# Ribonucleic Acids (RNAs)

## *Main Features*

1. Each RNA nucleotide consists of a nitrogenous base, a ribose sugar and a phosphate.

2. Each RNA molecule typically is a single strand, consisting of a relatively short chain of nucleotides. RNA can be shaped like a single helix, a straight molecule, or may be bent or twisted upon itself. DNA, in comparison, is double-stranded and consists of a very long chain of nucleotides.

3. In RNA, the base adenine binds to uracil. In DNA, adenine binds to thymine. RNA does not contain thymine, uracil is an unmethylated form of thymine. Guanine binds to cytosine in both DNA and RNA.

4. There are several types of RNA. The most important are transfer RNA (tRNA), messenger RNA (mRNA), and ribosomal RNA (rRNA), but many more RNA types are contained inside a cell. RNA performs many functions in an organism, such as coding, decoding, regulating, and expressing genes.

5. About 5% of the weight of a human cell is RNA. Only about 1% of a cell consists of DNA.

6. RNA is found in both the nucleus and cytoplasm of human cells. DNA is only found in the cell nucleus and, in small amounts, in mithocondrions

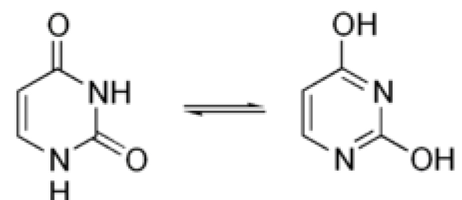
7. RNA is the genetic material for some organisms that don't have DNA. RNA exclusively containing Viruses are called Retroviruses.

8. Like DNA, most biologically active RNAs contain self-complementary sequences that allow parts of the RNA to fold and pair with itself to form double helices. Analysis of these RNAs has revealed that they are highly structured. Unlike DNA, their structures do not consist of long double helices but rather collections of short helices packed together into structures akin to proteins. In this fashion, RNAs can achieve chemical catalysis, like enzymes. For instance, determination of the structure of the ribosome—an enzyme that catalyzes peptide bond formation—revealed that its active site is composed entirely of RNA.

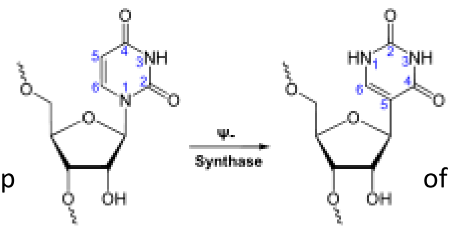
## *Bases and nucleotides*

Each nucleotide in RNA contains a ribose sugar.
A phosphate group is attached to the 3' position of one ribose and the 5' position of the next. The phosphate groups have a negative charge each at physiological pH, making RNA a charged molecule (polyanion).

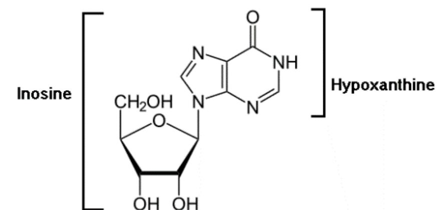
A base is attached to the 1' position, in general, adenine (A), cytosine (C), guanine (G), or uracil (U).
Uracil, the base not contained in DNA, undergoes amide-imidic acid tautomeric shifts. The amide tautomer is referred to as the lactam structure, while the imidic acid tautomer is referred to as the lactim structure. These tautomeric forms are predominant at pH 7. The lactam structure is the most common form of uracil.



RNA is transcribed with the only four bases stated above (adenine, cytosine, guanine and uracil), but these bases and attached sugars can be modified in numerous ways as the RNAs mature.
Pseudouridine (Ψ), in which the linkage between uracil and ribose changes from a C–N bond to a C–C bond, and ribothymidine (T) are found in various places (the most notable ones being in the TΨC loop of tRNA).

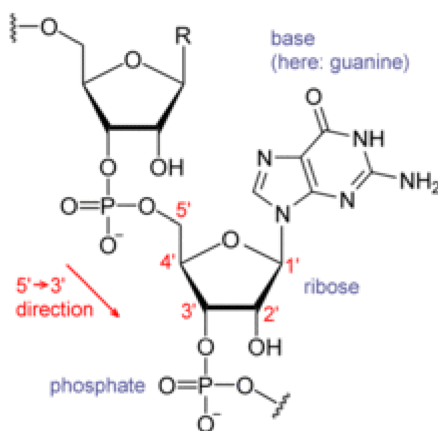



Another notable modified base is hypoxanthine, a deaminated adenine base whose nucleoside is called inosine (I).
There are more than 100 other naturally occurring modified nucleosides.
The greatest structural diversity of modifications can be found in tRNA, while pseudouridine and nucleosides with 2'-O-methylribose are the most common in rRNA.



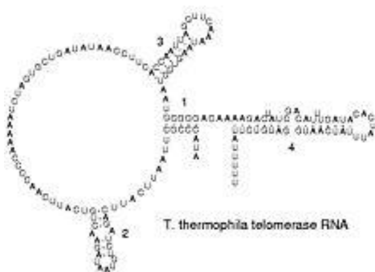

The specific roles of many of these modifications in RNA are not fully understood. However, it is notable that, in ribosomal RNA, many of the post-transcriptional modifications occur in highly functional regions, (peptidyl transferase center, subunit interface), implying that they are important for its functions.

## Structures



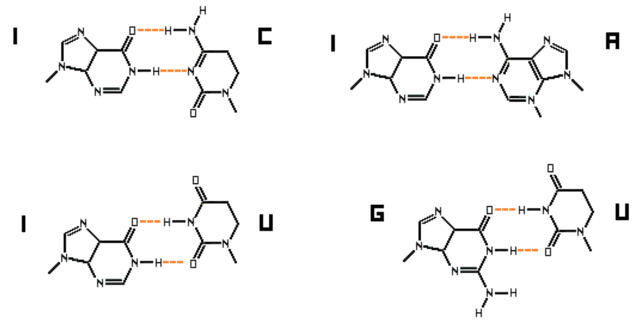

An important structural feature of RNA that distinguishes it from DNA is the presence of a hydroxyl group at the 2' position of the ribose sugar. The consequence of the presence of the 2'-hydroxyl group is that, in conformationally flexible regions of an RNA molecule (that is, not involved in formation of a double helix), it can chemically attack the adjacent phosphodiester bond to cleave the backbone (alkaline hydrolysis).

The functional form of single-stranded RNA molecules, just like proteins, frequently requires a specific tertiary structure. The scaffold for this structure is provided by secondary structural elements that are hydrogen bonds within the molecule. This leads to several recognizable "domains" of secondary structure like hairpin loops, bulges, and internal loops. Since RNA is charged, metal ions such as Mg2+ are needed to stabilise many secondary and tertiary structures.




Secondary structure of a telomerase RNA.

In RNA the bases form hydrogen bonds as in the standard Watson-Crick base pairing, that is cytosine with guanine and adenine with uracil.
In addition to that, RNA molecules show the so-called “wobble base pair”, that is a pairing between two nucleotides that does not follow Watson-Crick base pair rules. The four main wobble base pairs are guanine-uracil (G-U), hypoxanthine-uracil (I-U), hypoxanthine-adenine (I-A), and hypoxanthine-cytosine (I-C).

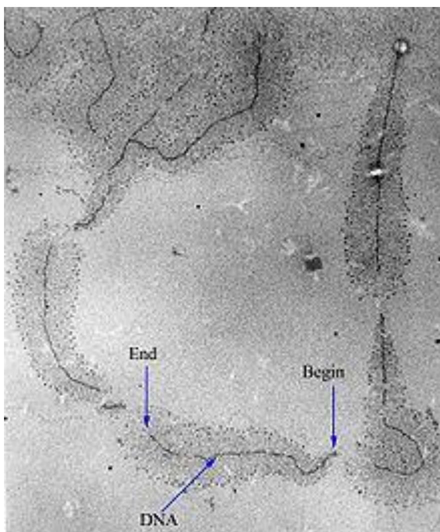



The thermodynamic stability of a wobble base pair is comparable to that of a Watson-Crick base pair. Wobble base pairs are both fundamental in RNA secondary structure and critical for the proper translation of the genetic code (see later)

Moreover, even other different sorts of interactions are possible, such as a group of adenine bases binding to each other in a bulge, or the “GNRA tetraloop” (a kind of hairpin) that has a guanine–adenine base-pair.

## *Synthesis and RNA polymerase*

Synthesis of RNA is usually catalyzed by an enzyme—RNA polymerase (RNAP)—using DNA as a template, a process known as transcription.
Initiation of transcription begins with the binding of the enzyme to a promoter sequence in the DNA (usually found "upstream" of a gene). The DNA double helix is unwound by the helicase activity of the enzyme. In contrast to DNA polymerase, RNAP includes helicase activity; therefore no separate enzyme is needed to unwind DNA.The enzyme then progresses along the template strand in the 3’ to 5’ direction, synthesizing a complementary RNA molecule with elongation occurring in the 5’ to 3’ direction.
The DNA sequence also dictates where termination of RNA synthesis will occur.




An electronic microscope photography of a DNA strand with hundreds RNAP molecules bound to it, too small to be seen. Every RNAP is transcribing an RNA strand, we can see them branching from the DNA strand. “Begin” indicates the DNA end where the synthesis begins; “End” indicates the 5’ end, where the longest RNA molecules are, completely transcribed.

Control of the process of gene transcription affects patterns of gene expression and, thereby, allows a cell to:

- adapt to a changing environment,
- perform specialized roles within an organism,
- maintain basic metabolic processes necessary for survival.

Therefore, it is hardly surprising that the activity of RNAP is both long and complex and highly regulated. For instance in *Escherichia coli* bacteria, more than 100 transcription factors have been identified, which modify the activity of RNAP.

RNAP can initiate transcription at specific DNA sequences known as **promoters**. It then produces an RNA chain, which is complementary to the template DNA strand. The process of adding nucleotides to the RNA strand is known as **elongation**; in eukaryotes, RNAP can build chains as long as 2.4 million nucleotides (the full length of the dystrophin gene). RNAP will release its RNA transcript at specific DNA sequences encoded at the end of genes, known as **terminators**.

RNAP accomplishes *de novo* synthesis. It is able to do this because specific interactions with the initiating nucleotide hold RNAP rigidly in place, facilitating chemical attack on the incoming nucleotide.

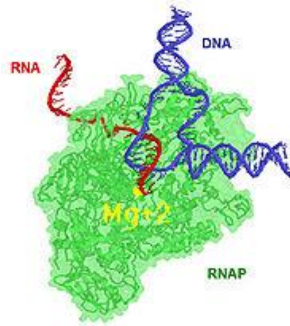
RNA polymerase binding in **bacteria** involves the **sigma factor** recognizing the core promoter region containing the -35 and -10 elements and the promoter upstream elements. There are multiple interchangeable sigma factors, each of which recognizes a distinct set of promoters. For example, in *E. coli*, $\sigma^{70}$ is expressed under normal conditions and recognizes promoters for genes required under normal conditions ("housekeeping genes"), while $\sigma^{32}$ recognizes promoters for genes required at high temperatures ("heat-shock genes").

After binding to the DNA, the RNA polymerase switches from a closed complex to an open complex. This change involves the separation of the DNA strands to form an unwound section of DNA of approximately 13 bp, referred to as the **transcription bubble**. Ribonucleotides are base-paired to the template DNA strand, according to Watson-Crick base-pairing interactions. Supercoiling plays an important part in polymerase activity because of the unwinding and rewinding of DNA. Because regions of DNA in front of RNAP are unwound, there is compensatory positive supercoils. Regions behind RNAP are rewound and negative supercoils are present.

As noted above, RNA polymerase makes contacts with the promoter region. However these stabilizing contacts inhibit the enzyme's ability to access DNA further downstream and thus the synthesis of the full-length product.
Once the open complex is stabilized, RNA polymerase synthesizes an RNA strand to establish a **DNA-RNA heteroduplex** (~8-9 bp) at the active center, which stabilizes the elongation complex.
In order to accomplish RNA synthesis, RNA polymerase must maintain promoter contacts while unwinding more downstream DNA for synthesis, "scrunching" more downstream DNA into the initiation complex. During the promoter escape transition, RNA polymerase is considered a "stressed intermediate." Thermodynamically the stress accumulates from the DNA-unwinding activities. Once the DNA-RNA heteroduplex is long enough, RNA polymerase releases its upstream contacts and effectively achieves the promoter escape transition into the elongation phase.

However, promoter escape is not the only outcome. RNA polymerase can also relieve the stress by releasing its downstream contacts, arresting transcription. The paused transcribing complex release the nascent transcript and begin anew at the promoter. Scientists have coined the term "**abortive initiation**" to explain the unproductive cycling of RNA polymerase before the promoter escape transition.

Once the RNAP starts forming longer transcripts it clears the promoter. At this point, the contacts with the -10 and -35 elements are disrupted, and the σ factor falls off RNAP. This allows the rest of the RNAP complex to move forward, as other σ factors held the RNAP complex in place.
As transcription progresses, ribonucleotides are added to the 3' end of the RNA transcript and the RNAP complex moves along the DNA.



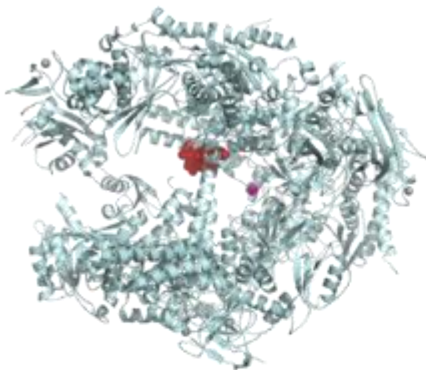

RNAP from made transparent so as to make the path of RNA and DNA more clear. The magnesium ion (yellow) is located at the enzyme active site. *T. aquaticus* pictured during elongation. Portions of the enzyme were

Although RNAP does not seem to have the 3'exonuclease activity that characterizes the ***proofreading*** activity found in DNA polymerase, there is evidence of that RNAP will halt at mismatched base-pairs and correct it.

Aspartyl (asp) residues in the RNAP will hold onto **$Mg^{2+}$ ions**, which will, in turn, coordinate the phosphates of the ribonucleotides adding to the chain.

Transcription of a palindrormic region of DNA causes the formation of a "**hairpin**" structure from the RNA transcription looping and binding upon itself. This hairpin structure is often rich in G-C base-pairs, making it more stable than the DNA-RNA hybrid itself. As a result, the 8 bp DNA-RNA hybrid in the transcription complex shifts to a 4 bp hybrid. These last 4 base pairs are weak A-U base pairs, and the entire RNA transcript will fall off the DNA causing the termination of transcription. In some other cases the termination is accomplished with the aid of a protein (**rho-protein**).

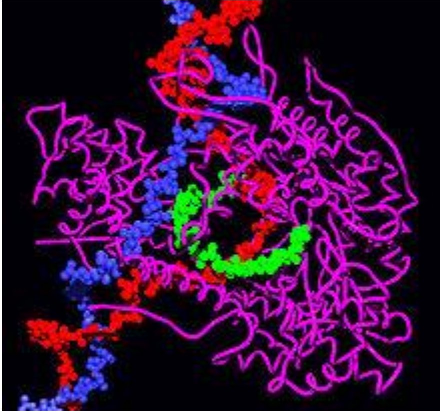
In bacteria the same enzyme catalyzes the synthesis of mRNA and ncRNA.
Bacterial RNAP is a large molecule. The core enzyme has five subunits (~400 kDa).
In order to bind promoters, RNAP core associates with the transcription initiation factor sigma (σ) to form RNA polymerase holoenzyme. Sigma reduces the affinity of RNAP for nonspecific DNA while increasing specificity for promoters, allowing transcription to initiate at correct sites. The complete holoenzyme therefore has 6 subunits (~450 kDa).



Structure of eukaryotic RNA polymerase II (light blue) in complex with α-amanitin (red), a strong poison found in death cap mushrooms that targets this vital enzyme

Eukaryotes have at least 5 types of nuclear RNAP, each responsible for synthesis of a distinct subset of RNA. All are structurally and mechanistically related to each other and to bacterial RNAP:

Given that DNA and RNA polymerases both carry out template-dependent nucleotide polymerization, it might be expected that the two types of enzymes would be structurally related. However, x-ray crystallographic studies of both types of enzymes reveal that, other than containing a critical Mg2+ ion at the catalytic site, they are virtually unrelated to each other.
Indeed template-dependent nucleotide polymerizing enzymes seem to have arisen independently twice during the early evolution of cells. One lineage led to the modern DNA Polymerases and reverse transcriptases. The other lineage formed all of the modern cellular RNA polymerases.



T7 RNA polymerase producing a mRNA (green) from a DNA template. The protein is shown as a purple ribbon. DNA strands are red and blue.

Viruses synthesize RNA using a wide range of mechanisms. Some of them are related to bacterial RNAP, eukaryotic nuclear, mitochondria or chloroplast RNAP. Other viruses use a RNA-dependent RNAP (an RNAP that employs RNA as a template instead of DNA).

## *Processing*

Primary transcript RNAs are often modified by enzymes after transcription.
The brief existence of an mRNA molecule begins with transcription, and ultimately ends in degradation. During its life, an mRNA molecule may also be processed, edited, and transported prior to translation. Eukaryotic mRNA molecules often require extensive processing and transport, while prokaryotic molecules do not.
The short-lived, unprocessed or partially processed product is termed precursor mRNA, or pre-mRNA; once completely processed, it is termed mature mRNA.

**5' cap addition**
A 5' cap is a modified guanine nucleotide that has been added to the "front"(5' end) of a eukaryotic messenger RNA shortly after the start of transcription. The 5' cap consists of a terminal 6-methylguanosine residue that is linked through a 5'-5'-triphosphate bond to the first transcribed nucleotide. Its presence is critical for recognition by the ribosome and protection from RNases.
Shortly after the start of transcription, the 5' end of the mRNA being synthesized is bound by a cap-synthesizing complex (associated with RNA polymerase) that catalyzes the chemical reactions required for mRNA capping.

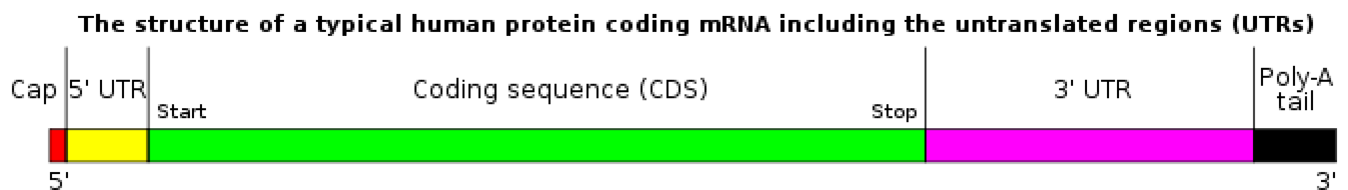
**Splicing**
Splicing is the process by which pre-mRNA is modified to remove certain stretches of non-coding sequences called **introns**; the stretches that remain include protein-coding sequences and are called **exons**.
Sometimes pre-mRNA messages may be spliced in several different ways, allowing a single gene to encode multiple proteins. This process is called alternative splicing. Splicing is usually performed by an RNA-protein complex called the spliceosome, but some RNA molecules are also capable of catalyzing their own splicing (e.g. ribozymes).

**Editing**
In some instances, an mRNA will be edited, changing the nucleotide composition of that mRNA. An example in humans is the apolipoprotein B mRNA, which is edited in some tissues, but not others. The editing creates an early stop codon, which, upon translation, produces a shorter protein.

**Polyadenylation**
Polyadenylation is the covalent linkage of a polyadenylyl moiety to a messenger RNA molecule.

In eukaryotic organisms, most messenger RNA (mRNA) molecules are polyadenylated at the 3' end, but recent studies have shown that short stretches of uridine (oligouridylation) are also common. The poly(A) tail and the protein bound to it aid in protecting mRNA from degradation by exonucleases. Polyadenylation is also important for transcription termination, export of the mRNA from the nucleus, and translation. mRNA can also be polyadenylated in prokaryotic organisms, where poly(A) tails act to facilitate, rather than impede, exonucleolytic degradation.
Polyadenylation occurs during and/or immediately after transcription of DNA into RNA. After transcription has been terminated, the mRNA chain is cleaved through the action of an endonuclease complex associated with RNA polymerase. After the mRNA has been cleaved, around 250 adenosine residues are added to the free 3' end at the cleavage site.

## *Mature RNA*




The picture above shows the structure of a mature eukaryotic mRNA. A fully processed mRNA includes a 5' cap, 5' UTR, coding region, 3' UTR, and poly(A) tail.

Coding regions are composed of codons, which are decoded and translated (in eukaryotes usually into one and in prokaryotes usually into several) into proteins by the ribosome. Coding regions begin with the start codon and end with a stop codon. In general, the start codon is an AUG triplet and the stop codon is UAA, UAG, or UGA. The coding regions tend to be stabilised by internal base pairs, this avoids degradation.

Untranslated regions (UTRs) are sections of the mRNA before the start codon and after the stop codon that are not translated: They are termed the five prime untranslated region (5' UTR) and three prime untranslated region (3' UTR), respectively. These regions are transcribed with the coding region and thus are exonic as they are present in the mature mRNA. Several roles in gene expression have been attributed to the untranslated regions, including mRNA stability and translational efficiency.
The stability of mRNAs may be controlled by the 5' UTR and/or 3' UTR due to varying affinity for RNA degrading enzymes called ribonucleases and for ancillary proteins that can promote or inhibit RNA degradation.
Translational efficiency, including sometimes the complete inhibition of translation, can be controlled by UTRs. Proteins that bind to either the 3' or 5' UTR may affect translation by influencing the ribosome's ability to bind to the mRNA. MicroRNAs bound to the 3' UTR also may affect translational efficiency or mRNA stability.
Some of the elements contained in untranslated regions form a characteristic secondary structure.
One class of mRNA element, the riboswitches, directly bind small molecules, changing their fold to modify levels of transcription or translation. In these cases, the mRNA regulates itself.

The 3' poly(A) tail is a long sequence of adenine nucleotides (often several hundred) added to the 3' end of the pre-mRNA. This tail promotes export from the nucleus and translation, and protects the mRNA from degradation.

## *mRNA transport*

A big difference between eukaryotes and prokaryotes is mRNA transport.
In prokaryotic cells, which do not have nucleus and cytoplasm compartments, mRNA can directly bind to ribosomes while it is being transcribed from DNA.

Because eukaryotic transcription and translation are compartmentally separated, eukaryotic mRNAs must be exported from the nucleus to the cytoplasm.
Mature mRNAs are recognized by their processed modifications and then exported through the nuclear pore by binding to cap-binding proteins, as well as the transcription/export complex (TREX).

In neurons, mRNA must be transported from the soma to the dendrites where local translation occurs in response to external stimuli. The mRNA associates to make a complex with a motor protein and is transported to the target location (neurite extension) along the cytoskeleton. Many messages are marked with so-called "zip codes," which target their transport to a specific location.

Because prokaryotic mRNA does not need to be processed or transported, translation by the ribosome can begin immediately after the end of transcription. Therefore, it can be said that prokaryotic translation is *coupled* to transcription and occurs *co-transcriptionally*.

Eukaryotic mRNA that has been processed and transported to the cytoplasm (i.e., mature mRNA) can then be translated by the ribosome. Translation may occur at ribosomes free-floating in the cytoplasm, or directed to the endoplasmic reticulum by the signal recognition particle. Therefore, unlike in prokaryotes, eukaryotic translation *is not* directly coupled to transcription.

## *Degradation*

Different mRNAs within the same cell have distinct lifetimes (stabilities).
In bacterial cells, individual mRNAs can survive from seconds to more than an hour; in mammalian cells, mRNA lifetimes range from several minutes to days.
The greater the stability of an mRNA the more protein may be produced from that mRNA. The limited lifetime of mRNA enables a cell to alter protein synthesis rapidly in response to its changing needs.
There are many mechanisms that lead to the destruction of an mRNA, some of which are described below.

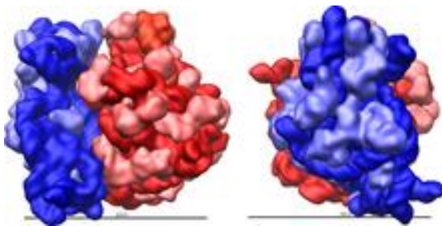
Prokaryotes degrade messages by using a combination of ribonucleases, including endonucleases, 3' exonucleases, and 5' exonucleases. In some instances, small RNA molecules (sRNA) tens to hundreds of nucleotides long can stimulate the degradation of specific mRNAs by base-pairing with complementary sequences and facilitating ribonuclease cleavage.

Inside eukaryotic cells, there is a balance between the processes of translation and mRNA decay.
Messages that are being actively translated are bound by ribosomes. Initiation factors block the decapping enzyme (DCP2), and poly(A)-binding protein blocks the exosome complex, protecting the ends of the message.
The poly(A) tail of the mRNA is shortened by specialized exonucleases. Poly(A) tail removal is thought to disrupt the circular structure of the message and destabilize the cap binding complex. The message is then subject to degradation by either the exosome complex or the decapping complex. In this way, translationally inactive messages can be destroyed quickly, while active messages remain intact. The mechanism by which translation stops and the message is handed-off to decay complexes is not understood in detail.

Eukaryotic messages are subject to surveillance by nonsense mediated decay (NMD), which checks for the presence of premature stop codons (nonsense codons) in the message. These can arise via incomplete splicing, mutations in DNA, transcription errors, leaky scanning by the ribosome causing a frame shift, and other causes. Detection of a premature stop codon triggers mRNA degradation by 5' decapping, 3' poly(A) tail removal, or endonucleolytic cleavage.

*Three-dimensional views of the ribosome, showing rRNA in dark blue (small subunit) and dark red (large subunit). Lighter colors represent ribosomal proteins.*

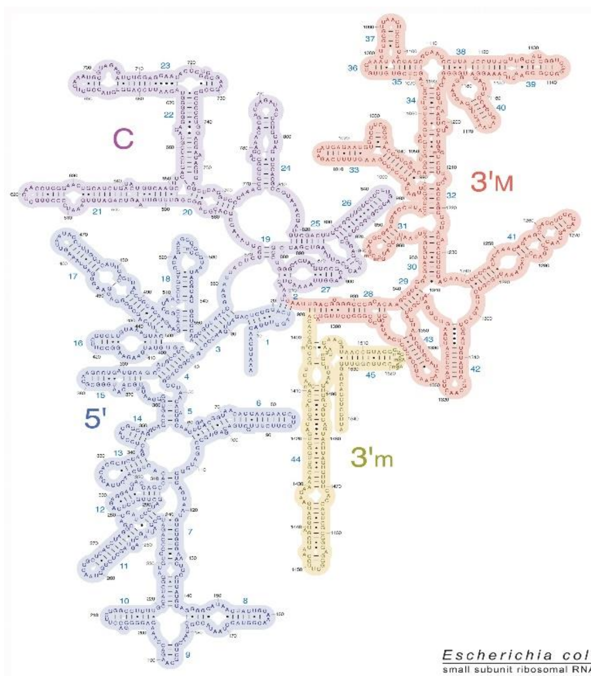
## *Ribosomal RNA (rRNA)*



rRNA is the catalytic component of the ribosomes.
Both prokaryotic and eukaryotic ribosomes can be broken down into two subunits. In the following table we report, for exemplary species, *Escherichia coli* (prokaryote) and human (eukaryote).

| **Type** | **Size** | **Large subunit** (rRNAs) | **Small subunit** (rRNA) |
|---|---|---|---|
| prokaryotic | 70S | 50S (5S : 120 nt, 23S : 2906 nt) | 30S (16S : 1542 nt) |
| eukaryotic | 80S | 60S (5S : 121 nt, 5.8S : 156 nt, 28S : 5070 nt[3]) | 40S (18S : 1869 nt]) |

S = Svedberg units, nt= length in nucleotides of the respective rRNAs.




Note that the S units of the subunits (or the rRNAs) cannot simply be added because they represent measures of sedimentation rate rather than of mass. The sedimentation rate of each subunit is affected by its shape, as well as by its mass. The nt units can be added as these represent the integer number of units in the linear rRNA polymers (for example, the total length of the human rRNA = 7216 nt).

Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA. Three of the rRNA molecules are synthesized in the nucleolus, and one is synthesized elsewhere. In the cytoplasm, ribosomal RNA and protein combine to form a nucleoprotein called a ribosome.
The ribosome binds mRNA and carries out protein synthesis. Several ribosomes may be attached to a single mRNA at any time. Nearly all the RNA found in a typical eukaryotic cell is rRNA.

## *Other types of RNA*

Besides mRNA, rRNA and tRNA, many other types are found inside human cells.
Most of them plays regulatory functions: several types of RNA can downregulate gene expression by being complementary to a part of an mRNA or a gene's DNA.

**Small nuclear ribonucleic acids** (snRNA) are found within the splicing speckles and cajal bodies of the cell nucleus in eukaryotic cells. The length of an average snRNA is approximately 150 nucleotides. They are transcribed by and studies have shown that their primary function is in the processing of pre-messenger RNA in the nucleus. They have also been shown to aid in the regulation of transcription factors and maintaining the telomeres.
snRNA are always associated with a set of specific proteins, and the complexes are referred to as small nuclear ribonucleoproteins (snRNP) often pronounced "snurps". Analysis has shown that these RNA were high in uridylate and were established in the nucleus.

**microRNAs** (abbreviated **miRNA**) are small non-coding RNA molecules (containing about 22 nucleotides) which function in RNA silencing and post-transcriptional regulation of gene expression.
miRNAs function via base-pairing with complementary sequences within mRNA molecules. As a result, these mRNA molecules are silenced by one or more of the following processes:

1) cleavage of the mRNA strand into two pieces,
2) destabilization of the mRNA through shortening of its poly(A) tail,
3) less efficient translation of the mRNA into proteins by ribosomes.

miRNAs derive from regions of RNA transcripts that fold back on themselves to form short hairpins. The human genome may encode over 1000 miRNAs, which appear to target about 60% of the genes of humans and other mammals.
In human cells, miRNAs are able to recognize their target mRNAs by using as little as 6–8 nucleotides (the seed region) at the 5' end of the miRNA, which is not enough pairing to induce cleavage of the target mRNAs. A given miRNA may have hundreds of different mRNA targets, and a given target might be regulated by multiple miRNAs.
Aberrant expression of miRNAs has been implicated in numerous disease states, and miRNA-based therapies are under investigation. Experiments show that a single miRNA may repress the production of hundreds of proteins, but that this repression often is relatively mild.

**Small interfering RNAs (siRNA;** 20-25 nt) are often produced by breakdown of viral RNA, there are also endogenous sources of siRNAs. siRNAs act through RNA interference in a fashion similar to miRNAs. Some miRNAs and siRNAs can cause genes they target to be methylated, thereby decreasing or increasing transcription of those genes.

**Antisense RNAs** are widespread; most downregulate a gene, but a few are activators of transcription.
The way antisense RNA can act is by binding to an mRNA, forming double-stranded RNA that is enzymatically degraded.
Reverse transcribing viruses replicate their genomes by reverse transcribing DNA copies from their RNA; these DNA copies are then transcribed to new RNA. Retrotransposons also spread by copying DNA and RNA from one another, and telomerase contains an RNA that is used as template for building the ends of eukaryotic chromosomes.

**Double-stranded RNA (dsRNA)** is RNA with two complementary strands, similar to the DNA found in all cells. dsRNA forms the genetic material of some viruses (double-stranded RNA viruses). Double-stranded RNA such as viral RNA or siRNA can trigger RNA interference in eukaryotes, as well as interferon response in vertebrates.

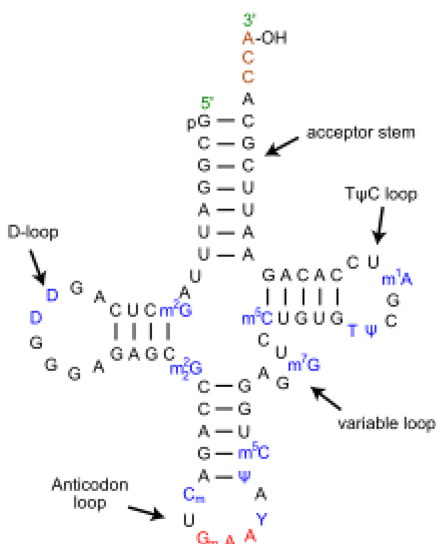
# *Transfer RNA (tRNA)*

A transfer RNA is a molecule composed of 76 to 90 nucleotides in length.
tRNAs are a necessary component of protein translation, the biological synthesis of new proteins according to the genetic code. They carry an amino acid to the ribosome directed by a three-nucleotide sequence (codon) in a messenger RNA (mRNA). Therefore, we can say that RNAs are the physical link between the nucleotide sequence of nucleic acids (DNA and RNA) and the amino acid sequence of proteins.
The nucleotide sequence of an mRNA specifies which amino acids are incorporated into the protein product of the gene from which the mRNA is transcribed.
The role of tRNA is to specify which sequence from the genetic code corresponds to which amino acid:

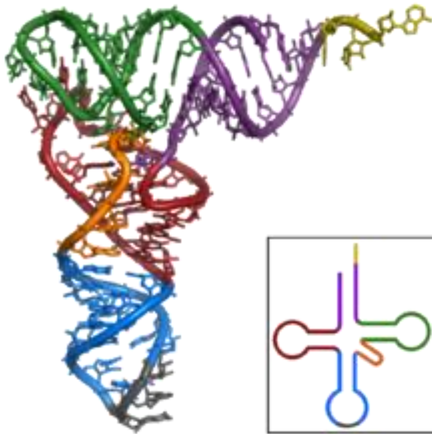
- One end of the tRNA matches the genetic code in a three-nucleotide sequence called the **anticodon**. The anticodon forms three base pairs with a codon in mRNA during protein biosynthesis. The mRNA encodes a protein as a series of contiguous codons, each of which is recognized by a particular tRNA.
- On the other end, the tRNA has a covalent bond to the **amino acid** that corresponds to the anticodon sequence. Each type of tRNA molecule can be attached to only one type of amino acid, so each organism has many types of tRNA (in fact, because the genetic code contains multiple codons that specify the same amino acid, there are many tRNA molecules bearing different anticodons which also carry the same amino acid).

The covalent bond of an a.a. to the tRNA 3' end is catalyzed by enzymes called **aminoacyl-tRNA synthetases**.
During protein synthesis, tRNAs with attached amino acids are delivered to the ribosome by proteins called **elongation factors**, which aid in decoding the mRNA codon sequence.
If the tRNA's anticodon matches the mRNA, another tRNA already bound to the ribosome transfers the growing polypeptide chain from its 3' end to the amino acid attached to the 3' end of the newly delivered tRNA, a reaction catalyzed by the ribosome.

## *tRNA structure*




Secondary *cloverleaf structure* of tRNA$^{Phe}$ from yeast.



Tertiary structure of tRNA. *CCA tail* in yellow, *Acceptor stem* in purple, *Variable loop* in orange, *D arm* in red, *Anticodon arm* in blue with *Anticodon* in black, *T arm* in green.

The structure of tRNA can be decomposed into its primary structure, its secondary structure (usually visualized as the *cloverleaf structure*), and its tertiary structure (all tRNAs have a similar L-shaped 3D structure that allows them to fit into the ribosome). The cloverleaf structure becomes the 3D L-shaped structure through coaxial stacking of the helices, which is a common RNA tertiary structure motif. The lengths of each arm, as well as the loop 'diameter', in a tRNA molecule vary from species to species.

The tRNA structure consists of the following parts:

1. A 5'-terminal phosphate group.
2. The acceptor stem is a 7- to 9-base pair (bp) stem made by the base pairing of the 5'-terminal nucleotide with the 3'-terminal nucleotide (which contains the CCA 3'-terminal group used to attach the amino acid). The acceptor stem may contain non-Watson-Crick base pairs.
3. The CCA tail is a cytosine-cytosine-adenine sequence at the 3' end of the tRNA molecule. The amino acid loaded onto the tRNA by aminoacyl tRNA synthetases, to form aminoacyl-tRNA, is covalently bonded to the 3'-hydroxyl group on the CCA tail. This sequence is important for the recognition of tRNA by enzymes and critical in translation. In most prokaryotic tRNAs and eukaryotic tRNAs, the CCA sequence is added during processing and therefore does not appear in the tRNA gene.
4. The D arm is a 4- to 6-bp stem ending in a loop that often contains dihydrouridine.
5. The anticodon arm is a 6-bp stem whose loop contains the anticodon. The tRNA 5'-to-3' primary structure contains the anticodon but in reverse order, since 3'-to-5' directionality is required to read the mRNA from 5'-to-3'.
6. The T arm is a 4- to 5- bp stem containing the sequence TΨC where Ψ is pseudouridine.
7. Bases that have been modified, especially by methylation occur in several positions throughout the tRNA. The first anticodon base, or wobble-position, is sometimes modified to inosine (derived from adenine), pseudouridine or lysidine (derived from cytosine).

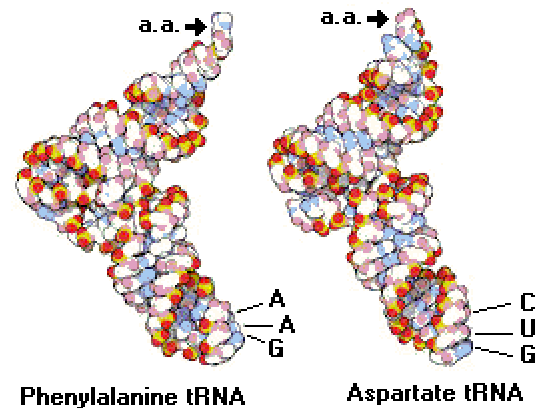
## *Anticodon*

An anticodon is a unit made up of three nucleotides that correspond to the three bases of the codon on the mRNA. Each tRNA contains a specific anticodon triplet sequence that can base-pair to one or more codons for an amino acid. Some anticodons can pair with more than one codon due to a phenomenon known as wobble base pairing. Frequently, the first nucleotide of the anticodon is one of two not found on mRNA: inosine and pseudouridine, which can hydrogen bond to more than one base in the corresponding codon position.

In the genetic code, it is common for a single amino acid to be specified by all four third-position possibilities, or at least by both pyrimidines and purines; for example, the amino acid glycine is coded for by the codon sequences GGU, GGC, GGA, and GGG.

To provide a one-to-one correspondence between tRNA molecules and codons that specify amino acids, 61 types of tRNA molecules would be required per cell. However, many cells contain fewer than 61 types of tRNAs because the wobble base is capable of binding to several, though not necessarily all, of the codons that specify a particular amino acid. A minimum of 31 tRNA are required to translate, unambiguously, all 61 sense codons of the standard genetic code

## *Aminoacylation*




Aminoacylation is the process of adding an aminoacyl group to a compound. It produces tRNA molecules with their CCA 3' ends covalently linked to an amino acid.
Each tRNA is aminoacylated (or charged) with a specific amino acid by an aminoacyl tRNA synthetase.
There is normally a single aminoacyl tRNA synthetase for each amino acid, despite the fact that there can be more than one tRNA, and more than one anticodon, for an amino acid.
Recognition of the appropriate tRNA by the synthetases is not mediated solely by the anticodon, and the acceptor stem often plays a prominent role.
ATP is needed for the acylation. Reactions are the following:

1. amino acid + ATP → aminoacyl-AMP + PPi
2. aminoacyl-AMP + tRNA → aminoacyl-tRNA + AMP

## *Binding to ribosome*

The **ribosome** is a large and complex molecular machine, found within all living cells, that serves as the primary site of biological protein synthesis (translation). Ribosomes consist of two major components:

- the small ribosomal subunit which reads the RNA
- the large subunit which joins amino acids to form a polypeptide chain.

Each subunit is composed of one or more ribosomal RNA molecules and a variety of proteins. A ribosome is therefore a ribonucleoprotein.
Ribosomes are ribozymes too, because the catalytic peptidyl transferase activity that links amino acids together is performed by the ribosomal RNA.
The sequence of translation can be outlined as follows:

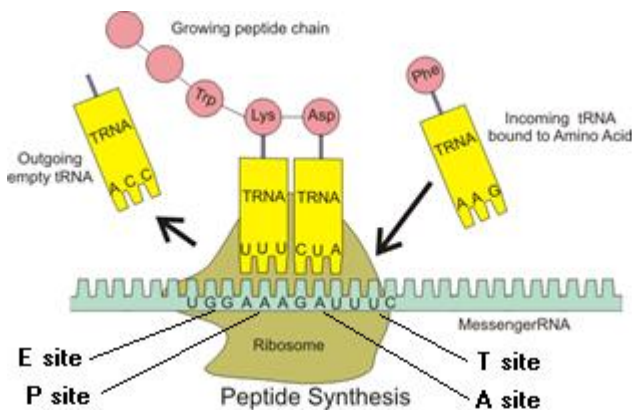
- Ribosomes' subunits bind to an mRNA chain and use it as a template for determining the correct sequence of amino acids in a particular protein.
- Amino acids are carried to the ribosome by tRNA molecules, which enter one part of the ribosome and bind to the messenger RNA chain.
- The attached amino acids are then linked together by another part of the ribosome.
- Once the protein is produced, it can then fold' to produce a specific functional three-dimensional structure.
- When a ribosome finishes reading an mRNA molecule, the two subunits split apart.

The ribosome has three binding sites for tRNA molecules that span the space between the two ribosomal subunits: the A (aminoacyl), P (peptidyl), and E (exit) sites. In addition, the ribosome has two other sites for tRNA binding that are used during mRNA decoding or during the initiation of protein synthesis. These are the T site (named elongation factor Tu) and I site (initiation).

More than one ribosome may move along a single mRNA chain at one time, each "reading" its sequence and producing a corresponding protein molecule.

The ribosomes in the mitochondria of eukaryotic cells functionally resemble many features of those in bacteria, reflecting the likely evolutionary origin of mitochondria

During translation elongation, tRNA first binds to the ribosome as part of a complex with elongation factor. This initial tRNA binding site is called the A/T site. In the A/T site, the A-site half resides in the small ribosomal subunit where the mRNA decoding site is located. The mRNA decoding site is where the mRNA codon is read out during translation. The T-site half resides mainly on the large ribosomal subunit where the E.F. interacts with the ribosome.
Once mRNA decoding is complete, the aminoacyl-tRNA is bound in the A/A site and is ready for the next peptide bond to be formed to its attached amino acid. The peptidyl-tRNA, which transfers the growing polypeptide to the aminoacyl-tRNA bound in the A/A site, is bound in the P/P site.




Once the peptide bond is formed, the tRNA in the P/P site is deacylated, or has a free 3' end, and the tRNA in the A/A site carries the growing polypeptide chain. To allow for the next elongation cycle, the tRNAs then move through hybrid A/P and P/E binding sites, before completing the cycle and residing in the P/P and E/E sites. Once the A/A and P/P tRNAs have moved to the P/P and E/E sites, the mRNA has also moved over by one codon and the A/T site is vacant, ready for the next round of mRNA decoding. The tRNA bound in the E/E site then leaves the ribosome.
The shifts of tRNA, mRNA and polypeptide from one site to another inside the ribosome are due to configuration changes in the ribosome 3-D structure.

The P/I site is actually the first to bind to aminoacyl tRNA, which is delivered by an initiation factor in bacteria. However, the existence of the P/I site in eukaryotic or archaeal ribosomes has not yet been confirmed.