Transcription and Translation

These terms describe the two steps used to transform the information carried in genes into useful products.

The final product of some genes are RNA molecules.
DNA ——Transcription——▶ RNA

The final product for other genes are protein molecules.
First the DNA gene is transcribed into a mRNA molecule and then the mRNA is translated into a protein.
DNA ——Transcription——▶ mRNA ——Translation——▶ protein
decoding the genetic code

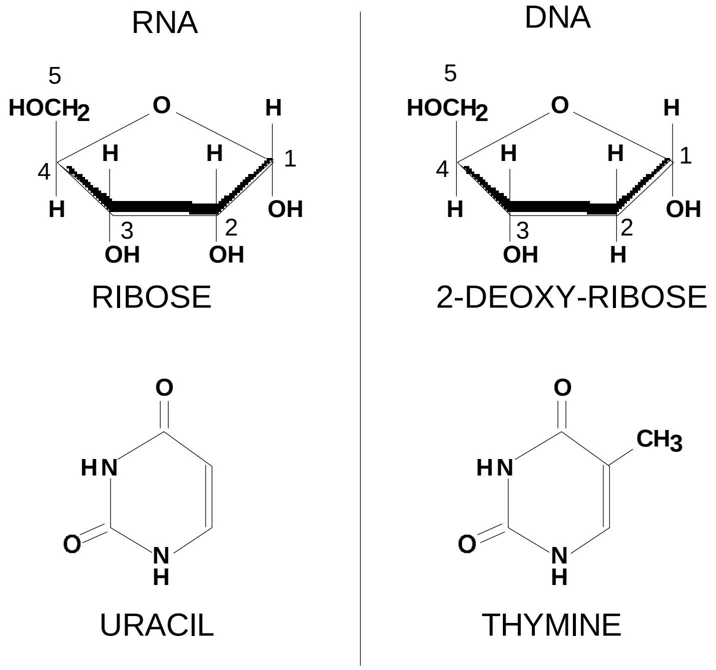
RNA

Ribonucleic acid (RNA) is very much like DNA. It has a phosphodiester linked sugar backbone and uses primarily 4 different nitrogenous bases. The bases are A, G, C and U. U stands for uracil. It is like a different dialect of the same language.

Two ways in which RNA is different than DNA.

1) The sugar is ribose NOT deoxyribose.
2) The nitrogenous base uracil is used in place of thymine. <u>Uracil can also base pair with adenine.</u> <u>Uracil is a pyrimidine.</u>

Molecular Differences between Ribonucleic Acid (RNA) & 2-deoxy-ribonucleic acid (DNA).



Structure of RNA

Some RNAs are essentially completely single stranded (mRNAs for instance).
Some RNAs have some double stranded regions (rRNAs & tRNAs).
IN RNA, G BASE PAIRS WITH C AND A BASE PAIRS WITH U (URACIL).

What is RNA used for?

The most well-known use of RNA is in transcription and translation. However, some viruses use RNA and not DNA as their genetic material.

Before we continue some terminology

Nucleotide Name Table

	Purines		Pyrimidines		
	Adenine (A)	Guanine (G)	Cytosine (C)	Thymine (T)	Uracil (U)
Nucleotides in DNA	deoxyadenylate	deoxyguanylate	deoxycytidylate	deoxythymidylate or thymidylate	
Nucleotides in RNA	adenylate	guanylate	cytidylate		uridylate

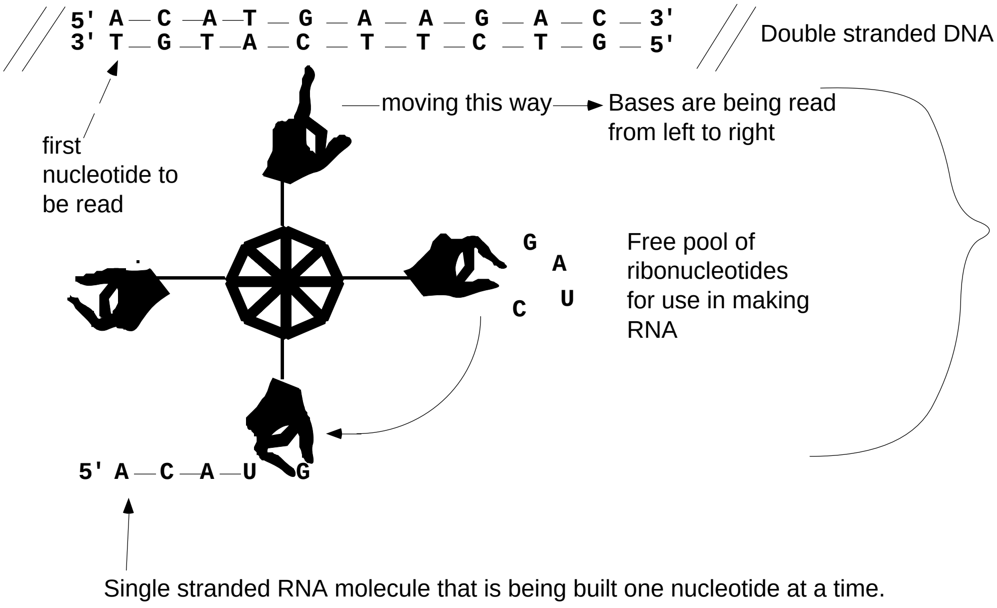
Abbreviations

Nucleoside monophosphates	AMP	GMP	CMP	TMP	UMP
Nucleoside diphosphates	ADP	GDP	CDP	TDP	UDP
Nucleoside triphosphates	ATP	GTP	CTP	TTP	UTP

For deoxynucleotides add 'd' in front of the above three.

Transcription

Transcription: DNA is transcribed into RNA



Transcripion
The transcription machinery reads the nucleotides found in one of the DNA strands and builds a **complementary** RNA molecule. If it reads G in the DNA then it will put a C on the end of the RNA molecule. If it reads A in the DNA then it will add a U on the end of the RNA molecule. That is: C is complementary to G and U is complementary to A.

Notice:

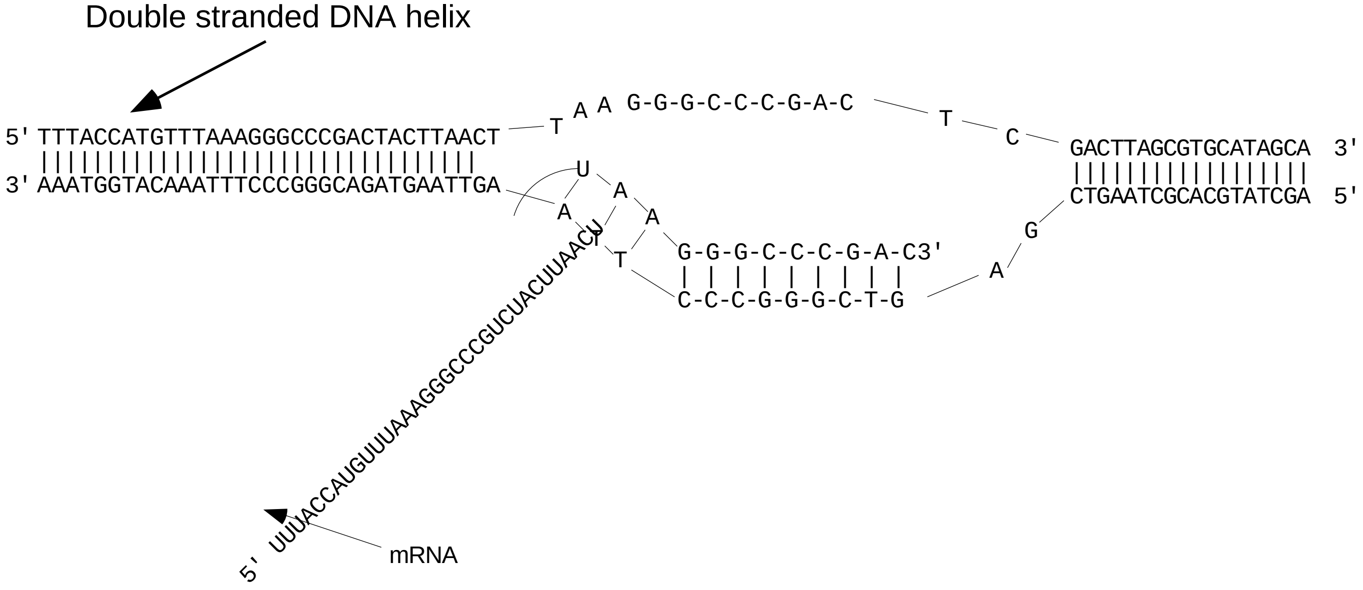
1. In the diagram above it is the bottom DNA strand that is actually being read.
2. The RNA that is being synthesized is complementary and antiparallel to this DNA strand.

Learn these two terms NOW!!!

- *Coding strand* refers to the DNA strand that has the same sequence as the transcribed RNA.
- *Template strand* refers the other strand. It is the template that was actually used to synthesize the RNA. It is antiparallel and complementary to the RNA.

Antisense and sense strands are also antiparallel and complementary to one another.

A bit closer look at transcription



This diagram is full of very important information. It shows a mRNA molecule being transcribed from a double stranded DNA molecule. The transcription machinery is not shown.

Notice that:

1. The DNA template strand is being read in the 3' to 5' direction.
2. The RNA molecule is being synthesized in the 5' to 3' direction.
3. The mRNA is antiparallel and complementary to the DNA strand that is being read.
4. About 12 nucleotides of the mRNA are H-bonded to the DNA strand.
   Actually, it is just such H-bonding that enables the transcription machinery to choose the correct ribonucleotides when synthesizing an RNA molecule. Ribonucleotides are chosen based on their ability to H-bond or base pair to the DNA template.

Translation

Translation: mRNA is translated into a protein.

What are proteins?

A protein is made of a specific sequence of amino acids. If the amino acid sequence of a protein is wrong then it won't function properly. Amino acids are substantially different than nucleotides and yet the sequence of amino acids in a protein is encoded in the DNA.

The amino acids typically found in proteins are:

phenylalanine	F	**serine**	S	**tyrosine**	Y	**cysteine**	C
leucine	L	**tryptophan**	W	**proline**	P	**histidine**	H
arginine	R	**glutamine**	Q	**methionine**	M	**isoleucine**	I
threonine	T	**asparagine**	N	**valine**	V	**alanine**	A
glutamic acid	E	**glycine**	G	**aspartic acid**	D	**lysine**	K

Next to the name of each amino acid its single-letter abbreviation.

How can the 4 nucleotides found in DNA encode the 20 different amino acids used to make a particular protein?

The sugar phosphate backbone is the same throughout the DNA molecule and so information cannot be carried here. The genetic information is carried in the order of the nitrogenous bases.

Genetic Code

Nucleic acids carry the amino acid sequence of proteins in the form of a genetic code. This code is read in groups of three nucleotides. Each group of three is called a codon (also referred to as a triplet codon). The genetic code is a language in which all words are 3 letters long, and in which there is no punctuation, or spaces between the words. The codons are non-overlapping.

Genetic Code Table

FIRST POSITION	SECOND POSITION				THIRD POSITION
	U	**C**	**A**	**G**	
U	F	S	Y	C	**U**
	F	S	Y	C	**C**
	L	S	Stop	Stop	**A**
	L	S	Stop	W	**G**
C	L	P	H	R	**U**
	L	P	H	R	**C**
	L	P	Q	R	**A**
	L	P	Q	R	**G**
A	I	T	N	S	**U**
	I	T	N	S	**C**
	I	T	K	R	**A**
	M	T	K	R	**G**
G	V	A	D	G	**U**
	V	A	D	G	**C**
	V	A	E	G	**A**
	V	A	E	G	**G**

UCAG on the left, top and right sides represent nitrogenous bases in the codon.
Examples of how to use this table:
The codon UUU and UUC both encode the amino acid F (phenylalanine).
The codons GCU, GCC, GCA and GCG all encode the amino acid A (alanine).
The amino acid M (methionine) is encoded only by the codon AUG.

Notice that we have written the genetic code in ribonucleotides. You can tell because we have used U in place of T. You will also see it written in terms of deoxynucleotides.

The genetic code is degenerate.

In this context, degeneracy means that some amino acids are encoded by more than one codon. That is, some codons have synonyms.

What the heck are stop codons?

Translation is performed by the ribosomes. When a ribosome encounters a stop codon it dissociates from the mRNA and releases the protein. Stop codons are found at the end of the protein coding region. The codons UAA, UGA and UAG are stop codons.

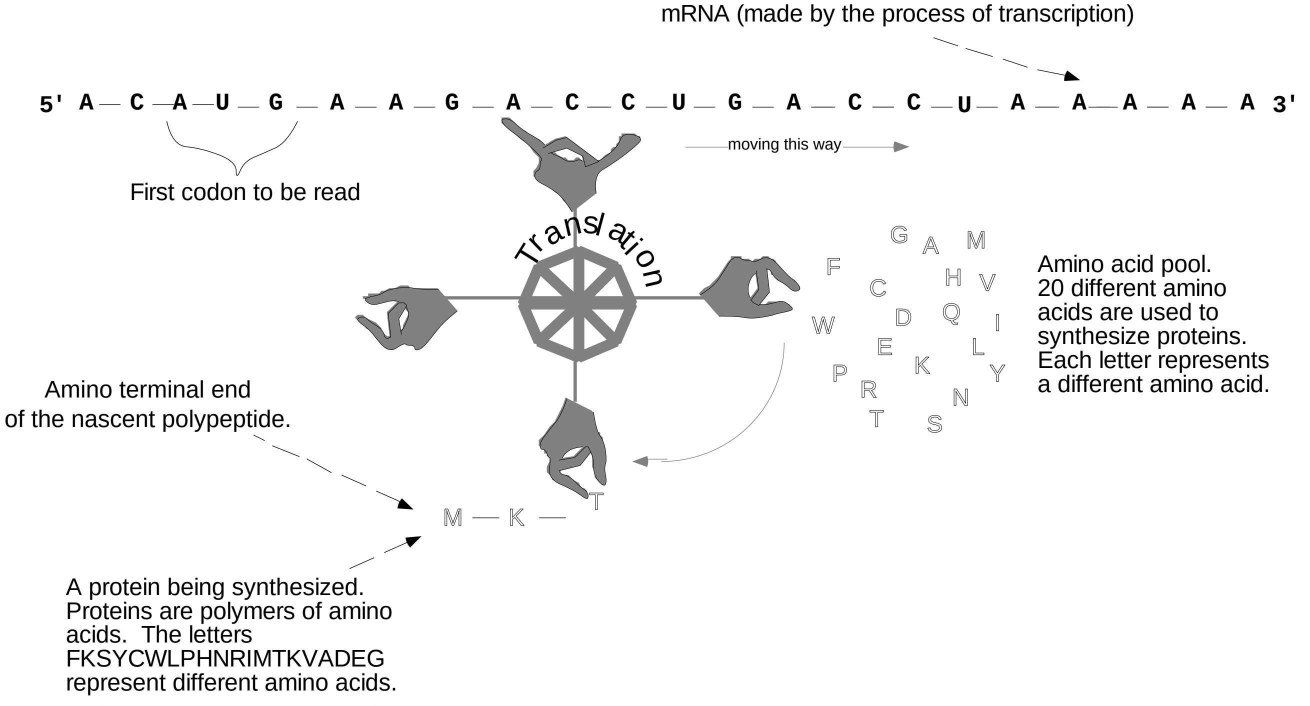
The genetic code is universal

The genetic code shown above is used, essentially unchanged, in all organisms. There are minor exceptions to this truth, but you should think of these as different dialects of the same language.

Be completely sure that you know how to read the genetic code table and could use it to translate a mRNA.

The Process of Translation

Translation is the process of reading the information carried in a mRNA molecule and using that information to direct the synthesis of a protein molecule. It is performed by the **ribosomes** using both **mRNA**, **tRNA** and **amino acids**.



mRNA (made by the process of transcription)

First codon to be read

Amino acid pool. 20 different amino acids are used to synthesize proteins. Each letter represents a different amino acid.

Amino terminal end of the nascent polypeptide.

A protein being synthesized. Proteins are polymers of amino acids. The letters FKSYCWLPHNRIMTKVADEG represent different amino acids.

The mRNA carries the information necessary to build a specific protein. Specifically, it describes the order of amino acids that the protein should contain. The mRNA is like a sentence. Each 'word' is 3 nucleotides long and is called a codon. There are no spaces and no punctuation. Each codon specifies a single amino acid. The translation machinery reads the mRNA a codon at time and joins together the amino acids specified by the mRNA.

Proteins are synthesized in the amino terminus to carboxy terminus direction.

In the example shown above:

The first amino acid M is methionine it is encoded by the codon AUG. The second amino acid K is lysine it is encoded by the codon AAG. The third amino acid T is threonine it is encoded by the codon ACC.

Not all RNAs are translated. In fact, only messenger RNA (mRNA) is translated. Both transfer RNA (tRNA) and the ribosomal RNAs (rRNA) are actually part of the translation machinery.

What are protein molecules composed of?

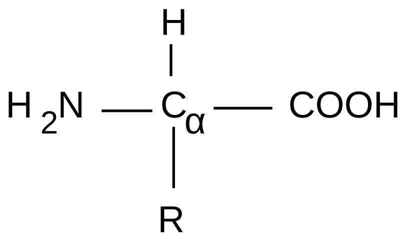
Proteins are polymers of amino acids.

What are amino acids?

These molecule are called amino acids because (with one exception) they each contain an amino group (NH_2) and an acidic carboxylic group (COOH).

General structure of amino acids---------->

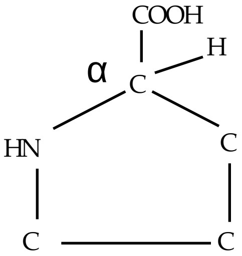
The alpha carbon is next to the COOH group.
Attached to the α carbon is a hydrogen and a nitrogen.
The nitrogen is part of an amino group.
The side chain marked R is different in each amino acid and gives them their characteristic properties.



$$\mathrm{H_2N-}\underset{\underset{\mathrm{R}}{|}}{\overset{\overset{\mathrm{H}}{|}}{\mathrm{C}_\alpha}}\mathrm{-COOH}$$

An exception to the above general structure.
In the amino acid, proline, the nitrogen is not part of an amino groups but is part of an imino ring that involves the α carbon.

The ring in proline causes it to be very rigid, and can produce a bend in a protein. None of the other R groups have rings that involve the α carbon.



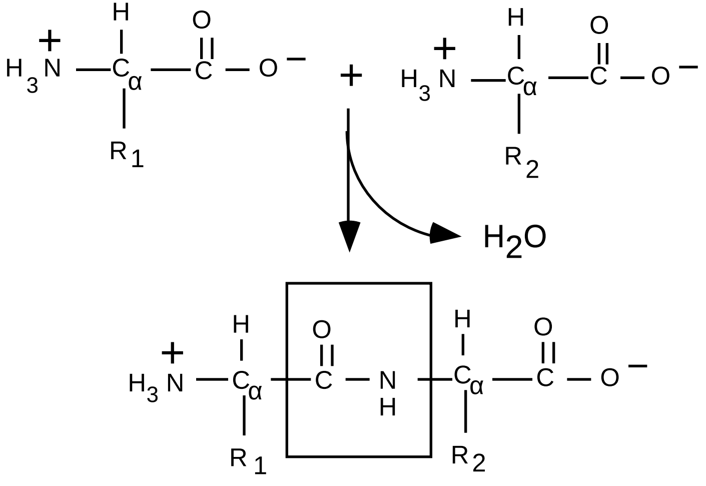
imino ring

Polypeptides or proteins are polymers of amino acids.
The term polypeptide is usually reserved for very small proteins.

In proteins, amino acids are joined to form polymers by the ribosomes.

Condensation reaction:
The amino acids are joined in a condensation reaction. This means that the reaction liberates a water molecule.

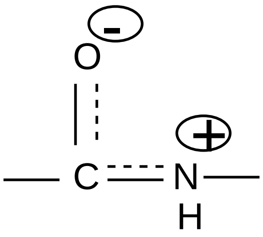
The newly formed bond between the amino acids is called a **peptide bond.**



Peptide Bond

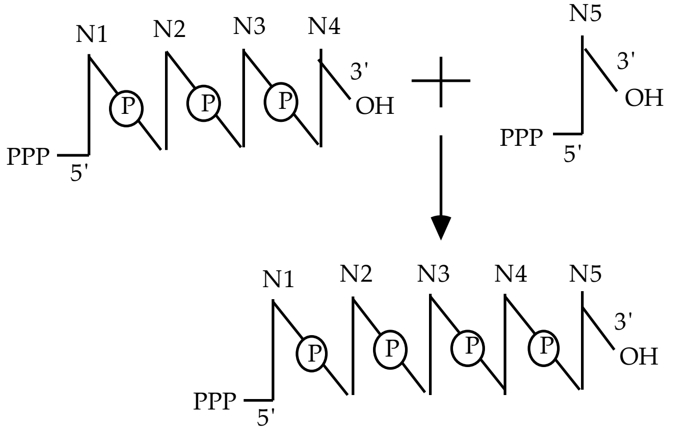
Actually we should draw the peptide bond as ---------->

Both O and N of the peptide bond have partial double bond characteristics.

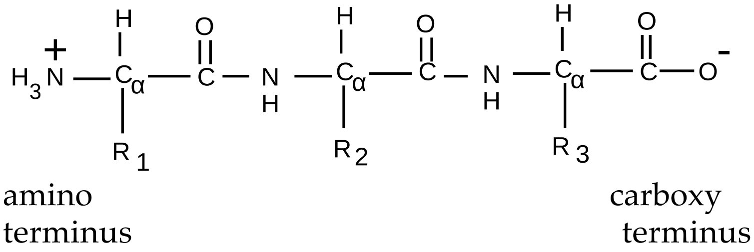


This double bond character causes the peptide bond to be flat (planar) and restricts the rotation about the peptide bond. Flexibility in proteins is primarily limited to the rotation around the α carbon's 2 single bonds. *This is not true for proline.* With proline the involvement of the α carbon in the imino ring further restriction the flexibility of the protein.

You remember that nucleic acids have an inherent polarity and that they are synthesized in a particular direction. ------------------> Right?

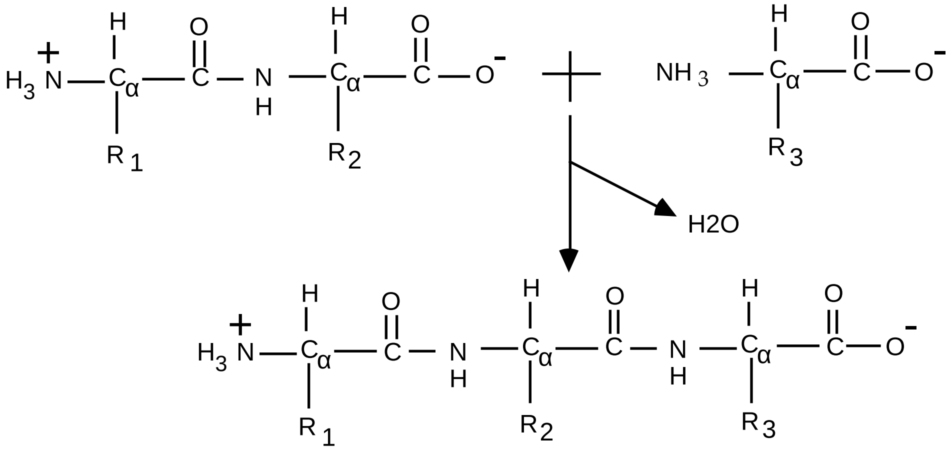


Well, proteins also have an inherent polarity and are also synthesized in a particular direction.
One end is called the amino terminus and the other the carboxy terminus.



$$H_3\overset{+}{N}-C_\alpha H(R_1)-C(=O)-NH-C_\alpha H(R_2)-C(=O)-NH-C_\alpha H(R_3)-C(=O)-O^-$$

Proteins are synthesized in the amino to carboxy direction.



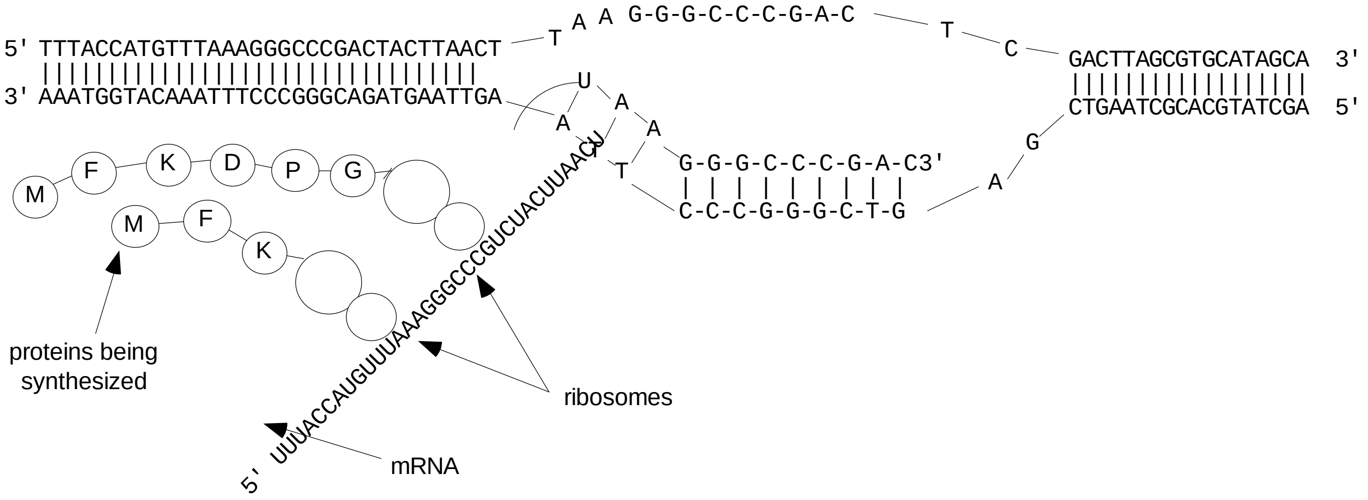
How do the ribosomes know which amino acids to join together to form a protein?

Well, of course, this information is carried in a gene. A mRNA copy of the gene is made via transcription and then this copy is translated into a protein. That is;

DNA ------------------> mRNA ------------------> protein

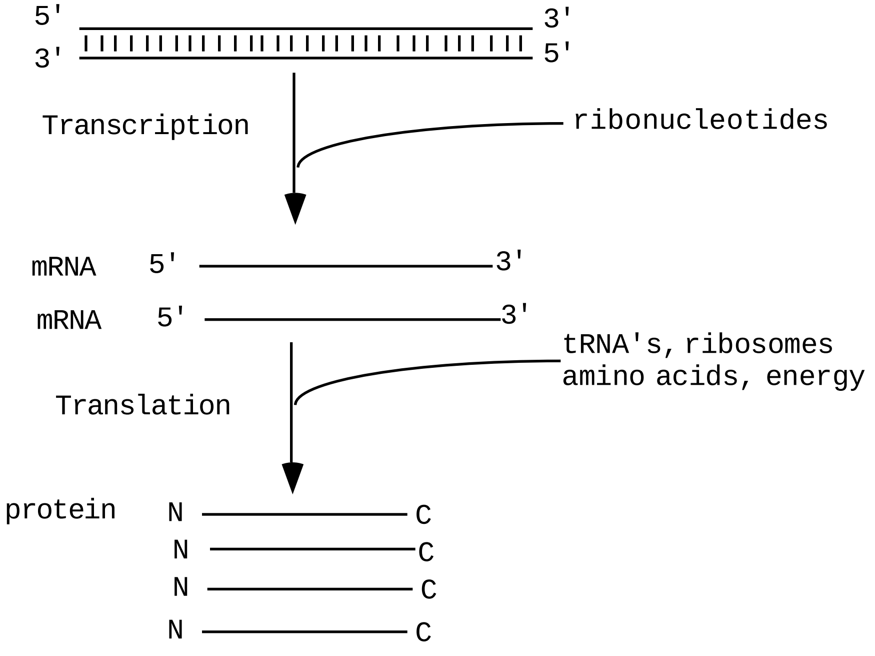
The translation machinery 'knows' which amino acids to incorporate into the growing polypeptide because it reads this information from a mRNA molecule.

Let's look at this in more detail.



This diagram show concomitant translation and transcription. Before transcription is finished, translation has begun. It is the kind of translation that occurs only in prokaryotic cells. Translation nearly always begins with the codon AUG. It is called the ***start codon***. Each trinucleotide codon is read in succession until a ***stop codon*** is reached. Stop codons are UAA, UAG and UGA. The amino terminus is the first to be synthesized and the carboxy terminus of the protein is the last part to be translated.

Many copies of an RNA and protein molecule can be made from a single gene.



<----Gene

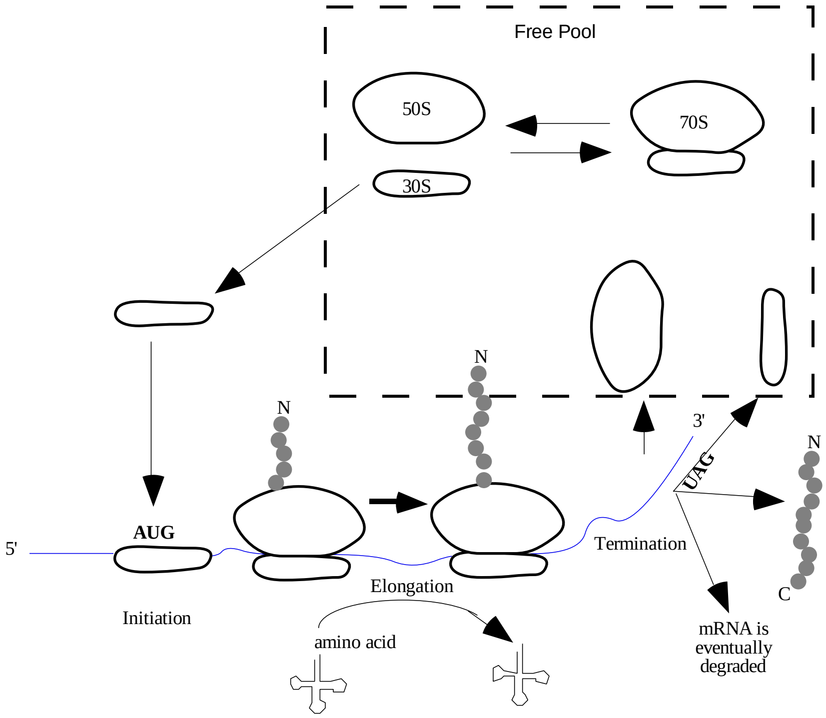
Transcription can produce more than one mRNA molecule from a gene.

Translation can synthesize many proteins from a single mRNA.

Three types of RNA molecules participate in translation.

1. mRNA carries the information from the DNA in a form that can be read by the ribosomes.
2. rRNA is a major component of the ribosomes.
3. tRNA carries the amino acid to the ribosomes and is actually the agent which deciphers the genetic code.

Translation Cycle



The Start codon

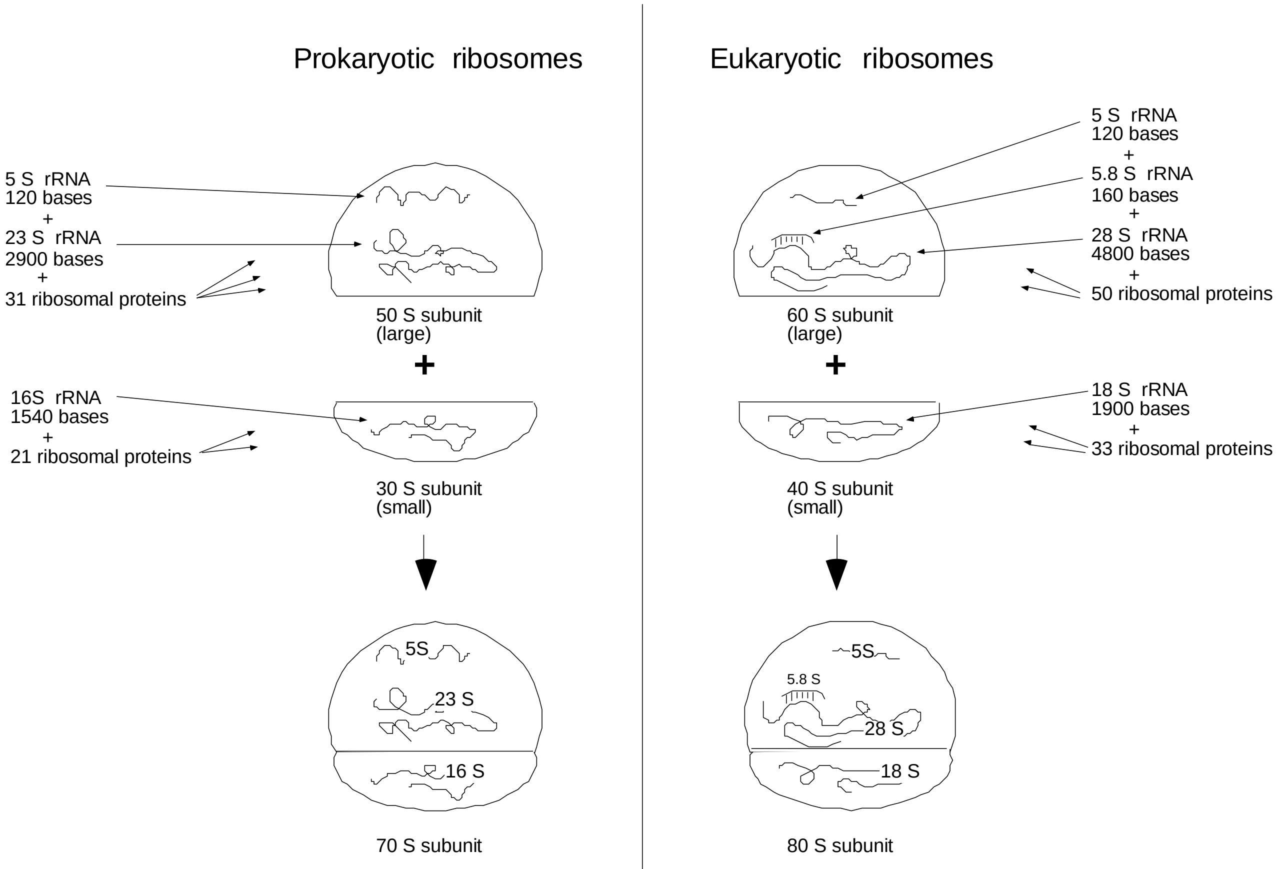
Ribosomes assemble on the mRNA around the start codon. Start codons are nearly always AUG. AUG encodes the amino acid methionine. Therefore, the first amino acid to be translated in a protein is methionine.

Stop codon

The ribosomes translate the mRNA one codon at a time until they hit a stop codon. Once the stop codon is encountered the ribosomes stop translating and dissociate into a large and small subunit and release the protein and the mRNA. The ribosomes re-enter the free ribosome pool and can be used again and again. Other ribosomes can continue to translate the mRNA or the mRNA may be degraded.

There are 3 stop codons. They are UAA, UAG and UGA. Stop codons are also called nonsense codons.

Composition of Ribosomes

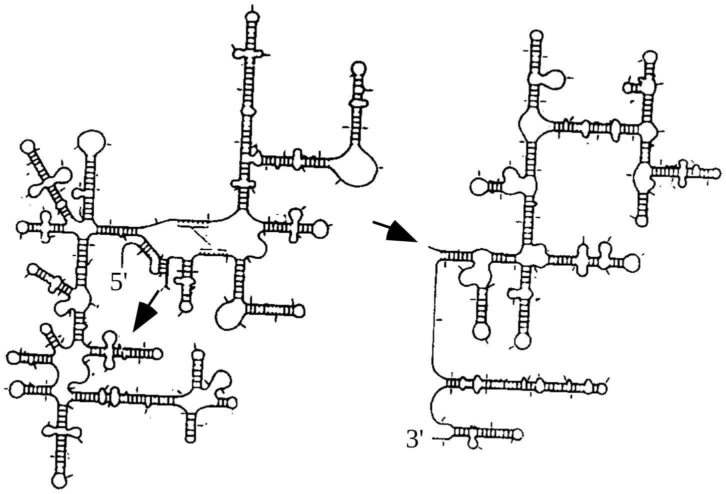


The ribosomal subunits are named according to their Svedberg coefficients (S). This is a unit of measure that describes the sedimentation rate of a particle in a centrifuge. It is particularly useful with very large macromolecular complexes. The greater the mass the larger is the Svedberg coefficient. Notice that the units are not additive.

What do ribosomes contain?

Ribosomes are composed to two different types of molecules: 1) RNA called rRNA or ribosomal RNA and 2) proteins, generically referred to as ribosomal proteins. In prokaryotes, the large ribosomal subunit contains two rRNAs (5S and 23S) and 31 different ribosomal proteins and the small subunit contains a single 16S rRNA and 21 different ribosomal proteins. Eukaryotic ribosomes are similar in composition to prokaryotic ribosomes. The exact size of the rRNAs and number of unique proteins differ between the two. One of the biggest differences is the presence of a third rRNA, the 5.8S rRNA, in the eukaryotic large subunit. Examine the diagram above very closely. *Learn the sizes of the subunits and the rRNAs.*

rRNA has extensive secondary structure



This diagram shows the secondary structure of the 16S rRNA from the bacteria, *Escherichia coli*. All rRNAs have extensive secondary and tertiary structure. The stems and loops are produced by the *intramolecular* base pairing between different parts of the molecule. This is a two-dimensional representation of the molecule. Actually, the molecule is folded in three dimensions to form a much more complex shape. This secondary structure is very similar, but not identical, in all organisms. This reflects the fact that the mechanisms of translation are evolutionarily strongly conserved.

Function of the ribosomal components

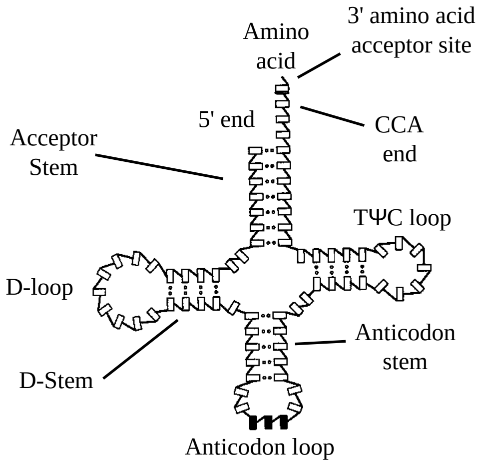
Translation requires both subunits. However, some of the labor is divided between the subunits. The small subunit recognizes the start codon and the large subunit synthesizes the peptide bond that joins the amino acids in a protein.

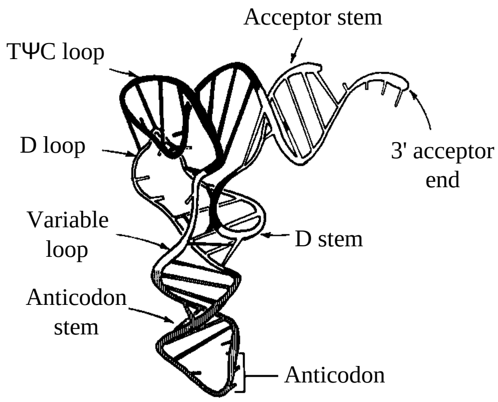
Notice that the ribosomes are composed of both RNA and protein. Originally, it was thought that the rRNA served only a structural role, that it formed a scaffold used solely for the purpose of organizing the various ribosomal proteins. Furthermore, it was thought that it was the ribosomal proteins that did all of the catalytic work of protein synthesis. *This is not entirely correct.*

It is true that the rRNA forms a scaffold, however, this is not its sole function. In the small subunit it is actually the 16S rRNA that recognizes the start codon of a mRNA. In the large subunit, it is the 23S rRNA that catalyzes peptide bond formation between amino acids. The various ribosomal proteins enhance both the specificity and efficiency of these processes.

tRNA

2-dimensional model 3-dimensional model





A tRNA is a single RNA chain that is folded into a two dimensional cloverleaf. This then folds in three dimensions to an L-like structure. tRNAs are small RNA molecules (usually in the range of 73 - 93 nucleotides) that participate in the translation of mRNAs. They are not part of the ribosome. But like rRNA they have extensive secondary structure produced by intramolecular base pairing between nucleotides. tRNAs are unique in that they contain many highly modified and unusual nucleotides. Immediately following transcription, the tRNA contains only the standard set of four nucleotides. The unusual nucleotides are produced by post-transcriptional enzymatic modifications. Some of the modifications help the cell to unambiguously recognize the tRNA. However, the function of most of the modifications is unknown.

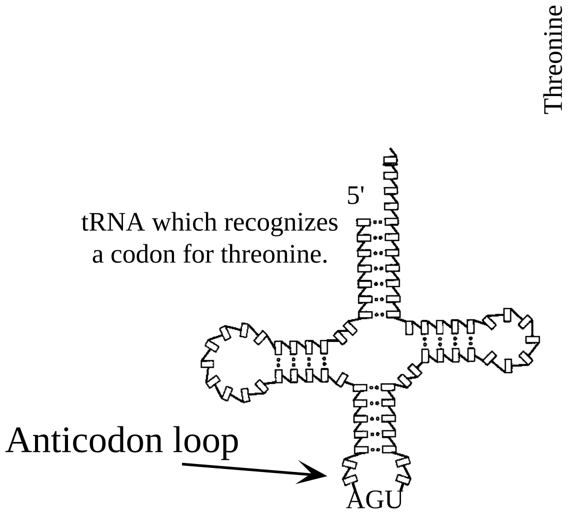
Take careful notice of the features common to all tRNAs. They are:

XCCA, D loop, TΨC loop and Anticodon loop

Notice that the tRNAs have 4 stems and 3 loops and that each of them has a name. Furthermore some tRNAs have a fourth loop called the *Variable Loop*. It is located between the TψC stem and the anticodon stem. The D loop is named for a modified nitrogenous base found within it; dihydrouracil. Similarly, the TψC loop, is named for four invariant nucleotides found within it. They are thymidylate, pseudouridylate, cytidylate, guanylate. The XCCA terminus is very important. XCCA refers to the nucleotides found at this position. X stand for any nucleotide, C for cytidylate and A for adenylate. ***The XCCA stem is the amino acid acceptor site.***

It is the tRNAs that actually decode the genetic code

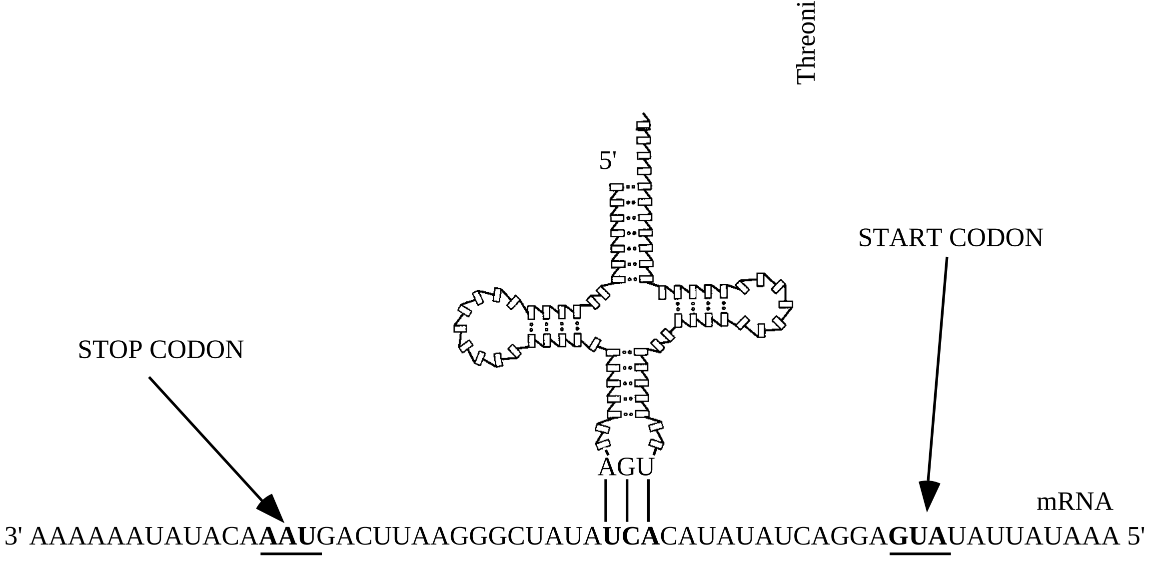
This is done via the anticodon loop.



This tRNA decodes one of the codons that specify threonine. This tRNA is referred to as: tRNA^{thr}.

The codons for threonine are ACU, ACC, ACA and ACG. This tRNA decodes the ACU codon. Notice that three of the nucleotides in the anticodon loop are complementary and antiparallel to the codon ACU. Notice that in the diagram that the tRNA is covalently attached to the amino acid threonine.

tRNAs decode codons by base pairing with them.



The anticodon of tRNA^{thr} base pairs with the threonine codon. Now, keep this straight. The anticodon is found in the tRNA and the codon is found in the mRNA. The anticodon base pairs with the codon in a complementary and antiparallel fashion. Notice that the tRNA and the mRNA are in opposite polarity. tRNAs actually decode codons within the confines of the ribosome, however, the ribosome is not shown here. We'll include it later on.

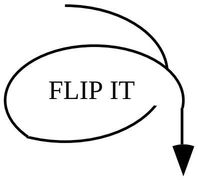
Every codon is recognized by at least one anticodon. There are 64 possible codons. Three of these are stop codons. Refer back to the Genetic Code Table to check this. Each of the codons is recognized by a tRNA except for the stop codons. The three stop codons are **NOT** recognized by any tRNA.

Wobble

There are 61 different codons that must be decoded. This would lead you to believe that cells have 61 different tRNAs, one for each codon. Right? Wrong! Prokaryotes typically have 30 to 40 different tRNAs and eukaryotes have about 50 tRNAs. How can 30 to 50 different tRNAs correctly decode 61 different codons. The answer is through wobble base pairing.

Wobble base pairing is a non-standard type of base pairing that occurs between the third nucleotide in a codon and the first nucleotide in an anticodon. It enables a single tRNA to recognize more than one codon for a particular amino acid.

```
CODONS ENCODE   M   R   R   I   Y   T   I   S
          5'//AU AUG CGA CGG AUA UAC ACU AUA UCG GG // 3'
```



```
          3'//GG GCU AUA UCA CAU AUA GGC AGC GUA UA // 5'
 CODONS ENCODE   S   I   T   Y   I   R   R   M
```

WOBBLE base pairing occurs between the 1st position of the anticodon and the third position of the codon.

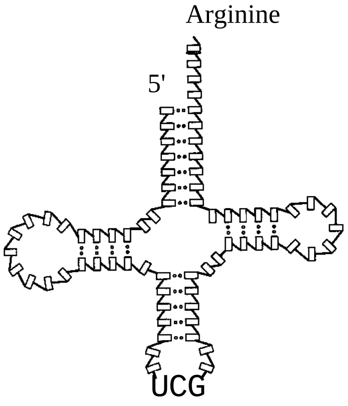
Wobble Base Pairing Table

Base in the **first** position of the **anticodon** can wobble base pair with →	Base recognized in the **third** position of the **codon**
U	A or G
C	G only
A	U only
G	C or U
I	C, A or U

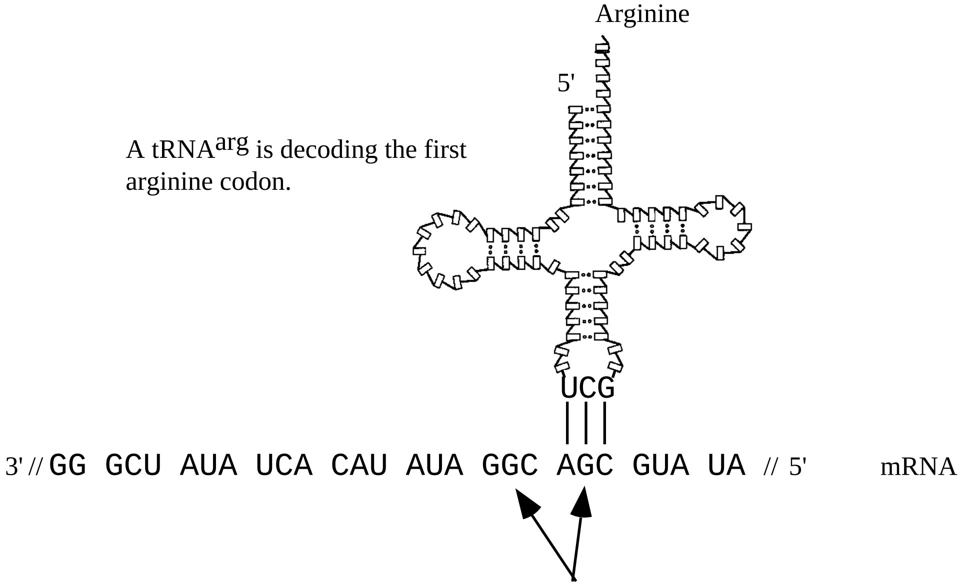
I = Inosine, one of the non-standard nitrogenous bases found in tRNAs.

Let's continue to examine this. Shown in the next figure is part of a mRNA and the amino acids that it encodes. I have spaced the codons for clarity. Usually, we write RNAs 5' to 3'. But for right now let's flip it over so that it is written 3' to 5'.
Notice that in the mRNA there are two different codons for arginine (R). If this does not make sense to you then please consult the Genetic Code Table.

Shown on the right is a tRNA^{arg} . We are going to see how this tRNA uses wobble base pairing to pair with both of the arginine codons shown above.



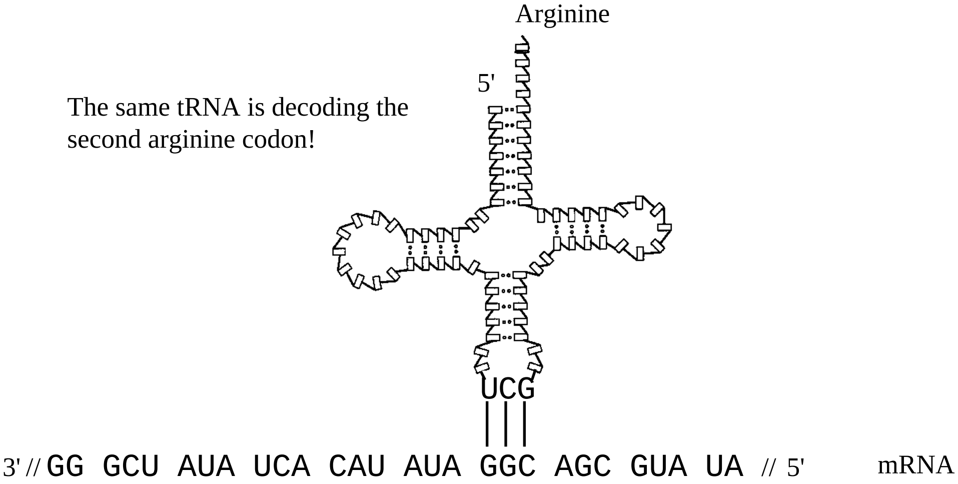
Here the tRNA^{arg} is shown base pairing with the CGA codon for arginine. This is the usual type of base pairing with which you are familiar.



A tRNA^{arg} is decoding the first arginine codon.

Both of these codons encode the amino acid arginine (R).

Here we see the same tRNA base pairing with a different arginine codon. This the non-standard type of base pairing called Wobble Base Pairing.



The same tRNA is decoding the second arginine codon!

You can consider this wobble base pairing to be unique to the base pairing between the first position of the anticodon and the third position of the codon.

The Genetic code table indicates that there are 6 different codons for arginine. They are CGU, CGC, CGA, CGG, AGA and AGG. Notice that these can be broken down into 2 groups that are identical in codon positions 1 and 2. Using wobble base pairing three different tRNAs can decode all 6 of these codons.

These are the six **CODONS** that can encode arginine.

1 2 3
5' ——CGU—— 3'

1 2 3
5' ——CGC—— 3'

1 2 3
5' ——CGA—— 3'

1 2 3
5' ——CGG—— 3'

1 2 3
5' ——AGA—— 3'

1 2 3
5' ——AGG—— 3'

Each figure represents an arginine codon in a mRNA. It is written in the 5' to 3' direction.

A minimum of three different tRNAs are required to decode all 6 of the above codons.

They are:

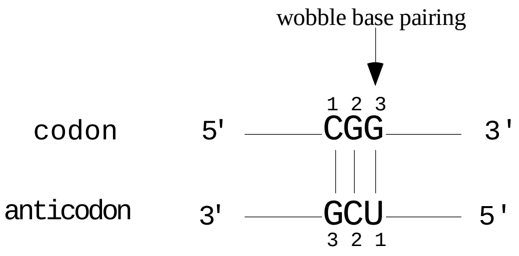
3' ——GCU—— 5'
3 2 1

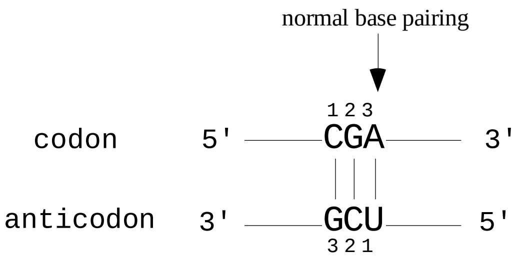
3' ——UCU—— 5'
3 2 1

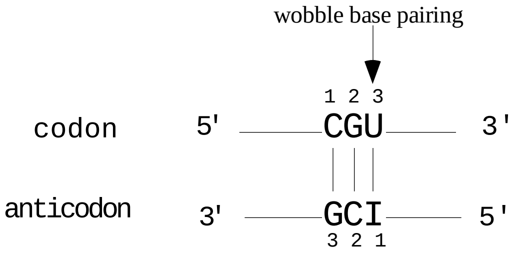
3' ——GCI—— 5'
3 2 1

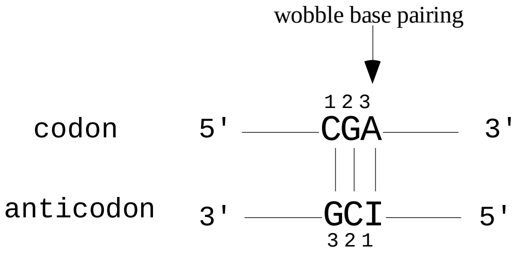
Each figure represents an anticodon in a tRNA. They are written in the 3' to 5' direction (backwards).

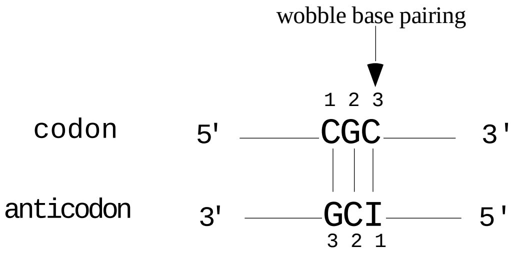
Below, I have spelled out how this happens. A minimum of two tRNAs are needed to recognize the CGX codon and a minimum of 1 tRNA is needed to recognize the AGA and AGG codons.

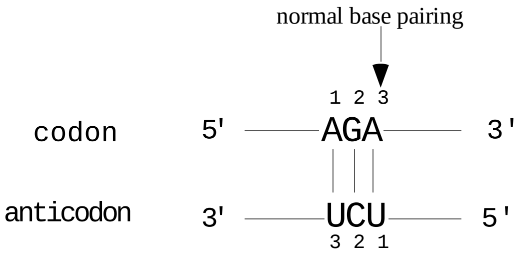


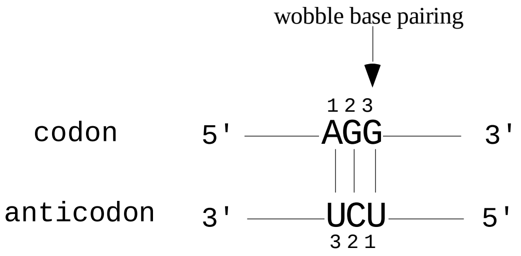








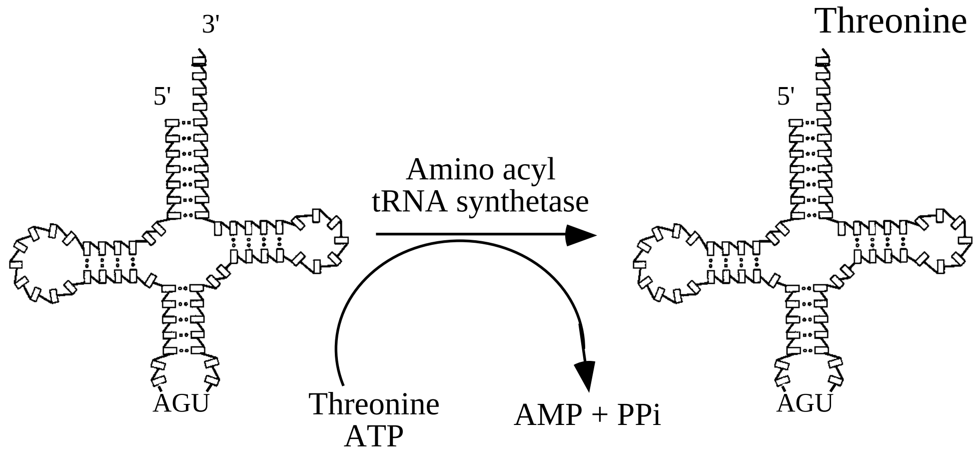




Refer back to the Wobble Base Pairing Table until this makes sense.

How are amino acids attached to a tRNA?

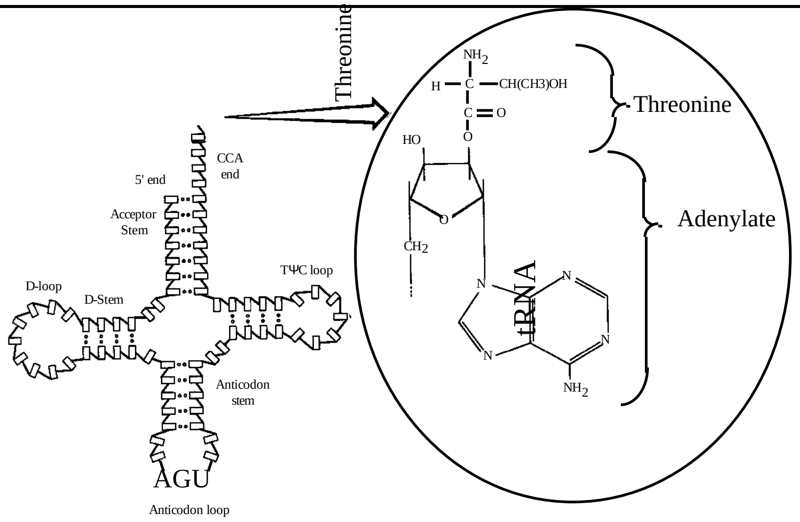
Each tRNA can only carry one specific amino acid.



A tRNA carrying an amino acid is said to be **charged**. But the amino acid that is attached to the tRNA is said to be **activated**. The enzyme that attaches the amino acid to the tRNA is called **aminoacyl tRNA synthetase.** Charging a tRNA consumes energy in the form of ATP. Energy from the ATP is conserved in bond between amino acid and tRNA.

Charged tRNA

Amino acid are attached to the tRNA through the terminal 3' adenosine residue found in tRNA.



The last three nucleotides on the 3' end of a tRNA are cytidylate (C) cytidylate (C) and adenylate (A). That is 5' CCA 3'. Amino acids are attached to the tRNA through the adenylate residue. They are attached through a carboxylic bond to the 2' or 3' carbon of the adenylate residue.

Remember that bacteria have 30 to 40 different tRNAs and eukaryotes have about 50 different tRNAs. There is one amino acid synthetase for each amino acid. That is, there are 20 different aminoacyl tRNA synthetases. Each individual synthetase charges all of the tRNAs that encode one particular amino acid. That is; each synthetase is specific for one amino acid. For example, if an organism has three tRNA^{arg} 's then the enzyme arginine aminoacyl tRNA synthetase will charge all of the three tRNAs. Some aminoacyl tRNA synthetases recognize their particular tRNAs based upon the shape of the molecule while others seem to recognize the anticodon loop itself.

Definitions

Escherichia coli	Gram-negative, rod-shaped bacteria typically found in the large intestine of warm-blooded animals. It is a prokaryote. A member of the coliform group of bacteria. Extensively used by molecular biologists for the propagation of plasmid and phage vectors. Those used in the laboratory cannot normally survive in nature. Abbreviated name is *E. coli*.
intramolecular	refers to interactions that occur between different parts of the same molecule.
intermolecular	refers to interactions that occur between 2 or more molecules.
nascent	in the process of emerging
post-translational	A modification that occurs after a protein molecule has been translated.
post-transcriptional	A modification that occurs after a RNA molecule has been transcribed.
Svedberg unit (S)	Abbreviation is "S". A unit describing the rate of sedimentation of a particle in a centrifuge. A particle with greater mass sediments faster (and has a larger S value) than a particle with lower mass. Shape also influences the S value. A more compact particle will sediment faster than a larger particle even if the two have the same mass. *Svedberg units are not additive*
primary structure	refers to the linear arrangement of monomers in a polymer. In a protein the monomers are amino acids and in a nucleic acid (DNA or RNA) the monomers are nucleotides.
Secondary structure	folding of a single polypeptide chain in three dimensions. Examples include the α-helix and ß-pleated sheet.
Tertiary structure	relative orientation of the side chains attached to a monomer backbone. In proteins this would be the side chains of each individual amino acid. In nucleic acids it would be the nitrogenous bases.
Quaternary structure	Folding of 2 or more molecules with respect to one another.