Introduction to RNA

RNA molecules are single-stranded nucleic acids composed of nucleotides carrying the genetioc code, which play a major role in protein synthesis, being translated to corresponding proteins. RNA stands for ribonucleic acid.

RNA nitrogenous bases include **adenine (A)**, **guanine (G)**, **cytosine (C)** and **uracil (U)**. The five-carbon (pentose) sugar in RNA is ribose. RNA molecules are polymers of nucleotides joined to one another by covalent bonds between the phosphate of one nucleotide and the sugar of another. These linkages are called phosphodiester linkages.

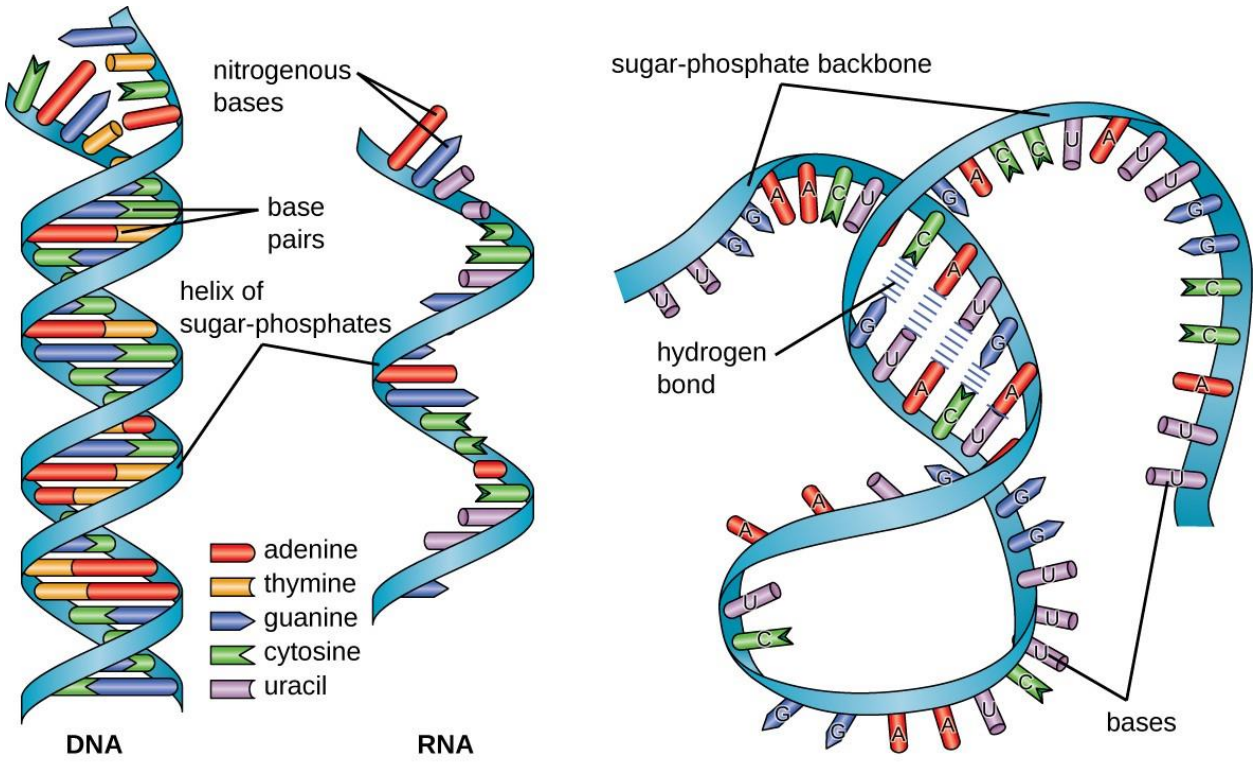
Although single-stranded, RNA is not always linear. It has the ability to fold into complex three-dimensional shapes and form **hairpin loops**. When this occurs, the nitrogenous bases bind to one another. Adenine pairs with uracil (A-U) and guanine pairs with cytosine (G-C). Hairpin loops are commonly observed in RNA molecules such as messenger RNA (mRNA) and transfer RNA (tRNA).

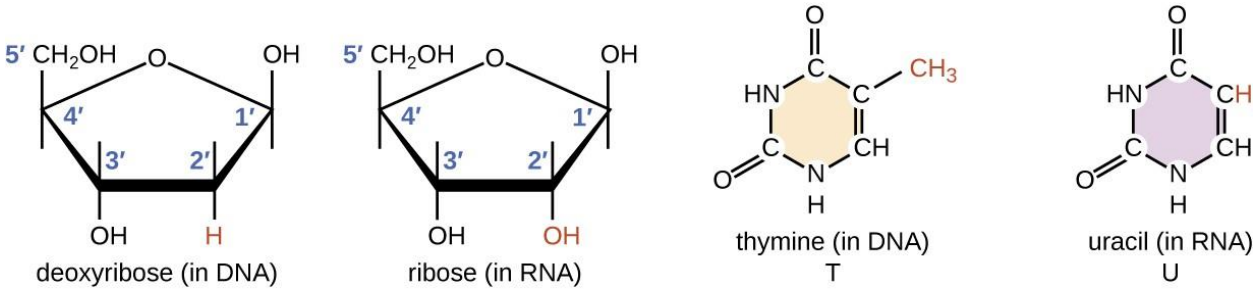
RNA structure

RNA typically is a single-stranded biopolymer. However, the presence of self-complementary sequences in the RNA strand leads to intrachain base-pairing and

folding of the ribonucleotide chain into complex structural forms consisting of bulges and helices. The three-dimensional structure of RNA is critical to its stability and function, allowing the ribose sugar and the nitrogenous bases to be modified in numerous different ways by cellular enzymes that attach chemical groups (e.g., methyl groups) to the chain. Such modifications enable the formation of chemical bonds between distant regions in the RNA strand, leading to complex contortions in the RNA chain, which further stabilizes the RNA structure.

RNA is made of ribonucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and a phosphate group. The subtle structural difference between the sugars gives DNA added stability, making DNA more suitable for storage of genetic information, whereas the relative instability of RNA makes it more suitable for its more short-term functions. The RNA-specific pyrimidine uracil forms a complementary base pair with adenine and is used instead of the thymine used in DNA.


nitrogenous bases
base pairs
helix of sugar-phosphates
adenine
thymine
guanine
cytosine
uracil
DNA
RNA
(a)
sugar-phosphate backbone
hydrogen bond
bases
(b)


5′ CH2OH
OH
4′
1′
3′
2′
OH
H
deoxyribose (in DNA)
5′ CH2OH
OH
4′
1′
3′
2′
OH
OH
ribose (in RNA)
(a)
CH3
thymine (in DNA)
T
uracil (in RNA)
U
(b)

Types of RNA

Messenger RNA (mRNA)

The mRNA carries the message from the DNA, which controls all of the cellular activities in a cell. If a cell requires a certain protein to be synthesized, the gene for this product is “turned on” and the mRNA is synthesized through the process of transcription. The mRNA then interacts with ribosomes and other cellular machinery to direct the synthesis of the protein it encodes during the process of translation. mRNA is relatively unstable and short-lived in the cell, especially in prokaryotic cells, ensuring that proteins are only made when needed.

Ribosomal RNA (rRNA)

rRNA is a stable type of RNA. In prokaryotes and eukaryotes, rRNA is encoded in the DNA, then copied into long RNA molecules that are cut to release smaller fragments containing the individual mature RNA species. In eukaryotes, synthesis, cutting, and assembly of rRNA into ribosomes takes place in the nucleolus, but these activities occur in the cytoplasm. It does not carry instructions to direct the synthesis of a polypeptide, but it play other important roles in protein synthesis.

Ribosomes are composed of rRNA and protein. As its name suggests, rRNA is a major constituent of ribosomes, composing up to about 60% of the ribosome by mass

and providing the location where the mRNA binds. The rRNA ensures the proper alignment of the mRNA, tRNA, and the ribosomes; the rRNA of the ribosome also has an enzymatic activity (peptidyl transferase) and catalyzes the formation of the peptide bonds between two aligned amino acids during protein synthesis. Although rRNA had long been thought to serve primarily a structural role, its catalytic role within the ribosome was proven in 2000.

Transfere RNA (tRNA)

Transfer RNA (tRNA) plays an important role in the translation portion of protein synthesis. Its job is to translate the message within the nucleotide sequences of mRNA into specific amino acid sequences. The amino acid sequences are joined together to form a protein. Transfer RNA is shaped like a clover leaf with three hairpin loops. It contains an amino acid attachment site on one end and a special section in the middle loop called the anticodon site. The anticodon recognizes a specific area on mRNA called a codon. A codon consists of three continuous nucleotide bases that code for an amino acid or signal the end of translation. Transfer RNA along with ribosomes read the mRNA codons and produce a polypeptide chain. The polypeptide chain undergoes several modifications before becoming a fully functioning protein.

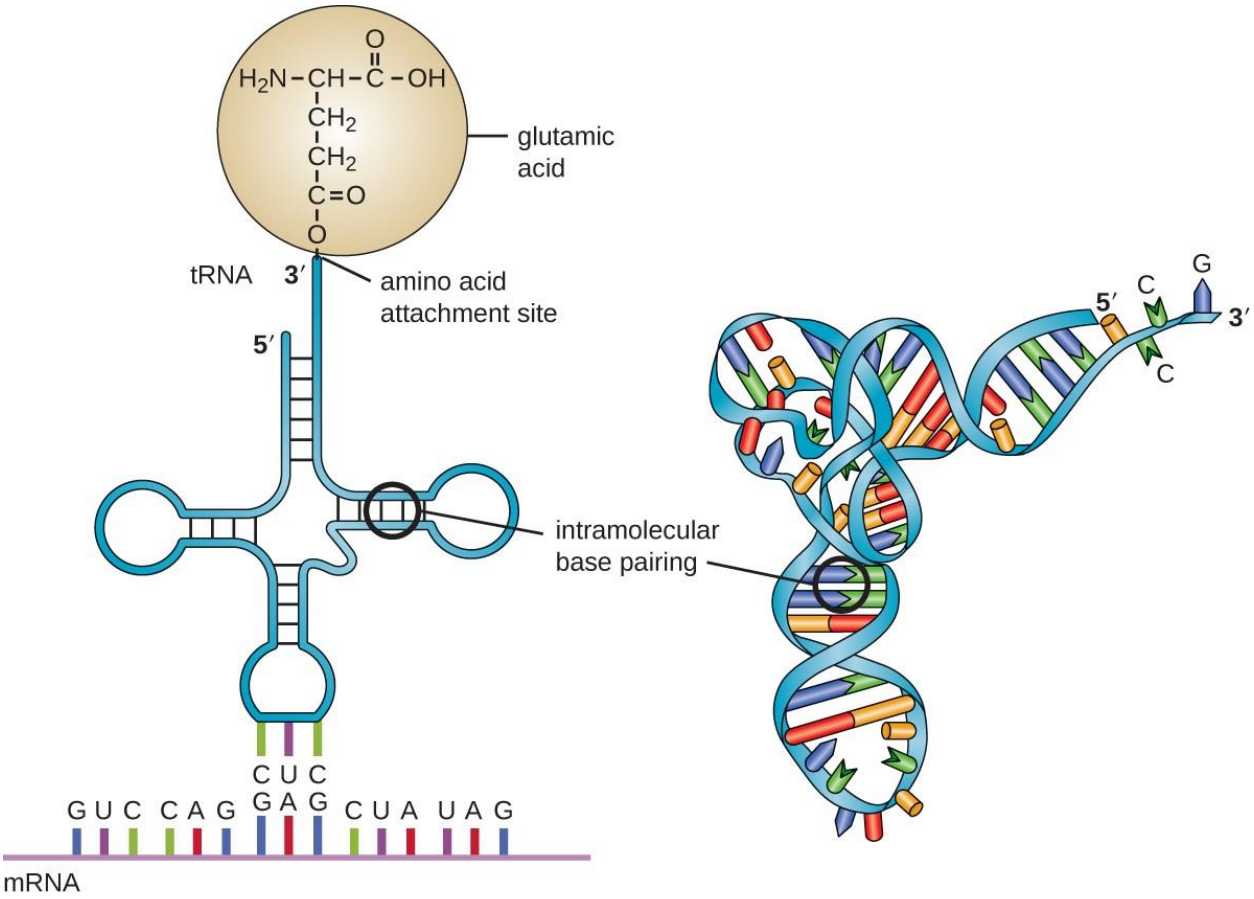


Table 1. Structure and Function of RNA

	mRNA	rRNA	tRNA
Structure	Short, unstable, single-stranded **RNA** corresponding to a gene encoded within DNA	Longer, stable RNA molecules composing 60% of ribosome's mass	Short (70-90 nucleotides), stable RNA with extensive intramolecular base pairing; contains an amino acid binding site and an mRNA binding site
Function	Serves as intermediary between **DNA** and protein; used by ribosome to direct synthesis of protein it encodes	Ensures the proper alignment of mRNA, tRNA, and ribosome during protein synthesis; catalyzes **peptide bond** formation between amino acids	Carries the correct amino acid to the site of protein synthesis in the ribosome

Differences between DNA and RNA

Comparison	DNA	RNA
Full Name	Deoxyribonucleic Acid	Ribonucleic Acid
Function	DNA replicates and stores genetic information. It is a blueprint for all genetic information contained within an organism	RNA converts the genetic information contained within DNA to a format used to build proteins, and then moves it to ribosomal protein factories.
Structure	DNA consists of two strands, arranged in a double helix. These strands are made up of subunits called nucleotides. Each nucleotide contains a phosphate, a 5-carbon sugar molecule and a nitrogenous base.	RNA only has one strand, but like DNA, is made up of nucleotides. RNA strands are shorter than DNA strands. RNA sometimes forms a secondary double helix structure, but only intermittently.
Length	DNA is a much longer polymer than RNA. A chromosome, for example, is a single, long DNA molecule, which would be several centimetres in length when unravelled.	RNA molecules are variable in length, but much shorter than long DNA polymers. A large RNA molecule might only be a few thousand base pairs long.
Sugar	The sugar in DNA is deoxyribose, which contains one less hydroxyl group than RNA's ribose.	RNA contains ribose sugar molecules, without the hydroxyl modifications of deoxyribose.
Bases	The bases in DNA are Adenine ('A'), Thymine ('T'), Guanine ('G') and Cytosine ('C').	RNA shares Adenine ('A'), Guanine ('G') and Cytosine ('C') with DNA, but contains Uracil ('U') rather than Thymine.
Base Pairs	Adenine and Thymine pair (A-T) Cytosine and Guanine pair (C-G)	Adenine and Uracil pair (A-U) Cytosine and Guanine pair (C-G)
Location	DNA is found in the nucleus, with a small amount of DNA also present in mitochondria.	RNA forms in the nucleolus, and then moves to specialised regions of the cytoplasm depending on the type of RNA formed.
Reactivity	Due to its deoxyribose sugar, which contains one less oxygen-containing hydroxyl group, DNA is a more stable molecule than RNA, which is useful for a molecule which has the task of keeping genetic information safe.	RNA, containing a ribose sugar, is more reactive than DNA and is not stable in alkaline conditions. RNA's larger helical grooves mean it is more easily subject to attack by enzymes.

Ultraviolet (UV) Sensitivity	DNA is vulnerable to damage by ultraviolet light.	RNA is more resistant to damage from UV light than DNA.

Ckinical significance of RNA

Various neurodegenerative diseases, including Alzheimer disease. In the case of other RNA types, tRNAs can bind to specialized proteins known as caspases, which are involved in apoptosis (programmed cell death). By binding to caspase proteins, tRNAs inhibit apoptosis; the ability of cells to escape programmed death signaling is a hallmark of cancer. Noncoding RNAs tRNA-derived fragments (tRFs) are also suspected to play a role in cancer. The emergence of techniques such as RNA sequencing has led to the identification of novel classes of tumour-specific RNA transcripts, such as MALAT1 (metastasis associated lung adenocarcinoma transcript 1), increased levels of which have been found in various cancerous tissues and are associated with the proliferation and metastasis (spread) of tumour cells.

RNA Transcription

General definitions

Promoter: A promoter is a region of DNA that initiates transcription of a particular gene.

Intron: An intron is a long sequence of noncoding DNA found between exons (or coding regions) in a gene.

Exons: An exon is any part of a gene that will encode a part of the final mature RNA produced by that gene after introns have been removed by RNA splicing. The term exon refers to both the DNA sequence within a gene and to the corresponding sequence in RNA transcripts.

RNA splicing: RNA splicing is a process that removes introns and joins exons in a primary transcript.

Transcription is the first step in gene expression. It involves copying a gene's DNA sequence to make an RNA molecule. Transcription is performed by enzymes called RNA polymerases, which link nucleotides to form an RNA strand (using a DNA strand as a template

In transcription, a region of DNA opens up. One strand, the template strand, serves as a template for synthesis of a complementary RNA

transcript. The other strand, the coding strand, is identical to the RNA transcript in sequence, except that it has uracil (U) bases in place of thymine (T) bases.

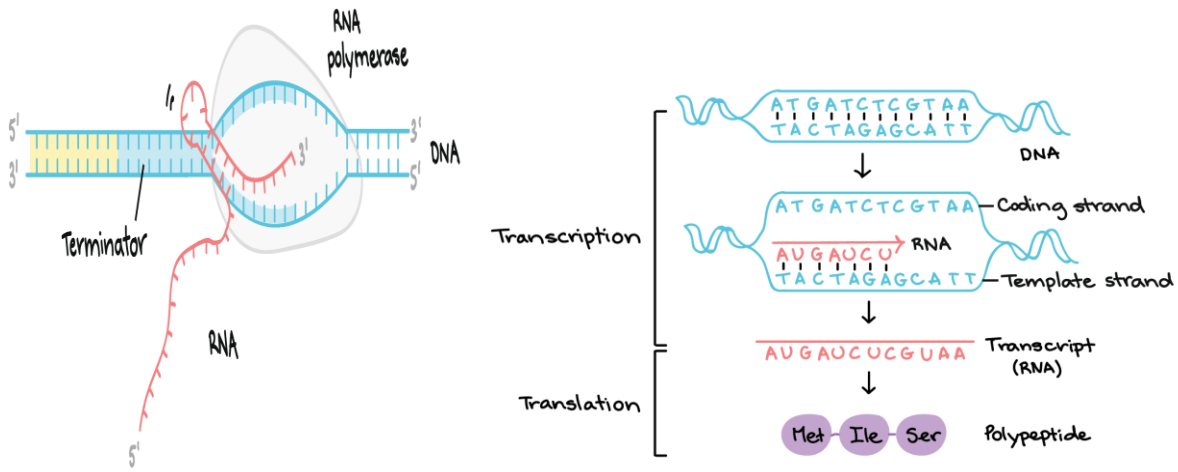
Coding strand: 5'-ATGATCTCGTAA-3'

Template strand: 3'-TACTAGAGCATT-5'

RNA transcript: 5'-AUGAUCUCGUAA-3'

RNA polymerase

The main enzyme involved in transcription is **RNA polymerase**, which uses a single-stranded DNA template to synthesize a complementary strand of RNA in a 5' to 3' direction. It also involves unwinding of the DNA double strand to provide space for the enzyme to bind the DNA at a specific location to undergo its function.



Steps of transcription

1- Pre-initiation

The first step of transcription is called pre-initiation. RNA polymerase and cofactors (general transcription factors) bind to DNA and unwind it, creating an initiation bubble. This space grants RNA polymerase access to a single strand of the DNA molecule. Approximately 14 base pairs are exposed at a time.

2- Initiation

The initiation of transcription in bacteria begins with the binding of RNA polymerase to the promoter in DNA. Transcription initiation is more complex in eukaryotes, where a group of proteins called transcription factors mediates the binding of RNA polymerase and the initiation of transcription. RNA polymerase must clear the promoter once the first bond has been synthesized.

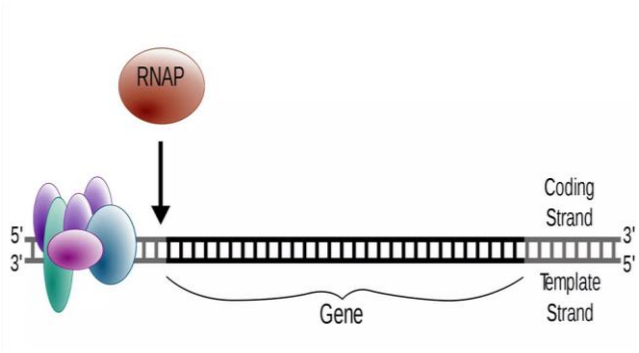
3- Elongation

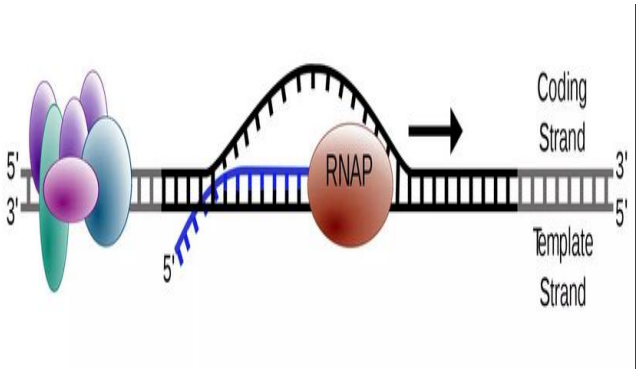
One strand of DNA serves as the template for RNA synthesis, but multiple rounds of transcription may occur so that many copies of a gene can be produced.

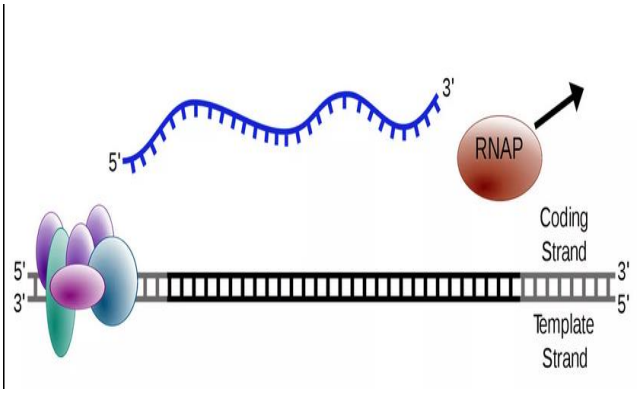
4- Termination

Termination is the final step of transcription. Termination results in the release of the newly synthesized mRNA from the elongation complex. In eukaryotes, the termination of transcription involves cleavage of the transcript, followed by a process called

polyadenylation. In polyadenylation, a series of adenine residues or poly(A) tail is added to the new 3' end of the end of the messenger RNA strand. Also, a structure of cap is added to the 5’ of the newly synthesized RNA. Both cap and polyA tail aid to protect the RNA from damage and play roles in protein synthesis.


RNAP
Coding
Strand
5'
3'
3'
5'
Template
Strand
Gene


Coding
Strand
RNAP
5'
3'
3'
5'
5'
Template
Strand


3'
5'
RNAP
Coding
Strand
5'
3'
3'
5'
Template
Strand

Genetic Code

Each amino acid, on DNA or RNA, is defined by a three-nucleotide sequence called the triplet codon. This means that when protein is synthetized, the ribosome reads each triplet of nucleotides on the mRNA, and consequently translate them to the corresponding amino acis. Scientists theorized that combinations of nucleotides must represent single amino acids. Nucleotide doublets would not be sufficient to specify every amino acid because there are only 16 possible two-nucleotide combinations (4^2). In contrast, there are 64 possible nucleotide triplets (4^3), which is far more than the number of amino acids. Nucleotide triplets encode amino acids and that the genetic code was "degenerate." In other words, a given amino acid could is encoded by more than one nucleotide triplet. This is confirmed experimentally. These nucleotide triplets are called codons. The insertion of one or two nucleotides completely changed the triplet reading frame, thereby altering the message for every subsequent amino acid. Though insertion of three nucleotides caused an extra amino acid to be inserted during translation, the integrity of the rest of the protein was maintained.

The flow of genetic information to cells of new generations is called central dogma. It includes DNA replication, RNA transcription and translation to produce the corresponding protein.

Amino Acid	SLC	DNA codons
Isoleucine	I	ATT, ATC, ATA
Leucine	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	V	GTT, GTC, GTA, GTG
Phenylalanine	F	TTT, TTC
Methionine	M	ATG
Cysteine	C	TGT, TGC
Alanine	A	GCT, GCC, GCA, GCG
Glycine	G	GGT, GGC, GGA, GGG
Proline	P	CCT, CCC, CCA, CCG
Threonine	T	ACT, ACC, ACA, ACG
Serine	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Y	TAT, TAC
Tryptophan	W	TGG
Glutamine	Q	CAA, CAG
Asparagine	N	AAT, AAC
Histidine	H	CAT, CAC
Glutamic acid	E	GAA, GAG
Aspartic acid	D	GAT, GAC
Lysine	K	AAA, AAG
Arginine	R	CGT, CGC, CGA, CGG, AGA, AGG
Stop codons	Stop	TAA, TAG, TGA

In addition to codons that instruct the addition of a specific amino acid to a polypeptide chain, three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called nonsense codons, or *stop codons (UAG, UGA and UAA)*. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves

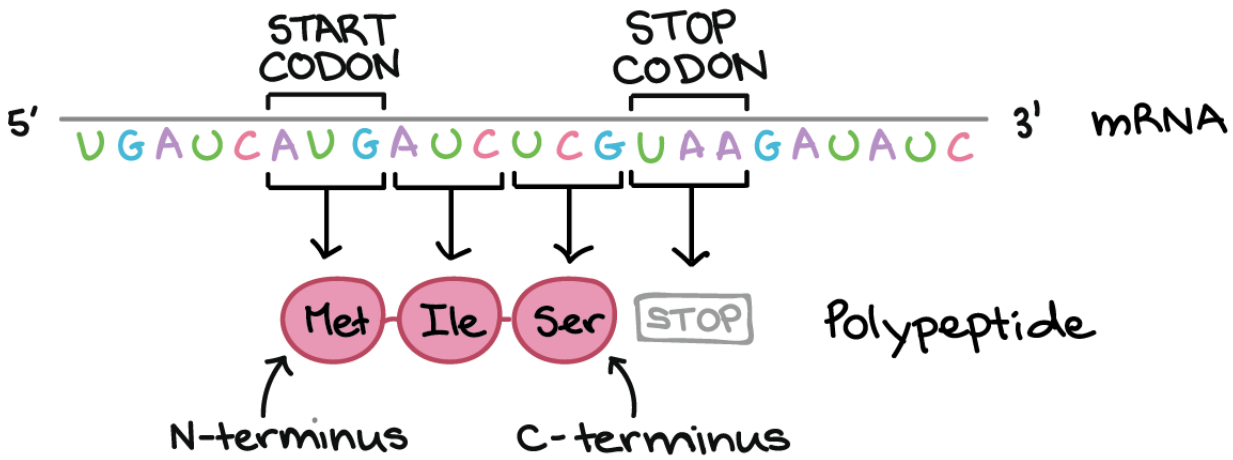
as the start codon to initiate translation. The reading frame for translation is set by the AUG start codon near the 5′ end of the mRNA. Following the start codon, the mRNA is read in groups of three until a stop codon is encountered.

The arrangement of the coding table reveals the structure of the code. There are sixteen "blocks" of codons, each specified by the first and second nucleotides of the codons within the block, e.g., the "AC*" block that corresponds to the amino acid threonine (Thr). Some blocks are divided into a pyrimidine half, in which the codon ends with U or C, and a purine half, in which the codon ends with A or G. Some amino acids get a whole block of four codons, like alanine (Ala), threonine (Thr) and proline (Pro). Some get the pyrimidine half of their block, like histidine (His) and asparagine (Asn). Others get the purine half of their block, like glutamate (Glu) and lysine (Lys). Note that some amino acids get a block and a half-block for a total of six codons.

The specification of a single amino acid by multiple similar codons is called "degeneracy." Degeneracy is believed to be a cellular mechanism to reduce the negative impact of random mutations. Codons that specify the same amino acid typically only differ by one nucleotide. In addition, amino acids with chemically similar side chains are encoded by similar codons. For example, aspartate (Asp) and glutamate (Glu), which occupy the GA* block, are both negatively charged. This

nuance of the genetic code ensures that a single-nucleotide substitution mutation might specify the same amino acid but have no effect or specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

The genetic code is nearly universal. With a few minor exceptions, virtually all species use the same genetic code for protein synthesis. Conservation of codons means that a purified mRNA encoding the globin protein in horses could be transferred to a tulip cell, and the tulip would synthesize horse globin. That there is only one genetic code is powerful evidence that all of life on Earth shares a common origin, especially considering that there are about 10^{84} possible combinations of 20 amino acids and 64 triplet codons.

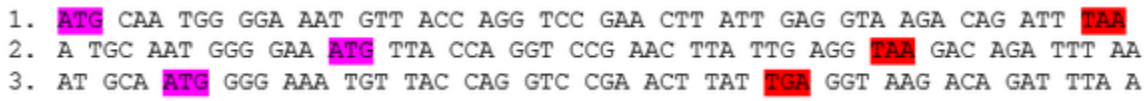


Open reading frame (ORF)

A reading frame is a non-overlapping set of three-nucleotide-codons (triplets) on DNA or RNA, which defines a gene mostly for a protein.

A nucleotide sequence that contains a start codon (initiation codon, typically AUG) and a stop codon (termination codons, UAA, UAG or UGA) is called an open reading frame (ORF).

An ORF is hence a part of an organism's genome which contains a sequence of bases that could potentially encode an individual protein. Once one knows the ORF for a gene or its mRNA, one can translate a nucleotide sequence into its corresponding amino acid sequence.



Mutations

A gene mutation is a permanent alteration in the DNA sequence that makes up a gene, such that multiple genes.

Gene the sequence differs from what is found in most people. Mutations range in size; they can affect anywhere from a single DNA building block (base pair) to a

large segment of a chromosome that includes mutations can be classified in two major ways:

- Hereditary mutations are inherited from a parent and are present throughout a person's life in virtually every cell in the body. These mutations are also called germline mutations because they are present in the parent's egg or sperm cells, which are also called germ cells. When an egg and a sperm cell unite, the resulting fertilized egg cell receives DNA from both parents. If this DNA has a mutation, the child that grows from the fertilized egg will have the mutation in each of his or her cells.
- Acquired (or somatic) mutations occur at some time during a person's life and are present only in certain cells, not in every cell in the body. These changes can be caused by environmental factors such as ultraviolet radiation from the sun, or can occur if an error is made as DNA copies itself during cell division. Acquired mutations in somatic cells (cells other than sperm and egg cells) cannot be passed to the next generation.