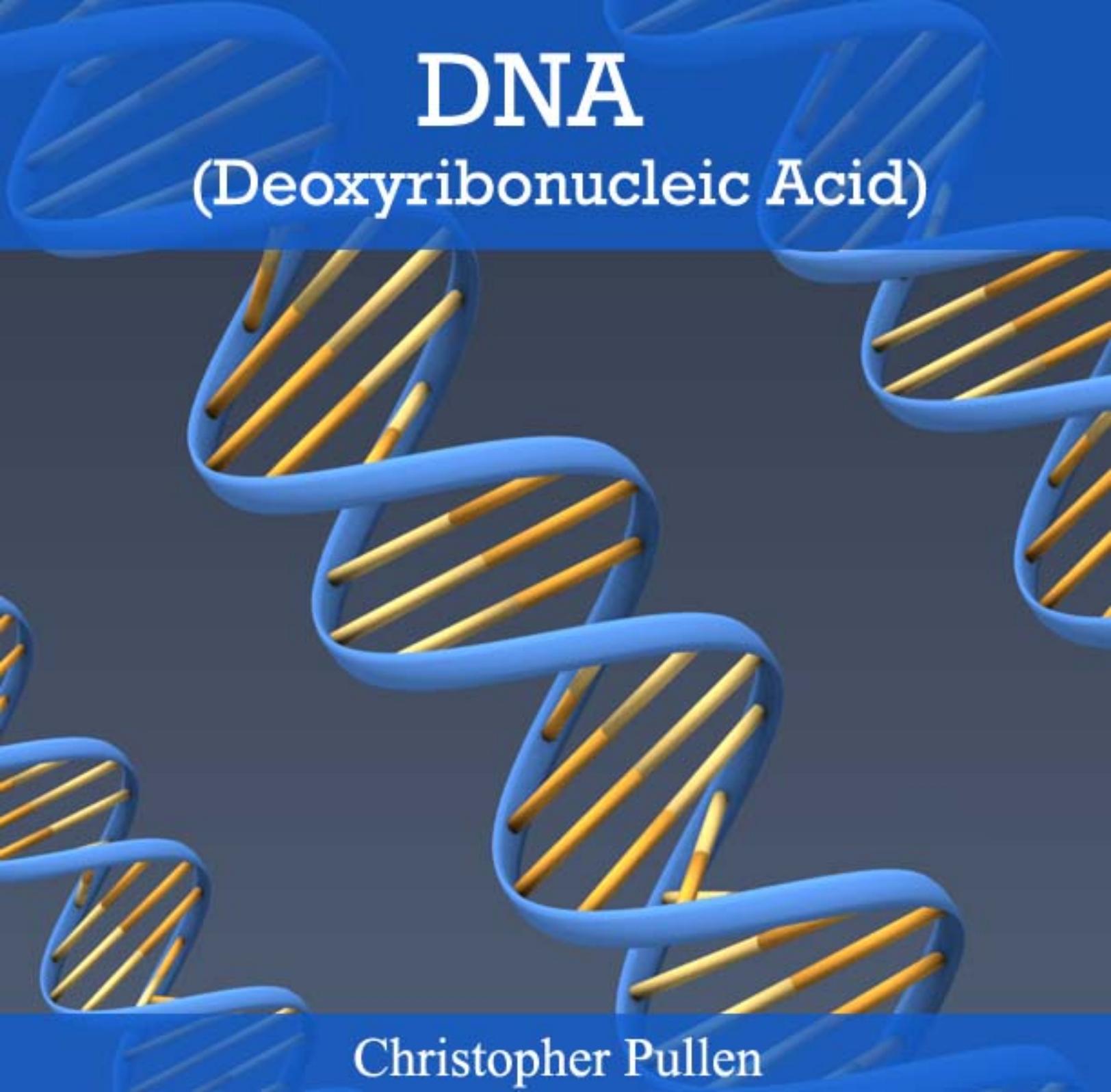


DNA

(Deoxyribonucleic Acid)



Christopher Pullen

First Edition, 2012

ISBN 978-81-323-3273-2

© All rights reserved.

Published by:

Research World

4735/22 Prakashdeep Bldg,

Ansari Road, Darya Ganj,

Delhi - 110002

Email: info@wtbooks.com

Table of Contents

Chapter 1 - DNA

Chapter 2 - Base Pair

Chapter 3 - Sense (Molecular Biology)

Chapter 4 - Molecular Models of DNA

Chapter 5 - DNA Microarray

Chapter 6 - DNA Nanotechnology

Chapter 7 - Chromatin

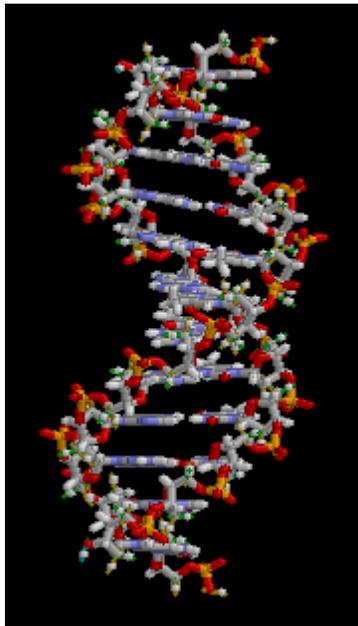
Chapter 8 - DNA Replication

Chapter 9 - DNA-Binding Protein and Genetic Recombination

Chapter 10 - DNA Repair

Chapter- 1

DNA



The structure of part of a DNA double helix

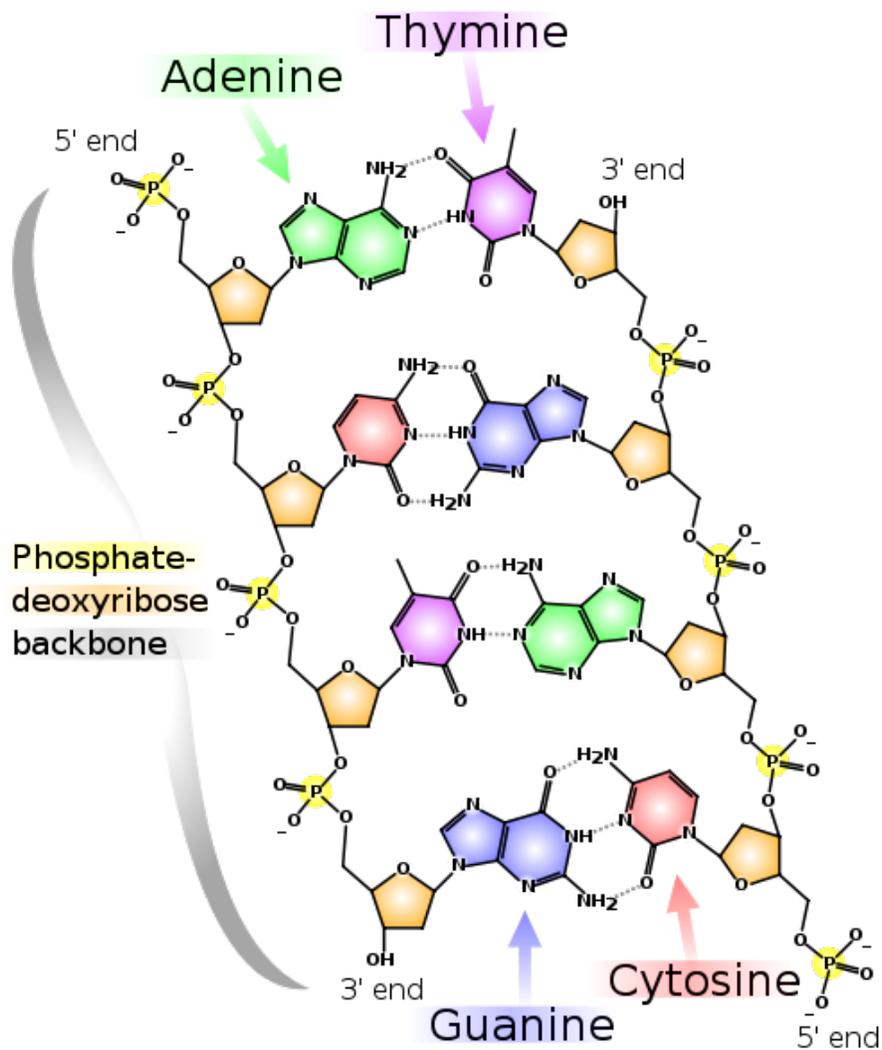
Deoxyribonucleic acid or **DNA** is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms (with the exception of RNA viruses). The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints, like a recipe or a code, since it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules. The DNA segments that carry this genetic information are called genes, but other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information.

DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases. It is the sequence of these four bases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by

copying stretches of DNA into the related nucleic acid RNA, in a process called transcription.

Within cells, DNA is organized into long structures called chromosomes. These chromosomes are duplicated before cells divide, in a process called DNA replication. 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.

Properties

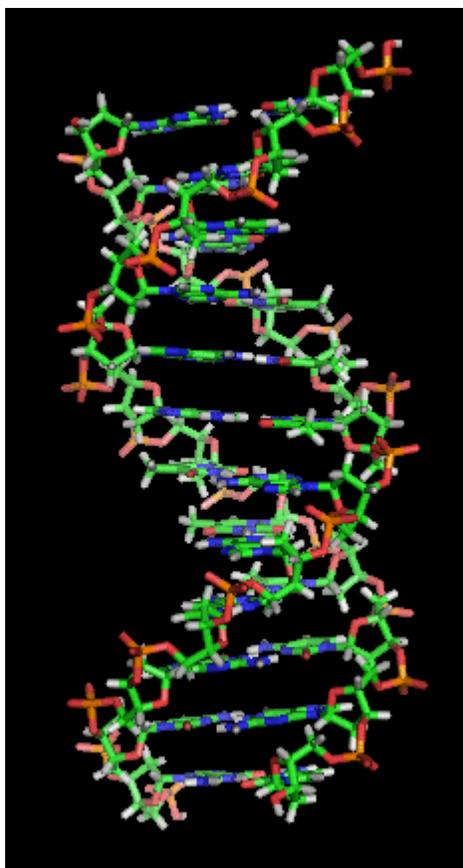


Chemical structure of DNA. Hydrogen bonds shown as dotted lines.

DNA is a long polymer made from repeating units called nucleotides. As first discovered by James D. Watson and Francis Crick, the structure of DNA of all species comprises two helical chains each coiled round the same axis, and each with a pitch of 34 Ångströms (3.4 nanometres) and a radius of 10 Ångströms (1.0 nanometres). According to another study, when measured in a particular solution, the DNA chain measured 22 to 26 Ångströms wide (2.2 to 2.6 nanometres), and one nucleotide unit measured 3.3 Å (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, is approximately 220 million base pairs long.

In living organisms, DNA does not usually exist as a single molecule, but instead as a pair of molecules that are held tightly together. These two long strands entwine like vines, in the shape of a double helix. The nucleotide repeats contain both the segment of the backbone of the molecule, which holds the chain together, and a base, which interacts with the other DNA strand in the helix. A base 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. If multiple nucleotides are linked together, as in DNA, this polymer is called a polynucleotide.

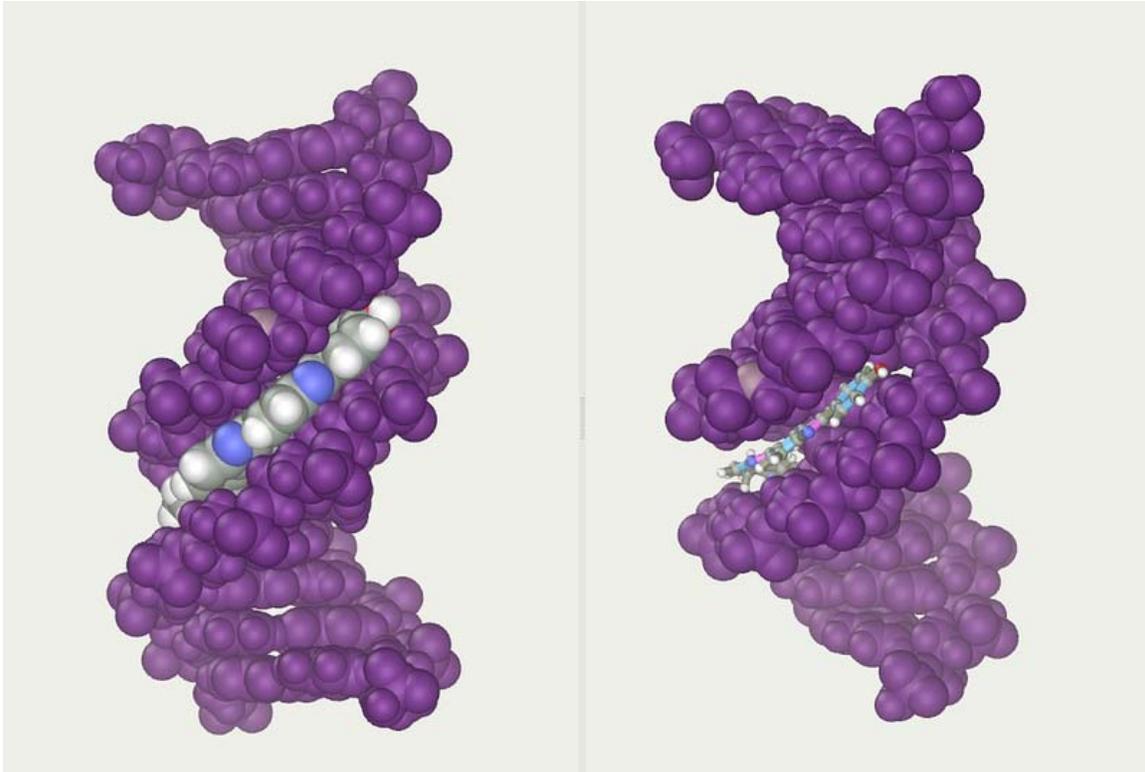
The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar 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 by hydrogen bonds between the bases attached to the two strands. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate.

These bases are classified into two types; adenine and guanine are fused five- and six-membered heterocyclic compounds called purines, while cytosine and thymine are six-membered rings called pyrimidines. A fifth pyrimidine base, called 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. In addition to RNA and DNA, a large number of artificial nucleic acid analogues have also been created to study the properties of nucleic acids, or for use in biotechnology.



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

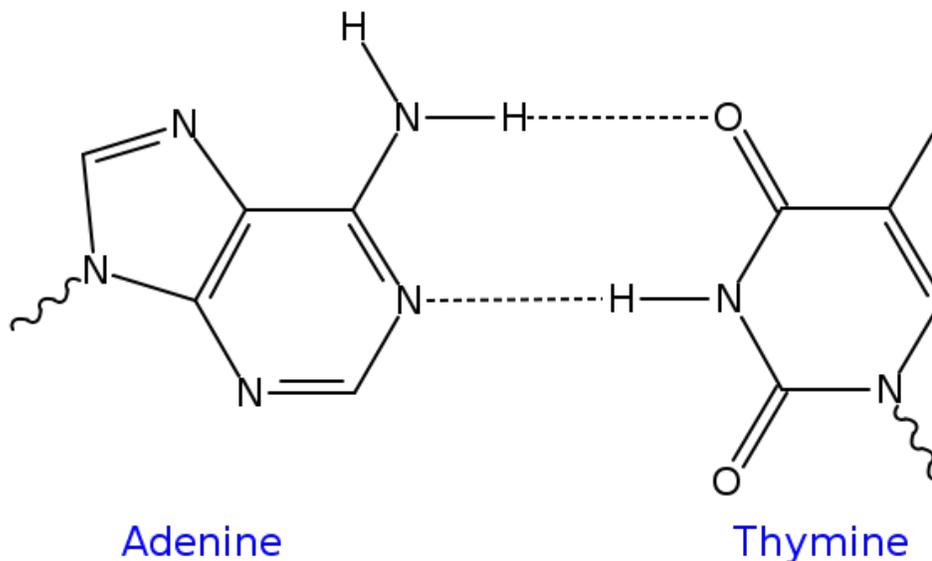
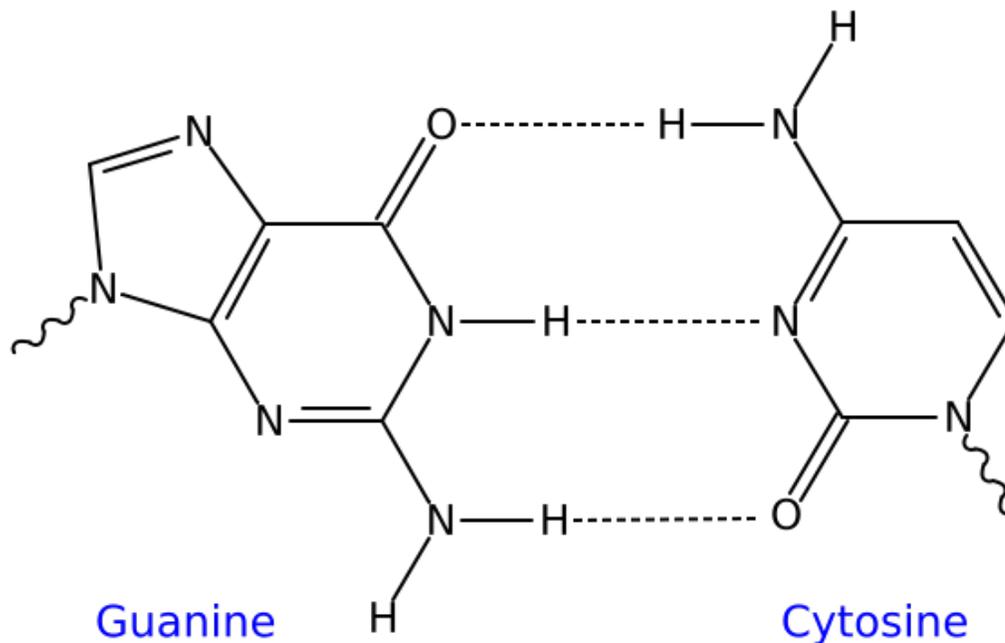
Grooves

Twin helical strands form the DNA backbone. Another double helix may be found by 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 directly opposite each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å 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.

Base pairing

Each type of base on one strand forms a bond with just one type of base on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with A bonding only to T, and C bonding only to G. This arrangement of two nucleotides binding together across the double helix is called a base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. 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. 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. Indeed, this reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.



Top, a **GC** base pair with three hydrogen bonds. Bottom, an **AT** base pair with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines.

The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds. DNA with high GC-content is more stable than DNA with low GC-content, but contrary to popular belief, this is not due to the extra hydrogen bond of a GC base pair but rather the contribution of stacking interactions (hydrogen bonding merely provides specificity of the pairing, not stability).

As a result, it is both the percentage of GC base pairs and the overall length of a DNA double helix that determine the strength of the association between the two strands of DNA. Long DNA helices with a high GC content have stronger-interacting strands, while short helices with high AT content have weaker-interacting strands. In biology, parts of the DNA double helix that need to separate easily, such as the TATAAT Pribnow box in some promoters, tend to have a high AT content, making the strands easier to pull apart.

In the laboratory, the strength of this interaction can be measured by finding the temperature required to break the hydrogen bonds, their melting temperature (also called T_m value). When all the base pairs in a DNA double helix melt, the strands separate and exist in solution as two entirely independent molecules. These single-stranded DNA molecules (*ssDNA*) have no single common shape, but some conformations are more stable than others.

Sense and antisense

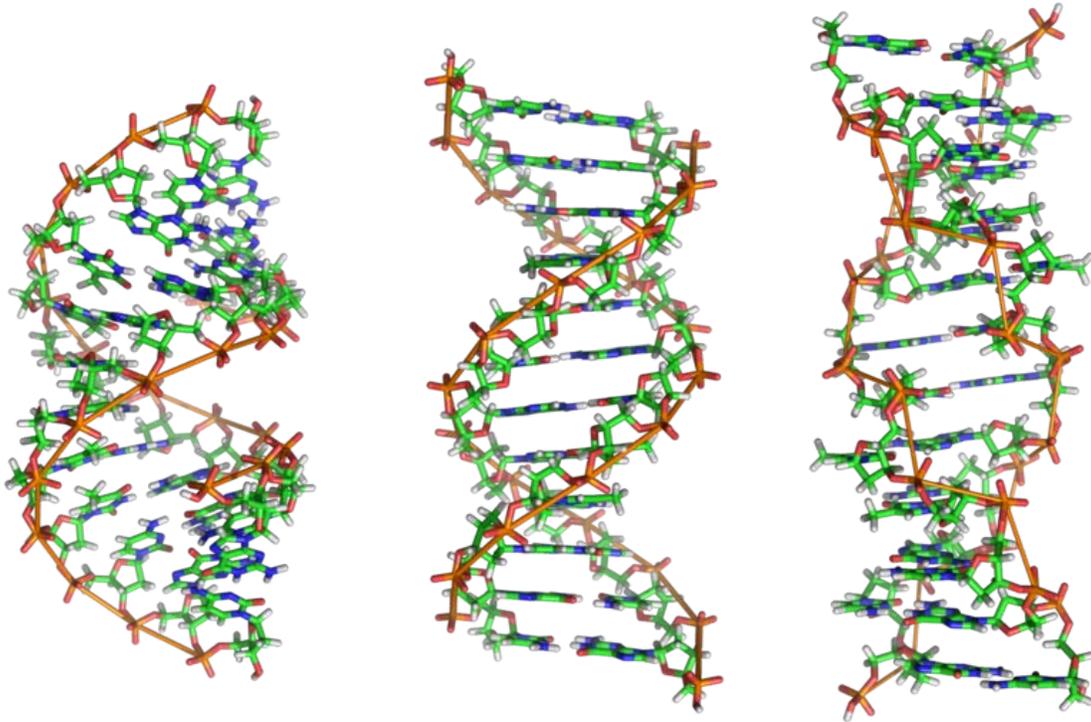
A DNA sequence is called "sense" if its sequence is the same as that of a messenger RNA copy that is translated into protein. The sequence on the opposite strand is called the "antisense" sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands 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

Alternate DNA structures

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.

The first published reports of A-DNA X-ray diffraction patterns— and also B-DNA used analyses based on Patterson transforms that provided only a limited amount of structural information for oriented fibers of DNA. An alternate analysis was then proposed by Wilkins *et al.*, in 1953, for the *in vivo* B-DNA X-ray diffraction/scattering patterns of highly hydrated DNA fibers in terms of squares of Bessel functions. In the same journal, James D. Watson and Francis Crick presented their molecular modeling analysis of the DNA X-ray diffraction patterns to suggest that the structure was a double-helix.

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.

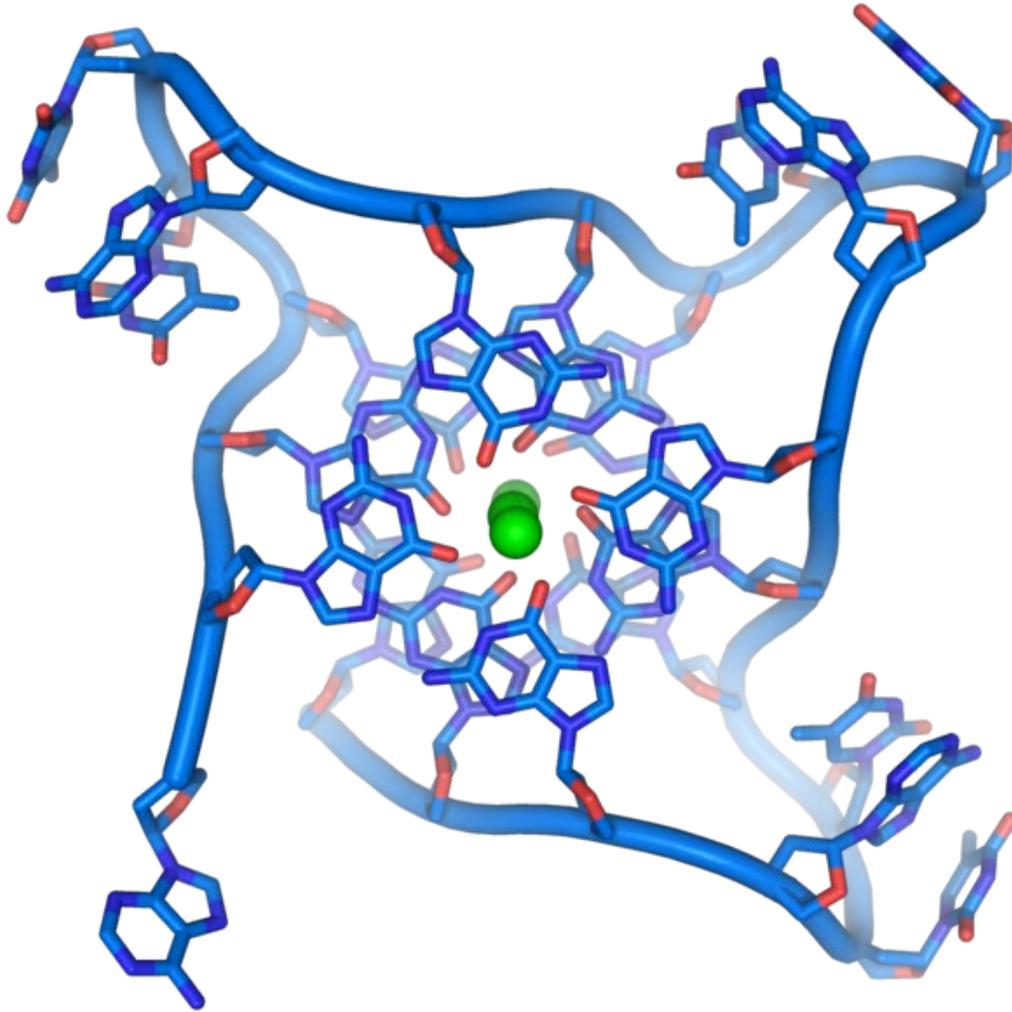
Alternate DNA chemistry

For a number of years exobiologists have proposed the existence of a shadow biosphere, a postulated microbial biosphere of Earth that uses radically different biochemical and molecular processes than currently known life. One of the proposals was the existence of lifeforms that use arsenic instead of phosphorus in DNA.

A December 2010 NASA press conference revealed that the bacterium GFAJ-1, which has evolved in an arsenic-rich environment, is the first terrestrial lifeform found which may have this ability. The bacterium was found in Mono Lake, east of Yosemite National Park. GFAJ-1 is a rod-shaped extremophile bacterium in the family Halomonadaceae that, when starved of phosphorus, may be capable of incorporating the usually poisonous element arsenic in its DNA. This discovery lends weight to the long-standing idea that extraterrestrial life could have a different chemical makeup from life on Earth. The research was carried out by a team led by Felisa Wolfe-Simon, a geomicrobiologist and geobiochemist, a Postdoctoral Fellow of the NASA Astrobiology Institute with Arizona State University.

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.

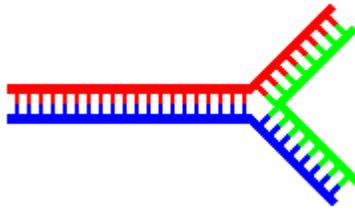


DNA quadruplex formed by telomere repeats. The looped conformation of the DNA backbone is very different from the typical DNA helix.

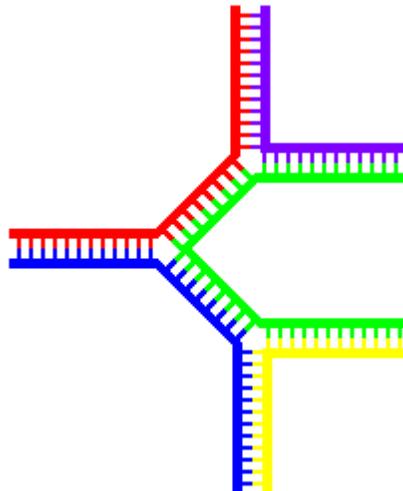
These guanine-rich sequences may stabilize chromosome ends by forming structures of stacked sets of four-base units, rather than the usual base pairs found in other DNA molecules. Here, four guanine bases form a flat plate and these flat four-base units then stack on top of each other, to form a stable *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 structures can also be formed, with the central set of four bases coming from either a single strand folded around the bases, or several different parallel strands, each contributing one base to the central structure.

In addition to these stacked structures, telomeres also form large loop structures called telomere loops, or T-loops. Here, the single-stranded DNA curls around in a long circle stabilized by telomere-binding proteins. At the very end of the T-loop, the single-stranded telomere DNA is held onto a region of double-stranded DNA by the telomere strand

disrupting the double-helical DNA and base pairing to one of the two strands. This triple-stranded structure is called a displacement loop or D-loop.



Single branch



Multiple branches

Branched DNA can form networks containing multiple branches.

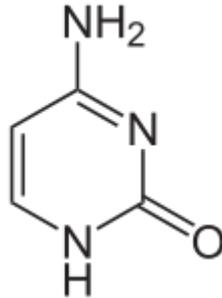
Branched DNA

In DNA fraying occurs when non-complementary regions exist at the end of an otherwise complementary double-strand of DNA. However, branched DNA can occur if a third strand of DNA is introduced and contains adjoining regions able to hybridize with the frayed regions of the pre-existing double-strand. Although the simplest example of branched DNA involves only three strands of DNA, complexes involving additional strands and multiple branches are also possible. Branched DNA can be used in nanotechnology to construct geometric shapes, see the section on uses in technology below.

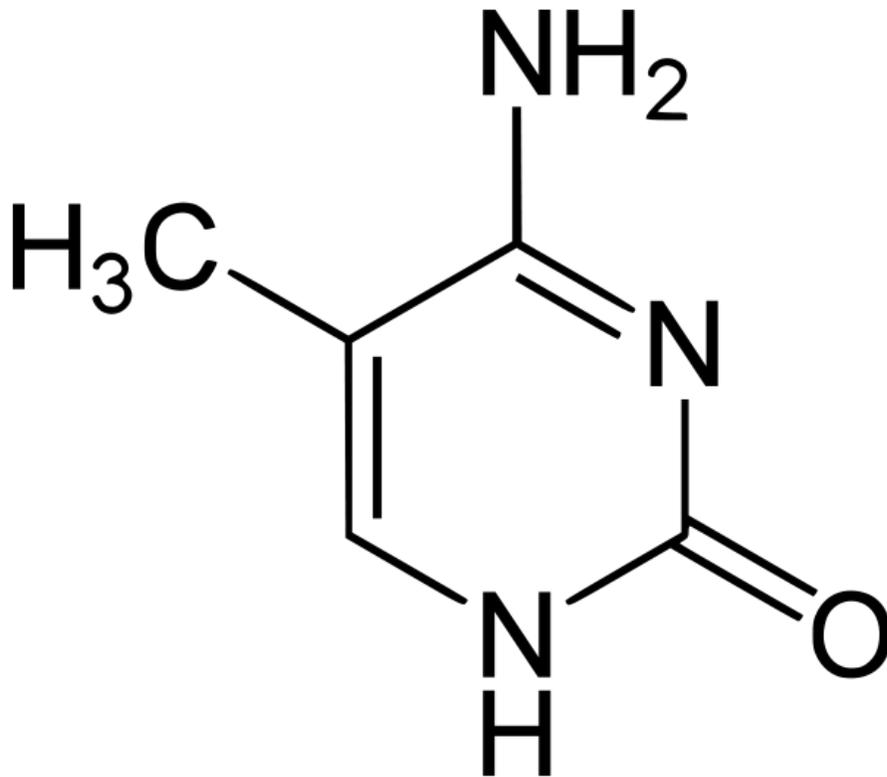
Vibration

DNA may carry out low-frequency collective motion as observed by the Raman spectroscopy and analyzed with a quasi-continuum model.

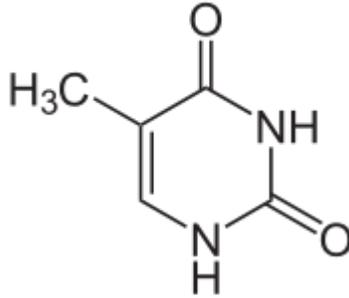
Chemical modifications



cytosine



5-methylcytosine



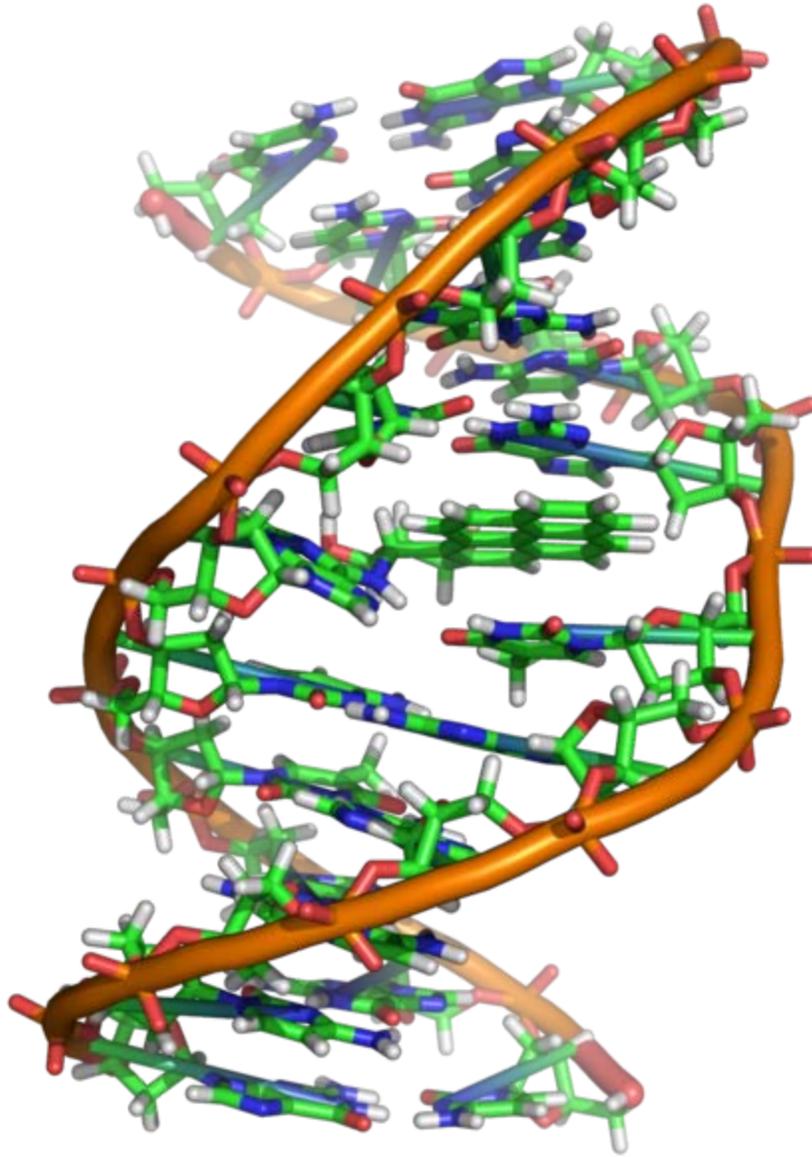
thymine

Structure of cytosine with and without the 5-methyl group. Deamination converts 5-methylcytosine into thymine.

Base modifications

The expression of genes is influenced by how the DNA is packaged in chromosomes, in a 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. For 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 kinetoplasts.

Damage



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

DNA can be damaged by many sorts of mutagens, which change the DNA sequence. Mutagens include oxidizing agents, alkylating agents and also high-energy electromagnetic radiation such as ultraviolet light and X-rays. The type of DNA damage produced depends on the type of mutagen. For example, 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 guanosine, 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 and deletions from the DNA sequence, as well as chromosomal translocations.

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, daunomycin, and doxorubicin. In order 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 are often carcinogens, and benzo[*a*]pyrene diol epoxide, acridines, aflatoxin and ethidium bromide are well-known examples. 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.

Biological functions

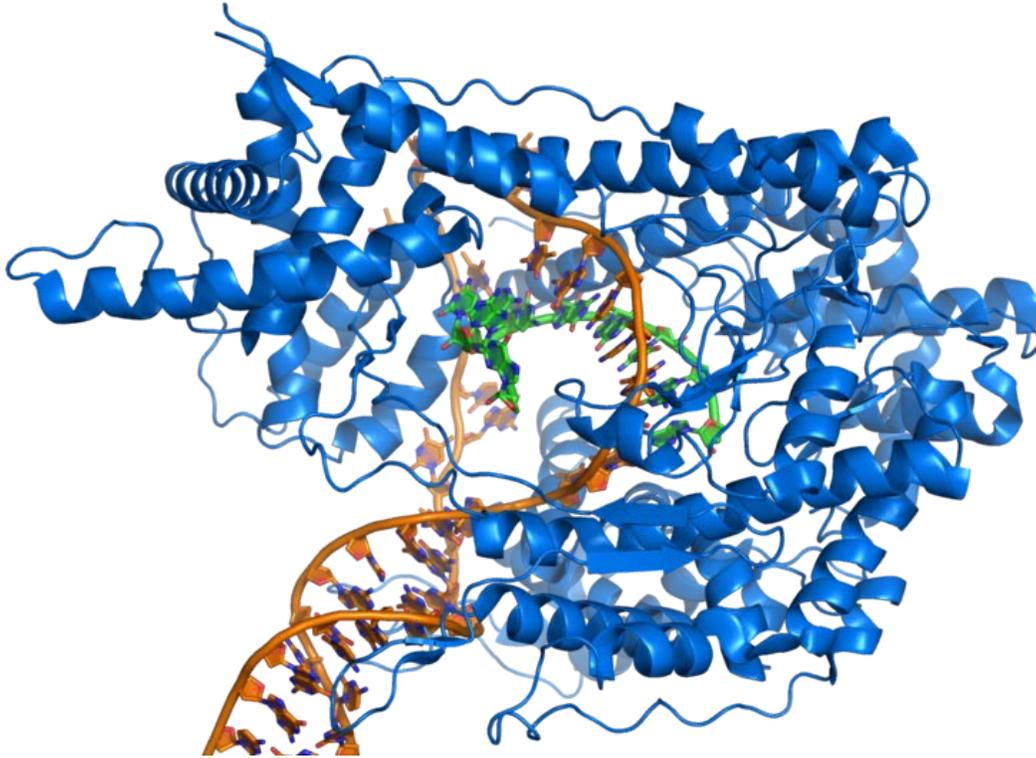
DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes. The information carried by DNA is held in the sequence of pieces of DNA called genes. Transmission of genetic information in genes is achieved via complementary base pairing. For example, 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. The details of these functions are covered in other articles; here we focus on the interactions between DNA and other molecules that mediate the function of the genome.

Genes and genomes

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, DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression.



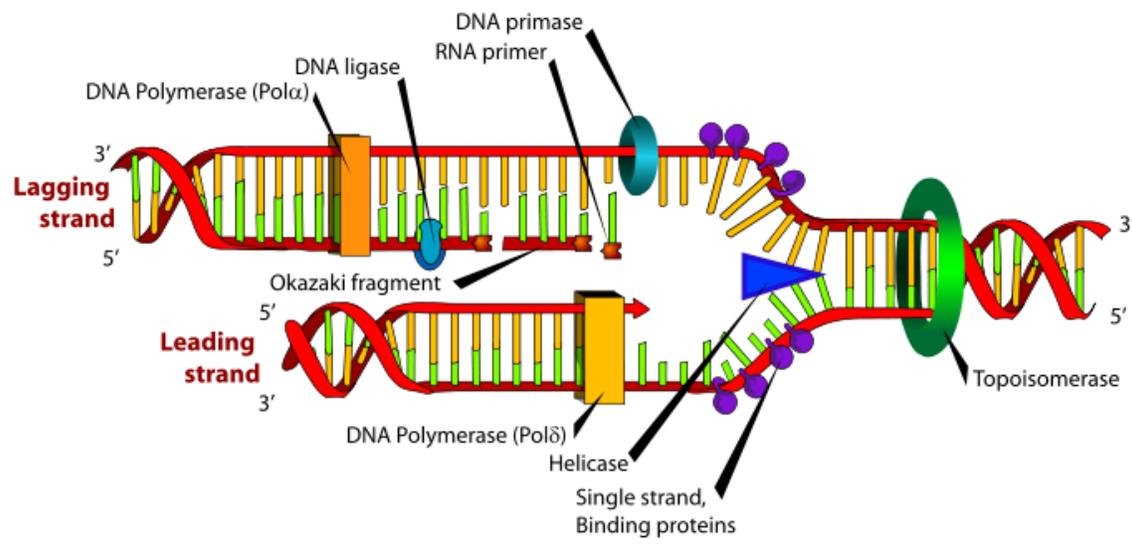
T7 RNA polymerase (blue) producing a mRNA (green) from a DNA template (orange).

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.

Transcription and translation

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



DNA replication. The double helix is unwound by a helicase and topoisomerase. 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.

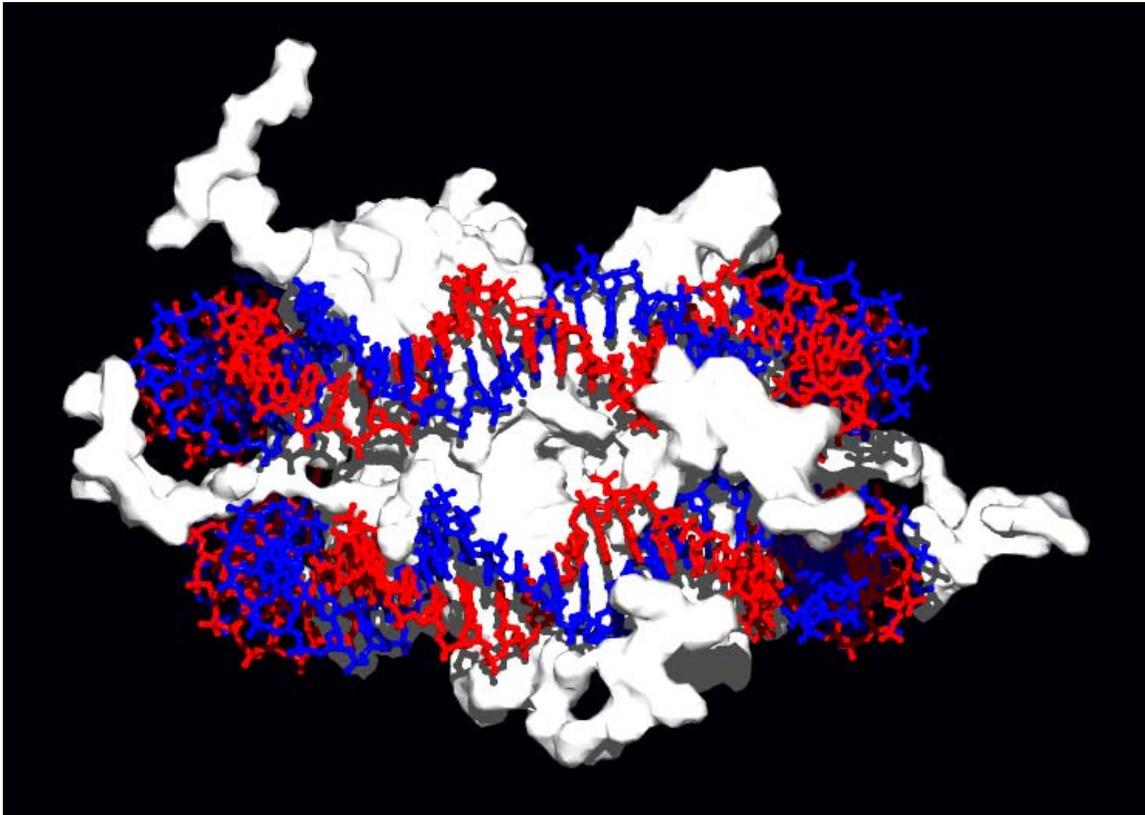
Replication

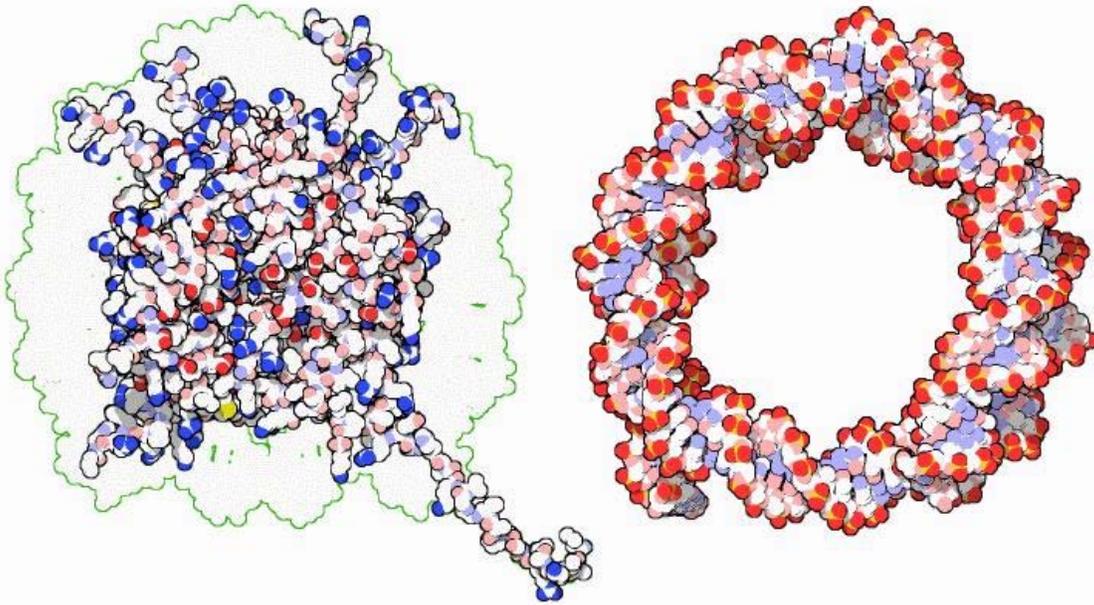
Cell division is essential for an organism to grow, but, when a cell divides, it must replicate the DNA in its genome so that the two daughter cells have the same genetic information as their parent. The double-stranded structure of DNA provides a simple mechanism for DNA replication. Here, the two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA polymerase. This enzyme makes the complementary strand by finding the correct base through complementary base pairing, and bonding it onto the original strand. As DNA polymerases can only extend a DNA strand in a 5' to 3' direction, different mechanisms are used to copy the antiparallel strands of the double helix. In this way, the base on the old strand dictates which base appears on the new strand, and the cell ends up with a perfect copy of its DNA.

Interactions with proteins

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

DNA-binding proteins

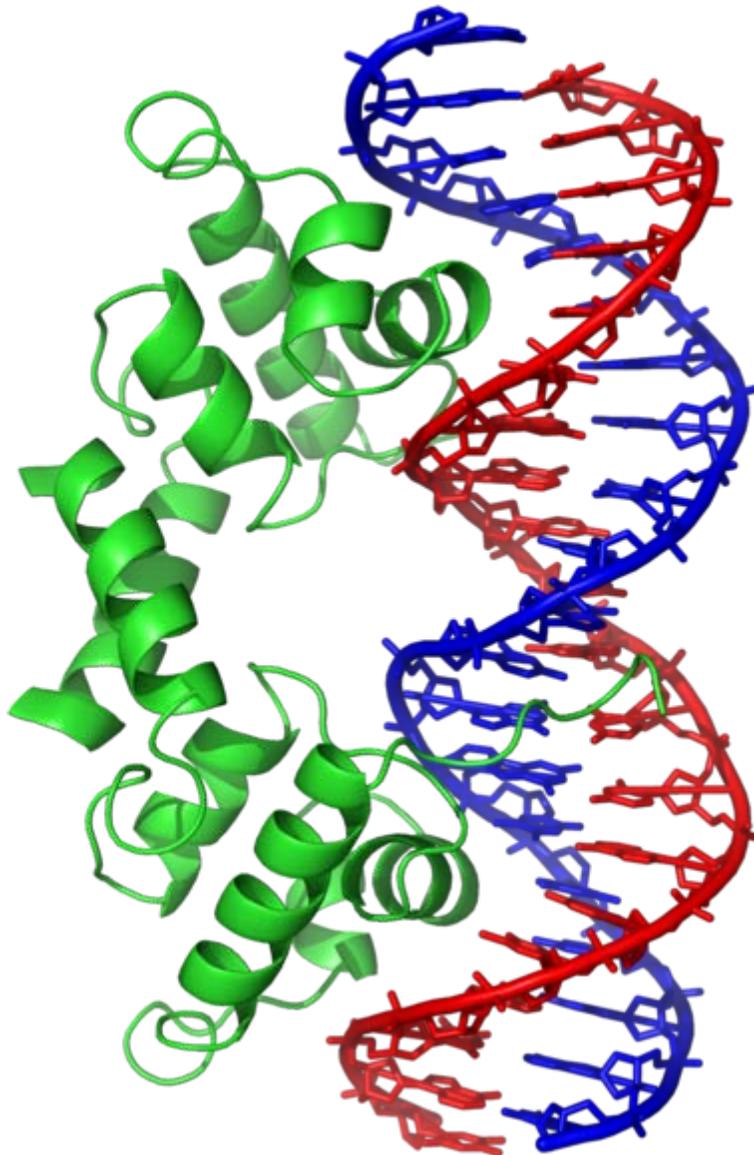




Interaction of DNA with histones (shown in white, top). These proteins' basic amino acids (below left, blue) bind to the acidic phosphate groups on DNA (below right, red).

Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. 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 its surface. These non-specific interactions are formed through basic residues in the histones making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are therefore largely independent of the base sequence. Chemical modifications of these basic amino acid residues include methylation, phosphorylation and acetylation. These chemical changes 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. Other non-specific DNA-binding proteins in chromatin include the high-mobility group proteins, which bind to bent or distorted DNA. These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.

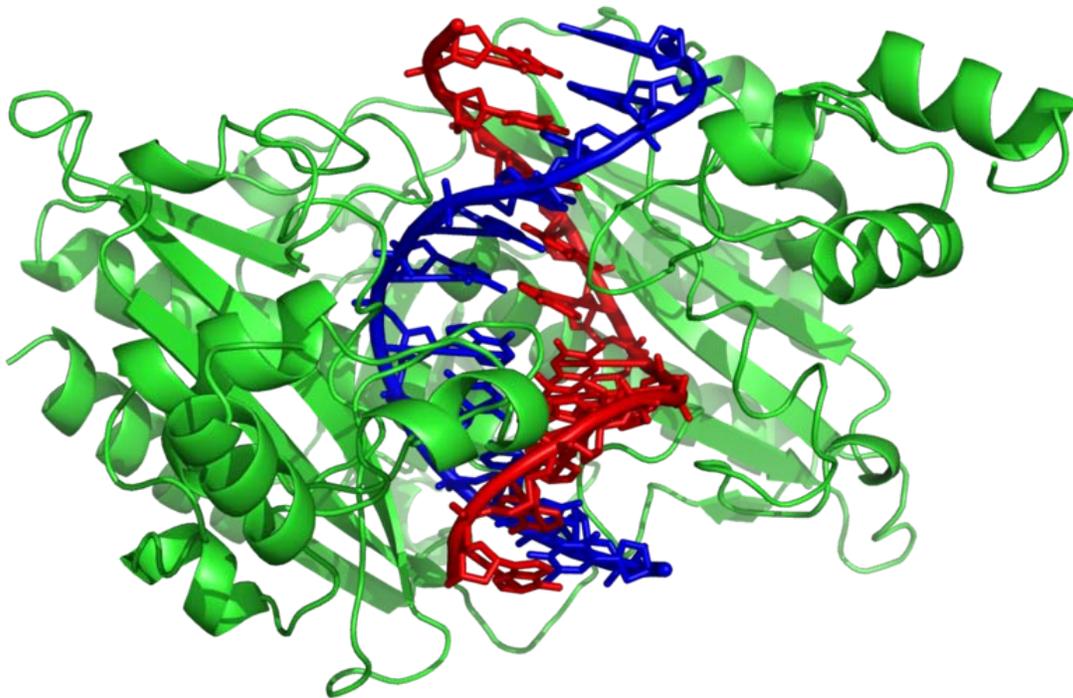
A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA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



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

In contrast, 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. The transcription factors do this in two ways. Firstly, 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. Alternatively, transcription factors can bind enzymes that modify the histones at the promoter; this will change the accessibility of the DNA template to the polymerase.

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.



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

DNA-modifying enzymes

Nucleases and ligases

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 to the left recognizes the 6-base sequence 5'-GAT|ATC-3' and makes a cut at the vertical line. 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 technology, these sequence-specific nucleases are used in molecular cloning and DNA fingerprinting.

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 and helicases

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 are capable of cutting one DNA helix and then passing a second strand of DNA through this break, before rejoining the helix. Topoisomerases are required for many processes involving DNA, such as DNA replication and transcription.

Helicases are proteins that are a type of molecular motor. They use the chemical energy in nucleoside triphosphates, predominantly ATP, 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.

Polymerases

Polymerases are enzymes that synthesize polynucleotide chains from nucleoside triphosphates. The sequence of their products are copies of existing polynucleotide chains - which are called *templates*. These enzymes function by adding nucleotides onto the 3' hydroxyl group of the previous nucleotide in a DNA strand. 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. Polymerases are classified according to the type of template that they use.

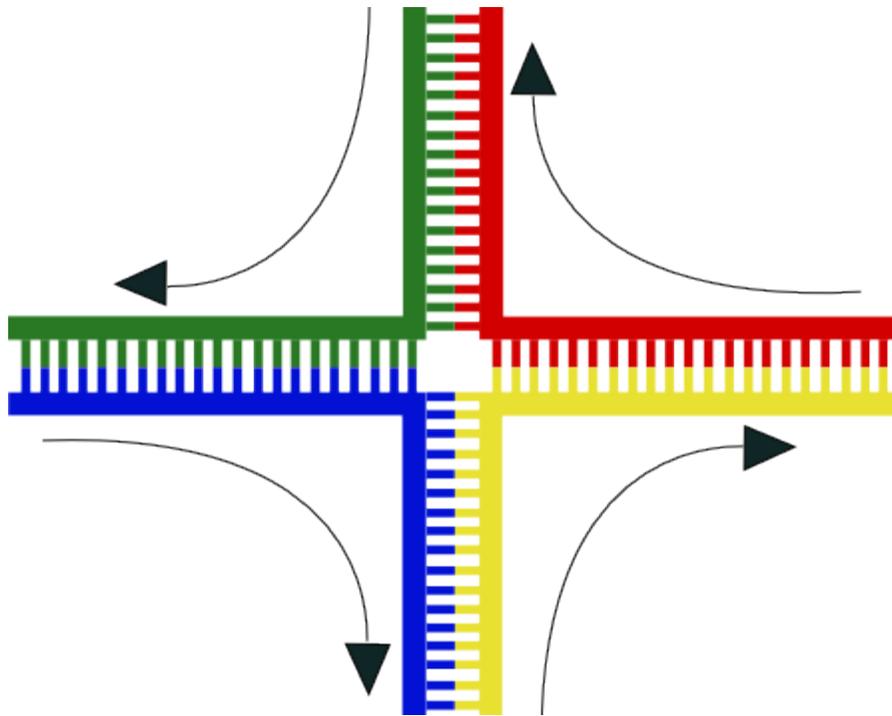
In DNA replication, a DNA-dependent DNA polymerase makes a copy of a DNA sequence. Accuracy is vital in this process, so many of these polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched 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.

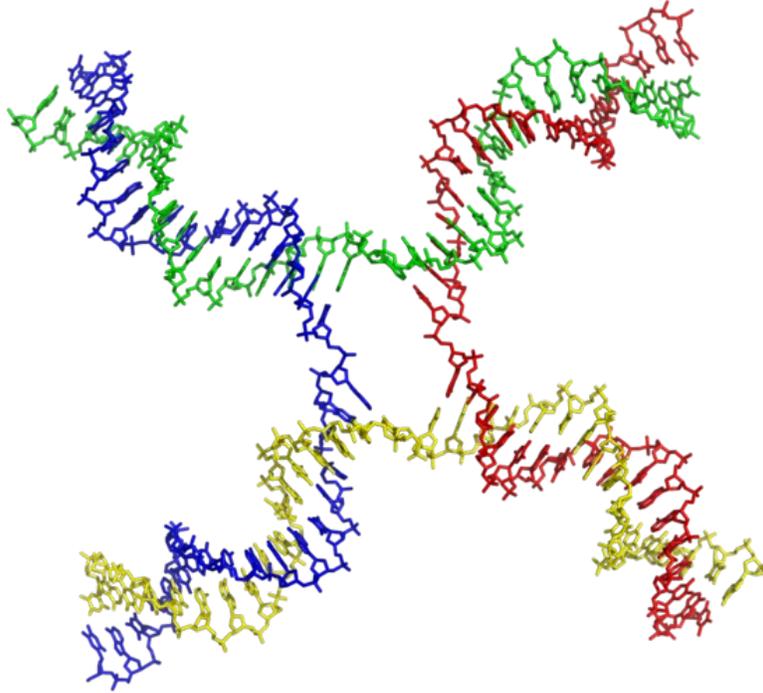
RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is

required for the replication of telomeres. Telomerase is an unusual polymerase because it contains its own RNA template as part of its structure.

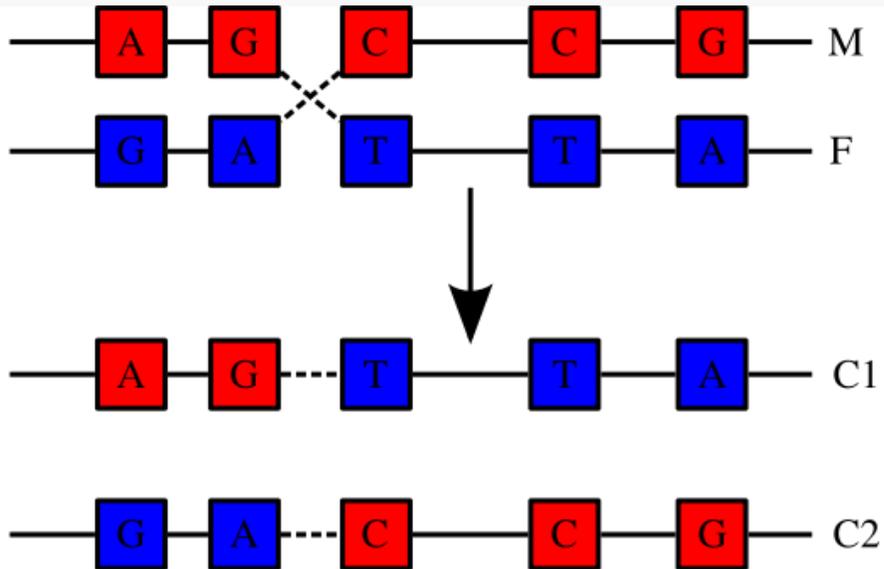
Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript 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





Structure of the Holliday junction intermediate in genetic recombination. The four separate DNA strands are coloured red, blue, green and yellow.



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, as one of the few times chromosomes interact is during chromosomal crossover when they recombine. 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 either caused by an endonuclease or damage to the DNA. A series of steps catalyzed in part by the recombinase then leads to joining of the two helices by at least one 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 will survive 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.

Uses in technology

Genetic engineering

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 transformed 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.

Forensics

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, such as short tandem repeats and minisatellites, are compared between people. This method is usually an extremely reliable technique for identifying a matching DNA. However, identification can be complicated if the scene is contaminated with DNA from several people. DNA profiling was developed in 1984 by British geneticist Sir Alec Jeffreys, and first used in forensic science to convict Colin Pitchfork in the 1988 Enderby murders case.

People convicted of certain types of crimes may be required to provide a sample of DNA for a database. This has helped investigators solve old cases where only a DNA sample was obtained from the scene. DNA profiling can also be used to identify victims of mass casualty incidents. On the other hand, many convicted people have been released from prison on the basis of DNA techniques, which were not available when a crime had originally been committed.

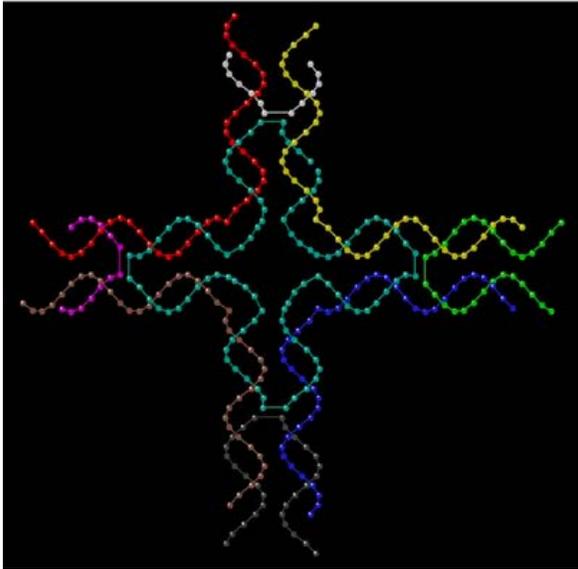
Bioinformatics

Bioinformatics involves the manipulation, searching, and data mining of biological data, and this includes DNA sequence data. The development of techniques to store and search DNA sequences have led to widely applied advances in computer science, especially string searching algorithms, machine learning and database theory. String searching or matching algorithms, which find an occurrence of a sequence of letters inside a larger sequence of letters, were developed to search for specific sequences of nucleotides. The DNA sequenced may be aligned with other DNA sequences to identify homologous sequences and locate the specific mutations that make them distinct. These techniques, especially multiple sequence alignment, are used in studying phylogenetic relationships and protein function. Data sets representing entire genomes' worth of DNA sequences, such as those produced by the Human Genome Project, are difficult to use without the

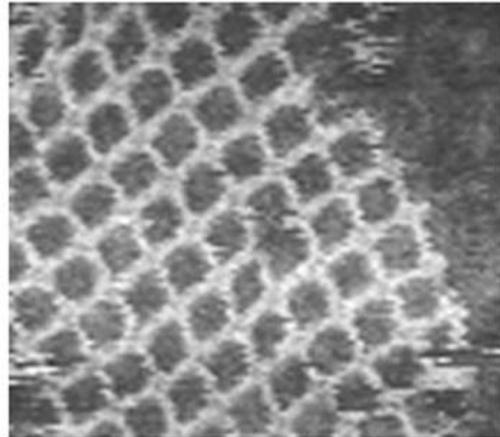
annotations that identify the locations of genes and regulatory elements on each chromosome. Regions of DNA sequence that have the characteristic patterns associated with protein- or RNA-coding genes can be identified by gene finding algorithms, which allow researchers to predict the presence of particular gene products and their possible functions in an organism even before they have been isolated experimentally. Entire genomes may also be compared which can shed light on the evolutionary history of particular organism and permit the examination of complex evolutionary events.

DNA nanotechnology

A



B



100 nm

The DNA structure at left (schematic shown) will self-assemble into the structure visualized by atomic force microscopy at right. DNA nanotechnology is the field that seeks to design nanoscale structures using the molecular recognition properties of DNA molecules.

DNA nanotechnology uses the unique molecular recognition properties of DNA and other nucleic acids to create self-assembling branched DNA complexes with useful properties. DNA is thus used as a structural material rather than as a carrier of biological information. This has led to the creation of two-dimensional periodic lattices (both tile-based as well as using the "DNA origami" method) as well as three-dimensional structures in the shapes of polyhedra. Nanomechanical devices and algorithmic self-assembly have also been demonstrated, and these DNA structures have been used to template the arrangement of other molecules such as gold nanoparticles and streptavidin proteins.

History and anthropology

Because DNA collects mutations over time, which are then inherited, it contains historical information, and, by comparing DNA sequences, geneticists can infer the

evolutionary history of organisms, their phylogeny. This field of phylogenetics is a powerful tool in evolutionary biology. If DNA sequences within a species are compared, population geneticists can learn the history of particular populations. This can be used in studies ranging from ecological genetics to anthropology; For example, DNA evidence is being used to try to identify the Ten Lost Tribes of Israel.

DNA has also been used to look at modern family relationships, such as establishing family relationships between the descendants of Sally Hemings and Thomas Jefferson. This usage is closely related to the use of DNA in criminal investigations detailed above. Indeed, some criminal investigations have been solved when DNA from crime scenes has matched relatives of the guilty individual.

History of DNA research



James D. Watson and Francis Crick (right), co-originators of the double-helix model, with Maclyn McCarty (left).

DNA was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein". In 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. However, Levene thought the chain was short and the bases repeated in a fixed order. In 1937 William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure.



Raymond Gosling, co-creator of the single X-ray diffraction image

In 1928, Frederick Griffith discovered that traits of the "smooth" form of the *Pneumococcus* could be transferred to the "rough" form of the same bacteria by mixing killed "smooth" bacteria with the live "rough" form. This system provided the first clear suggestion that DNA carries genetic information—the Avery–MacLeod–McCarty experiment—when Oswald Avery, along with coworkers Colin MacLeod and Maclyn McCarty, identified DNA as the transforming principle in 1943. DNA's role in heredity was confirmed in 1952, when Alfred Hershey and Martha Chase in the Hershey–Chase experiment showed that DNA is the genetic material of the T2 phage.

In 1953, James D. 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 (labeled 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. Chargaff's rules played a very important role in establishing double-helix configurations for B-DNA as well as A-DNA.

Experimental evidence supporting the Watson and Crick model were published in a series of five articles in the same issue of *Nature*. Of these, Franklin and Gosling's paper was the first publication of their own X-ray diffraction data and original analysis method that partially supported the Watson and Crick model; this issue also contained an article on DNA structure by Maurice Wilkins and two of his colleagues, whose analysis and *in vivo* B-DNA X-ray patterns also supported the presence *in vivo* of the double-helical DNA configurations as proposed by Crick and Watson for their double-helix molecular model of DNA in the previous two pages of *Nature*. In 1962, after Franklin's death, Watson, Crick, and Wilkins jointly received the Nobel Prize in Physiology or Medicine. However, Nobel rules of the time allowed only living recipients, but a vigorous debate continues on who should receive credit for the discovery.

In an influential presentation in 1957, Crick laid out the central dogma of molecular biology, which foretold the relationship between DNA, RNA, and proteins, and articulated the "adaptor hypothesis". Final confirmation of the replication mechanism that was implied by the double-helical structure followed in 1958 through the Meselson–Stahl experiment. Further work by Crick and coworkers showed that the genetic code was based on non-overlapping triplets of bases, called codons, allowing Har Gobind Khorana, Robert W. Holley and Marshall Warren Nirenberg to decipher the genetic code. These findings represent the birth of molecular biology.

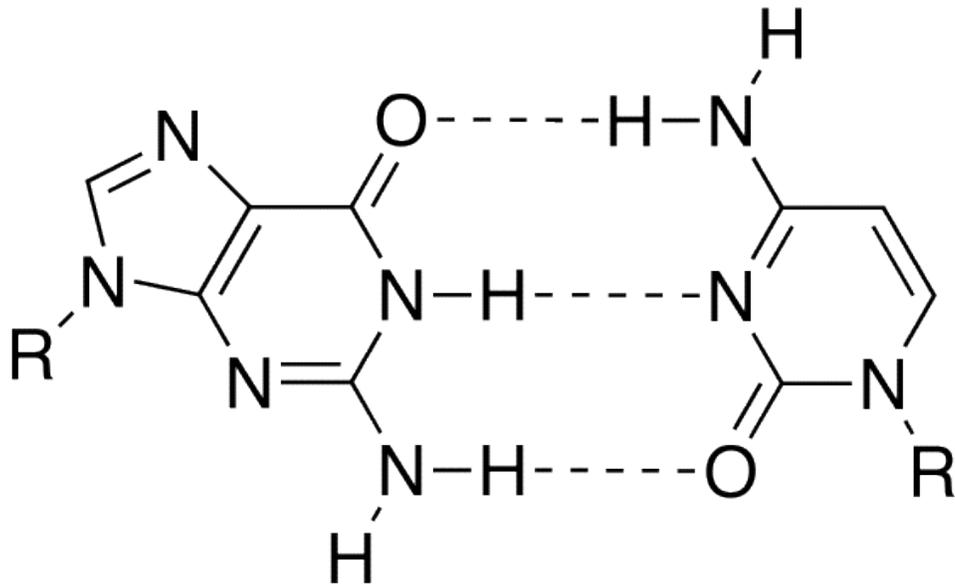
Chapter- 2

Base Pair

In molecular biology and genetics, two nucleotides on opposite complementary DNA or RNA strands that are connected via hydrogen bonds are called a **base pair** (often abbreviated bp). In the canonical Watson-Crick DNA base pairing, adenine (**A**) forms a base pair with thymine (**T**) and guanine (**G**) forms a base pair with cytosine (**C**). In RNA, thymine is replaced by uracil (**U**). Alternate hydrogen bonding patterns, such as the wobble base pair and Hoogsteen base pair, also occur—in particular, in RNA—giving rise to complex and functional tertiary structures. It is important to note that pairing is the mechanism by which codons on messenger RNA molecules are recognized by anticodons on transfer RNA during protein translation. Some DNA- or RNA-binding enzymes can recognize specific base pairing patterns that identify particular regulatory regions of genes.

The size of an individual gene or an organism's entire genome is often measured in base pairs because DNA is usually double-stranded. Hence, the number of total base pairs is equal to the number of nucleotides in one of the strands (with the exception of non-coding single-stranded regions of telomeres). The haploid human genome (23 chromosomes) is estimated to be about 3 billion base pairs long and to contain 20,000–25,000 distinct genes. A kilobase (kb) is a unit of measurement in molecular biology equal to 1000 base pairs of DNA or RNA.

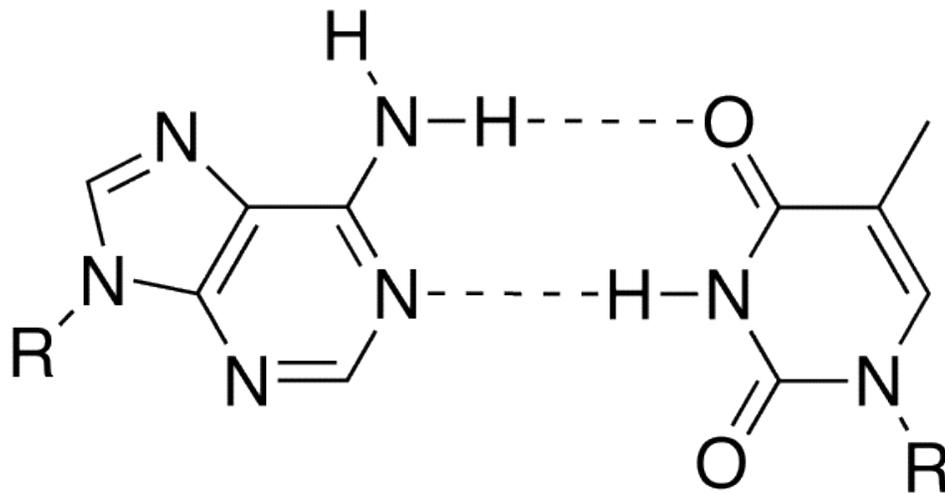
Hydrogen bonding and stability



Guanine

Cytosine

A GC base pair demonstrating three intermolecular hydrogen bonds



Adenine

Thymine

An AT base pair demonstrating two intermolecular hydrogen bonds

Hydrogen bonding is the chemical interaction that underlies the base-pairing rules described above. Appropriate geometrical correspondence of hydrogen bond donors and acceptors allows only the "right" pairs to form stably. DNA with high GC-content is more stable than DNA with low GC-content, but, contrary to popular belief, the hydrogen bonds do not stabilize the DNA significantly, and stabilization is mainly due to stacking interactions.

The larger nucleobases, adenine and guanine, are members of a class of double-ringed chemical structures called purines; the smaller nucleobases, cytosine and thymine (and uracil), are members of a class of single-ringed chemical structures called pyrimidines. Purines are complementary only with pyrimidines: pyrimidine-pyrimidine pairings are energetically unfavorable because the molecules are too far apart for hydrogen bonding to be established; purine-purine pairings are energetically unfavorable because the molecules are too close, leading to overlap repulsion. The only other possible pairings are GT and AC; these pairings are mismatches because the pattern of hydrogen donors and acceptors do not correspond. The GU pairing, with two hydrogen bonds, does occur fairly often in RNA.

Paired DNA and RNA molecules are comparatively stable at room temperature but the two nucleotide strands will separate above a melting point that is determined by the length of the molecules, the extent of mispairing (if any), and the GC content. Higher GC content results in higher melting temperatures; it is, therefore, unsurprising that the genomes of extremophile organisms such as *Thermus thermophilus* are particularly GC-rich. On the converse, regions of a genome that need to separate frequently — for example, the promoter regions for often-transcribed genes — are comparatively GC-poor. GC content and melting temperature must also be taken into account when designing primers for PCR reactions.

Base stacking

Base stacking interactions in DNA and RNA are due to dispersion attraction, short-range exchange repulsion, and electrostatic interactions, which also contribute to stability. Again, GC stacking interactions with adjacent bases tend to be more favorable. (Note, however, that a GC stacking interaction with the next base pair is geometrically different from a CG interaction.) Base stacking effects are especially important in the secondary structure and tertiary structure of RNA; for example, RNA stem-loop structures are stabilized by base stacking in the loop region.

Base analogs and intercalators

Chemical analogs of nucleotides can take the place of proper nucleotides and establish non-canonical base-pairing, leading to errors (mostly point mutations) in DNA replication and DNA transcription. One common mutagenic base analog is 5-bromouracil, which resembles thymine but can base-pair to guanine in its enol form.

Other chemicals, known as DNA intercalators, fit into the gap between adjacent bases on a single strand and induce frameshift mutations by "masquerading" as a base, causing the DNA replication machinery to skip or insert additional nucleotides at the intercalated site. Most intercalators are large polyaromatic compounds and are known or suspected carcinogens. Examples include ethidium bromide and acridine.

Examples

The following DNA sequences illustrate pair double-stranded patterns. By convention, the top strand is written from the 5' end to the 3' end; thus, the bottom strand is written 3' to 5'.

A base-paired DNA sequence:

```
ATCGATTGAGCTCTAGCG
TAGCTAACTCGAGATCGC
```

The corresponding RNA sequence, in which uracil is substituted for thymine where uracil takes its place in the RNA strand:

```
AUCGAUUGAGCUCUAGCG
UAGCUAACUCGAGAUCGC
```

Length measurements

The following abbreviations are commonly used to describe the length of a D/RNA molecule:

- bp = base pair(s)—one bp corresponds to circa 3.4 Å of length along the strand
- kb (= kbp) = kilo base pairs = 1,000 bp
- Mb = mega base pairs = 1,000,000 bp
- Gb = giga base pairs = 1,000,000,000 bp.

In case of single stranded DNA/RNA units of nucleotides are used, abbreviated nt (or knt, Mnt, Gnt), as they are not paired. For distinction between units of computer storage and bases kbp, Mbp, Gbp, etc. may be used for basepairs. The length of 16S rDNA for bacteria is 1542 base-pairs in length.

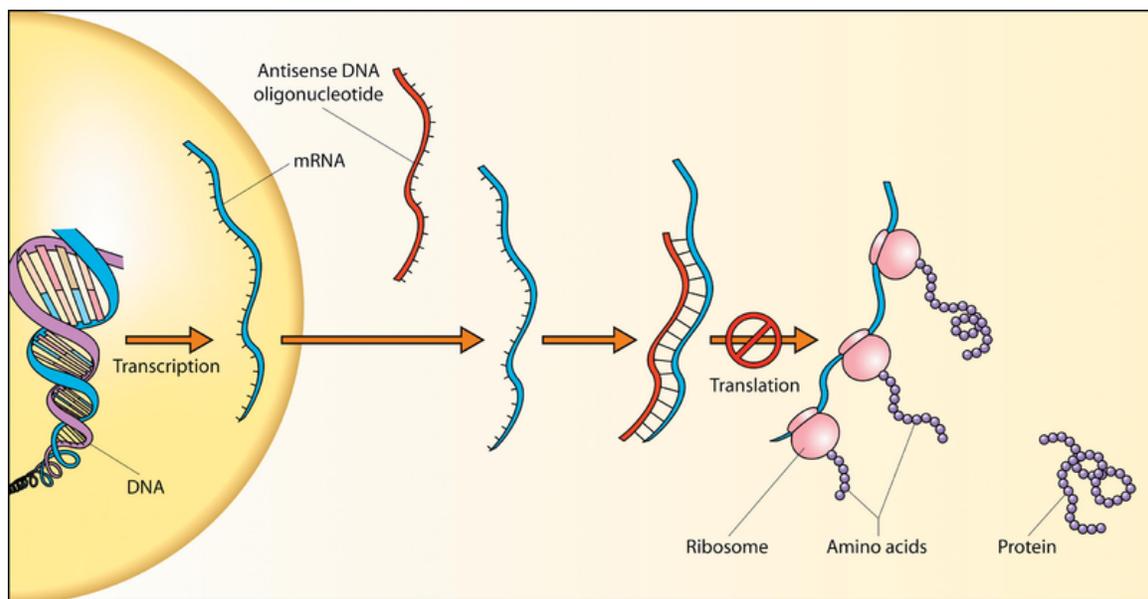
The Centimorgan is also often used to imply distance along a chromosome, but the number of base-pairs it corresponds to varies widely. In the Human genome, it is about 1 million base pairs

Chapter- 3

Sense (Molecular Biology)

Sense, when applied in a molecular biology context, is a general concept used to compare the polarity of nucleic acid molecules, such as DNA or RNA, to other nucleic acid molecules. Depending on the context within molecular biology, sense may have slightly different meanings.

DNA sense



Schematic showing how antisense DNA strands can interfere with protein translation

Molecular biologists call a single strand of DNA **sense** (or **positive (+)** sense) if an RNA version of the same sequence is translated or translatable into protein. Its complementary strand is called **antisense** (or **negative (-)** sense). Sometimes the phrase **coding strand** is encountered; however, protein coding and non-coding RNA's can be transcribed similarly from both strands, in some cases being transcribed in both directions from a common promoter region, or being transcribed from within introns, on both strands.

Antisense DNA

DNA normally has two strands, i.e., the sense strand and the antisense strand. In double-stranded DNA, only one strand codes for the RNA that is translated into protein. This DNA strand is referred to as the antisense strand. The strand that does not code for RNA is called the sense strand because it has a similar sequence to the messenger RNA (mRNA). Both the sense DNA strand and the mRNA transcript are complementary to the template DNA strand. Note that the DNA strands called "sense" and "antisense" are sometimes switched in older textbooks.

Antisense molecules interact with complementary strands of nucleic acids, modifying expression of genes.

Example with Double Stranded DNA

DNA strand 1: antisense strand (copied to) → RNA strand (sense)

DNA strand 2: sense strand

Some regions within a double strand of DNA code for genes, which are usually instructions specifying the order of amino acids in a protein along with regulatory sequences, splicing sites, noncoding introns, and other complicating details. For a cell to use this information, one strand of the DNA serves as a template for the synthesis of a complementary strand of RNA. The template DNA strand is called the transcribed strand with antisense sequence and the mRNA transcript is said to be sense sequence (the complement of antisense). Because the DNA is double-stranded, the strand complementary to the antisense sequence is called non-transcribed strand and has the same sense sequence as the mRNA transcript (though T bases in DNA are substituted with U bases in RNA).

5'CGCTATAGCGTTTCAT 3'	DNA template (noncoding) strand, antisense strand, Watson strand	Used as a template for transcription.
3'GCGATATCGCAAAGTA 5'	DNA nontemplate (coding) strand, sense strand, Crick strand	Complementary to the template strand.
3'GCGAUAUCGCAAAGUA 5'	Sense mRNA transcript	RNA strand that is transcribed from the antisense template strand. Note: it is <i>complementary</i> to the antisense DNA template and <i>identical</i> to the sense nontemplate

		strand, except all thymines are now uracils (T-->U).
5'CGCUAUAGCGUUUCAU 3'	Antisense RNA transcript	RNA strand that is transcribed from the nontemplate sense strand. Note: it is identical to the DNA antisense template strand, except all thymines are now uracils (T-->U).

A note on the confusion between "sense" and "antisense" strands: The strand names actually depend on which direction you are writing the sequence that contains the information for proteins (the "sense" information), not on which strand is on the top or bottom (that is arbitrary). The only real biological information that is important for labeling strands is the location of the 5' phosphate group and the 3' hydroxyl group because these ends determine the direction of transcription and translation. A sequence 5' CGCTAT 3' is equivalent to a sequence written 3' TATCGC 5' as long as the 5' and 3' ends are noted. If the ends are not labeled, convention is to assume that the sequence is written in the 5' to 3' direction. Good rule of thumb for figuring out the "sense" strand: Look for the start codon ATG (AUG in mRNA). In the table example, the sense mRNA has the AUG codon at the end (remember that translation proceeds in the 5' to 3' direction).

Ambisense

A single-stranded genome that contains both positive-sense and negative-sense is said to be **ambisense**. Bunya viruses have 3 single-stranded RNA (ssRNA) fragments containing both positive-sense and negative-sense sections; arenaviruses are also ssRNA viruses with an ambisense genome, as they have 2 fragments that are mainly negative-sense except for part of the 5' ends of the large and small segments of their genome.

Antisense RNA

Antisense RNA is an RNA transcript that is complementary to endogenous mRNA. In other words, it is a non-coding strand complementary to the coding sequence of RNA; this is similar to negative-sense viral RNA. Introducing a transgene coding for antisense RNA is a technique used to block expression of a gene of interest. Radioactively-labelled antisense RNA can be used to show the level of transcription of genes in various cell types. Some alternative antisense structural types are being experimentally applied as antisense therapy, with at least one antisense therapy approved for use in humans.

When mRNA forms a duplex with a complementary antisense RNA sequence, translation is blocked. This process is called RNA interference.

Antisense nucleic acid molecules have been used experimentally to bind to mRNA and prevent expression of specific genes. Antisense therapies are also in development; in the

USA, the Food and Drug Administration (FDA) has approved a phosphorothioate antisense oligo, fomivirsen (Vitravene), for human therapeutic use.

Cells can produce antisense RNA molecules naturally, which interact with complementary mRNA molecules and inhibit their expression.

RNA sense in viruses

In virology, the genome of an RNA virus can be said to be either **positive-sense**, also known as a "plus-strand", or **negative-sense**, also known as a "minus-strand". In most cases, the terms *sense* and *strand* are used interchangeably, making such terms as *positive-strand* equivalent to *positive-sense*, and *plus-strand* equivalent to *plus-sense*. Whether a virus genome is positive-sense or negative-sense can be used as a basis for classifying viruses.

Positive-sense

Positive-sense (5' to 3') viral RNA signifies that a particular viral RNA sequence may be directly translated into the desired viral proteins. Therefore, in positive-sense RNA viruses, the viral RNA genome can be considered viral mRNA, and can be immediately translated by the host cell. Unlike negative-sense RNA, positive-sense RNA is of the same sense as mRNA. Some viruses (e.g., Coronaviridae) have positive-sense genomes that can act as mRNA and be used directly to synthesize proteins without the help of a complementary RNA intermediate. Because of this, these viruses do not need to have an RNA polymerase packaged into the virion.

Negative-sense

Negative-sense (3' to 5') viral RNA is complementary to the viral mRNA and thus must be converted to positive-sense RNA by an RNA polymerase prior to translation. Negative-sense RNA (like DNA) has a nucleotide sequence complementary to the mRNA that it encodes. Like DNA, this RNA cannot be translated into protein directly. Instead, it must first be transcribed into a positive-sense RNA that acts as an mRNA. Some viruses (Influenza, for example) have negative-sense genomes and so must carry an RNA polymerase inside the virion.

Enzymes

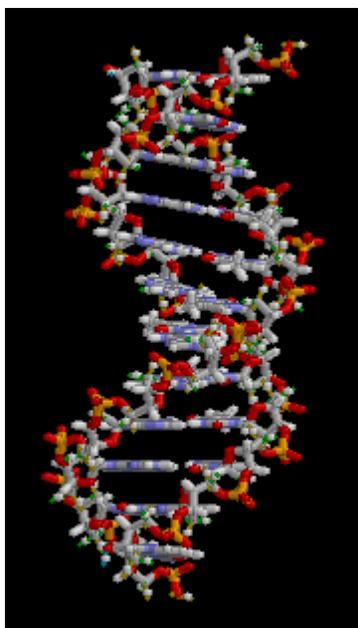
Enzyme-dependent antisense includes forms dependent on RNase H activity to degrade target mRNA, including single-stranded DNA, RNA, and phosphorothioate antisense. The R1 plasmid *hok/sok* system is an example of mRNA antisense regulation process, through enzymatic degradation of the resulting RNA duplex. Double stranded RNA acts as enzyme-dependent antisense through the RNAi/siRNA pathway, involving target mRNA recognition through sense-antisense strand pairing followed by target mRNA degradation by the RNA-induced silencing complex (RISC).

Steric blocking antisense (RNase-H-independent antisense) interferes with gene expression or other mRNA-dependent cellular processes by binding to a target sequence of mRNA and getting in the way of other processes. Steric blocking antisense includes 2'-O alkyl (usually in chimeras with RNase-H dependent antisense), peptide nucleic acid (PNA), locked nucleic acid (LNA), and Morpholino antisense.

Chapter- 4

Molecular Models of DNA





While this purified DNA precipitated in a water jug (top) appears to be a formless mass, nucleic acids actually possess intricate structure at the nanoscale (bottom).

Molecular models of DNA structures are representations of the molecular geometry and topology of Deoxyribonucleic acid (DNA) molecules using one of several means, with the aim of simplifying and presenting the essential, physical and chemical, properties of DNA molecular structures either *in vivo* or *in vitro*. These representations include closely packed spheres (CPK models) made of plastic, metal wires for 'skeletal models', graphic computations and animations by computers, artistic rendering. Computer molecular models also allow animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*.

The more advanced, computer-based molecular models of DNA involve molecular dynamics simulations as well as quantum mechanical computations of vibro-rotations, delocalized molecular orbitals (MOs), electric dipole moments, hydrogen-bonding, and so on. **DNA molecular dynamics modeling** involves simulations of DNA molecular geometry and topology changes with time as a result of both intra- and inter- molecular interactions of DNA. Whereas molecular models of Deoxyribonucleic acid (DNA) molecules such as closely packed spheres (CPK models) made of plastic or metal wires for 'skeletal models' are useful representations of static DNA structures, their usefulness is very limited for representing complex DNA dynamics. Computer molecular modeling allows both animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*.

History



The A-DNA double-helix molecular model of Crick and Watson (consistent with their X-ray data) for which they received a Nobel prize together with M.H.F. Wilkins.

From the very early stages of structural studies of DNA by X-ray diffraction and biochemical means, molecular models such as the Watson-Crick double-helix model were successfully employed to solve the 'puzzle' of DNA structure, and also find how the latter relates to its key functions in living cells. The first high quality X-ray diffraction patterns of A-DNA were reported by Rosalind Franklin and Raymond Gosling in 1953. The first calculations of the Fourier transform of an atomic helix were reported one year earlier by Cochran, Crick and Vand, and were followed in 1953 by the computation of the Fourier transform of a coiled-coil by Crick.

Structural information is generated from X-ray diffraction studies of oriented DNA fibers with the help of molecular models of DNA that are combined with crystallographic and mathematical analysis of the X-ray patterns.

The first reports of a double-helix molecular model of B-DNA structure were made by Watson and Crick in 1953. Last-but-not-least, Maurice F. Wilkins, A. Stokes and H.R. Wilson, reported the first X-ray patterns of *in vivo* B-DNA in partially oriented salmon sperm heads. The development of the first correct double-helix molecular model of DNA by Crick and Watson may not have been possible without the biochemical evidence for the nucleotide base-pairing ([A---T]; [C---G]), or Chargaff's rules. Although such initial studies of DNA structures with the help of molecular models were essentially static, their consequences for explaining the *in vivo* functions of DNA were significant in the areas of protein biosynthesis and the quasi-universality of the genetic code. Epigenetic transformation studies of DNA *in vivo* were however much slower to develop in spite of their importance for embryology, morphogenesis and cancer research. Such chemical dynamics and biochemical reactions of DNA are much more complex than the molecular dynamics of DNA physical interactions with water, ions and proteins/enzymes in living cells.

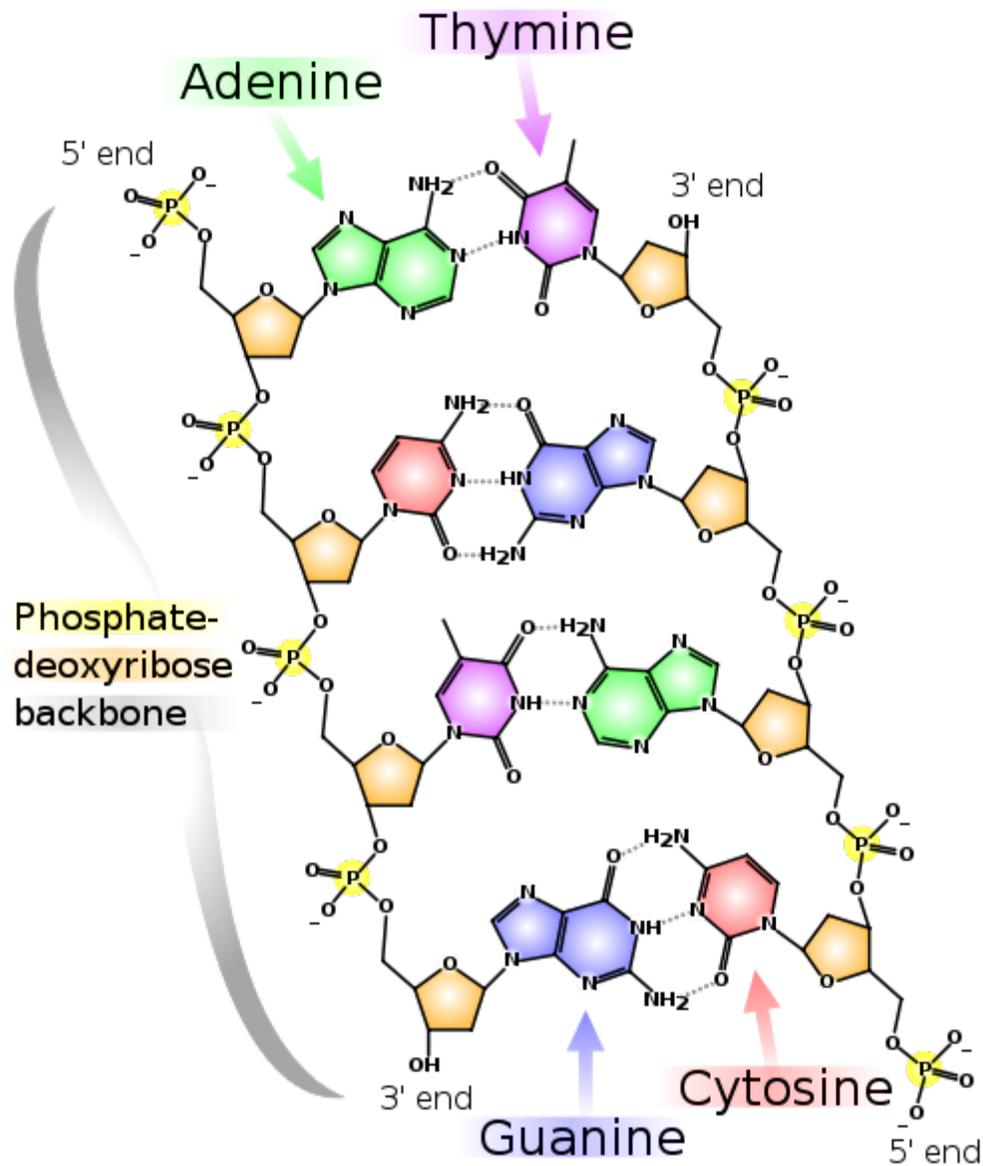
Importance

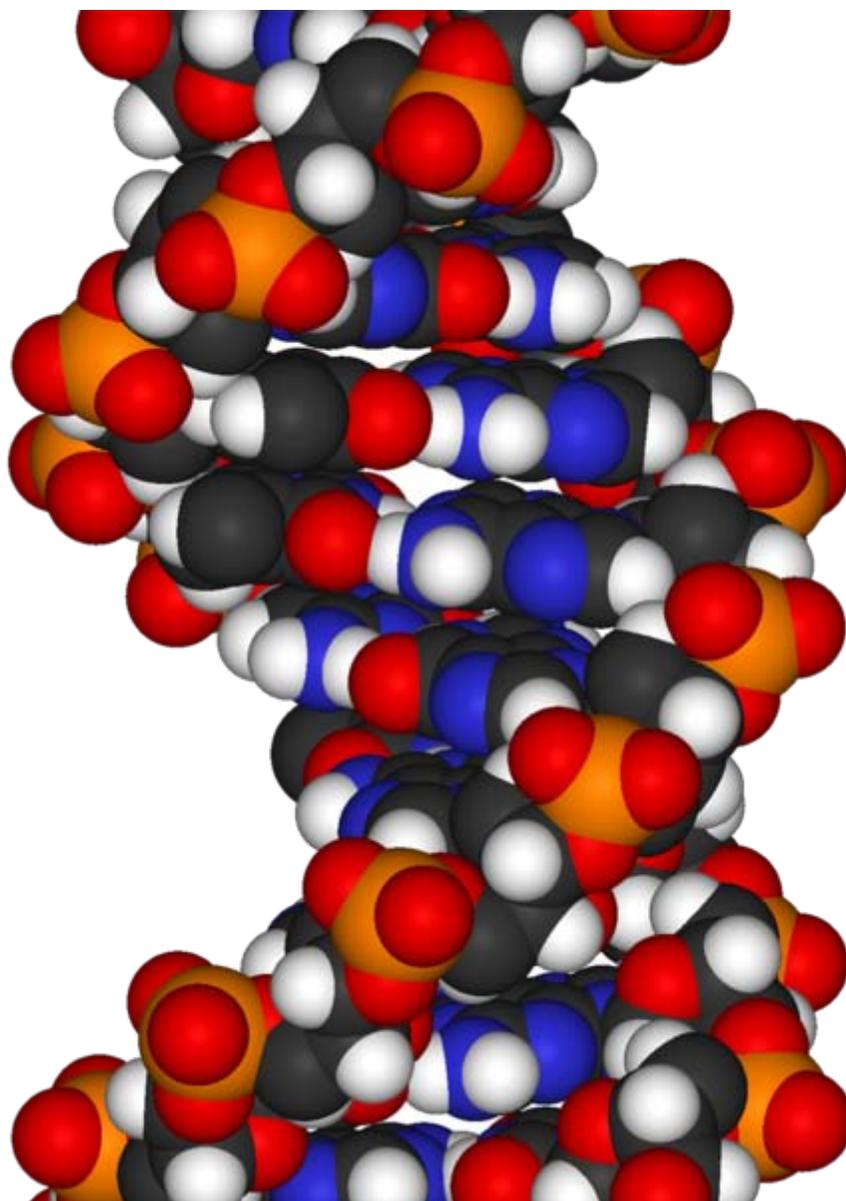
An old standing dynamic problem is how DNA "self-replication" takes place in living cells that should involve transient uncoiling of supercoiled DNA fibers. Although DNA consists of relatively rigid, very large elongated biopolymer molecules called "fibers" or chains (that are made of repeating nucleotide units of four basic types, attached to deoxyribose and phosphate groups), its molecular structure *in vivo* undergoes dynamic configuration changes that involve dynamically attached water molecules and ions. Supercoiling, packing with histones in chromosome structures, and other such supramolecular aspects also involve *in vivo* DNA topology which is even more complex than DNA molecular geometry, thus turning molecular modeling of DNA into an especially challenging problem for both molecular biologists and biotechnologists. Like other large molecules and biopolymers, DNA often exists in multiple stable geometries (that is, it exhibits conformational isomerism) and configurational, quantum states which are close to each other in energy on the potential energy surface of the DNA molecule.

Such varying molecular geometries can also be computed, at least in principle, by employing *ab initio* quantum chemistry methods that can attain high accuracy for small molecules, although claims that acceptable accuracy can be also achieved for polynucleotides, as well as DNA conformations, were recently made on the basis of VCD spectral data. Such quantum geometries define an important class of *ab initio* molecular models of DNA whose exploration has barely started especially in connection with results obtained by VCD in solutions. More detailed comparisons with such *ab initio* quantum computations are in principle obtainable through 2D-FT NMR spectroscopy and relaxation studies of polynucleotide solutions or specifically labeled DNA, as for example with deuterium labels.

In an interesting twist of roles, the DNA molecule itself was proposed to be utilized for quantum computing. Both DNA nanostructures as well as DNA 'computing' biochips have been built.

Fundamental concepts





At top, the chemical structure of DNA showing the base-pairing. This depiction of a DNA duplex lacks information about the molecule's three-dimensional structure, at bottom.

The chemical structure of DNA is insufficient to understand the complexity of the 3D structures of DNA. On the other hand, animated molecular models allow one to visually explore the three-dimensional (3D) structure of DNA. The DNA model shown (far right) is a space-filling, or CPK, model of the DNA double-helix.

The hydrogen bonding dynamics and proton exchange is very different by many orders of magnitude between the two systems of fully hydrated DNA and water molecules in ice. Thus, the DNA dynamics is complex, involving nanosecond and several tens of picosecond time scales, whereas that of liquid ice is on the picosecond time scale, and

that of proton exchange in ice is on the millisecond time scale. The proton exchange rates in DNA and attached proteins may vary from picosecond to nanosecond, minutes or years, depending on the exact locations of the exchanged protons in the large biopolymers.

A simple harmonic oscillator 'vibration' is only an oversimplified dynamic representation of the longitudinal vibrations of the DNA intertwined helices which were found to be anharmonic rather than harmonic as often assumed in quantum dynamic simulations of DNA.

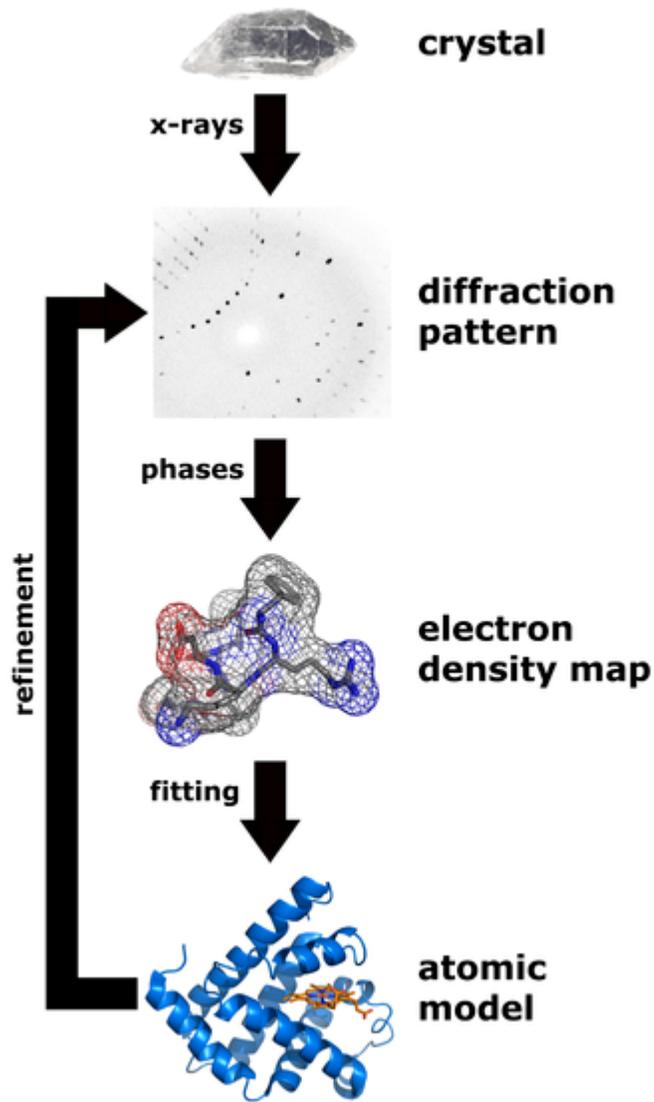
DNA structure

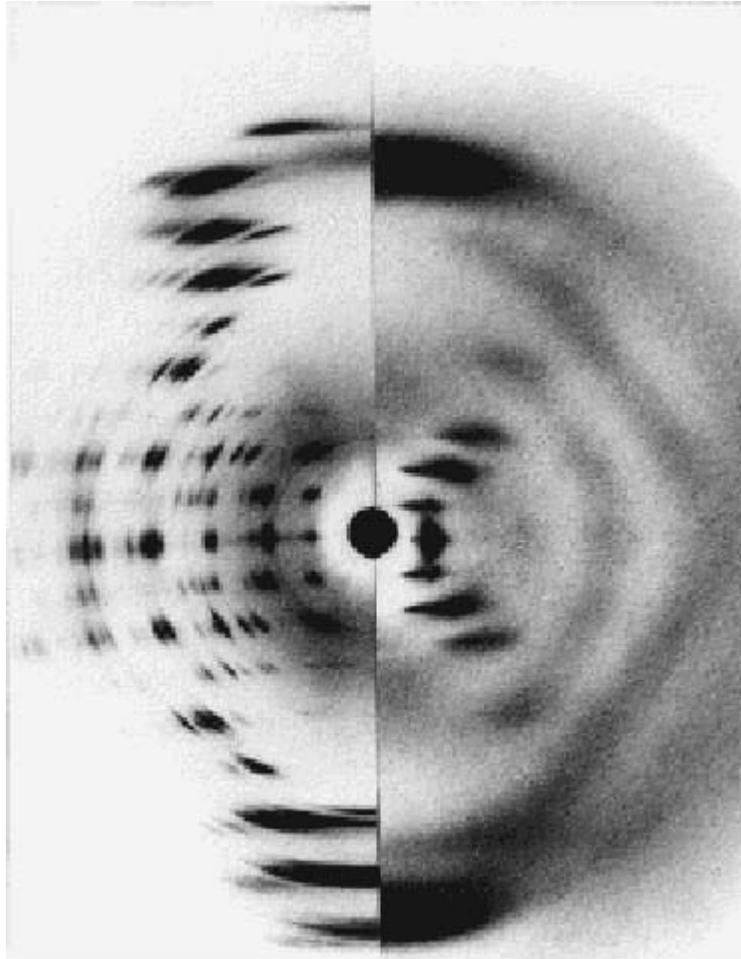
The structure of DNA shows a variety of forms, both double-stranded and single-stranded. The mechanical properties of DNA, which are directly related to its structure, are a significant problem for cells. Every process which binds or reads DNA is able to use or modify the mechanical properties of DNA for purposes of recognition, packaging and modification. The extreme length (a chromosome may contain a 10 cm long DNA strand), relative rigidity and helical structure of DNA has led to the evolution of histones and of enzymes such as topoisomerases and helicases to manage a cell's DNA. The properties of DNA are closely related to its molecular structure and sequence, particularly the weakness of the hydrogen bonds and electronic interactions that hold strands of DNA together compared to the strength of the bonds within each strand.

Experimental techniques which can directly measure the mechanical properties of DNA are relatively new, and high-resolution visualization in solution is often difficult. Nevertheless, scientists have uncovered large amount of data on the mechanical properties of this polymer, and the implications of DNA's mechanical properties on cellular processes is a topic of active current research.

The DNA found in many cells can be macroscopic in length - a few centimetres long for each human chromosome. Consequently, cells must compact or "package" DNA to carry it within them. In eukaryotes this is carried by spool-like proteins known as histones, around which DNA winds. It is the further compaction of this DNA-protein complex which produces the well known mitotic eukaryotic chromosomes.

DNA structure determination using molecular modeling and DNA X-ray patterns



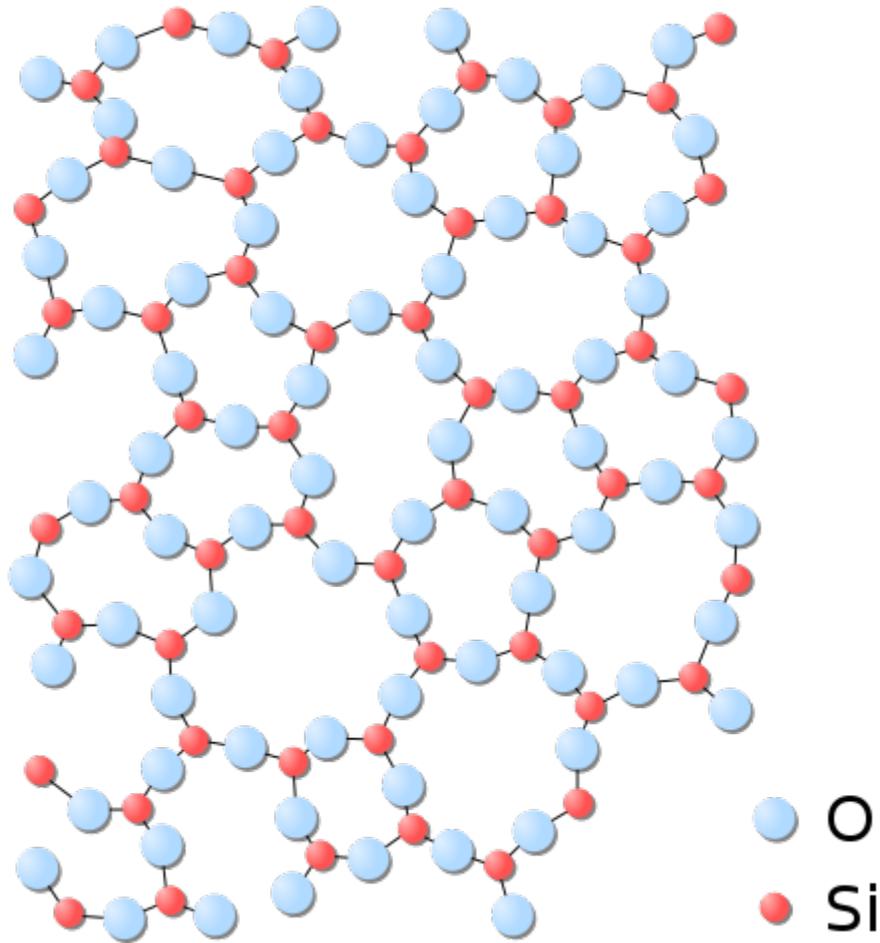


A-DNA B-DNA

Top, the major steps involved in DNA structure determination by X-ray crystallography showing the important role played by molecular models of DNA structure in this iterative process. Bottom, an image of actual A- and B- DNA X-ray patterns obtained from oriented and hydrated DNA fibers.

After DNA has been separated and purified by standard biochemical techniques one has a sample in a jar much like in the figure at the top. Below are the main steps involved in generating structural information from X-ray diffraction studies of oriented DNA fibers that are drawn from the hydrated DNA sample with the help of molecular models of DNA that are combined with crystallographic and mathematical analysis of the X-ray patterns.

Paracrystalline lattice models of B-DNA structures



Silica glass is another example of a material which is organized into a paracrystalline lattice.

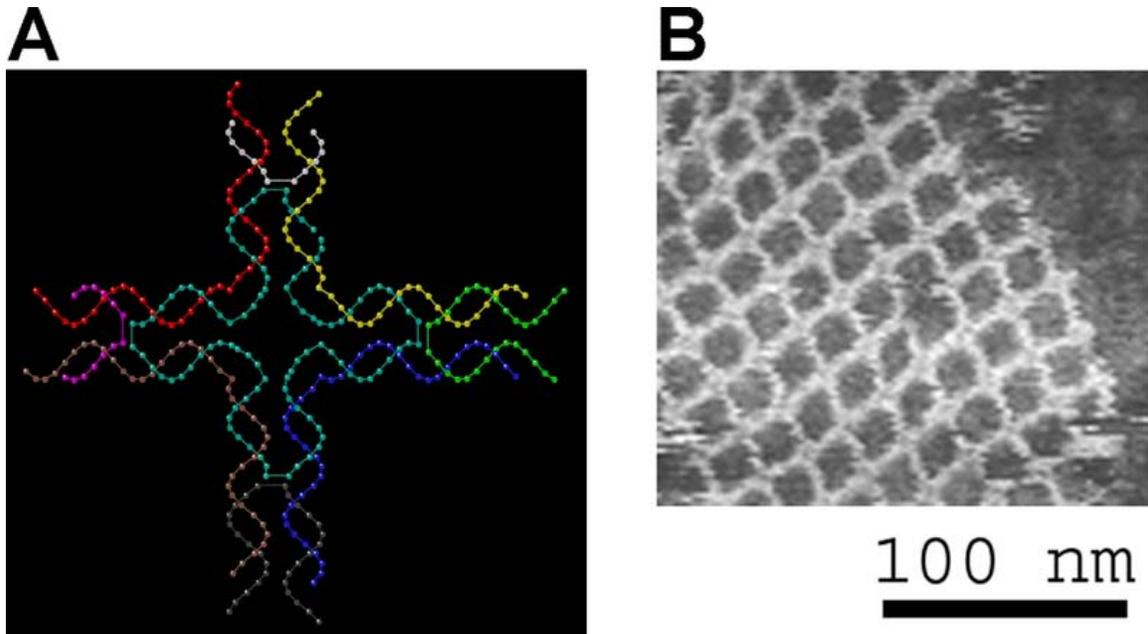
A paracrystalline lattice, or paracrystal, is a molecular or atomic lattice with significant amounts (e.g., larger than a few percent) of partial disordering of molecular arrangements. Limiting cases of the paracrystal model are nanostructures, such as glasses, liquids, etc., that may possess only local ordering and no global order. A simple example of a paracrystalline lattice is shown in the following figure for a silica glass:

Liquid crystals also have paracrystalline rather than crystalline structures.

Highly hydrated B-DNA occurs naturally in living cells in such a paracrystalline state, which is a dynamic one in spite of the relatively rigid DNA double-helix stabilized by parallel hydrogen bonds between the nucleotide base-pairs in the two complementary, helical DNA chains. For simplicity most DNA molecular models omit both water and ions dynamically bound to B-DNA, and are thus less useful for understanding the dynamic behaviors of B-DNA *in vivo*. The physical and mathematical analysis of X-ray

and spectroscopic data for paracrystalline B-DNA is therefore much more complicated than that of crystalline, A-DNA X-ray diffraction patterns. The paracrystal model is also important for DNA technological applications such as DNA nanotechnology. Novel techniques that combine X-ray diffraction of DNA with X-ray microscopy in hydrated living cells are now also being developed.

Genomic and biotechnology applications of DNA molecular modeling



Molecular models are useful in the design of structures for DNA nanotechnology. Here, individual DNA tiles (model at left) self-assemble into a highly ordered DNA 2D-nanogrid (AFM image at right).

There are various uses of DNA molecular modeling in Genomics and Biotechnology research applications, from DNA repair to PCR and DNA nanostructures. Two-dimensional DNA junction arrays have been visualized by Atomic force microscopy.

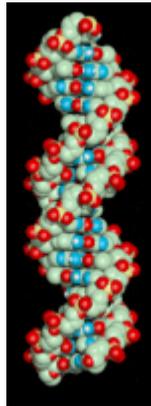
DNA molecular modeling has various uses in genomics and biotechnology, with research applications ranging from DNA repair to PCR and DNA nanostructures. These include computer molecular models of molecules as varied as RNA polymerase, an E. coli, bacterial DNA primase template suggesting very complex dynamics at the interfaces between the enzymes and the DNA template, and molecular models of the mutagenic, chemical interaction of potent carcinogen molecules with DNA. These are all represented in the gallery below.

Technological application include a DNA biochip and DNA nanostructures designed for DNA computing and other dynamic applications of DNA nanotechnology. The image at right is of self-assembled DNA nanostructures. The DNA “tile” structure in this image consists of four branched junctions oriented at 90° angles. Each tile consists of nine DNA

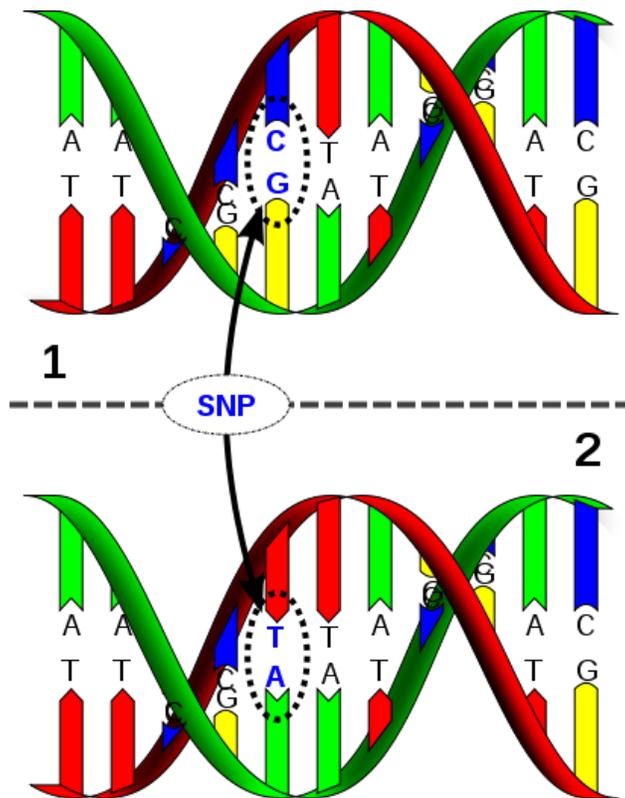
oligonucleotides as shown; such tiles serve as the primary “building block” for the assembly of the DNA nanogrids shown in the AFM micrograph.

Quadruplex DNA may be involved in certain cancers. Images of quadruplex DNA are in the gallery below.

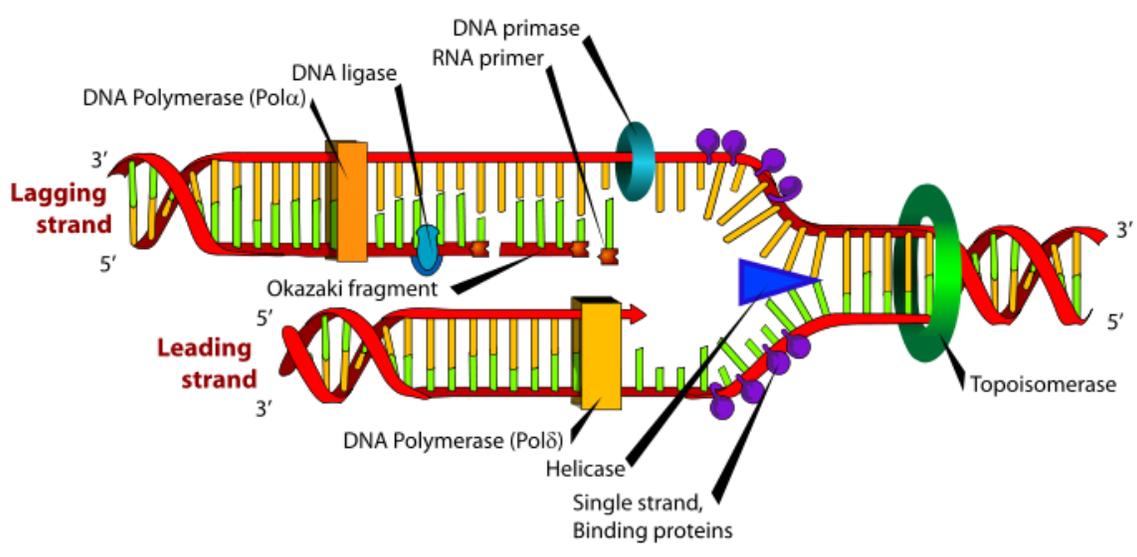
DNA models



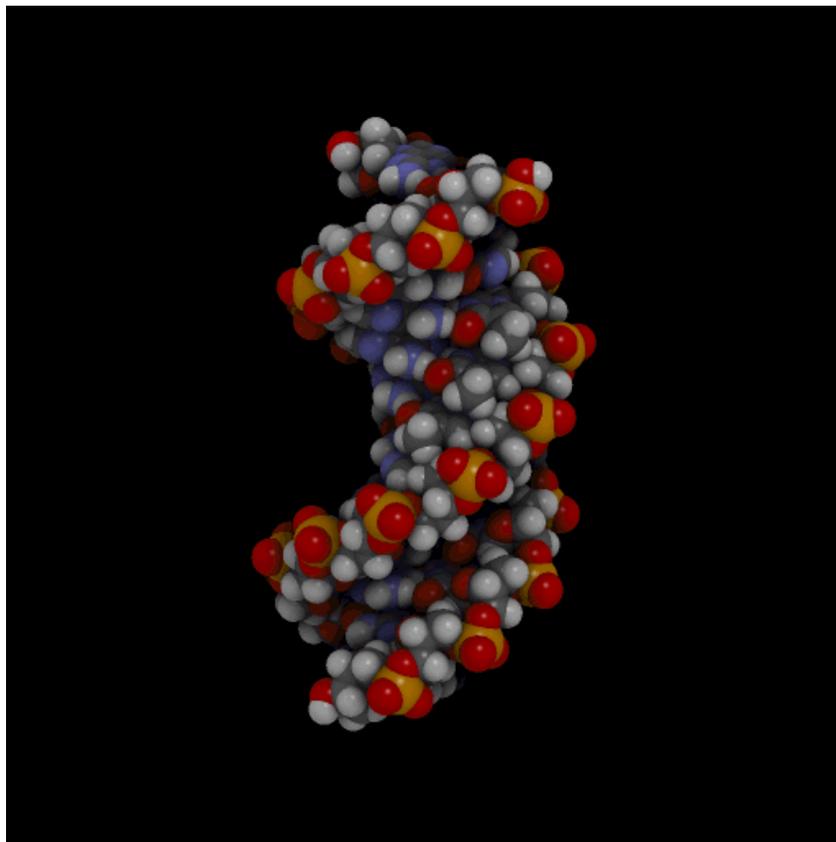
Spinning DNA generic model



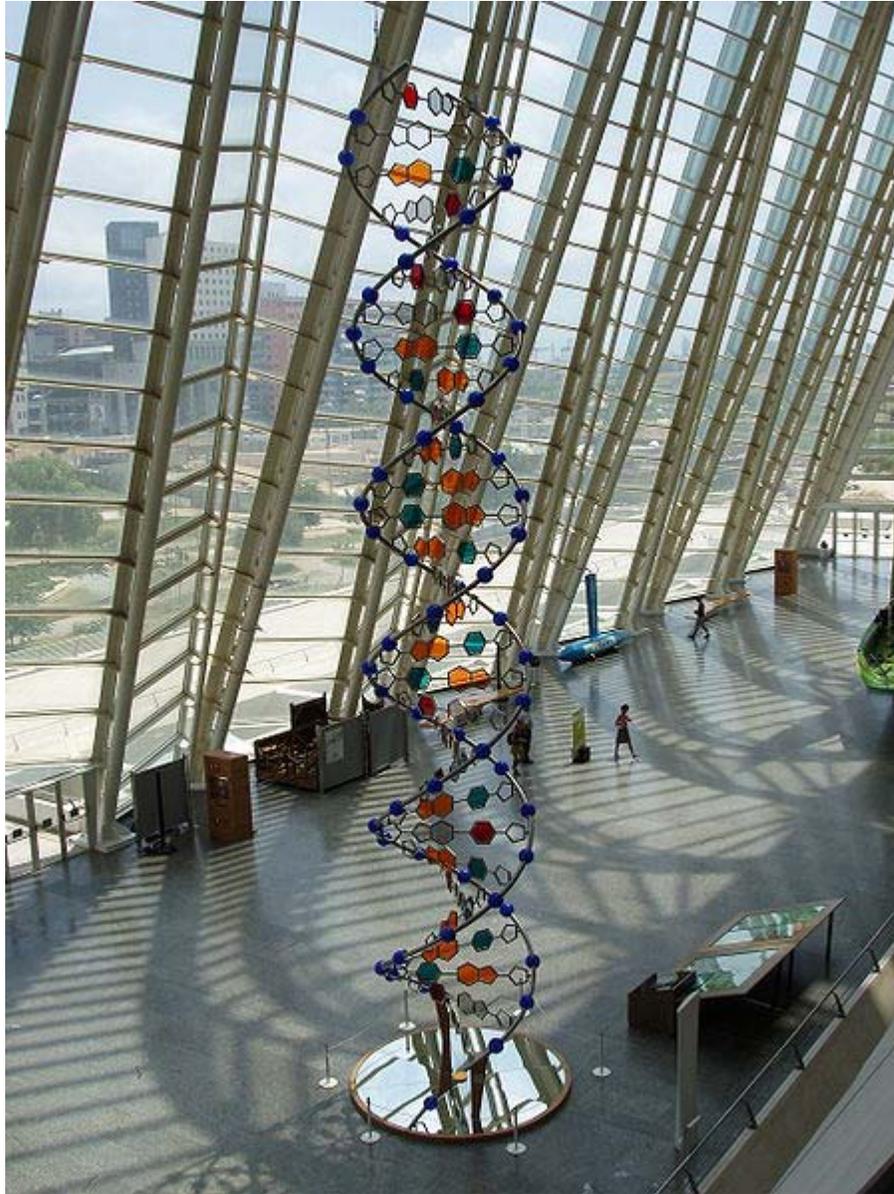
An oversimplified sketch of the double-helix structure of A-DNA



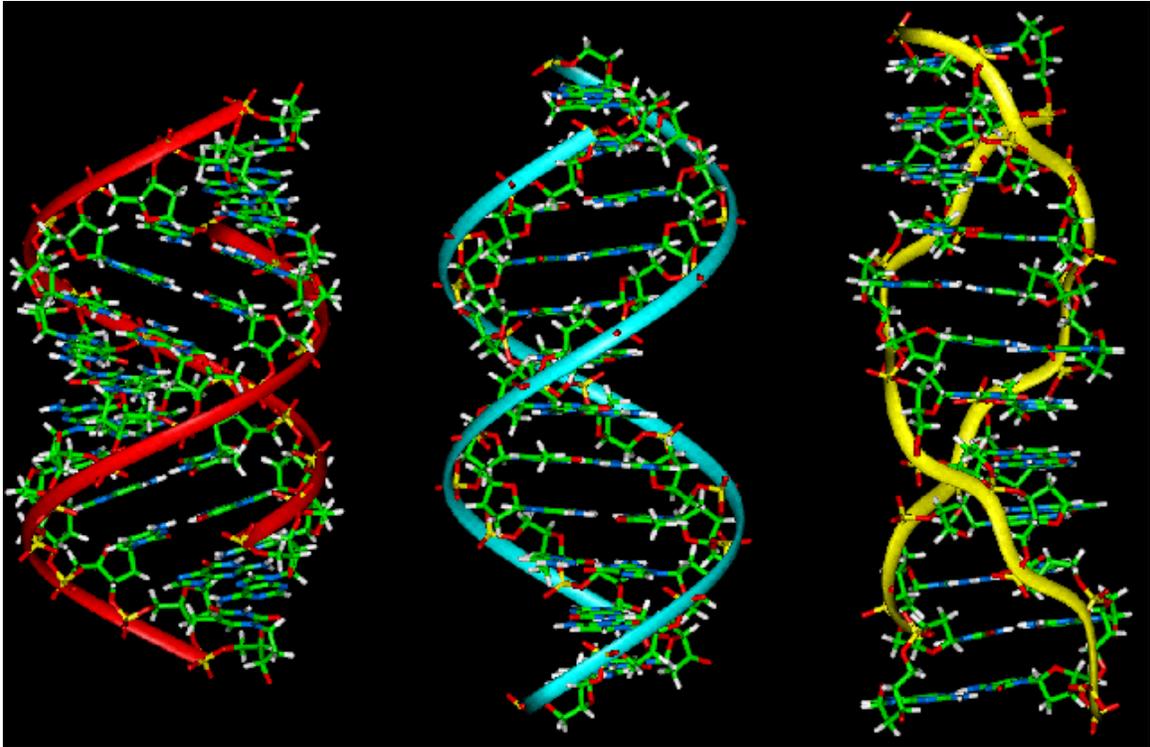
A model of DNA replication based on the double-helix concept



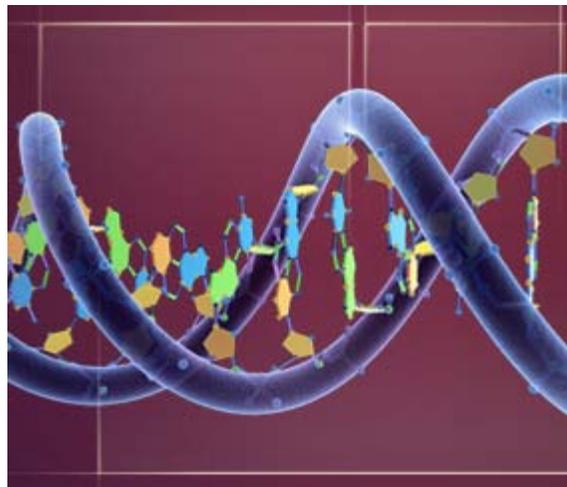
Space-filling molecular model of the A-DNA double helix



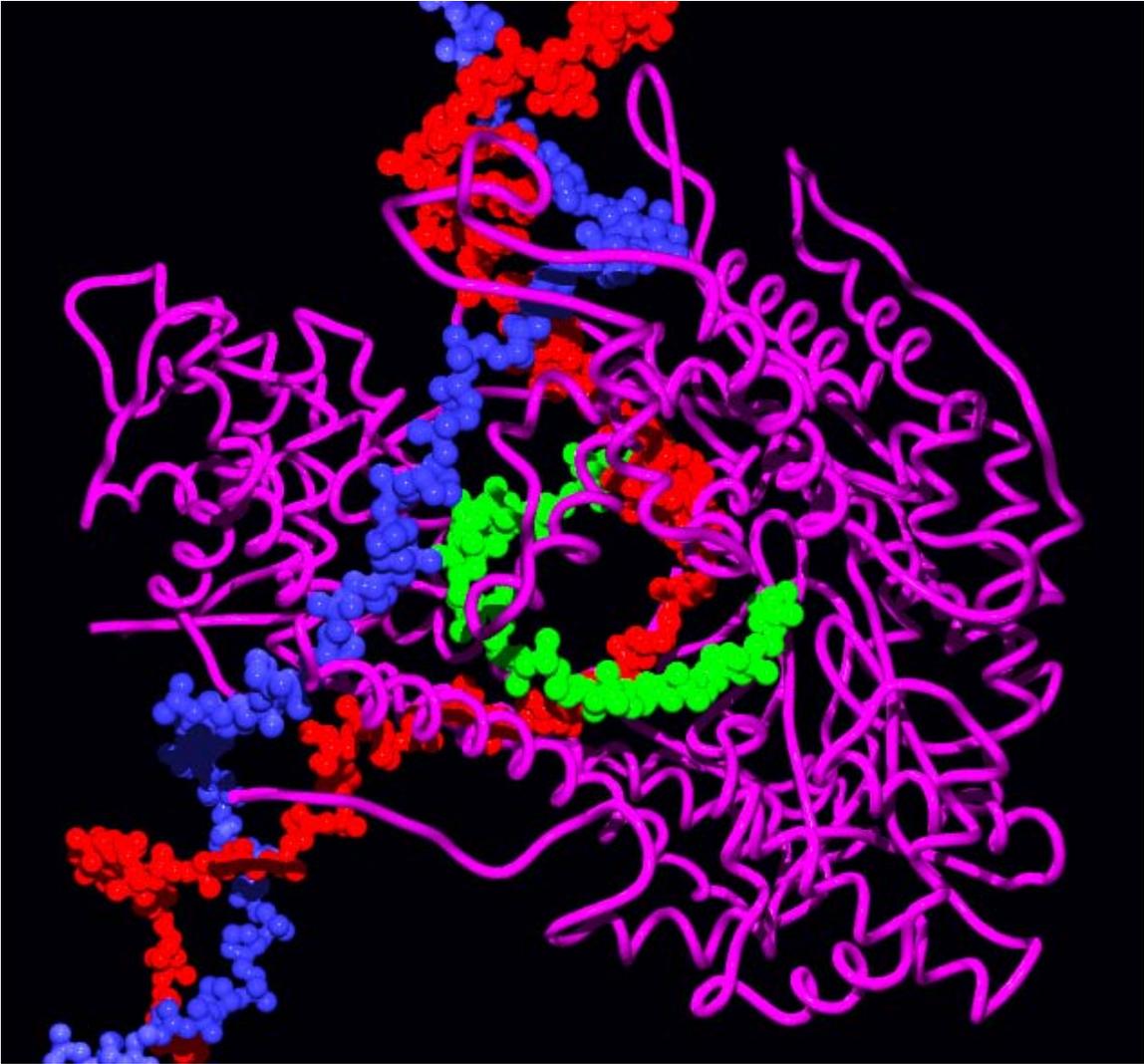
A large-scale Crick-Watson DNA Model shown in the Museum of Príncipe Felipe.



Side view of molecular models of A-, B-, Z- DNA



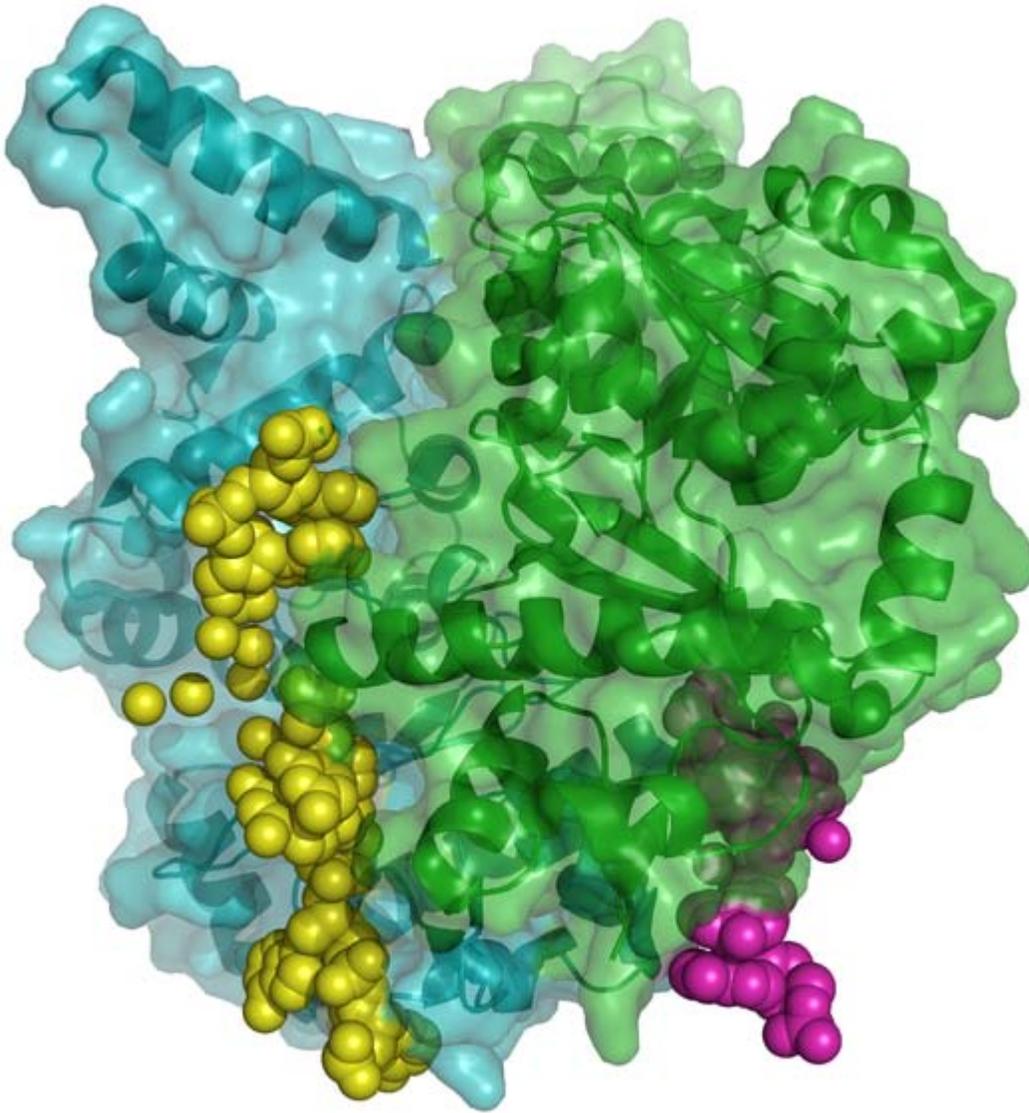
Oversimplified model of the A-DNA double-helix



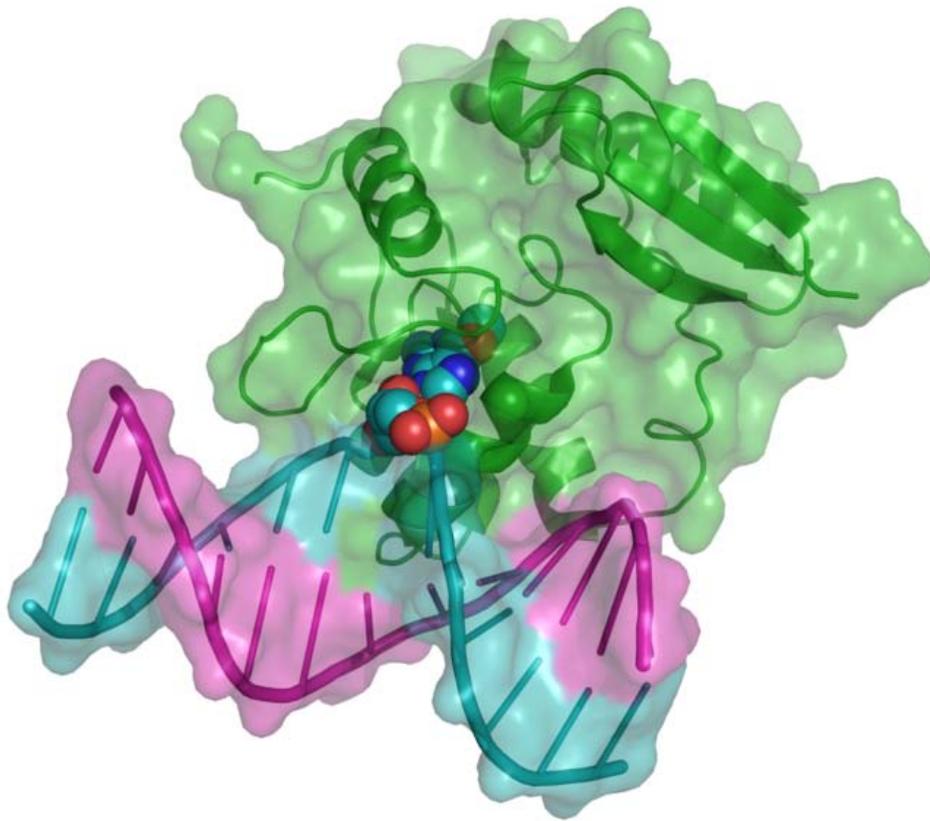
Molecular modeling of RNA Polymerase



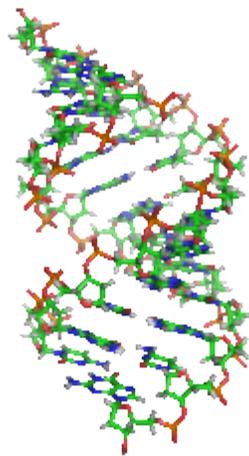
DNA spacefilling molecular model



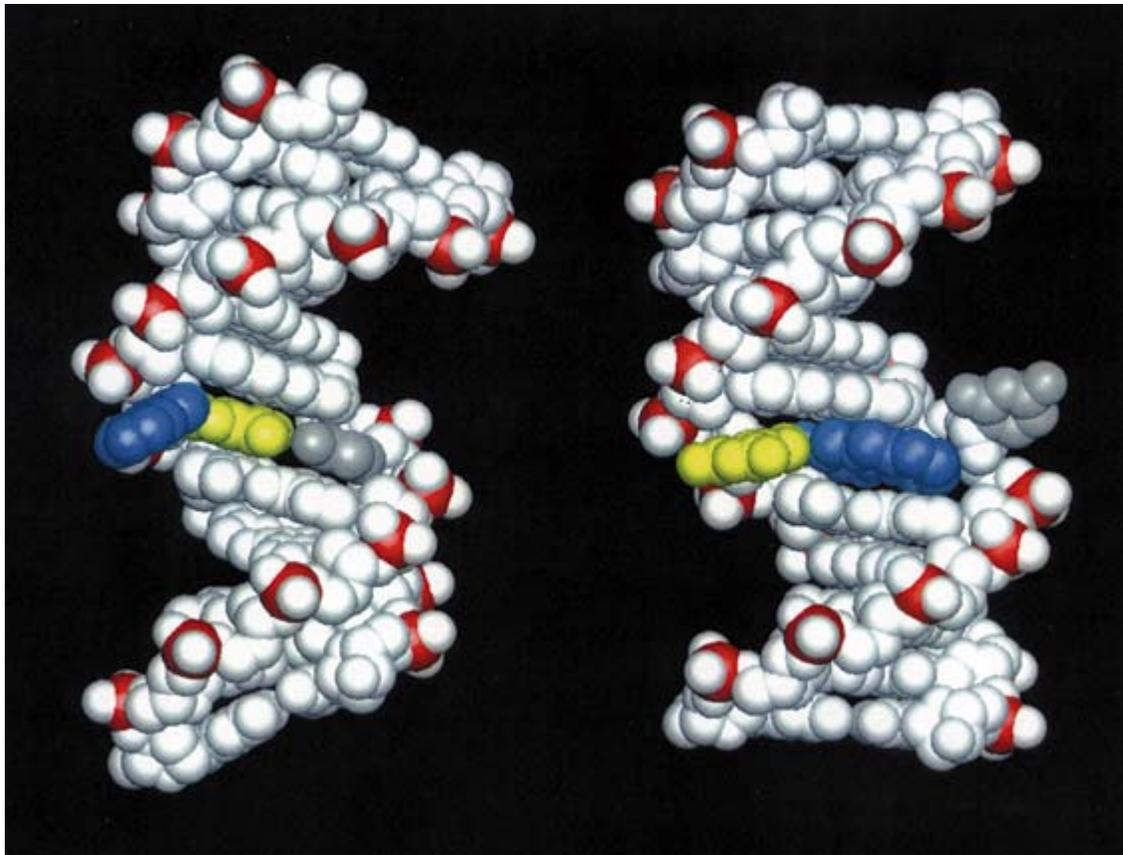
Molecular modeling of a bacterial DNA Primase Template



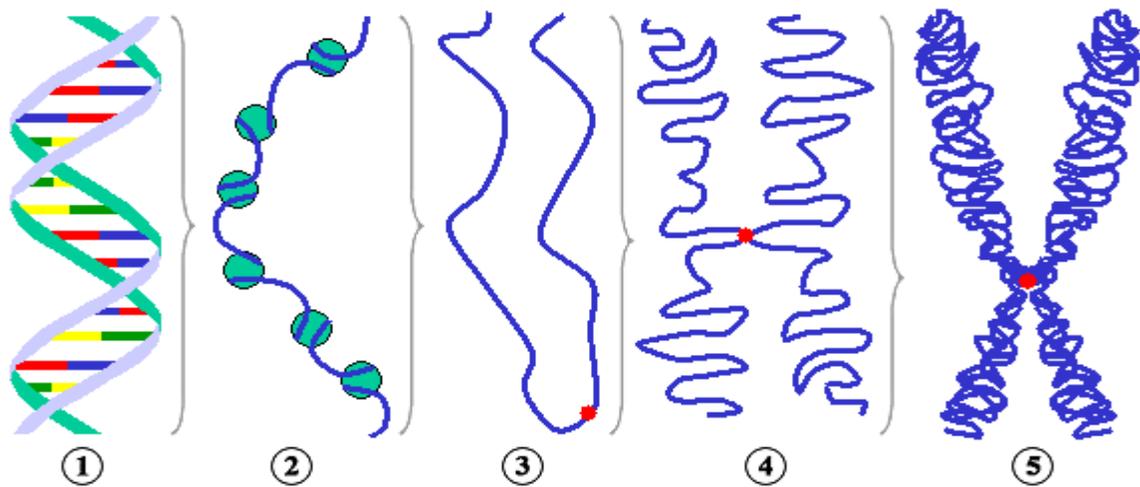
Molecular modeling of DNA interactions with the carcinogen molecule MGMT



Skeletal model of A-DNA



3D Molecular model of DNA damaged by carcinogenic 2-aminofluorene(AF)



Simplified models of chromatin

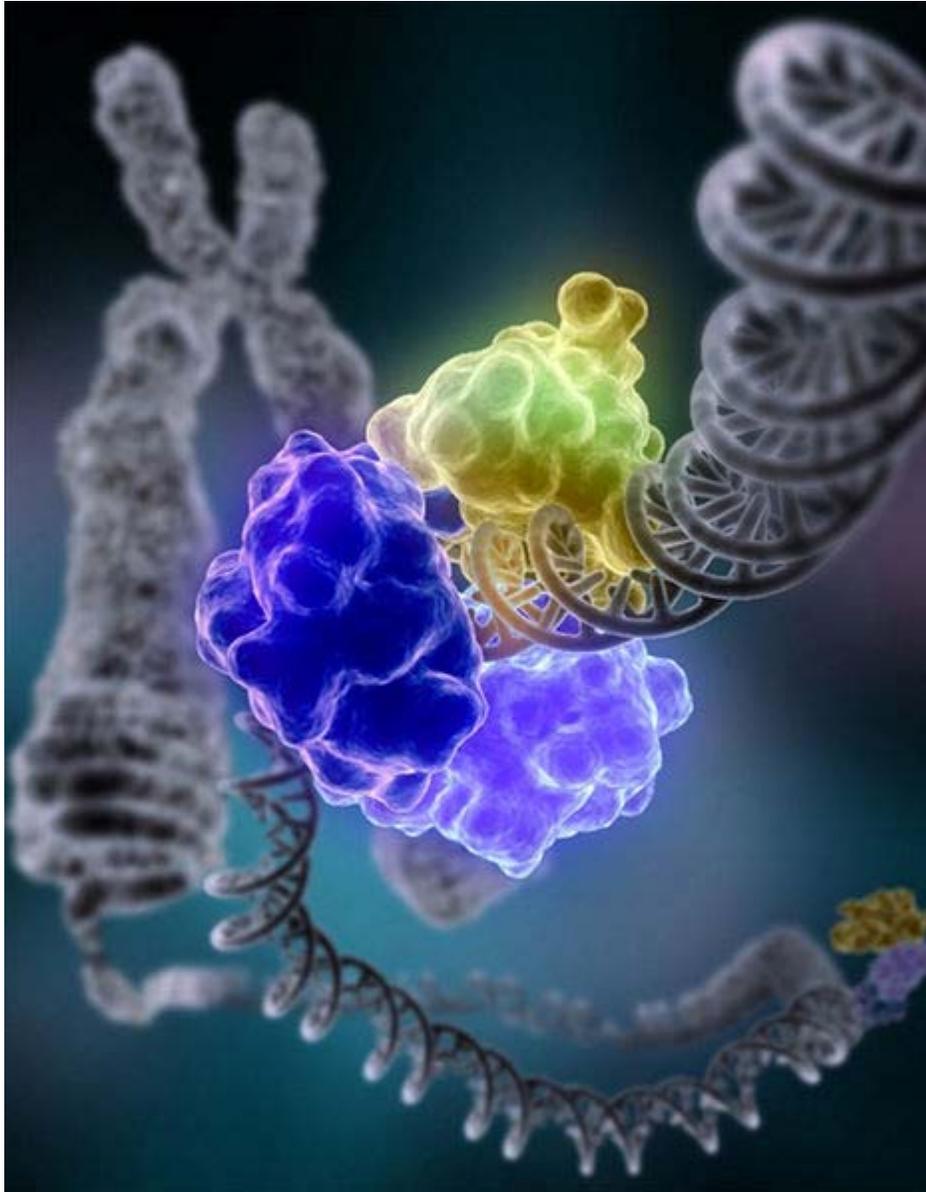
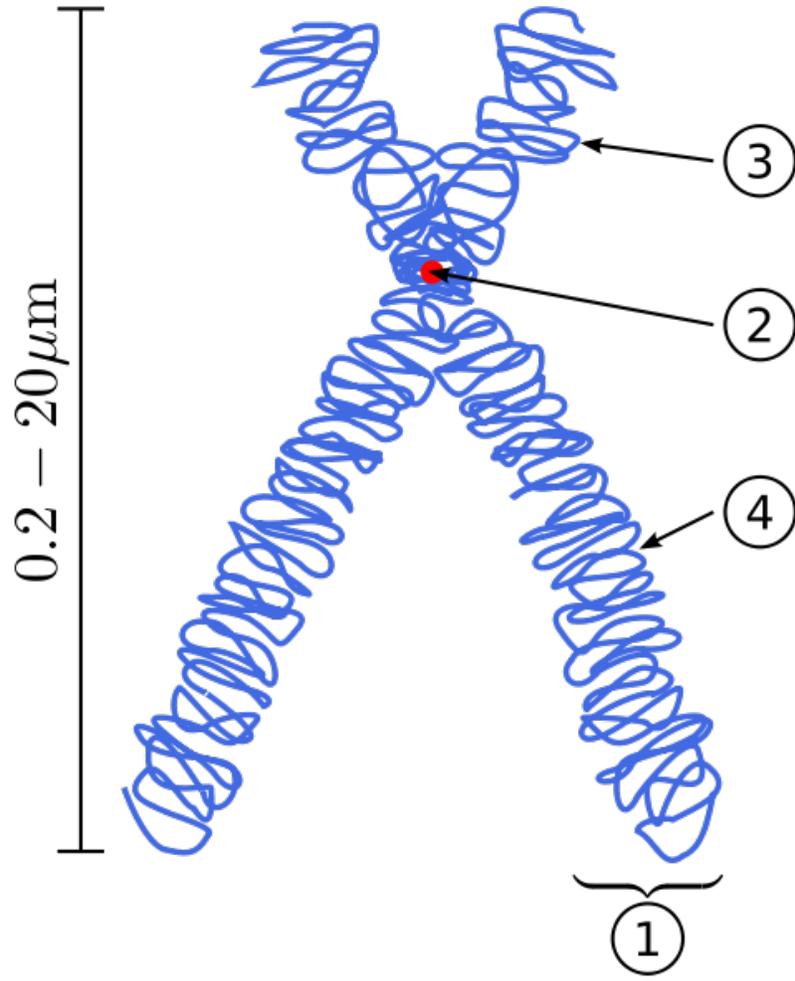
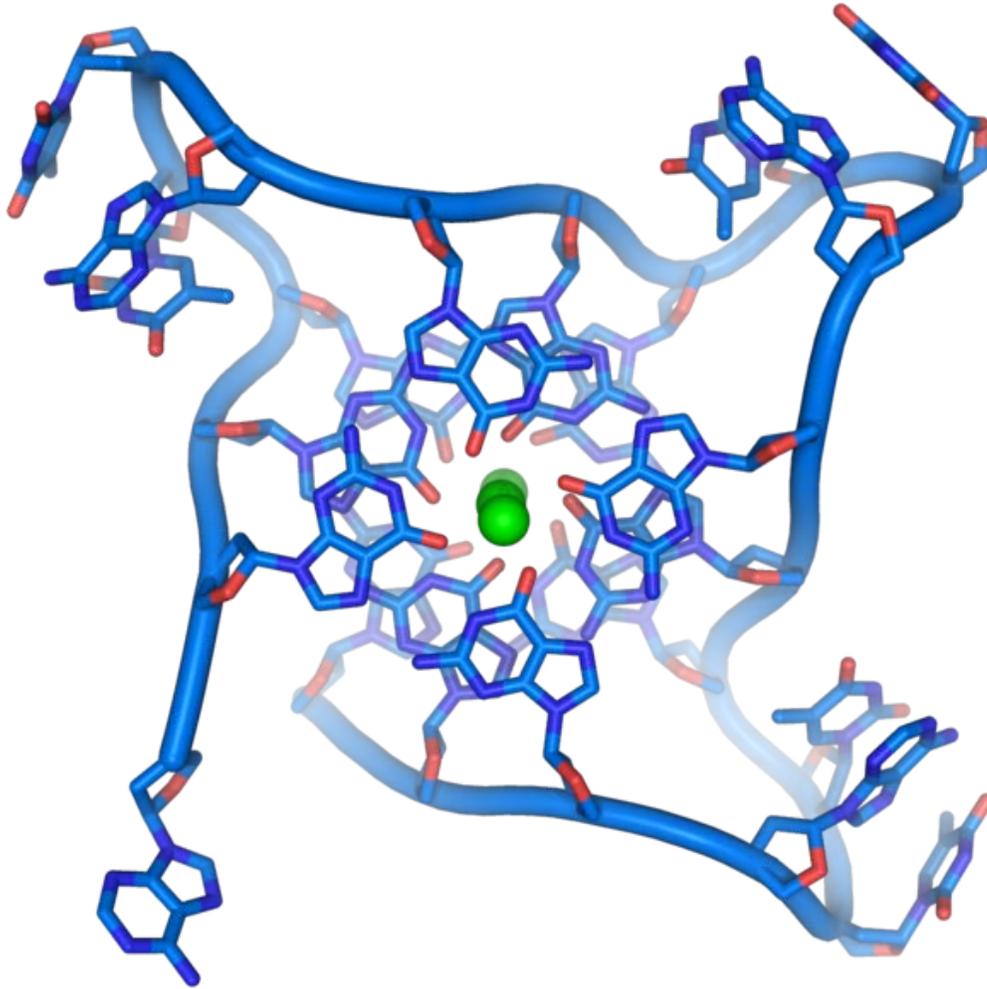


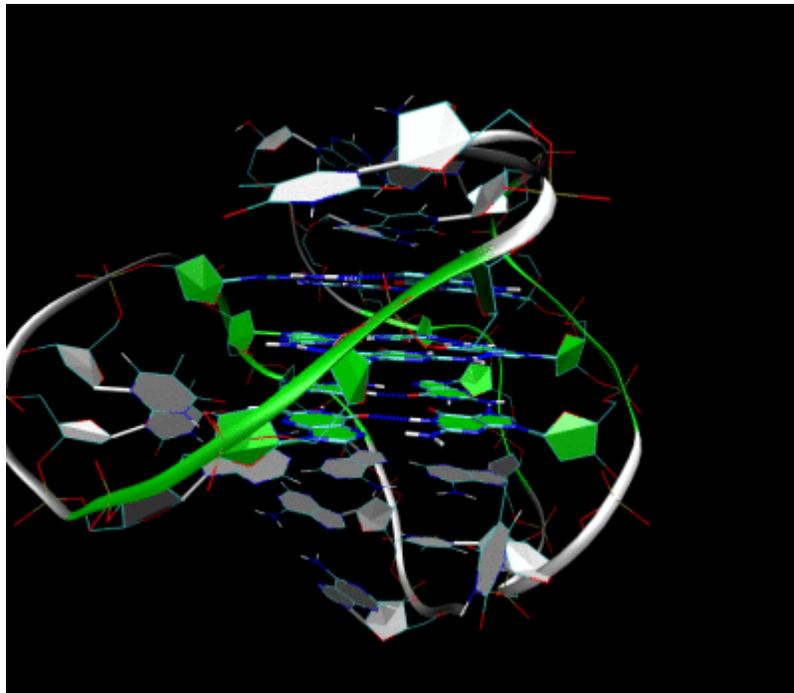
Fig.6. Molecular modeling of DNA repair



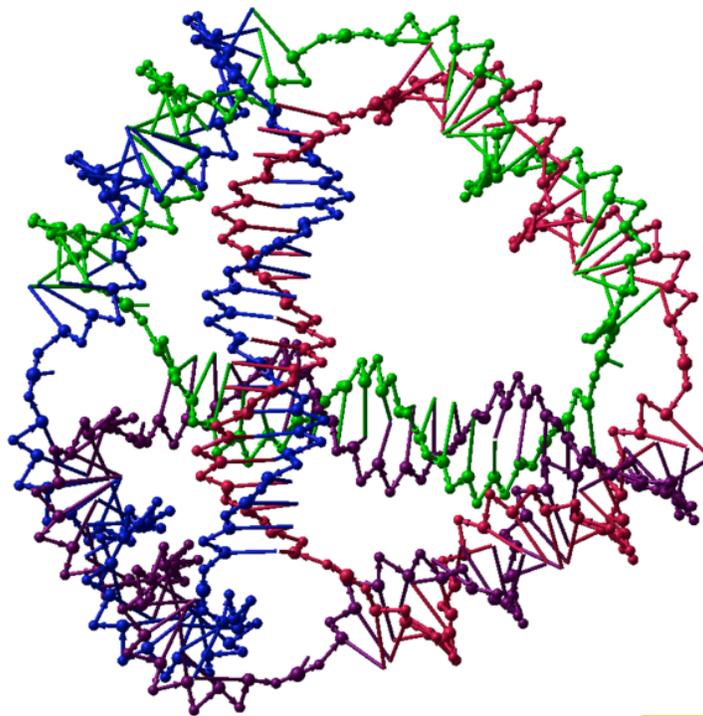
Simplified model of chromosome structure



A hypothetical quadruplex of guanine-rich DNA structures that may be involved in cancers.



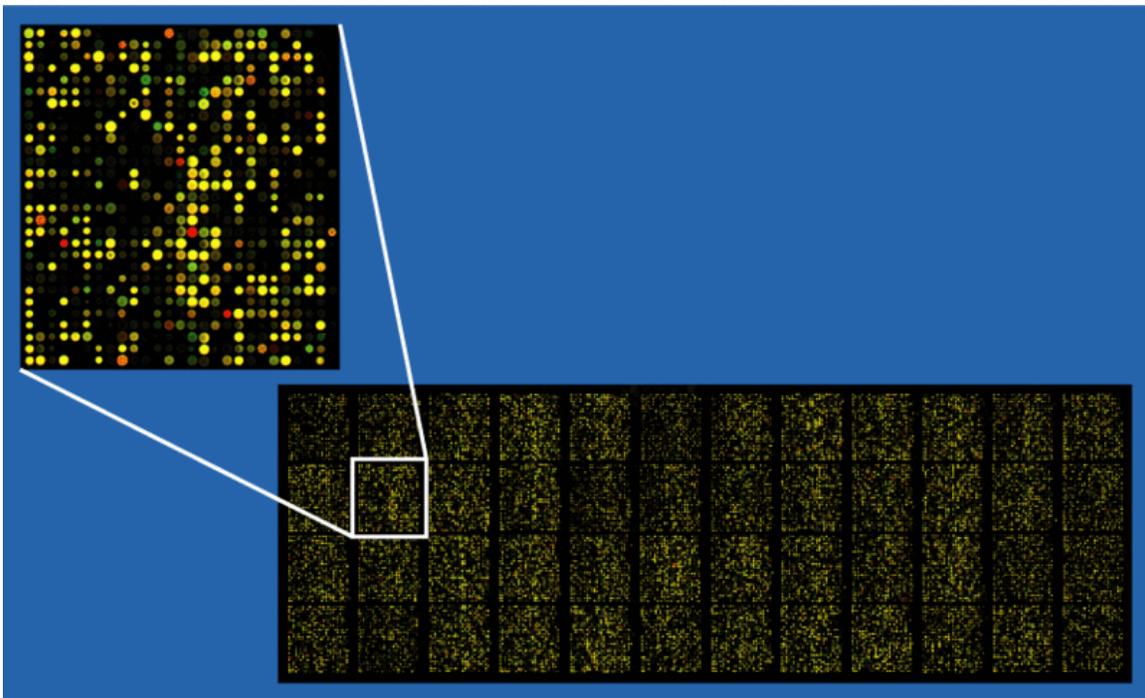
3D Molecular Structure of the intramolecular human telomeric G-quadruplex in potassium solution.



A model of a designed DNA tetrahedron

Chapter- 5

DNA Microarray



Example of an approximately 40,000 probe spotted oligo microarray with enlarged inset to show detail.

A **DNA microarray** is a multiplex technology used in molecular biology. It consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, called features, each containing picomoles (10^{-12} moles) of a specific DNA sequence, known as *probes* (or *reporters*). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA sample (called *target*) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel. Therefore arrays have dramatically accelerated many types of investigation.

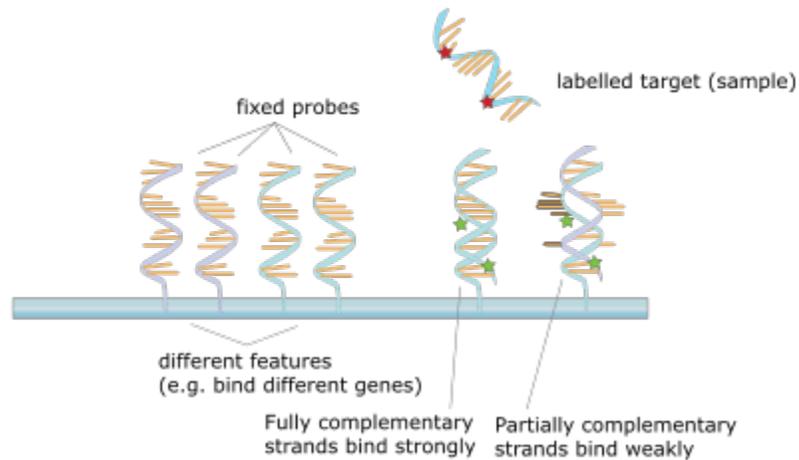
In standard microarrays, the probes are attached via surface engineering to a solid surface by a covalent bond to a chemical matrix (via epoxy-silane, amino-silane, lysine, polyacrylamide or others). The solid surface can be glass or a silicon chip, in which case they are colloquially known as an *Affy chip* when an Affymetrix chip is used. Other microarray platforms, such as Illumina, use microscopic beads, instead of the large solid support. DNA arrays are different from other types of microarray only in that they either measure DNA or use DNA as part of its detection system.

DNA microarrays can be used to measure changes in expression levels, to detect single nucleotide polymorphisms (SNPs), or to genotype or resequence mutant genomes. Microarrays also differ in fabrication, workings, accuracy, efficiency, and cost. Additional factors for microarray experiments are the experimental design and the methods of analyzing the data

History

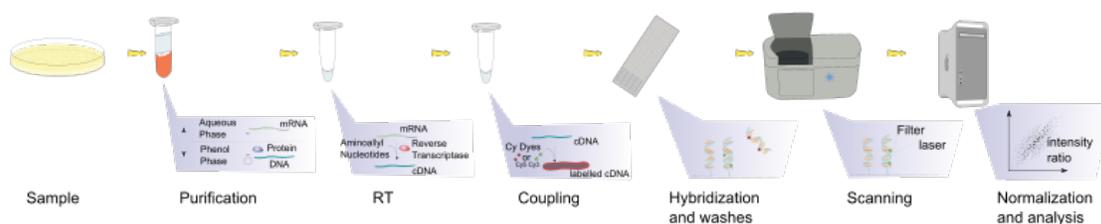
Microarray technology evolved from Southern blotting, where fragmented DNA is attached to a substrate and then probed with a known gene or fragment. Nucleic Acids Res. 1992 Apr 11;20(7):1679-84. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. Maskos U, Southern EM. The first reported use of this approach was the analysis of 378 arrayed lysed bacterial colonies each harboring a different sequence which were assayed in multiple replicas for expression of the genes in multiple normal and tumor tissue (Augenlicht and Kobrin, Cancer Research, 42, 1088–1093, 1982). This was expanded to analysis of more than 4000 human sequences with computer driven scanning and image processing for quantitative analysis of the sequences in human colonic tumors and normal tissue (Augenlicht *et al.*, Cancer Research, 47, 6017-6021, 1987) and then to comparison of colonic tissues at different genetic risk (Augenlicht *et al.*, Proceedings National Academy of Sciences, USA, 88, 3286-3289, 1991). The use of a collection of distinct DNAs in arrays for expression profiling was also described in 1987, and the arrayed DNAs were used to identify genes whose expression is modulated by interferon. These early gene arrays were made by spotting cDNAs onto filter paper with a pin-spotting device. The use of miniaturized microarrays for gene expression profiling was first reported in 1995, and a complete eukaryotic genome (*Saccharomyces cerevisiae*) on a microarray was published in 1997.

Principle



Hybridization of the target to the probe

The core principle behind microarrays is hybridization between two DNA strands, the property of complementary nucleic acid sequences to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs. A high number of complementary base pairs in a nucleotide sequence means tighter non-covalent bonding between the two strands. After washing off of non-specific bonding sequences, only strongly paired strands will remain hybridized. So fluorescently labeled target sequences that bind to a probe sequence generate a signal that depends on the strength of the hybridization determined by the number of paired bases, the hybridization conditions (such as temperature), and washing after hybridization. Total strength of the signal, from a spot (feature), depends upon the amount of target sample binding to the probes present on that spot. Microarrays use relative quantization in which the intensity of a feature is compared to the intensity of the same feature under a different condition, and the identity of the feature is known by its position. An alternative to microarrays is serial analysis of gene expression, where the transcriptome is sequenced allowing an absolute measurement.



The step required in a microarray experiment

Uses and types



Two Affymetrix chips

Many types of array exist and the broadest distinction is whether they are spatially arranged on a surface or on coded beads:

- The traditional solid-phase array is a collection of orderly microscopic "spots", called features, each with a specific probe attached to a solid surface, such as glass, plastic or silicon biochip (commonly known as a *genome chip*, *DNA chip* or *gene array*). Thousands of them can be placed in known locations on a single DNA microarray.
- The alternative bead array is a collection of microscopic polystyrene beads, each with a specific probe and a ratio of two or more dyes, which do not interfere with the fluorescent dyes used on the target sequence.

DNA microarrays can be used to detect DNA (as in comparative genomic hybridization), or detect RNA (most commonly as cDNA after reverse transcription) that may or may not

be translated into proteins. The process of measuring gene expression via cDNA is called expression analysis or expression profiling.

Applications include:

Application or technology	Synopsis
Gene expression profiling	In an mRNA or gene expression profiling experiment the expression levels of thousands of genes are simultaneously monitored to study the effects of certain treatments, diseases, and developmental stages on gene expression. For example, microarray-based gene expression profiling can be used to identify genes whose expression is changed in response to pathogens or other organisms by comparing gene expression in infected to that in uninfected cells or tissues.
Comparative genomic hybridization	Assessing genome content in different cells or closely related organisms.
GeneID	Small microarrays to check IDs of organisms in food and feed (like GMO), mycoplasmas in cell culture, or pathogens for disease detection, mostly combining PCR and microarray technology.
Chromatin immunoprecipitation on Chip	DNA sequences bound to a particular protein can be isolated by immunoprecipitating that protein (ChIP), these fragments can be then hybridized to a microarray (such as a tiling array) allowing the determination of protein binding site occupancy throughout the genome. Example protein to immunoprecipitate are histone modifications (H3K27me3, H3K4me2, H3K9me3, etc.), Polycomb-group protein (PRC2:Suz12, PRC1:YY1) and trithorax-group protein (Ash1) to study the epigenetic landscape or RNA Polymerase II to study the transcription landscape.
DamID	Analogously to ChIP, genomic regions bound by a protein of interest can be isolated and used to probe a microarray to determine binding site occupancy. Unlike ChIP, DamID does not require antibodies but makes use of adenine methylation near the protein's binding sites to selectively amplify those regions, introduced by expressing minute amounts of protein of interest fused to bacterial DNA adenine methyltransferase.
SNP detection	Identifying single nucleotide polymorphism among alleles within or between populations. Several applications of microarrays make use of SNP detection, including Genotyping, forensic analysis, measuring predisposition to disease, identifying drug-candidates, evaluating germline mutations in individuals or somatic mutations in cancers, assessing loss of heterozygosity, or genetic linkage analysis.
Alternative splicing detection	An <i>exon junction array</i> design uses probes specific to the expected or potential splice sites of predicted exons for a gene. It

is of intermediate density, or coverage, to a typical gene expression array (with 1-3 probes per gene) and a genomic tiling array (with hundreds or thousands of probes per gene). It is used to assay the expression of alternative splice forms of a gene. Exon arrays have a different design, employing probes designed to detect each individual exon for known or predicted genes, and can be used for detecting different splicing isoforms.

Fusion genes
microarray

A Fusion gene microarray can detect fusion transcripts, *e.g.* from cancer specimens. The principle behind this is building on the alternative splicing microarrays. The oligo design strategy enables combined measurements of chimeric transcript junctions with exon-wise measurements of individual fusion partners.

Tiling array

Genome tiling arrays consist of overlapping probes designed to densely represent a genomic region of interest, sometimes as large as an entire human chromosome. The purpose is to empirically detect expression of transcripts or alternatively splice forms which may not have been previously known or predicted.

Fabrication

Microarrays can be manufactured in different ways, depending on the number of probes under examination, costs, customization requirements, and the type of scientific question being asked. Arrays may have as few as 10 probes or up to 2.1 million micrometre-scale probes from commercial vendors.

Spotted vs. in situ synthesised arrays



A DNA microarray being printed by a robot at the University of Delaware

Microarrays can be fabricated using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks,

photolithography using dynamic micromirror devices, ink-jet printing, or electrochemistry on microelectrode arrays.

In *spotted microarrays*, the probes are oligonucleotides, cDNA or small fragments of PCR products that correspond to mRNAs. The probes are synthesized prior to deposition on the array surface and are then "spotted" onto glass. A common approach utilizes an array of fine pins or needles controlled by a robotic arm that is dipped into wells containing DNA probes and then depositing each probe at designated locations on the array surface. The resulting "grid" of probes represents the nucleic acid profiles of the prepared probes and is ready to receive complementary cDNA or cRNA "targets" derived from experimental or clinical samples. This technique is used by research scientists around the world to produce "in-house" printed microarrays from their own labs. These arrays may be easily customized for each experiment, because researchers can choose the probes and printing locations on the arrays, synthesize the probes in their own lab (or collaborating facility), and spot the arrays. They can then generate their own labeled samples for hybridization, hybridize the samples to the array, and finally scan the arrays with their own equipment. This provides a relatively low-cost microarray that may be customized for each study, and avoids the costs of purchasing often more expensive commercial arrays that may represent vast numbers of genes that are not of interest to the investigator. Publications exist which indicate in-house spotted microarrays may not provide the same level of sensitivity compared to commercial oligonucleotide arrays, possibly owing to the small batch sizes and reduced printing efficiencies when compared to industrial manufactures of oligo arrays.

In *oligonucleotide microarrays*, the probes are short sequences designed to match parts of the sequence of known or predicted open reading frames. Although oligonucleotide probes are often used in "spotted" microarrays, the term "oligonucleotide array" most often refers to a specific technique of manufacturing. Oligonucleotide arrays are produced by printing short oligonucleotide sequences designed to represent a single gene or family of gene splice-variants by synthesizing this sequence directly onto the array surface instead of depositing intact sequences. Sequences may be longer (60-mer probes such as the Agilent design) or shorter (25-mer probes produced by Affymetrix) depending on the desired purpose; longer probes are more specific to individual target genes, shorter probes may be spotted in higher density across the array and are cheaper to manufacture. One technique used to produce oligonucleotide arrays include photolithographic synthesis (Agilent and Affymetrix) on a silica substrate where light and light-sensitive masking agents are used to "build" a sequence one nucleotide at a time across the entire array. Each applicable probe is selectively "unmasked" prior to bathing the array in a solution of a single nucleotide, then a masking reaction takes place and the next set of probes are unmasked in preparation for a different nucleotide exposure. After many repetitions, the sequences of every probe become fully constructed. More recently, Maskless Array Synthesis from NimbleGen Systems has combined flexibility with large numbers of probes.

Two-channel vs. one-channel detection

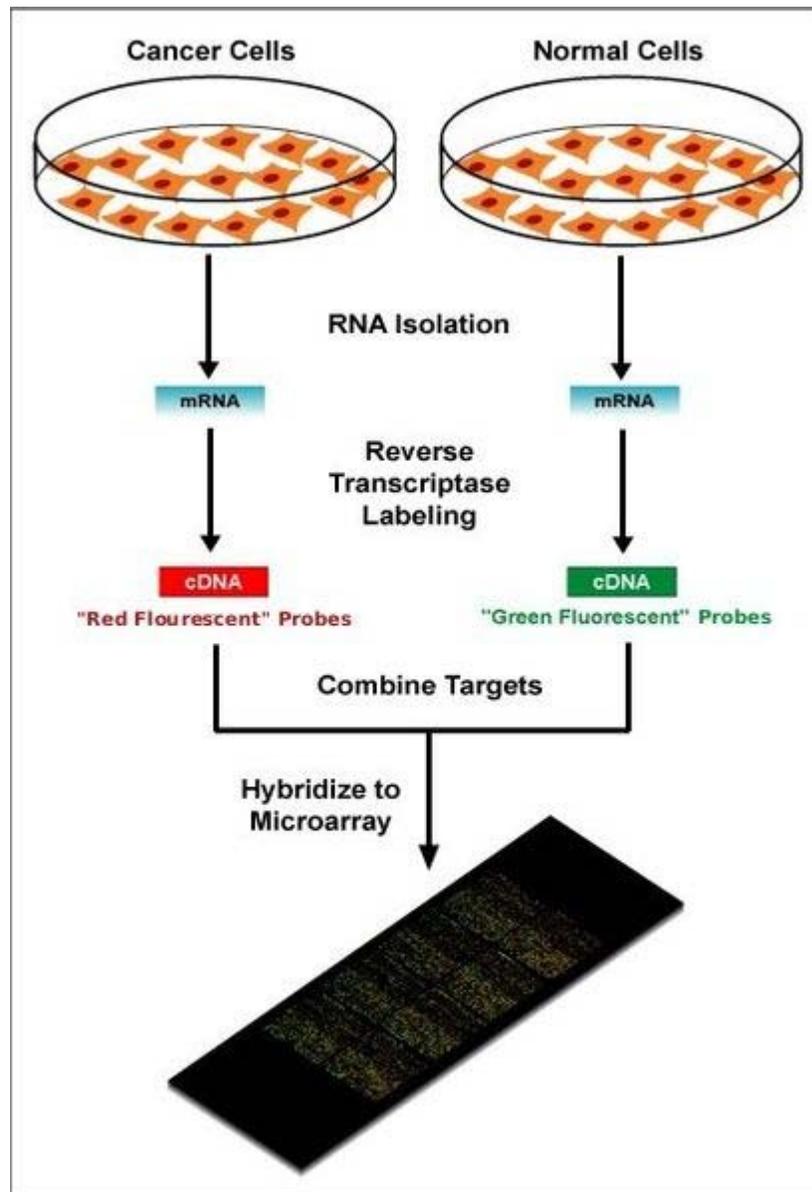


Diagram of typical dual-colour microarray experiment

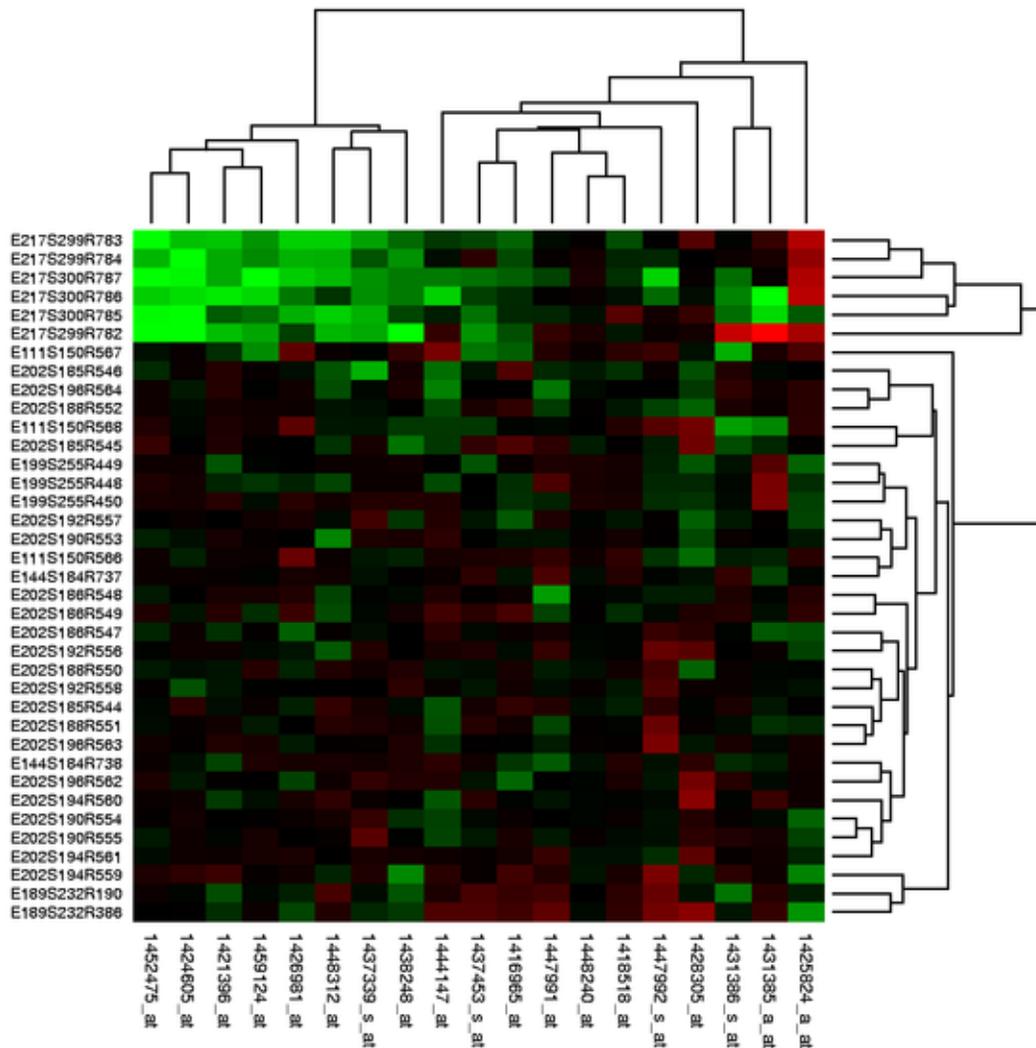
Two-color microarrays or *two-channel microarrays* are typically hybridized with cDNA prepared from two samples to be compared (e.g. diseased tissue versus healthy tissue) and that are labeled with two different fluorophores. Fluorescent dyes commonly used for cDNA labeling include Cy3, which has a fluorescence emission wavelength of 570 nm (corresponding to the green part of the light spectrum), and Cy5 with a fluorescence emission wavelength of 670 nm (corresponding to the red part of the light spectrum). The two Cy-labeled cDNA samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each

fluorophore may then be used in ratio-based analysis to identify up-regulated and down-regulated genes.

Oligonucleotide microarrays often carry control probes designed to hybridize with RNA spike-ins. The degree of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes. Although absolute levels of gene expression may be determined in the two-color array in rare instances, the relative differences in expression among different spots within a sample and between samples is the preferred method of data analysis for the two-color system. Examples of providers for such microarrays includes Agilent with their Dual-Mode platform, Eppendorf with their DualChip platform for colorimetric Silverquant labeling, and TeleChem International with Arrayit.

In *single-channel microarrays* or *one-color microarrays*, the arrays provide intensity data for each probe or probe set indicating a relative level of hybridization with the labeled target. However, they do not truly indicate abundance levels of a gene but rather relative abundance when compared to other samples or conditions when processed in the same experiment. Each RNA molecule encounters protocol and batch-specific bias during amplification, labeling, and hybridization phases of the experiment making comparisons between genes for the same microarray uninformative. The comparison of two conditions for the same gene requires two separate single-dye hybridizations. Several popular single-channel systems are the Affymetrix "Gene Chip", Illumina "Bead Chip", Agilent single-channel arrays, the Applied Microarrays "CodeLink" arrays, and the Eppendorf "DualChip & Silverquant". One strength of the single-dye system lies in the fact that an aberrant sample cannot affect the raw data derived from other samples, because each array chip is exposed to only one sample (as opposed to a two-color system in which a single low-quality sample may drastically impinge on overall data precision even if the other sample was of high quality). Another benefit is that data are more easily compared to arrays from different experiments so long as batch effects have been accounted for. A drawback to the one-color system is that, when compared to the two-color system, twice as many microarrays are needed to compare samples within an experiment.

Microarrays and bioinformatics



Gene expression values from microarray experiments can be represented as heat maps to visualize the result of data analysis.

The advent of inexpensive microarray experiments created several specific bioinformatics challenges:

- the multiple levels of replication in experimental design (Experimental design)
- the number of platforms and independent groups and data format (Standardization)
- the treatment of the data (Statistical analysis)
- accuracy and precision (Relation between probe and gene)
- the sheer volume of data and the ability to share it (Data warehousing)

Experimental design

Due to the biological complexity of gene expression, the considerations of experimental design that are discussed in the expression profiling article are of critical importance if statistically and biologically valid conclusions are to be drawn from the data.

There are three main elements to consider when designing a microarray experiment. First, replication of the biological samples is essential for drawing conclusions from the experiment. Second, technical replicates (two RNA samples obtained from each experimental unit) help to ensure precision and allow for testing differences within treatment groups. The technical replicates may be two independent RNA extractions or two aliquots of the same extraction. Third, spots of each cDNA clone or oligonucleotide are present as replicates (at least duplicates) on the microarray slide, to provide a measure of technical precision in each hybridization. It is critical that information about the sample preparation and handling is discussed, in order to help identify the independent units in the experiment and to avoid inflated estimates of statistical significance.

Standardization

Microarray data is difficult to exchange due to the lack of standardization in platform fabrication, assay protocols, and analysis methods. This presents an interoperability problem in bioinformatics. Various grass-roots open-source projects are trying to ease the exchange and analysis of data produced with non-proprietary chips:

- For example, the "Minimum Information About a Microarray Experiment" (MIAME) checklist helps define the level of detail that should exist and is being adopted by many journals as a requirement for the submission of papers incorporating microarray results. But MIAME does not describe the format for the information, so while many formats can support the MIAME requirements, as of 2007 no format permits verification of complete semantic compliance.
- The "MicroArray Quality Control (MAQC) Project" is being conducted by the US Food and Drug Administration (FDA) to develop standards and quality control metrics which will eventually allow the use of MicroArray data in drug discovery, clinical practice and regulatory decision-making.
- The MGED Society has developed standards for the representation of gene expression experiment results and relevant annotations.

Statistical analysis

Microarray data sets are commonly very large, and analytical precision is influenced by a number of variables. Statistical challenges include taking into account effects of background noise and appropriate normalization of the data. Normalization methods may be suited to specific platforms and, in the case of commercial platforms, the analysis may be proprietary. Algorithms that affect statistical analysis include:

- Image analysis: gridding, spot recognition of the scanned image (segmentation algorithm), removal or marking of poor-quality and low-intensity features (called *flagging*).
- Data processing: background subtraction (based on global or local background), determination of spot intensities and intensity ratios, visualisation of data, and log-transformation of ratios, global or local normalization of intensity ratios.
- Identification of statistically significant changes: t-test, ANOVA, Bayesian method Mann–Whitney test methods tailored to microarray data sets, which take into account multiple comparisons or cluster analysis. These methods assess statistical power based on the variation present in the data and the number of experimental replicates, and can help minimize Type I and type II errors in the analyses.
- Network-based methods: Statistical methods that take the underlying structure of gene networks into account, representing either associative or causative interactions or dependencies among gene products.

Microarray data may require further processing aimed at reducing the dimensionality of the data to aid comprehension and more focused analysis. Other methods permit analysis of data consisting of a low number of biological or technical replicates; for example, the Local Pooled Error (LPE) test pools standard deviations of genes with similar expression levels in an effort to compensate for insufficient replication.

Relation between probe and gene

The relation between a probe and the mRNA that it is expected to detect is not trivial. Some mRNAs may cross-hybridize probes in the array that are supposed to detect another mRNA. In addition, mRNAs may experience amplification bias that is sequence or molecule-specific. Thirdly, probes that are designed to detect the mRNA of a particular gene may be relying on genomic EST information that is incorrectly associated with that gene.

Data warehousing

Microarray data was found to be more useful when compared to other similar datasets. The sheer volume (in bytes), specialized formats (such as MIAME), and curation efforts associated with the datasets require specialized databases to store the data.

Chapter- 6

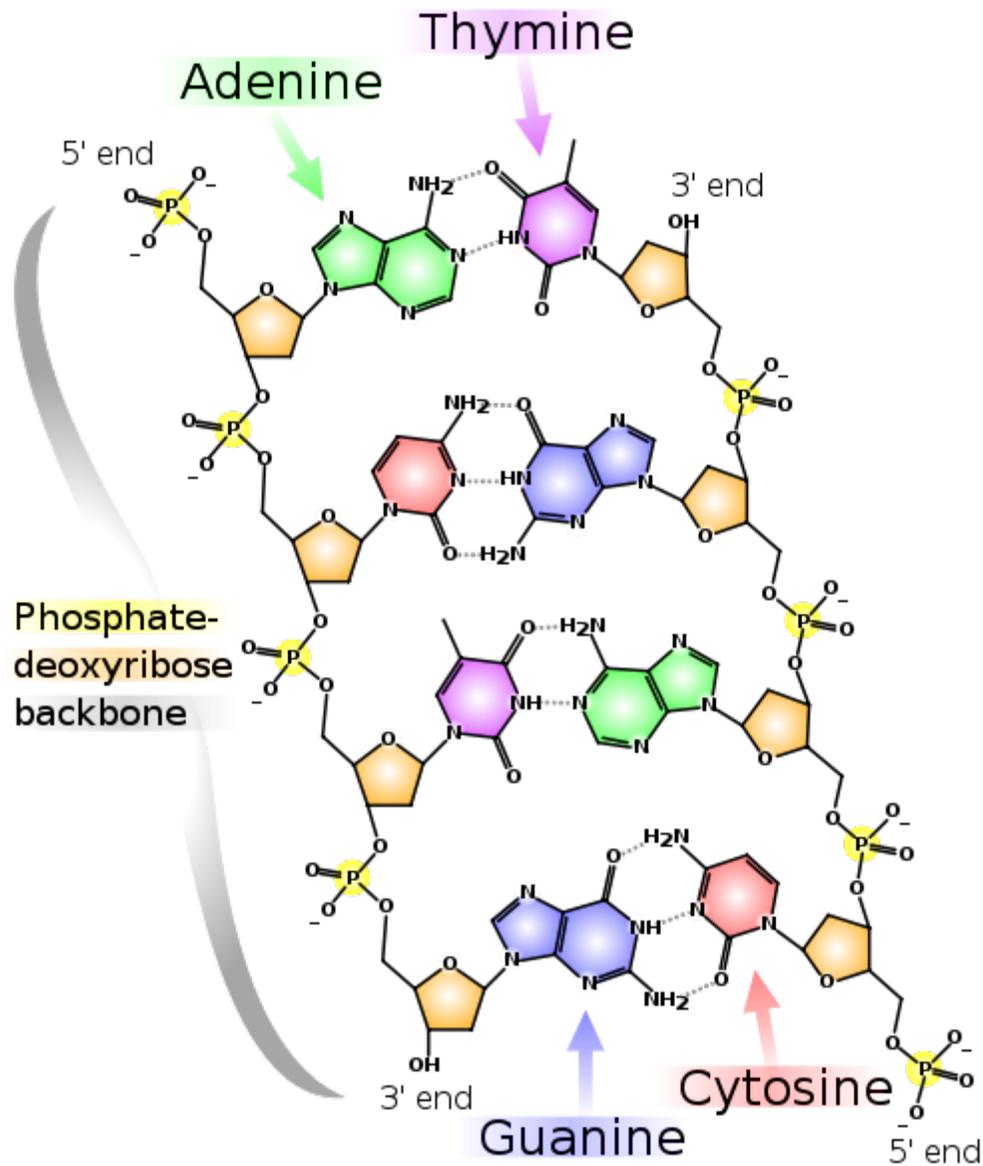
DNA Nanotechnology

DNA nanotechnology is a branch of nanotechnology which uses the molecular recognition properties of DNA and other nucleic acids to create designed, artificial structures out of DNA for technological purposes. In this field, DNA is used as a structural material rather than as a carrier of genetic information, making it an example of bionanotechnology. DNA nanotechnology has applications in molecular self-assembly and in DNA computing.

Although DNA is usually considered in the context of molecular biology as the carrier of genetic information in living cells, DNA nanotechnology considers DNA solely as a chemical and as a material, and is usually pursued outside of any biological context. DNA nanotechnology makes use of the fact that, due to the specificity of Watson-Crick base pairing, only portions of the strands which are complementary to each other will bind to each other to form duplex DNA. DNA nanotechnology attempts to rationally design sets of DNA strands so that desired portions of each strand will assemble in the correct positions to form some desired target structure.

Although the field is usually called DNA nanotechnology, its principles apply equally well to other nucleic acids such as RNA and PNA, and structures incorporating these have been made. For this reason the field is occasionally referred to as **nucleic acid nanotechnology**.

Fundamental concepts



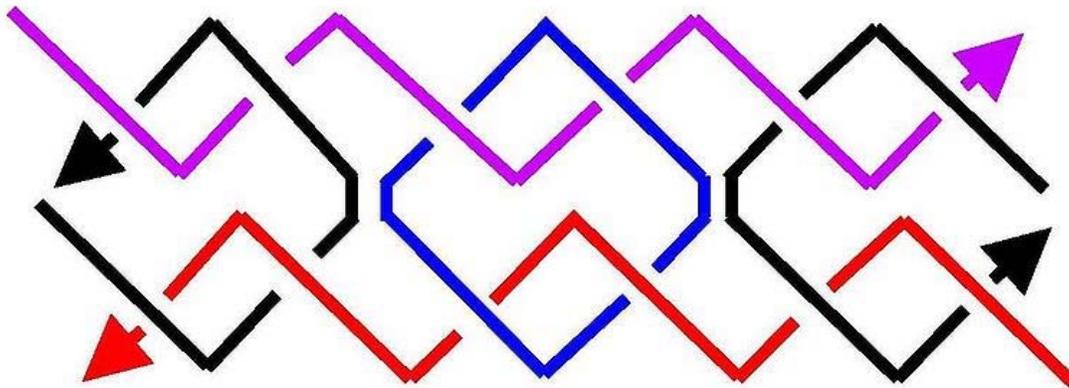
Chemical structure of DNA. Nucleic acid double helices will only form between two strands of complementary sequences, where the bases are matched into only A-T or G-C pairs.

DNA nanotechnology creates complex structures out of nucleic acids by making use of the specificity of base pairing in nucleic acid molecules. The structure of a nucleic acid molecule consists of a sequence of nucleotides, distinguished by which nucleobase they contain. In DNA, the four bases used are adenine (A), cytosine (C), guanine (G), and thymine (T). Nucleic acids have the property that two molecules will bind to each other to form a double helix only if the two sequences are complementary, meaning that they form matching sequences of base pairs, with A's only binding to T's, and C's only to G's.

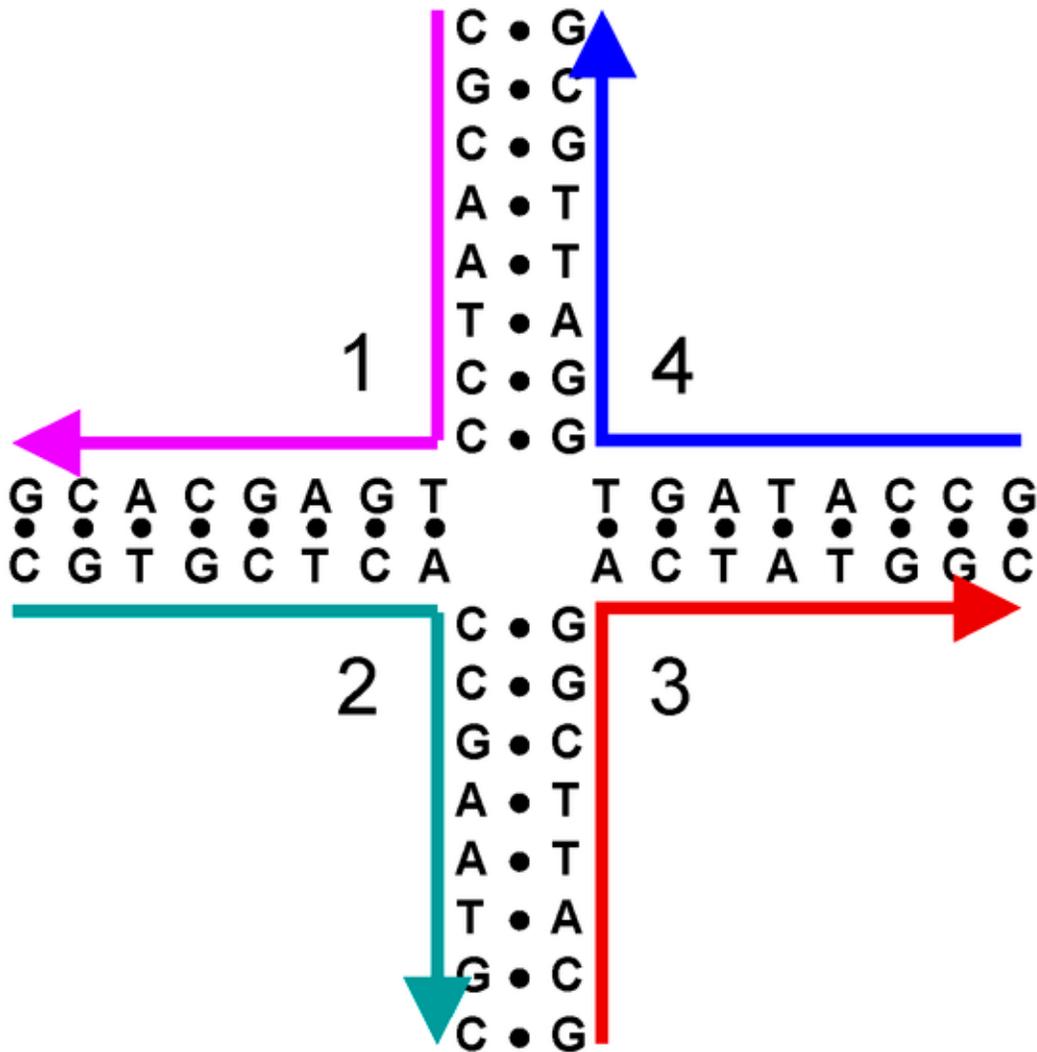
Because the formation of correctly matched base pairs is energetically favorable, nucleic acid strands are expected in most cases to bind to each other in the conformation that maximizes the number of correctly paired bases. This property, that the sequence determines the pattern of binding and the overall structure, is used by the field of DNA nanotechnology in that sequences are rationally designed so that a desired structure is favored to form.

Nearly all structures in DNA nanotechnology make use of branched DNA structures containing junctions, as opposed to most biological DNA which exists in a linear double helix form. One of the simplest branched structures, and the first made, is a four-arm junction which can be made using four individual DNA strands which are complementary to each other in the correct pattern. Unlike in natural Holliday junctions, in the artificial immobile four-arm junction shown below, the base sequence of each arm is different, meaning that the junction point is fixed in a certain position.

Junctions can be used in more complex molecules. One of the more widely-used of these is the "double-crossover" or DX motif. A DX molecule can be thought of as two DNA duplexes positioned parallel to each other, with two crossover points where strands cross from one duplex into the other. Each junction point is itself topologically a four-arm junction. This molecule has the advantage that the junction points are now constrained to a single orientation as opposed to being flexible as in the four-arm junction. This makes the DX motif suitable as a structural building block for larger DNA complexes.



A double-crossover (DX) molecule. This molecule consists of five DNA single strands which form two double-helical domains, on the left and the right in this image. There are two crossover points where the strands cross from one domain into the other.



These four strands associate into a DNA four-arm junction because this structure maximizes the number of correct base pairs, with A's matched to T's and C's matched to G's.

Design

DNA nanostructures must be designed so that they will assemble into the desired structures. This includes both the design of secondary structure, deciding which parts of which nucleic acid molecules should bind to each other, and primary structure, the specification of the identity of each individual base.

Structural design

The first step in designing a nucleic acid nanostructure is to decide how a given structure should be represented by a specific arrangement of nucleic acid strands. This design step thus determines the secondary structure of the nucleic acid complex which will assemble into the desired shape. There are several approaches which have been demonstrated:

- **Sequence symmetry minimization.** Most design in DNA nanotechnology focuses on designing sequences so that the target structure is a thermodynamic minimum, and mis-assembled structures have higher energies and are thus disfavored.
- **Folding structures.** An alternative to the tile-based approach, two-dimensional DNA structures can be made from a single, long DNA strand of arbitrary sequence which is folded into the desired shape by using shorter, "staple" strands. This allows the creation of two-dimensional shapes at the nanoscale using DNA. Demonstrated designs have included the smiley face and a coarse map of North America. DNA origami was the cover story of *Nature* on March 15, 2006.
- **Kinetic assembly.** Recently, there has been interest in controlling the kinetics of DNA self-assembly, so that transient dynamics can also be programmed into the assembly. Such a method also has the advantage of proceeding isothermally and thus not requiring a thermal annealing step required by solely thermodynamic approaches.

Sequence design

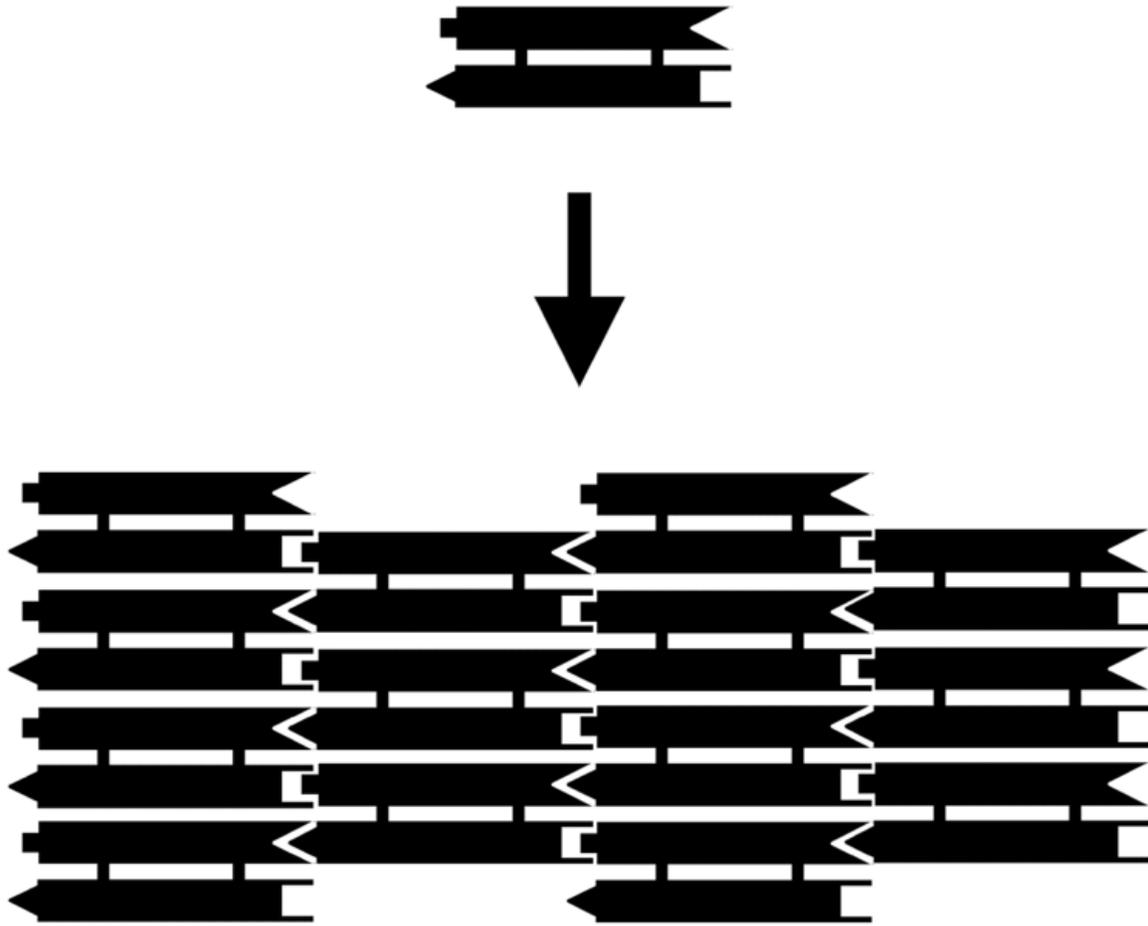
After any of the above approaches are used to design the secondary structure of a target molecule, an actual sequence of nucleotides must be devised which will form into the desired structure. Nucleic acid design is the process of generating a set of nucleic acid base sequences that will associate into a desired conformation (see, for example, RNA structure). Nucleic acid design is central to the field of DNA nanotechnology.

Nucleic acid design has similar goals to protein design: in both, the sequence of monomers is designed to favor the desired folded or associated structure and to disfavor alternate structures. Nucleic acid design has the advantage of being a much computationally simpler problem, since the simplicity of Watson-Crick base pairing rules leads to simple heuristic methods which yield experimentally robust designs. However, nucleic acid structures are less versatile than proteins in their functionality.

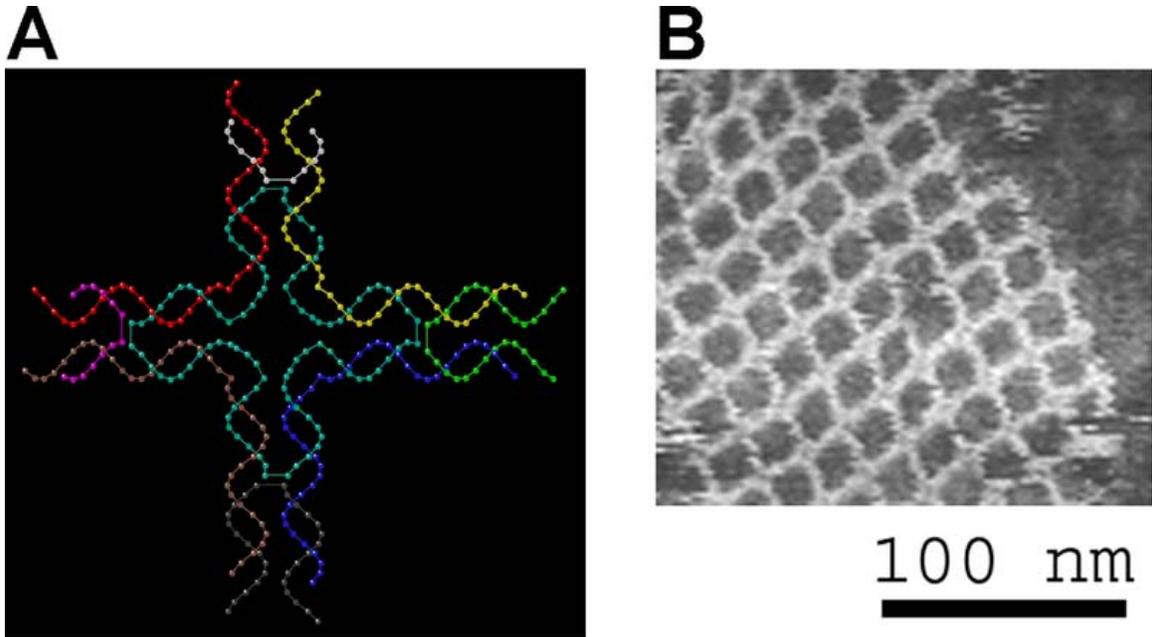
Types of structures

Many structures made from DNA have been synthesized and characterized.

Periodic lattices



Assembly of a DX array. Each bar represents a double-helical domain of DNA, with the shapes representing complimentary sticky ends. The DX molecule at top will combine into the two-dimensional DNA array shown at bottom. This is an example of the tile-based strategy for designing DNA nanostructures.



Left, a model of a DNA tile used to make a two-dimensional periodic lattice. Right, an atomic force micrograph of the assembled lattice.

The earliest method for creating DNA nanostructures was to construct them out of smaller discrete units. This method has the advantage of being able to conceptually separate the stronger interactions which form each tile from the assembly of the larger complete structure. It is often used to make periodic lattices, but can also be used to implement algorithmic self-assembly, making them one platform for DNA computing.

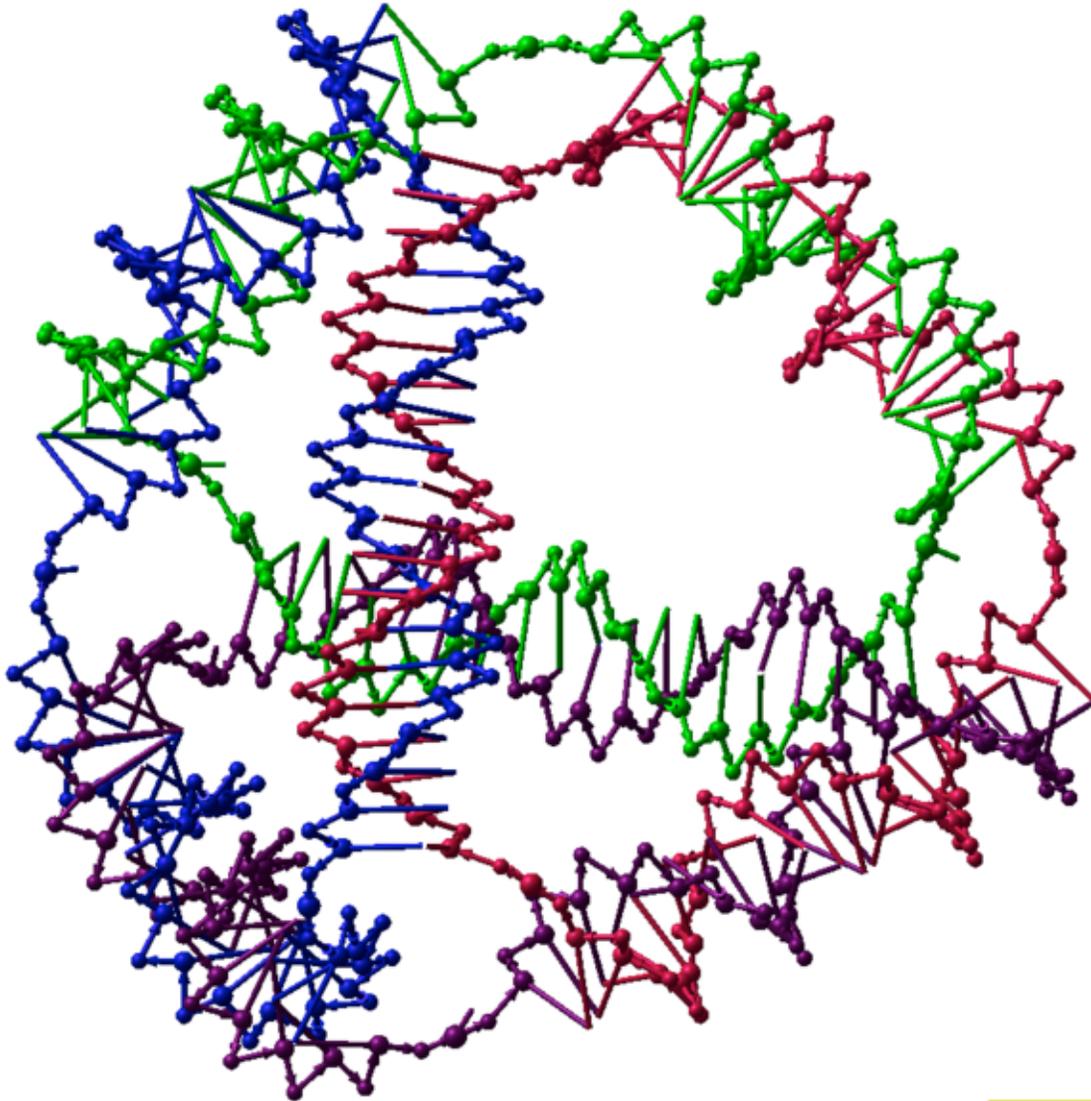
DX, or Double Crossover, molecules can be equipped with sticky ends in order to combine them into a two-dimensional periodic lattice. Each DX molecule has four termini, one at each end of the two double-helical domains, and these can be equipped with sticky ends that program them to combine into a specific pattern. More than one type of DX can be used which can be made to arrange in rows or any other tessellated pattern. They thus form extended flat sheets which are essentially two-dimensional crystals of DNA.

Two-dimensional arrays have been made out of other motifs as well, including the Holliday junction rhombus array as well as various DX-based arrays in the shapes of triangles and hexagons.

Creating three-dimensional lattices out of DNA was the earliest goal of DNA nanotechnology, but proved to be one of the most difficult to realize. Success in constructing three-dimensional DNA lattices was finally reported in 2009 using a motif based on the concept of tensegrity, a balance between tension and compression forces.

Nanotubes

In addition to flat sheets, DNA arrays have been made to form hollow nanotubes of 4-20 nm diameter. These DNA nanotubes are somewhat similar in size and shape to carbon nanotubes, but the carbon nanotubes are stronger and better conductors, whereas the DNA nanotubes are more easily modified and connected to other structures.



A model of a DNA tetrahedron described in Goodman, 2005. Each edge of the tetrahedron is a 20 base pair DNA duplex, and each vertex is a three-arm junction.

Polyhedra

A number of three-dimensional DNA molecules have been made which have the connectivity of a polyhedron such as an octahedron or cube. In other words, the DNA duplexes trace the edges of a polyhedron with a DNA junction at each vertex.

The earliest demonstrations of DNA polyhedra involved multiple ligations and solid-phase synthesis steps to create catenated polyhedra. More recent work has yielded polyhedra whose synthesis is much easier. These include a DNA octahedron made from a long single strand designed to fold into the correct conformation, as well as a tetrahedron which can be produced from four DNA strands in a single step.

Arbitrary shapes

DNA structures with solid faces have also been constructed, using the DNA origami method. These can be programmed to open and release their cargo in response to a stimulus, making them potentially useful as programmable molecular cages.

Functional nucleic acid nanostructures

DNA nanotechnology focuses on creating molecules with designed functionalities as well as structures. Many classes of functional systems have been demonstrated.

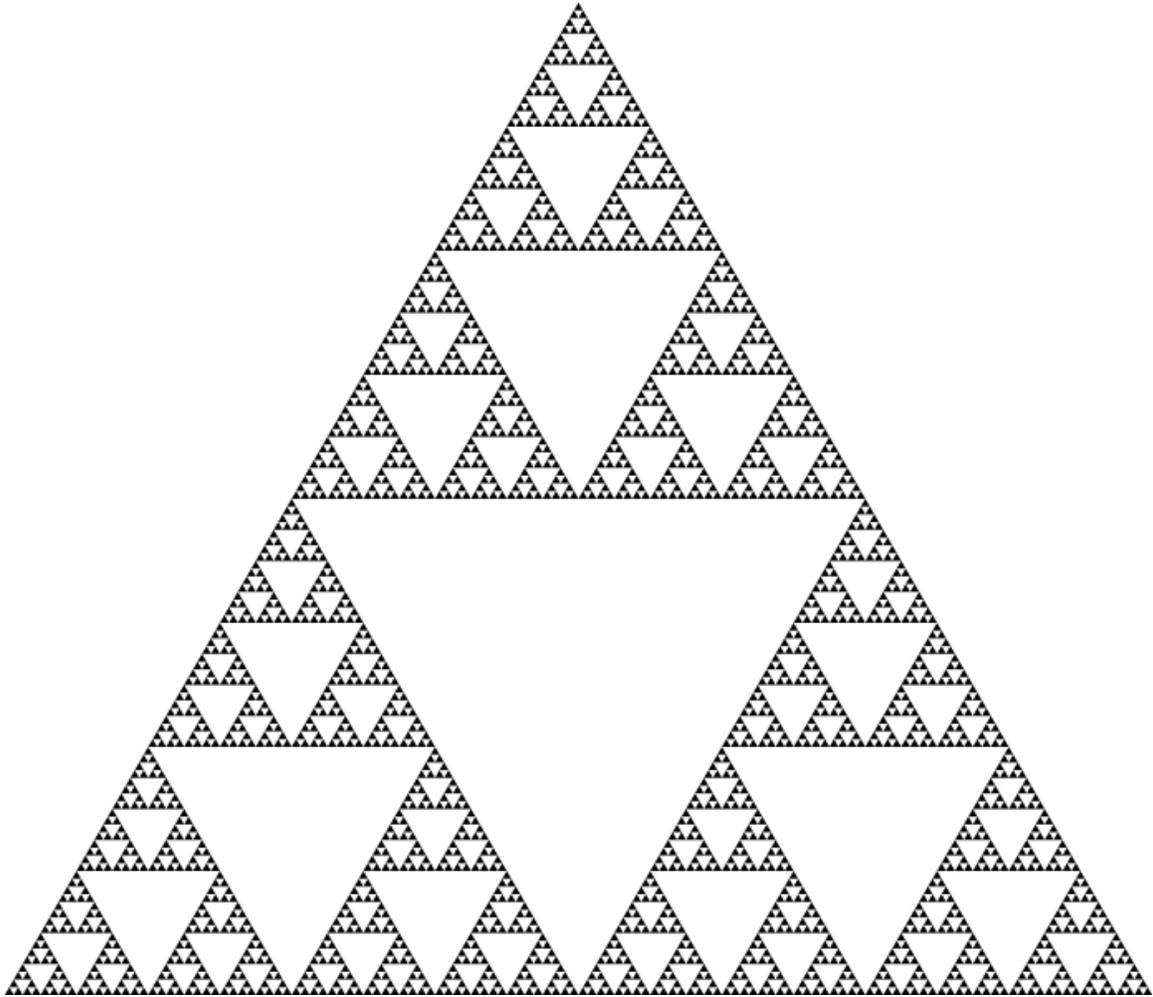
Nanoarchitecture

The idea of using DNA arrays to template the assembly of other functional molecules was first suggested by Nadrian Seeman in 1987, but only recently has progress been made in reducing these kinds of schemes to practice. In 2006, researchers covalently attached gold nanoparticles to a DX-based tile and showed that self-assembly of the DNA structures also assembled the nanoparticles hosted on them. A non-covalent hosting scheme was shown in 2007, using Dervan polyamides on a DX array to arrange streptavidin proteins on specific kinds of tiles on the DNA array.

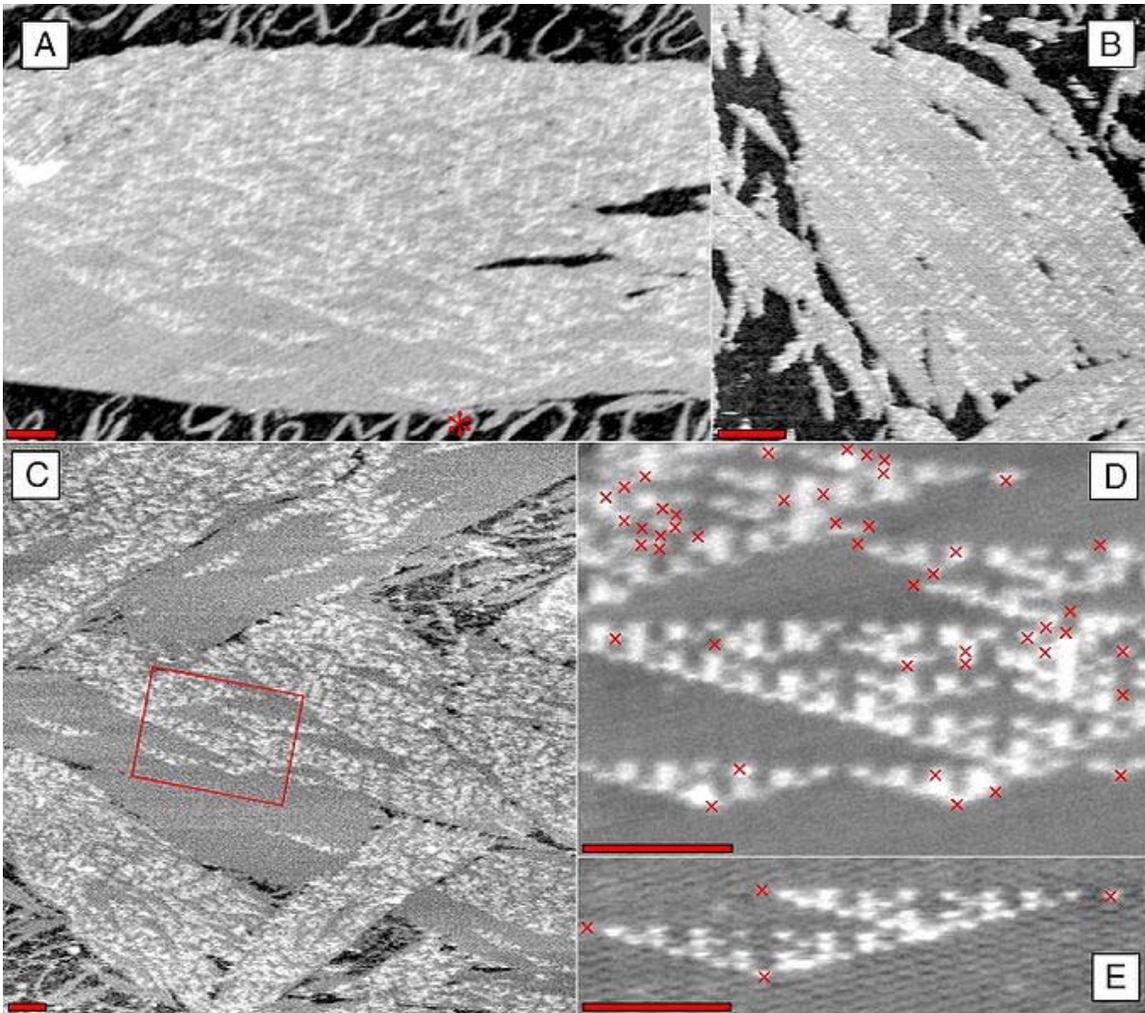
Previously in 2006, Dwyer and LaBean demonstrated the letters "D" "N" and "A" created on a 4x4 DX array using streptavidin. In 2007, a hierarchical assembly based on this approach was also demonstrated that scales to larger arrays (8X8 and 8.96 MD).

There has also been interest in using DNA nanotechnology to assemble molecular electronics devices. To this end, DNA has been used to assemble single walled carbon nanotubes into field-effect transistors.

Algorithmic self-assembly



The Sierpinski gasket



DNA arrays that display a representation of the Sierpinski gasket on their surfaces.

DNA nanotechnology has been applied to the related field of DNA computing. The DX tiles can have their sticky end sequences chosen so that they act as Wang tiles, allowing them to perform computation. A DX array has been demonstrated whose assembly encodes an XOR operation; this allows the DNA array to implement a cellular automaton which generates a fractal called the Sierpinski gasket. This shows that computation can be incorporated into the assembly of DNA arrays, increasing its scope beyond simple periodic arrays.

Note that DNA computing overlaps with, but is distinct from, DNA nanotechnology. The latter uses the specificity of Watson-Crick basepairing to make novel structures out of DNA. These structures can be used for DNA computing, but they do not have to be. Additionally, DNA computing can be done without using the types of molecules made possible by DNA nanotechnology.

DNA nanomechanical devices

DNA complexes have been made which change their conformation upon some stimulus. These are intended to have applications in nanorobotics. One of the first such devices, called "molecular tweezers," changes from an open to a closed state based upon the presence of control strands.

DNA machines have also been made which show a twisting motion. One of these makes use of the transition between the B-DNA and Z-DNA forms to respond to a change in buffer conditions. Another relies on the presence of control strands to switch from a paranemic-crossover (PX) conformation to a double-junction (JX2) conformation.

Materials and methods

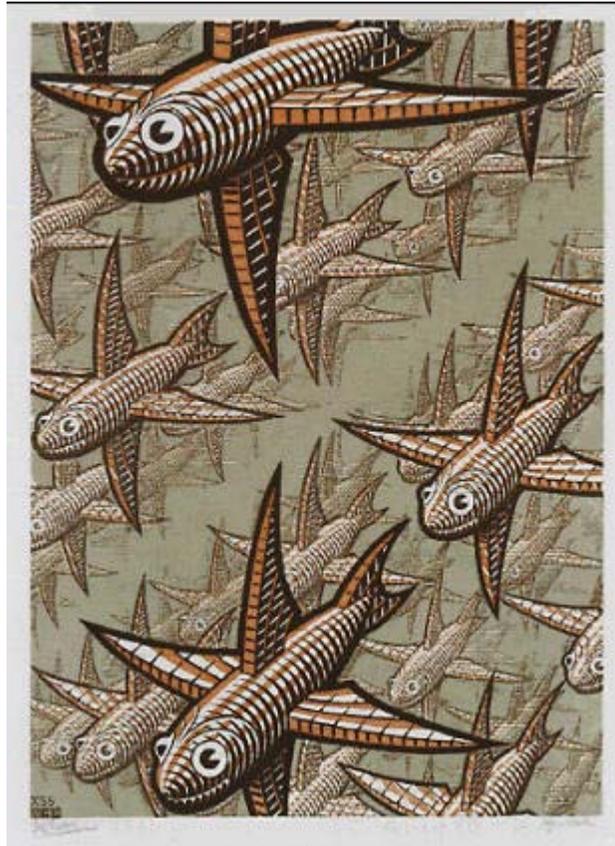
DNA of custom sequence is readily available through oligonucleotide synthesis. This process is usually automated by using a DNA synthesizing machine, and custom DNA is commercially available from many vendors.

The sequences of the individual DNA strands which make up the target structure are designed computationally. Molecular modelling and thermodynamic modelling are sometimes used to optimize the DNA sequences.

The DNA molecules created by DNA nanotechnology are usually characterized by gel electrophoresis, which provides information about the size and shape of DNA molecules, indicating whether they have formed properly. Fluorescent labelling and Fluorescence resonance energy transfer (FRET) are also used to characterize the structure of the molecules.

DNA structures can be directly imaged by atomic force microscopy, which images structures deposited on a flat surface. This method is well-suited to extended two-dimensional structures, but is less useful for discrete three-dimensional structures. For these latter structures transmission electron microscopy and cryo-electron microscopy are important methods. Extended three-dimensional lattices are analyzed by X-ray crystallography. Kinetics of DNA self assembly can be studied by real time techniques such as dual polarisation interferometry and QCMD.

History



The M. C. Escher woodcut *Depth* (pictured) inspired Nadrian Seeman to consider using three-dimensional lattices of DNA to orient hard-to-crystallize molecules. This led to the beginning of the field of DNA nanotechnology.

The concept of DNA nanotechnology was invented by Nadrian Seeman in the early 1980s. Seeman was originally concerned with using a three-dimensional DNA lattice to orient target molecules, which would simplify their crystallographic study by eliminating the difficult process of obtaining pure crystals. This idea had reportedly come to him in fall 1980, after realizing the similarity between the M. C. Escher woodcut *Depth* and an array of DNA six-arm junctions. To this end, Seeman's laboratory in 1991 published the synthesis of a cube made of DNA, the first three-dimensional nanoscale object, for which he received the 1995 Feynman Prize in Nanotechnology, which was followed by a DNA truncated octahedron. However, it soon became clear that these molecules, polygonal shapes with flexible junctions as their vertices, were not rigid enough to form extended three-dimensional lattices.

Seeman developed the more rigid double-crossover (DX) motif, and in collaboration with Erik Winfree, in 1998 published the creation of two-dimensional lattices of DX tiles. These tile-based structures had the advantage that they provided the capability to implement DNA computing, which was demonstrated by Winfree and Paul Rothemund in 2004, and for which they shared the 2006 Feynman Prize in Nanotechnology.

The field has continued to branch out. The first DNA nanomachine—a motif which changes its structure in response to an input—was demonstrated in 1999. Nanoarchitecture, first proposed by Seeman in 1987, was beginning to be demonstrated by 2006. Also in 2006, Rothemund first demonstrated the new DNA origami technique for easily and robustly creating folded DNA molecules of any shape. In 2009, Seeman published the synthesis of a three-dimensional lattice, nearly thirty years after he had set out to do so.

Chapter- 7

Chromatin

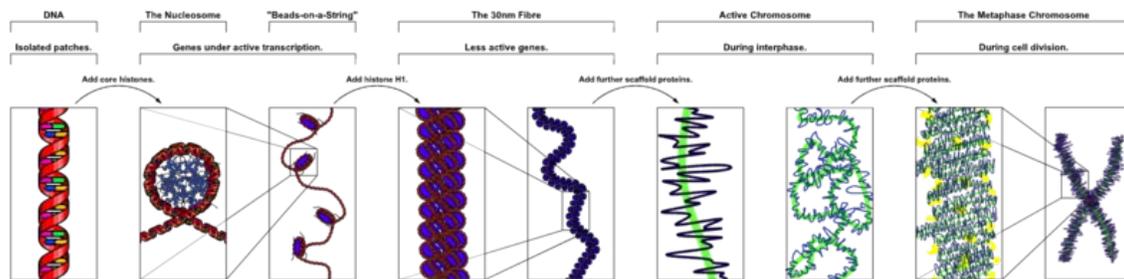


Fig. 1: The major structures in DNA compaction; DNA, the nucleosome, the 10nm "beads-on-a-string" fibre, the 30nm fibre and the metaphase chromosome.

Chromatin is the combination of DNA, histone, and other proteins that make up chromosomes. It is found inside the nuclear envelope of eukaryotic cells. It is divided between heterochromatin (condensed) and euchromatin (extended) forms. The functions of chromatin are to package DNA into a smaller volume to fit in the cell, to strengthen the DNA to allow mitosis and meiosis, and to control gene expression and DNA replication. Changes in chromatin structure are affected by chemical modifications of histone proteins, such as methylation and acetylation, and by other DNA-binding proteins.

Basic structure

In simple terms, there are three levels of chromatin organization (Fig. 1):

1. DNA wraps around histone proteins forming nucleosomes - the "beads on a string" structure.
2. A 30 nm condensed chromatin fiber consisting of nucleosome arrays in their most compact form.
3. Higher-level DNA packaging into the metaphase chromosome.

These structures do not occur in prokaryotic cells. Examples of cells with more extreme packaging are spermatozoa and avian red blood cells.

During spermiogenesis, the spermatid's chromatin is remodelled into a more spaced-packaged, widened, almost crystal-like structure. This process is associated with the cessation of transcription and involves nuclear protein exchange. The histones are mostly displaced, and replaced by protamines (small, arginine-rich proteins).

It should also be noted that, during mitosis, while most of the chromatin is tightly compacted, there are small regions that are not as tightly compacted. These regions often correspond to promoter regions of genes that were active in that cell type prior to entry into mitosis. The lack of compaction of these regions is called bookmarking, which is an epigenetic mechanism believed to be important for transmitting to daughter cells the "memory" of which genes were active prior to entry into mitosis. This bookmarking mechanism is needed to help transmit this memory because transcription ceases during mitosis.

During interphase

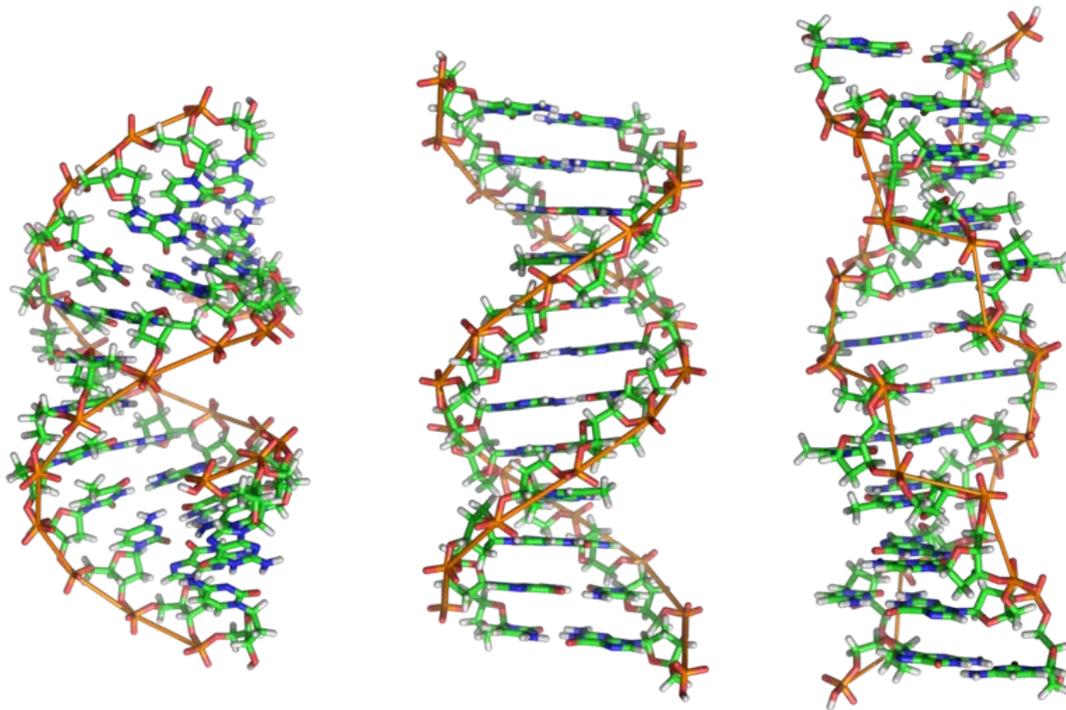
The structure of chromatin during interphase is optimised to allow easy access of transcription and DNA repair factors to the DNA while compacting the DNA into the nucleus. The structure varies depending on the access required to the DNA. Genes that require regular access by RNA polymerase require the looser structure provided by euchromatin.

Change in structure

Chromatin undergoes various forms of change in its structure. Histone proteins, the foundation blocks of chromatin, are modified by various post-translational modification to alter DNA packing. Acetylation results in the loosening of chromatin and lends itself to replication and transcription. When certain residues are methylated they hold DNA together strongly and restrict access to various enzymes. A recent study showed that there is a bivalent structure present in the chromatin: methylated lysine residues at location 4 and 27 on histone 3. It is thought that this may be involved in development; there is more methylation of lysine 27 in embryonic cells than in differentiated cells, whereas lysine 4 methylation positively regulates transcription by recruiting nucleosome remodeling enzymes and histone acetylases.

Polycomb-group proteins play a role in regulating genes through modulation of chromatin structure.

DNA structure

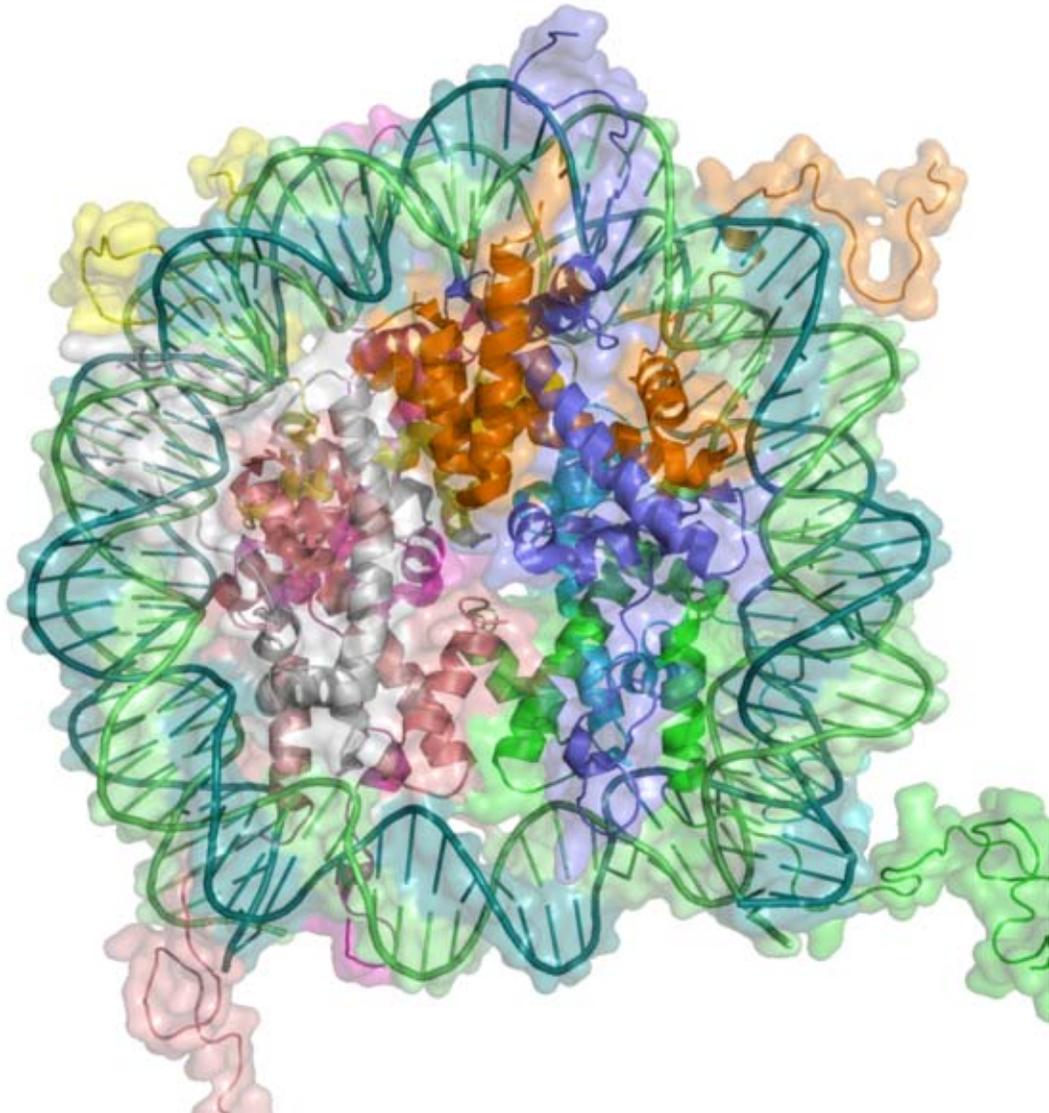


The structures of A-, B- and Z-DNA

The vast majority of DNA within the cell is the normal DNA structure. However in nature DNA can form three structures, A-, B- and Z-DNA. A and B chromosomes are very similar, forming right-handed helices, while Z-DNA is a more unusual left-handed helix with a zig-zag phosphate backbone. Z-DNA is thought to play a specific role in chromatin structure and transcription because of the properties of the junction between B- and Z-DNA.

At the junction of B- and Z-DNA one pair of bases is flipped out from normal bonding. These play a dual role of a site of recognition by many proteins and as a sink for torsional stress from RNA polymerase or nucleosome binding.

The nucleosome and "beads-on-a-string"



A cartoon representation of the nucleosome structure.

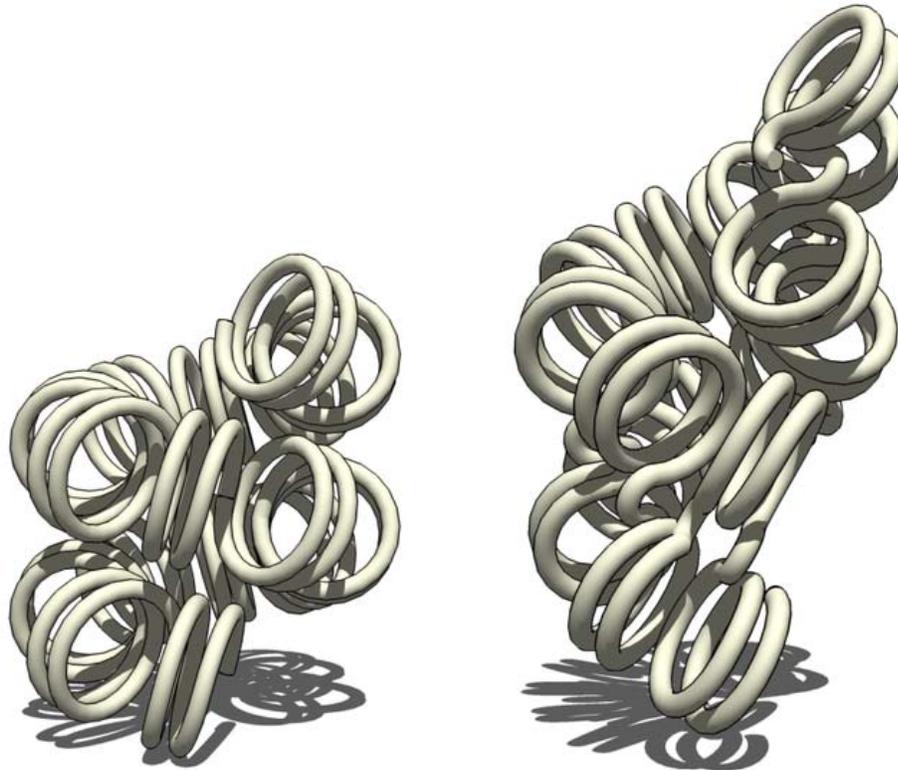
The basic repeat element of chromatin is the nucleosome, interconnected by sections of linker DNA, a far shorter arrangement than pure DNA in solution.

In addition to the core histones, there is the linker histone, H1, which contacts the exit/entry of the DNA strand on the nucleosome. The nucleosome core particle, together with histone H1, is known as a chromatosome. Nucleosomes, with about 20 to 60 base pairs of linker DNA, can form, under non-physiological conditions, an approximately 10 nm "beads-on-a-string" fibre. (Fig. 1-2).

The nucleosomes bind DNA non-specifically, as required by their function in general DNA packaging. There are, however, large DNA sequence preferences that govern

nucleosome positioning. This is due primarily to the varying physical properties of different DNA sequences: For instance, adenosine and thymine are more favorably compressed into the inner minor grooves. This means nucleosomes can bind preferentially at one position approximately every 10 base pairs (the helical repeat of DNA)- where the DNA is rotated to maximise the number of A and T bases that will lie in the inner minor groove.

30 nm chromatin fibre



Two proposed structures of the 30nm chromatin filament.

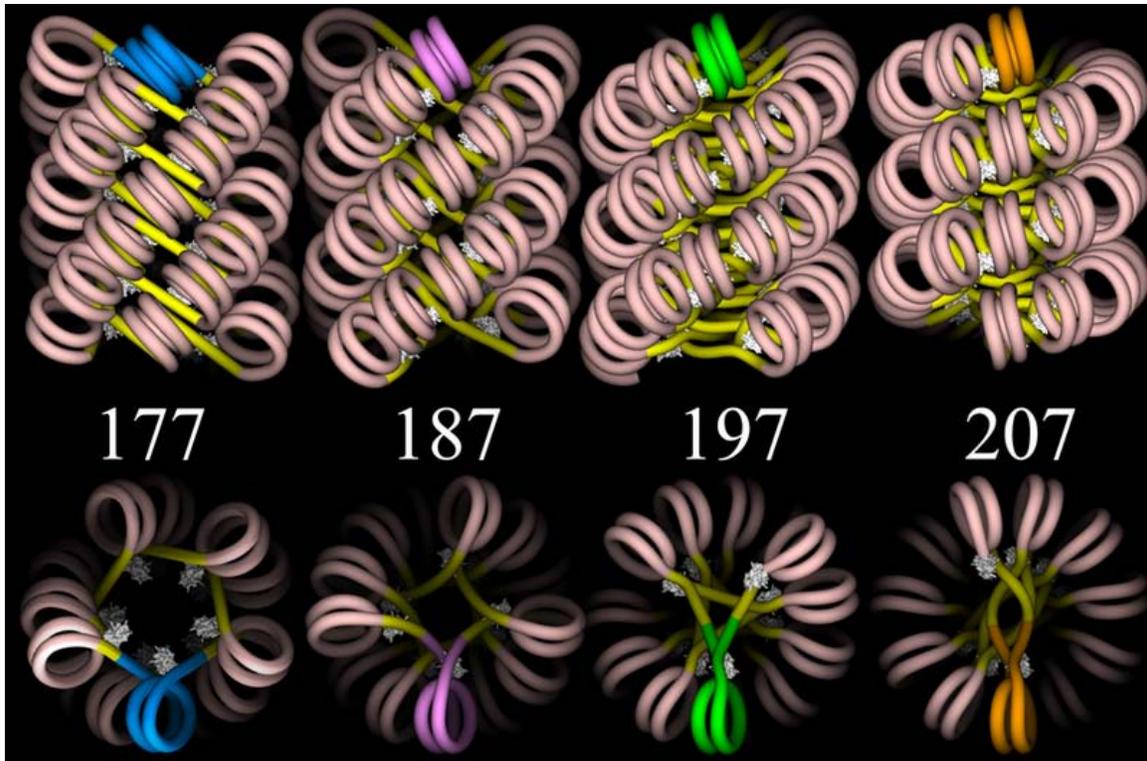
Left: 1 start helix "solenoid" structure.

Right: 2 start loose helix structure.

Note: the histones are omitted in this diagram - only the DNA is shown.

With addition of H1, the "beads-on-a-string" structure in turn coils into a 30 nm diameter helical structure known as the 30 nm fibre or filament. The precise structure of the chromatin fibre in the cell is not known in detail, and there is still some debate over this.

This level of chromatin structure is thought to be the form of euchromatin, which contains actively transcribed genes. EM studies have demonstrated that the 30 nm fibre is highly dynamic such that it unfolds into a 10 nm fiber ("beads-on-a-string") structure when transversed by an RNA polymerase engaged in transcription.



Four proposed structures of the 30 nm chromatin filament for DNA repeat length per nucleosomes ranging from 177 to 207 bp.
Linker DNA in yellow and nucleosomal DNA in pink.

The existing models commonly accept that the nucleosomes lie perpendicular to the axis of the fibre, with linker histones arranged internally. A stable 30 nm fibre relies on the regular positioning of nucleosomes along DNA. Linker DNA is relatively resistant to bending and rotation. This makes the length of linker DNA critical to the stability of the fibre, requiring nucleosomes to be separated by lengths that permit rotation and folding into the required orientation without excessive stress to the DNA. In this view, different length of the linker DNA should produce different folding topologies of the chromatin fiber. Recent theoretical work, based on electron-microscopy images of reconstituted fibers support this view.

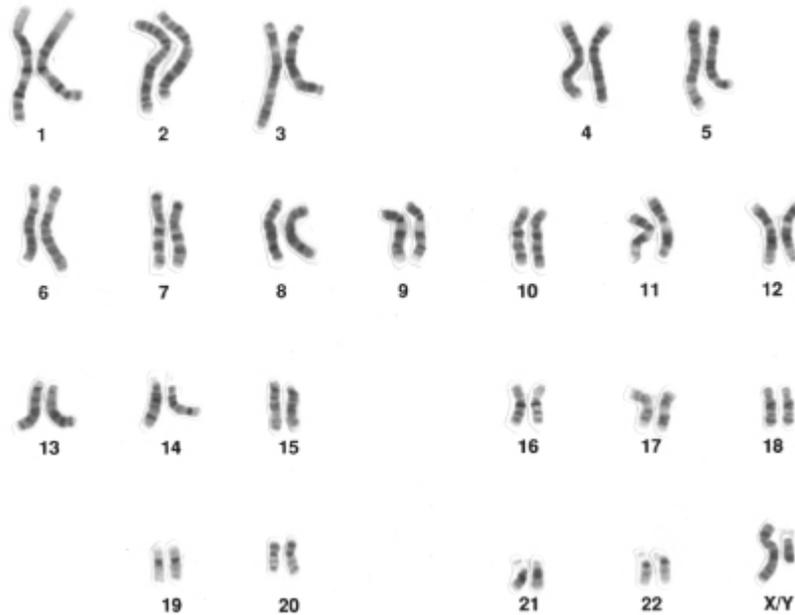
Spatial organization of chromatin in the cell nucleus

The layout of the genome within the nucleus is not random - specific regions of the genome have a tendency to be found in certain spaces. Specific regions of the chromatin are enriched at the nuclear membrane, while other regions are bound together by protein complexes. The layout of this is not, however, well characterised apart from the compaction of one of the two X chromosomes in mammalian females into the Barr body. This serves the role of permanently deactivating these genes, which prevents females getting a 'double dose' relative to males. The extent to which the inactive X is actually compacted is a matter of some controversy.

Chromatin and bursts of transcription

Fluctuations between open and closed chromatin may contribute discontinuity of transcription, or transcriptional bursting. Other factors are probably involved, such as the association and dissociation of transcription factor complexes with chromatin. The phenomenon, as opposed to simple probabilistic models of transcription, can account for the high variability in gene expression occurring between cells in isogenic populations.

Metaphase chromatin



Karyogram of human male using Giemsa staining, showing the classic metaphase chromatin structure.

The metaphase structure of chromatin differs vastly to that of interphase. It is optimised for physical strength and manageability, forming the classic chromosome structure seen in karyotypes. The structure of the condensed chromosome is thought to be loops of 30 nm fibre to a central scaffold of proteins. It is, however, not well characterised.

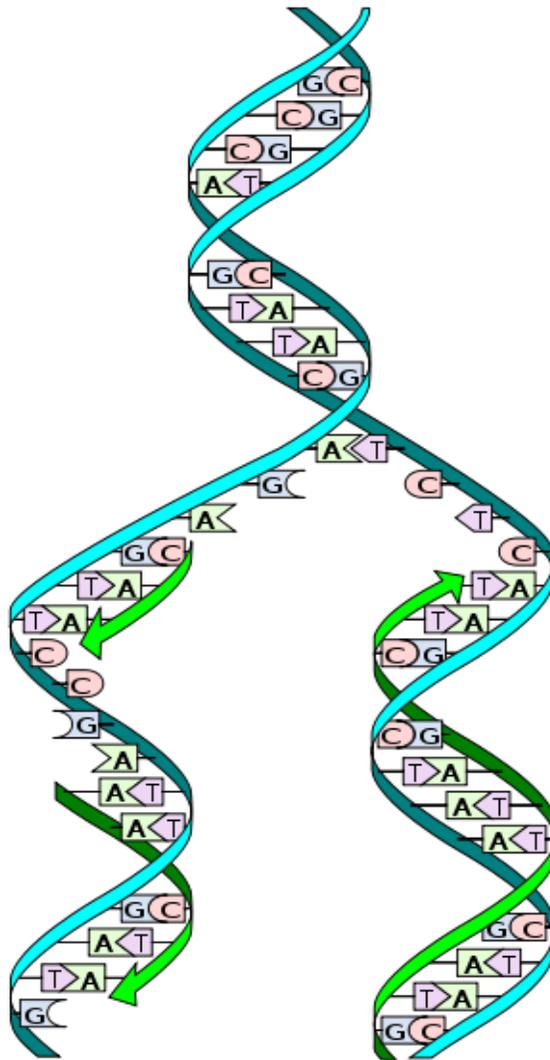
The physical strength of chromatin is vital for this stage of division to prevent shear damage to the DNA as the daughter chromosomes are separated. To maximise strength the composition of the chromatin changes as it approaches the centromere, primarily through alternative histone H1 analogues.

Chromatin: alternative definitions

1. **Simple and concise definition:** Chromatin is DNA plus the proteins (and RNA) that package DNA within the cell nucleus.
2. **A biochemists' operational definition:** Chromatin is the DNA/protein/RNA complex extracted from eukaryotic lysed interphase nuclei. Just which of the multitudinous substances present in a nucleus will constitute a part of the extracted material will depend in part on the technique each researcher uses. Furthermore, the composition and properties of chromatin vary from one cell type to the another, during development of a specific cell type, and at different stages in the cell cycle.
3. **The *DNA + histone = chromatin* definition:** The DNA double helix in the cell nucleus is packaged by special proteins termed histones. The formed protein/DNA complex is called chromatin. The structural entity of chromatin is the nucleosome.

Chapter- 8

DNA Replication



DNA replication. The double helix is unwound and each strand acts as a template. Bases are matched to synthesize the new partner strands.

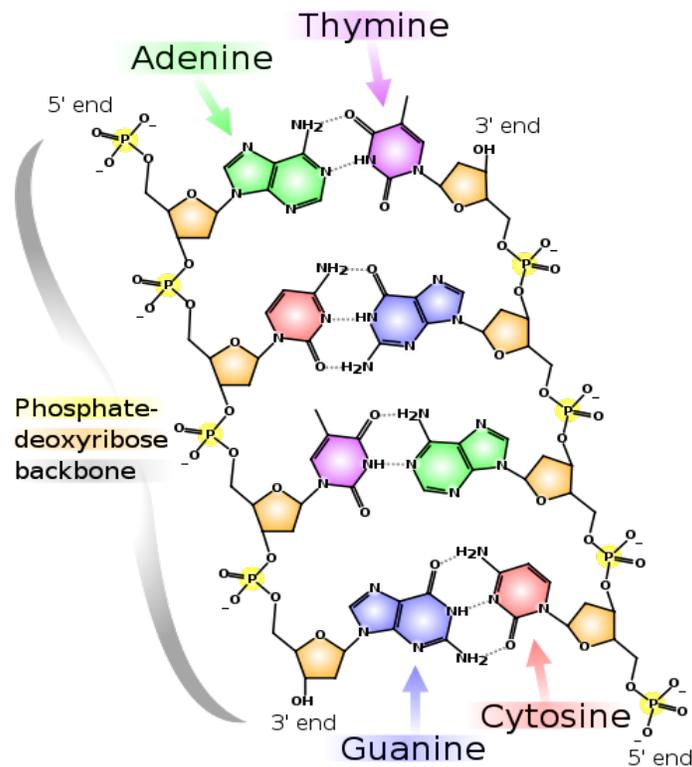
DNA replication, the basis for biological inheritance, is a fundamental process - that occurs in all living organisms - that copies their DNA. This process is *replication*, in that

each strand of the original double-stranded DNA molecule serves as template for the reproduction of the complementary strand. Therefore, following DNA replication, two identical DNA molecules have been produced from a single double-stranded DNA molecule. Cellular proofreading and error toe-checking mechanisms ensure near perfect fidelity for DNA replication.

In a cell, DNA replication begins at specific locations in the genome, called "origins". Unwinding of DNA at the origin, and synthesis of new strands, forms a replication fork. In addition to DNA polymerase, the enzyme that synthesizes the new DNA by adding nucleotides matched to the template strand, a number of other proteins are associated with the fork and assist in the initiation and continuation of DNA synthesis.

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

DNA structure



The chemical structure of DNA

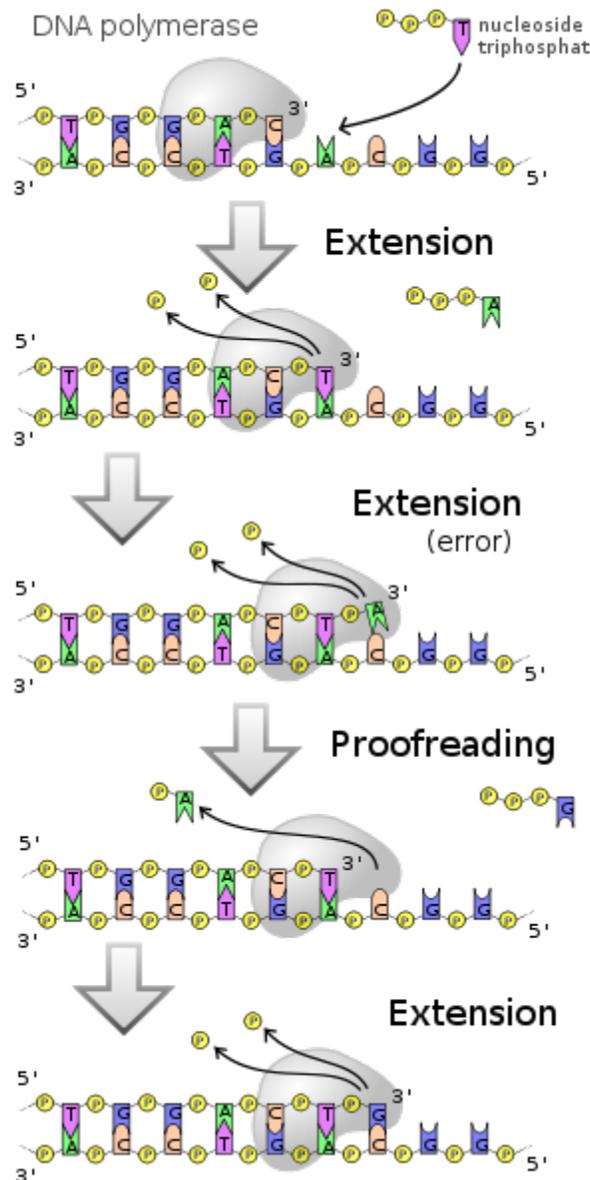


DNA usually exists as a double-stranded structure, with both strands coiled together to form the characteristic double-helix. Each single strand of DNA is a chain of four types of nucleotides having the bases: adenine, cytosine, guanine, and thymine. A nucleotide is a mono-, di-, or triphosphate deoxyribonucleoside; that is, a deoxyribose sugar is attached to one, two, or three phosphates. Chemical interaction of these nucleotides forms phosphodiester linkages, creating the phosphate-deoxyribose backbone of the DNA double helix with the bases pointing inward. Nucleotides (bases) are matched between strands through hydrogen bonds to form base pairs. Adenine pairs with thymine and cytosine pairs with guanine.

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." These terms refer to the carbon atom in deoxyribose to which the next phosphate in the chain attaches. In addition to being complementary, the two strands of DNA are antiparallel: They are orientated in opposite directions. This 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.

The pairing of bases in DNA through hydrogen bonding means that the information contained within each strand is redundant. The nucleotides on a single strand can be used to reconstruct nucleotides on a newly synthesized partner strand.

DNA polymerase



DNA polymerases add 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. A DNA polymerase can only extend an existing DNA strand paired with a template strand; it cannot begin the synthesis of a new strand. To begin synthesis of a new strand, a short fragment of DNA or RNA, called a primer, must be created and paired with the template strand before DNA polymerase can synthesize new DNA.

Once a primer pairs with DNA to be replicated, DNA polymerase synthesizes a new strand of DNA by extending the 3' end of an existing nucleotide chain, adding new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The energy for this process of DNA polymerization comes from two of the three total phosphates attached to each unincorporated base. (Free bases with their attached phosphate groups are called nucleoside triphosphates.) When a nucleotide is being added to a growing DNA strand, two of the phosphates are removed and the energy produced creates a phosphodiester (chemical) bond that attaches the remaining phosphate to the growing chain. The energetics of this process also help explain the directionality of synthesis - if DNA were synthesized in the 3' to 5' direction, the energy for the process would come from the 5' end of the growing strand rather than from free nucleotides.

In general, DNA polymerases are extremely accurate, making less than one error for every 10^7 nucleotides added. Even so, some DNA polymerases also have proofreading ability; they can remove nucleotides from the end of a strand in order to correct mismatched bases. If the 5' nucleotide needs to be removed during proofreading, the triphosphate end is lost. Hence, the energy source that usually provides energy to add a new nucleotide is also lost.

DNA replication within the cell

Origins of replication

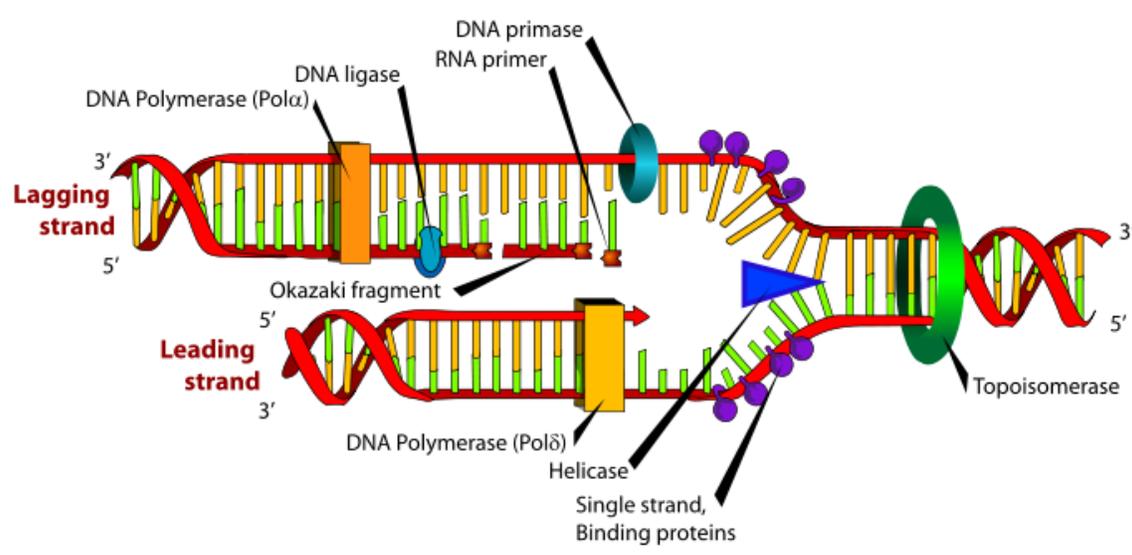
For a cell to divide, it must first replicate its DNA. This process is initiated at particular points within the DNA, known as "origins", which are targeted by proteins that separate the two strands and initiate DNA synthesis. Origins contain DNA sequences recognized by replication initiator proteins (e.g., *dnaA* in *E coli*' and the Origin Recognition Complex in yeast). These initiator proteins recruit other proteins to separate the two strands and initiate replication forks.

Initiator proteins recruit other proteins to separate the DNA strands at the origin, forming a bubble. Origins tend to be "AT-rich" (rich in adenine and thymine bases) to assist this process, because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair)—in general, strands rich in these nucleotides are easier to separate due the positive relationship between the number of hydrogen bonds and the difficulty of breaking these bonds. Once strands are separated, RNA primers are created on the template strands. To be more specific, the leading strand receives one RNA primer per active origin of replication while the lagging strand receives several; these several fragments of RNA primers found on the lagging strand of DNA are called Okazaki fragments, named after their discoverer. DNA Polymerase extends the leading strand in one continuous motion and the lagging strand in a discontinuous motion (due to the Okazaki fragments). RNase removes the RNA fragments used to initiate replication by DNA Polymerase, and another DNA 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.

As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming 2 replication forks. In bacteria, which have a single origin of replication on their circular chromosome, this process eventually 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.

The replication fork



Many enzymes are involved in the DNA replication fork.

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.

Leading strand

The leading strand is the template strand of the DNA double helix so that the replication fork moves along it in the 3' to 5' direction. This allows the new strand synthesized complementary to it to be synthesized 5' to 3' in the same direction as the movement of the replication fork.

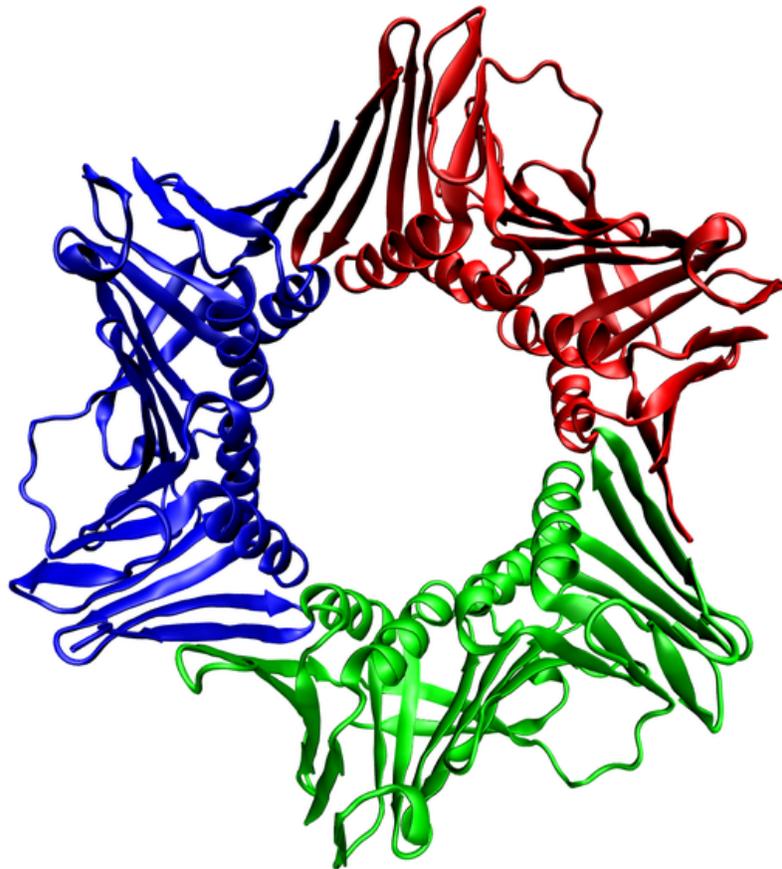
On the leading strand, a polymerase "reads" the DNA and adds nucleotides to it continuously. This polymerase is DNA polymerase III (DNA Pol III) in prokaryotes and presumably Pol ϵ in eukaryotes.

Lagging strand

The lagging strand is the strand of the template DNA double helix that is oriented so that the replication fork moves along it in a 5' to 3' manner. Because of its orientation, opposite to the working orientation of DNA polymerase III, which moves on a template in a 3' to 5' manner, replication of the lagging strand is more complicated than that of the leading strand.

On the lagging strand, primase "reads" the DNA and adds RNA to it in short, separated segments. In eukaryotes, primase is intrinsic to Pol α . DNA polymerase III or Pol δ lengthens the primed segments, forming Okazaki fragments. Primer removal in eukaryotes is also performed by Pol δ . In prokaryotes, DNA polymerase I "reads" the fragments, removes the RNA using its flap endonuclease domain (RNA primers are removed by 5'-3' exonuclease activity of polymerase I [weaver, 2005], and replaces the RNA nucleotides with DNA nucleotides (this is necessary because RNA and DNA use slightly different kinds of nucleotides). DNA ligase joins the fragments together.

Dynamics at the replication fork



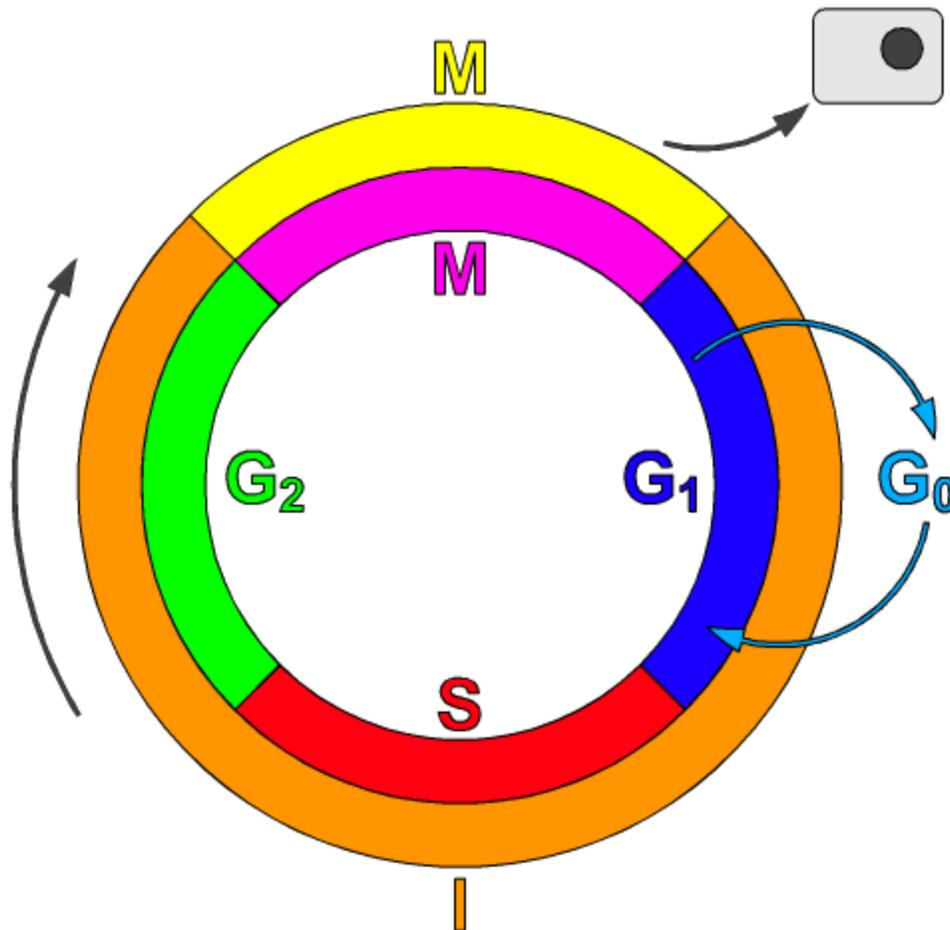
The assembled human DNA clamp, a trimer of the protein PCNA

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 would form a resistance that would eventually halt the progress of the replication fork. DNA topoisomerases are enzymes that solve these physical problems in the coiling of DNA. Topoisomerase I cuts a single backbone on the DNA, enabling the strands to swivel around each other to remove the build-up of twists. Topoisomerase II cuts both backbones, enabling one double-stranded DNA to pass through another, thereby removing knots and entanglements that can form within and between DNA molecules.

Bare single-stranded DNA has a tendency to fold back upon itself and form 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.

Regulation of replication



The cell cycle of eukaryotic cells

Eukaryotes

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 are quiescent in the "G₀" stage and do not replicate their DNA.

Replication of chloroplast and mitochondrial genomes occurs independent of the cell cycle, through the process of D-loop replication.

Bacteria

Most bacteria do not go through a well-defined cell cycle and, instead, continuously copy their DNA; during rapid growth, this can result in the concurrent occurrences of multiple rounds of replication. Within *E coli*, the best-characterized bacteria, regulation of DNA replication can be achieved through several mechanisms, including: the hemimethylation and sequestering of the origin sequence, the ratio of ATP to ADP, and the levels of protein DnaA. These all control the process of initiator proteins binding to the origin sequences.

Because *E coli* methylates GATC DNA sequences, DNA synthesis results in hemimethylated sequences. This hemimethylated DNA is recognized by a protein (SeqA), which binds and sequesters the origin sequence; in addition, *dnaA* (required for initiation of replication) binds less well to hemimethylated DNA. As a result, newly replicated origins are prevented from immediately initiating another round of DNA replication.

ATP builds up when the cell is in a rich medium, triggering DNA replication once the cell has reached a specific size. ATP competes with ADP to bind to DnaA, and the DnaA-ATP complex is able to initiate replication. A certain number of DnaA proteins are also required for DNA replication — each time the origin is copied the number of binding sites for DnaA doubles, requiring the synthesis of more DnaA to enable another initiation of replication.

Termination of replication

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* regulate 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 manner. Because eukaryotes have linear chromosomes, DNA replication is unable to synthesize to the very end of the chromosomes (telomeres), resulting in telomere shortening. This is a normal process in somatic cells — cells are able to divide only 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.

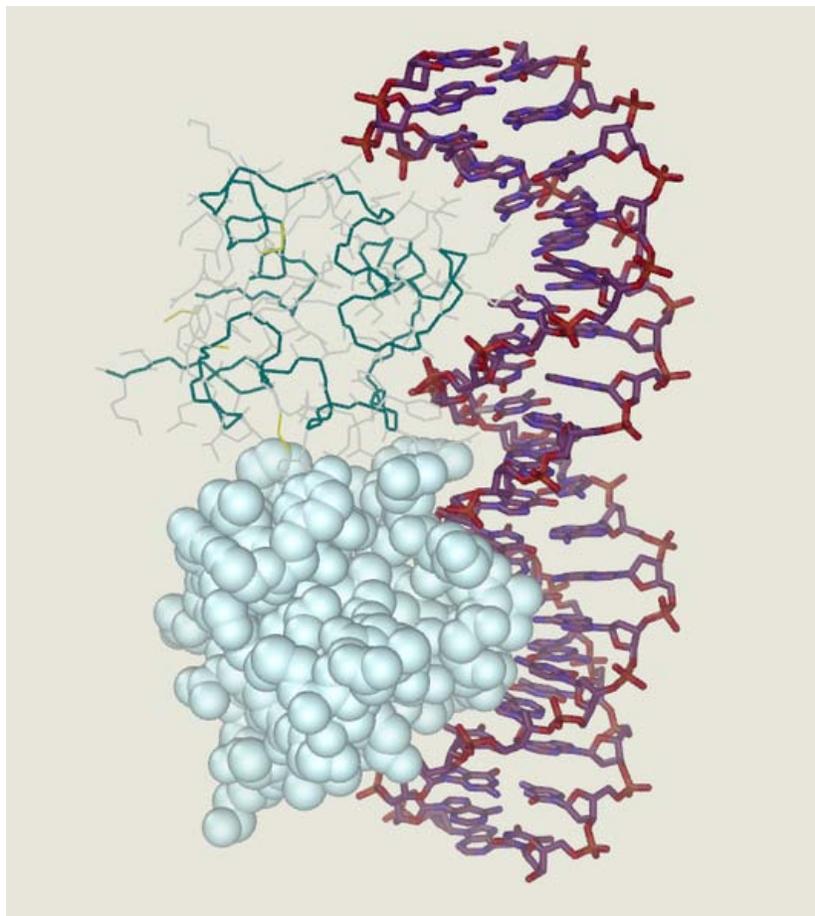
Polymerase chain reaction

Researchers commonly replicate DNA *in vitro* using the polymerase chain reaction (PCR). PCR uses a pair of primers to span a target region in template DNA, and then polymerizes partner strands in each direction from these primers using a thermostable DNA polymerase. Repeating this process through multiple cycles produces amplification of the targeted DNA region. At the start of each cycle, the mixture of template and primers is heated, separating the newly synthesized molecule and template. Then, as the mixture cools, both of these become templates for annealing of new primers, and the polymerase extends from these. As a result, the number of copies of the target region doubles each round, increasing exponentially.

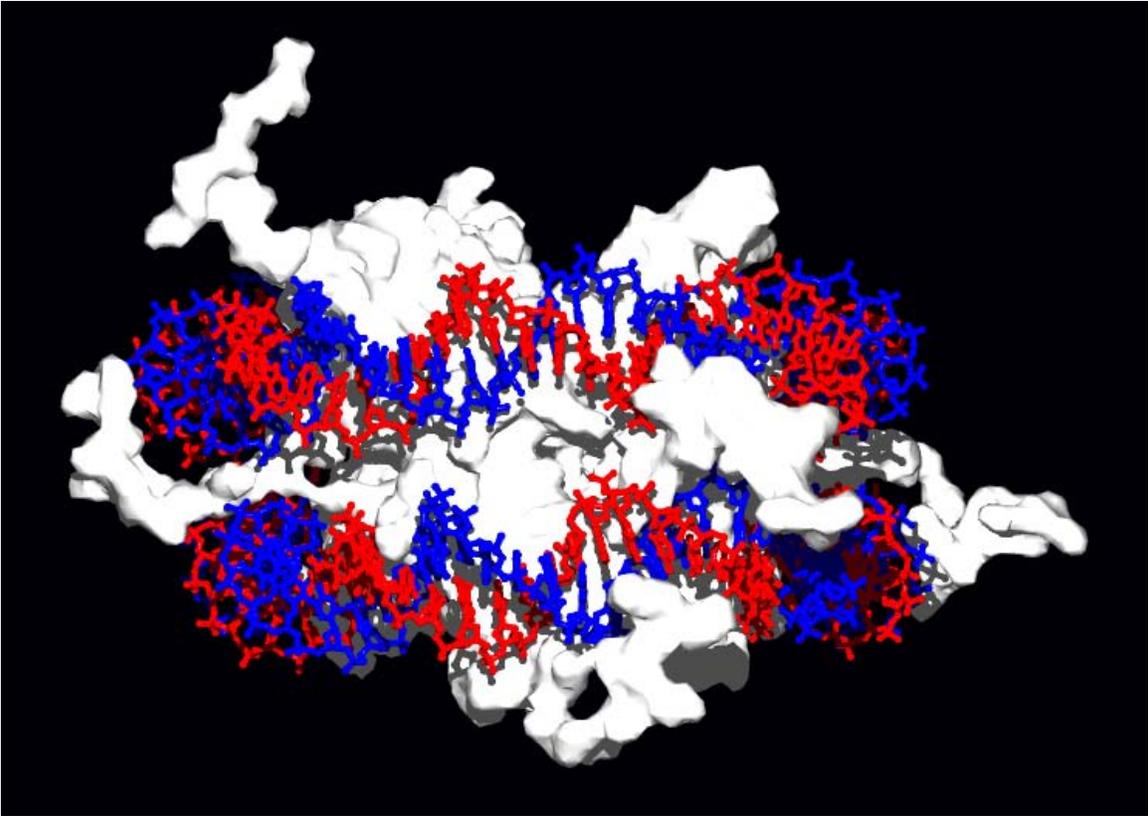
Chapter- 9

DNA-Binding Protein and Genetic Recombination

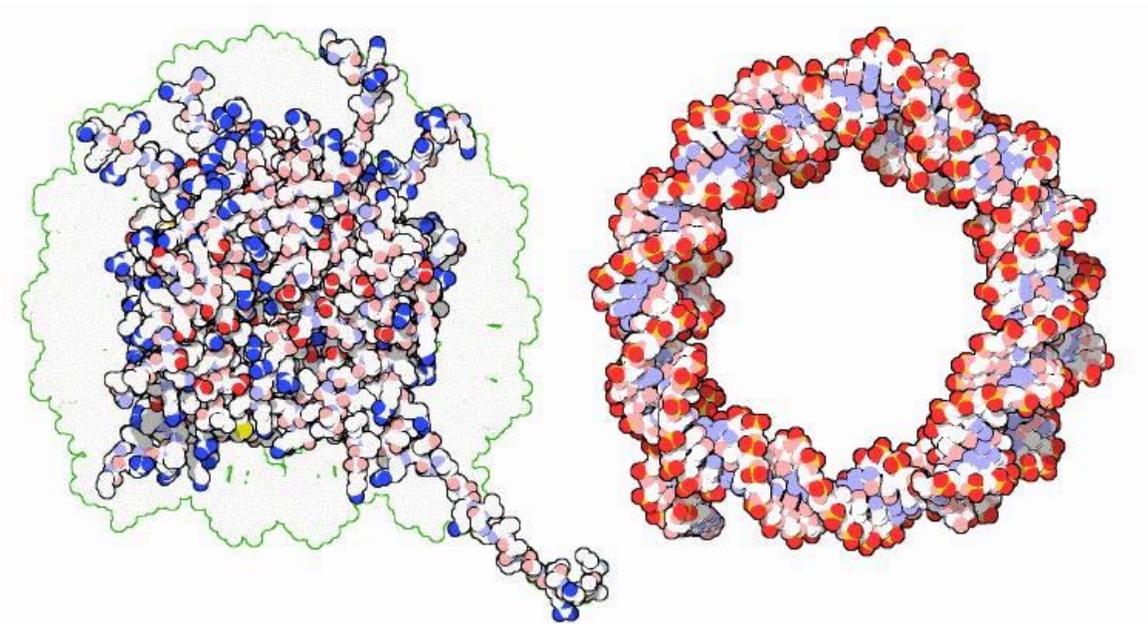
DNA-binding protein



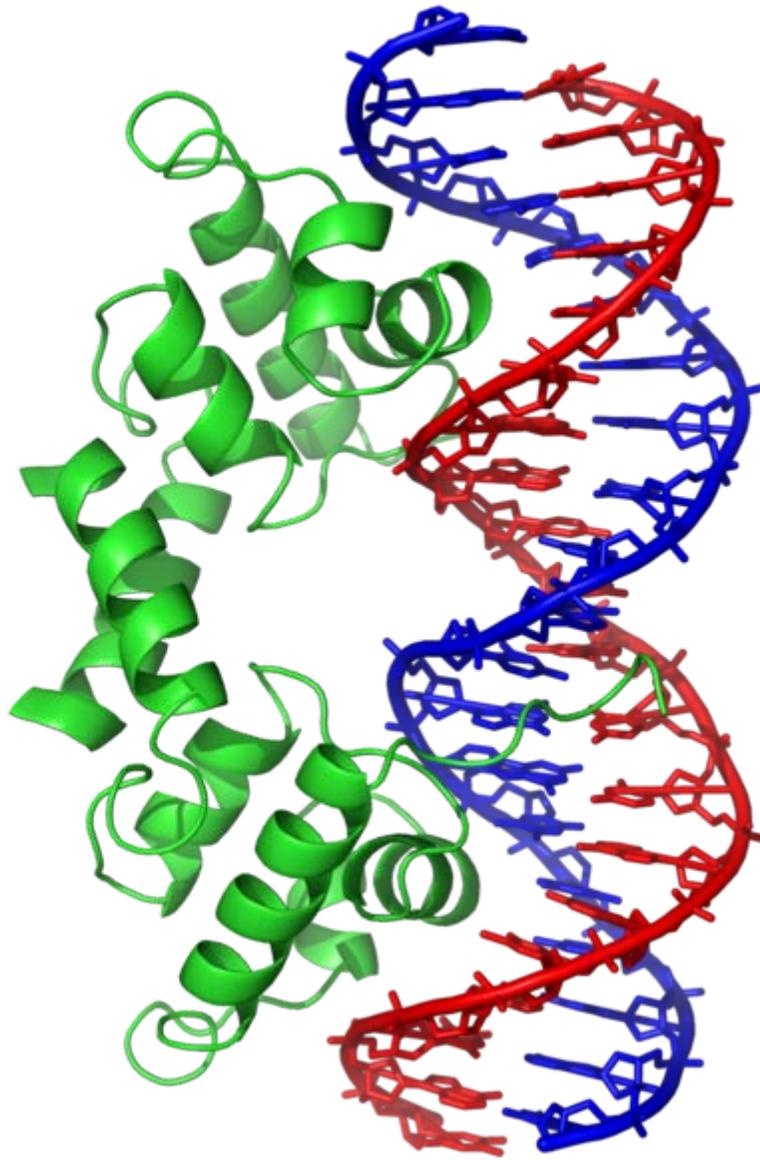
Cro protein complex with DNA



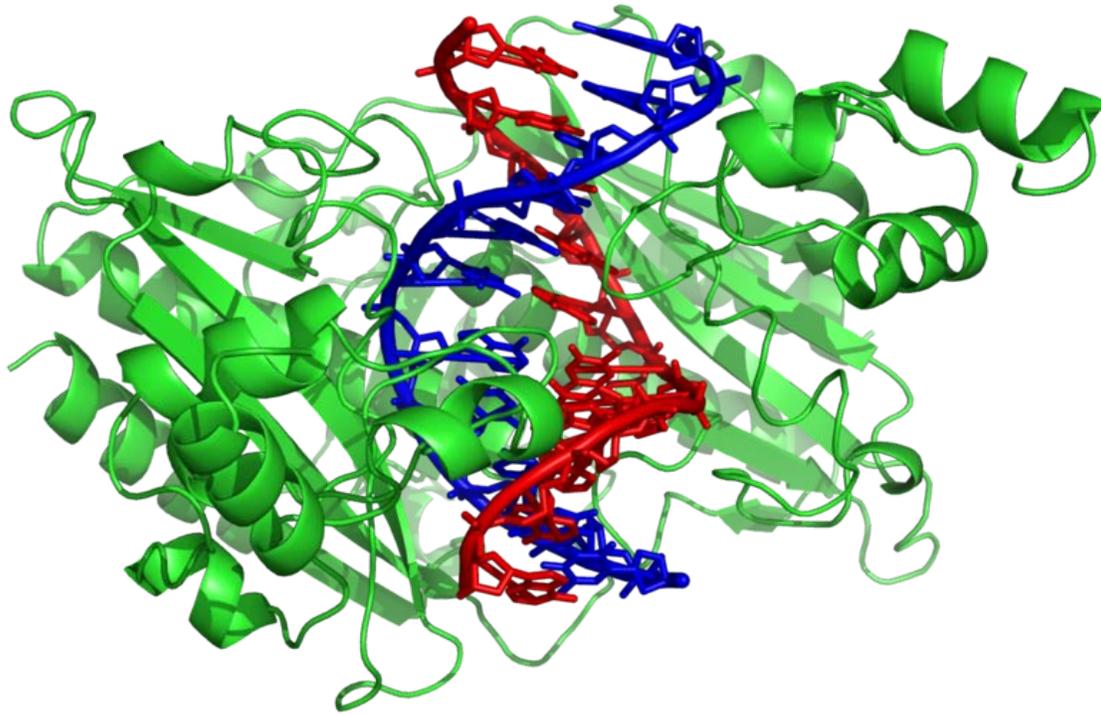
Interaction of DNA with histones (shown in white, top). These proteins' basic amino acids (below left, blue) bind to the acidic phosphate groups on DNA (below right, red).



Nucleosome (opposites attracts) (see above for color scheme)



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



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

DNA-binding proteins are proteins that are composed of DNA-binding domains and thus have a specific or general affinity for either single or double stranded DNA. Sequence-specific DNA-binding proteins generally interact with the major groove of B-DNA, because it exposes more functional groups that identify a base pair. However there are some known narrow-groove DNA-binding ligands such as Netropsin, Distamycin, Hoechst 33258, Pentamidine and others.

Examples

DNA-binding proteins include transcription factors which modulate the process of transcription, various polymerases, nucleases which cleave DNA molecules, and histones which are involved in chromosome packaging in the cell nucleus. DNA-binding proteins can incorporate such domains as the zinc finger, the helix-turn-helix, and the leucine zipper (among many others) that facilitate binding to nucleic acid.

Non-specific DNA-protein interactions

Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. 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 its surface. These non-specific interactions are formed through basic residues in the histones making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are therefore largely independent of the base sequence. Chemical modifications of these basic amino acid residues include methylation, phosphorylation and acetylation. These chemical changes 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. Other non-specific DNA-binding proteins in chromatin include the high-mobility group proteins, which bind to bent or distorted DNA. These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.

DNA-binding proteins that specifically bind single-stranded DNA

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 seem to stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

Binding to particular DNA sequences

In contrast, 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. The transcription factors do this in two ways. Firstly, 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. Alternatively, transcription factors can bind enzymes that modify the histones at the promoter; this will change the accessibility of the DNA template to the polymerase.

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. Mathematical descriptions of protein-DNA binding taking into account sequence-specificity, competitive and cooperative binding of proteins of different types are usually performed with the help of the lattice models.

Genetic recombination

Genetic recombination is a process by which a molecule of nucleic acid (usually DNA, but can also be RNA) is broken and then joined to a different one. Recombination can occur between similar molecules of DNA, as in homologous recombination, or dissimilar molecules, as in non-homologous end joining. Recombination is a common method of DNA repair in both bacteria and eukaryotes. In eukaryotes, recombination also occurs in meiosis, where it facilitates chromosomal crossover. The crossover process leads to offspring's having different combinations of genes from those of their parents, and can occasionally produce new chimeric alleles. In organisms with an adaptive immune system, a type of genetic recombination called V(D)J recombination helps immune cells rapidly diversify to recognize and adapt to new pathogens. The shuffling of genes brought about by genetic recombination is thought to have many advantages, as it is a major engine of genetic variation and also allows sexually reproducing organisms to avoid Muller's ratchet, in which the genomes of an asexual population accumulate deleterious mutations in an irreversible manner.

In genetic engineering, recombination can also refer to artificial and deliberate recombination of disparate pieces of DNA, often from different organisms, creating what is called recombinant DNA. A prime example of such a use of genetic recombination is gene targeting, which can be used to add, delete or otherwise change an organism's genes. This technique is important to biomedical researchers as it allows them to study the effects of specific genes. Techniques based on genetic recombination are also applied in protein engineering to develop new proteins of biological interest.

Genetic recombination is catalyzed by many different enzymes, called *recombinases*. RecA, the chief recombinase found in *Escherichia coli*, is responsible for the repair of DNA double strand breaks (DSBs). In yeast and other eukaryotic organisms there are two recombinases required for repairing DSBs. The RAD51 protein is required for mitotic and meiotic recombination, whereas the DMC1 protein is specific to meiotic recombination.

Chromosomal crossover

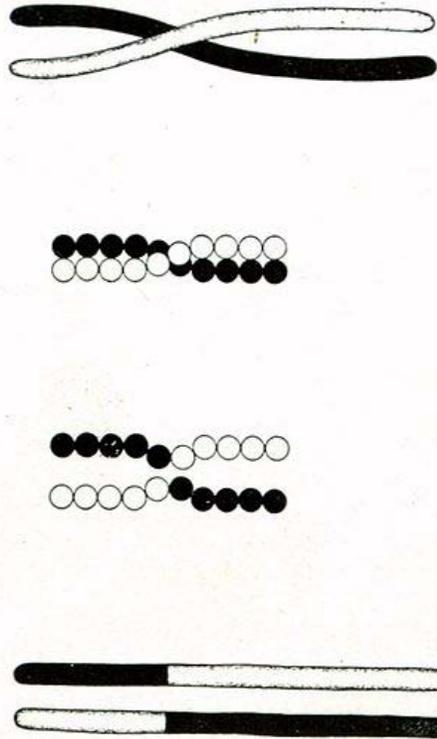


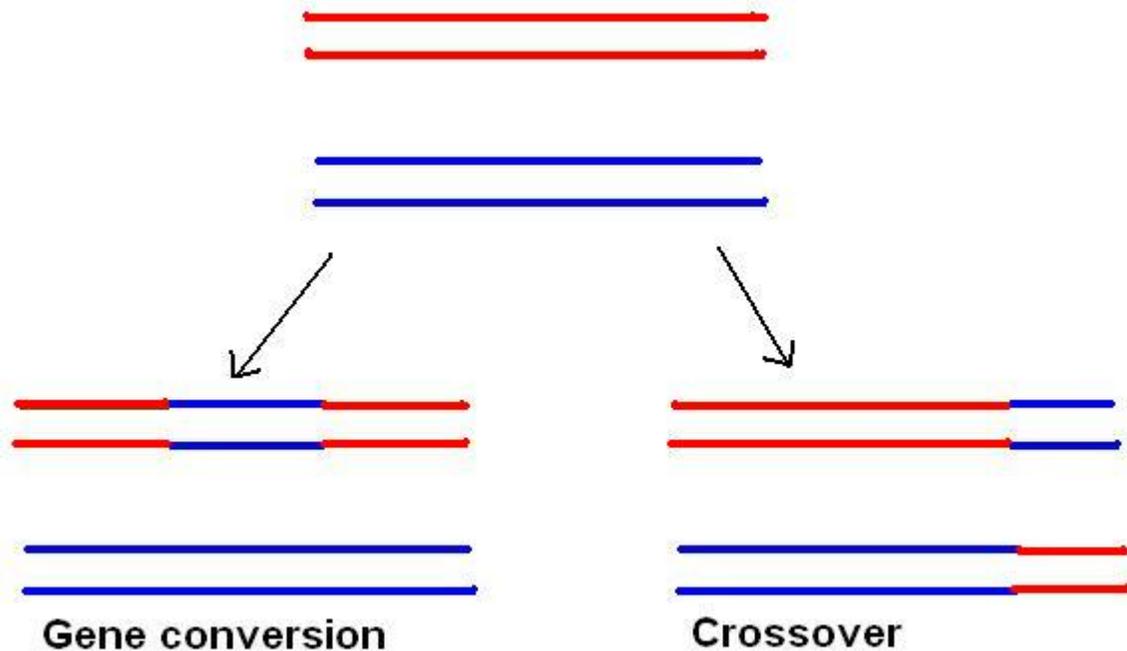
FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.

Thomas Hunt Morgan's illustration of crossing over (1916)

Chromosomal crossover refers to recombination between the paired chromosomes inherited from each of one's parents, generally occurring during meiosis. During prophase I the four available chromatids are in tight formation with one another. While in this formation, homologous sites on two chromatids can mesh with one another, and may exchange genetic information.

Because recombination can occur with small probability at any location along chromosome, the frequency of recombination between two locations depends on their distance. Therefore, for genes sufficiently distant on the same chromosome the amount of crossover is high enough to destroy the correlation between alleles.

Gene conversion



The difference between **gene conversion** and chromosomal crossover. Blue are the two chromatids of one chromosome and red are the two chromatids of another one.

In gene conversion, a section of genetic material is copied from one chromosome to another, without the donating chromosome being changed. Gene conversion occurs at high frequency during meiotic division and at low frequency in somatic cells. It is a process by which a DNA sequence is copied from one DNA helix (which remains unchanged) to another DNA helix, whose sequence is altered. It is one of the ways a gene may be mutated. Gene conversion may lead to non-Mendelian inheritance and has often been recorded in fungal crosses.

Nonhomologous recombination

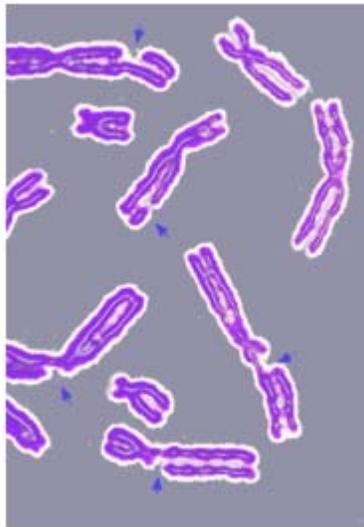
Recombination can occur between DNA sequences that contain no sequence homology. This is referred to as *nonhomologous recombination* or nonhomologous end joining.

In B cells

B cells of the immune system perform genetic recombination, called immunoglobulin class switching. It is a biological mechanism that changes an antibody from one class to another, for example, from an isotype called IgM to an isotype called IgG.

Chapter- 10

DNA Repair



DNA damage resulting in multiple broken chromosomes

DNA repair refers to a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome. In human cells, both normal metabolic activities and environmental factors such as UV light and radiation can cause DNA damage, resulting in as many as 1 million individual molecular lesions per cell per day. Many of these lesions cause structural damage to the DNA molecule and can alter or eliminate the cell's ability to transcribe the gene that the affected DNA encodes. Other lesions induce potentially harmful mutations in the cell's genome, which affect the survival of its daughter cells after it undergoes mitosis. As a consequence, the DNA repair process is constantly active as it responds to damage in the DNA structure. When normal repair processes fail, and when cellular apoptosis does not occur, irreparable DNA damage may occur, including double-strand breaks and DNA crosslinkages.

The rate of DNA repair is dependent on many factors, including the cell type, the age of the cell, and the extracellular environment. A cell that has accumulated a large amount of

DNA damage, or one that no longer effectively repairs damage incurred to its DNA, can enter one of three possible states:

1. an irreversible state of dormancy, known as senescence
2. cell suicide, also known as apoptosis or programmed cell death
3. unregulated cell division, which can lead to the formation of a tumor that is cancerous

The DNA repair ability of a cell is vital to the integrity of its genome and thus to its normal functioning and that of the organism. Many genes that were initially shown to influence life span have turned out to be involved in DNA damage repair and protection. Failure to correct molecular lesions in cells that form gametes can introduce mutations into the genomes of the offspring and thus influence the rate of evolution.

DNA damage

DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 1,000 to 1,000,000 molecular lesions per cell per day. While this constitutes only 0.000165% of the human genome's approximately 6 billion bases (3 billion base pairs), unrepaired lesions in critical genes (such as tumor suppressor genes) can impede a cell's ability to carry out its function and appreciably increase the likelihood of tumor formation.

The vast majority of DNA damage affects the primary structure of the double helix; that is, the bases themselves are chemically modified. These modifications can in turn disrupt the molecules' regular helical structure by introducing non-native chemical bonds or bulky adducts that do not fit in the standard double helix. Unlike proteins and RNA, DNA usually lacks tertiary structure and therefore damage or disturbance does not occur at that level. DNA is, however, supercoiled and wound around "packaging" proteins called histones (in eukaryotes), and both superstructures are vulnerable to the effects of DNA damage.

Sources of damage

DNA damage can be subdivided into two main types:

1. endogenous damage such as attack by reactive oxygen species produced from normal metabolic byproducts (spontaneous mutation), especially the process of oxidative deamination
 1. also includes replication errors
2. exogenous damage caused by external agents such as
 1. ultraviolet [UV 200-300nm] radiation from the sun
 2. other radiation frequencies, including x-rays and gamma rays
 3. hydrolysis or thermal disruption
 4. certain plant toxins

5. human-made mutagenic chemicals, especially aromatic compounds that act as DNA intercalating agents
6. cancer chemotherapy and radiotherapy
7. viruses

The replication of damaged DNA before cell division can lead to the incorporation of wrong bases opposite damaged ones. Daughter cells that inherit these wrong bases carry mutations from which the original DNA sequence is unrecoverable (except in the rare case of a back mutation, for example, through gene conversion).

Types of damage

There are five main types of damage to DNA due to endogenous cellular processes:

1. *oxidation* of bases [e.g. 8-oxo-7,8-dihydroguanine (8-oxoG)] and generation of DNA strand interruptions from reactive oxygen species,
2. *alkylation* of bases (usually methylation), such as formation of 7-methylguanine, 1-methyladenine, 6-O-Methylguanine
3. *hydrolysis* of bases, such as deamination, depurination, and depyrimidination.
4. "bulky adduct formation" (i.e., benzo[a]pyrene diol epoxide-dG adduct)
5. *mismatch* of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over or mistakenly inserted.

Damage caused by exogenous agents comes in many forms. Some examples are:

1. *UV-B light* causes crosslinking between adjacent cytosine and thymine bases creating *pyrimidine dimers*. This is called direct DNA damage.
2. *UV-A light* creates mostly free radicals. The damage caused by free radicals is called indirect DNA damage.
3. *Ionizing radiation* such as that created by radioactive decay or in *cosmic rays* causes breaks in DNA strands. Low-level ionizing radiation may induce irreparable DNA damage (leading to replicational and transcriptional errors needed for neoplasia or may trigger viral interactions) leading to pre-mature aging and cancer.
4. *Thermal disruption* at elevated temperature increases the rate of depurination (loss of purine bases from the DNA backbone) and single-strand breaks. For example, hydrolytic depurination is seen in the thermophilic bacteria, which grow in hot springs at 40-80 °C. The rate of depurination (300 purine residues per genome per generation) is too high in these species to be repaired by normal repair machinery, hence a possibility of an adaptive response cannot be ruled out.
5. *Industrial chemicals* such as vinyl chloride and hydrogen peroxide, and environmental chemicals such as polycyclic hydrocarbons found in smoke, soot and tar create a huge diversity of DNA adducts- ethenobases, oxidized bases, alkylated phosphotriesters and Crosslinking of DNA just to name a few.

UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of induced damage. Spontaneous damage can include the loss of a base, deamination, sugar ring puckering and tautomeric shift.

Nuclear versus mitochondrial DNA damage

In human cells, and eukaryotic cells in general, DNA is found in two cellular locations - inside the nucleus and inside the mitochondria. Nuclear DNA (nDNA) exists as chromatin during non-replicative stages of the cell cycle and is condensed into aggregate structures known as chromosomes during cell division. In either state the DNA is highly compacted and wound up around bead-like proteins called histones. Whenever a cell needs to express the genetic information encoded in its nDNA the required chromosomal region is unravelled, genes located therein are expressed, and then the region is condensed back to its resting conformation. Mitochondrial DNA (mtDNA) is located inside mitochondria organelles, exists in multiple copies, and is also tightly associated with a number of proteins to form a complex known as the nucleoid. Inside mitochondria, reactive oxygen species (ROS), or free radicals, byproducts of the constant production of adenosine triphosphate (ATP) via oxidative phosphorylation, create a highly oxidative environment that is known to damage mtDNA. A critical enzyme in counteracting the toxicity of these species is superoxide dismutase, which is present in both the mitochondria and cytoplasm of eukaryotic cells.

Senescence and apoptosis

Senescence, an irreversible state in which the cell no longer divides, is a protective response to the shortening of the chromosome ends. The telomeres are long regions of repetitive noncoding DNA that cap chromosomes and undergo partial degradation each time a cell undergoes division. In contrast, quiescence is a reversible state of cellular dormancy that is unrelated to genome damage. Senescence in cells may serve as a functional alternative to apoptosis in cases where the physical presence of a cell for spatial reasons is required by the organism, which serves as a "last resort" mechanism to prevent a cell with damaged DNA from replicating inappropriately in the absence of pro-growth cellular signaling. Unregulated cell division can lead to the formation of a tumor, which is potentially lethal to an organism. Therefore, the induction of senescence and apoptosis is considered to be part of a strategy of protection against cancer.

DNA damage and mutation

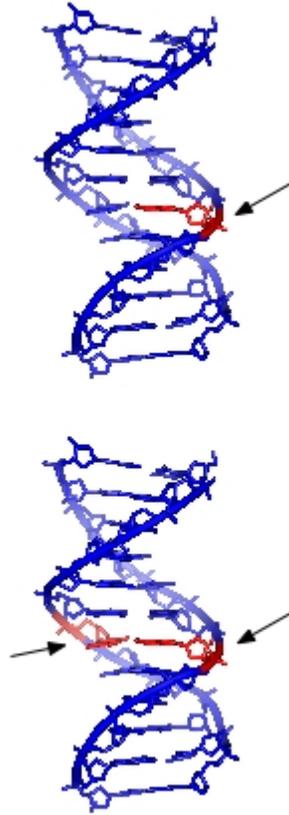
It is important to distinguish between DNA damage and mutation, the two major types of error in DNA. DNA damages and mutation are fundamentally different. Damages are physical abnormalities in the DNA, such as single- and double-strand breaks, 8-hydroxydeoxyguanosine residues, and polycyclic aromatic hydrocarbon adducts. DNA damages can be recognized by enzymes, and, thus, they can be correctly repaired if redundant information, such as the undamaged sequence in the complementary DNA strand or in a homologous chromosome, is available for copying. If a cell retains DNA

damage, transcription of a gene can be prevented, and, thus, translation into a protein will also be blocked. Replication may also be blocked and/or the cell may die.

In contrast to DNA damage, a mutation is a change in the base sequence of the DNA. A mutation cannot be recognized by enzymes once the base change is present in both DNA strands, and, thus, a mutation cannot be repaired. At the cellular level, mutations can cause alterations in protein function and regulation. Mutations are replicated when the cell replicates. In a population of cells, mutant cells will increase or decrease in frequency according to the effects of the mutation on the ability of the cell to survive and reproduce. Although distinctly different from each other, DNA damages and mutations are related because DNA damages often cause errors of DNA synthesis during replication or repair; these errors are a major source of mutation.

Given these properties of DNA damage and mutation, it can be seen that DNA damages are a special problem in non-dividing or slowly dividing cells, where unrepaired damages will tend to accumulate over time. On the other hand, in rapidly dividing cells, unrepaired DNA damages that do not kill the cell by blocking replication will tend to cause replication errors and thus mutation. The great majority of mutations that are not neutral in their effect are deleterious to a cell's survival. Thus, in a population of cells comprising a tissue with replicating cells, mutant cells will tend to be lost. However, infrequent mutations that provide a survival advantage will tend to clonally expand at the expense of neighboring cells in the tissue. This advantage to the cell is disadvantageous to the whole organism, because such mutant cells can give rise to cancer. Thus, DNA damages in frequently dividing cells, because they give rise to mutations, are a prominent cause of cancer. In contrast, DNA damages in infrequently dividing cells are likely a prominent cause of aging.

DNA repair mechanisms



Single-strand and double-strand DNA damage

Cells cannot function if DNA damage corrupts the integrity and accessibility of essential information in the genome (but cells remain superficially functional when so-called "non-essential" genes are missing or damaged). Depending on the type of damage inflicted on the DNA's double helical structure, a variety of repair strategies have evolved to restore lost information. If possible, cells use the unmodified complementary strand of the DNA or the sister chromatid as a template to recover the original information. Without access to a template, cells use an error-prone recovery mechanism known as translesion synthesis as a last resort.

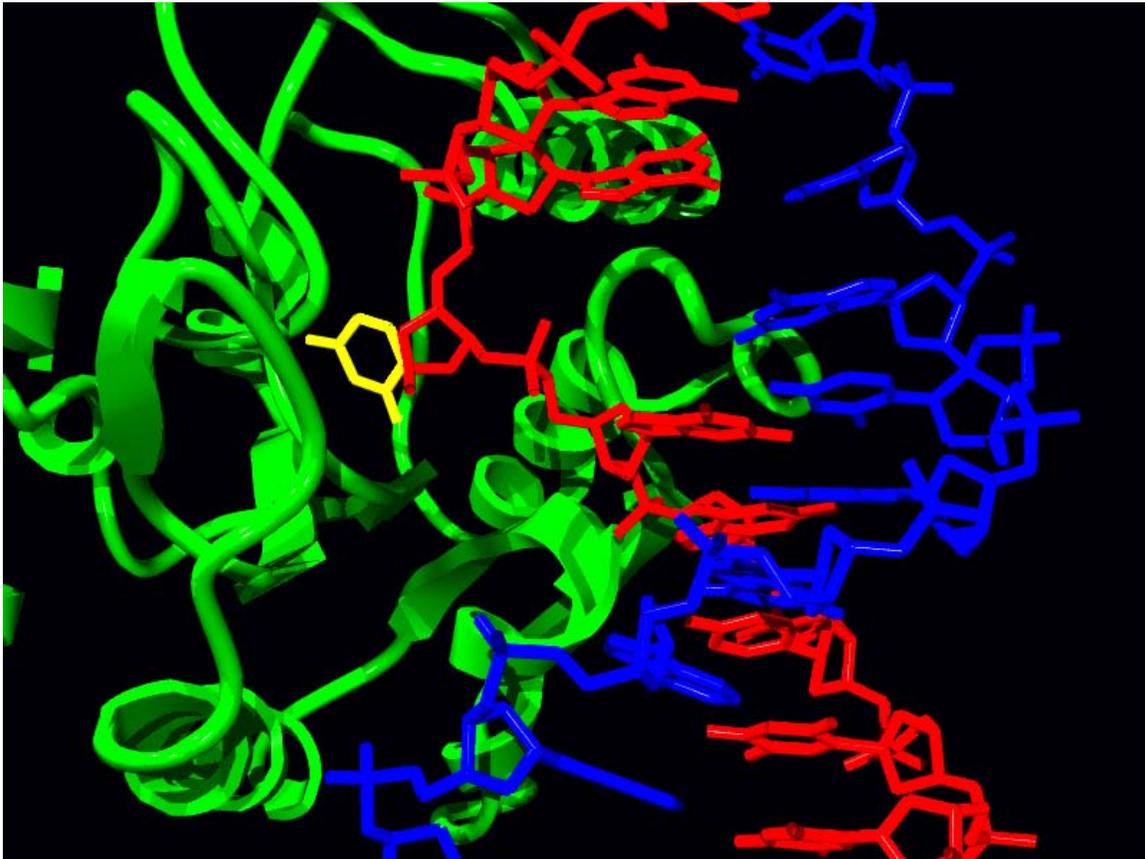
Damage to DNA alters the spatial configuration of the helix, and such alterations can be detected by the cell. Once damage is localized, specific DNA repair molecules bind at or near the site of damage, inducing other molecules to bind and form a complex that enables the actual repair to take place.

Direct reversal

Cells are known to eliminate three types of damage to their DNA by chemically reversing it. These mechanisms do not require a template, since the types of damage they

counteract can occur in only one of the four bases. Such direct reversal mechanisms are specific to the type of damage incurred and do not involve breakage of the phosphodiester backbone. The formation of pyrimidine dimers upon irradiation with UV light results in an abnormal covalent bond between adjacent pyrimidine bases. The photoreactivation process directly reverses this damage by the action of the enzyme photolyase, whose activation is obligately dependent on energy absorbed from blue/UV light (300–500 nm wavelength) to promote catalysis. Another type of damage, methylation of guanine bases, is directly reversed by the protein methyl guanine methyl transferase (MGMT), the bacterial equivalent of which is called ogt. This is an expensive process because each MGMT molecule can be used only once; that is, the reaction is stoichiometric rather than catalytic. A generalized response to methylating agents in bacteria is known as the adaptive response and confers a level of resistance to alkylating agents upon sustained exposure by upregulation of alkylation repair enzymes. The third type of DNA damage reversed by cells is certain methylation of the bases cytosine and adenine.

Single-strand damage



Structure of the base-excision repair enzyme uracil-DNA glycosylase. The uracil residue is shown in yellow.

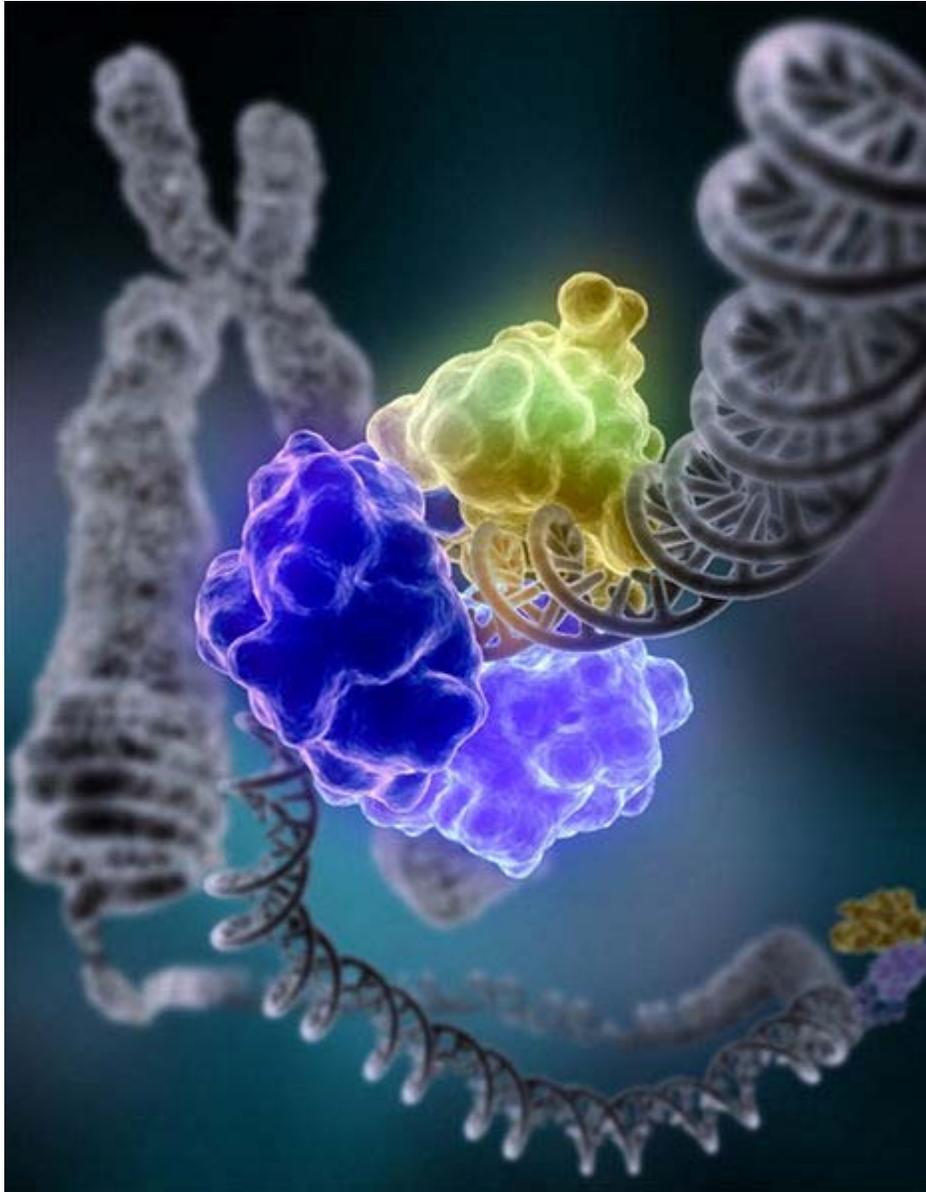
When only one of the two strands of a double helix has a defect, the other strand can be used as a template to guide the correction of the damaged strand. In order to repair

damage to one of the two paired molecules of DNA, there exist a number of excision repair mechanisms that remove the damaged nucleotide and replace it with an undamaged nucleotide complementary to that found in the undamaged DNA strand.

1. Base excision repair (BER), which repairs damage to a single base caused by oxidation, alkylation, hydrolysis, or deamination. The damaged base is removed by a DNA glycosylase. The "missing tooth" is then recognised by an enzyme called AP endonuclease, which cuts the Phosphodiester bond. The missing part is then resynthesized by a DNA polymerase, and a DNA ligase performs the final nick-sealing step.
2. Nucleotide excision repair (NER), which recognizes bulky, helix-distorting lesions such as pyrimidine dimers and 6,4 photoproducts. A specialized form of NER known as transcription-coupled repair deploys NER enzymes to genes that are being actively transcribed.
3. Mismatch repair (MMR), which corrects errors of DNA replication and recombination that result in mispaired (but undamaged) nucleotides.

Double-strand breaks

Double-strand breaks, in which both strands in the double helix are severed, are particularly hazardous to the cell because they can lead to genome rearrangements. Three mechanisms exist to repair double-strand breaks (DSBs): non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and homologous recombination. PVN Acharya noted that double-strand breaks and a "cross-linkage joining both strands at the same point is irreparable because neither strand can then serve as a template for repair. The cell will die in the next mitosis or in some rare instances, mutate."



DNA ligase, shown above repairing chromosomal damage, is an enzyme that joins broken nucleotides together by catalyzing the formation of an internucleotide ester bond between the phosphate backbone and the deoxyribose nucleotides.

In NHEJ, DNA Ligase IV, a specialized DNA ligase that forms a complex with the cofactor XRCC4, directly joins the two ends. To guide accurate repair, NHEJ relies on short homologous sequences called microhomologies present on the single-stranded tails of the DNA ends to be joined. If these overhangs are compatible, repair is usually accurate. NHEJ can also introduce mutations during repair. Loss of damaged nucleotides at the break site can lead to deletions, and joining of nonmatching termini forms translocations. NHEJ is especially important before the cell has replicated its DNA, since there is no template available for repair by homologous recombination. There are "backup" NHEJ pathways in higher eukaryotes. Besides its role as a genome caretaker,

NHEJ is required for joining hairpin-capped double-strand breaks induced during V(D)J recombination, the process that generates diversity in B-cell and T-cell receptors in the vertebrate immune system.

Homologous recombination requires the presence of an identical or nearly identical sequence to be used as a template for repair of the break. The enzymatic machinery responsible for this repair process is nearly identical to the machinery responsible for chromosomal crossover during meiosis. This pathway allows a damaged chromosome to be repaired using a sister chromatid (available in G2 after DNA replication) or a homologous chromosome as a template. DSBs caused by the replication machinery attempting to synthesize across a single-strand break or unrepaired lesion cause collapse of the replication fork and are typically repaired by recombination.

Topoisomerases introduce both single- and double-strand breaks in the course of changing the DNA's state of supercoiling, which is especially common in regions near an open replication fork. Such breaks are not considered DNA damage because they are a natural intermediate in the topoisomerase biochemical mechanism and are immediately repaired by the enzymes that created them.

A team of French researchers bombarded *Deinococcus radiodurans* to study the mechanism of double-strand break DNA repair in that organism. At least two copies of the genome, with random DNA breaks, can form DNA fragments through annealing. Partially overlapping fragments are then used for synthesis of homologous regions through a moving D-loop that can continue extension until they find complementary partner strands. In the final step there is crossover by means of RecA-dependent homologous recombination.

Translesion synthesis

Translesion synthesis is a DNA damage tolerance process that allows the DNA replication machinery to replicate past DNA lesions such as thymine dimers or AP sites. It involves switching out regular DNA polymerases for specialized translesion polymerases (e.g. DNA polymerase V), often with larger active sites that can facilitate the insertion of bases opposite damaged nucleotides. The polymerase switching is thought to be mediated by, among other factors, the post-translational modification of the replication processivity factor PCNA. Translesion synthesis polymerases often have low fidelity (high propensity to insert wrong bases) relative to regular polymerases. However, many are extremely efficient at inserting correct bases opposite specific types of damage. For example, Pol η mediates error-free bypass of lesions induced by UV irradiation, whereas Pol ζ introduces mutations at these sites. From a cellular perspective, risking the introduction of point mutations during translesion synthesis may be preferable to resorting to more drastic mechanisms of DNA repair, which may cause gross chromosomal aberrations or cell death. In short, the process involves specialized polymerases either bypassing or repairing lesions at locations of stalled DNA replication. A bypass platform is provided to these polymerases by Proliferating cell nuclear antigen (PCNA). Under normal circumstances, PCNA bound to polymerases replicates the DNA.

At a site of lesion, PCNA is ubiquitinated, or modified, by the RAD6/RAD18 proteins to provide a platform for the specialized polymerases to bypass the lesion and resume DNA replication.

Global response to DNA damage

Cells exposed to ionizing radiation, ultraviolet light or chemicals are prone to acquire multiple sites of bulky DNA lesions and double-strand breaks. Moreover, DNA damaging agents can damage other biomolecules such as proteins, carbohydrates, lipids, and RNA. The accumulation of damage, to be specific, double-strand breaks or adducts stalling the replication forks, are among known stimulation signals for a global response to DNA damage. The global response to damage is an act directed toward the cells' own preservation and triggers multiple pathways of macromolecular repair, lesion bypass, tolerance, or apoptosis. The common features of global response are induction of multiple genes, cell cycle arrest, and inhibition of cell division.

DNA damage checkpoints

After DNA damage, cell cycle checkpoints are activated. Checkpoint activation pauses the cell cycle and gives the cell time to repair the damage before continuing to divide. DNA damage checkpoints occur at the G1/S and G2/M boundaries. An intra-S checkpoint also exists. Checkpoint activation is controlled by two master kinases, ATM and ATR. ATM responds to DNA double-strand breaks and disruptions in chromatin structure, whereas ATR primarily responds to stalled replication forks. These kinases phosphorylate downstream targets in a signal transduction cascade, eventually leading to cell cycle arrest. A class of checkpoint mediator proteins including BRCA1, MDC1, and 53BP1 has also been identified. These proteins seem to be required for transmitting the checkpoint activation signal to downstream proteins.

P53 is an important downstream target of ATM and ATR, as it is required for inducing apoptosis following DNA damage. At the G1/S checkpoint, p53 functions by deactivating the CDK2/cyclin E complex. Similarly, p21 mediates the G2/M checkpoint by deactivating the CDK1/cyclin B complex.

The prokaryotic SOS response

The SOS response is the term used to describe changes in gene expression in *Escherichia coli* and other bacteria in response to extensive DNA damage. The prokaryotic SOS system is regulated by two key proteins: LexA and RecA. The LexA homodimer is a transcriptional repressor that binds to operator sequences commonly referred to as SOS boxes. In *Escherichia coli* it is known that LexA regulates transcription of approximately 48 genes including the *lexA* and *recA* genes. The SOS response is known to be widespread in the Bacteria domain, but it is mostly absent in some bacterial phyla, like the Spirochetes. The most common cellular signals activating the SOS response are regions of single-stranded DNA (ssDNA), arising from stalled replication forks or double-strand breaks, which are processed by DNA helicase to separate the two DNA

strands. In the initiation step, RecA protein binds to ssDNA in an ATP hydrolysis driven reaction creating RecA–ssDNA filaments. RecA–ssDNA filaments activate LexA autoprotease activity, which ultimately leads to cleavage of LexA dimer and subsequent LexA degradation. The loss of LexA repressor induces transcription of the SOS genes and allows for further signal induction, inhibition of cell division and an increase in levels of proteins responsible for damage processing.

In *Escherichia coli*, SOS boxes are 20-nucleotide long sequences near promoters with palindromic structure and a high degree of sequence conservation. In other classes and phyla, the sequence of SOS boxes varies considerably, with different length and composition, but it is always highly conserved and one of the strongest short signals in the genome. The high information content of SOS boxes permits differential binding of LexA to different promoters and allows for timing of the SOS response. The lesion repair genes are induced at the beginning of SOS response. The error-prone translesion polymerases, for example, UmuCD'2 (also called DNA polymerase V), are induced later on as a last resort. Once the DNA damage is repaired or bypassed using polymerases or through recombination, the amount of single-stranded DNA in cells is decreased, lowering the amounts of RecA filaments decreases cleavage activity of LexA homodimer, which then binds to the SOS boxes near promoters and restores normal gene expression.

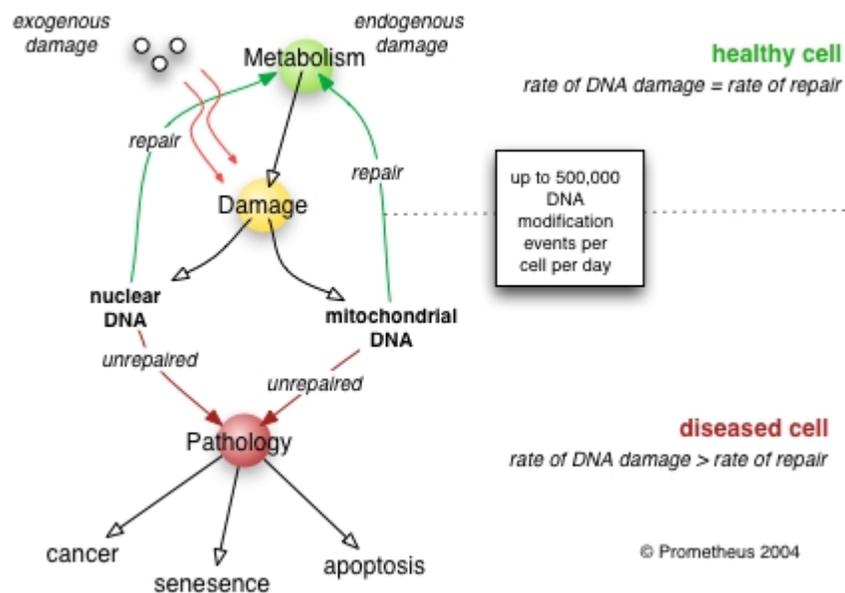
Eukaryotic transcriptional responses to DNA damage

Eukaryotic cells exposed to DNA damaging agents also activate important defensive pathways by inducing multiple proteins involved in DNA repair, cell cycle checkpoint control, protein trafficking and degradation. Such genome wide transcriptional response is very complex and tightly regulated, thus allowing coordinated global response to damage. Exposure of yeast *Saccharomyces cerevisiae* to DNA damaging agents results in overlapping but distinct transcriptional profiles. Similarities to environmental shock response indicates that a general global stress response pathway exist at the level of transcriptional activation. In contrast, different human cell types respond to damage differently indicating an absence of a common global response. The probable explanation for this difference between yeast and human cells may be in the heterogeneity of mammalian cells. In an animal different types of cells are distributed among different organs that have evolved different sensitivities to DNA damage.

In general global response to DNA damage involves expression of multiple genes responsible for postreplication repair, homologous recombination, nucleotide excision repair, DNA damage checkpoint, global transcriptional activation, genes controlling mRNA decay, and many others. A large amount of damage to a cell leaves it with an important decision: undergo apoptosis and die, or survive at the cost of living with a modified genome. An increase in tolerance to damage can lead to an increased rate of survival that will allow a greater accumulation of mutations. Yeast Rev1 and human polymerase η are members of [Y family translesion DNA polymerases present during global response to DNA damage and are responsible for enhanced mutagenesis during a global response to DNA damage in eukaryotes.

DNA repair and aging

Pathological effects of poor DNA repair

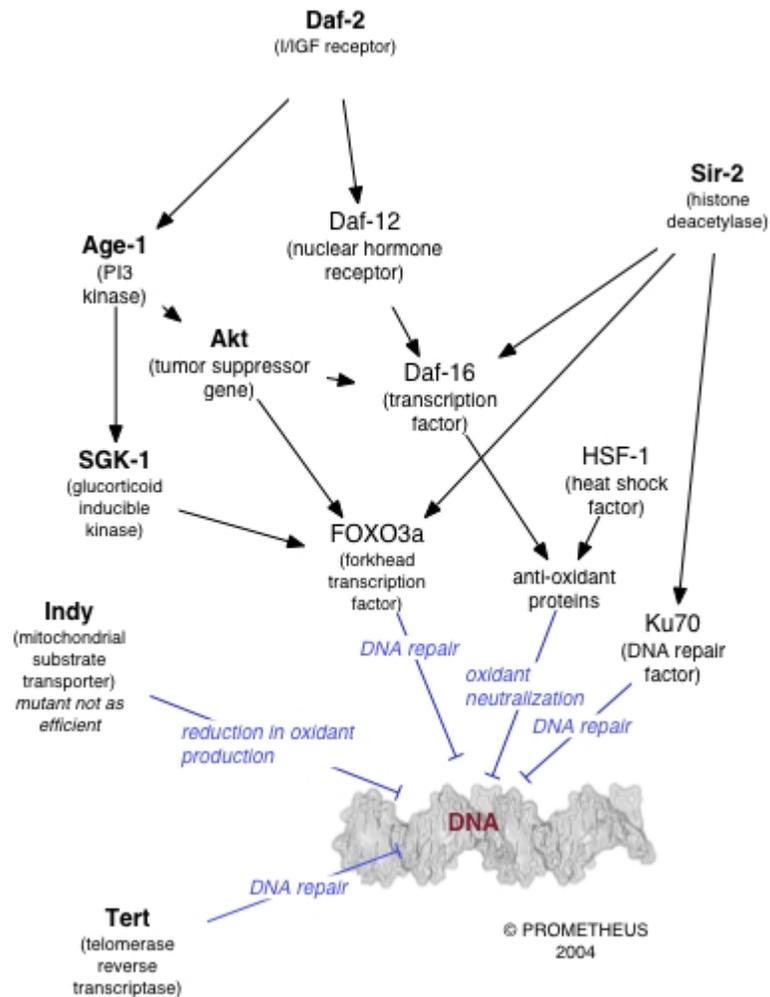


DNA repair rate is an important determinant of cell pathology

Experimental animals with genetic deficiencies in DNA repair often show decreased life span and increased cancer incidence. For example, mice deficient in the dominant NHEJ pathway and in telomere maintenance mechanisms get lymphoma and infections more often, and, as a consequence, have shorter lifespans than wild-type mice. In similar manner, mice deficient in a key repair and transcription protein that unwinds DNA helices have premature onset of aging-related diseases and consequent shortening of lifespan. However, not every DNA repair deficiency creates exactly the predicted effects; mice deficient in the NER pathway exhibited shortened life span without correspondingly higher rates of mutation.

If the rate of DNA damage exceeds the capacity of the cell to repair it, the accumulation of errors can overwhelm the cell and result in early senescence, apoptosis, or cancer. Inherited diseases associated with faulty DNA repair functioning result in premature aging, increased sensitivity to carcinogens, and correspondingly increased cancer risk (see below). On the other hand, organisms with enhanced DNA repair systems, such as *Deinococcus radiodurans*, the most radiation-resistant known organism, exhibit remarkable resistance to the double-strand break-inducing effects of radioactivity, likely due to enhanced efficiency of DNA repair and especially NHEJ.

Longevity and caloric restriction



Most life span influencing genes affect the rate of DNA damage

A number of individual genes have been identified as influencing variations in life span within a population of organisms. The effects of these genes is strongly dependent on the environment, in particular, on the organism's diet. Caloric restriction reproducibly results in extended lifespan in a variety of organisms, likely via nutrient sensing pathways and decreased metabolic rate. The molecular mechanisms by which such restriction results in lengthened lifespan are as yet unclear; however, the behavior of many genes known to be involved in DNA repair is altered under conditions of caloric restriction.

For example, increasing the gene dosage of the gene SIR-2, which regulates DNA packaging in the nematode worm *Caenorhabditis elegans*, can significantly extend lifespan. The mammalian homolog of SIR-2 is known to induce downstream DNA repair factors involved in NHEJ, an activity that is especially promoted under conditions of caloric restriction. Caloric restriction has been closely linked to the rate of base excision repair in the nuclear DNA of rodents, although similar effects have not been observed in mitochondrial DNA.

It is interesting to note that the *C. elegans* gene AGE-1, an upstream effector of DNA repair pathways, confers dramatically extended life span under free-feeding conditions but leads to a decrease in reproductive fitness under conditions of caloric restriction. This observation supports the pleiotropy theory of the biological origins of aging, which suggests that genes conferring a large survival advantage early in life will be selected for even if they carry a corresponding disadvantage late in life.

Medicine and DNA repair modulation

Hereditary DNA repair disorders

Defects in the NER mechanism are responsible for several genetic disorders, including:

- xeroderma pigmentosum: hypersensitivity to sunlight/UV, resulting in increased skin cancer incidence and premature aging
- Cockayne syndrome: hypersensitivity to UV and chemical agents
- trichothiodystrophy: sensitive skin, brittle hair and nails

Mental retardation often accompanies the latter two disorders, suggesting increased vulnerability of developmental neurons.

Other DNA repair disorders include:

- Werner's syndrome: premature aging and retarded growth
- Bloom's syndrome: sunlight hypersensitivity, high incidence of malignancies (especially leukemias).
- ataxia telangiectasia: sensitivity to ionizing radiation and some chemical agents

All of the above diseases are often called "segmental progerias" ("accelerated aging diseases") because their victims appear elderly and suffer from aging-related diseases at an abnormally young age, while not manifesting all the symptoms of old age.

Other diseases associated with reduced DNA repair function include Fanconi's anemia, hereditary breast cancer and hereditary colon cancer.

DNA repair and cancer

Inherited mutations that affect DNA repair genes are strongly associated with high cancer risks in humans. Hereditary nonpolyposis colorectal cancer (HNPCC) is strongly associated with specific mutations in the DNA mismatch repair pathway. BRCA1 and BRCA2, two famous mutations conferring a hugely increased risk of breast cancer on carriers, are both associated with a large number of DNA repair pathways, especially NHEJ and homologous recombination.

Cancer therapy procedures such as chemotherapy and radiotherapy work by overwhelming the capacity of the cell to repair DNA damage, resulting in cell death.

Cells that are most rapidly dividing - most typically cancer cells - are preferentially affected. The side-effect is that other non-cancerous but rapidly dividing cells such as stem cells in the bone marrow are also affected. Modern cancer treatments attempt to localize the DNA damage to cells and tissues only associated with cancer, either by physical means (concentrating the therapeutic agent in the region of the tumor) or by biochemical means (exploiting a feature unique to cancer cells in the body).

DNA repair and evolution

The basic processes of DNA repair are highly conserved among both prokaryotes and eukaryotes and even among bacteriophage (viruses that infect bacteria); however, more complex organisms with more complex genomes have correspondingly more complex repair mechanisms. The ability of a large number of protein structural motifs to catalyze relevant chemical reactions has played a significant role in the elaboration of repair mechanisms during evolution.

The fossil record indicates that single-cell life began to proliferate on the planet at some point during the Precambrian period, although exactly when recognizably modern life first emerged is unclear. Nucleic acids became the sole and universal means of encoding genetic information, requiring DNA repair mechanisms that in their basic form have been inherited by all extant life forms from their common ancestor. The emergence of Earth's oxygen-rich atmosphere (known as the "oxygen catastrophe") due to photosynthetic organisms, as well as the presence of potentially damaging free radicals in the cell due to oxidative phosphorylation, necessitated the evolution of DNA repair mechanisms that act specifically to counter the types of damage induced by oxidative stress.

Rate of evolutionary change

On some occasions, DNA damage is not repaired, or is repaired by an error-prone mechanism that results in a change from the original sequence. When this occurs, mutations may propagate into the genomes of the cell's progeny. Should such an event occur in a germ line cell that will eventually produce a gamete, the mutation has the potential to be passed on to the organism's offspring. The rate of evolution in a particular species (or, in a particular gene) is a function of the rate of mutation. As a consequence, the rate and accuracy of DNA repair mechanisms have an influence over the process of evolutionary change.