNA is present (Yoshizawa *et al.*, 1999). This is probably just one part of a series of safeguards that ensure the accuracy of the translation process (Rodnina and Wintermeyer, 2001a; 2001b).

When the aminoacyl-tRNA has entered the A site, a peptide bond is formed between the two amino acids. This involves a peptidyl transferase enzyme, which releases the amino acid from the initiator $tRNA_i^{Met}$ and then forms a peptide bond between this amino acid and the one attached to the second tRNA. In bacteria, the peptidyl transferase activity resides in the 23S rRNA of the large subunit, and so is an example of a ribozyme (Section 10.2.3; see Research Briefing 11.1). The reaction is energy dependent and requires hydrolysis of the GTP attached to EF-Tu (eEF-1 in eukaryotes). This inactivates EF-Tu, which is ejected from the ribosome and regenerated by EF-Ts. A eukaryotic equivalent of EF-Ts has not been identified, and it is possible that one of the subunits of eEF-1 has the regenerative activity.

Now the dipeptide corresponding to the first two codons in the open reading frame is attached to the tRNA in the A site. The next step is translocation, during which three things happen at once (see *Figure 11.18*):

- The ribosome moves along three nucleotides, so the next codon enters the A site.
- The dipeptide-tRNA in the A site moves to the P site.
- The deacylated tRNA in the P site moves to a third position, the **E** or exit site, in bacteria or, in eukaryotes, is simply ejected from the ribosome.

Translocation requires hydrolysis of a molecule of GTP and is mediated by EF-G in bacteria and by eEF-2 in eukaryotes. Electron microscopy of ribosomes at different intermediate stages in translocation suggests that the two subunits rotate slightly in opposite directions, opening up the space between them and enabling the ribosome to slide along the mRNA (Frank and Agrawal, 2000). Translocation results in the A site becoming vacant, allowing a new aminoacyl-tRNA to enter. The elongation cycle is now repeated, and continues until the end of the open reading frame is reached.

Frameshifting and other unusual events during elongation

The straightforward codon-by-codon translation of an mRNA is looked upon as the standard way in which proteins are synthesized. But an increasing number of unusual elongation events are being discovered. One of these is frameshifting, which occurs when a ribosome pauses in the middle of an mRNA, moves back one nucleotide or, less frequently, forward one nucleotide, and then continues translation (Farabaugh, 1996). The result is that the codons that are read after the pause are not contiguous with the preceding set of codons: they lie in a different reading frame (*Figure 11.20A*).

Spontaneous frameshifts occur randomly and are deleterious because the polypeptide synthesized after the frameshift has the incorrect amino acid sequence. But not all frameshifts are spontaneous: a few mRNAs utilize programmed frameshifting to induce the ribosome to change frame at a specific point within the transcript. Programmed frameshifting occurs in all types of organism, from bacteria through to humans, as well as during expression of a number of viral genomes. An example occurs during synthesis of DNA polymerase III in *E. coli*, the main enzyme involved in replication of DNA (Section 13.2.2). Two of the DNA polymerase III subunits, γ and τ, are coded by a single gene, *dnaX*. Subunit τ is the full-length translation product of the *dnaX*

mRNA, and subunit γ is a shortened version. Synthesis of γ involves a frameshift in the middle of the *dnaX* mRNA, the ribosome encountering a termination codon immediately after the frameshift and so producing the truncated γ version of the translation product. It is thought that the frameshift is induced by three features of the *dnaX* mRNA:

- A hairpin loop, located immediately after the frameshift position, which stalls the ribosome;
- A sequence similar to a ribosome binding site immediately upstream of the frameshift position, which is thought to base-pair with the 16S rRNA (as does an authentic ribosome binding site), again causing the ribosome to stall;
- The codon 5′-AAG-3′ at the frameshift position. The presence of a modified nucleotide at the wobble position of the $tRNA^{Lys}$ that decodes 5′-AAG-3′ means that the codon-anticodon interaction is relatively weak at this position, enabling the frameshift to occur.

A similar phenomenon - translational slippage - enables a single ribosome to translate an mRNA that contains copies of two or more genes (*Figure 11.20B*). This means that, for example, a single ribosome can synthesize each of the five proteins coded by the mRNA transcribed from the tryptophan operon of *E.* coli (see *Figure 2.20B*). When the ribosome reaches the end of one series of codons it releases the protein it has just made, slips to the next initiation codon, and begins synthesizing the next protein. A more extreme form of slippage is translational bypassing (Herr *et al.*, 2000) in which a larger part of the transcript, possibly a few tens of base pairs, is skipped, and elongation of the original protein continues after the bypassing event (*Figure 11.20C*). The bypass starts and ends either at two identical codons or at two codons that can be translated by the same tRNA by wobble. This suggests that the jump is controlled by the tRNA attached to the growing polypeptide, which scans the mRNA as the ribosome tracks along, and halts the bypass when a new codon to which it can base-pair is reached. Translational bypassing of 44 nucleotides occurs in *E.* coli during translation of the mRNA for gene *60* of T4 bacteriophage, which codes for a DNA topoisomerase subunit. Similar events have also been identified in a variety of other bacteria. Bypassing could result in two different proteins being synthesized from one mRNA - one protein from normal translation and one from bypassing - but whether this is its general function is not yet known.

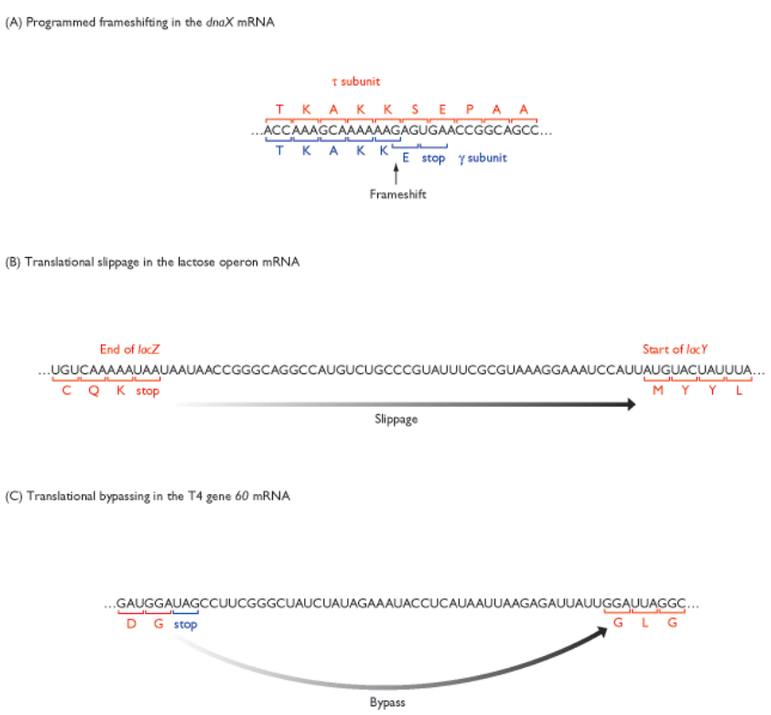

Figure 11.20. **Three unusual translation elongation events occurring in *Escherichia coli***. (A) Programmed frameshifting during translation of the *dnaX* mRNA. During synthesis of the γ subunit the ribosome shifts back one nucleotide, immediately after a series of As coding for two lysine amino acids. The ribosome inserts a glutamic acid into the polypeptide and then encounters a termination codon. (B) Slippage between the *lacZ* and *lacY* genes of the lactose operon mRNA. (C) Bypassing during translation of the T4 gene *60* mRNA involves a jump between two glycine codons. For the one-letter abbreviations of the amino acids see *Table 3.1* .

11.2.4. Termination of translation

Protein synthesis ends when one of the three termination codons is reached. The A site is now entered not by a tRNA but by a protein release factor (*Figure 11.21*). Bacteria have three of these: RF-1 which recognizes the termination codons 5′-UAA-3′ and 5′-UAG-3′, RF-2 which recognizes 5′-UAA-3′ and 5′-UGA-3′, and RF-3 which stimulates release of RF1 and RF2 from the ribosome after termination, in a reaction requiring energy from the hydrolysis of GTP. Eukaryotes have just two release factors (see *Table 11.4*): eRF-1, which recognizes the termination codon, and eRF-3, which might play the same role as RF-3 although this has not been proven (Kisselev and Buckingham, 2000). The structure of eRF-1 has been solved by X-ray crystallography, showing that the shape of this protein is very similar to that of a tRNA (*Figure 11.22*). This gives an indication of how the release factor is able to enter the A site when the termination codon is reached.

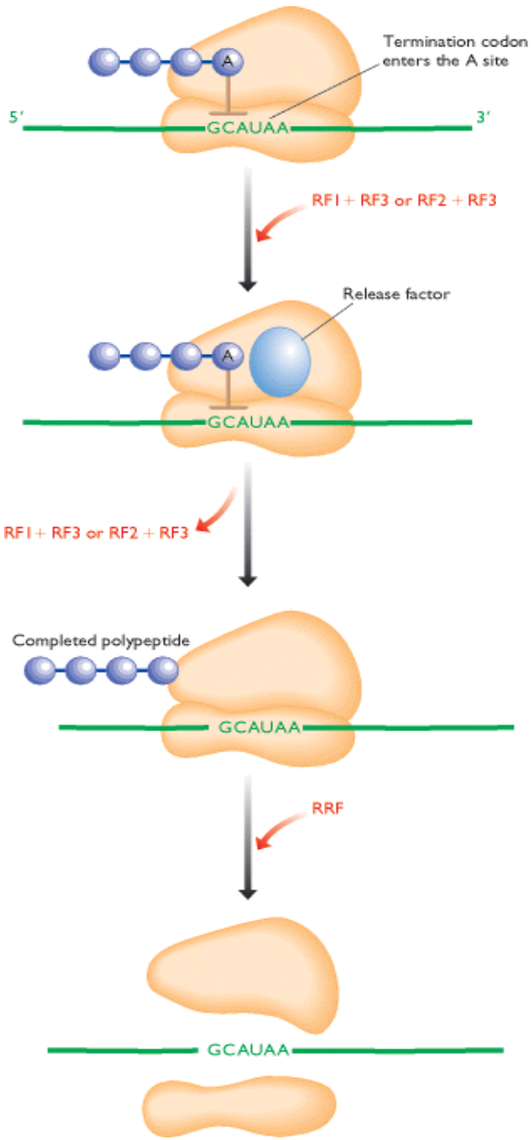

Figure 11.21. **Termination of translation**. Termination in *Escherichia coli* is illustrated. For differences in eukaryotes, see the text. The amino acid labeled with an ‘A' is an alanine. Abbreviations: RF, release factor; RRF, ribosome recycling factor

The release factors terminate translation but they do not appear to be responsible for disassociation of the ribosomal subunits, at least not in bacteria. This is the function of an additional protein called ribosome recycling factor (**RRF**) which, like eRF-1, has a tRNA-like structure (Selmer *et al*., 1999). RRF probably enters the P or A site and ‘unlocks' the ribosome (see *Figure 11.21*). Disassociation requires energy, which is released from GTP by EF-G, one of the elongation factors (see *Table 11.2*), and also requires the initiation factor IF-3 to prevent the subunits from attaching together again. A eukaryotic equivalent of RRF has not been identified,

and this may be one of the functions of eRF-3. The disassociated ribosome subunits enter the cytoplasmic pool, where they remain until used again in another round of translation.

Permission to reproduce this figure in this web version of ***Genomes 2*** is either pending or has not been granted.

Figure 11.22. **The structure of the eukaryotic release factor eRF-1 is similar to that of a tRNA**. The left panel shows eRF-1 and the right panel shows a tRNA. The part of eRF-1 that resembles the tRNA is highlighted in white. The purple segment of eRF-1 interacts with the second eukaryotic release factor, eRF-3. Reproduced with permission from Kisselev and Buckingham (2000)*Trends Biochem. Sci.*, **25**, 561–566

11.3. Post-translational Processing of Proteins

Translation is not the end of the genome expression pathway. The polypeptide that emerges from the ribosome is inactive, and before taking on its functional role in the cell must undergo at least the first of the following four types of post-translational processing (*Figure 11.23*):

- *Protein folding.* The polypeptide is inactive until it is folded into its correct tertiary structure.
- **Proteolytic cleavage.** Some proteins are processed by cutting events carried out by enzymes called proteases. These cutting events may remove segments from one or both ends of the polypeptide, resulting in a shortened form of the protein, or they may cut the polypeptide into a number of different segments, all or some of which are active.
- **Chemical modification.** Individual amino acids in the polypeptide might be modified by attachment of new chemical groups.
- **Intein splicing.** Inteins are intervening sequences in some proteins, similar in a way to introns in mRNAs. They have to be removed and the exteins ligated in order for the protein to become active.

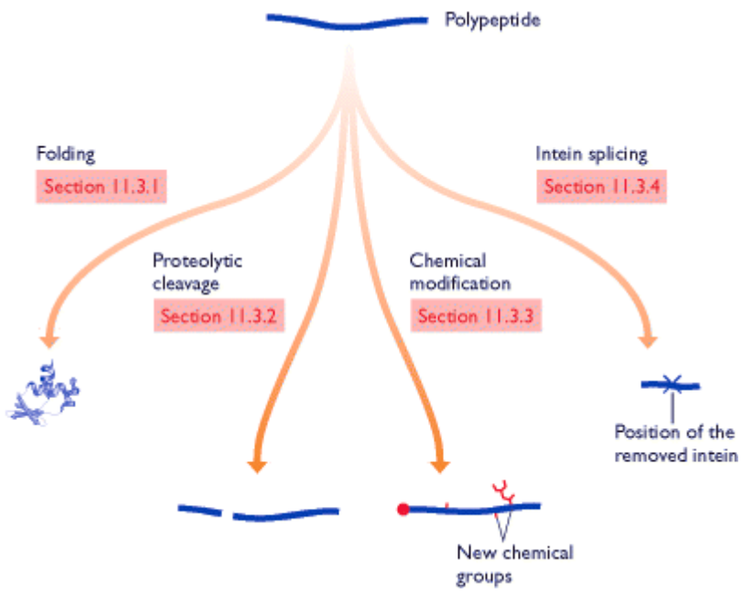

Figure 11.23. **Schematic representation of the four types of post-translational processing event**. Not all events occur in all organisms - see the text for details

Often these different types of processing occur together, the polypeptide being cut and modified at the same time that it is folded. If this is the case then the cutting, modification and/or splicing events may be necessary for the polypeptide to take up its correct three-dimensional conformation, because this is dependent in part on the relative positioning of the various chemical groups along the molecule. Alternatively, a cutting event or a chemical modification may occur

after the protein has been folded, possibly as part of a regulatory mechanism that converts a folded but inactive protein into an active form.

11.3.1. Protein folding

Protein folding was introduced in Chapter 3 when we examined the four levels of protein structure (primary, secondary, tertiary and quaternary) and learnt that all of the information that a polypeptide needs in order to adopt its correct three-dimensional structure is contained within its amino acid sequence (Section 3.3.3). This is one of the central principles of molecular biology. We must therefore examine its experimental basis and consider how the information contained in the amino acid sequence is utilized during the folding process for a newly translated polypeptide.

Not all proteins fold spontaneously in the test tube

The notion that the amino acid sequence contains all the information needed to fold the polypeptide into its correct tertiary structure derives from experiments carried out with ribonuclease in the 1960s (Anfinsen, 1973). Ribonuclease is a small protein, just 124 amino acids in length, containing four disulfide bridges and with a tertiary structure that is made up predominantly of β-sheet, with very little α-helix. Studies of its folding were carried out with ribonuclease that had been purified from cow pancreas and resuspended in buffer. Addition of urea, a compound that disrupts hydrogen bonding, resulted in a decrease in the activity of the enzyme (measured by testing its ability to cut RNA) and an increase in the viscosity of the solution (*Figure 11.24*), indicating that the protein was being **denatured** by unfolding into an unstructured polypeptide chain. The critical observation was that when the urea was removed by dialysis, the viscosity decreased and the enzyme activity reappeared. The conclusion is that the protein refolds spontaneously when the denaturant (in this case, urea) is removed. In these initial experiments the four disulfide bonds remained intact because they were not disrupted by urea, but the same result occurred when the urea treatment was combined with addition of a reducing agent to break the disulfide bonds: the activity was still regained on renaturation. This shows that the disulfide bonds are not critical to the protein's ability to refold, they merely stabilize the tertiary structure once it has been adopted.

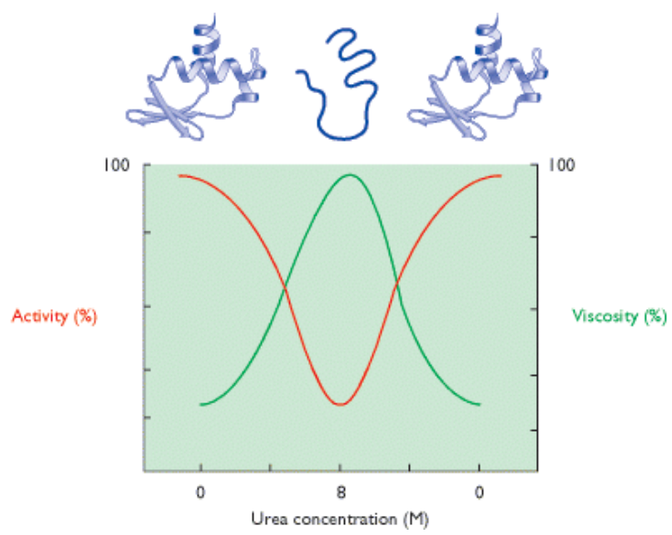

Figure 11.24. **Denaturation and spontaneous renaturation of a small protein**. As the urea concentration increases to 8M, the protein becomes denatured by unfolding: its activity decreases and the viscosity of the solution increases. When the urea is removed by dialysis, this small protein re-adopts its folded conformation. The activity of the protein increases back to the original level and the viscosity of the solution decreases

More detailed study of the spontaneous folding pathways for ribonuclease and other small proteins has led to the following general two-step description of the process (Hartl, 1996):

1. The secondary structural motifs along the polypeptide chain form within a few milliseconds of the denaturant being removed. This step is accompanied by the protein collapsing into a compact, but not folded, organization, with its hydrophobic groups on the inside, shielded from water.

2. During the next few seconds or minutes, the secondary structural motifs interact with one another and the tertiary structure gradually takes shape, often via a series of intermediate conformations. In other words, the protein follows a folding pathway. There may, however, be more than one possible pathway that a protein can follow to reach its correctly folded structure (Radford, 2000). The pathways may also have side-branches into which the protein can be diverted, leading to an incorrect structure. If an incorrect structure is sufficiently unstable then partial or complete unfolding may occur, allowing the protein a second opportunity to pursue a productive route towards its correct conformation (*Figure 11.25*).

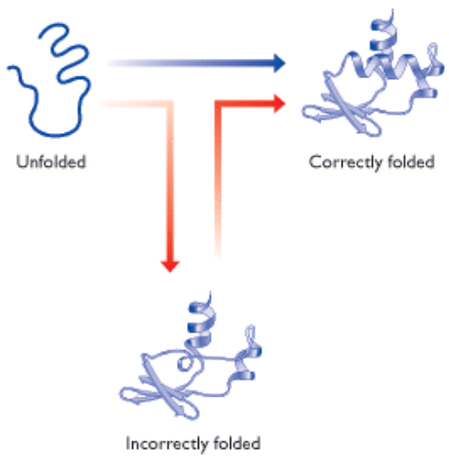

Figure 11.25. **An incorrectly folded protein might be able to refold into its correct conformation**. The blue arrow represents the correct folding pathway, leading from the unfolded protein on the left to the active protein on the right. The red arrow leads to an incorrectly folded conformation, but this conformation is unstable and the protein is able to unfold partially, return to its correct folding pathway and, eventually, reach its active conformation.

For several years it was more or less assumed that all proteins would fold spontaneously in the test tube, but experiments have shown that only smaller proteins with less complex structures possess this ability. Two factors seem to prevent larger proteins from folding spontaneously. The first of these is their tendency to form insoluble aggregates when the denaturant is removed: the polypeptides may collapse into interlocked networks when they attempt to protect their hydrophobic groups from water in step 1 of the general folding pathway. Experimentally, this can be avoided by using a low dilution of the protein, but this is not an option that the cell can take to prevent its unfolded proteins from aggregating. The second factor that prevents folding is that a large protein tends to get stuck in non-productive side branches of its folding pathway, taking on an intermediate form that is incorrectly folded but which is too stable to unfold to any significant extent. Concerns have also been raised about the relevance of *in vitro* folding, as studied with ribonuclease, to the folding of proteins in the cell, because a cellular protein might begin to fold before it has been fully synthesized. If the initial folding occurs when only part of the polypeptide is available, then there might be an increased possibility of incorrect branches of the folding pathway being followed. These various considerations prompted research into folding in living cells.

In cells, folding is aided by molecular chaperones

Most of our current understanding of protein folding in the cell is founded on the discovery of proteins that help other proteins to fold. These are called molecular chaperones and have been studied in most detail in *E.* coli. It is clear that both eukaryotes and archaea possess equivalent proteins, although some of the details of the way they work are different (Hartl, 1996; Slavotinek and Biesecker, 2001).

The molecular chaperones in *E.* coli can be divided into two groups:

- The Hsp70 chaperones, which include the proteins called Hsp70 (coded by the *dnaK* gene and sometimes called DnaK protein), Hsp40 (coded by *dnaJ*) and GrpE;
- The chaperonins, the main version of which in *E.* coli is the GroEL/GroES complex.

Molecular chaperones do not specify the tertiary structure of a protein, they merely help the protein find that correct structure. The two types of chaperone do this in different ways. The Hsp70 family bind to hydrophobic regions of proteins, including proteins that are still being translated (*Figure 11.26A*). They prevent protein aggregation by holding the protein in an open conformation until it is completely synthesized and ready to fold. The Hsp70 chaperones are also involved in other processes that require shielding of hydrophobic regions in proteins, such as transport through membranes and disaggregation of proteins that have been damaged by heat stress.

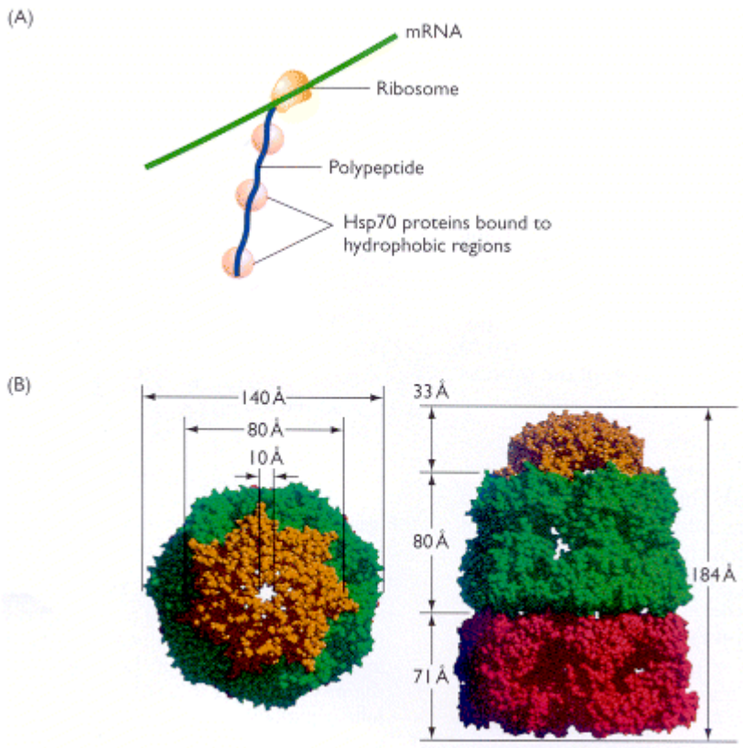

Figure 11.26. **Molecular chaperones of *Escherichia coli***. (A) Hsp70 chaperones bind to hydrophobic regions in unfolded polypeptides, including those that are still being translated, and hold the protein in an open conformation until it is ready to be folded. (B) The structure of the GroEL/GroES chaperonin. On the left is a view from the top and on the right a view from the side. 1Å is equal to 0.1 nm. The GroES part of the structure is made up of seven identical protein subunits and is shown in gold. The GroEL components consist of 14 identical proteins arranged into two rings (shown in red and green), each containing seven subunits. The main entrance into the central cavity is through the bottom of the structure shown on the right. Reprinted with permission from Xu *et al*., *Nature*, **388**, 741–750. Copyright 1997 Macmillan Magazines Limited. Original image kindly supplied by Dr Zhaohui Xu, Department of Biological Chemistry, The University of Michigan.

The chaperonins work in a quite different way. GroEL and GroES form a multi-subunit structure that looks like a hollowed-out bullet with a central cavity (*Figure 11.26B* ; Xu *et al*., 1997). A single unfolded protein enters the cavity and emerges folded. The mechanism for this is not known but it is postulated that GroEL/GroES acts as a cage that prevents the unfolded protein from aggregating with other proteins, and that the inside surface of the cavity changes from hydrophobic to hydrophilic in such a way as to promote the burial of hydrophobic amino acids within the protein. This is not the only hypothesis: other researchers hold that the cavity unfolds proteins that have folded incorrectly, passing these unfolded proteins back to the cytoplasm so they can have a second attempt at adopting their correct tertiary structure (Shtilerman *et al*., 1999).
Eukaryotic proteins equivalent to both the Hsp70 family of chaperones and the GroEL/GroES chaperonins have been found, but it seems that in eukaryotes protein folding depends mainly on the action of the Hsp70 proteins. This is probably true also of bacteria (Ellis, 2000) even though the GroEL/GroES chaperonins play a major role in the folding of metabolic enzymes, and proteins involved in transcription and translation (Houry *et al*., 1999).

11.3.2. Processing by proteolytic cleavage

Proteolytic cleavage has two functions in post-translational processing of proteins (*Figure 11.27*):

- It is used to remove short pieces from the N- and/or C-terminal regions of polypeptides, leaving a single shortened molecule that folds into the active protein.
- It is used to cut polyproteins into segments, all or some of which are active proteins.

These events are relatively common in eukaryotes but less frequent in bacteria.

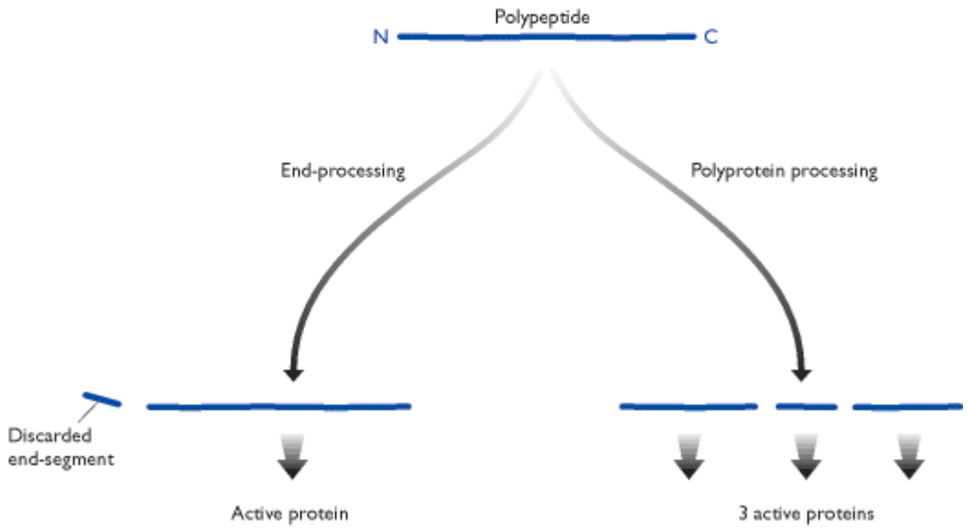

Figure 11.27. **Protein processing by proteolytic cleava**ge. On the left, the protein is processed by removal of the N-terminal segment. C-terminal processing also occurs with some proteins. On the right, a polyprotein is processed to give three different proteins. Not all proteins undergo proteolytic cleavage.

Cleavage of the ends of polypeptides

Processing by cleavage is common with secreted polypeptides whose biochemical activities might be deleterious to the cell producing the protein. An example is provided by melittin, the most abundant protein in bee venom and the one responsible for causing cell lysis after injection of the bee sting into the person or animal being stung. Melittin lyses cells in bees as well as animals and so must initially be synthesized as an inactive precursor. This precursor, promelittin, has 22 additional amino acids at its N terminus. The pre-sequence is removed by an extracellular protease that cuts it at 11 positions, releasing the active venom protein. The protease does not cleave within the active sequence because its mode of action is to release dipeptides with the sequence X-Y, where X is alanine, aspartic acid or glutamic acid, and Y is alanine or proline; these motifs do not occur in the active sequence (*Figure 11.28A*).

A similar type of processing occurs with insulin, the protein made in the islets of Langerhans in the vertebrate pancreas and responsible for controlling blood sugar levels. Insulin is synthesized as preproinsulin, which is 105 amino acids in length (*Figure 11.28B*). The processing pathway involves the removal of the first 24 amino acids to give proinsulin, followed by two additional cuts which excise a central segment, leaving two active parts of the protein, the A and B chains, which link together by formation of two disulfide bonds to form mature insulin. The first segment to be removed, the 24 amino acids from the N terminus, is a signal peptide, a highly hydrophobic stretch of amino acids that attaches the precursor protein to a membrane prior to transport across that membrane and out of the cell. Signal peptides are commonly found on proteins that bind to and/or cross membranes, in both eukaryotes and prokaryotes.

Proteolytic processing of polyproteins

In the examples shown in *Figure 11.28* , proteolytic processing results in a single mature protein. This is not always the case. Some proteins are initially synthesized as polyproteins, long polypeptides that contain a series of mature proteins linked together in head-to-tail fashion. Cleavage of the polyprotein releases the individual proteins, which may have very different functions from one another.

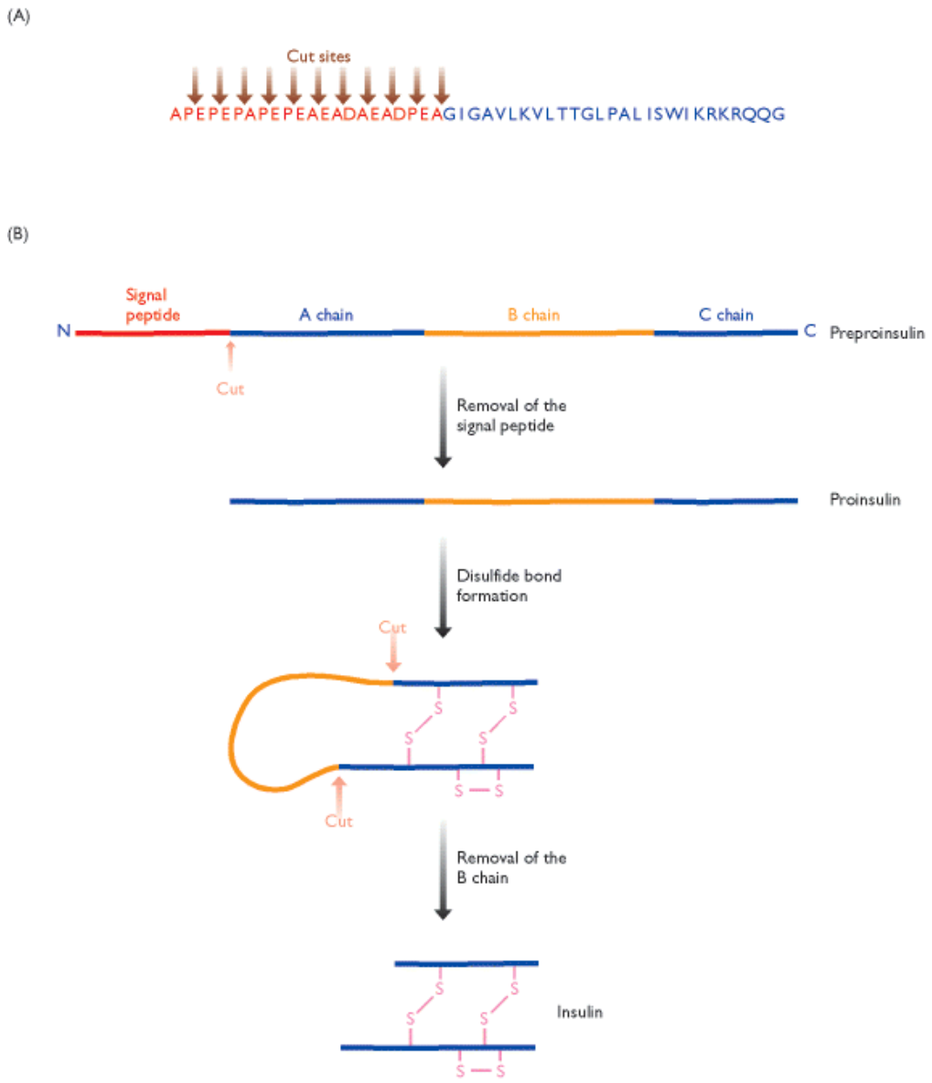

Figure 11.28. **Post-translational processing by proteolytic cleavage**. (A) Processing of promelittin, the bee-sting venom. Arrows indicate the cut sites. For the one-letter abbreviations of the amino acids see *Table 3.1* . (B) Processing of preproinsulin. See the text for details.

Polyproteins are not uncommon in eukaryotes. Several types of virus that infect eukaryotic cells use them as a way of reducing the sizes of their genomes, a single polyprotein gene with one promoter and one terminator taking up less space than a series of individual genes. Polyproteins are also involved in the synthesis of peptide hormones in vertebrates. For example, the polyprotein called pro-opiomelanocortin, made in the pituitary gland, contains at least ten different peptide hormones. These are released by proteolytic cleavage of the polyprotein (*Figure 11.29*), but not all can be produced at once because of overlaps between individual peptide sequences. Instead, the exact cleavage pattern is different in different cells.

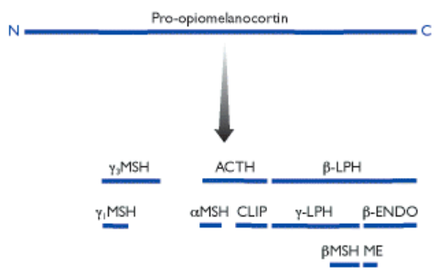

Figure 11.29. **Processing of the pro-opiomelanocortin polyprotein**. Abbreviations: ACTH, adrenocorticotropic hormone; CLIP, corticotropin-like intermediate lobe protein; ENDO, endorphin; LPH, lipotropin; ME, met-encephalin; MSH, melanotropin

11.3.3. Processing by chemical modification

The genome has the capacity to code for 21 different amino acids: the 20 specified by the standard genetic code, and selenocysteine, which is inserted into polypeptides by the context-dependent reading of a 5′-UGA-3′ codon (Section 3.3.2). This repertoire is increased dramatically by post-translational chemical modification of proteins, which results in a vast array of different amino acid types. The simpler types of modification occur in all organisms; the more complex ones, especially glycosylation, are rare in bacteria.

The simplest types of chemical modification involve addition of a small chemical group (e.g. an acetyl, methyl or phosphate group; *Table 11.5*) to an amino acid side chain, or to the amino or carboxyl groups of the terminal amino acids in a polypeptide (for an example see Bradshaw *et al.*, 1998). Over 150 different modified amino acids have been documented in different proteins, with each modification carried out in a highly specific manner, the same amino acids being modified in the same way in every copy of the protein. This is illustrated in *Figure 11.30* for histone H3. The example reminds us that chemical modification often plays an important role in determining the precise biochemical activity of the target protein: we saw in Section 8.2.1 how acetylation and methylation of H3 and other histones have an important influence on chromatin structure and hence on genome expression. Other types of chemical modification have important regulatory roles, an example being phosphorylation, which is used to activate many proteins involved in signal transduction (Section 12.1.2).

Table 11.5. Examples of post-translational chemical modifications

Modification	Amino acids that are modified	Examples of proteins
Addition of small chemical groups		
Acetylation	Lysine	Histones
Methylation	Lysine	Histones
Phosphorylation	Serine, threonine, tyrosine	Some proteins involved in signal transduction
Hydroxylation	Proline, lysine	Collagen
N-formylation	N-terminal glycine	Melittin
Addition of sugar side chains		
O-linked glycosylation	Serine, threonine	Many membrane proteins and secreted proteins
N-linked glycosylation	Asparagine	Many membrane proteins and secreted proteins
Addition of lipid side chains		
Acylation	Serine, threonine, cysteine	Many membrane proteins
N-myristoylation	N-terminal glycine	Some protein kinases involved in signal transduction
Addition of biotin		
Biotinylation	Lysine	Various carboxylase enzymes
See Section 12.1.2 for more information on the role of chemical modification during signal transduction.		

Figure 11.30. **Post-translational chemical modification of calf histone H3**. The first 30 amino acids of this 135-amino-acid protein are listed using the one-letter abbreviations (see *Table 3.1*). Five modifications occur: three methylations and two acetylations. For the role of methylation and acetylation of histones in determining chromatin structure see Section 8.2.1

A more complex type of modification is glycosylation, the attachment of large carbohydrate side chains to polypeptides (Drickamer and Taylor, 1998). There are two general types of glycosylation (*Figure 11.31*):

- *O-linked glycosylation* is the attachment of a sugar side chain via the hydroxyl group of a serine or threonine amino acid.
- *N-linked glycosylation* involves attachment through the amino group on the side chain of asparagine.

Glycosylation can result in attachment to the protein of grand structures comprising branched networks of 10–20 sugar units of various types. These side chains help to target proteins to particular sites in cells and determine the stability of proteins circulating in the bloodstream. Another type of large-scale modification involves attachment of long-chain lipids, often to serine or cysteine amino acids. This process is called acylation and occurs with many proteins that become associated with membranes. A less common modification is biotinylation, in which a molecule of biotin is attached to a small number of enzymes that catalyze the carboxylation of organic acids such as acetate and propionate (Chapman-Smith and Cronan, 1999).

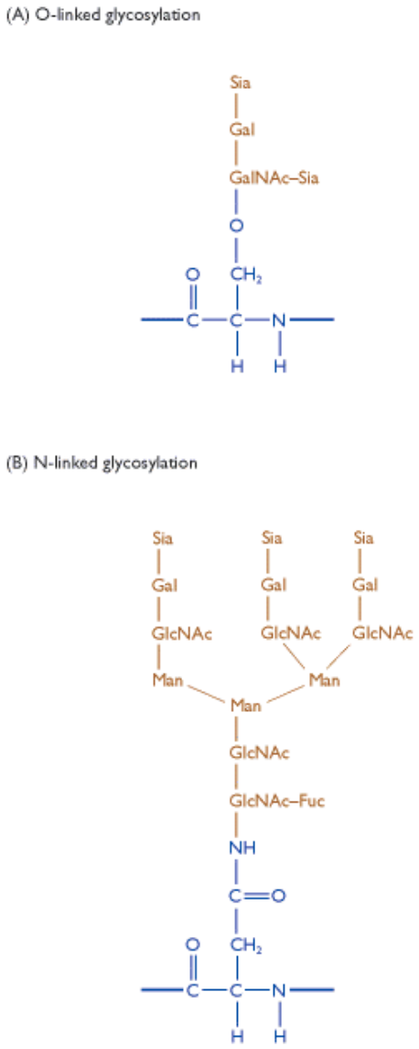

Figure 11.31. **Glycosylation**. (A) O-linked glycosylation. The structure shown is found in a number of glycoproteins. It is drawn here attached to a serine amino acid but it can also be linked to a threonine. (B) N-linked glycosylation usually results in larger sugar structures than are seen with O-linked glycosylation. The drawing shows a typical example of a complex glycan attached to an asparagine amino acid. Abbreviations: Fuc, fucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; Man, mannose; Sia, sialic acid.

11.3.4. Inteins

The final type of post-translational processing that we must consider is intein splicing, a protein version of the more extensive intron splicing that occurs with pre-RNAs. Inteins are internal segments of proteins that are removed soon after translation, the two external segments or exteins becoming linked together (*Figure 11.32*). The first intein was discovered in *S. cerevisiae* in 1990, and there have been only 100 confirmed identifications so far. Despite their scarcity, inteins are widespread. Most are known in bacteria and archaea but there are also examples in lower eukaryotes. In a few cases there is more than one intein in a single protein.

Most inteins are approximately 150 amino acids in length and, like pre-mRNA introns (Section 10.1.3), the sequences at the splice junctions of inteins have some similarity in most of the known examples. In particular, the first amino acid of the downstream extein is cysteine, serine or threonine. A few other amino acids within the intein sequence are also conserved. These conserved amino acids are involved in the splicing process, which is self-catalyzed by the intein itself (Paulus, 2000).

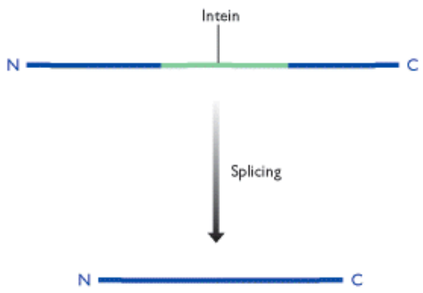

Figure 11.32. Intein splicing

Two interesting features of inteins have recently come to light. The first of these was discovered when the structures of two inteins were determined by X-ray crystallography (Duan *et al.*, 1997; Klabunde *et al.*, 1998). These structures are similar in some respects to that of a *Drosophila* protein called Hedgehog, which is involved in development of the segmentation pattern of the fly embryo. Hedgehog is an autoprocessing protein that cuts itself in two. The structural similarity with inteins lies in the part of the Hedgehog protein that catalyzes its self-cleavage. Possibly the same protein structure has evolved twice, or possibly inteins and Hedgehog shared a common link at some stage in the evolutionary past.

The second interesting feature is that with some inteins the excised segment is a sequence-specific endonuclease. The intein cuts DNA at the sequence corresponding to its insertion site in a gene coding for an intein-free version of the protein from which it is derived (*Figure 11.33*). If the cell also contains a gene coding for the intein-containing protein, then the DNA sequence for the intein is able to jump into the cut site, converting the intein-minus gene into an intein-plus version, a process called intein homing (Pietrokovski, 2001). The same type of event occurs with some Group I introns (Section 10.2.3), which code for proteins that direct intron homing. It is possible that transfer of inteins and Group I introns might also occur between cells or even between species (Cooper and Stevens, 1995). This is thought to be a mechanism by which selfish DNA is able to propagate (see Box 15.3).

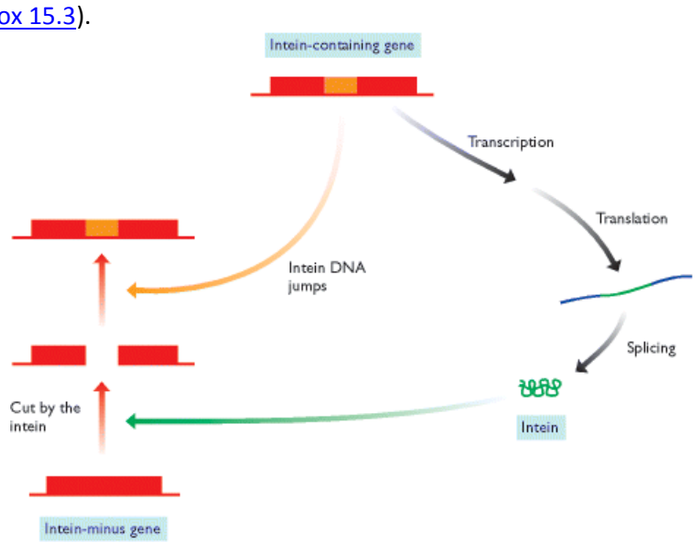

Figure 11.33. **Intein homing**. The cell is heterozygous for the intein-containing gene, possessing one allele with the intein and one allele without the intein. After protein splicing, the intein cuts the intein-minus gene at the appropriate place, allowing a copy of the intein DNA sequence to jump into this gene, converting it into the intein-plus version.

11.4. Protein Degradation

The protein synthesis and processing events that we have studied so far in this chapter result in new active proteins that take up their place in the cell's proteome. These proteins either replace existing ones that have reached the end of their working lives or provide new protein functions in response to the changing requirements of the cell. The concept that the proteome of a cell can change over time requires not only *de novo* protein synthesis but also the removal of proteins whose functions are no longer required. This removal must be highly selective so that only the correct proteins are degraded, and must also be rapid in order to account for the abrupt changes that occur under certain conditions, for example during key transitions in the cell cycle (Hunt, 1997).

For many years, protein degradation was an unfashionable subject and it was not until the 1990s that real progress was made in understanding how specific proteolysis events are linked with processes such as the cell cycle and differentiation. Even now, our knowledge centers largely on descriptions of general protein breakdown pathways and less on the regulation of the pathways and the mechanisms used to target specific proteins. There appear to be a number of different types of breakdown pathway whose interconnectivities have not yet been traced. This is particularly true in bacteria, which seem to have a range of proteases that work together in controlled degradation of proteins. In eukaryotes, most breakdown involves a single system, involving ubiquitin and the proteasome.

A link between ubiquitin and protein degradation was first established in 1975 when it was shown that this abundant 76-amino-acid protein is involved in energy-dependent proteolysis reactions in rabbit cells (Varshavsky, 1997). Subsequent research identified a series of three enzymes that attach ubiquitin molecules, singly or in chains, to lysine amino acids in proteins that are targeted for breakdown. Whether or not a protein becomes ubiquitinated depends on the presence or absence within it of amino acid motifs that act as degradation-susceptibility signals. These signals have not been completely characterized but there are thought to be at least ten different types in S. *cerevisiae*, including:

- **The N-degron**, a sequence element present at the N terminus of a protein;
- PEST sequences, internal sequences that are rich in proline (P), glutamic acid (E), serine (S) and threonine (T).

These sequences are permanent features of the proteins that contain them and so cannot be straightforward ‘degradation signals': if they were then these proteins would be broken down as soon as they are synthesized. Instead, they must determine susceptibility to degradation and hence the general stability of a protein in the cell. How this might be linked to the controlled breakdown of selected proteins at specific times, for instance during the cell cycle, is not yet clear.
The second component of the ubiquitin-dependent degradation pathway is the proteasome, the structure within which ubiquitinated proteins are broken down. In eukaryotes the proteasome is a large, multi-subunit structure with a sedimentation coefficient of 26S, comprising a hollow cylinder of 20S and two ‘caps' of 19S (Groll *et al.*, 1997; Ferrell *et al.*, 2000). Archaea also have proteasomes of about the same size but these are less complex, being composed of multiple copies of just two proteins; eukaryotic proteasomes contain 14 different types of protein subunit. The entrance into the cavity within the proteasome is narrow, and a protein must be unfolded before it can enter. This unfolding probably occurs through an energy-dependent process and may involve structures similar to chaperonins (Section 11.3.1) but with unfolding rather than folding activity (Lupas *et al.*, 1997). After unfolding, the protein can enter the proteasome within which it is cleaved into short peptides 4–10 amino acids in length. These are released back into the cytoplasm where they are broken down into individual amino acids which can be re-utilized in protein synthesis.

BOX 11.1. Translation in the archaea

In most respects, translation in the archaea more closely resembles the equivalent events in the eukaryotic cytoplasm rather than in bacteria. The one apparent exception is that the archaeal ribosome, at 70S, is comparable in size to the bacterial ribosome and, like bacterial ribosomes, contains 23S, 16S and 5S rRNAs. This apparent similarity is illusory because the archaeal rRNAs form base-paired secondary structures that are significantly different from the equivalent bacterial structures. The archaeal structures are also different from the eukaryotic versions, but the ribosomal proteins that attach to the rRNAs are homologs of the eukaryotic proteins. Archaeal mRNAs are capped and polyadenylated, and translation initiation is thought to involve a scanning process similar to that described for eukaryotic mRNAs. Archaeal tRNAs display a few unique features, including the absence of thymine in the so-called TΨC arm of the cloverleaf, and the presence at various positions of modified nucleotides not seen in either bacteria or eukaryotes. The methionine carried by the initiator tRNA is not *N*-formylated and the initiation and elongation factors resemble the eukaryotic molecules

Research Briefing 11.1 Peptidyl transferase is a ribozyme

A ribosome-associated protein that has the peptidyl transferase activity needed to synthesize peptide bonds during translation has never been isolated. The reason for this lack of success is now known: the enzyme activity is specified by part of the 23S rRNA.

When the base-paired structures of rRNAs (see *Figure 11.11*) were first determined in the early 1980s, the possibility that an RNA molecule could have enzymatic activity was unheard of, the breakthrough discoveries with regard to ribozymes not being made until the period 1982-86. Ribosomal RNAs were therefore initially assigned purely structural roles in the ribosome, their base-paired conformations being looked upon as scaffolds to which the important components of the ribosome - the proteins - were attached. Problems with this interpretation began to arise in the late 1980s when difficulties were encountered in identifying the protein or proteins responsible for the central catalytic activity of the ribosome - the formation of peptide bonds. By now the existence of ribozymes had been established and molecular biologists began to take seriously the possibility that rRNAs might have an enzymatic role in protein synthesis.

Locating the site of peptidyl transferase activity in the ribosome

Over the years, antibiotics and other inhibitors of protein synthesis have played an important role in studies of ribosome function. In 1995, a new inhibitor called CCdA-phosphate-puromycin was synthesized, this compound being an analog of the intermediate structure formed when two amino acids are joined by formation of a peptide bond during protein synthesis. CCdA-phosphate-puromycin binds tightly to the bacterial ribosome and, because of its structure, this binding site must be at precisely the position where peptide bonds are formed in the functioning ribosome. Would it be possible to use the inhibitor to find out where in the ribosome peptide bonds are made?

X-ray crystallography (Section 9.1.3) has revealed exactly where CCdA-phosphate-puromycin binds within the 50S subunit. Its position is deep down within the body of the subunit. The view shown here depicts the critical part of CCdA-phosphate-puromycin as a red dot, marking the position where the chemical reaction that creates a dipeptide must occur. This position is closely associated with the 23S rRNA of the large subunit (the rRNA is not shown in the figure) but is 18.4 Å away from the nearest protein, L3, and slightly more distant from L2, L4 and L10 (10 Å = 1 nm).

In atomic terms, 18–24 Å is a massive distance and it is inconceivable that any biochemical activity occurring at such a position could be catalyzed by one of the four proteins shown in the figure. The positioning of CCdA-phosphate-puromycin, and hence of the active site for peptide bond formation, provides convincing evidence that peptidyl transferase must be a ribozyme.

Permission to reproduce this figure in this web version of ***Genomes 2*** is either pending or has not been granted.

Now that the evidence has finally been obtained, researchers are moving on to determine exactly how the rRNA backbone acts as a ribozyme in peptide bond formation. Attention was initially concentrated on an adenine nucleotide at position 2451 in the *E. coli* 23S rRNA, because this adenine has unusual charge properties compared with other nucleotides. The hypothesis was that an interaction between this adenine and a nearby guanine, at position 2447, is the key to protein synthesis. But this model has been thrown into disarray by mutational studies, which have shown that both A2451 and G2447 can be replaced by other nucleotides without a detectable effect on the ability of the ribosome to carry out peptide bond synthesis.

These results have prompted a re-evaluation of the roles of A2451 and G2447 in peptide bond formation, and attention is now turning to other nucleotides present in the parts of the 23S rRNA that are located in the vicinity of the active site. Much work still needs to be done, but the ribozymal basis for peptidyl transferase activity is gradually being tracked down.
References

P. Nissen, J. Hansen, N. Ban, PB. Moore, and TA. Steitz. (2000). The structural basis of ribosome activity in peptide bond synthesis *Science* 289: 920-930. (PubMed)

N. Polacek, M. Gaynor, A. Yassin, and AS. Mankin. (2001). Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide *Nature* 411: 498-501. (PubMed)

SELF STUDY QUESTIONS

1. Outline the terminology used to distinguish between isoaccepting tRNAs.
2. Draw the cloverleaf structure of a tRNA. Indicate the key features of the structure, including those parts of the molecule that form attachments with the amino acid and the codon.
3. Write a short essay on aminoacyl-tRNA synthetases. Make sure that you cover the following points: the two classes of synthetase; fidelity of aminoacylation; modification of attached amino acids after aminoacylation.
4. Draw a series of diagrams to illustrate the codon-anticodon interactions that occur during wobble involving G-U base pairs and inosine.
5. Compare and contrast the usage of wobble during translation of mRNAs in (a) *Escherichia coli*, and (b) humans.
6. Construct a table giving details of the RNA and protein components of bacterial and eukaryotic ribosomes.
7. List the various techniques that have been used to study the structure of the bacterial ribosome. What are the applications and limitations of each technique?
8. Give a detailed description of initiation of translation in (a) *Escherichia coli*, and (b) eukaryotes.
9. How are eukaryotic mRNAs that lack a cap structure translated?
10. How is translation initiation regulated?
11. Give a detailed description of the elongation phase of translation in bacteria and eukaryotes.
12. Using examples, distinguish between the terms ‘frameshifting', ‘programmed frameshifting', ‘slippage' and ‘translational bypassing'.
13. Outline the roles of the three release factors and the ribosome recycling factor during termination of translation in *Escherichia coli*. Which proteins play the equivalent roles during termination in eukaryotes?
14. Describe the experiments which showed that a small protein such as ribonuclease can fold spontaneously *in vitro*. Why are larger proteins unable to fold spontaneously?

15. Distinguish between the activities of Hsp70 chaperones and chaperonins in protein folding.
16. Give examples of proteins that are processed by proteolytic cleavage.
17. Construct a table listing, with examples, the various types of post-translational chemical modification that occur with different proteins.
18. What is an intein? How are inteins spliced?
19. Describe the processes thought to be responsible for protein degradation in eukaryotes.

Problem-based learning

1. Why are there two classes of aminoacyl-tRNA synthetases? (*Hint: a good starting point for tackling this difficult problem is de Pouplana LR and Schimmel P [2001] Aminoacyl-tRNA synthetases: potential markers of genetic code development.* Trends Biochem. Sci.,***26****, 591–596.*)
2. To what extent have studies of ribosome structure been of value in understanding the detailed process by which proteins are synthesized?
3. Evaluate the information suggesting that peptidyl transferase is a ribozyme.
4. It is thought that translational bypassing ‘is controlled by the tRNA attached to the growing polypeptide, which scans the mRNA as the ribosome tracks along, and halts the bypass when a new codon to which it can base-pair is reached.' Devise an experiment to test this hypothesis.
5. Are protein folding studies that are conducted *in vitro* good models for protein folding *in vivo*?
6. Are the similarities between inteins and introns purely coincidental?
7. Using the current information on protein degradation, devise a hypothesis to explain how specific proteins could be individually degraded. Can your hypothesis be tested?