USING THE GENETIC CODE

2.7.1 Introduction

The sequence of a coding strand of DNA, read in the direction from 5′ to 3′, consists of nucleotide triplets (codons) corresponding to the amino acid sequence of a protein read from N-terminus to C-terminus. Sequencing of DNA and proteins makes it possible to compare corresponding nucleotide and amino acid sequences directly. There are 64 codons (each of 4 possible nucleotides can occupy each of the three positions of the codon, making $4^3 = 64$ possible trinucleotide sequences). Each of these codons has a specific meaning in protein synthesis: 61 codons represent amino acids; 3 codons cause the termination of protein synthesis.

The meaning of a codon that represents an amino acid is determined by the tRNA that corresponds to it; the meaning of the termination codons is determined directly by protein factors.

The breaking of the genetic code originally showed that genetic information is stored in the form of nucleotide triplets, but did not reveal how each codon specifies its corresponding amino acid. Before the advent of sequencing, codon assignments were deduced on the basis of two types of *in vitro* studies. A system involving the translation of synthetic polynucleotides was introduced in 1961, when Nirenberg showed that polyuridylic acid [poly(U)] directs the assembly of phenylalanine into polyphenylalanine. This result means that UUU must be a codon for phenylalanine. A second system was later introduced in which a trinucleotide was used to mimic a codon, thus causing the corresponding aminoacyl-tRNA to bind to a ribosome. By identifying the amino acid component of the aminoacyl-tRNA, the meaning of the codon can be found. The two techniques together assigned meaning to all of the codons that represent amino acids (418; 423).

61 of the 64 codons represent amino acids. The other three cause termination of protein synthesis. The assignment of amino acids to codons is not random, but shows relationships in which the third base has less effect on codon meaning; also related amino acids are often represented by related codons.

References

418. Nirenberg, M. W. and Matthaei, H. J. (1961). *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides*. Proc. Natl. Acad. Sci. USA 47, 1588-1602.

423. Nirenberg, M. W. and Leder, P. (1964). *The effect of trinucleotides upon the binding of sRNA to ribosomes*. Science 145, 1399-1407.

USING THE GENETIC CODE

2.7.2 Related codons represent related amino acids

Key Terms

Synonym codons have the same meaning in the genetic code. Synonym tRNAs bear the same amino acid and respond to the same codon.

Third base degeneracy describes the lesser effect on codon meaning of the nucleotide present in the third codon position.

A **stop codon (Termination codon)** is one of three triplets (UAG, UAA, UGA) that causes protein synthesis to terminate. They are also known historically as *nonsense codons*. The UAA codon is called ochre, and the UAA codon is called amber, after the names of the nonsense mutations by which they were originally identified.

Key Concepts

- 61 of the 64 possible triplets code for 20 amino acids.
- 3 codons do not represent amino acids and cause termination.
- The genetic code was frozen at an early stage of evolution and is universal.
- Most amino acids are represented by more than one codon.
- The multiple codons for an amino acid are usually related.
- Related amino acids often have related codons, minimizing the effects of mutation.

The code is summarized in **Figure 7.1**. Because there are more codons (61) than there are amino acids (20), almost all amino acids are represented by more than one codon. The only exceptions are methionine and tryptophan. Codons that have the same meaning are called **synonyms**. Because the genetic code is actually read on the mRNA, usually it is described in terms of the four bases present in RNA: U, C, A, and G.

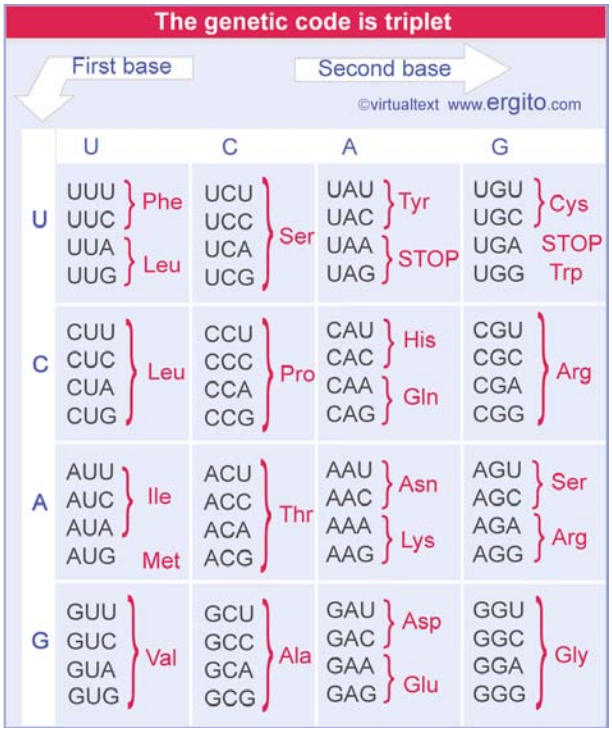

Figure 7.1 All the triplet codons have meaning: 61 represent amino acids, and 3 cause termination (STOP).

Codons representing the same or related amino acids tend to be similar in sequence. Often the base in the third position of a codon is not significant, because the four codons differing only in the third base represent the same amino acid. Sometimes a distinction is made only between a purine versus a pyrimidine in this position. The reduced specificity at the last position is known as **third base degeneracy**.

The interpretation of a codon requires base pairing with the anticodon of the corresponding aminoacyl-tRNA. The reaction occurs within the ribosome: complementary trinucleotides in isolation would usually be too short to pair in a stable manner, but the interaction is stabilized by the environment of the ribosomal A site. Also, base pairing between codon and anticodon is not solely a matter of A·U and G·C base pairing. The ribosome controls the environment in such a way that conventional pairing occurs at the first two positions of the codon, but additional reactions are permitted at the third base. As a result, a single aminoacyl-tRNA may recognize more than one codon, corresponding with the pattern of degeneracy. Furthermore, pairing interactions may also be influenced by the introduction of special bases into tRNA, especially by modification in or close to the anticodon.

The tendency for similar amino acids to be represented by related codons minimizes the effects of mutations. It increases the probability that a single random base change will result in no amino acid substitution or in one involving amino acids of similar character. For example, a mutation of CUC to CUG has no effect, since both codons represent leucine; and a mutation of CUU to AUU results in replacement of leucine with isoleucine, a closely related amino acid.

Figure 7.2 plots the number of codons representing each amino acid against the

frequency with which the amino acid is used in proteins (in *E. coli*). There is only a slight tendency for amino acids that are more common to be represented by more codons, and therefore it does not seem that the genetic code has been optimized with regard to the utilization of amino acids.

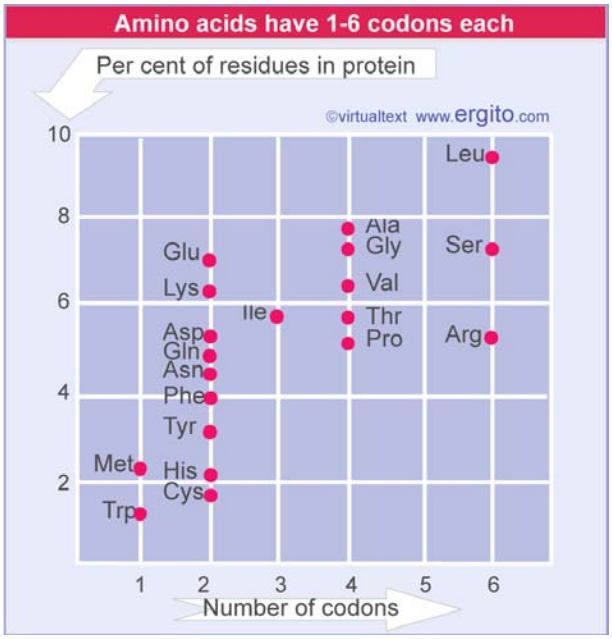

Figure 7.2 The number of codons for each amino acid does not correlate closely with its frequency of use in proteins.

The three codons (UAA, UAG, and UGA) that do not represent amino acids are used specifically to terminate protein synthesis. One of these **stop codons** marks the end of every gene.

Is the genetic code the same in all living organisms?

Comparisons of DNA sequences with the corresponding protein sequences reveal that the identical set of codon assignments is used in bacteria and in eukaryotic cytoplasm. As a result, mRNA from one species usually can be translated correctly *in vitro* or *in vivo* by the protein synthetic apparatus of another species. So the codons used in the mRNA of one species have the same meaning for the ribosomes and tRNAs of other species.

The universality of the code argues that it must have been established very early in evolution. Perhaps the code started in a primitive form in which a small number of codons were used to represent comparatively few amino acids, possibly even with one codon corresponding to any member of a group of amino acids. More precise codon meanings and additional amino acids could have been introduced later. One possibility is that at first only two of the three bases in each codon were used; discrimination at the third position could have evolved later. (Originally there might have been a stereochemical relationship between amino acids and the codons representing them. Then a more complex system evolved.)

Evolution of the code could have become "frozen" at a point at which the system had become so complex that any changes in codon meaning would disrupt existing proteins by substituting unacceptable amino acids. Its universality implies that this must have happened at such an early stage that all living organisms are descended from a single pool of primitive cells in which this occurred.

Exceptions to the universal genetic code are rare. Changes in meaning in the principal genome of a species usually concern the termination codons. For example, in a mycoplasma, UGA codes for tryptophan; and in certain species of the ciliates Tetrahymena and Paramecium, UAA and UAG code for glutamine. Systematic alterations of the code have occurred only in mitochondrial DNA (see *Molecular Biology 2.7.7 There are sporadic alterations of the universal code*).

USING THE GENETIC CODE

2.7.3 Codon-anticodon recognition involves wobbling

Key Terms

The **wobble hypothesis** accounts for the ability of a tRNA to recognize more than one codon by unusual (non-G·C, non-A·T) pairing with the third base of a codon.

Key Concepts

- Multiple codons that represent the same amino acid most often differ at the third base position.
- The wobble in pairing between the first base of the anticodon and the third base of the codon results from the structure of the anticodon loop.

The function of tRNA in protein synthesis is fulfilled when it recognizes the codon in the ribosomal A site. The interaction between anticodon and codon takes place by base pairing, but under rules that extend pairing beyond the usual G·C and A·U partnerships.

We can deduce the rules governing the interaction from the sequences of the anticodons that correspond to particular codons. The ability of any tRNA to respond to a given codon can be measured directly by the trinucleotide binding assay or by its use in an *in vitro* protein synthetic system.

The genetic code itself yields some important clues about the process of codon recognition. The pattern of third-base degeneracy is drawn in **Figure 7.3**, which shows that in almost all cases either the third base is irrelevant or a distinction is made only between purines and pyrimidines.

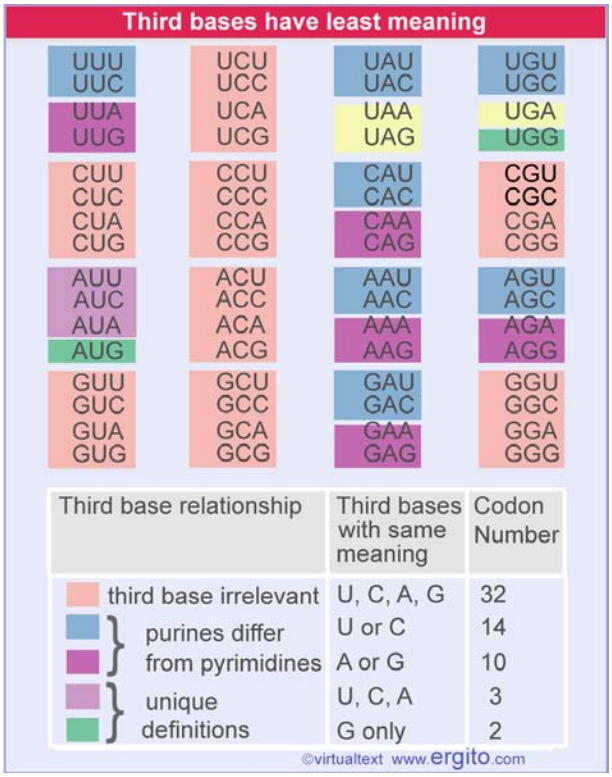

Figure 7.3 Third bases have the least influence on codon meanings. Boxes indicate groups of codons within which third-base degeneracy ensures that the meaning is the same.

There are eight codon families in which all four codons sharing the same first two bases have the same meaning, so that the third base has no role at all in specifying the amino acid. There are seven codon pairs in which the meaning is the same whichever pyrimidine is present at the third position; and there are five codon pairs in which either purine may be present without changing the amino acid that is coded.

There are only three cases in which a unique meaning is conferred by the presence of a particular base at the third position: AUG (for methionine), UGG (for tryptophan), and UGA (termination). So C and U never have a unique meaning in the third position, and A never signifies a unique amino acid.

Because the anticodon is complementary to the codon, it is the first base in the anticodon sequence written conventionally in the direction from 5′ to 3′ that pairs with the third base in the codon sequence written by the same convention. So the combination

Codon 5′ A C G 3′

Anticodon 3′ U G C 5′

is usually written as codon ACG/anticodon CGU, where the anticodon sequence must be read backward for complementarity with the codon.

To avoid confusion, we shall retain the usual convention in which all sequences are written 5′–3′, but indicate anticodon sequences with a backward arrow as a reminder of the relationship with the codon. So the codon/anticodon pair shown above will be written as ACG and CGU←, respectively.

Does each triplet codon demand its own tRNA with a complementary anticodon? Or can a single tRNA respond to both members of a codon pair and to all (or at least some) of the four members of a codon family?

Often one tRNA can recognize more than one codon. This means that the base in the first position of the anticodon must be able to partner alternative bases in the corresponding third position of the codon. Base pairing at this position cannot be limited to the usual G·C and A·U partnerships.

The rules governing the recognition patterns are summarized in the **wobble hypothesis**, which states that the pairing between codon and anticodon at the first two codon positions always follows the usual rules, but that exceptional "wobbles" occur at the third position. Wobbling occurs because the conformation of the tRNA anticodon loop permits flexibility at the first base of the anticodon (449). **Figure 7.4** shows that G·U pairs can form in addition to the usual pairs.

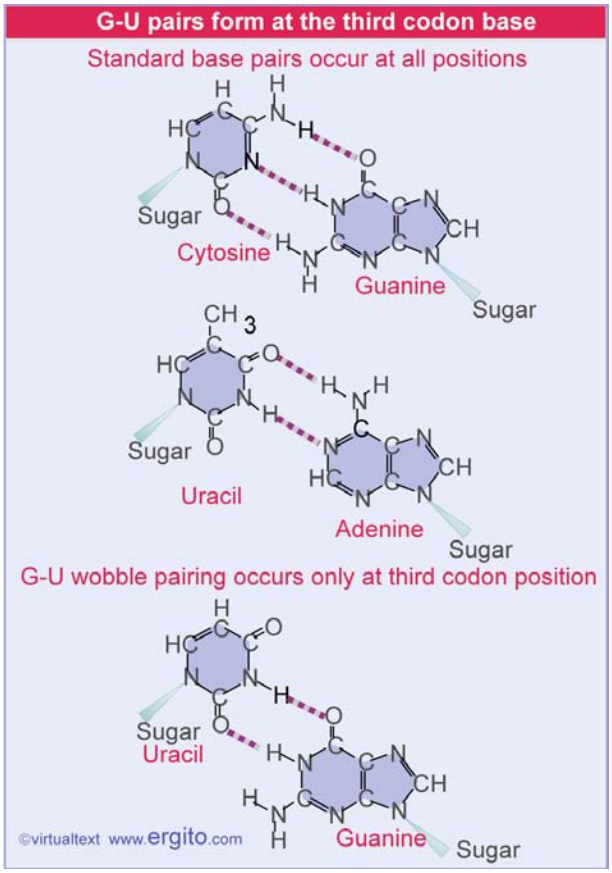

Figure 7.4 Wobble in base pairing allows G-U pairs to form between the third base of the codon and the first base of the anticodon.

This single change creates a pattern of base pairing in which A can no longer have a unique meaning in the codon (because the U that recognizes it must also recognize G). Similarly, C also no longer has a unique meaning (because the G that recognizes it also must recognize U). **Figure 7.5** summarizes the pattern of recognition.

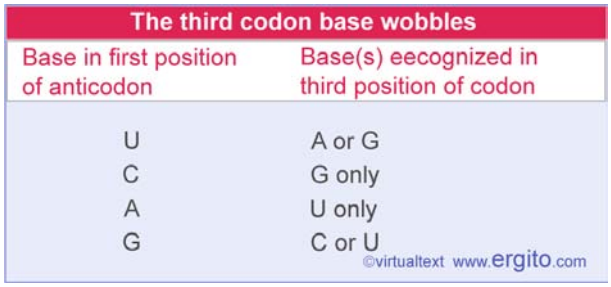

Figure 7.5 Codon-anticodon pairing involves wobbling at the third position.

It is therefore possible to recognize unique codons only when the third bases are G or U; this option is not used often, since UGG and AUG are the only examples of the first type, and there is none of the second type.

(G·U pairs are common in RNA duplex structures. But the formation of stable contacts between codon and anticodon, when only 3 base pairs can be formed, is more constrained, and thus G·U pairs can contribute only in the last position of the codon.)

References

449. Crick, F. H. C. (1966). *Codon-anticodon pairing: the wobble hypothesis*. J. Mol. Biol. 19, 548-555.

USING THE GENETIC CODE

2.7.4 tRNAs are processed from longer precursors

Key Concepts

- A mature tRNA is generated by processing a precursor.
- The 5′ end is generated by cleavage by the endonuclease RNAase P.
- The 3′ end is generated by cleavage followed by trimming of the last few bases, followed by addition of the common terminal trinucleotide sequence CCA.

tRNAs are commonly synthesized as precursor chains with additional material at one or both ends. **Figure 7.6** shows that the extra sequences are removed by combinations of endonucleolytic and exonucleolytic activities. One feature that is common to all tRNAs is that the three nucleotides at the 3′ terminus, always the triplet sequence CCA, are not coded in the genome, but are added as part of tRNA processing (for review see 3461).

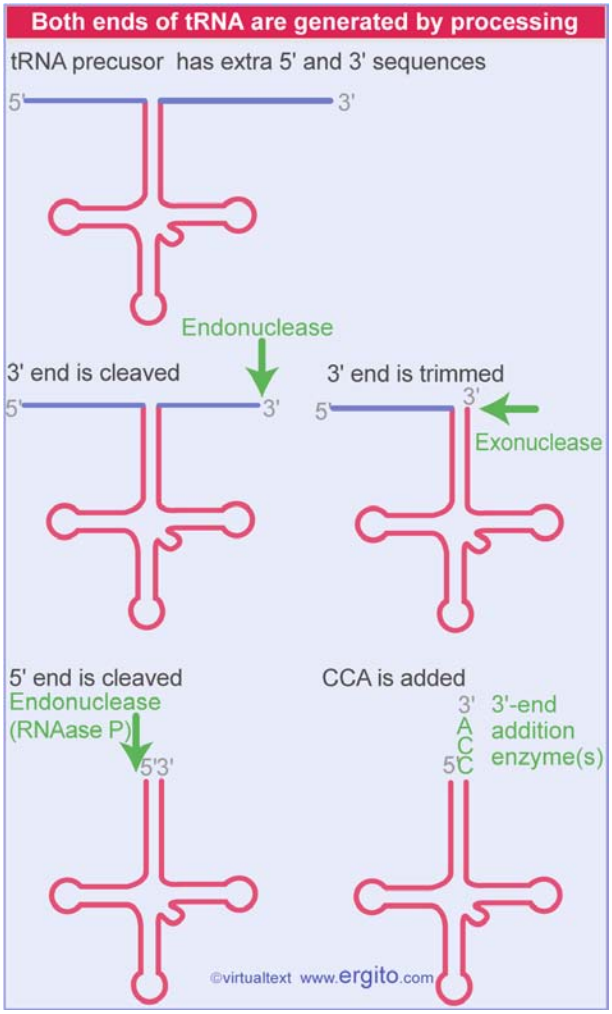

Figure 7.6 The tRNA 3 ′ end is generated by cutting and trimming followed by addition of CCA; the 5 ′ end is generated by cutting.

The 5 ′ end of tRNA is generated by a cleavage action catalyzed by the enzyme ribonuclease P.

The enzymes that process the 3 ′ end are best characterized in *E. coli*, where an endonuclease triggers the reaction by cleaving the precursor downstream, and several exonucleases then trim the end by degradation in the 3 ′ -5 ′ direction. The reaction also involves several enzymes in eukaryotes. It generates a tRNA that needs the CCA trinucleotide sequence to be added to the 3 ′ end.

The addition of CCA is the result solely of an enzymatic process, that is, the enzymatic activity carries the specificity for the sequence of the trinucleotide, which is not determined by a template. There are several models for the process, which may be different in different organisms.

In some organisms, the process is catalyzed by a single enzyme. One model for its action proposes that a single enzyme binds to the 3 ′ end, and sequentially adds C, C,

and A, the specificity at each stage being determined by the structure of the 3′ end. Other models propose that the enzyme has different active sites for CTP and ATP.

In other organisms, different enzymes are responsible for adding the C and A residues, and they function sequentially.

When a tRNA is not properly processed, it attracts the attention of a quality control system that degrades it. This ensures that the protein synthesis apparatus does not become blocked by nonfunctional tRNAs.

Last updated on 1-30-2003

Reviews

3461. Hopper, A. K. and Phizicky, E. M. (2003). *tRNA transfers to the limelight.* Genes Dev. 17, 162-180.

USING THE GENETIC CODE

2.7.5 tRNA contains modified bases

Key Terms

Modification of DNA or RNA includes all changes made to the nucleotides after their initial incorporation into the polynucleotide chain.

Key Concepts

- tRNAs contain >50 modified bases.
- Modification usually involves direct alteration of the primary bases in tRNA, but there are some exceptions in which a base is removed and replaced by another base.

Transfer RNA is unique among nucleic acids in its content of "unusual" bases. An unusual base is any purine or pyrimidine ring except the usual A, G, C, and U from which all RNAs are synthesized. All other bases are produced by **modification** of one of the four bases after it has been incorporated into the polyribonucleotide chain.

All classes of RNA display some degree of modification, but in all cases except tRNA this is confined to rather simple events, such as the addition of methyl groups. In tRNA, there is a vast range of modifications, ranging from simple methylation to wholesale restructuring of the purine ring. Modifications occur in all parts of the tRNA molecule There are >50 different types of modified bases in tRNA.

Figure 7.7 shows some of the more common modified bases. Modifications of pyrimidines (C and U) are less complex than those of purines (A and G). In addition to the modifications of the bases themselves, methylation at the 2′–O position of the ribose ring also occurs.

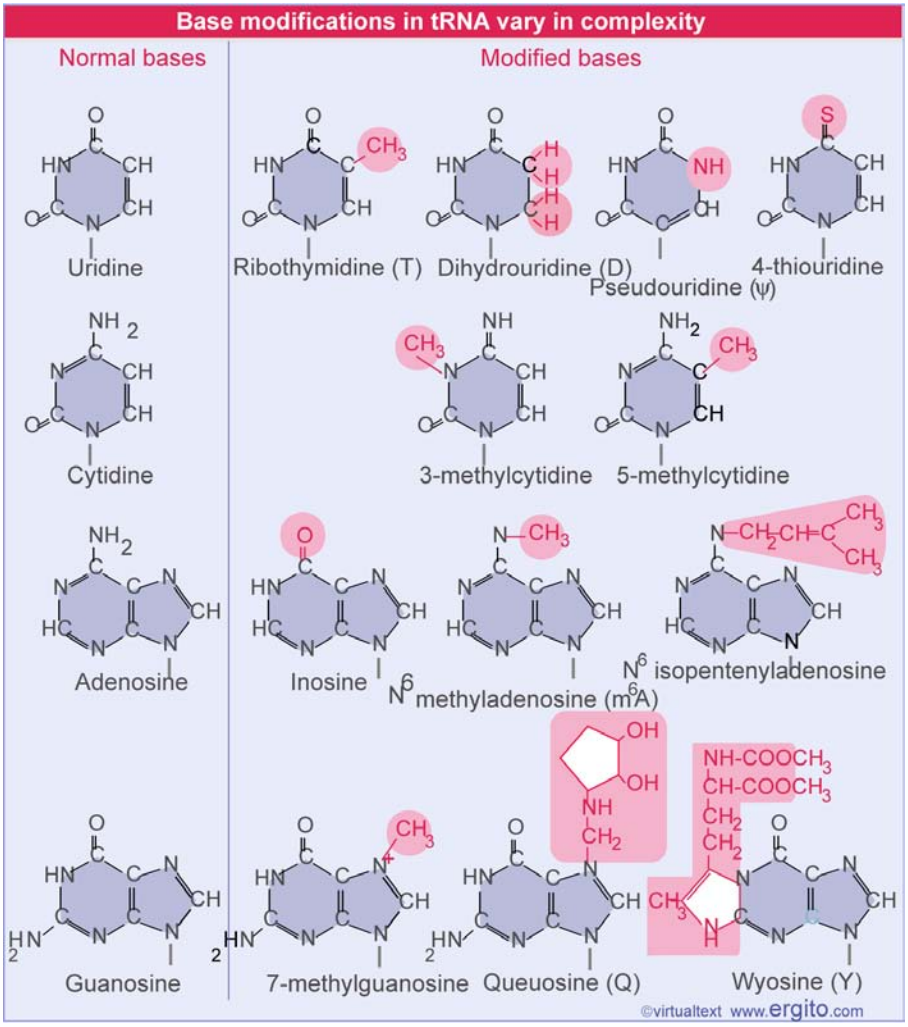

Figure 7.7 All of the four bases in tRNA can be modified.

The most common modifications of uridine are straightforward. Methylation at position 5 creates ribothymidine (T). The base is the same commonly found in DNA; but here it is attached to ribose, not deoxyribose. In RNA, thymine constitutes an unusual base, originating by modification of U.

Dihydrouridine (D) is generated by the saturation of a double bond, changing the ring structure. Pseudouridine (ψ) interchanges the positions of N and C atoms (see **Figure 24.40**). And 4-thiouridine has sulfur substituted for oxygen.

The nucleoside inosine is found normally in the cell as an intermediate in the purine biosynthetic pathway. However, it is not incorporated directly into RNA, where instead its existence depends on modification of A to create I. Other modifications of A include the addition of complex groups.

Two complex series of nucleotides depend on modification of G. The Q bases, such as queuosine, have an additional pentenyl ring added via an NH linkage to the methyl group of 7-methylguanosine. The pentenyl ring may carry various further groups.

The Y bases, such as wyosine, have an additional ring fused with the purine ring itself; the extra ring carries a long carbon chain, again to which further groups are added in different cases.

The modification reaction usually involves the alteration of, or addition to, existing bases in the tRNA. An exception is the synthesis of Q bases, where a special enzyme exchanges free queuosine with a guanosine residue in the tRNA. The reaction involves breaking and remaking bonds on either side of the nucleoside.

The modified nucleosides are synthesized by specific tRNA-modifying enzymes. The original nucleoside present at each position can be determined either by comparing the sequence of tRNA with that of its gene or (less efficiently) by isolating precursor molecules that lack some or all of the modifications. The sequences of precursors show that different modifications are introduced at different stages during the maturation of tRNA.

Some modifications are constant features of all tRNA molecules – for example, the D residues that give rise to the name of the D arm, and the ψ found in the T ψC sequence. On the 3′ side of the anticodon there is always a modified purine, although the modification varies widely.

Other modifications are specific for particular tRNAs or groups of tRNAs. For example, wyosine bases are characteristic of $tRNA^{Phe}$ in bacteria, yeast, and mammals. There are also some species-specific patterns.

The many tRNA-modifying enzymes (~60 in yeast) vary greatly in specificity (for review see 3461). In some cases, a single enzyme acts to make a particular modification at a single position. In other cases, an enzyme can modify bases at several different target positions. Some enzymes undertake single reactions with individual tRNAs; others have a range of substrate molecules. The features recognized by the tRNA-modifying enzymes are unknown, but probably involve recognition of structural features surrounding the site of modification. Some modifications require the successive actions of more than one enzyme.

Last updated on 1-30-2003

Reviews

3461. Hopper, A. K. and Phizicky, E. M. (2003). *tRNA transfers to the limelight.* Genes Dev. 17, 162-180.

USING THE GENETIC CODE

2.7.6 Modified bases affect anticodon-codon pairing

Key Concepts

- Modifications in the anticodon affect the pattern of wobble pairing and therefore are important in determining tRNA specificity.

The most direct effect of modification is seen in the anticodon, where change of sequence influences the ability to pair with the codon, thus determining the meaning of the tRNA. Modifications elsewhere in the vicinity of the anticodon also influence its pairing.

When bases in the anticodon are modified, further pairing patterns become possible in addition to those predicted by the regular and wobble pairing involving A, C, U, and G. **Figure 7.8** shows the use of inosine (I), which is often present at the first position of the anticodon. Inosine can pair with any one of three bases, U, C, and A.

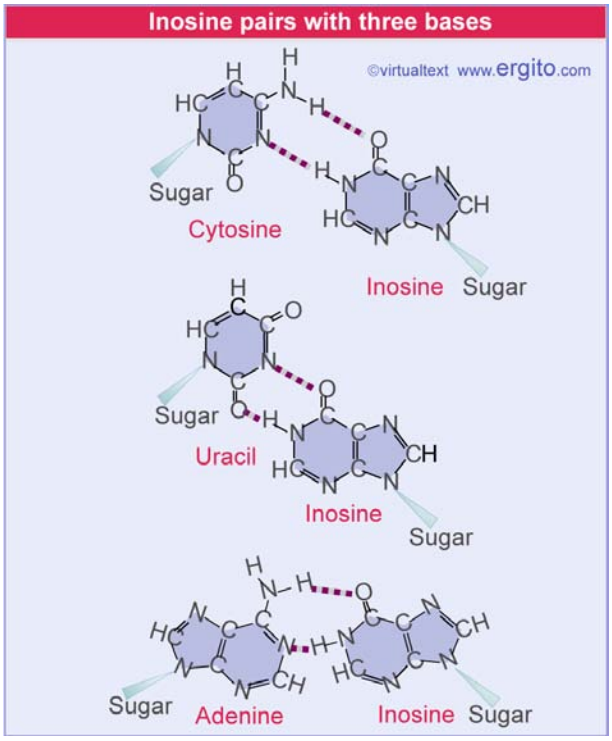

Figure 7.8 Inosine can pair with any of U, C, and A.

This ability is especially important in the isoleucine codons, where AUA codes for isoleucine, while AUG codes for methionine. Because with the usual bases it is not possible to recognize A alone in the third position, any tRNA with U starting its anticodon would have to recognize AUG as well as AUA. So AUA must be read together with AUU and AUC, a problem that is solved by the existence of tRNA

with I in the anticodon.

Actually, some of the predicted regular combinations do not occur, because some bases are always modified. There seems to be an absolute ban on the employment of A; usually it is converted to I. And U at the first position of the anticodon is usually converted to a modified form that has altered pairing properties.

Some modifications create preferential readings of some codons with respect to others. Anticodons with uridine-5-oxyacetic acid and 5-methoxyuridine in the first position recognize A and G efficiently as third bases of the codon, but recognize U less efficiently. Another case in which multiple pairings can occur, but with some preferred to others, is provided by the series of queuosine and its derivatives. These modified G bases continue to recognize both C and U, but pair with U more readily.

A restriction not allowed by the usual rules can be achieved by the employment of 2-thiouridine in the anticodon. **Figure 7.9** shows that its modification allows the base to continue to pair with A, but prevents it from indulging in wobble pairing with G (for review see 32).

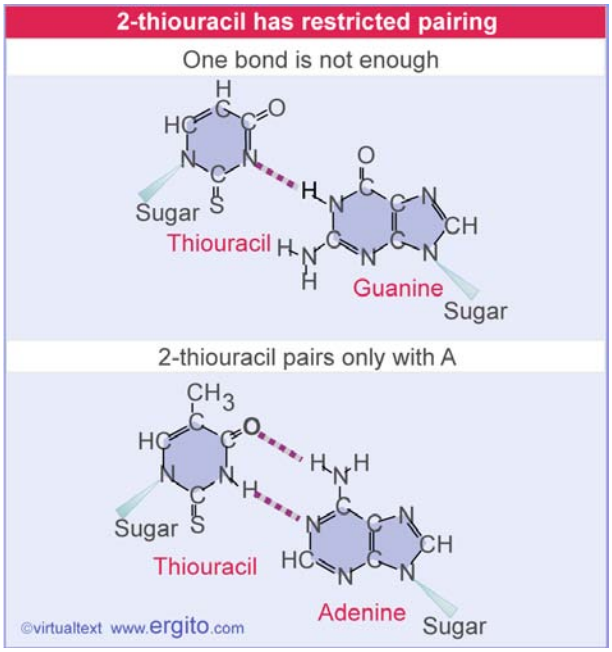

Figure 7.9 Modification to 2-thiouridine restricts pairing to A alone because only one H-bond can form with G.

These and other pairing relationships make the general point that there are multiple ways to construct a set of tRNAs able to recognize all the 61 codons representing amino acids. No particular pattern predominates in any given organism, although the absence of a certain pathway for modification can prevent the use of some recognition patterns. So a particular codon family is read by tRNAs with different anticodons in different organisms.

Often the tRNAs will have overlapping responses, so that a particular codon is read by more than one tRNA. In such cases there may be differences in the efficiencies of

the alternative recognition reactions. (As a general rule, codons that are commonly used tend to be more efficiently read.) And in addition to the construction of a set of tRNAs able to recognize all the codons, there may be multiple tRNAs that respond to the same codons.

The predictions of wobble pairing accord very well with the observed abilities of almost all tRNAs. But there are exceptions in which the codons recognized by a tRNA differ from those predicted by the wobble rules. Such effects probably result from the influence of neighboring bases and/or the conformation of the anticodon loop in the overall tertiary structure of the tRNA. Indeed, the importance of the structure of the anticodon loop is inherent in the idea of the wobble hypothesis itself. Further support for the influence of the surrounding structure is provided by the isolation of occasional mutants in which a change in a base in some other region of the molecule alters the ability of the anticodon to recognize codons.

Another unexpected pairing reaction is presented by the ability of the bacterial initiator, fMet-tRNA$_f$, to recognize both AUG and GUG. This misbehavior involves the third base of the anticodon.

Reviews

32. Bjork, G. R (1987). *Transfer RNA modification.* Annu. Rev. Biochem. 56, 263-287.

USING THE GENETIC CODE

2.7.7 There are sporadic alterations of the universal code

Key Concepts

- Changes in the universal genetic code have occurred in some species.
- They are more common in mitochondrial genomes, where a phylogenetic tree can be constructed for the changes.
- In nuclear genomes, they are sporadic and usually affect only termination codons.

The universality of the genetic code is striking, but some exceptions exist. They tend to affect the codons involved in initiation or termination and result from the production (or absence) of tRNAs representing certain codons. The changes found in principal (bacterial or nuclear) genomes are summarized in **Figure 7.10**.

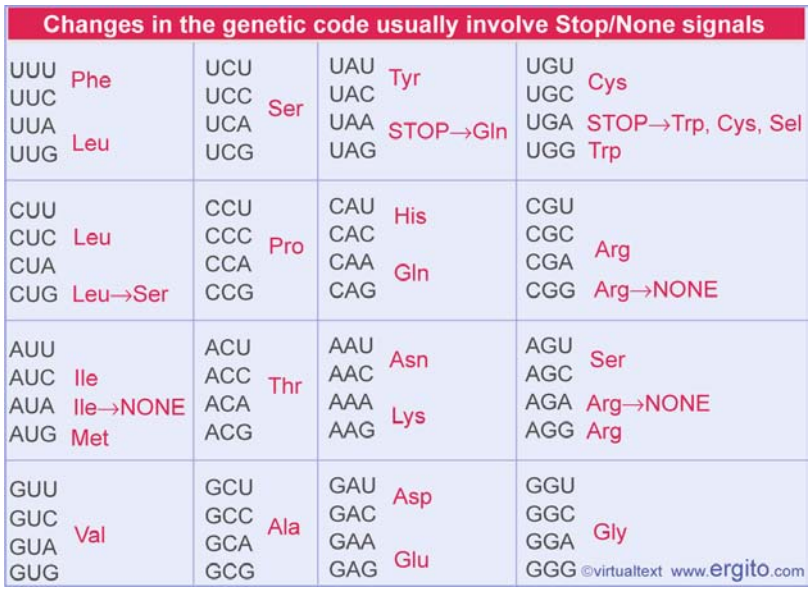

Figure 7.10 Changes in the genetic code in bacterial or eukaryotic nuclear genomes usually assign amino acids to stop codons or change a codon so that it no longer specifies an amino acid. A change in meaning from one amino acid to another is unusual.

Almost all of the changes that allow a codon to represent an amino acid affect termination codons:

- In the prokaryote *Mycoplasma capricolum*, UGA is not used for termination, but instead codes for tryptophan. In fact, it is the predominant Trp codon, and UGG is used only rarely. Two Trp-tRNA species exist, with the anticodons UCA ←

(reads UGA and UGG) and CCA$^{\leftarrow}$ (reads only UGG).

- Some ciliates (unicellular protozoa) read UAA and UAG as glutamine instead of termination signals. *Tetrahymena thermophila*, one of the ciliates, contains three $tRNA^{Glu}$ species. One recognizes the usual codons CAA and CAG for glutamine, one recognizes both UAA and UAG (in accordance with the wobble hypothesis), and the last recognizes only UAG. We assume that a further change is that the release factor eRF has a restricted specificity, compared with that of other eukaryotes.

- In another ciliate (*Euplotes octacarinatus*), UGA codes for cysteine. Only UAA is used as a termination codon, and UAG is not found. The change in meaning of UGA might be accomplished by a modification in the anticodon of $tRNA^{Cys}$ to allow it to read UGA with the usual codons UGU and UGC.

- The only substitution in coding for amino acids occurs in a yeast (*Candida*), where CUG means serine instead of leucine (and UAG is used as a sense codon).

Acquisition of a coding function by a termination codon requires two types of change: a tRNA must be mutated so as to recognize the codon; and the class 1 release factor must be mutated so that it does not terminate at this codon.

The other common type of change is loss of the tRNA that responds to a codon, so that the codon no longer specifies any amino acid. What happens at such a codon will depend on whether the termination factor evolves to recognize it.

All of these changes are sporadic, which is to say that they appear to have occurred independently in specific lines of evolution. They may be concentrated on termination codons, because these changes do not involve substitution of one amino acid for another. Once the genetic code was established, early in evolution, any general change in the meaning of a codon would cause a substitution in all the proteins that contain that amino acid. It seems likely that the change would be deleterious in at least some of these proteins, with the result that it would be strongly selected against. The divergent uses of the termination codons could represent their "capture" for normal coding purposes. If some termination codons were used only rarely, they could be recruited to coding purposes by changes that allowed tRNAs to recognize them.

Exceptions to the universal genetic code also occur in the mitochondria from several species. **Figure 7.11** constructs a phylogeny for the changes. It suggests that there was a universal code that was changed at various points in mitochondrial evolution. The earliest change was the employment of UGA to code for tryptophan, which is common to all (non-plant) mitochondria (for review see 41).

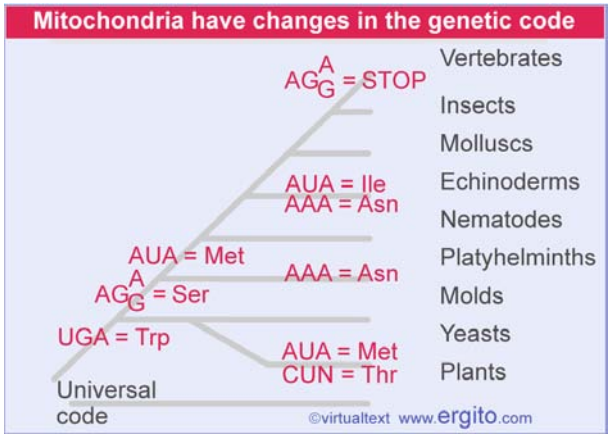

Figure 7.11 Changes in the genetic code in mitochondria can be traced in phylogeny. The minimum number of independent changes is generated by supposing that the AUA=Met and the AAA=Asn changes each occurred independently twice, and that the early AUA=Met change was reversed in echinoderms.

Some of these changes make the code simpler, by replacing two codons that had different meanings with a pair that has a single meaning. Pairs treated like this include UGG and UGA (both Trp instead of one Trp and one termination) and AUG and AUA (both Met instead of one Met and the other Ile).

Why have changes been able to evolve in the mitochondrial code? Because the mitochondrion synthesizes only a small number of proteins (~10), the problem of disruption by changes in meaning is much less severe. Probably the codons that are altered were not used extensively in locations where amino acid substitutions would have been deleterious. The variety of changes found in mitochondria of different species suggests that they have evolved separately, and not by common descent from an ancestral mitochondrial code.

According to the wobble hypothesis, a minimum of 31 tRNAs (excluding the initiator) are required to recognize all 61 codons (at least 2 tRNAs are required for each codon family and 1 tRNA is needed per codon pair or single codon). But an unusual situation exists in (at least) mammalian mitochondria in which there are only 22 different tRNAs. How does this limited set of tRNAs accommodate all the codons?

The critical feature lies in a simplification of codon-anticodon pairing, in which one tRNA recognizes all four members of a codon family. This reduces to 23 the minimum number of tRNAs required to respond to all usual codons. The use of AG^{A}_{G} for termination reduces the requirement by one further tRNA, to 22.

In all eight codon families, the sequence of the tRNA contains an unmodified U at the first position of the anticodon. The remaining codons are grouped into pairs in which all the codons ending in pyrimidines are read by G in the anticodon, and all the codons ending in purines are read by a modified U in the anticodon, as predicted by the wobble hypothesis. The complication of the single UGG codon is avoided by the change in the code to read UGA with UGG as tryptophan; and in mammals, AUA ceases to represent isoleucine and instead is read with AUG as methionine.

This allows all the nonfamily codons to be read as 14 pairs.

The 22 identified tRNA genes therefore code for 14 tRNAs representing pairs, and 8 tRNAs representing families. This leaves the two usual termination codons UAG and UAA unrecognized by tRNA, together with the codon pair AG^{A}_{G}. Similar rules are followed in the mitochondria of fungi (for review see 33).

Last updated on 12-17-2001

Reviews

33. Fox, T. D. (1987). *Natural variation in the genetic code.* Annu. Rev. Genet. 21, 67-91.

41. Osawa, S. et al. (1992). *Recent evidence for evolution of the genetic code.* Microbiol. Rev. 56, 229-264.

USING THE GENETIC CODE

2.7.8 Novel amino acids can be inserted at certain stop codons

Key Concepts

- Changes in the reading of specific codons can occur in individual genes.
- The insertion of seleno-Cys-tRNA at certain UGA codons requires several proteins to modify the Cys-tRNA and insert it into the ribosome.
- Pyrrolysine can be inserted at certain UAG codons.

Specific changes in reading the code occur in individual genes. The specificity of such changes implies that the reading of the particular codon must be influenced by the surrounding bases.

A striking example is the incorporation of the modified amino acid seleno-cysteine at certain UGA codons within the genes that code for selenoproteins in both prokaryotes and eukaryotes. Usually these proteins catalyze oxidation-reduction reactions, and contain a single seleno-cysteine residue, which forms part of the active site. The most is known about the use of the UGA codons in three *E. coli* genes coding for formate dehydrogenase isozymes. The internal UGA codon is read by a seleno-Cys-tRNA. This unusual reaction is determined by the local secondary structure of mRNA, in particular by the presence of a hairpin loop downstream of the UGA.

Mutations in 4 *sel* genes create a deficiency in selenoprotein synthesis. *selC* codes for tRNA (with the anticodon ACU$^{\leftarrow}$) that is charged with serine. *selA* and *selD* are required to modify the serine to seleno-cysteine. SelB is an alternative elongation factor. It is a guanine nucleotide-binding protein that acts as a specific translation factor for entry of seleno-Cys-tRNA into the A site; it thus provides (for this single tRNA) a replacement for factor EF-Tu. The sequence of SelB is related to both EF-Tu and IF-2 (for review see 39).

Why is seleno-Cys-tRNA inserted only at certain UGA codons? These codons are followed by a stem-loop structure in the mRNA. **Figure 7.12** shows that the stem of this structure is recognized by an additional domain in SelB (one that is not present in EF-Tu or IF-2). A similar mechanism interprets some UGA codons in mammalian cells, except that two proteins are required to identify the appropriate UGA codons. One protein (SBP2) binds a stem-loop structure far downstream from the UGA codon, while the counterpart of SelB (called SECIS) binds to SBP2 and simultaneously binds the tRNA to the UGA codon (1186).

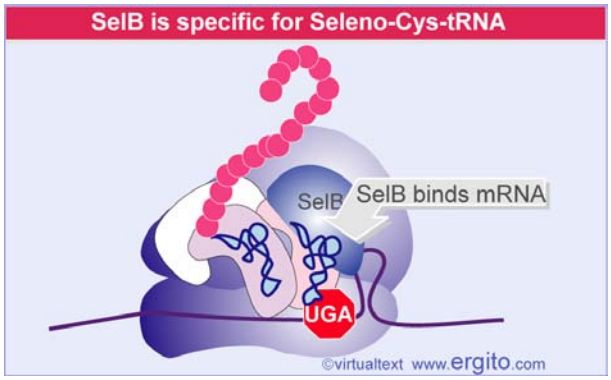

Figure 7.12 SelB is an elongation factor that specifically binds Seleno-Cys-tRNA to a UGA codon that is followed by a stem-loop structure in mRNA.

Another example of the insertion of a special amino acid is the placement of pyrrolysine at a UAG codon. This happens in both an archaea and a bacterium (2492; 2493). The mechanism is probably similar to the insertion of seleno-cysteine. An unusual tRNA is charged with lysine, which is presumably then modified. The tRNA has a CUA anticodon, which responds to UAG. There must be other components of the system that restricts its response to the appropriate UAG codons.

Last updated on 5-28-2002

Reviews

39. Bock, A. (1991). *Selenoprotein synthesis: an expansion of the genetic code.* Trends Biochem. Sci. 16, 463-467.

References

1186. Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., and Krol, A. (2000). *Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation.* EMBO J. 19, 4796-4805.

2492. Srinivasan, G., James, C. M., and Krzycki, J. A. (2002). *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.* Science 296, 1459-1462.

2493. Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002). *A new UAG-encoded residue in the structure of a methanogen methyltransferase.* Science 296, 1462-1466.

USING THE GENETIC CODE

2.7.9 tRNAs are charged with amino acids by synthetases

Key Terms

Cognate tRNAs (Isoaccepting tRNA) are those recognized by a particular aminoacyl-tRNA synthetase. They all are charged with the same amino acid.

Key Concepts

- Aminoacyl-tRNA synthetases are enzymes that charge tRNA with an amino acid to generate aminoacyl-tRNA in a two-stage reaction that uses energy from ATP.
- There are 20 aminoacyl-tRNA synthetases in each cell. Each charges all the tRNAs that represent a particular amino acid.
- Recognition of a tRNA is based on a small number of points of contact in the tRNA sequence.

It is necessary for tRNAs to have certain characteristics in common, yet be distinguished by others. The crucial feature that confers this capacity is the ability of tRNA to fold into a specific tertiary structure. Changes in the details of this structure, such as the angle of the two arms of the "L" or the protrusion of individual bases, may distinguish the individual tRNAs.

All tRNAs can fit in the P and A sites of the ribosome, where at one end they are associated with mRNA via codon-anticodon pairing, while at the other end the polypeptide is being transferred. Similarly, all tRNAs (except the initiator) share the ability to be recognized by the translation factors (EF-Tu or eEF1) for binding to the ribosome. The initiator tRNA is recognized instead by IF-2 or eIF2. So the tRNA set must possess common features for interaction with elongation factors, but the initiator tRNA can be distinguished.

Amino acids enter the protein synthesis pathway through the aminoacyl-tRNA synthetases, which provide the interface for connection with nucleic acid. All synthetases function by the two-step mechanism depicted in **Figure 7.13**:

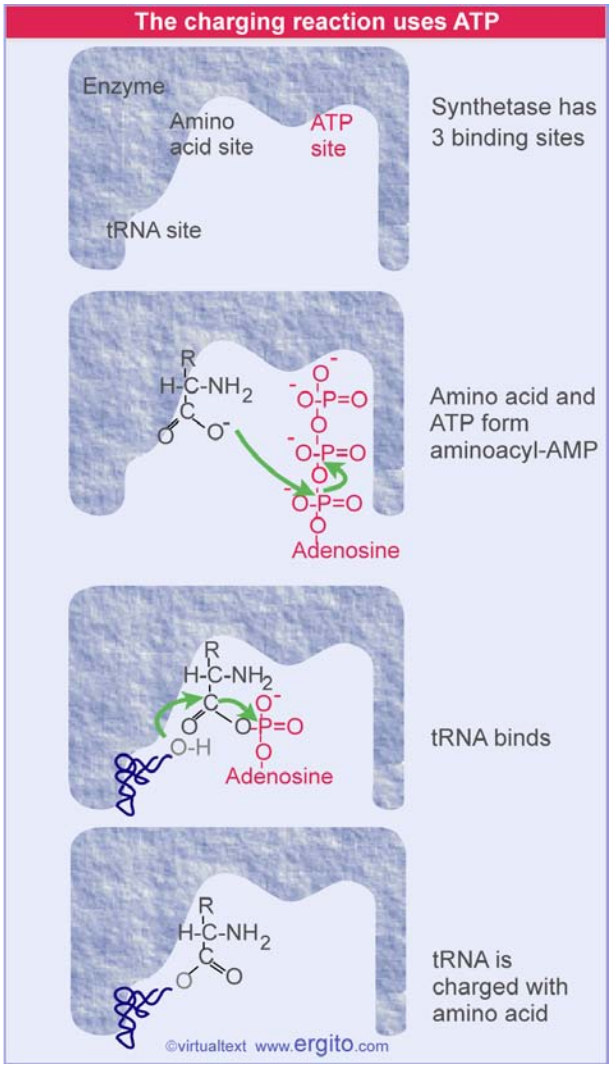

Figure 7.13 An aminoacyl-tRNA synthetase charges tRNA with an amino acid.

- First, the amino acid reacts with ATP to form aminoacyl~adenylate, releasing pyrophosphate. Energy for the reaction is provided by cleaving the high energy bond of the ATP.
- Then the activated amino acid is transferred to the tRNA, releasing AMP.

The synthetases sort the tRNAs and amino acids into corresponding sets. Each synthetase recognizes a single amino acid and all the tRNAs that should be charged with it. Usually, each amino acid is represented by more than one tRNA. Several tRNAs may be needed to respond to synonym codons, and sometimes there are multiple species of tRNA reacting with the same codon. Multiple tRNAs representing the same amino acid are called **isoaccepting tRNAs**; because they are all recognized by the same synthetase, they are also described as its **cognate tRNAs**.

Many attempts to deduce similarities in sequence between cognate tRNAs, or to

induce chemical alterations that affect their charging, have shown that the basis for recognition is different for different tRNAs, and does not necessarily lie in some feature of primary or secondary structure alone. We know from the crystal structure that the acceptor stem and the anticodon stem make tight contacts with the synthetase, and mutations that alter recognition of a tRNA are found in these two regions. (The anticodon itself is not necessarily recognized as such; for example, the "suppressor" mutations discussed later in this chapter change a base in the anticodon, and therefore the codons to which a tRNA responds, without altering its charging with amino acids.)

A group of isoaccepting tRNAs must be charged only by the single aminoacyl-tRNA synthetase specific for their amino acid. So isoaccepting tRNAs must share some common feature(s) enabling the enzyme to distinguish them from the other tRNAs. The entire complement of tRNAs is divided into 20 isoaccepting groups; each group is able to identify itself to its particular synthetase.

tRNAs are identified by their synthetases by contacts that recognize a small number of bases, typically from 1-5. Three types of feature commonly are used:

- Usually (but not always), at least one base of the anticodon is recognized. Sometimes all the positions of the anticodon are important.
- Often one of the last three base pairs in the acceptor stem is recognized. An extreme case is represented by alanine tRNA, which is identified by a single unique base pair in the acceptor stem.
- The so-called discriminator base, which lies between the acceptor stem and the CCA terminus, is always invariant among isoacceptor tRNAs.

No one of these features constitutes a unique means of distinguishing 20 sets of tRNAs, or provides sufficient specificity, so it appears that recognition of tRNAs is idiosyncratic, each following its own rules.

Several synthetases can specifically charge a "minihelix" consisting only of the acceptor and T ψ C arms (equivalent to one arm of the L-shaped molecule) with the correct amino acid. For certain tRNAs, specificity depends exclusively upon the acceptor stem. However, it is clear that there are significant variations between tRNAs, and in some cases the anticodon region is important. Mutations in the anticodon can affect recognition by the class II Phe-tRNA synthetase. Multiple features may be involved; minihelices from the $tRNA^{Val}$ and $tRNA^{Met}$ (where we know that the anticodon is important *in vivo*) can react specifically with their synthetases.

So recognition depends on an interaction between a few points of contact in the tRNA, concentrated at the extremities, and a few amino acids constituting the active site in the protein. The relative importance of the roles played by the acceptor stem and anticodon is different for each tRNA·synthetase interaction (for review see 35).

Reviews

35. Schimmel, P. (1989). *Parameters for the molecular recognition of tRNAs.* Biochemistry 28, 2747-2759.

USING THE GENETIC CODE

2.7.10 Aminoacyl-tRNA synthetases fall into two groups

Key Concepts

- Aminoacyl-tRNA synthetases are divided into the class I and class II groups by sequence and structural similarities.

In spite of their common function, synthetases are a rather diverse group of proteins. The individual subunits vary from 40-110 kD, and the enzymes may be monomeric, dimeric, or tetrameric. Homologies between them are rare. Of course, the active site that recognizes tRNA comprises a rather small part of the molecule. It is interesting to compare the active sites of different synthetases.

Synthetases have been divided into two general groups, each containing 10 enzymes, on the basis of the structure of the domain that contains the active site. A general type of organization that applies to both groups is represented in **Figure 7.14**. The catalytic domain includes the binding sites for ATP and amino acid. It can be recognized as a large region that is interrupted by an insertion of the domain that binds the acceptor helix of the tRNA. This places the terminus of the tRNA in proximity to the catalytic site. A separate domain binds the anticodon region of tRNA. Those synthetases that are multimeric also possess an oligomerization domain (for review see 34).

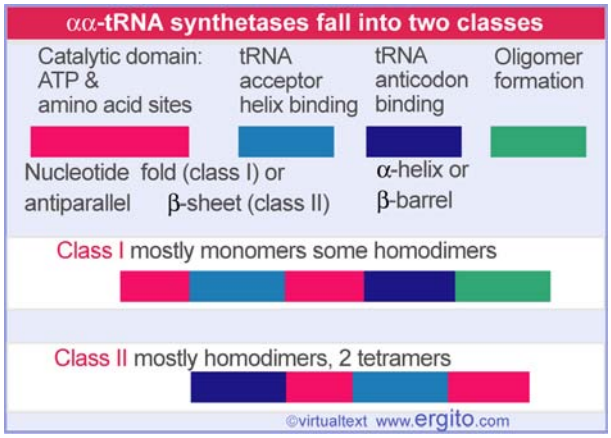

Figure 7.14 An aminoacyl-tRNA synthetase contains three or four regions with different functions. (Only multimeric synthetases possess an oligomerization domain.)

Class I synthetases have an N-terminal catalytic domain that is identified by the presence of two short, partly conserved sequences of amino acids, sometimes called "signature sequences." The catalytic domain takes the form of a motif called a nucleotide-binding fold (which is also found in other classes of enzymes that bind nucleotides). The nucleotide fold consists of alternating parallel β-strands and

-helices; the signature sequence forms part of the ATP-binding site. The insertion that contacts the acceptor helix of tRNA differs widely between different class I enzymes. The C-terminal domains of the class I synthetases, which include the tRNA anticodon-binding domain and any oligomerization domain, also are quite different from one another.

Class II enzymes share three rather general similarities of sequence in their catalytic domains. The active site contains a large antiparallel β-sheet surrounded by α-helices. Again, the acceptor helix-binding domain that interrupts the catalytic domain has a structure that depends on the individual enzyme. The anticodon-binding domain tends to be N-terminal. The location of any oligomerization domain is widely variable.

The lack of any apparent relationship between the two groups of synthetases is a puzzle. Perhaps they evolved independently of one another. This makes it seem possible even that an early form of life could have existed with proteins that were made up of just the 10 amino acids coded by one type or the other.

A general model for synthetase·tRNA binding suggests that the protein binds the tRNA along the "side" of the L-shaped molecule. The same general principle applies for all synthetase·tRNA binding: the tRNA is bound principally at its two extremities, and most of the tRNA sequence is not involved in recognition by a synthetase. However, the detailed nature of the interaction is different between class I and class II enzymes, as can be seen from the models of **Figure 7.15**, which are based on crystal structures. The two types of enzyme approach the tRNA from opposite sides, with the result that the tRNA-protein models look almost like mirror images of one another.

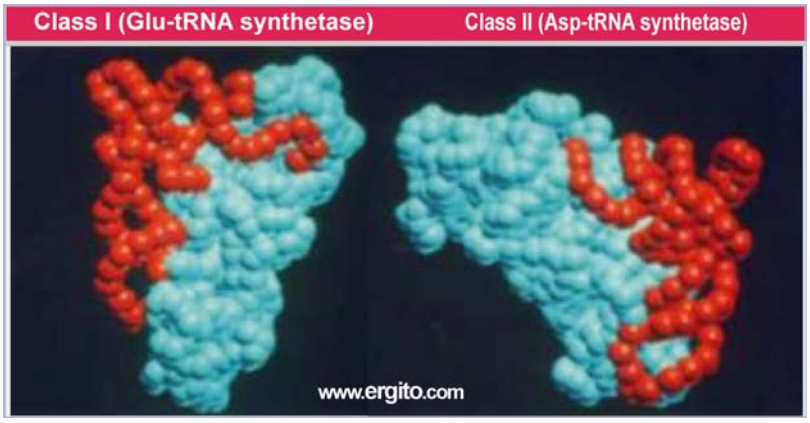

Figure 7.15 Crystal structures show that class I and class II aminoacyl-tRNA synthetases bind the opposite faces of their tRNA substrates. The tRNA is shown in red, and the protein in blue. Photographs kindly provided by Dino Moras.

A class I enzyme (Gln-tRNA synthetase) approaches the D-loop side of the tRNA. It recognizes the minor groove of the acceptor stem at one end of the binding site, and interacts with the anticodon loop at the other end. **Figure 7.16** is a diagrammatic representation of the crystal structure of the $tRNA^{Gln}$·synthetase complex. A revealing feature of the structure is that contacts with the enzyme change the structure of the tRNA at two important points. These can be seen by comparing the

dotted and solid lines in the anticodon loop and acceptor stem:

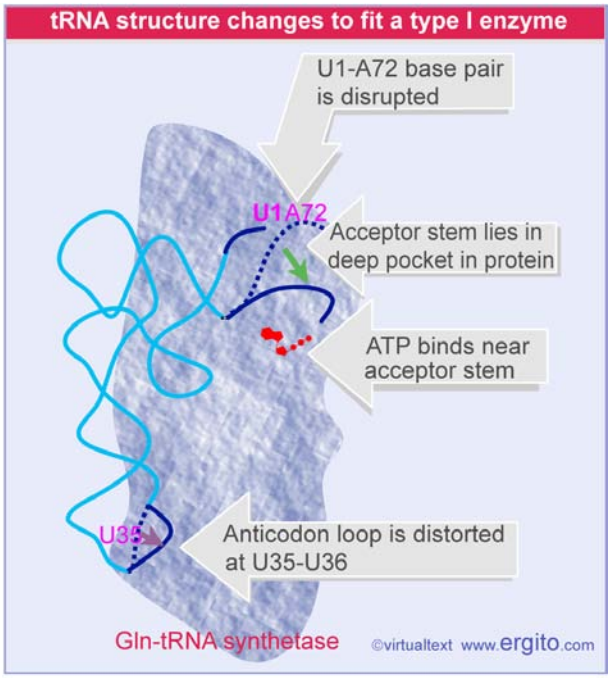

Figure 7.16 A class I tRNA synthetase contacts tRNA at the minor groove of the acceptor stem and at the anticodon.

- Bases U35 and U36 in the anticodon loop are pulled farther out of the tRNA into the protein.
- The end of the acceptor stem is seriously distorted, with the result that base pairing between U1 and A72 is disrupted. The single-stranded end of the stem pokes into a deep pocket in the synthetase protein, which also contains the binding site for ATP.

This structure explains why changes in U35, G73, or the U1-A72 base pair affect the recognition of the tRNA by its synthetase. At all of these positions, hydrogen bonding occurs between the protein and tRNA (447).

A class II enzyme (Asp-tRNA synthetase) approaches the tRNA from the other side, and recognizes the variable loop, and the major groove of the acceptor stem, as drawn in **Figure 7.17**. The acceptor stem remains in its regular helical conformation. ATP is probably bound near to the terminal adenine. At the other end of the binding site, there is a tight contact with the anticodon loop, which has a change in conformation that allows the anticodon to be in close contact with the protein (448).

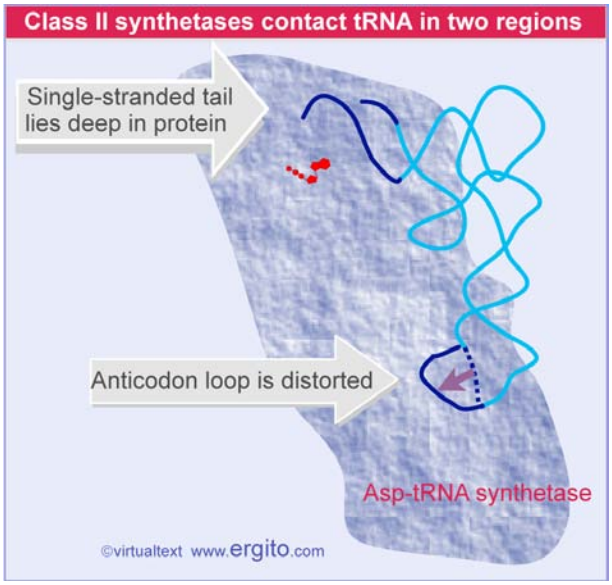

Figure 7.17 A class II aminoacyl-tRNA synthetase contacts tRNA at the major groove of the acceptor helix and at the anticodon loop.

Reviews

34. Schimmel, P. (1987). *Aminoacyl-tRNA synthetases: general scheme of structure-function relationships on the polypeptides and recognition of tRNAs.* Annu. Rev. Biochem. 56, 125-158.

References

447. Rould, M. A. et al. (1989). *Structure of E. coli glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 28*Å *resolution*. Science 246, 1135-1142.

448. Ruff, M. et al. (1991). *Class II aminoacyl tRNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexes with $tRNA^{Asp}$*. Science 252, 1682-1689.

USING THE GENETIC CODE

2.7.11 Synthetases use proofreading to improve accuracy

Key Terms

Proofreading refers to any mechanism for correcting errors in protein or nucleic acid synthesis that involves scrutiny of individual units *after* they have been added to the chain.

Kinetic proofreading describes a proofreading mechanism that depends on incorrect events proceeding more slowly than correct events, so that incorrect events are reversed before a subunit is added to a polymeric chain.

Chemical proofreading describes a proofreading mechanism in which the correction event occurs after addition of an incorrect subunit to a polymeric chain, by reversing the addition reaction.

Key Concepts

- Specificity of recognition of both amino acid and tRNA is controlled by aminoacyl-tRNA synthetases by proofreading reactions that reverse the catalytic reaction if the wrong component has been incorporated.

Aminoacyl-tRNA synthetases have a difficult job. Each synthetase must distinguish 1 out of 20 amino acids, and and must differentiate cognate tRNAs (typically 1-3) from the total set (perhaps 100 in all).

Many amino acids are closely related to one another, and all amino acids are related to the metabolic intermediates in their particular synthetic pathway. It is especially difficult to distinguish between two amino acids that differ only in the length of the carbon backbone (that is, by one CH_2 group). Intrinsic discrimination based on relative energies of binding two such amino acids would be only ~1/5. The synthetase enzymes improve this ratio ~1000 fold.

Intrinsic discrimination between tRNAs is better, because the tRNA offers a larger surface with which to make more contacts, but it is still true that all tRNAs conform to the same general structure, and there may be a quite limited set of features that distinguish the cognate tRNAs from the noncognate tRNAs.

We can imagine two general ways in which the enzyme might select its substrate:

- The cycle of admittance, scrutiny, rejection/acceptance could represent a single binding step that precedes all other stages of whatever reaction is involved. This is tantamount to saying that the affinity of the binding site is sufficient to control the entry of substrate. In the case of synthetases, this would mean that only the correct amino acids and cognate tRNAs could form a stable attachment at the site.

- Alternatively, the reaction proceeds through some of its stages, after which a decision is reached on whether the correct species is present. If it is not present, the reaction is reversed, or a bypass route is taken, and the wrong member is expelled. This sort of postbinding scrutiny is generally described as **proofreading**. In the example of synthetases, it would require that the charging reaction proceeds through certain stages even if the wrong tRNA or amino acid is present.

Synthetases use proofreading mechanisms to control the recognition of both types of substrates. They improve significantly on the intrinsic differences among amino acids or among tRNAs, but, consistent with the intrinsic differences in each group, make more mistakes in selecting amino acids (error rates are 10^{-4} - 10^{-5}) than in selecting tRNAs (error rates are ~10^{-6}) (see **Figure 6.8**).

Transfer RNA binds to synthetase by the two stage reaction depicted in **Figure 7.18**. Cognate tRNAs have a greater intrinsic affinity for the binding site, so they are bound more rapidly and dissociate more slowly. Following binding, the enzyme scrutinizes the tRNA that has been bound. If the correct tRNA is present, binding is stabilized by a conformational change in the enzyme. This allows aminoacylation to occur rapidly. If the wrong tRNA is present, the conformational change does not occur. As a result, the reaction proceeds much more slowly; this increases the chance that the tRNA will dissociate from the enzyme before it is charged. This type of control is called **kinetic proofreading** (450).

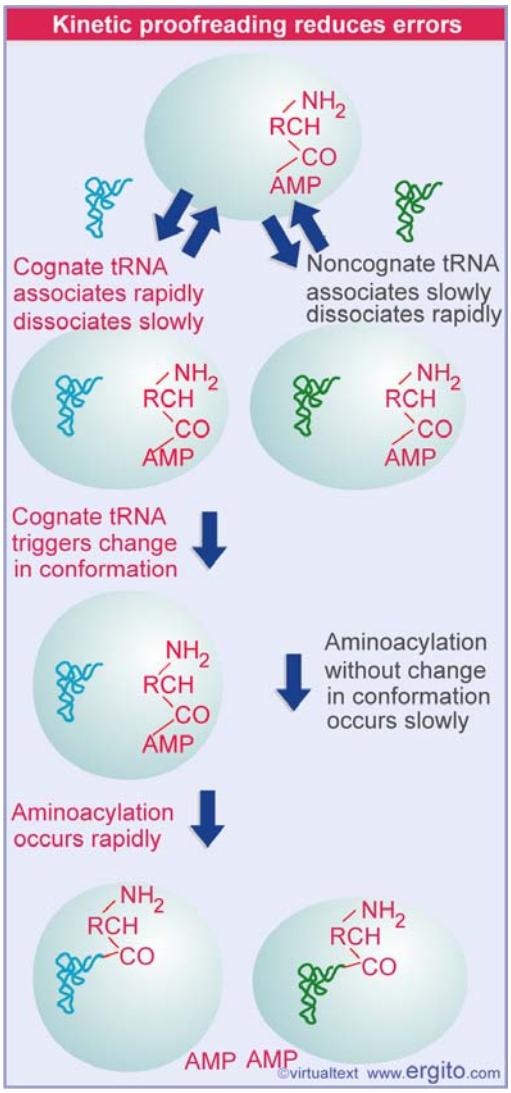

Figure 7.18 Recognition of the correct tRNA by synthetase is controlled at two steps. First, the enzyme has a greater affinity for its cognate tRNA. Second, the aminoacylation of the incorrect tRNA is very slow.

Specificity for amino acids varies among the synthetases. Some are highly specific for initially binding a single amino acid, but others can also activate amino acids closely related to the proper substrate. Although the analog amino acid can sometimes be converted to the adenylate form, in none of these cases is an incorrectly activated amino acid actually used to form a stable aminoacyl-tRNA.

The presence of the cognate tRNA usually is needed to trigger proofreading, even if the reaction occurs at the stage before formation of aminoacyl-adenylate. (An exception is provided by Met-tRNA synthetase, which can reject noncognate aminoacyl-adenylate complexes even in the absence of tRNA.)

There are two stages at which proofreading of an incorrect aminoacyl-adenylate may

occur during formation of aminoacyl-tRNA. **Figure 7.19** shows that both use **chemical proofreading**, in which the catalytic reaction is reversed. The extent to which one pathway or the other predominates varies with the individual synthetase:

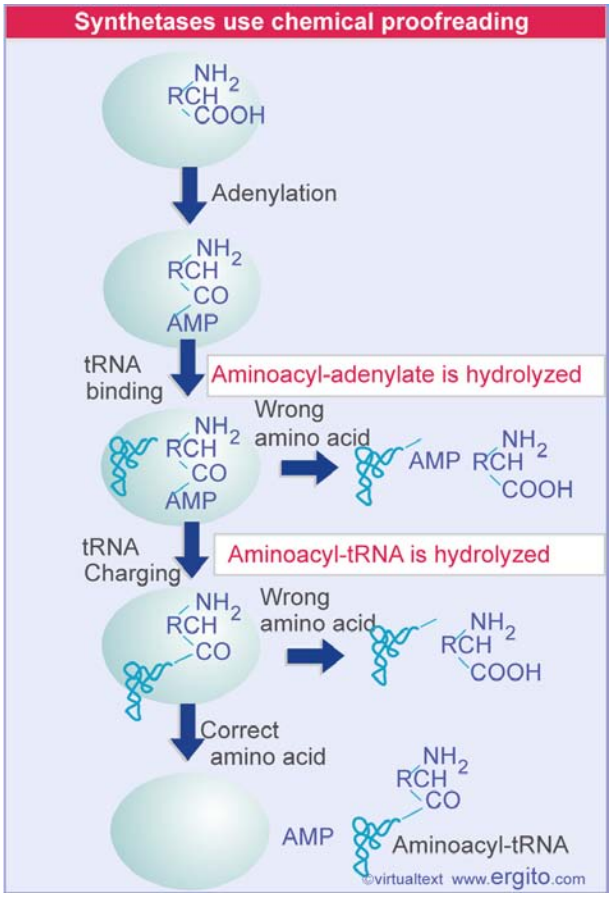

Figure 7.19 When a synthetase binds the incorrect amino acid, proofreading requires binding of the cognate tRNA. It may take place either by a conformation change that causes hydrolysis of the incorrect aminoacyl-adenylate, or by transfer of the amino acid to tRNA, followed by hydrolysis.

- The noncognate aminoacyl-adenylate may be hydrolyzed when the cognate tRNA binds. This mechanism is used predominantly by several synthetases, including those for methionine, isoleucine, and valine. (Usually, the reaction cannot be seen *in vivo*, but it can be followed for Met-tRNA synthetase when the incorrectly activated amino acid is homocysteine, which lacks the methyl group of methionine). Proofreading releases the amino acid in an altered form, as homocysteine thiolactone. In fact, homocysteine thiolactone is produced in *E. coli* as a by-product of the charging reaction of Met-tRNA synthetase. This shows that continuous proofreading is part of the process of charging a tRNA with its amino acid (451).

- Some synthetases use chemical proofreading at a later stage. The wrong amino acid is actually transferred to tRNA, is then recognized as incorrect by its structure in the tRNA binding site, and so is hydrolyzed and released. The

process requires a continual cycle of linkage and hydrolysis until the correct amino acid is transferred to the tRNA.

A classic example in which discrimination between amino acids depends on the presence of tRNA is provided by the Ile-tRNA synthetase of *E. coli*. The enzyme can charge valine with AMP, but hydrolyzes the valyl-adenylate when $tRNA^{Ile}$ is added. The overall error rate depends on the specificities of the individual steps, as summarized in **Figure 7.20**. The overall error rate of 1.5×10^{-5} is less than the measured rate at which valine is substituted for isoleucine (in rabbit globin), which is $2\text{-}5 \times 10^{-4}$. So mischarging probably provides only a small fraction of the errors that actually occur in protein synthesis.

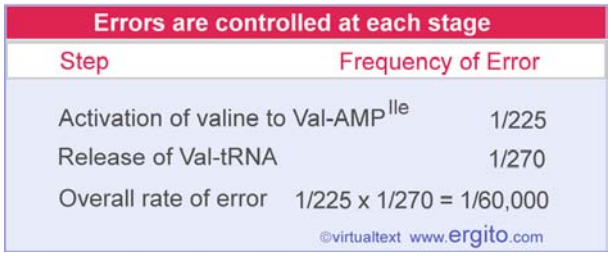

Figure 7.20 The accuracy of charging $tRNA^{Ile}$ by its synthetase depends on error control at two stages.

Ile-tRNA synthetase uses size as a basis for discrimination among amino acids. **Figure 7.21** shows that it has two active sites: the synthetic (or activation) site and the editing (or hydrolytic) site. The crystal structure of the enzyme shows that the synthetic site is too small to allow leucine (a close analog of isoleucine) to enter. All amino acids large than isoleucine are excluded from activation because they cannot enter the synthetic site. An amino acid that can enter the synthetic site is placed on tRNA. Then the enzyme tries to transfer it to the editing site. Isoleucine is safe from editing because it is too large to enter the editing site. However, valine can enter this site, and as a result an incorrect $Val\text{-}tRNA^{Ile}$ is hydrolyzed. Essentially the enzyme provides a double molecular sieve, in which size of the amino acid is used to discriminate between closely related species (452; for review see 40).

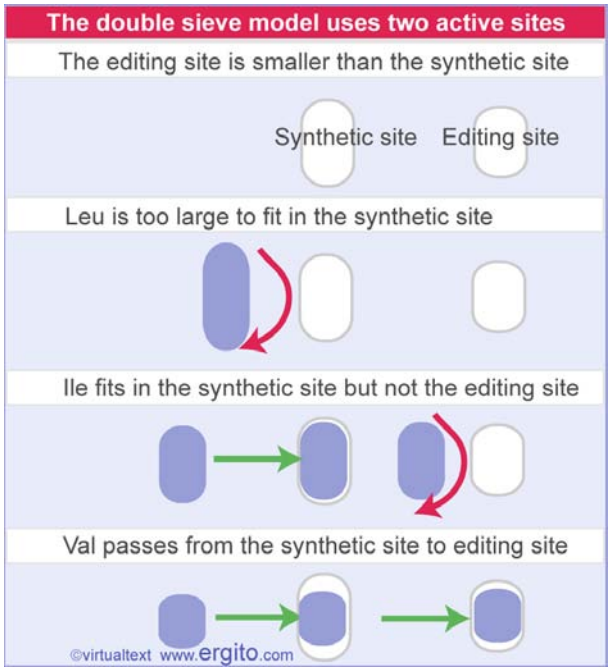

Figure 7.21 Ile-tRNA synthetase has two active sites. Amino acids larger than Ile cannot be activated because they do not fit in the synthetic site. Amino acids smaller than Ile are removed because they are able to enter the editing site.

One interesting feature of Ile-tRNA synthetase is that the synthetic and editing sites are a considerable distance apart, ~34Å. A crystal structure of the enzyme complexed with an edited analog of isoleucine shows that the amino acid is transported from the synthetic site to the editing site (2164). **Figure 7.22** shows that this involves a change in the conformation of the tRNA. The amino acid acceptor stem of $tRNA^{Ile}$ can exist in alternative conformations. It adopts an unusual hairpin in order to be aminoacylated by an amino acid in the synthetic site. Then it returns to the more common helical structure in order to move the amino acid to the editing site. The translocation between sites is the rate-limiting step in proofreading (2165). Ile-tRNA synthetase is a class I synthetase, but the double sieve mechanism is used also by class II synthetases (2166).

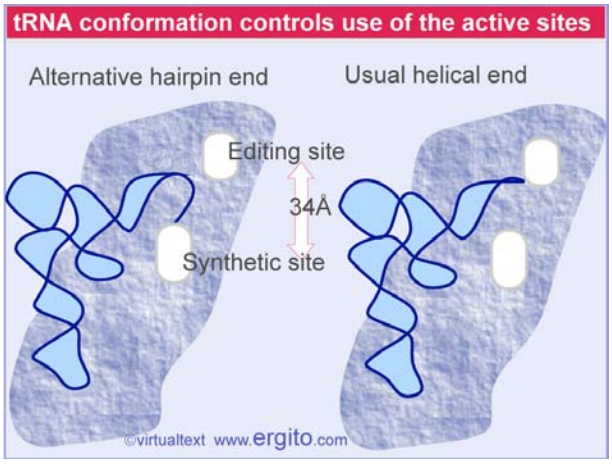

Figure 7.22 An amino acid is transported from the synthetic site to the editing site of Ile-tRNA synthetase by a change in the conformation of the amino acceptor stem of tRNA.

Last updated on 10-31-2001

Reviews

40. Jakubowski, H. and Goldman, E. (1992). *Editing of errors in selection of amino acids for protein synthesis.* Microbiol. Rev. 56, 412-429.

References

450. Hopfield, J. J. (1974). *Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity*. Proc. Natl. Acad. Sci. USA 71, 4135-4139.

451. Jakubowski, H. (1990). *Proofreading in vivo: editing of homocysteine by methionyl-tRNA synthetase in E. coli*. Proc. Natl. Acad. Sci. USA 87, 4504-4508.

452. Nureki, O. et al. (1998). *Enzyme structure with two catalytic sites for double sieve selection of substrate*. Science 280, 578-581.

2164. Silvian, L. F., Wang, J., and Steitz, T. A. (1999). *Insights into editing from an ile-tRNA synthetase structure with $tRNA^{Ile}$ and mupirocin.* Science 285, 1074-1077.

2165. Nomanbhoy, T. K., Hendrickson, T. L., and Schimmel, P. (1999). *Transfer RNA-dependent translocation of misactivated amino acids to prevent errors in protein synthesis.* Mol. Cell 4, 519-528.

2166. Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C., and Moras, D. (2000). *Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem.* Cell 103, 877-884.

USING THE GENETIC CODE

2.7.12 Suppressor tRNAs have mutated anticodons that read new codons

Key Terms

A **suppressor** is a second mutation that compensates for or alters the effects of a primary mutation.

A **nonsense suppressor** is a gene coding for a mutant tRNA able to respond to one or more of the termination codons and insert an amino acid at that site.

A **missense suppressor** codes for a tRNA that has been mutated so as to recognize a different codon. By inserting a different amino acid at a mutant codon, the tRNA suppresses the effect of the original mutation.

Key Concepts

- A suppressor tRNA typically has a mutation in the anticodon that changes the codons to which it responds.
- When the new anticodon corresponds to a termination codon, an amino acid is inserted and the polypeptide chain is extended beyond the termination codon. This results in nonsense suppression at a site of nonsense mutation or in readthrough at a natural termination codon.
- Missense suppression occurs when the tRNA recognizes a different codon from usual, so that one amino acid is substituted for another.

Isolation of mutant tRNAs has been one of the most potent tools for analyzing the ability of a tRNA to respond to its codon(s) in mRNA, and for determining the effects that different parts of the tRNA molecule have on codon-anticodon recognition.

Mutant tRNAs are isolated by virtue of their ability to overcome the effects of mutations in genes coding for proteins. In general genetic terminology, a mutation that is able to overcome the effects of another mutation is called a **suppressor**.

In tRNA suppressor systems, the primary mutation changes a codon in an mRNA so that the protein product is no longer functional. The secondary, suppressor mutation changes the anticodon of a tRNA, so that it recognizes the mutant codon instead of (or as well as) its original target codon. The amino acid that is now inserted restores protein function. The suppressors are described as **nonsense suppressors** or **missense suppressors**, depending on the nature of the original mutation.

In a wild-type cell, a nonsense mutation is recognized only by a release factor, terminating protein synthesis. The suppressor mutation creates an aminoacyl-tRNA that can recognize the termination codon; by inserting an amino acid, it allows protein synthesis to continue beyond the site of nonsense mutation. This new

capacity of the translation system allows a full-length protein to be synthesized, as illustrated in **Figure 7.23**. If the amino acid inserted by suppression is different from the amino acid that was originally present at this site in the wild-type protein, the activity of the protein may be altered.

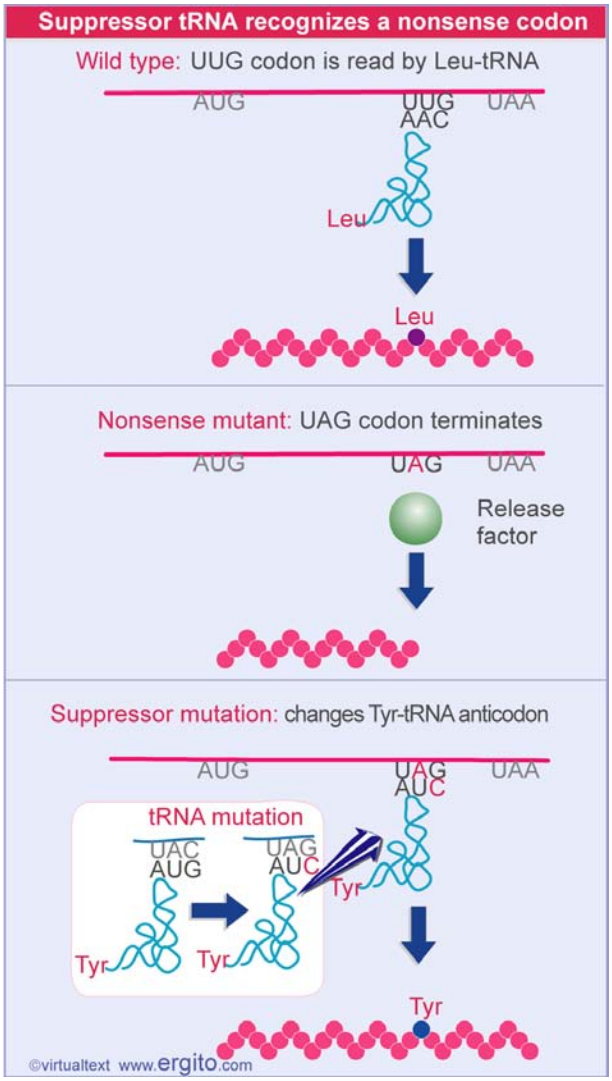

Figure 7.23 Nonsense mutations can be suppressed by a tRNA with a mutant anticodon, which inserts an amino acid at the mutant codon, producing a full length protein in which the original Leu residue has been replaced by Tyr.

Missense mutations change a codon representing one amino acid into a codon representing another amino acid, one that cannot function in the protein in place of the original residue. (Formally, any substitution of amino acids constitutes a missense mutation, but in practice it is detected only if it changes the activity of the protein.) The mutation can be suppressed by the insertion either of the original amino acid or of some other amino acid that is acceptable to the protein.

Figure 7.24 demonstrates that missense suppression can be accomplished in the

same way as nonsense suppression, by mutating the anticodon of a tRNA carrying an acceptable amino acid so that it responds to the mutant codon. So missense suppression involves a change in the meaning of the codon from one amino acid to another.

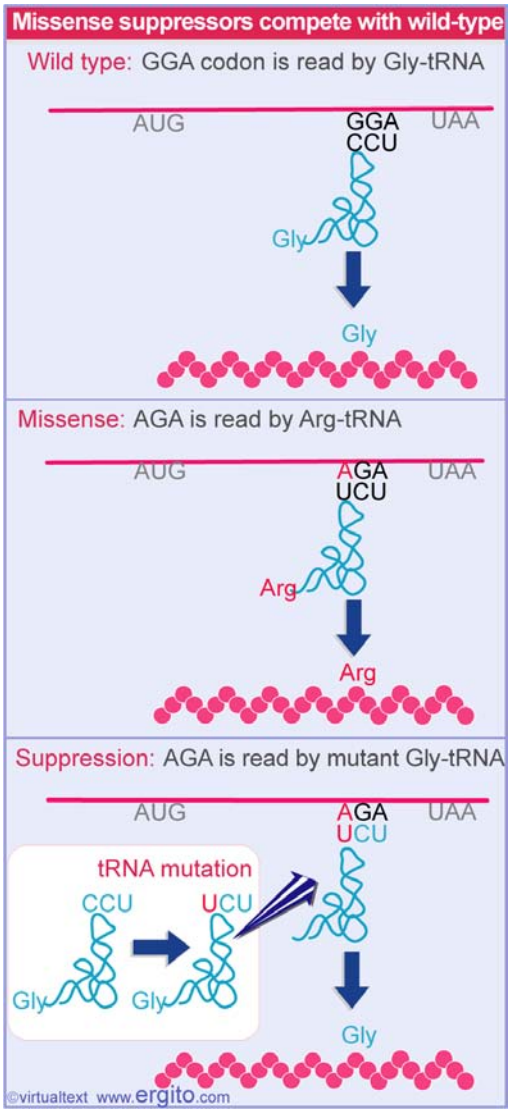

Figure 7.24 Missense suppression occurs when the anticodon of tRNA is mutated so that it responds to the wrong codon. The suppression is only partial because both the wild-type tRNA and the suppressor tRNA can respond to AGA.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.2.7.12

USING THE GENETIC CODE

2.7.13 There are nonsense suppressors for each termination codon

Key Concepts

- Each type of nonsense codon is suppressed by tRNAs with mutant anticodons.
- Some rare suppressor tRNAs have mutations in other parts of the molecule.

Nonsense suppressors fall into three classes, one for each type of termination codon. **Figure 7.25** describes the properties of some of the best characterized suppressors.

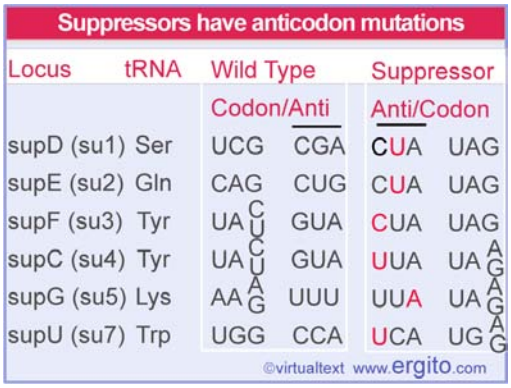

Figure 7.25 Nonsense suppressor tRNAs are generated by mutations in the anticodon.

The easiest to characterize have been amber suppressors. In *E. coli*, at least 6 tRNAs have been mutated to recognize UAG codons. All of the amber suppressor tRNAs have the anticodon CUA$^{\leftarrow}$, in each case derived from wild type by a single base change. The site of mutation can be any one of the three bases of the anticodon, as seen from *supD*, *supE*, and *supF*. Each suppressor tRNA recognizes only the UAG codon, instead of its former codon(s). The amino acids inserted are serine, glutamine, or tyrosine, the same as those carried by the corresponding wild-type tRNAs.

Ochre suppressors also arise by mutations in the anticodon. The best known are *supC* and *supG*, which insert tyrosine or lysine in response to both ochre (UAA) and amber (UAG) codons. This conforms with the prediction of the wobble hypothesis that UAA cannot be recognized alone.

A UGA suppressor has an unexpected property. It is derived from tRNA^{Trp}, but its only mutation is the substitution of A in place of G at position 24. This change replaces a G·U pair in the D stem with an A·U pair, increasing the stability of the helix. The sequence of the anticodon remains the same as the wild type, CCA$^{\leftarrow}$. So the mutation in the D stem must in some way alter the conformation of the anticodon loop, allowing CCA$^{\leftarrow}$ to pair with UGA in an unusual wobble pairing of C with A.

The suppressor tRNA continues to recognize its usual codon, UGG.

A related response is seen with a eukaryotic tRNA. Bovine liver contains a $tRNA^{Ser}$ with the anticodon $^{m}CCA^{\leftarrow}$. The wobble rules predict that this tRNA should respond to the tryptophan codon UGG; but in fact it responds to the termination codon UGA. So it is possible that UGA is suppressed naturally in this situation.

The general importance of these observations lies in the demonstration that codon-anticodon recognition of either wild-type or mutant tRNA cannot be predicted entirely from the relevant triplet sequences, but is influenced by other features of the molecule.

USING THE GENETIC CODE

2.7.14 Suppressors may compete with wild-type reading of the code

Key Terms

The **context** of a codon in mRNA refers to the fact that neighboring sequences may change the efficiency with which a codon is recognized by its aminoacyl-tRNA or is used to terminate protein synthesis.

Readthrough at transcription or translation occurs when RNA polymerase or the ribosome, respectively, ignores a termination signal because of a mutation of the template or the behavior of an accessory factor.

Key Concepts

- Suppressor tRNAs compete with wild-type tRNAs that have the same anticodon to read the corresponding codon(s).
- Efficient suppression is deleterious because it results in readthrough past normal termination codons.
- The UGA codon is leaky and is misread by Trp-tRNA at 1-3% frequency.

There is an interesting difference between the usual recognition of a codon by its proper aminoacyl-tRNA and the situation in which mutation allows a suppressor tRNA to recognize a new codon. In the wild-type cell, only one meaning can be attributed to a given codon, which represents either a particular amino acid or a signal for termination. But in a cell carrying a suppressor mutation, the mutant codon has the alternatives of being recognized by the suppressor tRNA or of being read with its usual meaning.

A nonsense suppressor tRNA must compete with the release factors that recognize the termination codon(s). A missense suppressor tRNA must compete with the tRNAs that respond properly to its new codon. The extent of competition influences the efficiency of suppression; so the effectiveness of a particular suppressor depends not only on the affinity between its anticodon and the target codon, but also on its concentration in the cell, and on the parameters governing the competing termination or insertion reactions.

The efficiency with which any particular codon is read is influenced by its location. So the extent of nonsense suppression by a given tRNA can vary quite widely, depending on the **context** of the codon. We do not understand the effect that neighboring bases in mRNA have on codon-anticodon recognition, but the context can change the frequency with which a codon is recognized by a particular tRNA by more than an order of magnitude. The base on the 3′ side of a codon appears to have a particularly strong effect.

A nonsense suppressor is isolated by its ability to respond to a mutant nonsense

codon. But the same triplet sequence constitutes one of the normal termination signals of the cell! The mutant tRNA that suppresses the nonsense mutation must in principle be able to suppress natural termination at the end of any gene that uses this codon. **Figure 7.26** shows that this **readthrough** results in the synthesis of a longer protein, with additional C-terminal material. The extended protein will end at the next termination triplet sequence found in the phase of the reading frame. Any extensive suppression of termination is likely to be deleterious to the cell by producing extended proteins whose functions are thereby altered.

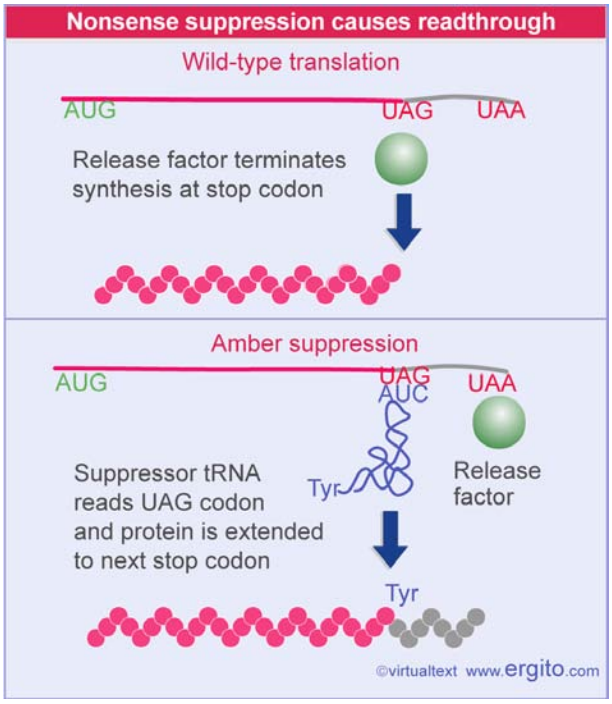

Figure 7.26 Nonsense suppressors also read through natural termination codons, synthesizing proteins that are longer than wild-type.

Amber suppressors tend to be relatively efficient, usually in the range of 10-50%, depending on the system. This efficiency is possible because amber codons are used relatively infrequently to terminate protein synthesis in *E. coli*.

Ochre suppressors are difficult to isolate. They are always much less efficient, usually with activities below 10%. All ochre suppressors grow rather poorly, which indicates that suppression of both UAA and UAG is damaging to *E. coli*, probably because the ochre codon is used most frequently as a natural termination signal.

UGA is the least efficient of the termination codons in its natural function; it is misread by Trp-tRNA as frequently as 1-3% in wild-type situations. In spite of this deficiency, however, it is used more commonly than the amber triplet to terminate bacterial genes.

One gene's missense suppressor is likely to be another gene's mutator. A suppressor corrects a mutation by substituting one amino acid for another at the mutant site. But in other locations, the same substitution will replace the wild-type amino acid with a

new amino acid. The change may inhibit normal protein function.

This poses a dilemma for the cell: it must suppress what is a mutant codon at one location, while failing to change too extensively its normal meaning at other locations. The absence of any strong missense suppressors is therefore explained by the damaging effects that would be caused by a general and efficient substitution of amino acids.

A mutation that creates a suppressor tRNA can have two consequences. First, it allows the tRNA to recognize a new codon. Second, sometimes it prevents the tRNA from recognizing the codons to which it previously responded. It is significant that all the high-efficiency amber suppressors are derived by mutation of one copy of a redundant tRNA set. In these cases, the cell has several tRNAs able to respond to the codon originally recognized by the wild-type tRNA. So the mutation does not abolish recognition of the old codons, which continue to be served adequately by the tRNAs of the set. In the unusual situation in which there is only a single tRNA that responds to a particular codon, any mutation that prevents the response is lethal (for review see 31; 36; 37; 38).

Suppression is most often considered in the context of a mutation that changes the reading of a codon. However, there are some situations in which a stop codon is read as an amino acid at a low frequency in the wild-type situation. The first example to be discovered was the coat protein gene of the RNA phage Q β. The formation of infective Q β particles requires that the stop codon at the end of this gene is suppressed at a low frequency to generate a small proportion of coat proteins with a C-terminal extension. In effect, this stop codon is leaky. The reason is that Trp-tRNA recognizes the codon at a low frequency (3062; 3063).

Readthrough past stop codons occurs also in eukaryotes, where it is employed most often by RNA viruses. This may involve the suppression of UAG/UAA by Tyr-tRNA, Gln-tRNA, or Leu-tRNA, or the suppression of UGA by Trp-tRNA or Arg-tRNA. The extent of partial suppression is dictated by the context surrounding the codon (for review see 3061).

Last updated on 11-2-2002

Reviews

31. Murgola, E. J. (1985). *tRNA, suppression, and the code.* Annu. Rev. Genet. 19, 57-80.

36. Eggertsson, G. and Soll, D. (1988). *Transfer RNA-mediated suppression of termination codons in E. coli.* Microbiol. Rev. 52, 354-374.

37. Normanly, J. and Abelson, J. (1989). *Transfer RNA identity.* Annu. Rev. Biochem. 58, 1029-1049.

38. Atkins, J. F (1991). *Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frameshifts and hops.* Annu. Rev. Genet. 25, 201-228.

3061. Beier, H. and Grimm, M. (2001). *Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs.* Nucleic Acids Res. 29, 4767-4782.

References

3062. Hirsh, D. (1971). *Tryptophan transfer RNA as the UGA suppressor*. J. Mol. Biol. 58, 439-458.

3063. Weiner, A. M. and Weber, K. (1973). *A single UGA codon functions as a natural termination signal in the coliphage q beta coat protein cistron.* J. Mol. Biol. 80, 837-855.

USING THE GENETIC CODE

2.7.15 The ribosome influences the accuracy of translation

Key Concepts

- The structure of the 16S rRNA at the P and A sites of the ribosome influences the accuracy of translation.

The lack of detectable variation when the sequence of a protein is analyzed demonstrates that protein synthesis must be extremely accurate. Very few mistakes are apparent in the form of substitutions of one amino acid for another. There are two general stages in protein synthesis at which errors might be made (see **Figure 6.8** in *Molecular Biology 2.6.3 Special mechanisms control the accuracy of protein synthesis*):

- Charging a tRNA only with its correct amino acid clearly is critical. This is a function of the aminoacyl-tRNA synthetase. Probably the error rate varies with the particular enzyme, but generally mistakes occur in $<1/10^5$ aminoacylations.
- The specificity of codon-anticodon recognition is crucial, but puzzling. Although binding constants vary with the individual codon-anticodon reaction, the specificity is always much too low to provide an error rate of $<10^{-5}$. When free in solution, tRNAs bind to their trinucleotide codon sequences only rather weakly. Related, but erroneous, triplets (with two correct bases out of three) are recognized 10^{-1}–10^{-2} times as efficiently as the correct triplets.

Codon-anticodon base pairing therefore seems to be a weak point in the accuracy of translation. The ribosome has an important role in controlling the specificity of this interaction, functioning directly or indirectly as a "proofreader," to distinguish correct and incorrect codon-anticodon pairs, and thus amplifying the rather modest intrinsic difference by ~1000×. And in addition to the role of the ribosome itself, the factors that place initiator- and aminoacyl-tRNAs in the ribosome also may influence the pairing reaction.

So there must be some mechanism for stabilizing the correct aminoacyl-tRNA, allowing its amino acid to be accepted as a substrate for receipt of the polypeptide chain; contacts with an incorrect aminoacyl-tRNA must be rapidly broken, so that the complex leaves without reacting. Suppose that there is no specificity in the initial collision between the aminoacyl-tRNA·EF-Tu·GTP complex and the ribosome. If any complex, irrespective of its tRNA, can enter the A site, the number of incorrect entries must far exceed the number of correct entries.

There are two basic models for how the ribosome might discriminate between correctly and incorrectly paired aminoacyl-tRNAs. The actual situation incorporates elements of both models.

- The direct recognition model supposes that the structure of the ribosome is designed to recognize aminoacyl-tRNAs that are correctly paired. This would mean that the correct pairing results in some small change in the conformation of the aminoacyl-tRNA that the ribosome can recognize. Discrimination occurs before any further reaction occurs.

- The kinetic proofreading model proposes that there are two (or more) stages in the process, so that the aminoacyl-tRNA has multiple opportunities to disengage. An incorrectly paired aminoacyl-tRNA may pass through some stages of the reaction before it is rejected. Overall selectivity can in principle be the product of the selectivities at each stage.

Figure 7.27 illustrates diagrammatically what happens to correctly and incorrectly paired aminoacyl-tRNAs. A correctly paired aminoacyl-tRNA is able to make stabilizing contacts with rRNA. An incorrectly paired aminoacyl-tRNA does not make these contacts, and therefore is able to diffuse out of the A site.

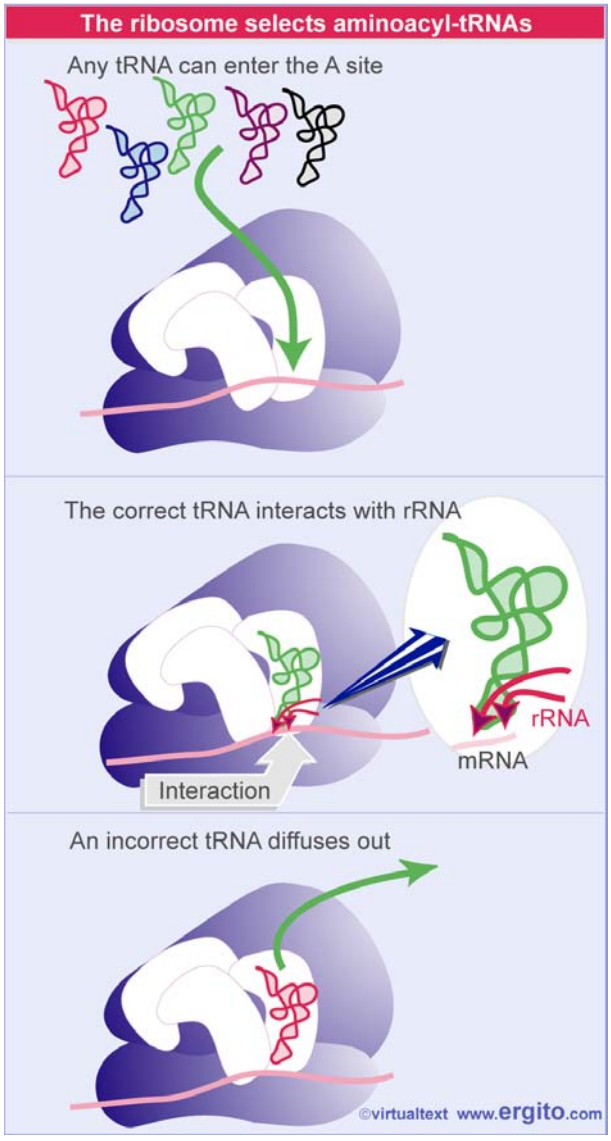

Figure 7.27 Any aminoacyl-tRNA can be placed in the A site (by EF-Tu), but only one that pairs with the anticodon can make stabilizing contacts with rRNA. In the absence of these contacts, the aminoacyl-tRNA diffuses out of the A site.

The path to discovering these interactions started with investigations of the effects of the antibiotic streptomycin in the 1960s. Streptomycin inhibits protein synthesis by binding to 16S rRNA and inhibiting the ability of EF-G to catalyze translocation. It also increases the level of misreading of the pyrimidines U and C (usually one is mistaken for the other, occasionally for A). The site at which streptomycin acts is influenced by the S12 protein; the sequence of this protein is altered in resistant mutants. Ribosomes with an S12 protein derived from resistant bacteria show a reduction in the level of misreading compared with wild-type ribosomes. In effect, S12 controls the level of misreading. When it is mutated to decrease misreading, it suppresses the effect of streptomycin.

S12 stabilizes the structure of 16S rRNA in the region that is bound by streptomycin.

The important point to note here is that the P/A site region influences the accuracy of translation: translation can be made more or less accurate by changing the structure of 16S rRNA. The combination of the effects of the S12 protein and streptomycin on the rRNA structure explains the behavior of different mutants in S12, some of which even make the ribosome *dependent* on the presence of streptomycin for correct translation (1185).

We now know from the crystal structure of the ribosome that 16S rRNA is in a position to make contacts with aminoacyl-tRNA (for review see 2390). Two bases of 16S rRNA can contact the minor groove of the helix formed by pairing between the anticodon in tRNA with the first two bases of the codon in mRNA (2392). This directly stabilizes the structure when the correct codon-anticodon contacts are made at the first two codon positions, but it does not monitor contacts at the third position.

The stabilization of correctly paired aminoacyl-tRNA may have two effects. By holding the aminoacyl-tRNA in the A site, it prevents it from escaping before the next stage of protein synthesis. And the conformational change in the rRNA may help to trigger the next stage of the reaction, which is the hydrolysis of GTP by EF-Tu.

Part of the proofreading effect is determined by timing. An aminoacyl-tRNA in the A site may in effect be trapped if the next stage of protein synthesis occurs while it is there. So a delay between entry into the A site and peptidyl transfer may give more opportunity for a mismatched aminoacyl-tRNA to dissociate. Mismatched aminoacyl-tRNA dissociates more rapidly than correctly matched aminoacyl-tRNA, probably by a factor of ~5×. Its chance of escaping is therefore increased when the peptide transfer step is slowed (for review see 437).

The specificity of decoding has been assumed to reside with the ribosome itself, but some recent results suggest that translation factors influence the process at both the P site and A site. An indication that EF-Tu is involved in maintaining the reading frame is provided by mutants of the factor that suppress frameshifting. This implies that EF-Tu does not merely bring aminoacyl-tRNA to the A site, but also is involved in positioning the incoming aminoacyl-tRNA relative to the peptidyl-tRNA in the P site.

A striking case where factors influence meaning is found at initiation. Mutation of the AUG initiation codon to UUG in the yeast gene *HIS4* prevents initiation. Extragenic suppressor mutations can be found that allow protein synthesis to be initiated at the mutant UUG codon. Two of these suppressors prove to be in genes coding for the α and β subunits of eIF2, the factor that binds Met-tRNA$_i$ to the P site. The mutation in eIF β2 resides in a part of the protein that is almost certainly involved in binding nucleic acid. It seems likely that its target is either the initiation sequence of mRNA as such or the base-paired association between the mRNA codon and tRNA$_i^{Met}$ anticodon. This suggests that eIF2 participates in the discrimination of initiation codons as well as bringing the initiator tRNA to the P site.

The cost of protein synthesis in terms of high-energy bonds may be increased by proofreading processes. An important question in calculating the cost of protein synthesis is the stage at which the decision is taken on whether to accept a tRNA. If a decision occurs immediately to release an aminoacyl-tRNA·EF-Tu·GTP complex,

there is little extra cost for rejecting the large number of incorrect tRNAs that are likely (statistically) to enter the A site before the correct tRNA is recognized. But if GTP is hydrolyzed before the mismatched aminoacyl-tRNA dissociates, the cost will be greater. A mismatched aminoacyl-tRNA can be rejected either before or after the cleavage of GTP, although we do not know yet where on average it is rejected. There is some evidence that the use of GTP *in vivo* is greater than the three high-energy bonds that are used in adding every (correct) amino acid to the chain

Last updated on 3-21-2002

Reviews

437. Kurland, C. G. (1992). *Translational accuracy and the fitness of bacteria*. Annu. Rev. Genet. 26, 29-50.

2390. Ramakrishnan, V. (2002). *Ribosome structure and the mechanism of translation.* Cell 108, 557-572.

References

1185. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). *Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics*. Nature 407, 340-348.

2392. Ogle, J. M., Brodersen, D. E., Clemons , W. M., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001). *Recognition of cognate transfer RNA by the 30S ribosomal subunit.* Science 292, 897-902.

USING THE GENETIC CODE

2.7.16 Recoding changes codon meanings

Key Terms

Recoding events occur when the meaning of a codon or series of codons is changed from that predicted by the genetic code. It may involve altered interactions between aminoacyl-tRNA and mRNA that are influenced by the ribosome.

Key Concepts

- Changes in codon meaning can be caused by mutant tRNAs or by tRNAs with special properties.
- The reading frame can be changed by frameshifting or bypassing, both of which depend on properties of the mRNA.

The reading frame of a messenger usually is invariant. Translation starts at an AUG codon and continues in triplets to a termination codon. Reading takes no notice of sense: insertion or deletion of a base causes a frameshift mutation, in which the reading frame is changed beyond the site of mutation. Ribosomes and tRNAs continue ineluctably in triplets, synthesizing an entirely different series of amino acids.

There are some exceptions to the usual pattern of translation that enable a reading frame with an interruption of some sort – such as a nonsense codon or frameshift – to be translated into a full-length protein. **Recoding** events are responsible for making exceptions to the usual rules, and can involve several types of events.

Changing the meaning of a single codon allows one amino acid to be substituted in place of another, or for an amino acid to be inserted at a termination codon. **Figure 7.28** shows that these changes rely on the properties of an individual tRNA that responds to the codon:

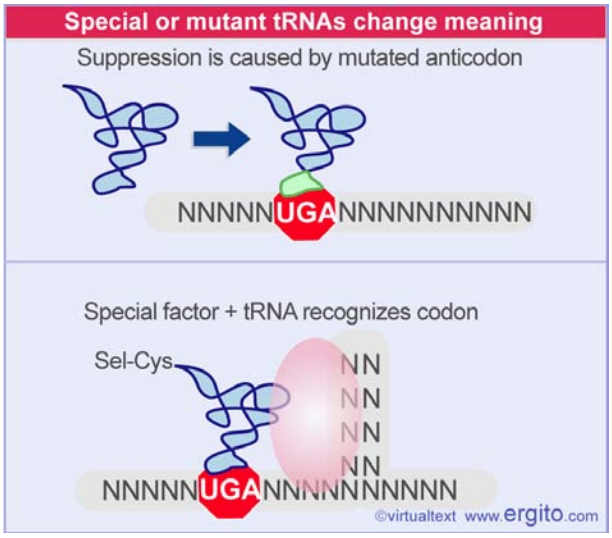

Figure 7.28 A mutation in an individual tRNA (usually in the anticodon) can suppress the usual meaning of that codon. In a special case, a specific tRNA is bound by an unusual elongation factor to recognize a termination codon adjacent to a hairpin loop.

- Suppression involves recognition of a codon by a (mutant) tRNA that usually would respond to a different codon (see *Molecular Biology 2.7.12 Suppressor tRNAs have mutated anticodons that read new codons*).
- Redefinition of the meaning of a codon occurs when an aminoacyl-tRNA is modified (see *Molecular Biology 2.7.8 Novel amino acids can be inserted at certain stop codons*).

Changing the reading frame occurs in two types of situation:

- Frameshifting typically involves changing the reading frame when aminoacyl-tRNA slips by one base (+1 forward or –1 backward) (see *Molecular Biology 2.7.17 Frameshifting occurs at slippery sequences*). The result shown in **Figure 7.29** is that translation continues past a termination codon.
- Bypassing involves a movement of the ribosome to change the codon that is paired with the peptidyl-tRNA in the P site. The sequence between the two codons fails to be represented in protein. As shown in **Figure 7.30**, this allows translation to continue past any termination codons in the intervening region.

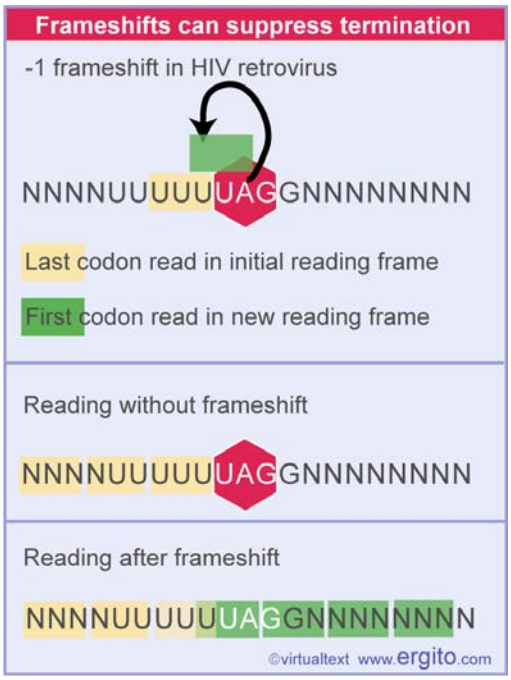

Figure 7.29 A tRNA that slips one base in pairing with a codon causes a frameshift that can suppress termination. The efficiency is usually ~5%.

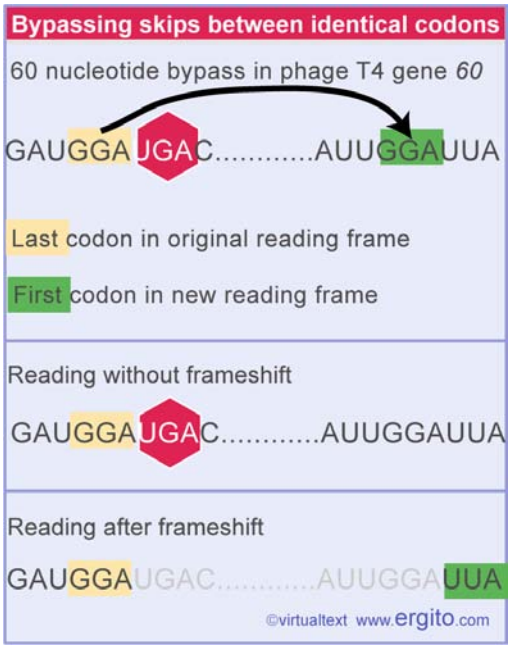

Figure 7.30 Bypassing occurs when the ribosome moves along mRNA so that the peptidyl-tRNA in the P site is released from pairing with its codon and then repairs with another codon farther along.

Last updated on 2-1-2002

USING THE GENETIC CODE

2.7.17 Frameshifting occurs at slippery sequences

Key Terms

Programmed frameshifting is required for expression of the protein sequences coded beyond a specific site at which a +1 or -1 frameshift occurs at some typical frequency.

Key Concepts

- The reading frame may be influenced by the sequence of mRNA and the ribosomal environment.
- Slippery sequences allow a tRNA to shift by 1 base after it has paired with its anticodon, thereby changing the reading frame.
- Translation of some genes depends upon the regular occurrence of programmed frameshifting.

Frameshifting is associated with specific tRNAs in two circumstances (for review see 3064):

- Some mutant tRNA suppressors recognize a "codon" for 4 bases instead of the usual 3 bases.
- Certain "slippery" sequences allow a tRNA to move a base up or down mRNA in the A site.

Frameshift mutants result from the insertion or deletion of a base. They can be suppressed by restoring the original reading frame. This can be achieved by compensating base deletions and insertions within a gene (see *Molecular Biology 1.1.21 The genetic code is triplet*). However, extragenic frameshift suppressors also can be found in the form of tRNAs with aberrant properties.

The simplest type of external frameshift suppressor corrects the reading frame when a mutation has been caused by inserting an additional base within a stretch of identical residues. For example, a G may be inserted in a run of several contiguous G bases. The frameshift suppressor is a $tRNA^{Gly}$ that has an extra base inserted in its anticodon loop, converting the anticodon from the usual triplet sequence CCC $\leftarrow$ to the quadruplet sequence CCCC $\leftarrow$. The suppressor tRNA recognizes a 4-base "codon".

Some frameshift suppressors can recognize more than one 4-base "codon". For example, a bacterial $tRNA^{Lys}$ suppressor can respond to either AAAA or AAAU, instead of the usual codon AAA. Another suppressor can read any 4-base "codon" with ACC in the first three positions; the next base is irrelevant. In these cases, the alternative bases that are acceptable in the fourth position of the longer "codon" are

not related by the usual wobble rules. The suppressor tRNA probably recognizes a 3 base codon, but for some other reason – most likely steric hindrance – the adjacent base is blocked. This forces one base to be skipped before the next tRNA can find a codon.

Situations in which frameshifting is a normal event are presented by phages and viruses. Such events may affect the continuation or termination of protein synthesis, and result from the intrinsic properties of the mRNA.

In retroviruses, translation of the first gene is terminated by a nonsense codon in phase with the reading frame. The second gene lies in a different reading frame, and (in some viruses) is translated by a frameshift that changes into the second reading frame and therefore bypasses the termination codon (see **Figure 7.29**) (2324) (see *Molecular Biology 4.17.3 Retroviral genes codes for polyproteins*). The efficiency of the frameshift is low, typically ~5%. In fact, this is important in the biology of the virus; an increase in efficiency can be damaging. **Figure 7.31** illustrates the similar situation of the yeast Ty element, in which the termination codon of *tya* must be bypassed by a frameshift in order to read the subsequent *tyb* gene.

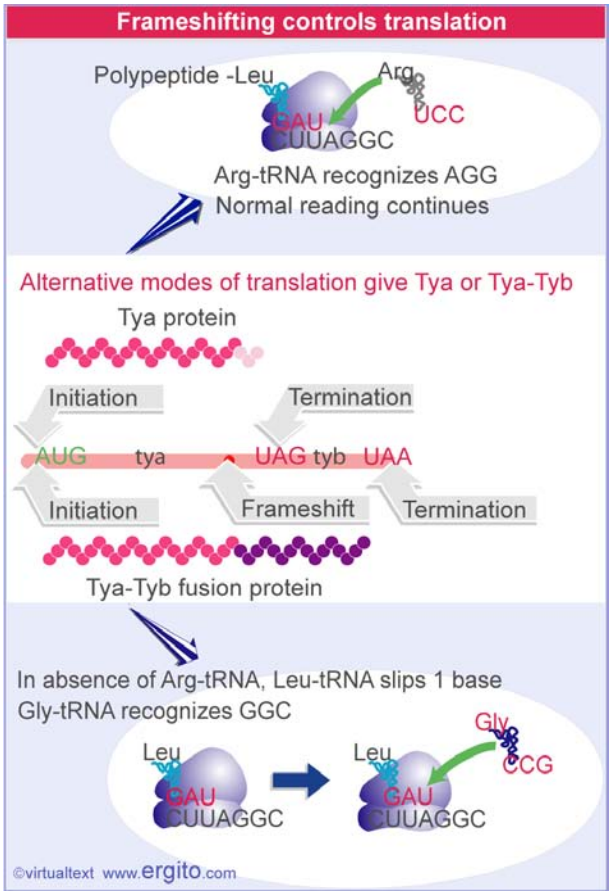

Figure 7.31 A +1 frameshift is required for expression of the *tyb* gene of the yeast Ty element. The shift occurs at a 7 base sequence at which two Leu codon(s) are followed by a scarce Arg codon.

Such situations makes the important point that the rare (but predictable) occurrence of "misreading" events can be relied on as a necessary step in natural translation. This is called **programmed frameshifting** (for review see 43; 44). It occurs at particular sites at frequencies that are 100-1000× greater than the rate at which errors are made at nonprogrammed sites (~3×10^{-5} per codon).

There are two common features in this type of frameshifting:

- A "slippery" sequence allows an aminoacyl-tRNA to pair with its codon and then to move +1 (rare) or –1 base (more common) to pair with an overlapping triplet sequence that can also pair with its anticodon.
- The ribosome is delayed at the frameshifting site to allow time for the aminoacyl-tRNA to rearrange its pairing. The cause of the delay can be an adjacent codon that requires a scarce aminoacyl-tRNA, a termination codon that is recognized slowly by its release factor, or a structural impediment in mRNA (for example, a "pseudoknot," a particular conformation of RNA) that impedes the ribosome.

Slippery events can involve movement in either direction; a –1 frameshift is caused when the tRNA moves backwards, and a +1 frameshift is caused when it moves forwards. In either case, the result is to expose an out-of-phase triplet in the A site for the next aminoacyl-tRNA. The frameshifting event occurs before peptide bond synthesis. In the most common type of case, when it is triggered by a slippery sequence in conjunction with a downstream hairpin in mRNA, the surrounding sequences influence its efficiency.

The frameshifting in **Figure 7.31** shows the behavior of a typical slippery sequence. The 7 nucleotide sequence CUUAGGC is usually recognized by Leu-tRNA at CUU followed by Arg-tRNA at AGC. However, the Arg-tRNA is scarce, and when its scarcity results in a delay, the Leu-tRNA slips from the CUU codon to the overlapping UUA triplet. This causes a frameshift, because the next triplet in phase with the new pairing (GGC) is read by Gly-tRNA. Slippage usually occurs in the P site (when the Leu-tRNA actually has become peptidyl-tRNA, carrying the nascent chain).

Frameshifting at a stop codon causes readthrough of the protein. The base on the 3′ side of the stop codon influences the relative frequencies of termination and frameshifting, and thus affects the efficiency of the termination signal. This helps to explain the significance of context on termination.

Last updated on 2-1-2002

Reviews

43. Farabaugh, P. J. (1995). *Programmed translational frameshifting.* Microbiol. Rev. 60, 103-134.

44. Gesteland, R. F. and Atkins, J. F. (1996). *Recoding: dynamic reprogramming of translation.* Annu. Rev. Biochem. 65, 741-768.

3064. Farabaugh, P. J. and Bjorkk, G. R. (1999). *How translational accuracy influences reading frame maintenance.* EMBO J. 18, 1427-1434.

References

2324. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988). *Characterization of ribosomal frameshifting in HIV-1 gag-pol expression.* Nature 331, 280-283.

USING THE GENETIC CODE

2.7.18 Bypassing involves ribosome movement

Certain sequences trigger a bypass event, when a ribosome stops translation, slides along mRNA with peptidyl-tRNA remaining in the P site, and then resumes translation (see **Figure 7.30**). This is a rather rare phenomenon, with only ~3 authenticated examples (for review see 2265). The most dramatic example of bypassing is in gene *60* of phage T4, where the ribosome moves 60 nucleotides along the mRNA (2325).

The key to the bypass system is that there are identical (or synonymous) codons at either end of the sequence that is skipped. They are sometimes referred to as the "take-off" and "landing" sites. Before bypass, the ribosome is positioned with a peptidyl-tRNA paired with the take-off codon in the P site, with an empty A site waiting for an aminoacyl-tRNA to enter. **Figure 7.32** shows that the ribosome slides along mRNA in this condition until the peptidyl-tRNA can become paired with the codon in the landing site. A remarkable feature of the system is its high efficiency, ~50%.

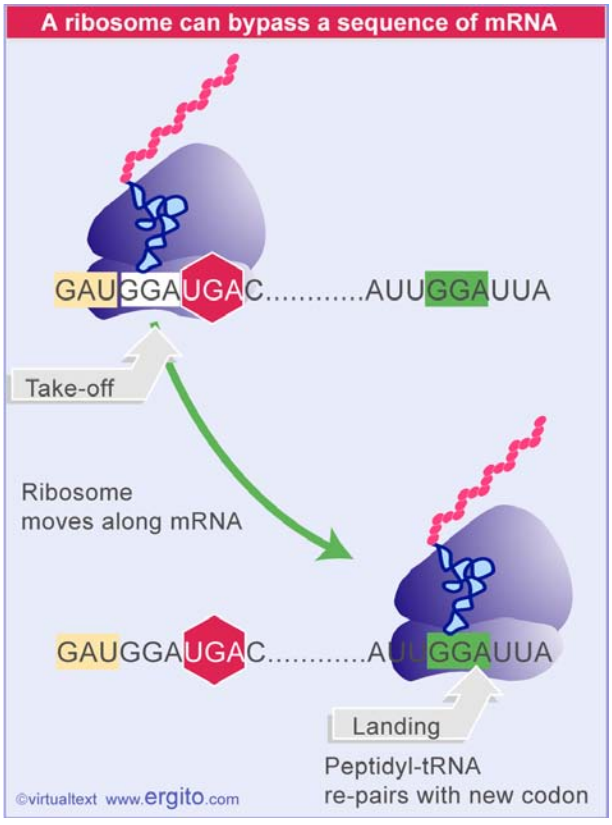

Figure 7.32 In bypass mode, a ribosome with its P site occupied can stop translation. It slides along mRNA to a site where peptidyl-tRNA pairs with a new codon in the P site. Then protein synthesis is resumed.

The sequence of the mRNA triggers the bypass. The important features are the two GGA codons for take-off and landing, the spacing between them, a stem-loop structure that includes the take-off codon, and the stop codon adjacent to the take-off codon. The protein under synthesis is also involved.

The take-off stage requires the peptidyl-tRNA to unpair from its codon. This is followed by a movement of the mRNA that prevents it from re-pairing. Then the ribosome scans the mRNA until the peptidyl-tRNA can repair with the codon in the landing reaction. This is followed by the resumption of protein synthesis when aminoacyl-tRNA enters the A site in the usual way.

Like frameshifting, the bypass reaction depends on a pause by the ribosome. The probability that peptidyl-tRNA will dissociate from its codon in the P site is increased by delays in the entry of aminoacyl-tRNA into the A site. Starvation for an amino acid can trigger bypassing in bacterial genes because of the delay that occurs when there is no aminoacyl-tRNA available to enter the A site (2326). In phage T4 gene *60*, one role of mRNA structure may be to reduce the efficiency of termination, thus creating the delay that is needed for the take-off reaction.

Last updated on 10-2-2003

Reviews

2265. Herr, A. J., Atkins, J. F., and Gesteland, R. F. (2000). *Coupling of open reading frames by translational bypassing.* Annu. Rev. Biochem. 69, 343-372.

References

2325. Huang, W. M., Ao, S. Z., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R., Winge, D., and Fang, M. (1988). *A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60.* Science 239, 1005-1012.

2326. Gallant, J. A. and Lindsley, D. (1998). *Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream.* Proc. Natl. Acad. Sci. USA 95, 13771-13776.

USING THE GENETIC CODE

2.7.19 Summary

The sequence of mRNA read in triplets 5′ → 3′ is related by the genetic code to the amino acid sequence of protein read from N- to C-terminus. Of the 64 triplets, 61 code for amino acids and 3 provide termination signals. Synonym codons that represent the same amino acids are related, often by a change in the third base of the codon. This third-base degeneracy, coupled with a pattern in which related amino acids tend to be coded by related codons, minimizes the effects of mutations. The genetic code is universal, and must have been established very early in evolution. Changes in nuclear genomes are rare, but some changes have occurred during mitochondrial evolution.

Multiple tRNAs may respond to a particular codon. The set of tRNAs responding to the various codons for each amino acid is distinctive for each organism. Codon-anticodon recognition involves wobbling at the first position of the anticodon (third position of the codon), which allows some tRNAs to recognize multiple codons. All tRNAs have modified bases, introduced by enzymes that recognize target bases in the tRNA structure. Codon-anticodon pairing is influenced by modifications of the anticodon itself and also by the context of adjacent bases, especially on the 3′ side of the anticodon. Taking advantage of codon-anticodon wobble allows vertebrate mitochondria to use only 22 tRNAs to recognize all codons, compared with the usual minimum of 31 tRNAs; this is assisted by the changes in the mitochondrial code.

Each amino acid is recognized by a particular aminoacyl-tRNA synthetase, which also recognizes all of the tRNAs coding for that amino acid. Aminoacyl-tRNA synthetases have a proofreading function that scrutinizes the aminoacyl-tRNA products and hydrolyzes incorrectly joined aminoacyl-tRNAs.

Aminoacyl-tRNA synthetases vary widely, but fall into two general groups according to the structure of the catalytic domain. Synthetases of each group bind the tRNA from the side, making contacts principally with the extremities of the acceptor stem and the anticodon stem-loop; the two types of synthetases bind tRNA from opposite sides. The relative importance attached to the acceptor stem and the anticodon region for specific recognition varies with the individual tRNA.

Mutations may allow a tRNA to read different codons; the most common form of such mutations occurs in the anticodon itself. Alteration of its specificity may allow a tRNA to suppress a mutation in a gene coding for protein. A tRNA that recognizes a termination codon provides a nonsense suppressor; one that changes the amino acid responding to a codon is a missense suppressor. Suppressors of UAG and UGA codons are more efficient than those of UAA codons, which is explained by the fact that UAA is the most commonly used natural termination codon. But the efficiency of all suppressors depends on the context of the individual target codon.

Frameshifts of the +1 type may be caused by aberrant tRNAs that read "codons" of 4 bases. Frameshifts of either +1 or –1 may be caused by slippery sequences in mRNA that allow a peptidyl-tRNA to slip from its codon to an overlapping sequence that

can also pair with its anticodon. This frameshifting also requires another sequence that causes the ribosome to delay. Frameshifts determined by the mRNA sequence may be required for expression of natural genes. Bypassing occurs when a ribosome stops translation and moves along mRNA with its peptidyl-tRNA in the P site until the peptidyl-tRNA pairs with an appropriate codon; then translation resumes.