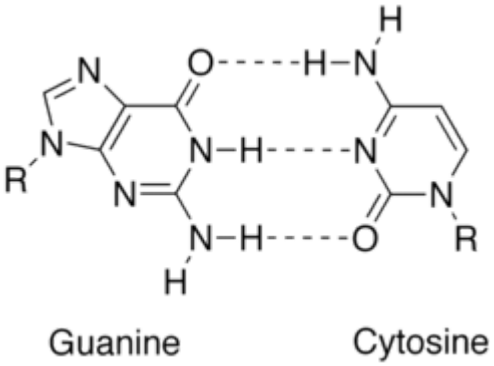
WIKIPEDIA

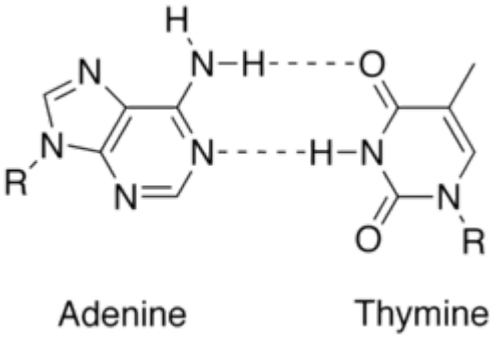
Complementarity (molecular biology)

In molecular biology, **complementarity** describes a relationship between two structures each following the lock-and-key principle. In nature complementarity is the base principle of DNA replication and transcription as it is a property shared between two DNA or RNA sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position in the sequences will be **complementary**, much like looking in the mirror and seeing the reverse of things. This complementary base pairing allows cells to copy information from one generation to another and even find and repair damage to the information stored in the sequences.



Match up between two DNA bases (guanine and cytosine) showing hydrogen bonds (dashed lines) holding them together

The degree of complementarity between two nucleic acid strands may vary, from complete complementarity (each nucleotide is across from its opposite) to no complementarity (each nucleotide is not across from its opposite) and determines the stability of the sequences to be together. Furthermore, various DNA repair functions as well as regulatory functions are based on base pair complementarity. In biotechnology, the principle of base pair complementarity allows the generation of DNA hybrids between RNA and DNA, and opens the door to modern tools such as cDNA libraries. While most complementarity is seen between two separate strings of DNA or RNA, it is also possible for a sequence to have internal complementarity resulting in the sequence binding to itself in a folded configuration.



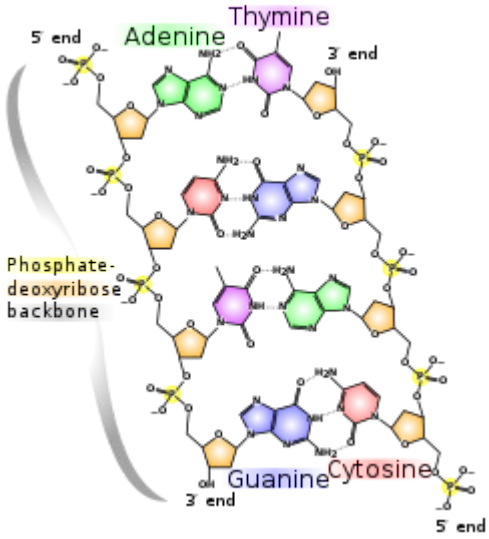
Match up between two DNA bases (adenine and thymine) showing hydrogen bonds (dashed lines) holding them together

Contents

DNA and RNA base pair complementarity

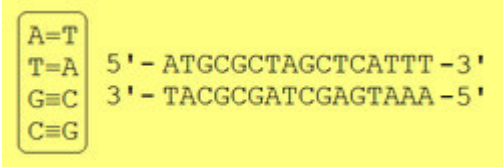
Complementarity is achieved by distinct interactions between nucleobases: adenine, thymine (uracil in RNA), guanine and cytosine. Adenine and guanine are purines, while thymine, cytosine and uracil are pyrimidines. Purines are larger than pyrimidines. Both types of molecules complement each other and can only base pair with the opposing type of nucleobase. In nucleic acid, nucleobases are held together by hydrogen bonding, which only works efficiently between adenine and thymine and between guanine and cytosine. The base complement A=T shares two hydrogen bonds, while the base pair G≡C has three hydrogen bonds. All other configurations between nucleobases would hinder double helix formation. DNA strands are oriented in opposite directions, they are said to be antiparallel.[1]

Nucleic Acid	Nucleobases	Base complement
DNA	adenine(A), thymine(T), guanine(G), cytosine(C)	A=T, G≡C
RNA	adenine(A), uracil(U), guanine(G), cytosine(C)	A=U, G≡C



Complementarity between two antiparallel strands of DNA. The top strand goes from the left to the right and the lower strand goes from the right to the left lining them up.

A complementary strand of DNA or RNA may be constructed based on nucleobase complementarity.[2] Each base pair, A=T vs. G≡C, takes up roughly the same space, thereby enabling a twisted DNA double helix formation without any spatial distortions. Hydrogen bonding between the nucleobases also stabilizes the DNA double helix.[3]



Left: the nucleotide base pairs that can form in double-stranded DNA. Between A and T there are two hydrogen bonds, while there are three between C and G. *Right:* two complementary strands of DNA.

Complementarity of DNA strands in a double helix make it possible to use one strand as a template to construct the other. This principle plays an important role in DNA replication, setting the foundation of heredity by explaining how genetic information can be passed down to the next generation. Complementarity is also utilized in DNA transcription, which generates an RNA strand from a DNA template.[4] In addition, human immunodeficiency virus, a single-stranded RNA virus, encodes an RNA-dependent DNA polymerase (reverse transcriptase) that uses complementarity to catalyze genome replication. The reverse transcriptase can switch between two parental RNA genomes by copy-choice recombination during replication.[5]

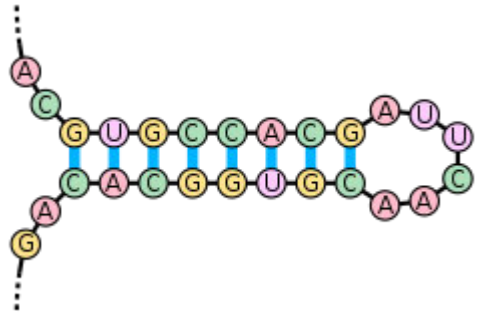
DNA repair mechanisms such as proof reading are complementarity based and allow for error correction during DNA replication by removing mismatched nucleobases.[1] In general, damages in one strand of DNA can be repaired by removal of the damaged section and its replacement by using complementarity to copy information from the other strand, as occurs in the processes of mismatch repair, nucleotide excision repair and base excision repair.[6]

Nucleic acids strands may also form hybrids in which single stranded DNA may readily anneal with complementary DNA or RNA. This principle is the basis of commonly performed laboratory techniques such as the polymerase chain reaction, PCR.[1]

Two strands of complementary sequence are referred to as sense and anti-sense. The sense strand is, generally, the transcribed sequence of DNA or the RNA that was generated in transcription. While the anti-sense strand is the strand that is complementary to the sense sequence.

Self-complementarity and hairpin loops

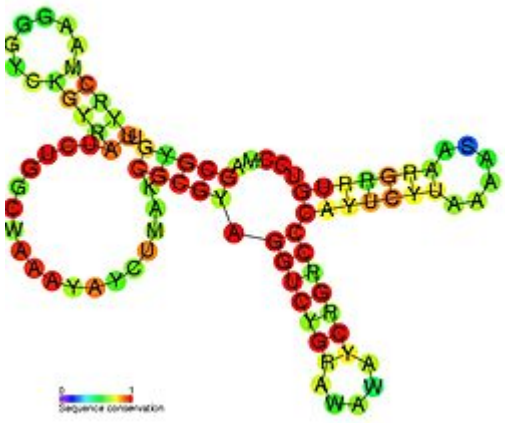
Self-complementarity refers to the fact that a sequence of DNA or RNA may fold back on itself, creating a double-strand like structure. Depending on how close together the parts of the sequence are that are self-complementary, the strand may form hairpin loops, junctions, bulges or internal loops.[1] RNA is more likely to form these kinds of structures due to base pair binding not seen in DNA, such as guanine binding with uracil.[1]



A sequence of RNA that has internal complementarity which results in it folding into a hairpin

Regulatory functions

Complementarity can be found between short nucleic acid stretches and a coding region or an transcribed gene, and results in base pairing. These short nucleic acid sequences are commonly found in nature and have regulatory functions such as gene silencing.[1]



A sequence of RNA showing hairpins (far right and far upper left), and internal loops (lower left structure)

Antisense transcripts

Antisense transcripts are stretches of non coding mRNA that are complementary to the coding sequence.[7] Genome wide studies have shown that RNA antisense transcripts occur commonly within nature. They are generally believed to increase the coding potential of the genetic code and add an overall layer of complexity to gene regulation. So far, it is known that 40% of the human genome is transcribed in both directions, underlining the potential significance of reverse transcription.[8] It has been suggested that complementary regions between sense and antisense transcripts would allow generation of double stranded RNA hybrids, which may play an important role in gene regulation. For example, *hypoxia-induced factor 1α mRNA* and *β-secretase mRNA* are transcribed bidirectionally, and it has been shown that the antisense transcript acts as a stabilizer to the sense script.[9]

miRNAs and siRNAs

miRNAs, microRNA, are short RNA sequences that are complementary to regions of a transcribed gene and have regulatory functions. Current research indicates that circulating miRNA may be utilized as novel biomarkers, hence show promising evidence to be utilized in disease diagnostics.[10] MiRNAs are formed from longer sequences of RNA that are cut free by a Dicer enzyme from an RNA sequence that is from a regulator gene. These short strands bind to a RISC complex. They match up with sequences in the upstream region of a transcribed gene due to their complementarity to act as a silencer for the gene in three ways. One is by preventing a ribosome from binding and initiating translation. Two is by degrading the mRNA that the complex has bound to. And three is by providing a new double-stranded RNA (dsRNA) sequence that Dicer can act upon to create more miRNA to find