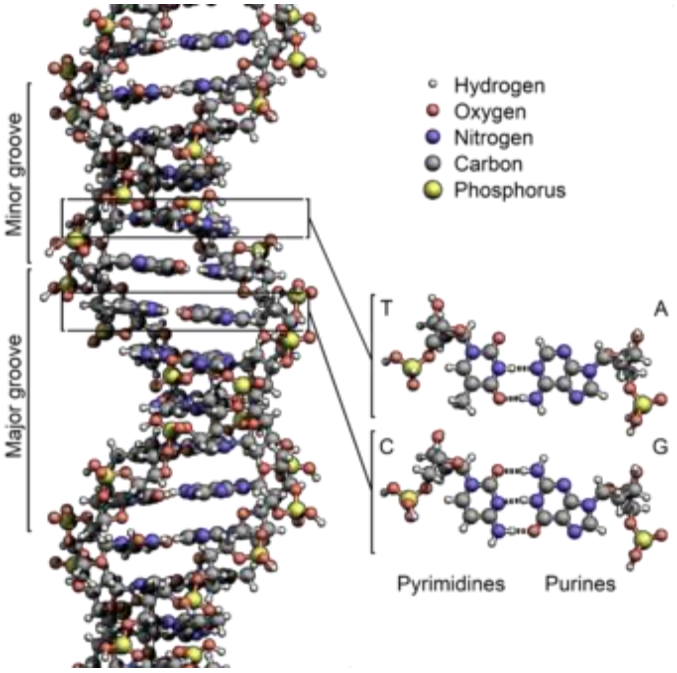
Struttura del capitolo (Pagine)

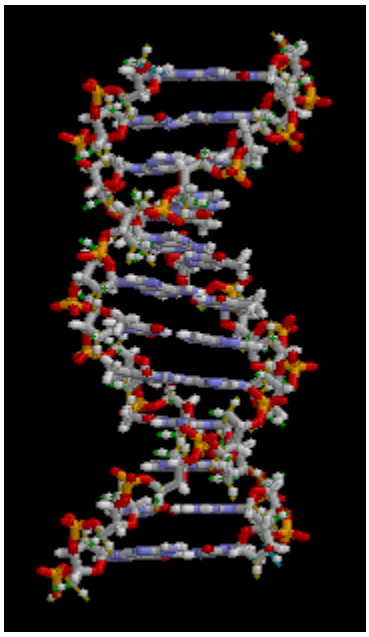
2 - DNA

- 3 - STRUCTURE
  - 4 - BASE PARING
  - 6 - GROOVES
  - 7 - SENSE AND ANTISENSE
  - 7 - SUPERCOILING
  - 8 - QUADRUPLEX STRUCTURES
  - 8 - PACKAGING
  - 9 - DAMAGES
- 11 - GENES AND GENOMES
  - 12 - GENETIC CODE
- 14 - REPLICATION
  - 15 - DNA POLYMERASE
  - 16 - INITIATION
  - 17 - ELONGATION
  - 17 - REPLICATION FORK
  - 19 - LEADING STRAND
  - 19 - LAGGING STRAND
  - 19 - REPLICATION MACHINERY
  - 20 - TERMINATION
  - 20 - REGULATION
- 21 - DNA BINDING PROTEINS
  - 22 - SINGLE STRAND BINDING PROTEINS
  - 22 - TRANSCRIPTION FACTORS
  - 23 - DNA MODIFYNG ENZYMES
    - 23 - NUCLEASES
    - 24 - LIGASES
    - 24 - TOPOISOMERASES
    - 25 - HELICASES
    - 25 - POLYMERASES
- 26 - GENETIC RECOMBINATION
- 26 - EVOLUTION
- 27 - HISTORY OF DNA RESEARCHES
- 28 - GENETIC DISORDERS

DNA

.



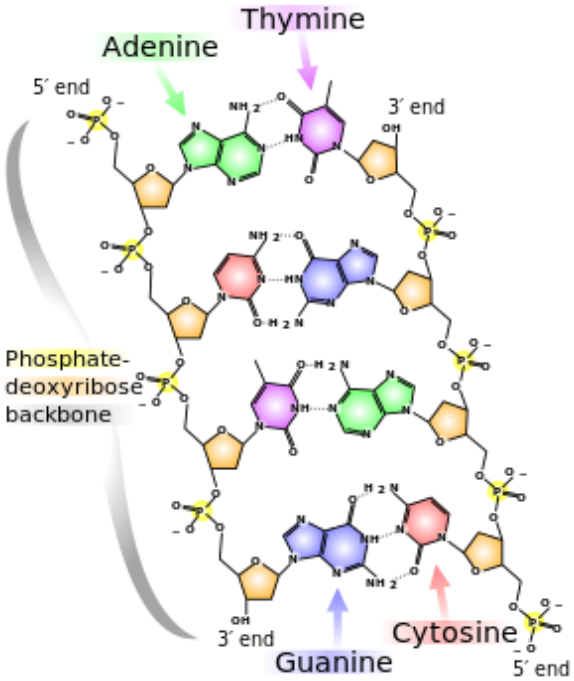


Deoxyribonucleic acid (**DNA**) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses.
Most DNA molecules consist of two biopolymer strands coiled around each other to form a double helix. The two DNA strands are known as polynucleotides since they are composed of simpler units called nucleotides.
Each nucleotide is composed of a nitrogen-containing nucleobase—either guanine (G), adenine (A), thymine (T), or cytosine (C)—as well as a monosaccharide sugar called deoxyribose and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds between the sugar of one nucleotide and the phosphate of the next, resulting in an alternating sugar-phosphate backbone.
Hydrogen bonds bind the nitrogenous bases of the two separate polynucleotide strands to make double-stranded DNA.

DNA is well-suited for biological information storage. The DNA backbone is resistant to cleavage, while the hydrogen bonds are easy to break permitting the strand separations whenever needed. Both strands of the double-stranded structure store the same biological information. Biological information is replicated as the two strands are separated. A significant portion of DNA (more than 98% for humans) is non-coding, meaning that these sections do not serve a function of encoding proteins.
The two strands of DNA run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of nucleobases (informally, *bases*). It is the sequence of these four nucleobases along the backbone that encodes biological information. Under the genetic code, RNA strands are translated to specify the sequence of amino acids within proteins. These RNA strands are initially created using DNA strands as a template in a process called transcription.

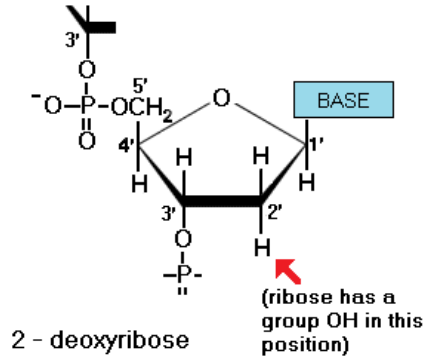
Within cells, DNA is organized into long structures called **chromosomes**. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as **histones** compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

Structure



DNA is a long polymer made from repeating units called nucleotides.
The structure of DNA of all species is formed by two helical chains each coiled round the same axis, and each with a pitch of 3.4 nanometres and a radius of 1.0 nanometres.
When measured in a particular solution, the DNA chain measured 2.2 to 2.6 nanometres wide, and one nucleotide unit measured 0.33 nm long. Although each individual repeating unit is very small, DNA polymers can be very large molecules containing millions of nucleotides. For instance, the largest human chromosome, chromosome number 1, consists of approximately 220 million base pairs and is 85 nm long.
A nucleobase linked to a sugar is called a nucleoside and a base linked to a sugar and one or more phosphate groups is called a nucleotide. A polymer comprising multiple linked nucleotides (as in DNA) is called a polynucleotide.

The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings.



These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. The asymmetric ends of DNA strands are called the 5′ (*five prime*) and 3′ (*three prime*) ends, with the 5′ end having a terminal phosphate group and the 3′ end a terminal hydroxyl group. One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.

A section of DNA. The bases lie horizontally between the two spiraling strands.

The DNA double helix is stabilized primarily by two forces:

- hydrogen bonds between nucleotides
- base-stacking interactions among aromatic nucleobases.

In the aqueous environment of the cell, the conjugated π bonds of nucleotide bases align perpendicular to the axis of the DNA molecule, minimizing their interaction with the solvation shell and therefore the Gibbs free energy.

Base pairing

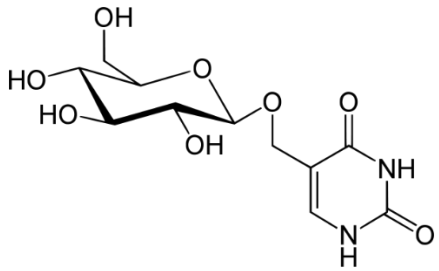
The four bases found in DNA are

1. adenine (abbreviated A),
2. cytosine (C),
3. guanine (G)
4. thymine (T).

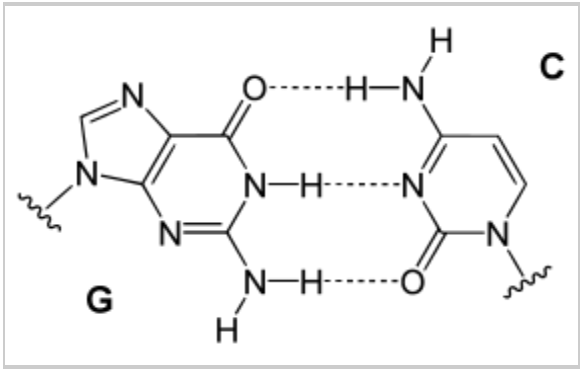
These four bases are attached to the sugar/phosphate to form the complete nucleotide.
The nucleobases are classified into two types: the purines, A and G, being fused five- and six-membered heterocyclic compounds, and the pyrimidines, the six-membered rings C and T.

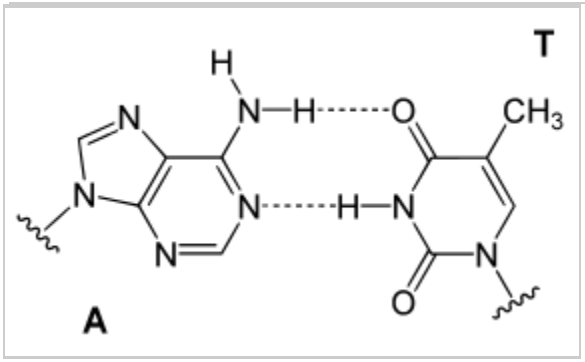
A fifth pyrimidine nucleobase, uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring.
Uracil is not usually found in DNA, occurring only as a breakdown product of cytosine. However, in a number of bacteriophages – *Bacillus subtilis* bacteriophages PBS1 and PBS2 , *Yersinia* bacteriophage piR1-37, Staphylococcal phage S6 - thymine has been replaced by uracil.

Another base, called Base J (beta-d-glucopyranosyloxymethyluracil), a modified form of uracil, is also found in a number of organisms including human pathogenic T*ripanosoma*. Proteins that bind specifically to this base J have been identified. Base J appears to act as a termination signal for RNA polymerase II.



In a DNA double helix, each type of nucleobase on one strand bonds with just one type of nucleobase on the other strand. This is called complementary base pairing.
Purines form hydrogen bonds to pyrimidines. Adenine is able to bond only to thymine and cytosine only to guanine. This happens because the base pair A-T can form only two hydrogen bonds, while the G-C can form three hydrogen bonds.



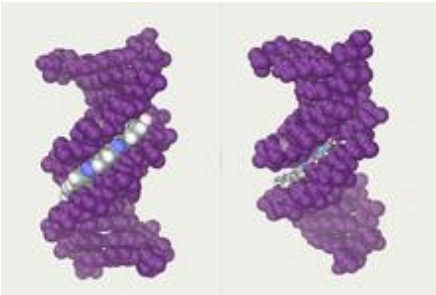


As a result of this complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication.
As hydrogen bonds are not covalent, they can be broken and rejoined with relative ease. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature. This reversible and specific interaction between complementary base pairs is actually critical for all the functions of DNA in living organisms.

The double stranded helical structure (**dsDNA**) is instead maintained largely by the intrastrand base stacking interactions, which are strongest for G,C stacks.
The two strands can come apart – a process known as “melting” – to form two single-stranded DNA molecules (**ssDNA**) molecules. Melting occurs at high temperature, low salt concentration

and high pH (low pH also melts DNA, but since DNA is unstable due to acid depurination, low pH is rarely used).
DNA with high GC-content is more stable than DNA with low GC-content.
The stability of the dsDNA form depends on sequence (since stacking is sequence specific) and length (longer molecules are more stable). The stability can be measured in various ways; a common way is the "melting temperature" T_m, which is the temperature at which 50% of the ds molecules are converted to ss molecules; melting temperature is dependent on ionic strength and the concentration of DNA.
The parts of DNA double helix that need to separate easily, such as the TATAAT Pribnow box (an essential part of a promoter site on DNA for transcription in bacteria) for exemple, tend to have a high AT content, making the strands easier to pull apart.

Grooves



Major and minor grooves of DNA. Minor groove is a binding site for the dye Hoechst 33258.

Twin helical strands form the DNA backbone. Another double helix may be found tracing the spaces, or grooves, between the strands. These voids are adjacent to the base pairs and may provide a binding site.
As the strands are not symmetrically located with respect to each other, the grooves are unequally sized.
One groove, the major groove, is 2,2 nm wide and the other, the minor groove, is 1,2 nm wide.
The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result proteins, like transcription factors, that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove. This situation varies in unusual conformations of DNA within the cell *(see below)*, but the major and minor grooves are always named to reflect the differences in size that would be seen if the DNA is twisted back into the ordinary B form.

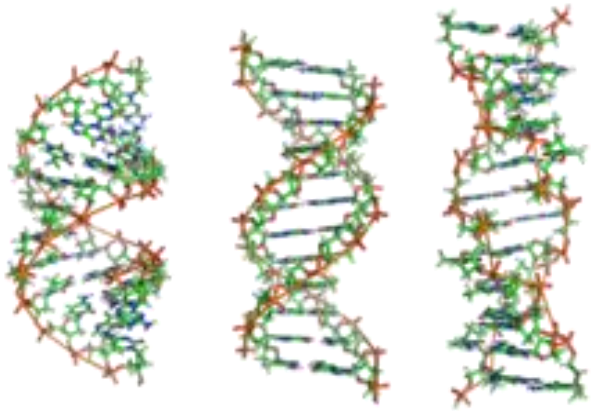
Sense and antisense

A DNA sequence is called "sense" if its sequence is the same as that of a messenger RNA copy that is translated into protein. The sequence on the opposite strand is called the "antisense" sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands can contain both sense and antisense sequences).
In both prokaryotes and eukaryotes, antisense RNA sequences are produced, but the functions of these RNAs are not entirely clear. One proposal is that antisense RNAs are involved in regulating gene expression through RNA-RNA base pairing.
A few DNA sequences in prokaryotes and eukaryotes, and more in plasmids and viruses, blur the distinction between sense and antisense strands by having overlapping genes. In these cases, some DNA sequences do double duty, encoding one protein when read along one strand, and a

second protein when read in the opposite direction along the other strand. In bacteria, this overlap may be involved in the regulation of gene transcription, while in viruses, overlapping genes increase the amount of information that can be encoded within the small viral genome.

Supercoiling

DNA can be twisted like a rope in a process called DNA supercoiling.
With DNA in its "relaxed" state, a strand usually circles the axis of the double helix once every 10.4 base pairs, but if the DNA is twisted the strands become more tightly or more loosely wound. If the DNA is twisted in the direction of the helix, this is positive supercoiling, and the bases are held more tightly together. If they are twisted in the opposite direction, this is negative supercoiling, and the bases come apart more easily. In nature, most DNA has slight negative supercoiling that is introduced by enzymes called topoisomerases. These enzymes are also needed to relieve the twisting stresses introduced into DNA strands during processes such as transcription and DNA replication.

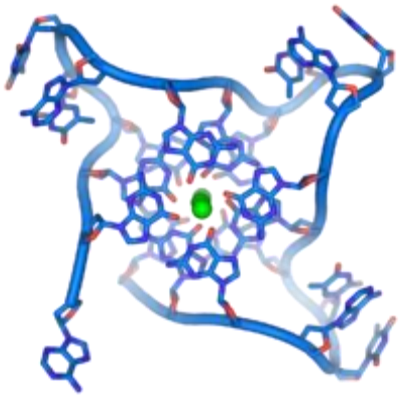


From left to right, the structures of A, B and Z DNA

DNA exists in many possible conformations that include A-DNA, B-DNA, and Z-DNA forms, although, only B-DNA and Z-DNA have been directly observed in functional organisms.
The conformation that DNA adopts depends on the hydration level, DNA sequence, the amount and direction of supercoiling, chemical modifications of the bases, the type and concentration of metal ions, as well as the presence of polyamines in solution.
Although the "B-DNA form" is most common under the conditions found in cells, it is not a well-defined conformation but a family of related DNA conformations that occur at the high hydration levels present in living cells. Their corresponding X-ray diffraction and scattering patterns are characteristic of molecular paracrystals with a significant degree of disorder.
Compared to B-DNA, the A-DNA form is a wider right-handed spiral, with a shallow, wide minor groove and a narrower, deeper major groove. The A form occurs under non-physiological conditions in partially dehydrated samples of DNA, while in the cell it may be produced in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes.
Segments of DNA where the bases have been chemically modified by methylation may undergo a larger change in conformation and adopt the Z form. Here, the strands turn about the helical axis in a left-handed spiral, the opposite of the more common B form. These unusual structures can be recognized by specific Z-DNA binding proteins and may be involved in the regulation of transcription.

Quadruplex structures

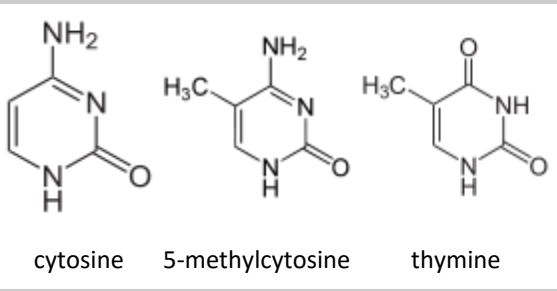
At the ends of the linear chromosomes are specialized regions of DNA called **telomeres**. The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, as the enzymes that normally replicate DNA cannot copy the extreme 3' ends of chromosomes.
These specialized chromosome caps also help protect the DNA ends, and stop the DNA repair systems in the cell from treating them as damage to be corrected. In human cells, telomeres are usually lengths of single-stranded DNA containing several thousand repeats of a simple TTAGGG sequence.



DNA quadruplex formed by telomere repeats. The looped conformation of the DNA backbone is very different from the typical DNA helix.

These guanine-rich sequences may stabilize chromosome ends by forming structures of stacked sets of four-base units, rather than the usual base pairs found in other DNA molecules. Here, four guanine bases form a flat plate and these flat four-base units then stack on top of each other, to form a stable G-quadruplex structure. These structures are stabilized by hydrogen bonding between the edges of the bases and chelation of a metal ion in the centre of each four-base unit. Other stacked structures can also be formed and, in addition to these stacked structures, telomeres also form large loop structures called telomere loops, or T-loops.

Chemical modifications and altered DNA packaging



cytosine 5-methylcytosine thymine

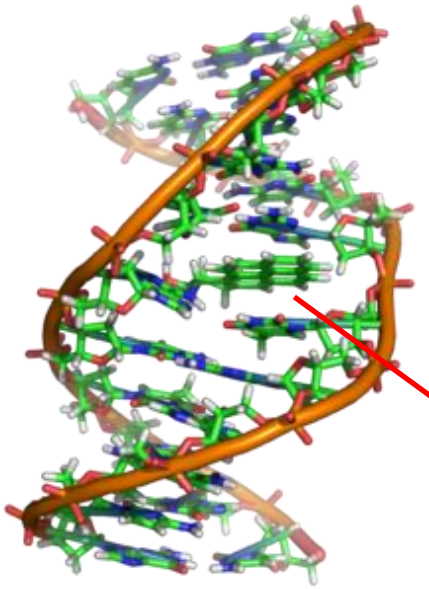
Structure of cytosine with and without the 5-methyl group.

Deamination converts 5-methylcytosine into thymine.

The expression of genes is influenced by how the DNA is packaged in chromosomes, in the structure called chromatin. Base modifications can be involved in packaging, with regions that have low or no gene expression usually containing high levels of methylation of cytosine bases. DNA packaging and its influence on gene expression can also occur by covalent modifications of the histone protein core around which DNA is wrapped in the chromatin structure or else by remodeling carried out by chromatin remodeling complexes. There is, further, crosstalk between DNA methylation and histone modification, so they can coordinately affect chromatin and gene expression.

For one example, cytosine methylation, produces 5-methylcytosine, which is important for X-chromosome inactivation. The average level of methylation varies between organisms – the worm *Caenorhabditis elegans* lacks cytosine methylation, while vertebrates have higher levels, with up to 1% of their DNA containing 5-methylcytosine. Despite the importance of 5-methylcytosine, it can deaminate to leave a thymine base, so methylated cytosines are particularly prone to mutations.
Other base modifications include adenine methylation in bacteria, the presence of 5-hydroxymethylcytosine in the brain, and the glycosylation of uracil to produce the "J-base" in kinetoplastids.

Damages



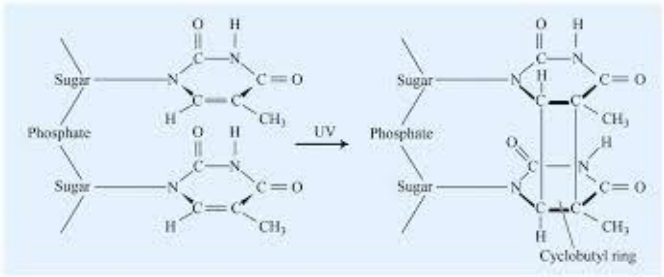
A covalent adduct between a metabolically activated form of benzo[*a*]pyrene, the major mutagen in tobacco smoke, and DNA

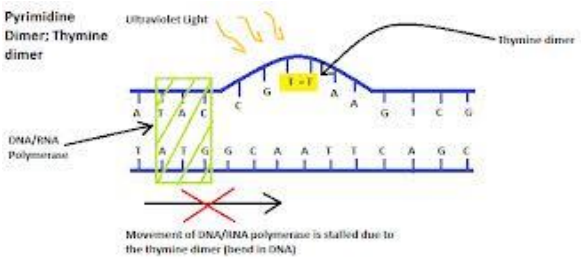
benzo[*a*]pyrene

DNA can be damaged by many sorts of mutagens, agents (molecules or radiations) which change the DNA sequence. **Mutagens** include:

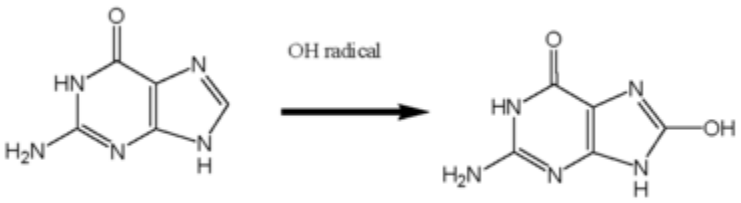
- oxidizing agents,
- alkylating agents,
- high-energy electromagnetic radiations such as ultraviolet light, X-rays and γ-rays.

The type of DNA damage produced depends on the type of mutagen.
UV light can damage DNA by producing thymine dimers, which are cross-links between pyrimidine bases.





On the other hand, oxidants such as free radicals or hydrogen peroxide produce multiple forms of damage, including base modifications (particularly of guanine) and double-strand breaks.

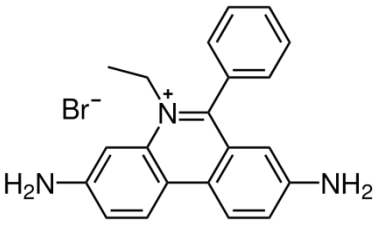


A typical human cell contains about 150,000 bases that have suffered oxidative damage. Of these oxidative lesions, the most dangerous are double-strand breaks, as these are difficult to repair and can produce

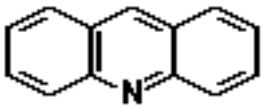
- point mutations,
- insertions,
- deletions,
- chromosomal translocations.

Insertion and deletion can cause frame-shift mutations.

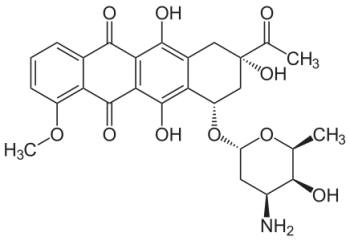
These mutations can cause cancer. Because of inherent limitations in the DNA repair mechanisms, if humans lived long enough, they would all eventually develop cancer.
DNA damages that are naturally occurring, due to normal cellular processes that produce reactive oxygen species, the hydrolytic activities of cellular water, etc. Although most of these damages are repaired, in any cell some DNA damage may remain despite the action of repair processes. These remaining DNA damages accumulate in the time lapse and appear to be an important underlying cause of aging.



Ethidium bromide

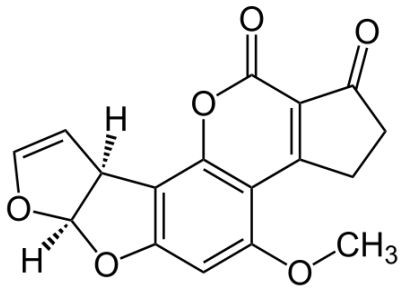


Acridine

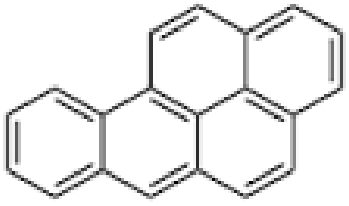


Daunomycin(cis)Doxorubicin(trans)

Many mutagens fit into the space between two adjacent base pairs, this is called *intercalation*. Most intercalators are aromatic and planar molecules; examples include ethidium bromide (a fluorescent tag), acridines (once use as dyes), daunomycin and doxorubicin(chemoterapeutics). For an intercalator to fit between base pairs, the bases must separate, distorting the DNA strands by unwinding of the double helix. This inhibits both transcription and DNA replication, causing toxicity and mutations. As a result, DNA intercalators may be carcinogens, and in the case of thalidomide, a teratogen.
Others such as benzo[*a*]pyrene diol epoxide and aflatoxin form DNA adducts that induce errors in replication.

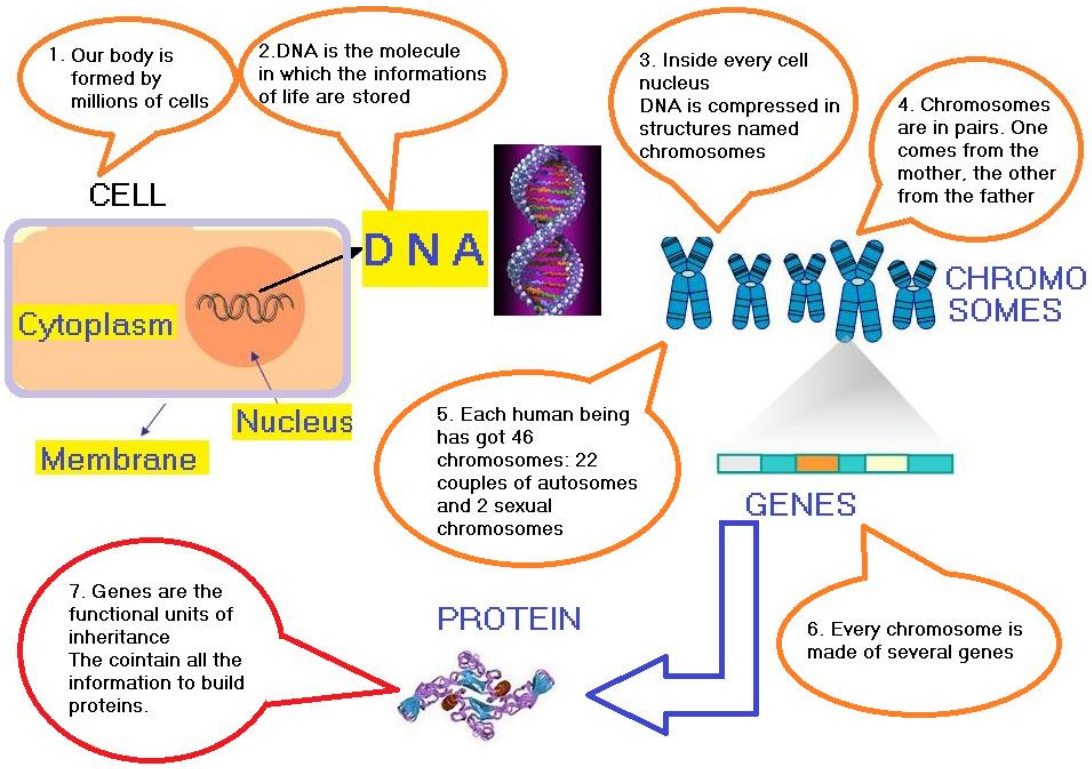


← aflatoxin B1 – benzo[a]pyrene →

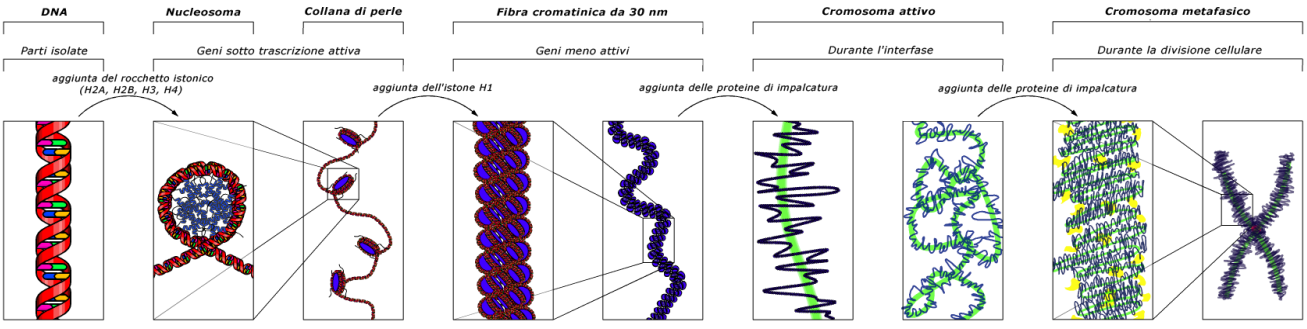


Nevertheless, due to their ability to inhibit DNA transcription and replication, other similar toxins are also used in chemotherapy to inhibit rapidly growing cancer cells.

Genes and genomes



DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes.
The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46
chromosomes.



The information carried by DNA is held in the sequence of pieces of DNA called genes.
Transmission of genetic information in genes is achieved via complementary base pairing.
In transcription, when a cell uses the information in a gene, the DNA sequence is copied into a complementary RNA sequence through the attraction between the DNA and the correct RNA nucleotides. Usually, this RNA copy is then used to make a matching protein sequence in a process called translation, which depends on the same interaction between RNA nucleotides.

In alternative fashion, a cell may simply copy its genetic information in a process called DNA replication.

Genomic DNA is tightly and orderly packed in the process called DNA condensation to fit the small available volumes of the cell. In eukaryotes, DNA is located in the cell nucleus, as well as small amounts in mitochondria and chloroplasts. In prokaryotes, the DNA is held within an irregularly shaped body in the cytoplasm called the nucleoid. The genetic information in a genome is held within genes, and the complete set of this information in an organism is called its genotype.
A gene is a unit of heredity and is a region of DNA that influences a particular characteristic in an organism.
Genes contain an open reading frame that can be transcribed, as well as regulatory sequences such as promoters and enhancers, which control the transcription of the open reading frame.
In many species, only a small fraction of the total sequence of the genome encodes protein. For example, only about 1.5% of the human genome consists of protein-coding exons, with over 50% of human DNA consisting of non-coding repetitive sequences.
The reasons for the presence of so much noncoding DNA in eukaryotic genomes and the extraordinary differences in genome size, or *C-value*, among species represent a long-standing puzzle known as the "C-value enigma".
However, some DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression.
Some noncoding DNA sequences play structural roles in chromosomes. Telomeres and centromeres typically contain few genes, but are important for the function and stability of chromosomes. An abundant form of noncoding DNA in humans are pseudogenes, which are copies of genes that have been disabled by mutation. These sequences are usually just molecular fossils, although they can occasionally serve as raw genetic material for the creation of new genes through the process of gene duplication and divergence.

Genetic Code

A gene is a sequence of DNA that contains genetic information and can influence the phenotype of an organism.

Within a gene, the sequence of bases along a DNA strand defines a messenger RNA sequence, which then defines one or more protein sequences.

The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by the rules of translation, known collectively as the genetic code.
The genetic code consists of three-letter 'words' called *codons* formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).
In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries amino acids.
Since there are 4 bases in 3-letter combinations, there are 64 possible codons (4^3 combinations). These encode the twenty standard amino acids, giving most amino acids more than one possible codon.
As it can be easily seen from the table above, nearly all a.acids are encoded by 4 triplets which have the same nucleotides in the first two positions. The third position can be occupied by each of

the four nucleotides and the encoding doesn't change. That's a kind of insurance against mutation (they are reduced by one third).

Did codons evolve to correspond to particular amino acids on chemical bases, or did the code evolve at random? The code evolved at random, in that there is no direct chemical connection between, say, GGG and Glycine. But the code appears to have evolved along certain lines for logical reasons. E.g. The two most "fundamental" amino acids are Gly and Ala, in biochemical pathways and in natural occurrence in prebiotic systems. Both are specified by G/C pairing at the first two positions--the strongest possible interaction. Early life, under high-heat conditions, would have needed extra strong codon-anticodon pairing. The first code may even have been a two-base code.

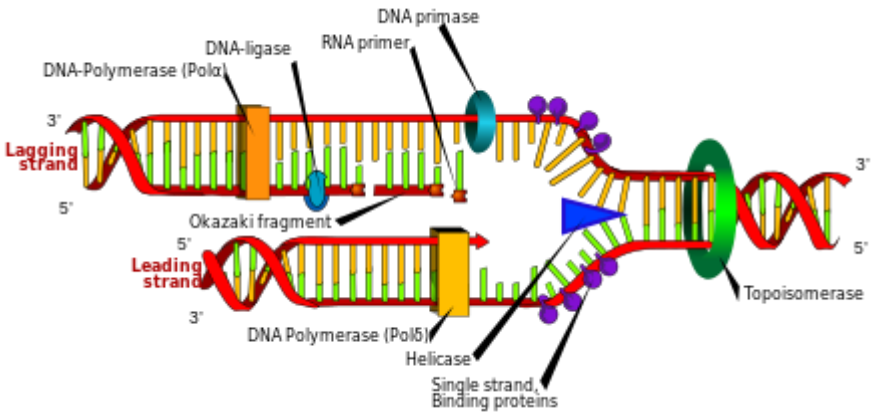
There are also three 'stop' or 'nonsense' codons signifying the end of the coding region; these are the TAA, TGA, and TAG codons.

There is only one start codon: AUG. It encodes Methionine and this is the only amino acid specified by just one codon.

1st letter	Second Letter: U	Second Letter: C	Second Letter: A	Second Letter: G	3rd letter
U	UUU Phe UUC UUA Leu UUG	UCU UCC UCA Ser UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
C	CUU CUC Leu CUA CUG	CCU CCC CCA Pro CCG	CAU His CAC CAA Gln CAG	CGU CGC CGA Arg CGG	U C A G
A	AUU AUC Ile AUA AUG Met	ACU ACC ACA Thr ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
G	GUU GUC Val GUA GUG	GCU GCC GCA Ala GCG	GAU Asp GAC GAA Glu GAG	GGU GGC GGA Gly GGG	U C A G

The sickle-cell disease occurs when, in hemoglobin, the sixth amino acid, glutamic acid, is replaced by valine. This substitution changes hemoglobin structure and function; that's possible simply by substituting an A with an U in the corresponding triplet.

Replication



DNA replication. The double helix is unwound by topoisomerase and a helicase. Next, one DNA polymerase produces the leading strand copy. Another DNA polymerase binds to the lagging strand. This enzyme makes discontinuous segments (called Okazaki fragments) before DNA ligase joins them together.

DNA replication is the process of producing two identical replicas from one original DNA molecule. This biological process occurs in all living organisms and is the basis for biological inheritance. DNA is made up of two strands and each strand of the original DNA molecule serves as template for the production of the complementary strand, a process referred to as semiconservative replication.

In a cell, DNA replication begins at specific locations, or **origins of replication**, in the genome. Unwinding of DNA at the origin and synthesis of new strands results in **replication forks** growing bidirectional from the origin. A number of proteins are associated with the replication fork which helps in terms of the initiation and continuation of DNA synthesis. DNA polymerase synthesizes the new DNA by adding complementary nucleotides to the template strand.

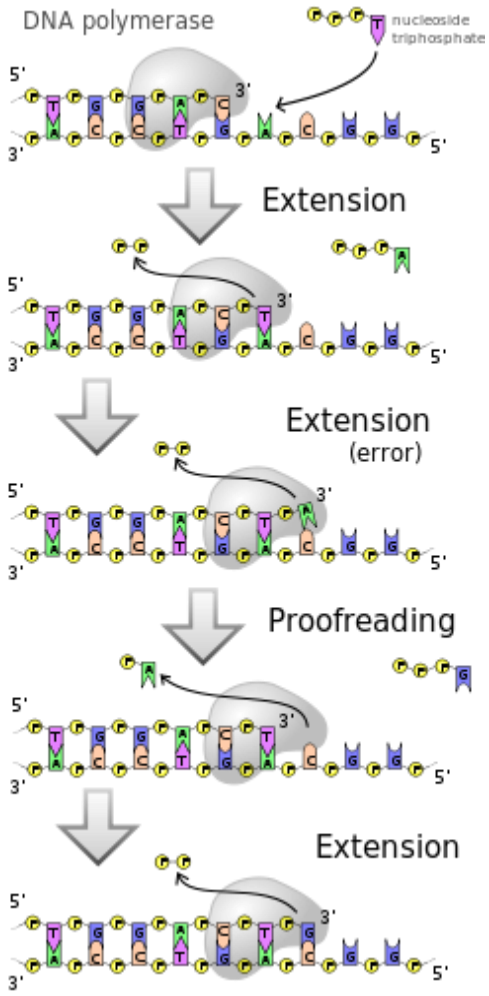
DNA replication can also be performed *in vitro* (artificially, outside a cell). DNA polymerases isolated from cells and artificial DNA primers can be used to initiate DNA synthesis at known sequences in a template DNA molecule.
The polymerase chain reaction (PCR), a common laboratory technique, cyclically applies such artificial synthesis to amplify a specific target DNA fragment from a pool of DNA.

As already stated, DNA strands have a directionality, and the different ends of a single strand are called the "3' (three-prime) end" and the "5' (five-prime) end". By convention, if the base sequence of a single strand of DNA is given, the left end of the sequence is 5' end, while the right end of the sequence is the 3' end. The strands of the double helix are anti-parallel with one being 5' to 3', and the opposite strand 3' to 5'. Directionality has consequences in DNA synthesis, because DNA polymerase can synthesize DNA in only one direction by adding nucleotides to the 3' end of a DNA strand.

DNA Replication, like all biological polymerization processes, proceeds in three enzymatically catalized and coordinated steps:

- initiation,
- elongation
- termination.

DNA Polymerase



DNA polymerases adds nucleotides to the 3' end of a strand of DNA.
If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension.
Proofreading removes the mismatched nucleotide and extension continues.

DNA polymerases are a family of enzymes that carry out all forms of DNA replication.

DNA polymerases in general cannot initiate synthesis of new strands, but can only extend an existing DNA or RNA strand paired with a template strand. To begin synthesis, a short fragment of RNA, called a primer, must be sythesized and paired with the template DNA strand.

DNA polymerase adds new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The energy for this process of polymerization comes from hydrolysis of the high-energy phosphate bonds between the three phosphates attached to each nucleotide. When a nucleotide is being added to a growing DNA strand, the formation of a phosphodiester bond between the proximal phosphate of the nucleotide to the growing chain is accompanied by hydrolysis of a high-energy phosphate bond with release of the two distal phosphates as a pyrophosphate. Enzymatic hydrolysis of the resulting pyrophosphate into inorganic phosphate consumes a second high-energy phosphate bond and renders the reaction effectively irreversible.

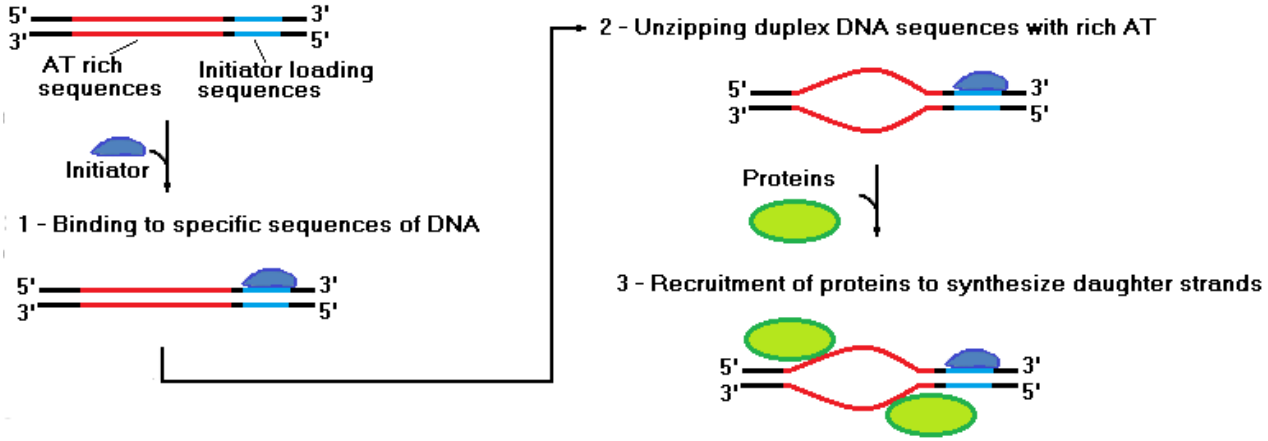
In general, DNA polymerases are highly accurate, with an intrinsic error rate of less than one mistake for every 10^7 nucleotides added. In addition, some DNA polymerases also have

proofreading ability; they can remove nucleotides from the end of a growing strand in order to correct mismatched bases.

Finally, post-replication mismatch repair mechanisms monitor the DNA for errors, being capable of distinguishing mismatches in the newly synthesized DNA strand from the original strand sequence. Together, these three discrimination steps enable replication fidelity of less than one mistake for every 10^9 nucleotides added.

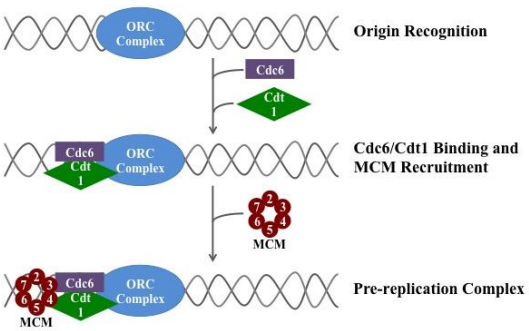
The rate of DNA replication in a living cell was first measured as the rate of phage T4 DNA elongation in phage-infected E. coli. During the period of exponential DNA increase at 37 °C, the rate was 749 nucleotides per second. The mutation rate per base pair per replication during phage T4 DNA synthesis is 1.7 per 10^8. Thus DNA replication is both impressively fast and accurate.

Initiation



Role of initiators for initiation of DNA replication.

For a cell to divide, it must first replicate its DNA. This process is initiated at particular points in the DNA, known as "origins", which are targeted by initiator proteins. In *E. coli* this protein is DnaA; in yeast, this is the origin recognition complex. Sequences used by initiator proteins tend to be "AT-rich" (rich in adenine and thymine bases), because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair) which are easier to unzip. Once the origin has been located, these initiators recruit other proteins and form the pre-replication complex, which unzips the double-stranded DNA.



Formation of pre-replication complex.

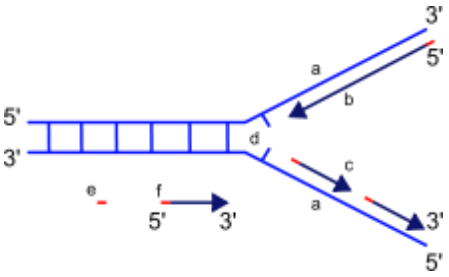
Elongation

DNA polymerase has 5'-3' activity. All known DNA replication systems require a free 3' hydroxyl group before synthesis can be initiated (Important note: DNA is read in 3' to 5' direction whereas a new strand is synthesized in the 5' to 3' direction—this is often confused). Distinct mechanisms for synthesis have been described for different life forms.
All cellular life forms and many DNA viruses, phages and plasmids use a primase to synthesize a short RNA primer with a free 3' OH group which is subsequently elongated by a DNA polymerase. Retroviruses employ a transfer RNA that primes DNA replication by providing a free 3′ OH that is used for elongation by the reverse transcriptase (few different replication mechanisms concerning viruses and plasmids are aldso known).
The first is used by the cellular organisms. In this mechanism, once the two strands are separated, primase adds RNA primers to the template strands. The leading strand receives one RNA primer while the lagging strand receives several. The leading strand is continuously extended from the primer by a high processivity, replicative DNA polymerase, while the lagging strand is extended discontinuously from each primer, forming Okazaki fragments. RNase removes the primer RNA fragments, and a low processivity DNA polymerase distinct from the replicative polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.
Multiple DNA polymerases take on different roles in the DNA replication process.
In *E. coli*, DNA Pol III is the polymerase enzyme primarily responsible for DNA replication. It assembles into a replication complex at the replication fork that exhibits extremely high processivity, remaining intact for the entire replication cycle.
In contrast, DNA Pol I is the enzyme responsible for replacing RNA primers with DNA. DNA Pol I has a 5' to 3' exonuclease activity in addition to its polymerase activity, and uses its exonuclease activity to degrade the RNA primers ahead of it as it extends the DNA strand behind it, in a process called nick translation. Pol I is much less processive than Pol III because its primary function in DNA replication is to create many short DNA regions rather than a few very long regions.
In eukaryotes, the low-processivity enzyme, Pol α, helps to initiate replication. The high-processivity extension enzymes are Pol δ and Pol ε.
As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming a replication fork with two prongs. In bacteria, which have a single origin of replication on their circular chromosome, this process creates a "theta structure" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

Processivity is a characteristic of enzymes that function on polymeric substrates. We can define it as remaining attached to their substrates and performing multiple rounds of catalysis before dissociating.
In the meantime, **Sliding** can be defined as remaining attached to the substrate and binding at multiple sites before dissociating. As such, it closely relates to processivity. Sliding along a linear polymer, such as DNA, reduces a three-dimensional search for a target site to a search in one dimension. The general value of reduction of dimensionality in biological search and diffusion processes has long been recognized.
In the case of DNA polymerase, the degree of processivity refers to the average number of nucleotides added each time the enzyme binds a template.
The average DNA polymerase requires about one second locating and binding a primer/template junction. Once it is bound, processive DNA polymerases add multiple nucleotides per second, drastically increasing the rate of DNA synthesis. The degree of processivity is directly proportional to the rate of DNA synthesis.

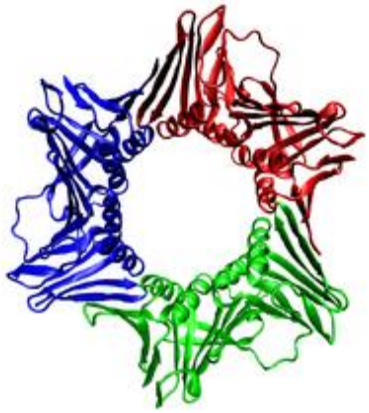
The rate of DNA synthesis was first determined as the rate of phage T4 DNA elongation in phage infected *E. coli*. During the period of exponential DNA increase at 37 °C, the rate was 749 nucleotides per second. DNA polymerase's ability to slide along the DNA template allows increased processivity.

Replication fork



Scheme of the replication fork.
a: template, b: leading strand, c: lagging strand, d: replication fork, e: primer, f: Okazaki fragments

The replication fork is a structure that forms within the nucleus during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together. The resulting structure has two branching "prongs", each one made up of a single strand of DNA. These two strands serve as the template for the leading and lagging strands, which will be created as DNA polymerase matches complementary nucleotides to the templates; the templates may be properly referred to as the leading strand template and the lagging strand template.
DNA is always synthesized in the 5' to 3' direction. Since the leading and lagging strand templates are oriented in opposite directions at the replication fork, a major issue is how to achieve synthesis of nascent (new) lagging strand DNA, whose direction of synthesis is opposite to the direction of the growing replication fork.



The assembled human DNA clamp, a trimer of the protein PCNA.

As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead. This build-up forms a torsional resistance that would eventually halt the progress of the replication fork. Topoisomerases are enzymes that temporarily break the strands of DNA, relieving the tension caused by unwinding the two strands of the DNA helix; topoisomerases achieve this by adding negative supercoils to the DNA helix. Bare single-stranded DNA tends to fold back on itself forming secondary structures; these structures can interfere with the movement of DNA polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.
Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template, thereby assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA

polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.

Leading strand

The leading strand is the strand of nascent DNA which is being synthesized in the same direction as the growing replication fork. A polymerase "reads" the leading strand *template* and adds complementary nucleotides to the nascent leading strand on a continuous basis.
The polymerase involved in leading strand synthesis is DNA polymerase III (DNA Pol III) in prokaryotes and Pol ε in yeasts. In human cells the leading and lagging strands are synthesized by Pol ε and Pol δ, respectively, within the nucleus and Pol γ in the mitochondria.

Lagging strand

The lagging strand is the strand of nascent DNA whose direction of synthesis is opposite to the direction of the growing replication fork. Because of its orientation, replication of the lagging strand is more complicated than that of the leading strand.
The lagging strand is synthesized in short, separated segments. On the lagging strand *template*, a primase "reads" the template DNA and initiates synthesis of a short complementary RNA primer. A DNA polymerase extends the primed segments, forming Okazaki fragments. The RNA primers are then removed and replaced with DNA, and the fragments of DNA are joined together by DNA ligase.
In eukaryotes, primase is intrinsic to Pol α. DNA polymerase III (in prokaryotes) or Pol δ/Pol ε (in eukaryotes) is/are responsible for extension of the primed segments. Primer removal in eukaryotes is also performed by Pol δ. In prokaryotes, primer removal is performed by DNA polymerase I, which "reads" the fragments, removes the RNA using its flap endonuclease domain (RNA primers are removed by 5'-3' exonuclease activity of polymerase I, and replaces the RNA nucleotides with DNA nucleotides.)

Replication machineries

At the replication fork, many replication enzymes assemble on the DNA into a complex molecular machine called the replisome. The following is a list of major DNA replication enzymes that participate in the replisome:

Enzyme	Function in DNA replication
DNA Helicase	Also known as helix destabilizing enzyme. Unwinds the DNA double helix at the Replication Fork.
DNA Polymerase	Builds a new duplex DNA strand by adding nucleotides in the 5' to 3' direction. Also performs proof-reading and error correction. There exist many different types of DNA Polymerase, each of which perform different functions in different types of cells.
DNA clamp	A protein which prevents DNA polymerase III from dissociating from the DNA parent strand.
Single-Strand Binding (SSB) Proteins	Bind to ssDNA and prevent the DNA double helix from re-annealing after DNA helicase unwinds it, thus maintaining the strand separation.
Topoisomerase	Relaxes the DNA from its super-coiled nature.
DNA Gyrase	Relieves strain of unwinding by DNA helicase; this is a specific type of topoisomerase

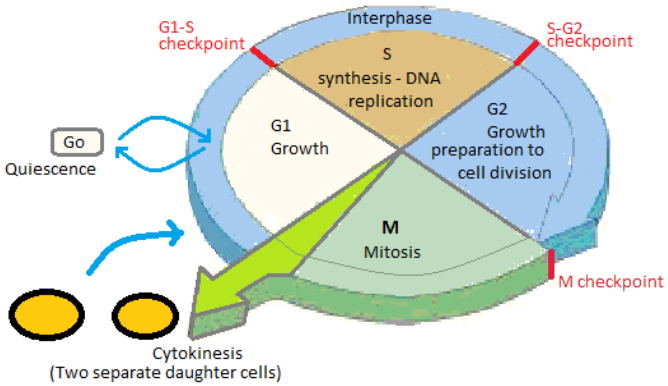
DNA Ligase	Re-anneals the semi-conservative strands and joins Okazaki Fragments of the lagging strand.
Primase	Provides a starting point of RNA (or DNA) for DNA polymerase to begin synthesis of the new DNA strand.
Telomerase	Lengthens telomeric DNA by adding repetitive nucleotide sequences to the ends of **eukaryotic chromosomes**.

Replication machineries mean complexes consisting of those factors involved in DNA replication and appearing on template ssDNAs.
The factors are replication enzymes; DNA polymerase, DNA helicases, DNA clamps and DNA topoisomerases, and replication proteins; e.g. single-stranded DNA binding proteins (SSB). In the replication machineries these components coordinate.

Termination

Termination requires that the progress of the DNA replication fork must stop or be blocked. Termination at a specific locus, when it occurs, involves the interaction between two components:
(1) one termination site sequence in the DNA
(2) one protein which binds to this sequence to physically stop DNA replication.
In various bacterial species, this is named the DNA replication terminus site-binding protein. Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome. *E. coli* regulates this process through the use of termination sequences that, when bound by the Tus protein, enable only one direction of replication fork to pass through. As a result, the replication forks are constrained to always meet within the termination region of the chromosome. Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome; these are not known to be regulated in any particular way. Because eukaryotes have linear chromosomes, DNA replication is unable to reach the very end of the chromosomes, but ends at the telomere region of repetitive DNA close to the end. This shortens the telomere of the daughter DNA strand. Shortening of the telomeres is a normal process in somatic cells. As a result, cells can only divide a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick limit.) Within the germ cell line, which passes DNA to the next generation, telomerase extends the repetitive sequences of the telomere region to prevent degradation. Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation.

Regulation



Within eukaryotes, DNA replication is controlled within the context of the cell cycle.
As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication occurs during the S phase (synthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases. The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA replication and subsequent division. Cells that do not proceed through this checkpoint remain in the G0 stage and do not replicate their DNA.
Replication of chloroplast and mitochondrial genomes occurs independently of the cell cycle, through a different process named D-loop replication.
In vertebrate cells, replication sites concentrate into positions called **replication foci**. Replication foci of varying size and positions appear in S phase of cell division and their number per nucleus is far smaller than the number of genomic replication forks.
Spatial juxtaposition of replication sites brings **clustering** of replication forks. The clustering do **rescue of stalled replication forks** and favours normal progress of replication forks. Progress of replication forks is inhibited by many factors (collision with proteins or with complexes binding strongly on DNA, deficiency of dNTPs, nicks on template DNAs and so on). If replication forks stall and remaining sequences from the stalled forks are not replicated, daughter strands have nick obtained un-replicated sites. The un-replicated sites on one parents strand hold other strand together but not daughter strands. Therefore, resulting sister chromatids cannot separate and cannot divide into 2 daughter cells. When neighbouring origins fire and a fork from one origin is stalled, fork from other origin access on an opposite direction of the stalled fork and duplicate the un-replicated sites. As other mechanism of the rescue there is application of **dormant replication origins** that excess origins don't fire in normal DNA replication.

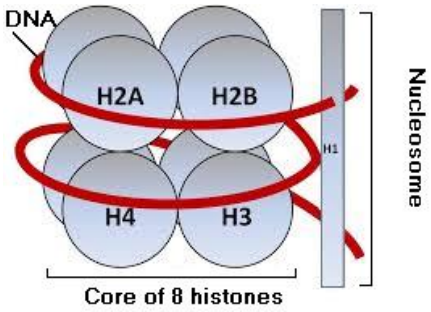
DNA-binding proteins

All the functions of DNA depend on interactions with proteins. These protein interactions can be non-specific (i.e. the protein can bind to every DNA sequence) or specific (i.e. the protein can bind specifically only to a single DNA sequence).
Enzymes can also bind to DNA and among these the polymerases, that copy the DNA base sequence in transcription and DNA replication, are particularly important.

Histones

Within chromosomes, DNA is held in complexes with structural proteins. These proteins organize the DNA into a compact structure called chromatin. In eukaryotes this structure involves DNA binding to a complex of small basic proteins called histones, while in prokaryotes multiple types of proteins are involved.

The histones form a disk-shaped complex called a nucleosome, which contains two complete turns of double-stranded DNA wrapped around eight histones and blocked by another one. These non-specific interactions are formed through basic residues (positive) in the histones making ionic bonds to the acidic sugar-phosphate backbone (negative) of the DNA, and are therefore largely independent of the base sequence.



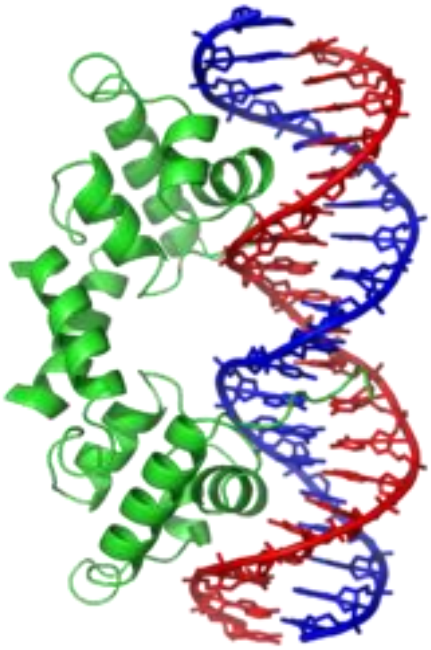
Chemical modifications of histones basic amino acid residues alter the strength of the interaction between the DNA and the histones, making the DNA more or less accessible to transcription factors and changing the rate of transcription. Those chemical modifications include methylation, phosphorylation and acetylation.

Single strand binding proteins

A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA.

In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA repair. These binding proteins stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

Transcription factors



The lambda repressor helix-turn-helix transcription factor bound to its DNA target

In contrast with non-specific DNA binding proteins, other proteins have evolved to bind to particular DNA sequences.

The most intensively studied of these are the various transcription factors, which are proteins that regulate transcription.

Each transcription factor binds to one particular set of DNA sequences and activates or inhibits the transcription of genes that have these sequences close to their promoters.

Transcription factors bind to either enhancer or promoter regions of DNA adjacent to the genes that they regulate. Depending on the transcription factor, the transcription of the adjacent gene is either up- or down-regulated.

The transcription factors do this in two ways.

- They can bind the RNA polymerase responsible for transcription, either directly or through other mediator proteins; this locates the polymerase at the promoter and allows it to begin transcription.
- Transcription factors can bind enzymes that modify the histones at the promoter. This changes the accessibility of the DNA template to the polymerase. Generally these enzymes catalyze the acetylation (weakens the association of DNA with histones, which make the DNA more accessible to transcription) or deacetylation of histone proteins (strengthens the association of DNA with histones, which make the DNA less accessible to transcription) weakens the association of DNA with histones, which make the DNA more accessible to transcription)

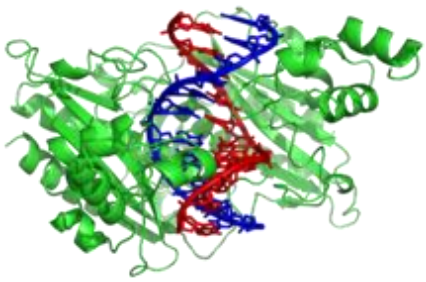
As these DNA targets can occur throughout an organism's genome, changes in the activity of one type of transcription factor can affect thousands of genes. Consequently, these proteins are often the targets of the signal transduction processes that control responses to environmental changes or cellular differentiation and development.

The specificity of these transcription factors' interactions with DNA come from the proteins making multiple contacts to the edges of the DNA bases, allowing them to "read" the DNA sequence. Most of these base-interactions are made in the major groove, where the bases are most accessible.

DNA-modifying enzymes

Enzymes which interact with DNA includes:

- Nucleases
- Ligases
- Topoisomerases
- Helicases
- Polymerases



The restriction enzyme EcoRV (green) in a complex with its substrate DNA

Nucleases

Nucleases are enzymes that cut DNA strands by catalyzing the hydrolysis of the phosphodiester bonds. Nucleases that hydrolyse nucleotides from the ends of DNA strands are called exonucleases, while endonucleases cut within strands.

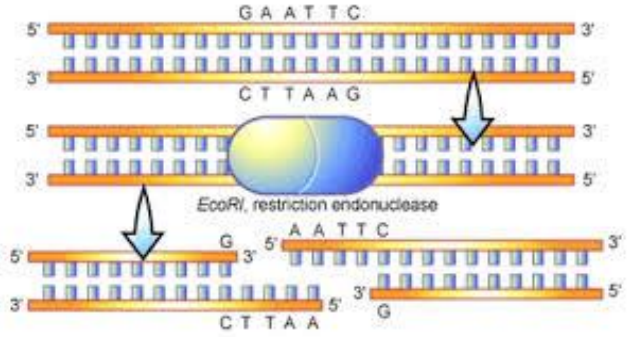
The most frequently used nucleases in molecular biology are the restriction endonucleases, which cut DNA at specific sequences.
For instance, the EcoRV enzyme shown above recognizes the 6-base sequence 5’-GATATC-3’ and makes a cut at the vertical line.
On the other hand the EcoRI enzyme recognize the sequence:

```
5'GAATTC
3'CTTAAG
```

This enzyme makes a “staggered cut” like that:

```
5'---G     AATTC---3'
3'---CTTAA     G---5'
```

leaving a single strain at the end of both parts of the cut DNA (“sticky ends”).

1. 

In nature, these enzymes protect bacteria against phage infection by digesting the phage DNA when it enters the bacterial cell, acting as part of the restriction modification system.
In biotechnology these sequence-specific nucleases are very useful as they allow, with the help of ligases, to “cut and paste” different DNA sequences and built new man-made DNA molecules (recombinant DNA)
Methods have been developed to purify DNA from organisms, such as phenol-chloroform extraction, and to manipulate it in the laboratory, such as restriction digests and the polymerase chain reaction. Modern biology and biochemistry make intensive use of these techniques in recombinant DNA technology. Recombinant DNA is a man-made DNA sequence that has been assembled from other DNA sequences. They can be transferred into organisms in the form of plasmids or in the appropriate format, by using a viral vector. The genetically modified organisms produced can be used to produce products such as recombinant proteins, used in medical research, or be grown in agriculture.

Restriction enzymes are used in “genetic fingerprinting” too.
Forensic scientists can use DNA in blood, semen, skin, saliva or hair found at a crime scene to identify a matching DNA of an individual, such as a perpetrator. This process is formally termed DNA profiling, but may also be called "genetic fingerprinting".
In DNA profiling, the lengths of variable sections of repetitive DNA are compared between people. This method is usually an extremely reliable technique for identifying a matching DNA.

Ligases

Enzymes called DNA ligases can rejoin cut or broken DNA strands. Ligases are particularly important in lagging strand DNA replication, as they join together the short segments of DNA produced at the replication fork into a complete copy of the DNA template.
They are also used in DNA repair and genetic recombination.

Topoisomerases

Topoisomerases are enzymes with both nuclease and ligase activity. These proteins change the amount of supercoiling in DNA.
Some of these enzymes work by cutting the DNA helix and allowing one section to rotate, thereby reducing its level of supercoiling; the enzyme then seals the DNA break.
Other types of these enzymes can cut one DNA helix and then pass a second strand of DNA through this break, before rejoining the helix. Topoisomerases are required for DNA replication and transcription.

Helicases

Helicases are proteins that are a type of molecular motor. They use the chemical energy in nucleoside triphosphates (ATP mainly) to break hydrogen bonds between bases and unwind the DNA double helix into single strands.
These enzymes are essential for most processes where enzymes need to access the DNA bases (replication, transcription, translation, recombination, DNA repair ecc).
It's not surprising then that a great variety of helicases is present in the cell nuclei. In humans 95 helicases are coded for in the genome (approximately 1% of eukaryotic genes code for helicases)

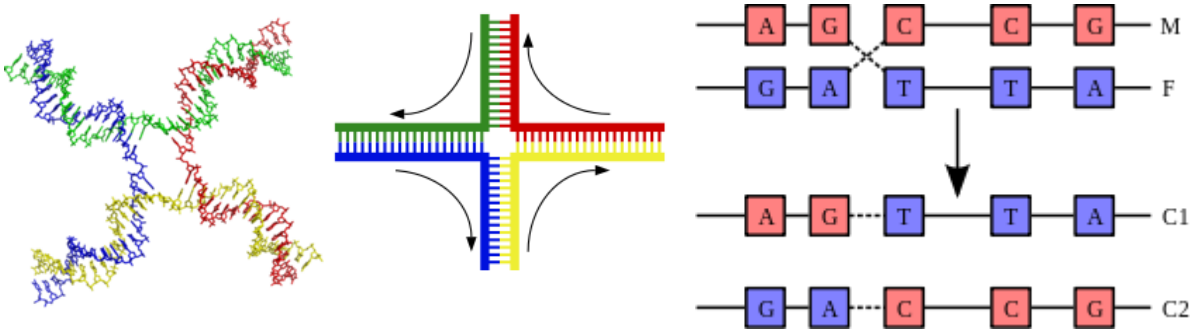
Polymerases

Polymerases are enzymes that synthesize polynucleotide chains from nucleoside triphosphates. The sequence of their products are synthesized using existing polynucleotide chains as mould (*templates)*. These enzymes function by repeatedly adding a nucleotide to the 3' hydroxyl group at the end of the growing polynucleotide chain. As a consequence, all polymerases work in a 5' to 3' direction.
In the active site of these enzymes, the incoming nucleoside triphosphate base-pairs to the template: this allows polymerases to accurately synthesize the complementary strand of their template.
There are different polymerases:

1. In DNA replication, DNA-dependent DNA polymerases make copies of DNA polynucleotide chains. In order to preserve biological information, it is essential that the sequence of bases in each copy are precisely complementary to the sequence of bases in the template strand. Many DNA polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched nucleotides. If a mismatch is detected, a 3' to 5' exonuclease activity is activated and the incorrect base removed. In most organisms, DNA polymerases function in a large complex called the replisome that contains multiple accessory subunits, such as the DNA clamp or helicases.
2. RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is required for the replication of telomeres. Telomerase is an unusual polymerase because it contains its own RNA template as part of its structure.
3. Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript until it reaches a region of DNA called the terminator, where it halts and detaches from the DNA. As with human DNA-dependent

DNA polymerases, RNA polymerase II, the enzyme that transcribes most of the genes in the human genome, operates as part of a large protein complex with multiple regulatory and accessory subunits.

Genetic recombination



Recombination involves the breakage and rejoining of two chromosomes (M and F) to produce two re-arranged chromosomes (C1 and C2).

A DNA helix usually does not interact with other segments of DNA, and in human cells the different chromosomes even occupy separate areas in the nucleus called "chromosome territories". This physical separation of different chromosomes is important for the ability of DNA to function as a stable repository for information.

One of the few times chromosomes interact is during prophase of meiosis. Duplicated homologous chromosomes pairs align end-to-end at that moment and chromosomal crossover can take place. Chromosomal crossover is when two DNA helices break, swap a section and then rejoin.

Recombination allows chromosomes to exchange genetic information and produces new combinations of genes, which increases the efficiency of natural selection and can be important in the rapid evolution of new proteins.

Genetic recombination can also be involved in DNA repair, particularly in the cell's response to double-strand breaks.

The most common form of chromosomal crossover is homologous recombination, where the two chromosomes involved share very similar sequences. Non-homologous recombination can be damaging to cells, as it can produce chromosomal translocations and genetic abnormalities.

The recombination reaction is catalyzed by enzymes known as recombinases, such as RAD51. The first step in recombination is a double-stranded break caused by either an endonuclease or damage to the DNA. A series of steps catalyzed in part by the recombinase then leads to the joining of the two helices by at least one junction (“Holliday junction”), in which a segment of a single strand in each helix is annealed to the complementary strand in the other helix. The Holliday junction is a tetrahedral junction structure that can be moved along the pair of chromosomes, swapping one strand for another. The recombination reaction is then halted by cleavage of the junction and re-ligation of the released DNA.

Evolution

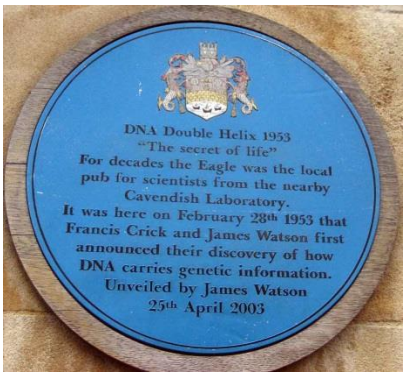
DNA contains the genetic information that allows all modern living things to function, grow and reproduce. However, it is unclear how long in the 4-billion-year history of life DNA has performed this function, as it has been proposed that the earliest forms of life may have used RNA as their genetic material.

RNA may have acted as the central part of early cell metabolism as it can both transmit genetic information and carry out catalysis as part of ribozymes. This ancient RNA world where nucleic acid would have been used for both catalysis and genetics may have influenced the evolution of the current genetic code based on four nucleotide bases. This would occur, since the number of different bases in such an organism is a trade-off between a small number of bases increasing replication accuracy and a large number of bases increasing the catalytic efficiency of ribozymes. However, there is no direct evidence of ancient genetic systems, as recovery of DNA from most fossils is impossible. This is because DNA survives in the environment for less than one million years, and slowly degrades into short fragments in solution.

Claims for older DNA have been made, most notably a report of the isolation of a viable bacterium from a salt crystal 250 million years old, but these claims are controversial.

On 8 August 2011, a report, based on NASA studies with meteorites found on Earth, was published suggesting building blocks of DNA (adenine, guanine and related organic molecules) may have been formed extraterrestrially in outer space.

History of DNA research



1869 - DNA was first isolated by the Swiss physician Friedrich Miescher who, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein".

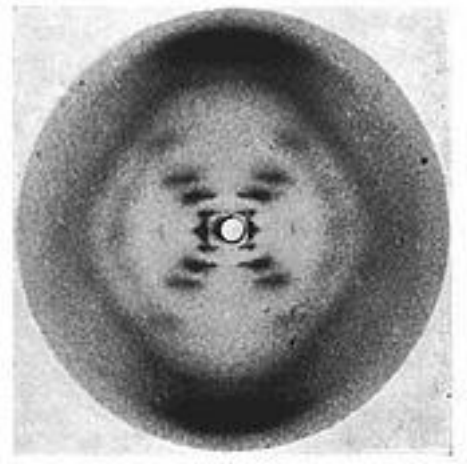
1878 - Albrecht Kossel isolated the non-protein component of "nuclein", nucleic acid, and later isolated its five primary nucleobases.

1919 - Phoebus Levene identified the base, sugar and phosphate nucleotide unit. Levene suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups. Levene thought the chain was short and the bases repeated in a fixed order.

1927 - Nikolai Koltsov proposed that inherited traits would be inherited via a "giant hereditary molecule" made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template".

1928 - Frederick Griffith in his experiment discovered that traits of the "smooth" form of *Pneumococcus* could be transferred to the "rough" form of the same bacteria by mixing killed "smooth" bacteria with the live "rough" form.

1937 - William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure.
1943 - The Avery–MacLeod–McCarty experiment provided the first clear suggestion that DNA carries genetic information. Identified DNA as the transforming principle.
1952 - Alfred Hershey and Martha Chase in the Hershey–Chase experiment showed that DNA is the genetic material of the T2 phage.



1953 - James Watson and Francis Crick suggested what is now accepted as the first correct double-helix model of DNA structure in the journal *Nature*. Their double-helix, molecular model of DNA was then based on a single X-ray diffraction image (labelled as "Photo 51") taken by Rosalind Franklin and Raymond Gosling in May 1952, as well as the information that the DNA bases are paired — also obtained through private communications from Erwin Chargaff in the previous years. In the same year, Maurice Wilkins and two of his colleagues, with X-ray analysis *in vivo* B-DNA also supported the presence *in vivo* of the double-helical DNA configurations.
1958 - Final confirmation of the replication mechanism that was implied by the double-helical structure followed through the Meselson–Stahl experiment.
1962 - After Franklin's death, Watson, Crick, and Wilkins jointly received the Nobel Prize in Physiology or Medicine. Nobel Prizes were awarded only to living recipients at the time.
1968 - Har Gobind Khorana, Robert W. Holley and Marshall Warren Nirenberg received the Nobel prize for deciphering the genetic code. These findings represent the birth of molecular biology.

Genetic Disorder

A **genetic disorder** is an illness caused by one or more abnormalities in the genome, especially a condition that is present from birth (congenital).
Most genetic disorders are quite rare and affect one person in every several thousands or millions.
Genetic disorders may or may not be heritable, i.e., passed down from the parents' genes.
In non-heritable genetic disorders, defects may be caused by new mutations or changes to the DNA. In such cases, the defect will only be heritable if it occurs in the germ line.
The same disease, such as some forms of cancer, may be caused by an inherited genetic condition in some people, by new mutations in other people, and mainly by environmental causes in still other people. Whether, when and to what extent a person with the genetic defect or abnormality will actually suffer from the disease is almost always affected by environmental factors and events in the person's development.
Some types of recessive gene disorders confer an advantage in certain environments when only one copy of the gene is present.

Single gene disorder

A **single gene disorder** is the result of a single mutated gene. Over 4000 human diseases are caused by single gene defects.
Single gene disorders can be passed on to subsequent generations in several ways. Genomic imprinting and uniparental disomy, however, may affect inheritance patterns. The divisions between recessive and dominant types are not "hard and fast", although the divisions between autosomal and X-linked types are (since the latter types are distinguished purely based on the chromosomal location of the gene).
For example, Sickle-cell anemia is also considered a recessive condition, but heterozygous carriers have increased resistance to malaria in early childhood, which could be described as a related dominant condition.

(I seguenti paragrafi sono Facoltativi)

Autosomal dominant

Only one mutated copy of the gene will be necessary for a person to be affected by an autosomal dominant disorder. Each affected person usually has one affected parent. The chance a child will inherit the mutated gene is 50%. Autosomal dominant conditions sometimes have reduced penetrance, which means although only one mutated copy is needed, not all individuals who inherit that mutation go on to develop the disease. Examples of this type of disorder are Huntington's disease, neurofibromatosis, Marfan syndrome, hereditary nonpolyposis colorectal cancer. Birth defects are also called congenital anomalies.

Autosomal recessive

Two copies of the gene must be mutated for a person to be affected by an autosomal recessive disorder. An affected person usually has unaffected parents who each carry a single copy of the mutated gene (and are referred to as carriers).
Two unaffected people who each carry one copy of the mutated gene have a 25% risk with each pregnancy of having a child affected by the disorder. Examples of this type of disorder are Medium-chain acyl-CoA dehydrogenase deficiency, cystic fibrosis, sickle-cell disease, Tay-Sachs disease, Niemann-Pick disease, spinal muscular atrophy, and Roberts syndrome.

X-linked dominant

X-linked dominant disorders are caused by mutations in genes on the X chromosome. Only a few disorders have this inheritance pattern, with a prime example being X-linked hypophosphatemic rickets. Males and females are both affected in these disorders, with males typically being more severely affected than females. Some X-linked dominant conditions, such as Rett syndrome, incontinentia pigmenti type 2, and Aicardi syndrome, are usually fatal in males either *in utero* or shortly after birth, and are therefore predominantly seen in females.
The chance of passing on an X-linked dominant disorder differs between men and women. The sons of a man with an X-linked dominant disorder will all be unaffected (since they receive their father's Y chromosome), and his daughters will all inherit the condition. A woman with an X-linked dominant disorder has a 50% chance of having an affected fetus with each pregnancy. In addition, although these conditions do not alter fertility *per se*, individuals with Rett syndrome or Aicardi syndrome rarely reproduce.

X-linked recessive

X-linked recessive conditions are also caused by mutations in genes on the X chromosome. Males are more frequently affected than females, and the chance of passing on the disorder differs

between men and women. The sons of a man with an X-linked recessive disorder will not be affected, and his daughters will carry one copy of the mutated gene. A woman who is a carrier of an X-linked recessive disorder (X^RX^r) has a 50% chance of having sons who are affected and a 50% chance of having daughters who carry one copy of the mutated gene and are therefore carriers. X-linked recessive conditions include the serious diseases hemophilia A, Duchenne muscular dystrophy, and Lesch-Nyhan syndrome, as well as common and less serious conditions such as male pattern baldness and red-green color blindness.

Y-linked

Y-linked disorders, also called holandric disorders, are caused by mutations on the Y chromosome. Y-linked disorders in humans can only be passed from men to their sons; females can never be affected because they do not possess Y allosomes.

Y-linked disorders are exceedingly rare but the most well-known examples typically cause infertility. Reproduction in such conditions is only possible through the circumvention of infertility by medical intervention.