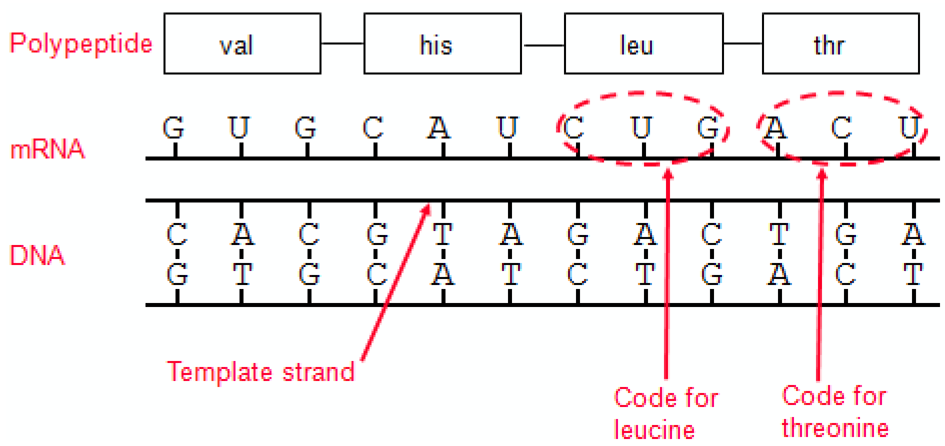
Gene Expression: Transcription and Translation

The Central Dogma

Protein synthesis requires two steps: ***transcription*** and ***translation***.

DNA contains codes

The simplistic diagram below illustrates the concept that three bases in DNA code for one amino acid. The DNA code is copied to produce mRNA. Later in this chapter, we will learn that RNA may be modified. The order of amino acids in the polypeptide is determined by the sequence of 3-letter codes in mRNA.



DNA vs RNA

	DNA	**RNA**
Sugar:	deoxyribose	Ribose
Bonds with Adenine:	thymine	Uracil
# of Strands:	two	One

Kinds of RNA

RNA has a variety of different functions in the cell. Three of these are listed below.

Messenger RNA (mRNA)

Messenger RNA contains genetic information. It is a copy of a portion of the DNA.

It carries genetic information from the gene (DNA) out of the nucleus, into the cytoplasm of the cell where it is translated to produce protein.

Ribosomal RNA (rRNA)

This type of RNA is a structural component of the ribosomes. It does not contain a genetic message.

Transfer RNA (tRNA)

Transfer RNA functions to transport amino acids to the ribosomes during protein synthesis.

Small nuclear RNA (snRNA)

These strands of RNA are complexed with protein producing small nuclear ribonucleoproteins (snRNP). One function, described later in this chapter, is the modification of the RNA transcript.

Transcription

Transcription is the synthesis of RNA from a DNA template.

Only one strand of DNA is copied.

A single gene may be transcribed thousands of times.

After transcription, the DNA strands rejoin.

Some of the RNA produced by transcription is not used for protein synthesis. These RNA molecules have other functions in the cell.

Steps involved in transcription

The enzyme ***RNA polymerase*** is responsible for creating RNA by copying the template strand of DNA.

Before transcription can begin in eukaryotes, proteins called ***transcription factors*** must bind to a region of the DNA called the ***promoter***. The promoter identifies the start of a gene, which strand is to be copied, and the direction that it is to be copied.

RNA polymerase binds to the transcription factors and the promoter.

In bacteria, RNA polymerase binds directly to the promoter without the assistance of transcription factors.

RNA polymerase unwinds the DNA.

RNA polymerase arranges nucleotides that are complimentary to the DNA strand being copied. RNA contains uracil instead of thymine.

The direction of synthesis is 5' to 3'.

A gene can be transcribed many times by multiple RNA polymerase molecules all transcribing at the same time. One RNA polymerase molecule follows another as transcription proceeds.

In bacteria and in eukaryotes, transcription ends after a specific code is transcribed. In bacteria, a ***termination sequence*** in the DNA indicates where transcription will stop. In eukaryotes, transcription stops shortly after a sequence of bases called the polyadenylation signal.

The strand of RNA that is initially produced by transcription is called a ***primary transcript***.

Processing the mRNA Transcript

Some primary transcripts are never translated into protein. These RNA molecules have other functions in the cell.

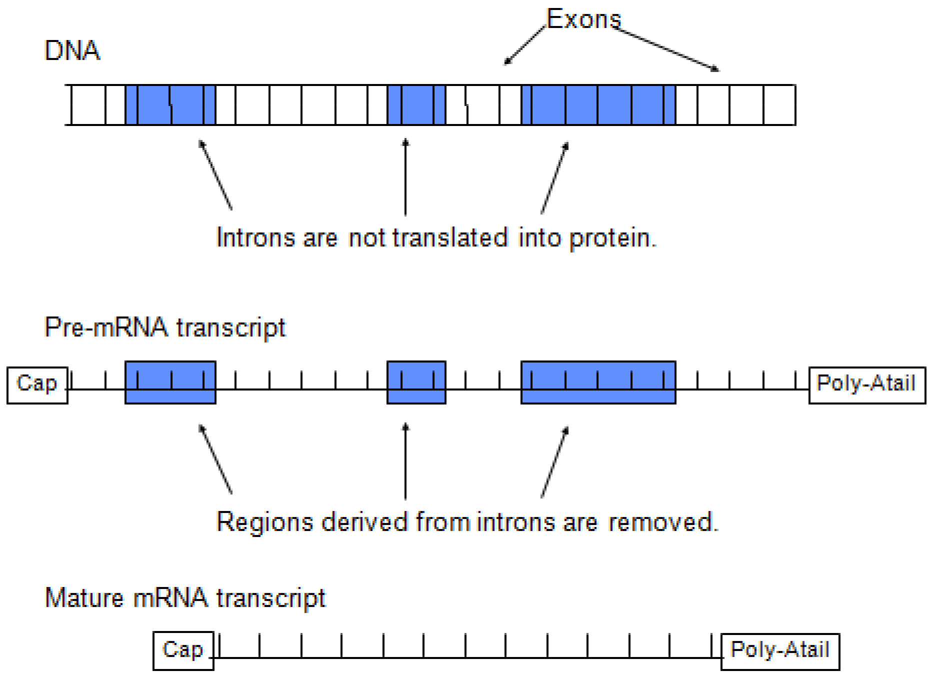
In eukaryotic cells, primary transcripts that are to be translated into protein are modified. These transcripts are called precursor mRNA (or ***pre-mRNA***).

A modified guanine nucleotide "cap" is added to the 5’ end and a poly-A tail (50 to 250 adenines) is added to the 3’end of the molecule. These modifications are thought to 1) enhance the movement of mRNA through the nuclear pores into the cytoplasm, 2) prevent the destruction of mRNA by hydrolytic enzymes, and 3) help ribosomes attach during translation.

The 5' end and the 3' end each contain nucleotides that are not translated into protein. These two regions are called the 5' UTR (untranslated region) and the 3' UTR.

Eukaryotic genes contain regions that are not translated into proteins. These regions of DNA are called ***introns*** (intervening sequences) and must be removed from mRNA by a process called RNA splicing. Their function is not well understood.

The remaining portions of DNA that are translated into protein are called ***exons*** (expressed). After intron-derived regions are removed from mRNA, the remaining fragments- derived from exons- are spliced together to form a ***mature mRNA transcript***.



The process of RNA splicing is carried out by complexes of proteins and small RNA molecules called spliceosomes. The RNA component of spliceosomes is called small nuclear RNA or snRNA. The snRNA is joined together with protein to form small nuclear ribonuclearprotein (snRNP). Small ribonuclearproteins and other proteins together form spliceosomes.

Some introns have catalytic (enzyme) capabilities and they are able to catalyze their own removal from the primary transcript.

Transcription and mRNA processing occur in the nucleus.

Alternative RNA Splicing

A single gene is capable of producing more than one different polypeptide by removing different introns from the primary RNA transcript.

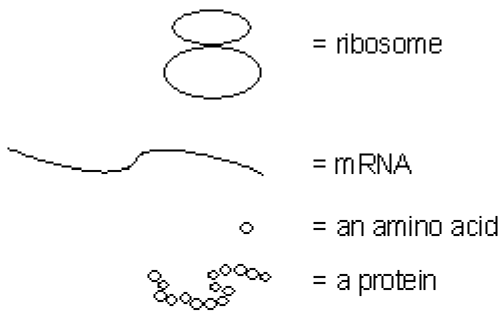
For example, humans have an estimated 20,000 genes. These genes produce as many as 100,000 different proteins due to alternative RNA splicing.

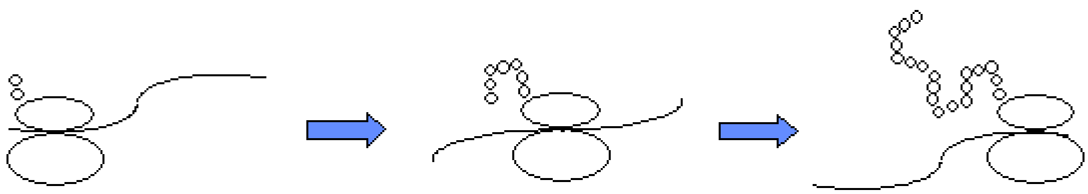
Translation

Translation is the process where ribosomes synthesize proteins using the mature mRNA transcript produced during transcription.

Overview

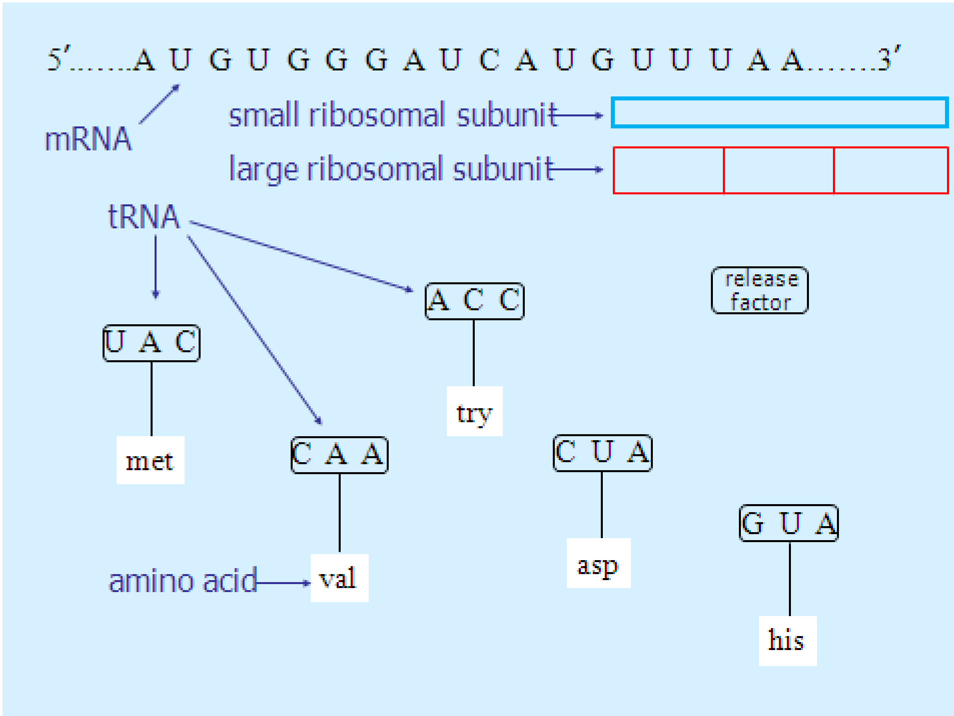
The diagram below shows a ribosome attach to mRNA, and then move along the mRNA adding amino acids to the growing polypeptide chain.





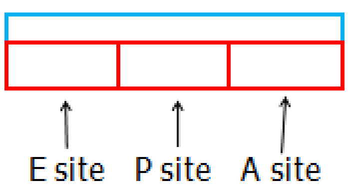
Translation - Details

A mature mRNA transcript, ribosomal subunits, several tRNA molecules and attached amino acids are shown.

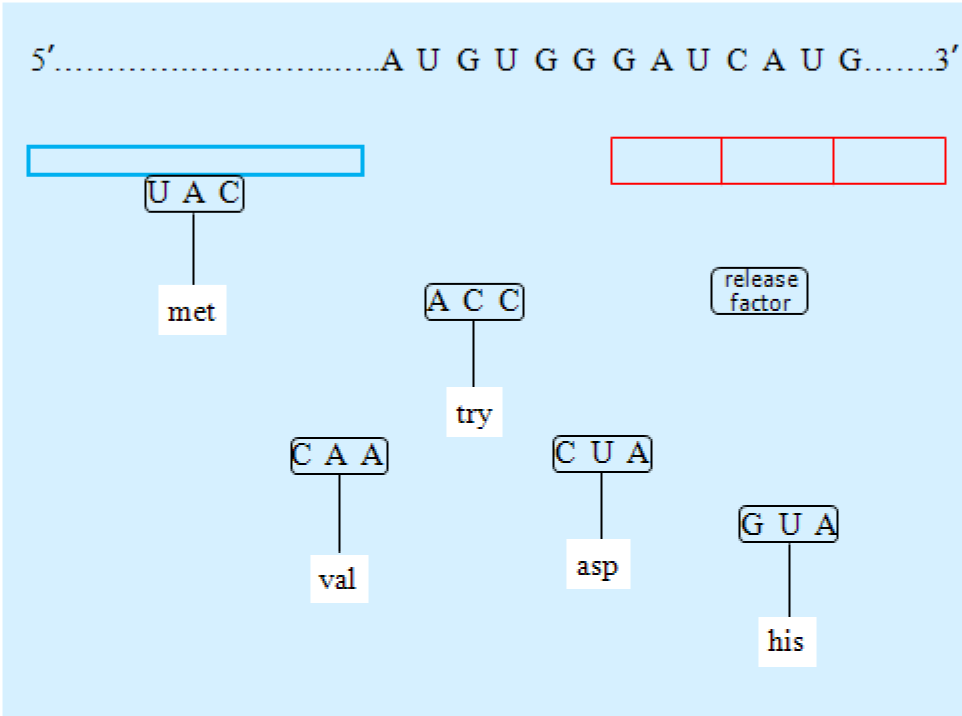


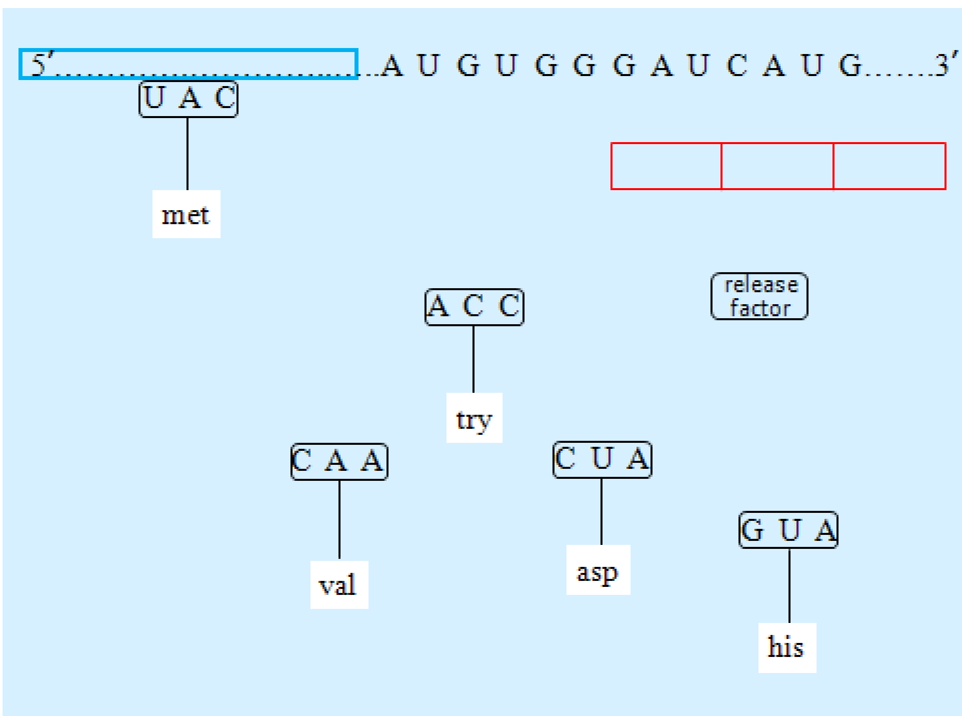
Each three-letter code in the mRNA is a codon. The tRNA molecules have anticodons that are complimentary to the codons in RNA.

Before translation begins, a ribosome will be assembled from two ribosomal subunits. The ribosome contains three attachment sites for tRNA molecules.

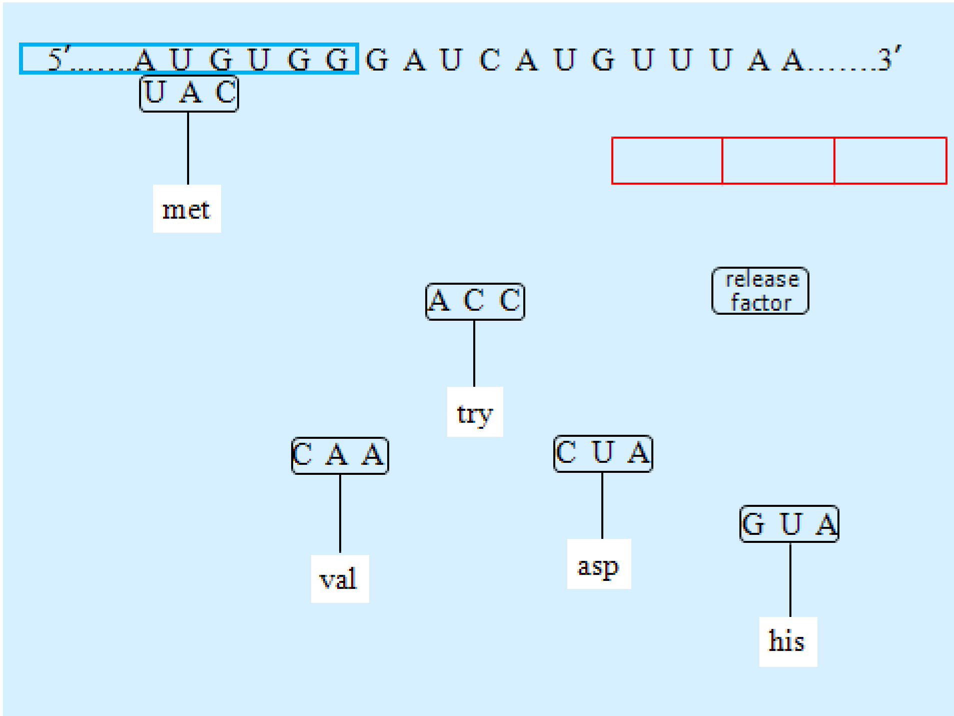


Below: A ribosome attaches to the 5' end of the mRNA transcript. In eukaryotes, the small ribosomal subunit first binds to a tRNA carrying methionine and then to the 5' end of the mRNA.



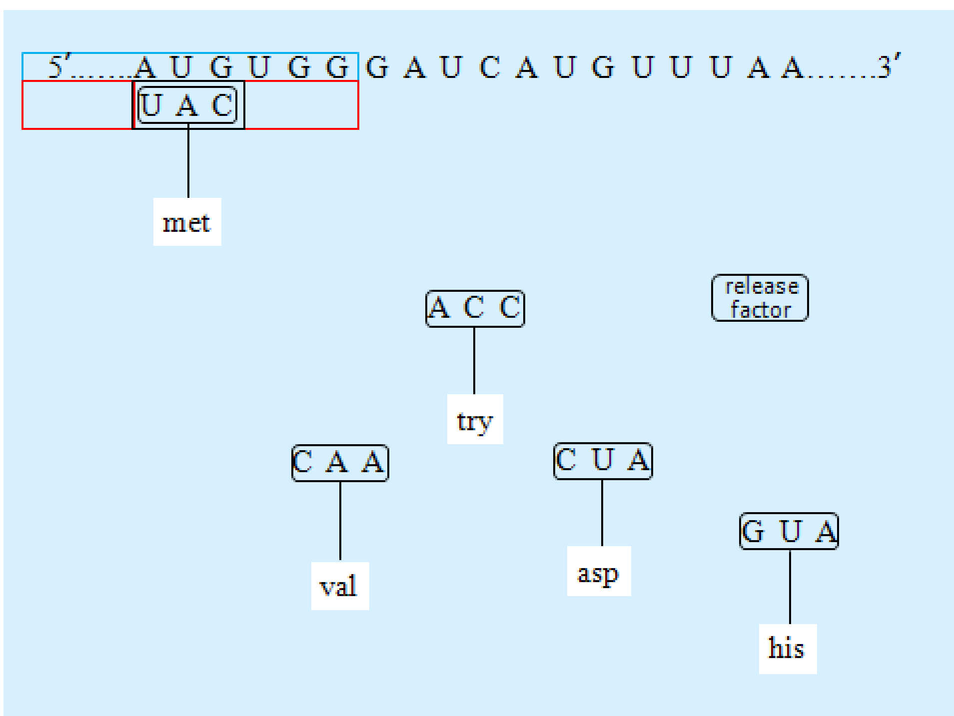


The ribosome moves along the mRNA until it reaches the start codon (AUG).

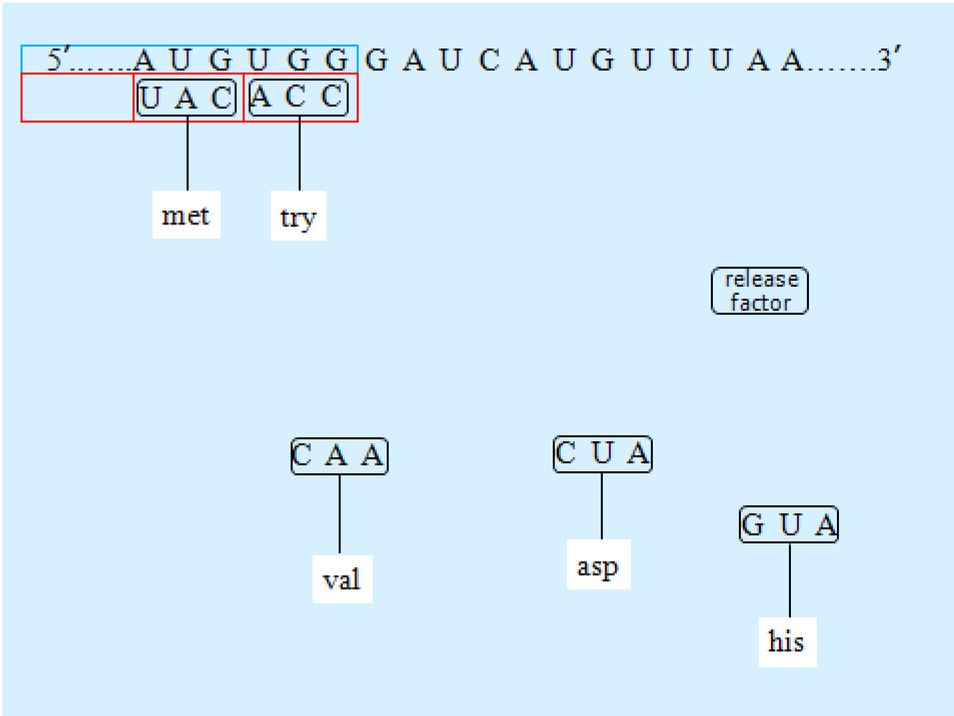


.

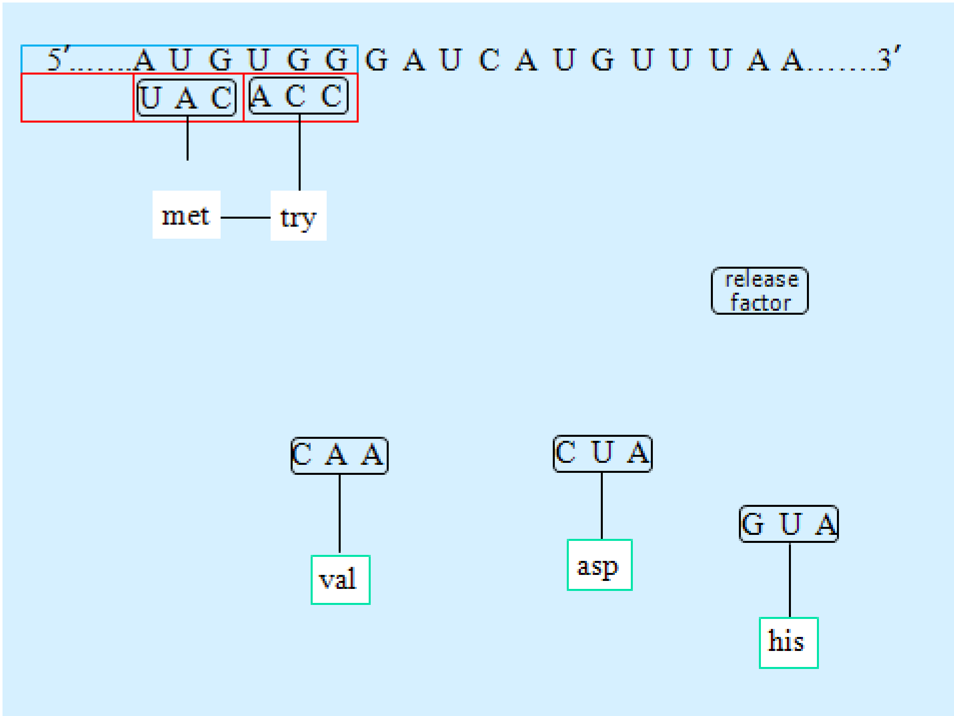
At this point, the tRNA becomes attached to the mRNA and the large ribosomal subunit attaches.



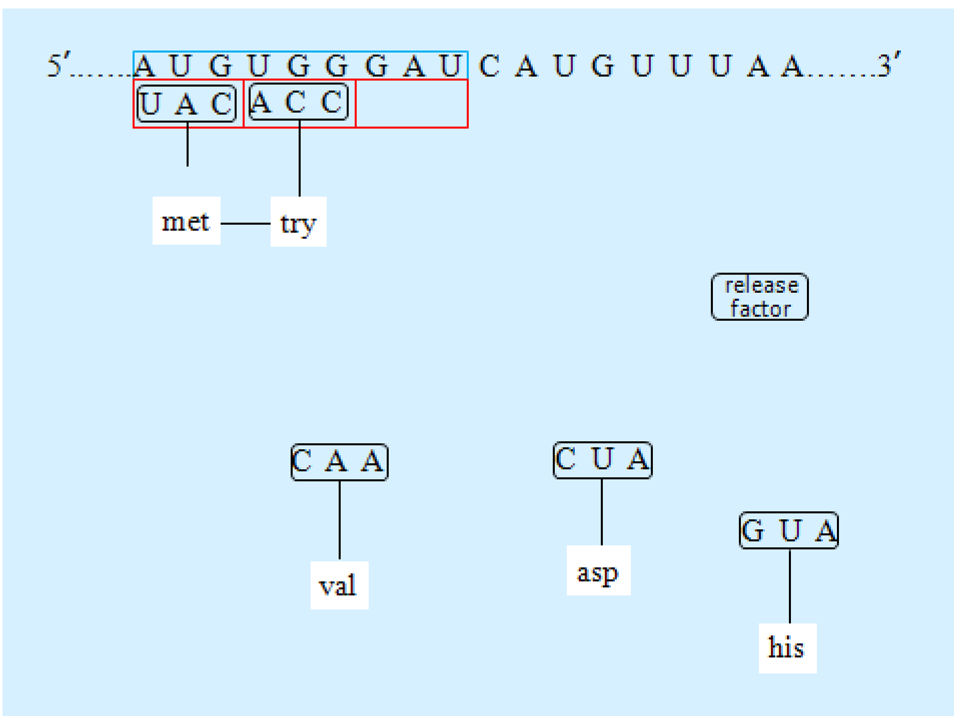
A tRNA molecule transports the next amino acid to the ribosome. Notice that the 3-letter ***anticodon*** on the tRNA molecule matches the 3-letter code (called a ***codon***) in the mRNA. The tRNA with the anticodon "ACC" bonds with tryptophan. It always transports tryptophan. Transfer RNA molecules with different anticodons transport other amino acids.



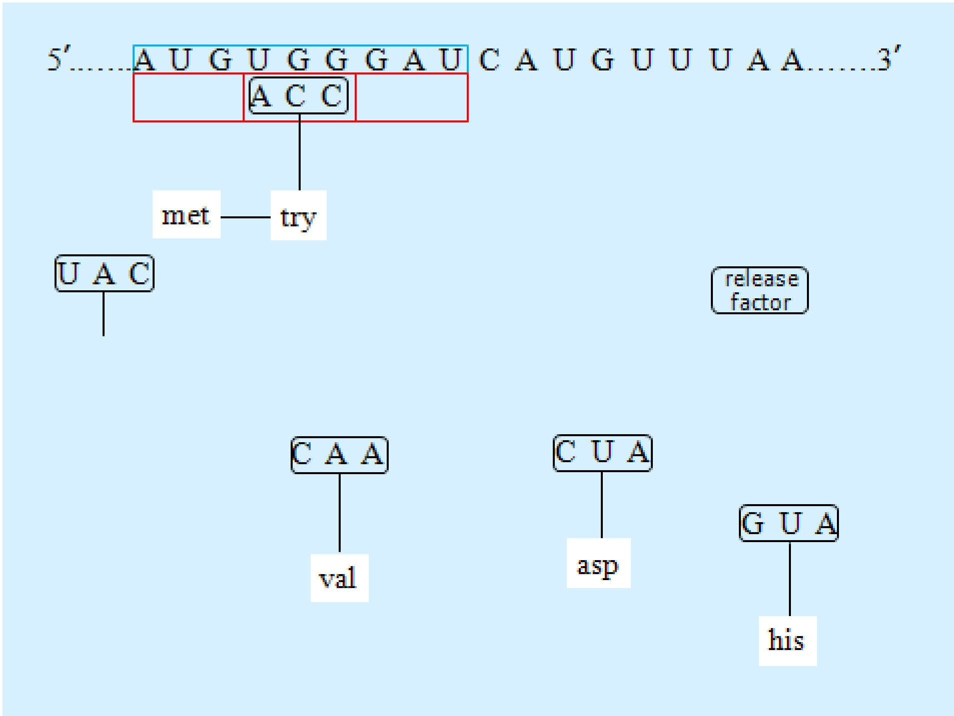
A peptide bond forms between the amino acid in the P site and the amino acid in the A site. The growing polypeptide chain is now attached to the tRNA in the A site.



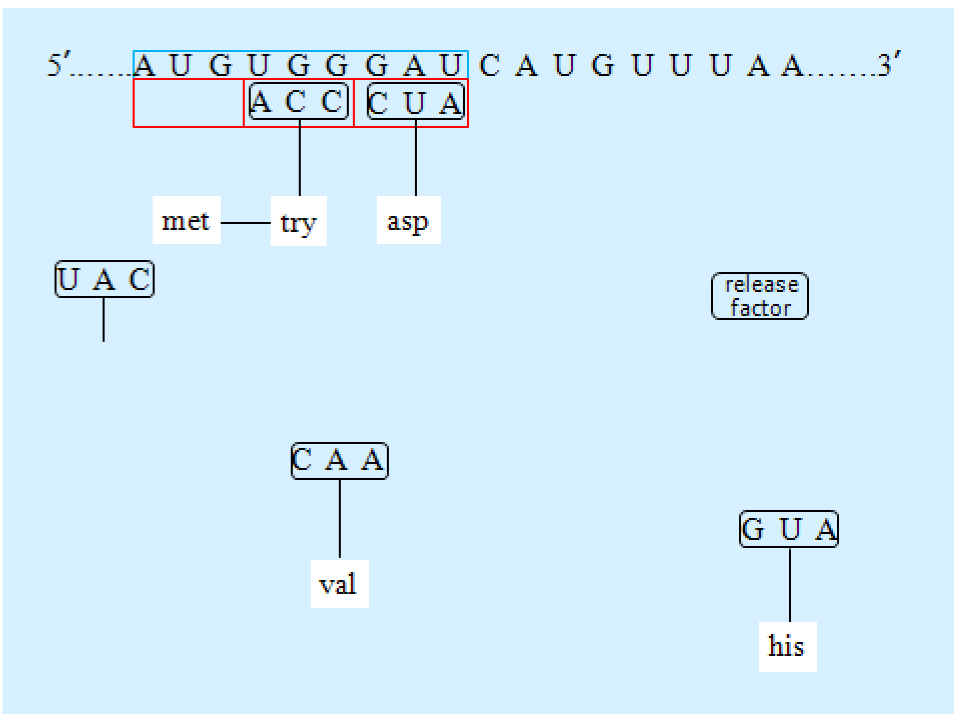
The ribosome moves along the mRNA to expose another codon (GAU) for another tRNA molecule.



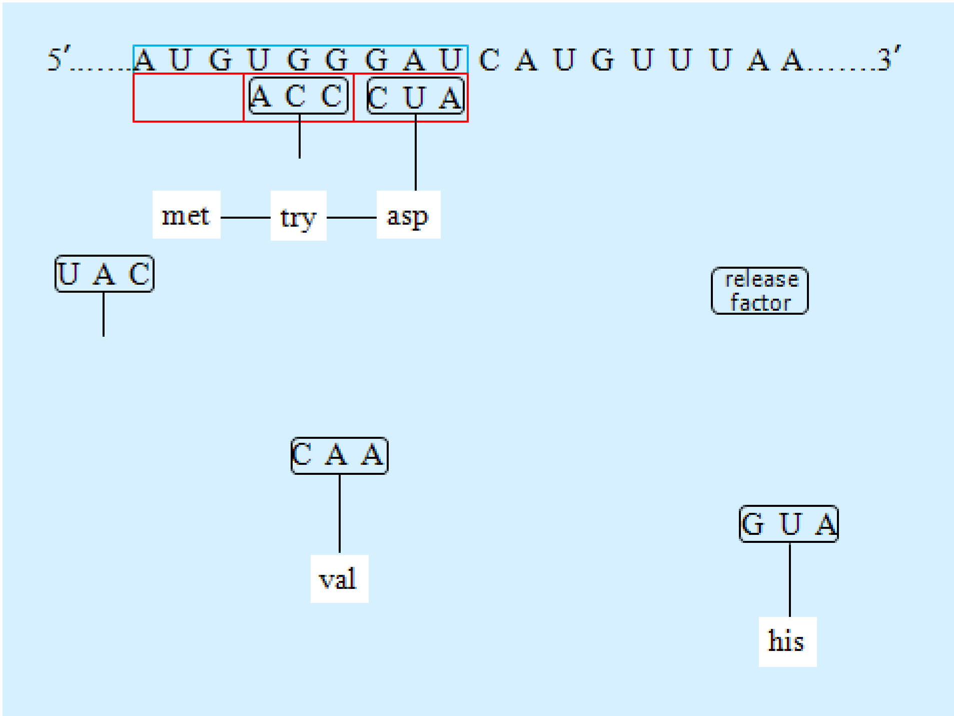
When the tRNA in the P site moves into the E site, it is released. This tRNA can now become attached to another amino acid.

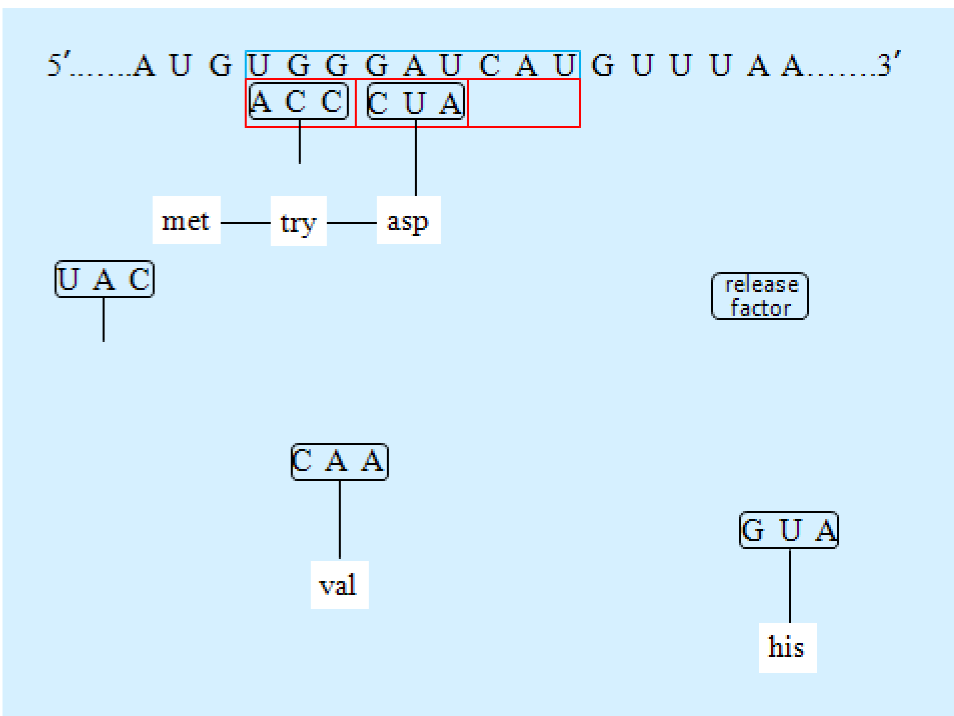


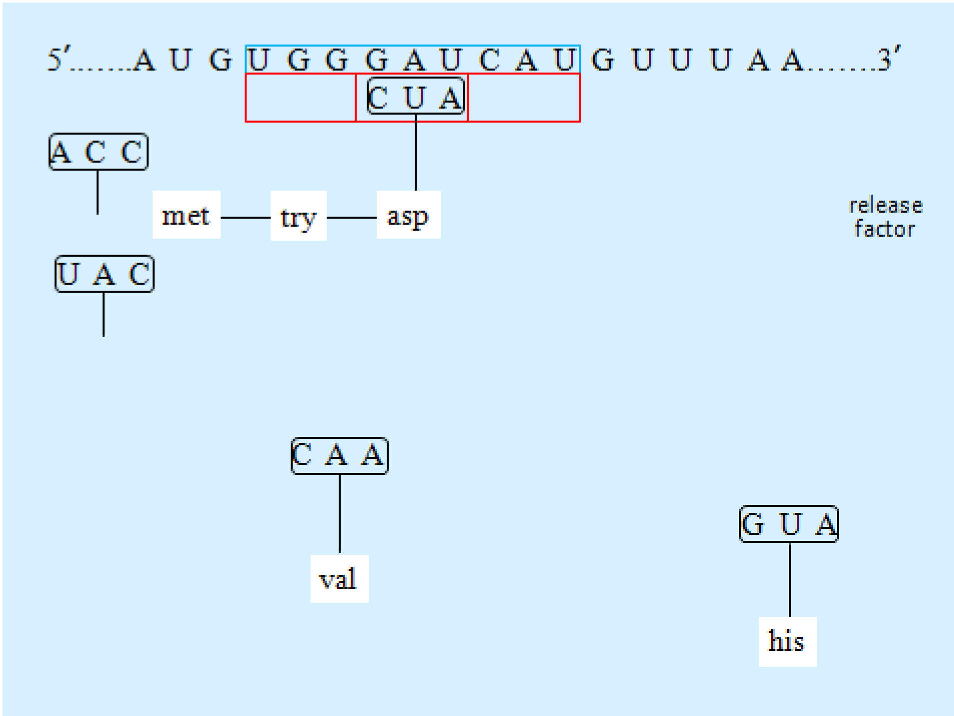
The mRNA codon in the A site is able to bind with the corresponding tRNA (CUA). The tRNA with the CUA anticodon always transports asparagine.

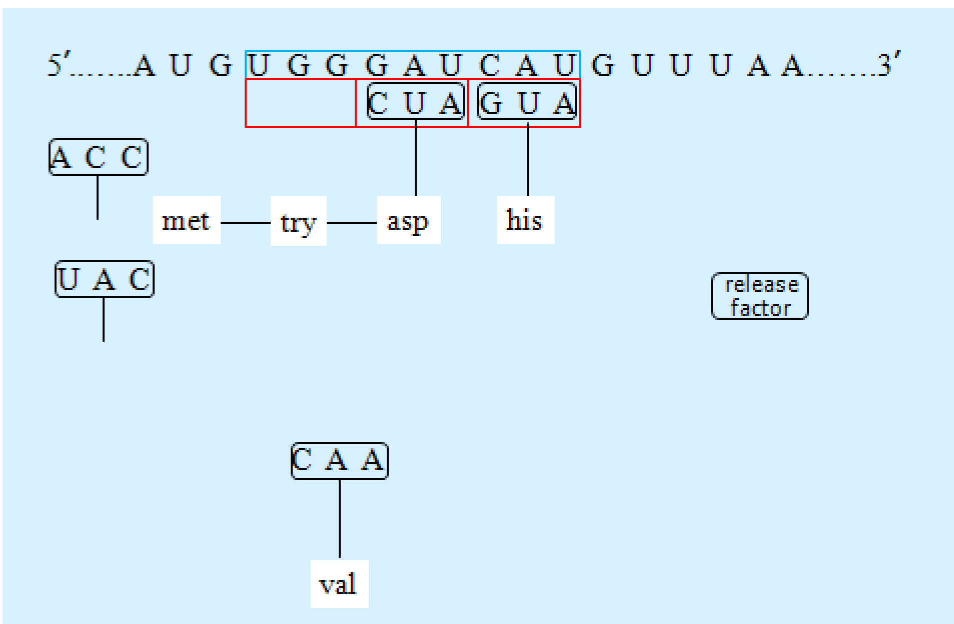


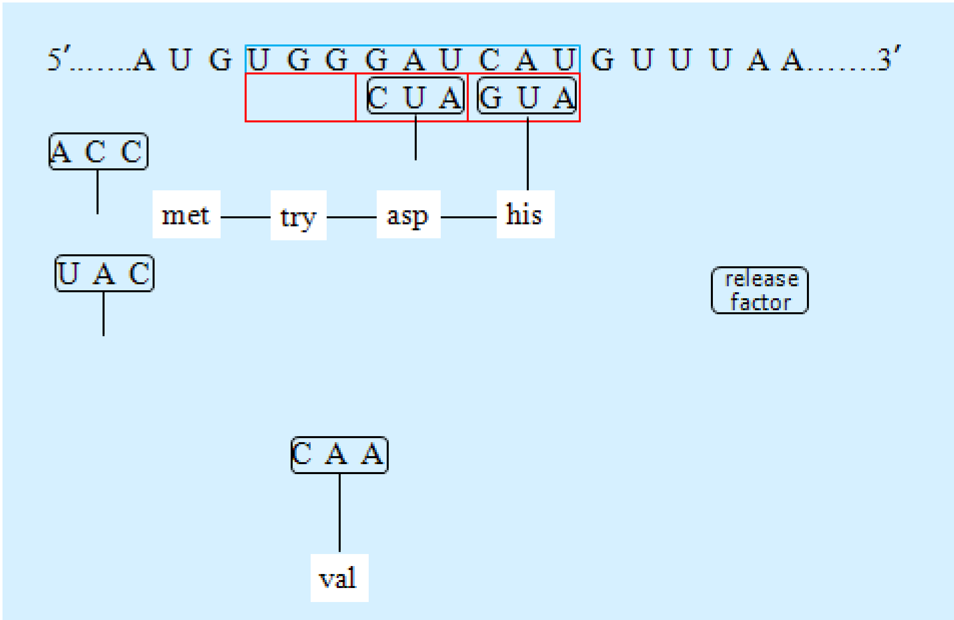
Asparagine is now added to the growing amino acid chain.

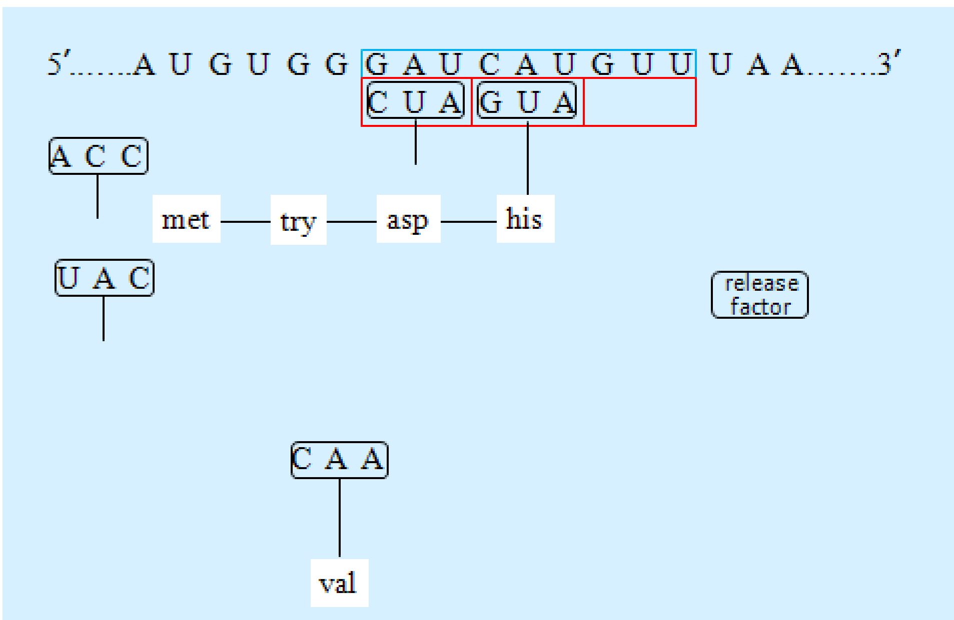

5′……A U G U G G G A U C A U G U U U A A……3′
A C C
C U A
met
try
asp
U A C
release factor
C A A
val
G U A
his

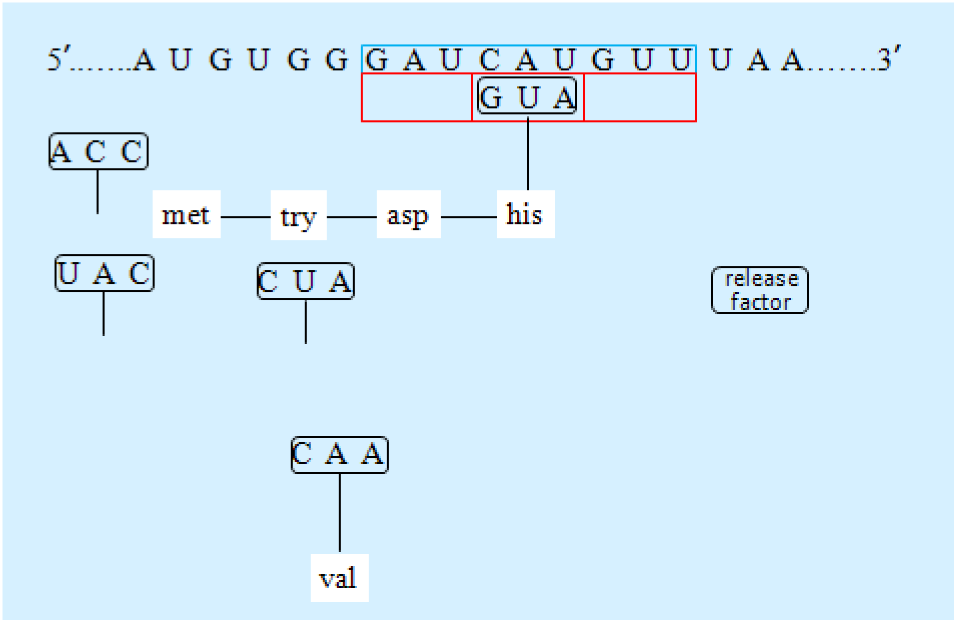

5′……A U G U G G G A U C A U G U U U A A……3′
A C C
C U A
met
try
asp
U A C
release factor
C A A
val
G U A
his

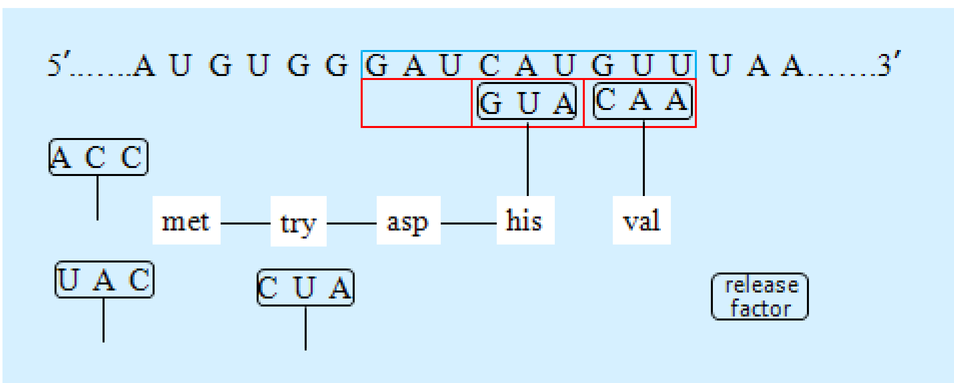

5′......A U G U G G G A U C A U G U U U A A.......3′
C U A
A C C
met
try
asp
release factor
U A C
C A A
val
G U A
his

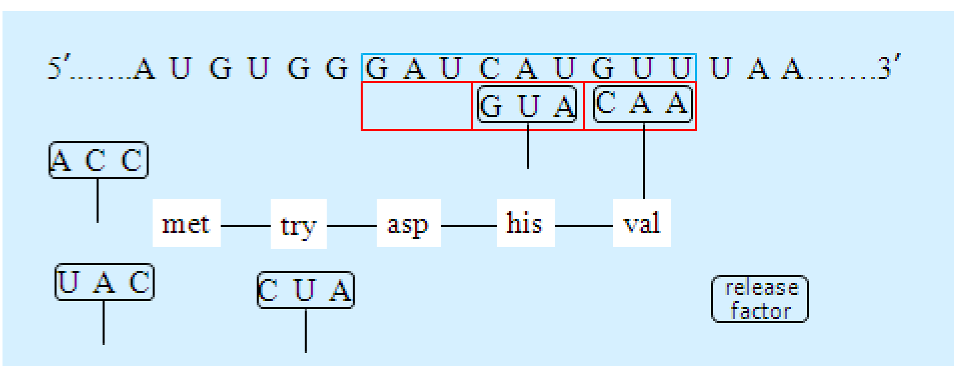

5′......A U G U G G G A U C A U G U U U A A.......3′
C U A
G U A
A C C
met
try
asp
his
U A C
release factor
C A A
val

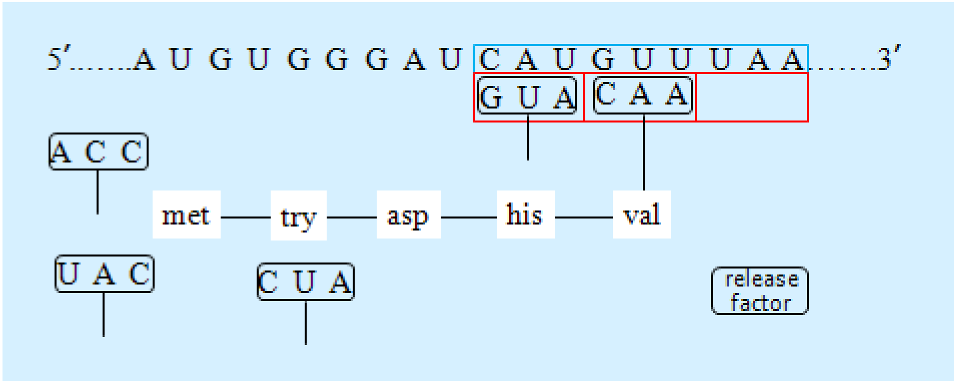

5′……A U G U G G G A U C A U G U U U A A…….3′
C U A
G U A
A C C
met
try
asp
his
U A C
release factor
C A A
val

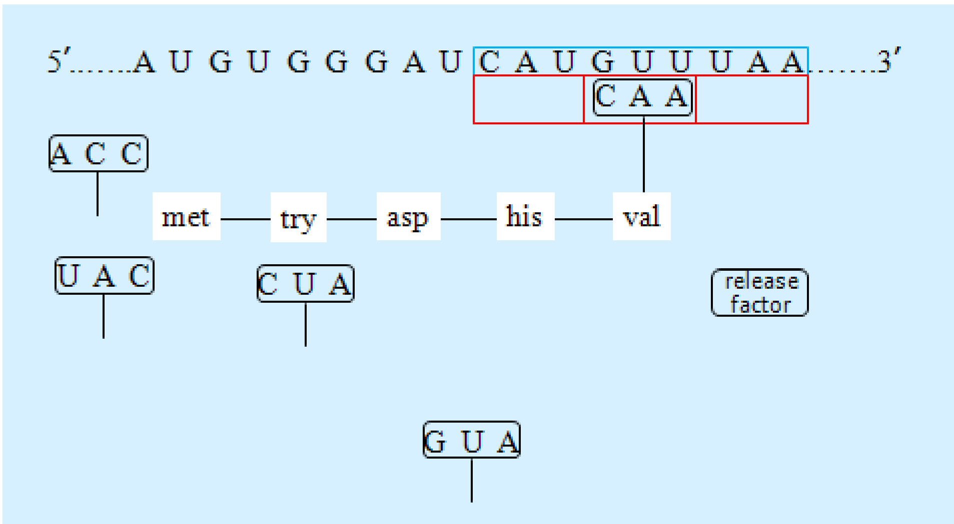

5′……A U G U G G G A U C A U G U U U A A…….3′
C U A
G U A
A C C
met
try
asp
his
U A C
release factor
C A A
val


5′……A U G U G G G A U C A U G U U U A A……3′
G U A
A C C
met
try
asp
his
U A C
C U A
release factor
C A A
val

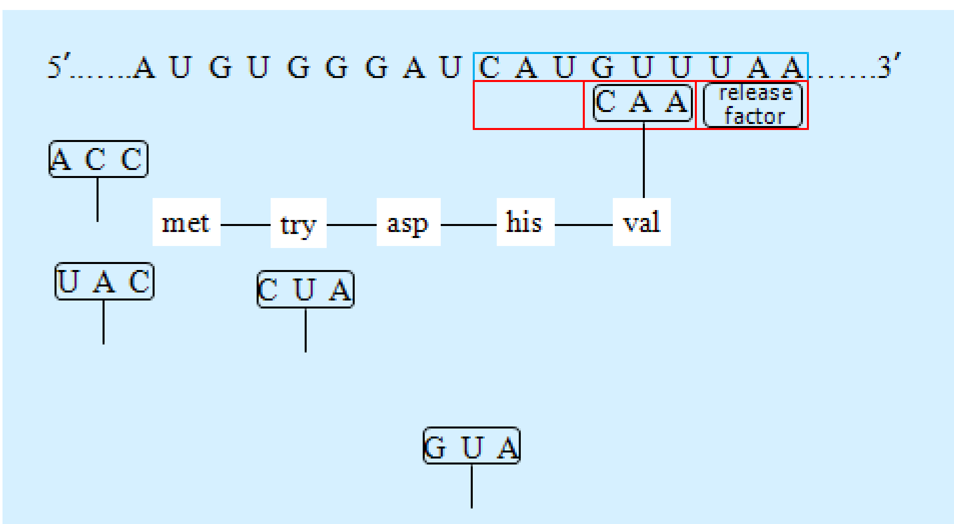

5′……A U G U G G G A U C A U G U U U A A……3′
G U A
C A A
A C C
met
try
asp
his
val
U A C
C U A
release factor

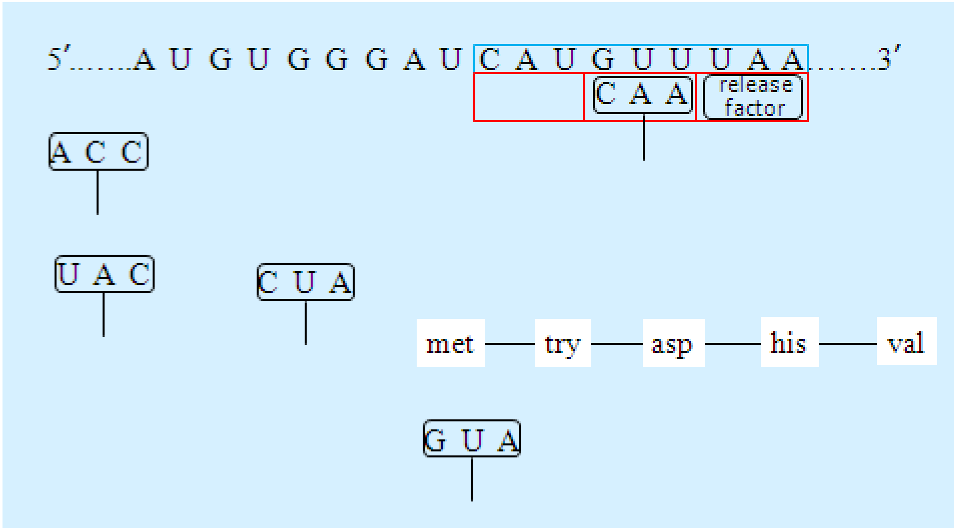

5′……A U G U G G G A U C A U G U U U A A……3′
G U A
C A A
A C C
met
try
asp
his
val
U A C
C U A
release factor

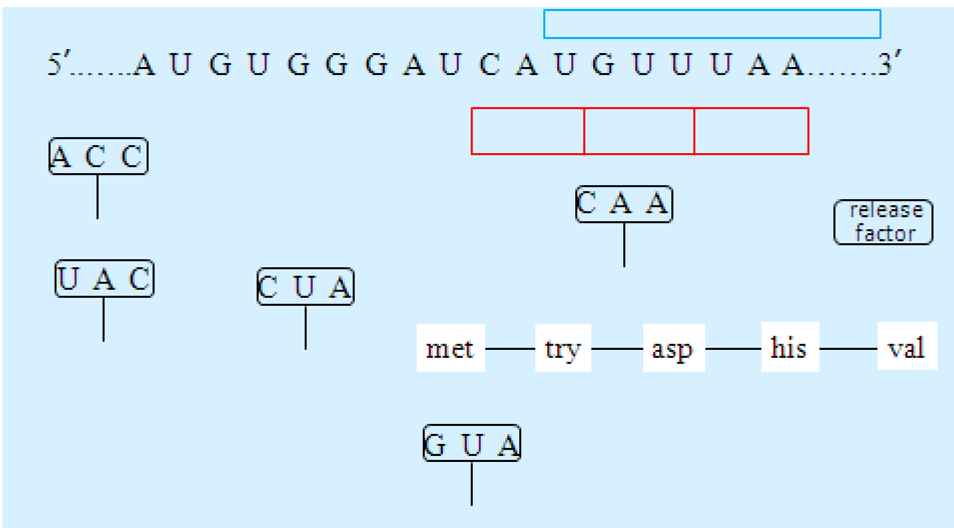




A ***release factor*** binds to a stop codon causing the polypeptide chain to be released and causing the ribosomal subunits and mRNA to dissociate.







Other Proteins

Aminoacyl-tRNA synthetase catylizes the covalent bonding between tRNA and amino acids. There are 20 different aminoacyl-tRNA synthetases, one for each amino acid. The process requires ATP as an energy source.

Initiation factors participate in binding the two ribosomal units to mRNA.

Folding and Modification

The sequence of amino acids in a protein (the primary structure) determines how it will fold. Folding may also require the assistance of other proteins called chaperones.

The completed protein may be chemically modified before it becomes functional. For example, some proteins require the attachment of a carbohydrate chain.

In some cases, amino acids may need to be removed or the polypeptide may need to be cut into shorter segments in order to produce a functional protein.

Genetic Code

The table below can be used to determine what amino acid corresponds to any 3-letter codon.

First Base	**Second Base**				**Third Base**
	U	**C**	**A**	**G**	
U	UUU phenylalanine	UCU serine	UAU tyrosine	UGU cysteine	**U**
U	UUC phenylalanine	UCC serine	UAC tyrosine	UGC cysteine	**C**
U	UUA leucine	UCA serine	UAA stop	UGA stop	**A**
U	UUG leucine	UCG serine	UAG stop	UGG tryptophan	**G**
C	CUU leucine	CCU proline	CAU histidine	CGU arginine	**U**
C	CUC leucine	CCC proline	CAC histidine	CGC arginine	**C**
C	CUA leucine	CCA proline	CAA glutamine	CGA arginine	**A**
C	CUG leucine	CCG proline	CAG glutamine	CGG arginine	**G**
A	AUU isoleucine	ACU threonine	AAU asparagine	AGU serine	**U**
A	AUC isoleucine	ACC threonine	AAC asparagine	AGC serine	**C**
A	AUA isoleucine	ACA threonine	AAA lysine	AGA arginine	**A**
A	AUG (start) methionine	ACG threonine	AAG lysine	AGG arginine	**G**
G	GUU valine	GCU alanine	GAU aspartate	GGU glycine	**U**
G	GUC valine	GCC alanine	GAC aspartate	GGC glycine	**C**

	GUA valine	GCA alanine	GAA glutamate	GGA glycine	**A**
	GUG valine	GCG alanine	GAG glutamate	GGG glycine	**G**

Wobble

The pairing of bases between the tRNA and mRNA does not always follow the standard base pairing rules (A-U and G-C) for the third base pair. For example, in some cases, if the third letter is G, it could pair with U or with C. This phenomenon, called wobble, enables a specific tRNA to pair with more than one codon.

Mutation

Mutations are changes in the DNA.

Frameshift

A frameshift mutation is usually severe, producing a completely nonfunctional protein.

The priniciple of a frameshift can be explained using the sentence below. If the letters are read three at a time and one is deleted, the second sentence becomes meaningless.

Original DNA:THE BIG RED ANT ATE ONE FAT BUG
Frameshift mutation: THB IGR EDA NTA TEO NEF ATB UG?

Point Mutation

Point mutations involve a single nucleotide, thus a single amino acid.

In the sentence below, eliminating one letter does not change in the remaining three-letter words and therefore may not cause a significant change in the meaning of the sentence.

Original DNA:THE BIG RED ANT ATE ONE FAT BUG
Point mutation:THA BIG RED ANT ATE ONE FAT BUG

Silent, Missense, and Nonsense Mutations

Silent mutations are those that do not change the amino acid sequence. This happens because amino acids have more than one spelling. Silent mutations code for functional proteins.

A mutation that results in an amino acid substitution is called a missense mutation.

A mutation that results in a stop codon so that incomplete proteins are produced is a nonsense mutation