

= RNA =

Ribonucleic acid ( RNA ) is a polymeric molecule implicated in various biological roles in coding , decoding , regulation , and expression of genes . RNA and DNA are nucleic acids , and , along with proteins and carbohydrates , constitute the three major macromolecules essential for all known forms of life . Like DNA , RNA is assembled as a chain of nucleotides , but unlike DNA it is more often found in nature as a single @-@ strand folded onto itself , rather than a paired double @-@ strand . Cellular organisms use messenger RNA ( mRNA ) to convey genetic information ( using the letters G , U , A , and C to denote the nitrogenous bases guanine , uracil , adenine , and cytosine ) that directs synthesis of specific proteins . Many viruses encode their genetic information using an RNA genome .

Some RNA molecules play an active role within cells by catalyzing biological reactions , controlling gene expression , or sensing and communicating responses to cellular signals . One of these active processes is protein synthesis , a universal function wherein mRNA molecules direct the assembly of proteins on ribosomes . This process uses transfer RNA ( tRNA ) molecules to deliver amino acids to the ribosome , where ribosomal RNA ( rRNA ) then links amino acids together to form proteins .

= = Comparison with DNA = =

Like DNA , most biologically active RNAs , including mRNA , tRNA , rRNA , snRNAs , and other non @-@ coding 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 .

= = Structure = =

Each nucleotide in RNA contains a ribose sugar , with carbons numbered 1 ' through 5 ' . A base is attached to the 1 ' position , in general , adenine ( A ) , cytosine ( C ) , guanine ( G ) , or uracil ( U ) . Adenine and guanine are purines , cytosine and uracil are pyrimidines . 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 , making RNA a charged molecule ( polyanion ) . The bases form hydrogen bonds between cytosine and guanine , between adenine and uracil and between guanine and uracil . However , other interactions are possible , such as a group of adenine bases binding to each other in a bulge , or the GNRA tetraloop that has a guanine ? adenine base @-@ pair .

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 presence of this functional group causes the helix to mostly adopt the A @-@ form geometry , although in single strand dinucleotide contexts , RNA can rarely also adopt the B @-@ form most commonly observed in DNA . The A @-@ form geometry results in a very deep and narrow major groove and a shallow and wide minor groove . A second 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 .

RNA is transcribed with only four bases ( adenine , cytosine , guanine and uracil ) , but these bases and attached sugars can be modified in numerous ways as the RNAs mature . Pseudouridine ( ? ) , in which the linkage between uracil and ribose is changed 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 T?C loop of tRNA ) . Another notable modified base is hypoxanthine , a deaminated adenine base whose nucleoside is called inosine ( I ) . Inosine plays a key role in the wobble hypothesis of the genetic

code .

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 often present in rRNA are the most common . 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 , such as the peptidyl transferase center and the subunit interface , implying that they are important for normal function .

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  $Mg^{2+}$  are needed to stabilise many secondary and tertiary structures .

The naturally occurring enantiomer of RNA is D RNA composed of D ribonucleotides . All chirality centers are located in the D ribose . By the use of L ribose or rather L ribonucleotides , L RNA can be synthesized . L RNA is much more stable against degradation by RNase .

Like other structured biopolymers such as proteins , one can define topology of a folded RNA molecule . This is often done based on arrangement of intra chain contacts within a folded RNA , termed as circuit topology .

= = Synthesis = =

Synthesis of RNA is usually catalyzed by an enzyme ? RNA polymerase ? 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 . 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 .

Primary transcript RNAs are often modified by enzymes after transcription . For example , a poly ( A ) tail and a 5' cap are added to eukaryotic pre mRNA and introns are removed by the spliceosome .

There are also a number of RNA dependent RNA polymerases that use RNA as their template for synthesis of a new strand of RNA . For instance , a number of RNA viruses ( such as poliovirus ) use this type of enzyme to replicate their genetic material . Also , RNA dependent RNA polymerase is part of the RNA interference pathway in many organisms .

= = Types of RNA = =

= = = Overview = = =

Messenger RNA ( mRNA ) is the RNA that carries information from DNA to the ribosome , the sites of protein synthesis ( translation ) in the cell . The coding sequence of the mRNA determines the amino acid sequence in the protein that is produced . However , many RNAs do not code for protein ( about 97 % of the transcriptional output is non protein coding in eukaryotes ) .

These so called non coding RNAs ( " ncRNA " ) can be encoded by their own genes ( RNA genes ) , but can also derive from mRNA introns . The most prominent examples of non coding RNAs are transfer RNA ( tRNA ) and ribosomal RNA ( rRNA ) , both of which are involved in the process of translation . There are also non coding RNAs involved in gene regulation , RNA processing and other roles . Certain RNAs are able to catalyse chemical reactions such as cutting

and ligating other RNA molecules , and the catalysis of peptide bond formation in the ribosome ; these are known as ribozymes .

== In length ==

According to the length of RNA chain , RNA includes small RNA and long RNA . Usually , small RNAs are < 200 nt in length , and long RNA are > 200 nt . Long RNAs , also called large RNAs , mainly include long non @-@ coding RNA ( lncRNA ) and mRNA . Small RNAs mainly include 5.8S ribosomal RNA ( rRNA ) , 5S rRNA , transfer RNA ( tRNA ) , microRNA ( miRNA ) , small interfering RNA ( siRNA ) , small nucleolar RNA ( snoRNAs ) , Piwi @-@ interacting RNA ( piRNA ) , tRNA @-@ derived small RNA ( tsRNA ) and small rDNA @-@ derived RNA ( srRNA ) .

== In translation ==

Messenger RNA ( mRNA ) carries information about a protein sequence to the ribosomes , the protein synthesis factories in the cell . It is coded so that every three nucleotides ( a codon ) correspond to one amino acid . In eukaryotic cells , once precursor mRNA ( pre @-@ mRNA ) has been transcribed from DNA , it is processed to mature mRNA . This removes its introns ? non @-@ coding sections of the pre @-@ mRNA . The mRNA is then exported from the nucleus to the cytoplasm , where it is bound to ribosomes and translated into its corresponding protein form with the help of tRNA . In prokaryotic cells , which do not have nucleus and cytoplasm compartments , mRNA can bind to ribosomes while it is being transcribed from DNA . After a certain amount of time the message degrades into its component nucleotides with the assistance of ribonucleases .

Transfer RNA ( tRNA ) is a small RNA chain of about 80 nucleotides that transfers a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation . It has sites for amino acid attachment and an anticodon region for codon recognition that binds to a specific sequence on the messenger RNA chain through hydrogen bonding .

Ribosomal RNA ( rRNA ) is the catalytic component of the ribosomes . 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 .

Transfer @-@ messenger RNA ( tmRNA ) is found in many bacteria and plastids . It tags proteins encoded by mRNAs that lack stop codons for degradation and prevents the ribosome from stalling .

== Regulatory RNAs ==

Several types of RNA can downregulate gene expression by being complementary to a part of an mRNA or a gene 's DNA . MicroRNAs ( miRNA ; 21 @-@ 22 nt ) are found in eukaryotes and act through RNA interference ( RNAi ) , where an effector complex of miRNA and enzymes can cleave complementary mRNA , block the mRNA from being translated , or accelerate its degradation .

While 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 . Animals have Piwi @-@ interacting RNAs ( piRNA ; 29 @-@ 30 nt ) that are active in germline cells and are thought to be a defense against transposons and play a role in gametogenesis .

Many prokaryotes have CRISPR RNAs , a regulatory system similar to RNA interference . Antisense RNAs are widespread ; most downregulate a gene , but a few are activators of transcription . One way antisense RNA can act is by binding to an mRNA , forming double @-@ stranded RNA that is enzymatically degraded . There are many long noncoding RNAs that regulate genes in eukaryotes , one such RNA is Xist , which coats one X chromosome in female mammals

and inactivates it .

An mRNA may contain regulatory elements itself , such as riboswitches , in the 5 ' untranslated region or 3 ' untranslated region ; these cis @-@ regulatory elements regulate the activity of that mRNA . The untranslated regions can also contain elements that regulate other genes .

== In RNA processing ==

Many RNAs are involved in modifying other RNAs . Introns are spliced out of pre @-@ mRNA by spliceosomes , which contain several small nuclear RNAs ( snRNA ) , or the introns can be ribozymes that are spliced by themselves . RNA can also be altered by having its nucleotides modified to other nucleotides than A , C , G and U. In eukaryotes , modifications of RNA nucleotides are in general directed by small nucleolar RNAs ( snoRNA ; 60 @-@ 300 nt ) , found in the nucleolus and cajal bodies. snoRNAs associate with enzymes and guide them to a spot on an RNA by basepairing to that RNA . These enzymes then perform the nucleotide modification. rRNAs and tRNAs are extensively modified , but snRNAs and mRNAs can also be the target of base modification . RNA can also be methylated .

== RNA genomes ==

Like DNA , RNA can carry genetic information . RNA viruses have genomes composed of RNA that encodes a number of proteins . The viral genome is replicated by some of those proteins , while other proteins protect the genome as the virus particle moves to a new host cell . Viroids are another group of pathogens , but they consist only of RNA , do not encode any protein and are replicated by a host plant cell 's polymerase .

== In reverse transcription ==

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 ==

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 .

== Circular RNA ==

Recently , it was shown that there is a single stranded covalently closed , i.e. circular form of RNA expressed throughout the animal and plant kingdom ( see circRNA ) . circRNAs are thought to arise via a " back @-@ splice " reaction where the spliceosome joins a downstream donor to an upstream acceptor splice site . So far the function of circRNAs is largely unknown , although for few examples a microRNA sponging activity has been demonstrated .

= Key discoveries in RNA biology =

Research on RNA has led to many important biological discoveries and numerous Nobel Prizes . Nucleic acids were discovered in 1868 by Friedrich Miescher , who called the material ' nuclein ' since it was found in the nucleus . It was later discovered that prokaryotic cells , which do not have a nucleus , also contain nucleic acids . The role of RNA in protein synthesis was suspected already in

1939 . Severo Ochoa won the 1959 Nobel Prize in Medicine ( shared with Arthur Kornberg ) after he discovered an enzyme that can synthesize RNA in the laboratory . However , the enzyme discovered by Ochoa ( polynucleotide phosphorylase ) was later shown to be responsible for RNA degradation , not RNA synthesis . In 1956 Alex Rich and David Davies hybridized two separate strands of RNA to form the first crystal of RNA whose structure could be determined by X @-@ ray crystallography .

The sequence of the 77 nucleotides of a yeast tRNA was found by Robert W. Holley in 1965 , winning Holley the 1968 Nobel Prize in Medicine ( shared with Har Gobind Khorana and Marshall Nirenberg ) . In 1967 , Carl Woese hypothesized that RNA might be catalytic and suggested that the earliest forms of life ( self @-@ replicating molecules ) could have relied on RNA both to carry genetic information and to catalyze biochemical reactions ? an RNA world .

During the early 1970s , retroviruses and reverse transcriptase were discovered , showing for the first time that enzymes could copy RNA into DNA ( the opposite of the usual route for transmission of genetic information ) . For this work , David Baltimore , Renato Dulbecco and Howard Temin were awarded a Nobel Prize in 1975 . In 1976 , Walter Fiers and his team determined the first complete nucleotide sequence of an RNA virus genome , that of bacteriophage MS2 .

In 1977 , introns and RNA splicing were discovered in both mammalian viruses and in cellular genes , resulting in a 1993 Nobel to Philip Sharp and Richard Roberts . Catalytic RNA molecules ( ribozymes ) were discovered in the early 1980s , leading to a 1989 Nobel award to Thomas Cech and Sidney Altman . In 1990 , it was found in Petunia that introduced genes can silence similar genes of the plant 's own , now known to be a result of RNA interference .

At about the same time , 22 nt long RNAs , now called microRNAs , were found to have a role in the development of *C. elegans* . Studies on RNA interference gleaned a Nobel Prize for Andrew Fire and Craig Mello in 2006 , and another Nobel was awarded for studies on the transcription of RNA to Roger Kornberg in the same year . The discovery of gene regulatory RNAs has led to attempts to develop drugs made of RNA , such as siRNA , to silence genes .

= = Evolution = =

In March 2015 , complex DNA and RNA organic compounds of life , including uracil , cytosine and thymine , were reportedly formed in the laboratory under outer space conditions , using starting chemicals , such as pyrimidine , found in meteorites . Pyrimidine , like polycyclic aromatic hydrocarbons ( PAHs ) , the most carbon @-@ rich chemical found in the Universe , may have been formed in red giants or in interstellar dust and gas clouds , according to the scientists .