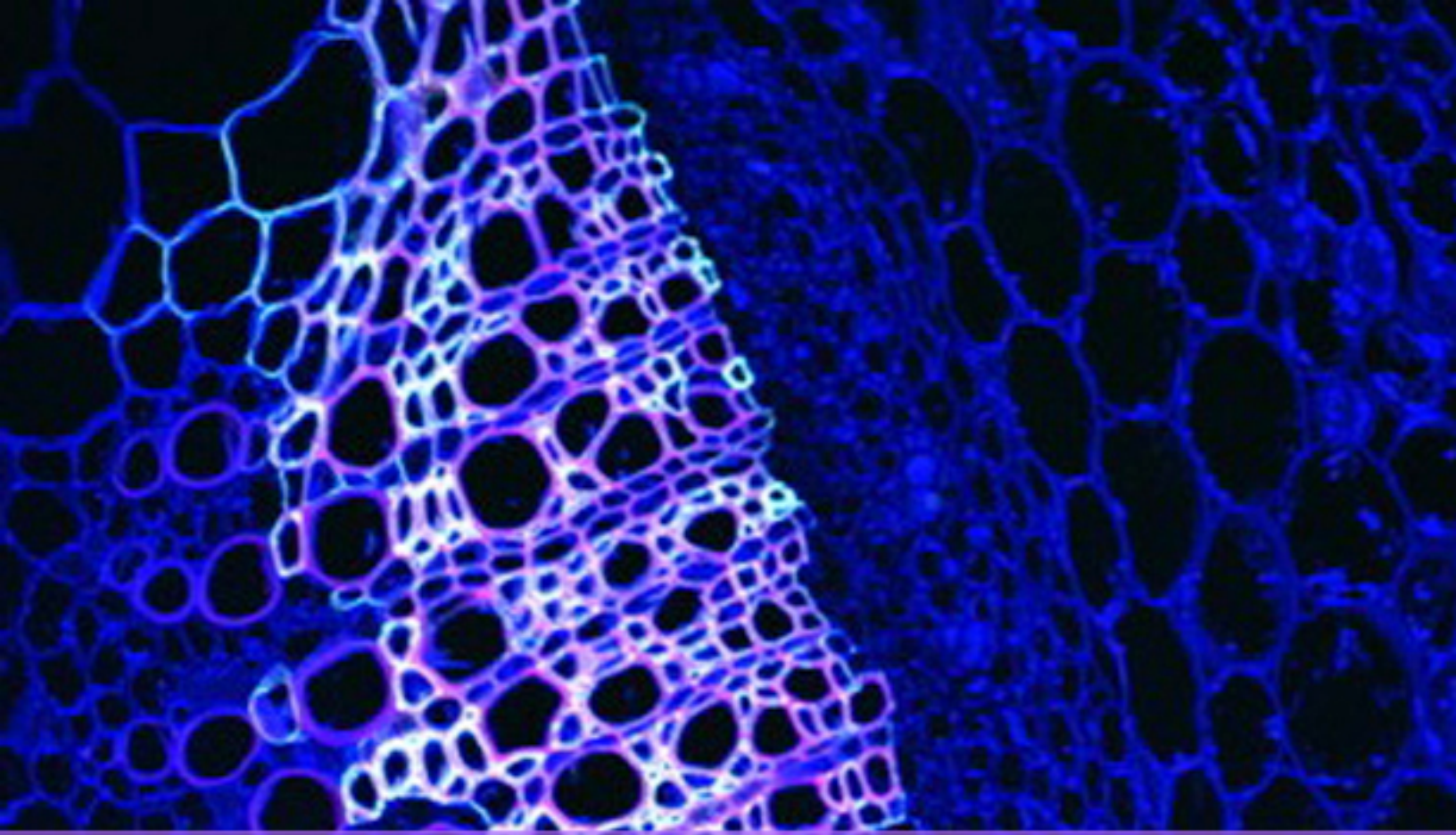


Cellular Anatomy

Lurlene Lemus

First Edition, 2012

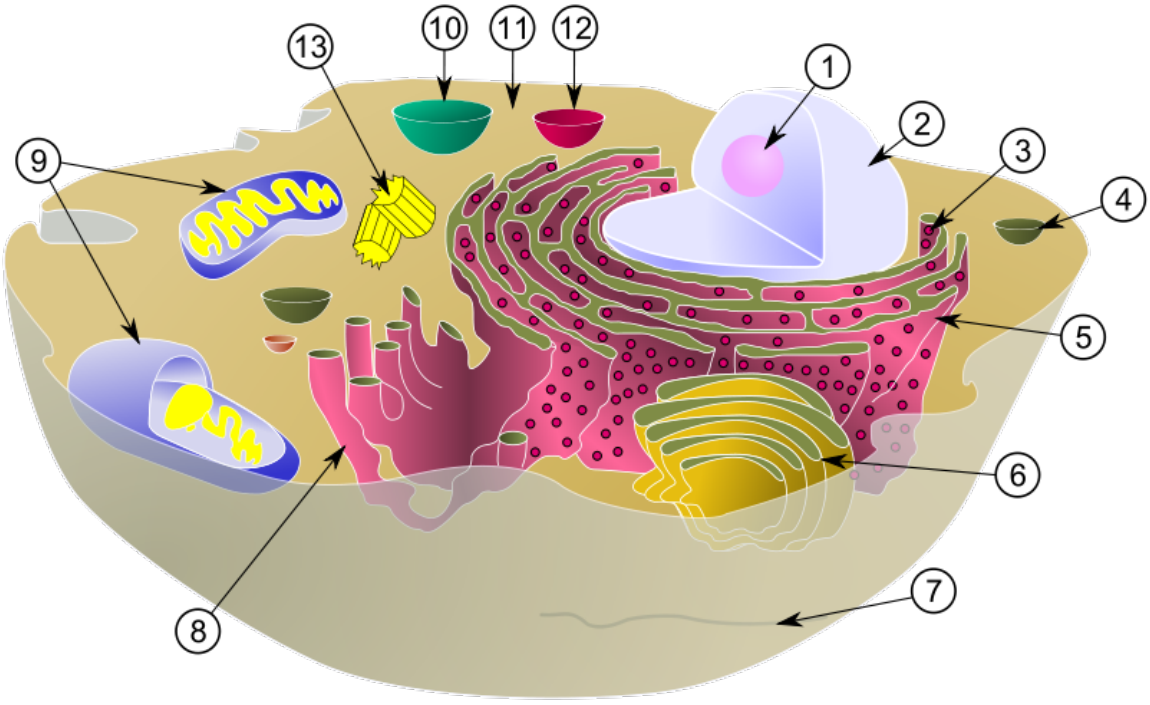
ISBN 978-81-323-3136-0

Published by:
Research World
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter- 1

Organelle



A typical animal cell. Within the cytoplasm, the major organelles and cellular structures include: (1) nucleolus (2) nucleus (3) ribosome (4) vesicle (5) rough endoplasmic reticulum (6) Golgi apparatus (7) cytoskeleton (8) smooth endoplasmic reticulum (9) mitochondria (10) vacuole (11) cytosol (12) lysosome (13) centriole.

In cell biology, an **organelle** is a specialized subunit within a cell that has a specific function, and is usually separately enclosed within its own lipid bilayer.

The name *organelle* comes from the idea that these structures are to cells what an organ is to the body (hence the name *organelle,* the suffix *-elle* being a diminutive). Organelles are identified by microscopy, and can also be purified by cell fractionation. There are many types of organelles, particularly in eukaryotic cells. Prokaryotes were once thought not to have organelles, but some examples have now been identified.

History and terminology

In biology, *organs* are defined as confined functional units within an organism. The analogy of bodily organs to microscopic cellular substructures is obvious, as from even early works, authors of respective textbooks rarely elaborate on the distinction between the two.

Credited as the first to use a diminutive of *organ* (*i.e.* little organ) for cellular structures was German zoologist Karl August Möbius (1884), who used the term "organula" (plural form of *organulum*, the diminutive of latin *organum*). From the context, it is clear that he referred to reproduction related structures of protists. In a footnote, which was published as a correction in the next issue of the journal, he justified his suggestion to call organs of unicellular organisms "organella" since they are only differently formed parts of one cell, in contrast to multicellular organs of multicellular organisms. Thus, the original definition was limited to structures of unicellular organisms.

It would take several years before *organulum*, or the later term *organelle*, became accepted and expanded in meaning to include subcellular structures in multicellular organisms. Books around 1900 from Valentin Häcker, Edmund Wilson and Oscar Hertwig still referred to cellular *organs*. Later, both terms came to be used side by side: Bengt Lidforss wrote 1915 (in German) about "Organs or Organells".

Around 1920, the term organelle was used to describe propulsion structures ("motor organelle complex", i.e., flagella and their anchoring) and other protist structures, such as ciliates. Alfred Kühn wrote about centrioles as division organelles, although he stated that, for Vahlkampfias, the alternative 'organelle' or 'product of structural build-up' had not yet been decided, without explaining the difference between the alternatives.

In his 1953 textbook, Max Hartmann used the term for extracellular (pellicula, shells, cell walls) and intracellular skeletons of protists.

Later, the now-widely-used definition of organelle emerged, after which only cellular structures with surrounding membrane had been considered organelles. However, the more original definition of subcellular functional unit in general still coexists.

In 1978, Albert Frey-Wyssling suggested that the term organelle should refer only to structures that convert energy, such as centrosomes, ribosomes, and nucleoli. This new definition, however, did not win wide recognition.

Examples

While most cell biologists consider the term **organelle** to be synonymous with "cell compartment", other cell biologists choose to limit the term organelle to include only those that are DNA-containing, having originated from formerly-autonomous microscopic organisms acquired via endosymbiosis.

The most notable of these organelles having originated from endosymbiont bacteria are:

- mitochondria (in almost all eukaryotes)
- chloroplasts (in plants, algae and protists).

Other organelles are also suggested to have endosymbiotic origins.

Under the more restricted definition of membrane-bound structures, some parts of the cell do not qualify as organelles. Nevertheless, the use of organelle to refer to non-membrane bound structures such as ribosomes is common. This has led some texts to delineate between membrane-bound and non-membrane bound organelles. These structures are large assemblies of macromolecules that carry out particular and specialized functions, but they lack membrane boundaries. Such cell structures include:

- ribosome
- cytoskeleton
- flagellum
- centriole and microtubule-organizing center (MTOC).

Eukaryotic organelles

Eukaryotes are one of the structurally complex cell type, and by definition are in part organized by smaller interior compartments, that are themselves enclosed by lipid membranes that resemble the outermost cell membrane. The larger organelles, such as the nucleus and vacuoles, are easily visible with the light microscope. They were among the first biological discoveries made after the invention of the microscope.

Not all eukaryotic cells have each of the organelles listed below. Exceptional organisms have cells which do not include some organelles that might otherwise be considered universal to eukaryotes (such as mitochondria). There are also occasional exceptions to the number of membranes surrounding organelles, listed in the tables below (e.g., some that are listed as double-membrane are sometimes found with single or triple membranes). In addition, the number of individual organelles of each type found in a given cell varies depending upon the function of that cell.

Major eukaryotic organelles

Organelle	Main function	Structure	Organisms	Notes
chloroplast (plastid)	photosynthesis	double-membrane compartment	plants, protists (rare kleptoplastic organisms)	has some genes; theorized to be engulfed by the ancestral eukaryotic cell (endosymbiosis)
endoplasmic reticulum	translation and folding of new proteins (rough	single-membrane compartment	all eukaryotes	rough endoplasmic reticulum is covered with ribosomes, has

	endoplasmic reticulum), expression of lipids (smooth endoplasmic reticulum)			folds that are flat sacs; smooth endoplasmic reticulum has folds that are tubular
Golgi apparatus	sorting and modification of proteins	single-membrane compartment	all eukaryotes	cis-face (convex) nearest to rough endoplasmic reticulum; trans-face (concave) farthest from rough endoplasmic reticulum
mitochondria	energy production	double-membrane compartment	most eukaryotes	has some DNA; theorized to be engulfed by an ancestral eukaryotic cell (endosymbiosis)
vacuole	storage, helps maintain homeostasis	single-membrane compartment	eukaryotes	
nucleus	DNA maintenance, RNA transcription	double-membrane compartment	all eukaryotes	contains bulk of genome

Mitochondria and chloroplasts, which have double-membranes and their own DNA, are believed to have originated from incompletely consumed or invading prokaryotic organisms, which were adopted as a part of the invaded cell. This idea is supported in the Endosymbiotic theory.

Minor eukaryotic organelles and cell components

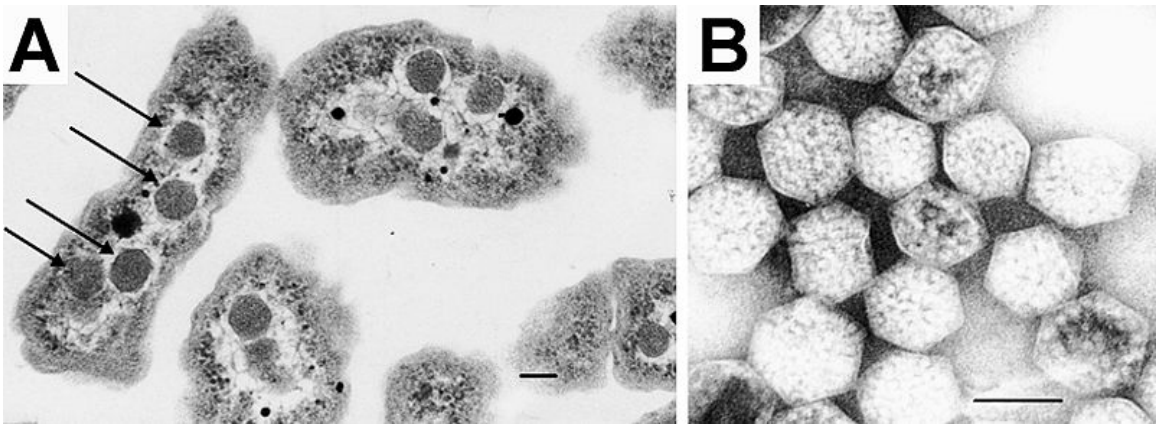
Organelle/Macromolecule	Main function	Structure	Organisms
acrosome	helps spermatoza fuse with ovum	single-membrane compartment	many animals
autophagosome	vesicle which sequesters cytoplasmic material and organelles for degradation	double-membrane compartment	all eukaryotic cells
centriole	anchor for cytoskeleton, helps in cell division	Microtubule protein	animals
cilium	movement in or of external medium; "critical developmental signaling pathway".	Microtubule protein	animals, protists, few plants

eyespot apparatus	detects light, allowing phototaxis to take place		green algae and other unicellular photosynthetic organisms such as euglenids
glycosome	carries out glycolysis	single-membrane compartment	Some protozoa, such as *Trypanosomes*.
glyoxysome	conversion of fat into sugars	single-membrane compartment	plants
hydrogenosome	energy & hydrogen production	double-membrane compartment	a few unicellular eukaryotes
lysosome	breakdown of large molecules (e.g., proteins + polysaccharides)	single-membrane compartment	most eukaryotes
melanosome	pigment storage	single-membrane compartment	animals
mitosome	not characterized	double-membrane compartment	a few unicellular eukaryotes
myofibril	muscular contraction	bundled filaments	animals
nucleolus	ribosome production	protein-DNA-RNA	most eukaryotes
parenthesome	not characterized	not characterized	fungi
peroxisome	breakdown of metabolic hydrogen peroxide	single-membrane compartment	all eukaryotes
ribosome	translation of RNA into proteins	RNA-protein	eukaryotes, prokaryotes
vesicle	material transport	single-membrane compartment	all eukaryotes

Other related structures:

- cytosol
- endomembrane system
- nucleosome

- microtubule
- cell membrane



(A) Electron micrograph of *Halothiobacillus neapolitanus* cells, arrows highlight carboxysomes. (B) Image of intact carboxysomes isolated from *H. neapolitanus*. Scale bars are 100 nm.

Prokaryotic organelles

Prokaryotes are not as structurally complex as eukaryotes, and were once thought not to have any internal structures enclosed by lipid membranes. In the past, they were often viewed as having little internal organization; but, slowly, details are emerging about prokaryotic internal structures. An early false turn was the idea developed in the 1970s that bacteria might contain membrane folds termed mesosomes, but these were later shown to be artifacts produced by the chemicals used to prepare the cells for electron microscopy.

However, more recent research has revealed that at least some prokaryotes have microcompartments such as carboxysomes. These subcellular compartments are 100 - 200 nm in diameter and are enclosed by a shell of proteins. Even more striking is the description of membrane-bound magnetosomes in bacteria, as well as the nucleus-like structures of the *Planctomycetes* that are surrounded by lipid membranes.

Prokaryotic organelles and cell components

Organelle/Macromolecule	Main function	Structure	Organisms
carboxysome	carbon fixation	protein-shell compartment	some bacteria
chlorosome	photosynthesis	light harvesting complex	green sulfur bacteria
flagellum	movement in external medium	protein filament	some prokaryotes and eukaryotes
magnetosome	magnetic orientation	inorganic crystal, lipid membrane	magnetotactic bacteria

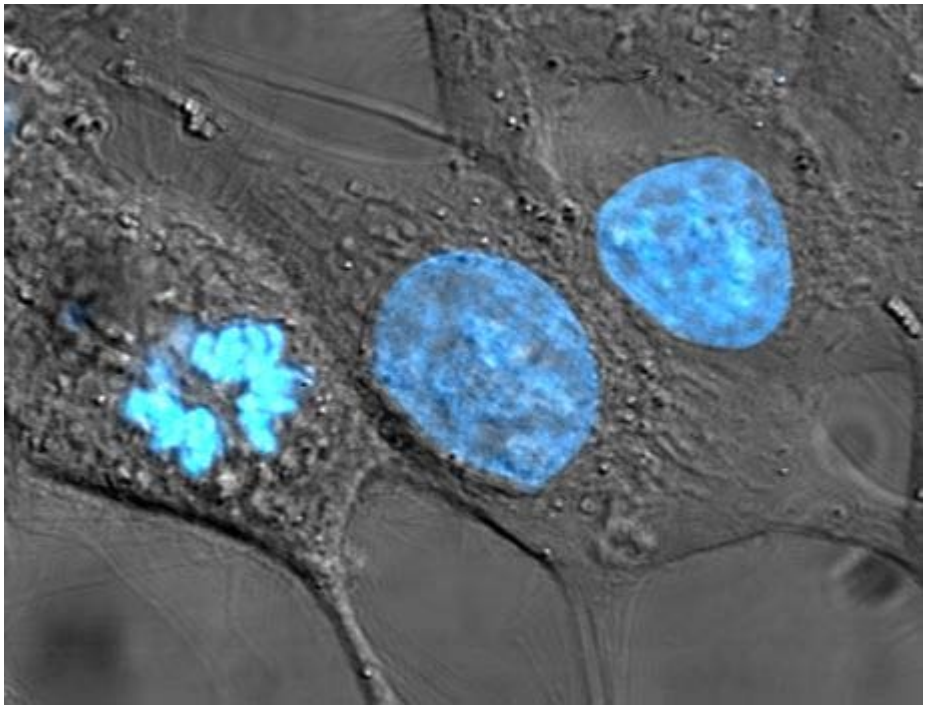
nucleoid	DNA maintenance, transcription to RNA	DNA-protein	prokaryotes
plasmid	DNA exchange	circular DNA	some bacteria
ribosome	translation of RNA into proteins	RNA-protein	eukaryotes, prokaryotes
thylakoid	photosynthesis	photosystem proteins and pigments	mostly cyanobacteria

Proteins and organelles

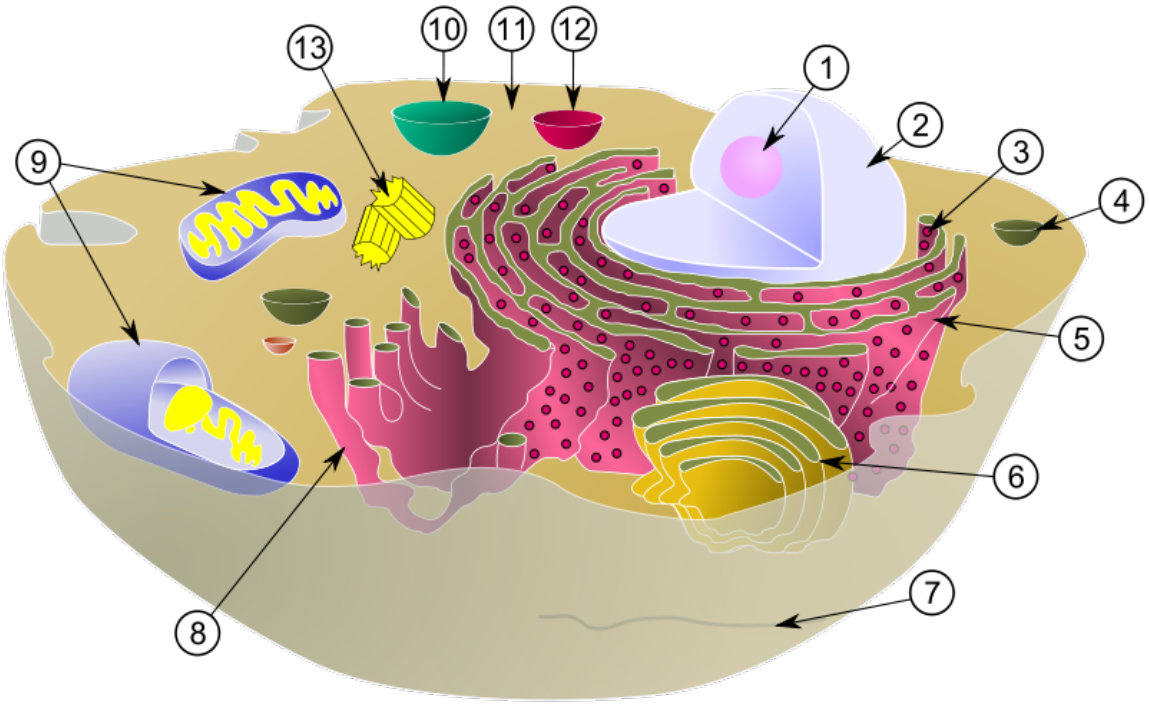
The function of a protein is closely correlated with the organelle in which it resides. Some methods were proposed for predicting the organelle in which an uncharacterized protein is located according to its amino acid composition and some methods were based on pseudo amino acid composition.

Chapter- 2

Cell Nucleus



HeLa cells stained for DNA with the Blue Hoechst dye. The central and rightmost cell are in interphase, thus teir entire nuclei are labeled. On the left a cell is going through mitosis and its DNA has condensed ready for division.



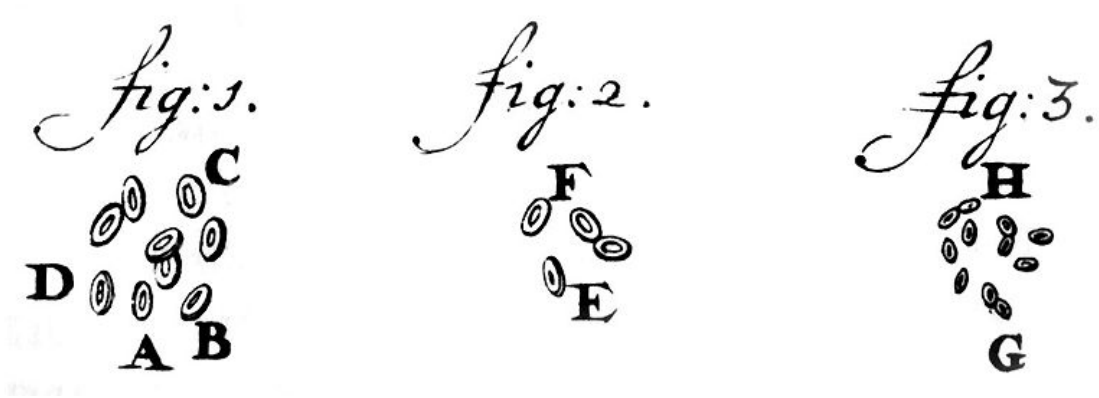
Schematic of typical animal cell, showing subcellular components. Organelles: (1) nucleolus (2) **nucleus** (3) ribosome (4) vesicle (5) rough endoplasmic reticulum (ER) (6) Golgi apparatus (7) Cytoskeleton (8) smooth ER (9) mitochondria (10) vacuole (11) cytoplasm (12) lysosome (13) centrioles

In cell biology, the **nucleus** (pl. *nuclei*; from Latin *nucleus* or *nuculeus*, meaning kernel) is a membrane enclosed organelle found in eukaryotic cells. It contains most of the cell's genetic material, organized as multiple long linear DNA molecules in complex with a large variety of proteins, such as histones, to form chromosomes. The genes within these chromosomes are the cell's nuclear genome. The function of the nucleus is to maintain the integrity of these genes and to control the activities of the cell by regulating gene expression — the nucleus is therefore the control center of the cell. The main structures making up the nucleus are the nuclear envelope, a double membrane that encloses the entire organelle and separates its contents from the cellular cytoplasm, and the nuclear lamina, a meshwork within the nucleus that adds mechanical support, much like the cytoskeleton supports the cell as a whole. Because the nuclear membrane is impermeable to most molecules, nuclear pores are required to allow movement of molecules across the envelope. These pores cross both of the membranes, providing a channel that allows free movement of small molecules and ions. The movement of larger molecules such as proteins is carefully controlled, and requires active transport regulated by carrier proteins. Nuclear transport is crucial to cell function, as movement through the pores is required for both gene expression and chromosomal maintenance.

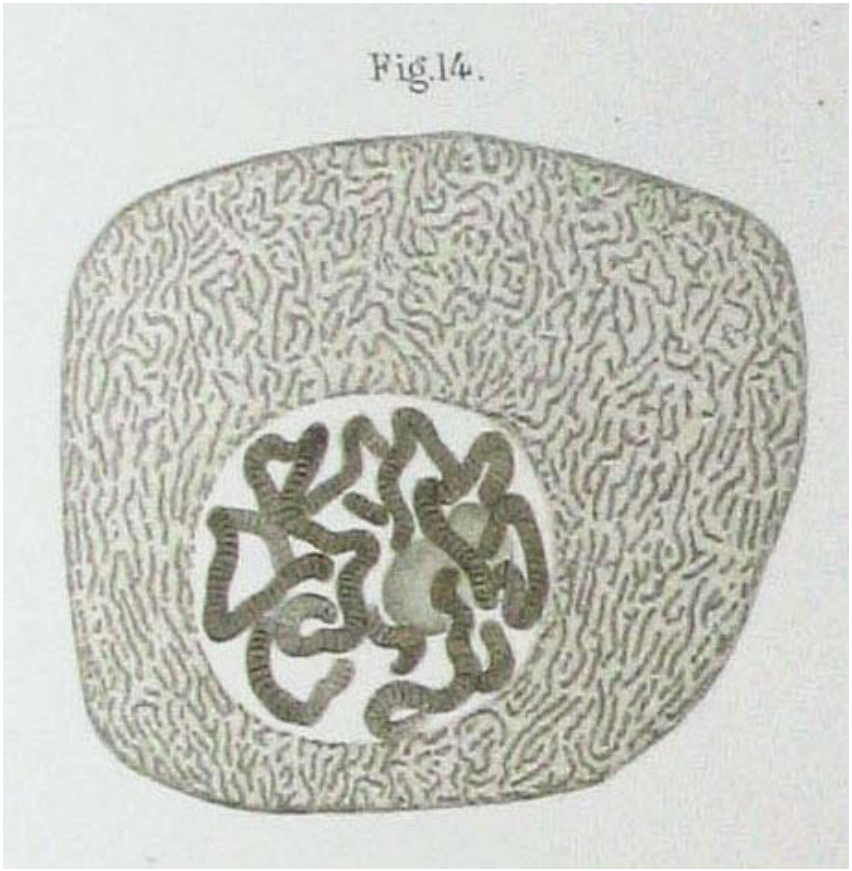
Although the interior of the nucleus does not contain any membrane-bound subcompartments, its contents are not uniform, and a number of *subnuclear bodies* exist, made up of unique proteins, RNA molecules, and particular parts of the chromosomes.

The best known of these is the nucleolus, which is mainly involved in the assembly of ribosomes. After being produced in the nucleolus, ribosomes are exported to the cytoplasm where they translate mRNA.

History



Oldest known depiction of cells and their nuclei by Antonie van Leeuwenhoek, 1719.



Drawing of a *Chironomus* salivary gland cell published by Walther Flemming in 1882. The nucleus contains Polytene chromosomes.

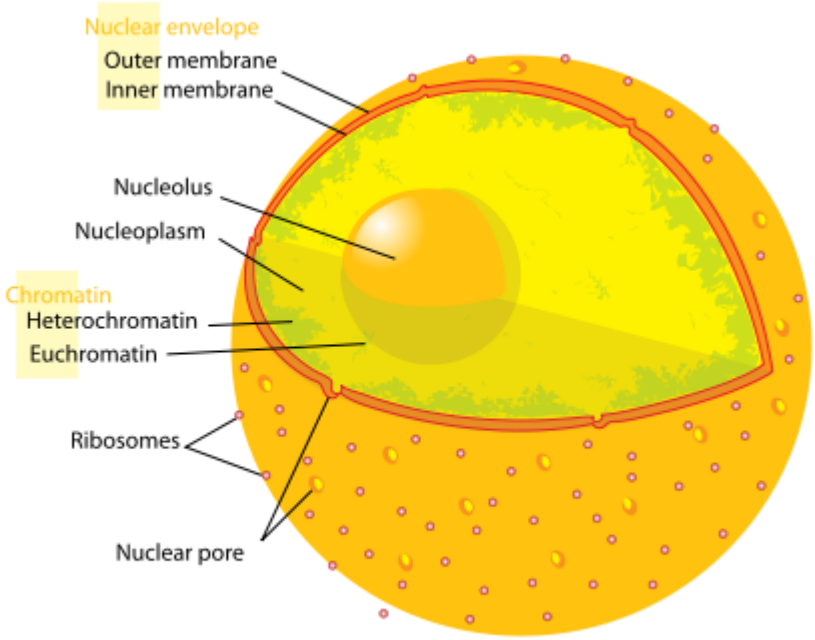
The nucleus was the first organelle to be discovered. The probably oldest preserved drawing dates back to the early microscopist Antonie van Leeuwenhoek (1632 – 1723). He observed a "Lumen", the nucleus, in the red blood cells of salmon. Unlike mammalian red blood cells, those of other vertebrates still possess nuclei. The nucleus was also described by Franz Bauer in 1804 and in more detail in 1831 by Scottish botanist Robert Brown in a talk at the Linnean Society of London. Brown was studying orchids microscopically when he observed an opaque area, which he called the areola or nucleus, in the cells of the flower's outer layer. He did not suggest a potential function. In 1838 Matthias Schleiden proposed that the nucleus plays a role in generating cells, thus he introduced the name "Cytoblast" (cell builder). He believed that he had observed new cells assembling around "cytoblasts". Franz Meyen was a strong opponent of this view having already described cells multiplying by division and believing that many cells would have no nuclei. The idea that cells can be generated de novo, by the "cytoblast" or otherwise, contradicted work by Robert Remak (1852) and Rudolf Virchow (1855) who decisively propagated the new paradigm that cells are generated solely by cells ("Omnis cellula e cellula"). The function of the nucleus remained unclear.

Between 1876 and 1878 Oscar Hertwig published several studies on the fertilization of sea urchin eggs, showing that the nucleus of the sperm enters the oocyte and fuses with its nucleus. This was the first time it was suggested that an individual develops from a (single) nucleated cell. This was in contradiction to Ernst Haeckel's theory that the complete phylogeny of a species would be repeated during embryonic development, including generation of the first nucleated cell from a "Monerula", a structureless mass of primordial mucus ("Urschleim"). Therefore, the necessity of the sperm nucleus for fertilization was discussed for quite some time. However, Hertwig confirmed his observation in other animal groups, e.g. amphibians and molluscs. Eduard Strasburger produced the same results for plants (1884). This paved the way to assign the nucleus an important role in heredity. In 1873 August Weismann postulated the equivalence of the maternal and paternal germ *cells* for heredity. The function of the nucleus as carrier of genetic information became clear only later, after mitosis was discovered and the Mendelian rules were rediscovered at the beginning of the 20th century; the chromosome theory of heredity was developed.

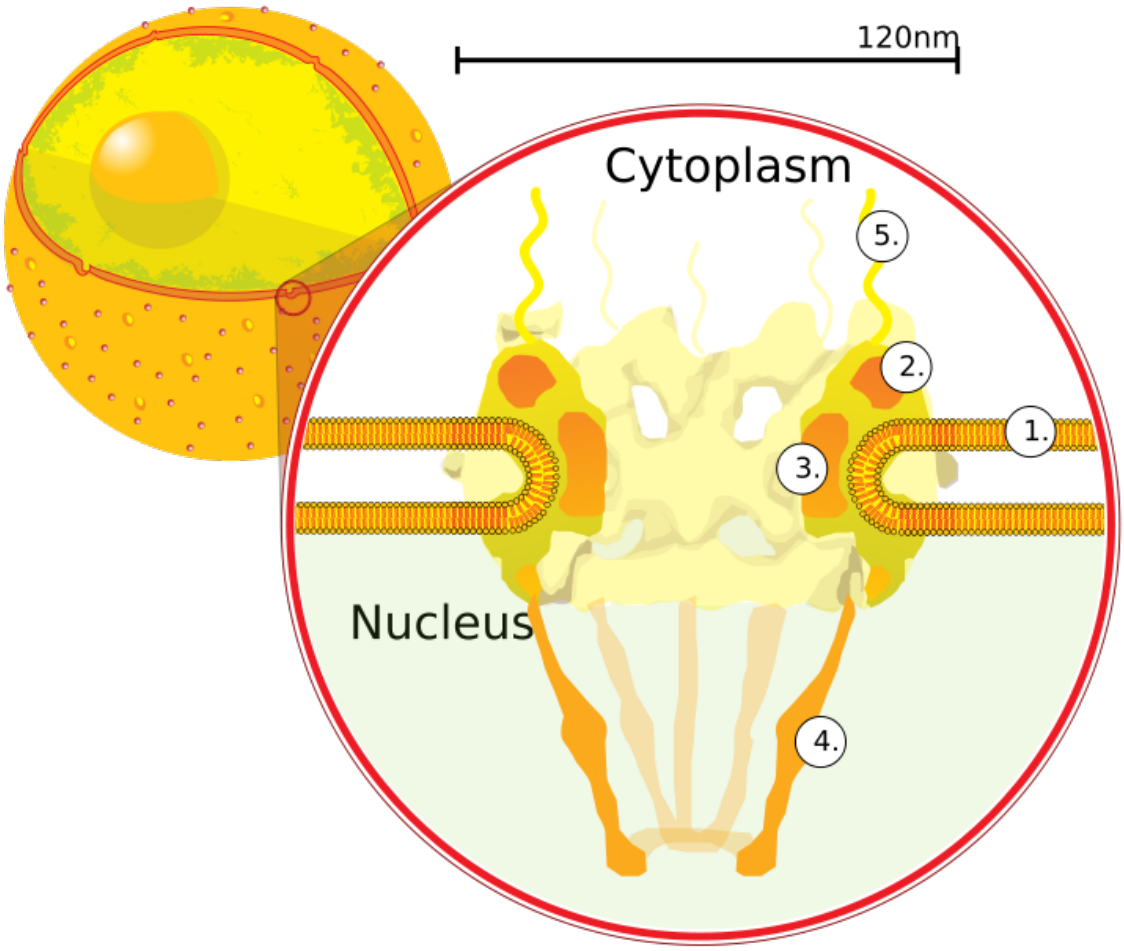
Structures

The nucleus is the largest cellular organelle in animals. In mammalian cells, the average diameter of the nucleus is approximately 6 micrometers (μm), which occupies about 10% of the total cell volume. The viscous liquid within it is called nucleoplasm, and is similar in composition to the cytosol found outside the nucleus. It appears as a dense, roughly spherical organelle.

Nuclear envelope and pores



The eukaryotic cell nucleus. Visible in this diagram are the ribosome-studded double membranes of the nuclear envelope, the DNA (complexed as chromatin), and the nucleolus. Within the cell nucleus is a viscous liquid called nucleoplasm, similar to the cytoplasm found outside the nucleus.



A cross section of a nuclear pore on the surface of the nuclear envelope (1). Other diagram labels show (2) the outer ring, (3) spokes, (4) basket, and (5) filaments.

The nuclear envelope otherwise known as nuclear membrane consists of two cellular membranes, an inner and an outer membrane, arranged parallel to one another and separated by 10 to 50 nanometers (nm). The nuclear envelope completely encloses the nucleus and separates the cell's genetic material from the surrounding cytoplasm, serving as a barrier to prevent macromolecules from diffusing freely between the nucleoplasm and the cytoplasm. The outer nuclear membrane is continuous with the membrane of the rough endoplasmic reticulum (RER), and is similarly studded with ribosomes. The space between the membranes is called the perinuclear space and is continuous with the RER lumen.

Nuclear pores, which provide aqueous channels through the envelope, are composed of multiple proteins, collectively referred to as nucleoporins. The pores are about 125 million daltons in molecular weight and consist of around 50 (in yeast) to 100 proteins (in vertebrates). The pores are 100& nbsp;nm in total diameter; however, the gap through which molecules freely diffuse is only about 9 nm wide, due to the presence of regulatory systems within the center of the pore. This size allows the free passage of small water-soluble molecules while preventing larger molecules, such as nucleic acids and larger

proteins, from inappropriately entering or exiting the nucleus. These large molecules must be actively transported into the nucleus instead. The nucleus of a typical mammalian cell will have about 3000 to 4000 pores throughout its envelope,(ref name="Rhoades")Rodney Rhoades, Richard Pflanzer, ed (1996). "Ch3". *Human Physiology* (3rd ed.). Saunders College Publishing.</ref> each of which contains a donut-shaped, eightfold-symmetric ring-shaped structure at a position where the inner and outer membranes fuse. Attached to the ring is a structure called the *nuclear basket* that extends into the nucleoplasm, and a series of filamentous extensions that reach into the cytoplasm. Both structures serve to mediate binding to nuclear transport proteins.

Most proteins, ribosomal subunits, and some RNAs are transported through the pore complexes in a process mediated by a family of transport factors known as karyopherins. Those karyopherins that mediate movement into the nucleus are also called importins, while those that mediate movement out of the nucleus are called exportins. Most karyopherins interact directly with their cargo, although some use adaptor proteins. Steroid hormones such as cortisol and aldosterone, as well as other small lipid-soluble molecules involved in intercellular signaling can diffuse through the cell membrane and into the cytoplasm, where they bind nuclear receptor proteins that are trafficked into the nucleus. There they serve as transcription factors when bound to their ligand; in the absence of ligand many such receptors function as histone deacetylases that repress gene expression.

Nuclear lamina

In animal cells, two networks of intermediate filaments provide the nucleus with mechanical support: the nuclear lamina forms an organized meshwork on the internal face of the envelope, while less organized support is provided on the cytosolic face of the envelope. Both systems provide structural support for the nuclear envelope and anchoring sites for chromosomes and nuclear pores.

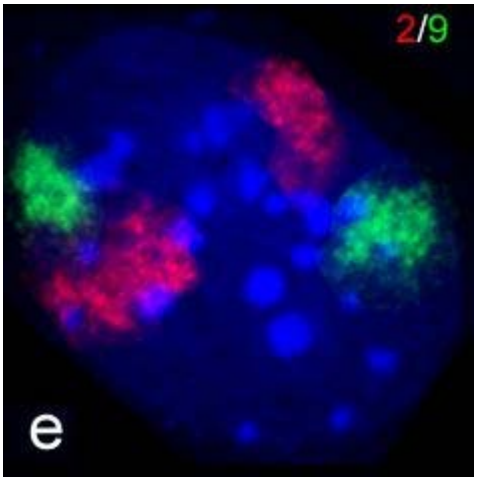
The nuclear lamina is mostly composed of lamin proteins. Like all proteins, lamins are synthesized in the cytoplasm and later transported into the nucleus interior, where they are assembled before being incorporated into the existing network of nuclear lamina. Lamins are also found inside the nucleoplasm where they form another regular structure, known as the *nucleoplasmic veil*, that is visible using fluorescence microscopy. The actual function of the veil is not clear, although it is excluded from the nucleolus and is present during interphase. The lamin structures that make up the veil bind chromatin and disrupting their structure inhibits transcription of protein-coding genes.

Like the components of other intermediate filaments, the lamin monomer contains an alpha-helical domain used by two monomers to coil around each other, forming a dimer structure called a coiled coil. Two of these dimer structures then join side by side, in an antiparallel arrangement, to form a tetramer called a *protofilament*. Eight of these protofilaments form a lateral arrangement that is twisted to form a ropelike *filament*. These filaments can be assembled or disassembled in a dynamic manner, meaning that

changes in the length of the filament depend on the competing rates of filament addition and removal.

Mutations in lamin genes leading to defects in filament assembly are known as *laminopathies*. The most notable laminopathy is the family of diseases known as progeria, which causes the appearance of premature aging in its sufferers. The exact mechanism by which the associated biochemical changes give rise to the aged phenotype is not well understood.

Chromosomes



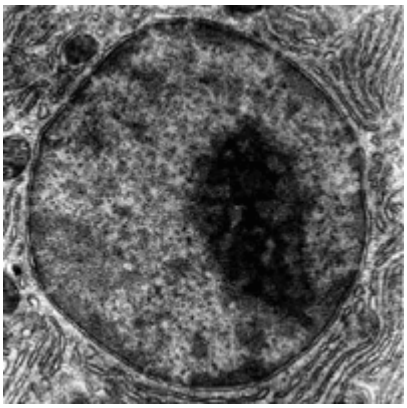
A mouse fibroblast nucleus in which DNA is stained blue. The distinct chromosome territories of chromosome 2 (red) and chromosome 9 (green) are visible stained with fluorescent in situ hybridization.

The cell nucleus contains the majority of the cell's genetic material, in the form of multiple linear DNA molecules organized into structures called chromosomes. During most of the cell cycle these are organized in a DNA-protein complex known as chromatin, and during cell division the chromatin can be seen to form the well defined chromosomes familiar from a karyotype. A small fraction of the cell's genes are located instead in the mitochondria.

There are two types of chromatin. Euchromatin is the less compact DNA form, and contains genes that are frequently expressed by the cell. The other type, heterochromatin, is the more compact form, and contains DNA that are infrequently transcribed. This structure is further categorized into *facultative* heterochromatin, consisting of genes that are organized as heterochromatin only in certain cell types or at certain stages of development, and *constitutive* heterochromatin that consists of chromosome structural components such as telomeres and centromeres. During interphase the chromatin organizes itself into discrete individual patches, called *chromosome territories*. Active genes, which are generally found in the euchromatic region of the chromosome, tend to be located towards the chromosome's territory boundary.

Antibodies to certain types of chromatin organization, particularly nucleosomes, have been associated with a number of autoimmune diseases, such as systemic lupus erythematosus. These are known as anti-nuclear antibodies (ANA) and have also been observed in concert with multiple sclerosis as part of general immune system dysfunction. As in the case of progeria, the role played by the antibodies in inducing the symptoms of autoimmune diseases is not obvious.

Nucleolus



An electron micrograph of a cell nucleus, showing the darkly stained nucleolus.

The nucleolus is a discrete densely stained structure found in the nucleus. It is not surrounded by a membrane, and is sometimes called a *suborganelle*. It forms around tandem repeats of rDNA, DNA coding for ribosomal RNA (rRNA). These regions are called nucleolar organizer regions (NOR). The main roles of the nucleolus are to synthesize rRNA and assemble ribosomes. The structural cohesion of the nucleolus depends on its activity, as ribosomal assembly in the nucleolus results in the transient association of nucleolar components, facilitating further ribosomal assembly, and hence further association. This model is supported by observations that inactivation of rDNA results in intermingling of nucleolar structures.

The first step in ribosomal assembly is transcription of the rDNA, by a protein called RNA polymerase I, forming a large pre-rRNA precursor. This is cleaved into the subunits 5.8S, 18S, and 28S rRNA. The transcription, post-transcriptional processing, and assembly of rRNA occurs in the nucleolus, aided by small nucleolar RNA (snoRNA) molecules, some of which are derived from spliced introns from messenger RNAs encoding genes related to ribosomal function. The assembled ribosomal subunits are the largest structures passed through the nuclear pores.

When observed under the electron microscope, the nucleolus can be seen to consist of three distinguishable regions: the innermost *fibrillar centers* (FCs), surrounded by the *dense fibrillar component* (DFC), which in turn is bordered by the *granular component* (GC). Transcription of the rDNA occurs either in the FC or at the FC-DFC boundary, and therefore when rDNA transcription in the cell is increased more FCs are detected. Most

of the cleavage and modification of rRNAs occurs in the DFC, while the latter steps involving protein assembly onto the ribosomal subunits occur in the GC.

Other subnuclear bodies

Subnuclear structure sizes	
Structure name	**Structure diameter**
Cajal bodies	0.2–2.0 μm
PIKA	5 μm
PML bodies	0.2–1.0 μm
Paraspeckles	0.2–1.0 μm
Speckles	20–25 nm

Besides the nucleolus, the nucleus contains a number of other non-membrane delineated bodies. These include Cajal bodies, Gemini of coiled bodies, polymorphic interphase karyosomal association (PIKA), promyelocytic leukaemia (PML) bodies, paraspeckles and splicing speckles. Although little is known about a number of these domains, they are significant in that they show that the nucleoplasm is not uniform mixture, but rather contains organized functional subdomains.

Other subnuclear structures appear as part of abnormal disease processes. For example, the presence of small intranuclear rods have been reported in some cases of nemaline myopathy. This condition typically results from mutations in actin, and the rods themselves consist of mutant actin as well as other cytoskeletal proteins.

Cajal bodies and gems

A nucleus typically contains between 1 and 10 compact structures called Cajal bodies or coiled bodies (CB), whose diameter measures between 0.2 μm and 2.0 μm depending on the cell type and species. When seen under an electron microscope, they resemble balls of tangled thread and are dense foci of distribution for the protein coilin. CBs are involved in a number of different roles relating to RNA processing, specifically small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA) maturation, and histone mRNA modification.

Similar to Cajal bodies are Gemini of coiled bodies, or gems, whose name is derived from the Gemini constellation in reference to their close "twin" relationship with CBs. Gems are similar in size and shape to CBs, and in fact are virtually indistinguishable under the microscope. Unlike CBs, gems do not contain small nuclear ribonucleoproteins (snRNPs), but do contain a protein called *survivor of motor neurons* (SMN) whose function relates to snRNP biogenesis. Gems are believed to assist CBs in snRNP biogenesis, though it has also been suggested from microscopy evidence that CBs and gems are different manifestations of the same structure.

PIKA and PTF domains

PIKA domains, or polymorphic interphase karyosomal associations, were first described in microscopy studies in 1991. Their function was and remains unclear, though they were not thought to be associated with active DNA replication, transcription, or RNA processing. They have been found to often associate with discrete domains defined by dense localization of the transcription factor PTF, which promotes transcription of snRNA.

PML bodies

Promyelocytic leukaemia bodies (PML bodies) are spherical bodies found scattered throughout the nucleoplasm, measuring around 0.2–1.0 μm. They are known by a number of other names, including nuclear domain 10 (ND10), Kremer bodies, and PML oncogenic domains. They are often seen in the nucleus in association with Cajal bodies and cleavage bodies. It has been suggested that they play a role in regulating transcription.

Paraspeckles

Discovered by Fox et al. in 2002, paraspeckles are irregularly shaped compartments in the nucleus' interchromatin space. First documented in HeLa cells, where there are generally 10–30 per nucleus, paraspeckles are now known to also exist in all human primary cells, transformed cell lines and tissue sections. Their name is derived from their distribution in the nucleus; the "para" is short for parallel and the "speckles" refers to the splicing speckles to which they are always in close proximity.

Paraspeckles are dynamic structures that are altered in response to changes in cellular metabolic activity. They are transcription dependent and in the absence of RNA Pol II transcription, the paraspeckle disappears and all of its associated protein components (PSP1, p54nrb, PSP2, CFI(m)68 and PSF) form a crescent shaped perinucleolar cap in the nucleolus. This phenomenon is demonstrated during the cell cycle. In the cell cycle, paraspeckles are present during interphase and during all of mitosis except for telophase. During telophase, when the two daughter nuclei are formed, there is no RNA Pol II transcription so the protein components instead form a perinucleolar cap.

Splicing speckles

Sometimes referred to as *interchromatin granule clusters* or as *splicing-factor compartments*, speckles are rich in splicing snRNPs and other splicing proteins necessary for pre-mRNA processing. Because of a cell's changing requirements, the composition and location of these bodies changes according to mRNA transcription and regulation via phosphorylation of specific proteins.

Function

The main function of the cell nucleus is to control gene expression and mediate the replication of DNA during the cell cycle. The nucleus provides a site for genetic transcription that is segregated from the location of translation in the cytoplasm, allowing levels of gene regulation that are not available to prokaryotes.

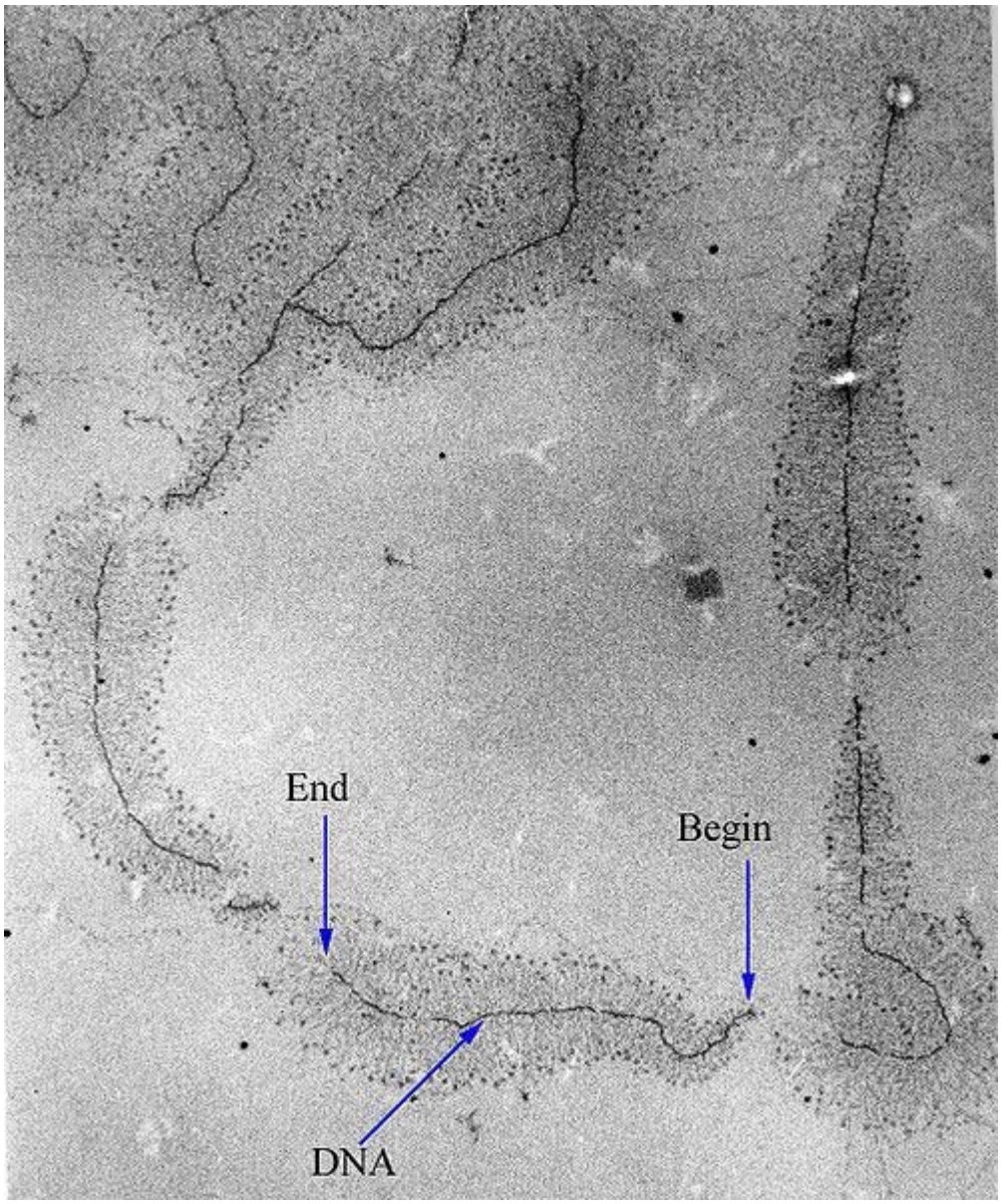
Cell compartmentalization

The nuclear envelope allows the nucleus to control its contents, and separate them from the rest of the cytoplasm where necessary. This is important for controlling processes on either side of the nuclear membrane. In some cases where a cytoplasmic process needs to be restricted, a key participant is removed to the nucleus, where it interacts with transcription factors to downregulate the production of certain enzymes in the pathway. This regulatory mechanism occurs in the case of glycolysis, a cellular pathway for breaking down glucose to produce energy. Hexokinase is an enzyme responsible for the first the step of glycolysis, forming glucose-6-phosphate from glucose. At high concentrations of fructose-6-phosphate, a molecule made later from glucose-6-phosphate, a regulator protein removes hexokinase to the nucleus, where it forms a transcriptional repressor complex with nuclear proteins to reduce the expression of genes involved in glycolysis.

In order to control which genes are being transcribed, the cell separates some transcription factor proteins responsible for regulating gene expression from physical access to the DNA until they are activated by other signaling pathways. This prevents even low levels of inappropriate gene expression. For example in the case of NF-κB-controlled genes, which are involved in most inflammatory responses, transcription is induced in response to a signal pathway such as that initiated by the signaling molecule TNF-α, binds to a cell membrane receptor, resulting in the recruitment of signalling proteins, and eventually activating the transcription factor NF-κB. A nuclear localisation signal on the NF-κB protein allows it to be transported through the nuclear pore and into the nucleus, where it stimulates the transcription of the target genes.

The compartmentalization allows the cell to prevent translation of unspliced mRNA. Eukaryotic mRNA contains introns that must be removed before being translated to produce functional proteins. The splicing is done inside the nucleus before the mRNA can be accessed by ribosomes for translation. Without the nucleus ribosomes would translate newly transcribed (unprocessed) mRNA resulting in misformed and nonfunctional proteins.

Gene expression



A micrograph of ongoing gene transcription of ribosomal RNA illustrating the growing primary transcripts. "Begin" indicates the 3' end of the DNA, where new RNA synthesis begins; "end" indicates the 5' end, where the primary transcripts are almost complete.

Gene expression first involves transcription, in which DNA is used as a template to produce RNA. In the case of genes encoding proteins, that RNA produced from this process is messenger RNA (mRNA), which then needs to be translated by ribosomes to form a protein. As ribosomes are located outside the nucleus, mRNA produced needs to be exported.

Since the nucleus is the site of transcription, it also contains a variety of proteins which either directly mediate transcription or are involved in regulating the process. These

proteins include helicases that unwind the double-stranded DNA molecule to facilitate access to it, RNA polymerases that synthesize the growing RNA molecule, topoisomerases that change the amount of supercoiling in DNA, helping it wind and unwind, as well as a large variety of transcription factors that regulate expression.

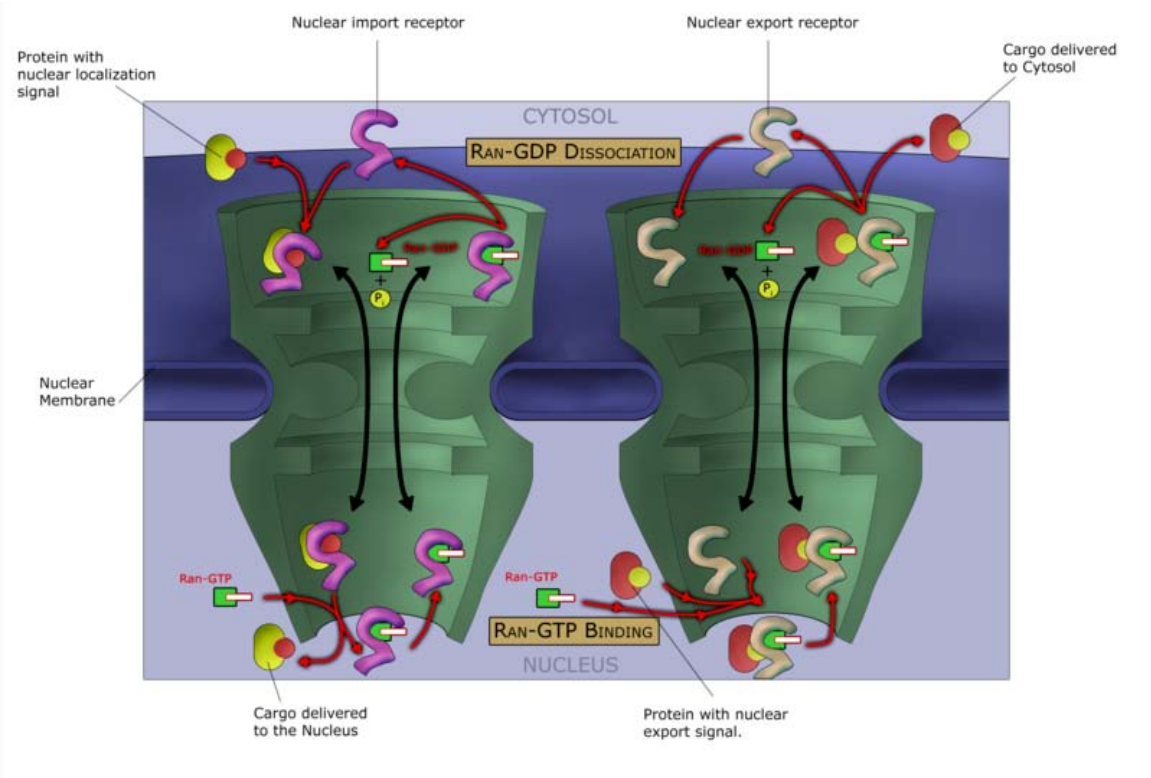
Processing of pre-mRNA

Newly synthesized mRNA molecules are known as primary transcripts or pre-mRNA. They must undergo post-transcriptional modification in the nucleus before being exported to the cytoplasm; mRNA that appears in the cytoplasm without these modifications is degraded rather than used for protein translation. The three main modifications are 5' capping, 3' polyadenylation, and RNA splicing. While in the nucleus, pre-mRNA is associated with a variety of proteins in complexes known as heterogeneous ribonucleoprotein particles (hnRNPs). Addition of the 5' cap occurs co-transcriptionally and is the first step in post-transcriptional modification. The 3' poly-adenine tail is only added after transcription is complete.

RNA splicing, carried out by a complex called the spliceosome, is the process by which introns, or regions of DNA that do not code for protein, are removed from the pre-mRNA and the remaining exons connected to re-form a single continuous molecule. This process normally occurs after 5' capping and 3' polyadenylation but can begin before synthesis is complete in transcripts with many exons. Many pre-mRNAs, including those encoding antibodies, can be spliced in multiple ways to produce different mature mRNAs that encode different protein sequences. This process is known as alternative splicing, and allows production of a large variety of proteins from a limited amount of DNA.

Dynamics and regulation

Nuclear transport



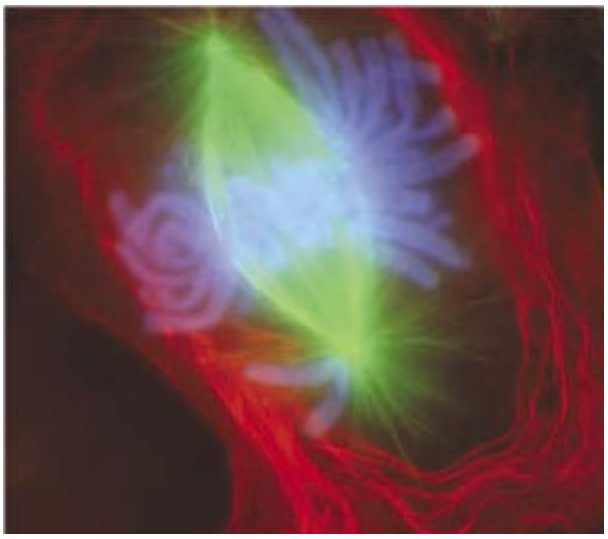
Macromolecules, such as RNA and proteins, are actively transported across the nuclear membrane in a process called the Ran-GTP nuclear transport cycle.

The entry and exit of large molecules from the nucleus is tightly controlled by the nuclear pore complexes. Although small molecules can enter the nucleus without regulation, macromolecules such as RNA and proteins require association karyopherins called importins to enter the nucleus and exportins to exit. "Cargo" proteins that must be translocated from the cytoplasm to the nucleus contain short amino acid sequences known as nuclear localization signals which are bound by importins, while those transported from the nucleus to the cytoplasm carry nuclear export signals bound by exportins. The ability of importins and exportins to transport their cargo is regulated by GTPases, enzymes that hydrolyze the molecule guanosine triphosphate to release energy. The key GTPase in nuclear transport is Ran, which can bind either GTP or GDP (guanosine diphosphate) depending on whether it is located in the nucleus or the cytoplasm. Whereas importins depend on RanGTP to dissociate from their cargo, exportins require RanGTP in order to bind to their cargo.

Nuclear import depends on the importin binding its cargo in the cytoplasm and carrying it through the nuclear pore into the nucleus. Inside the nucleus, RanGTP acts to separate the cargo from the importin, allowing the importin to exit the nucleus and be reused. Nuclear export is similar, as the exportin binds the cargo inside the nucleus in a process facilitated by RanGTP, exits through the nuclear pore, and separates from its cargo in the cytoplasm.

Specialized export proteins exist for translocation of mature mRNA and tRNA to the cytoplasm after post-transcriptional modification is complete. This quality-control mechanism is important due to the these molecules' central role in protein translation; mis-expression of a protein due to incomplete excision of exons or mis-incorporation of amino acids could have negative consequences for the cell; thus incompletely modified RNA that reaches the cytoplasm is degraded rather than used in translation.

Assembly and disassembly



An image of a newt lung cell stained with fluorescent dyes during metaphase. The mitotic spindle can be seen, stained green, attached to the two sets of chromosomes, stained light blue. All chromosomes but one are already at the metaphase plate.

During its lifetime a nucleus may be broken down, either in the process of cell division or as a consequence of apoptosis, a regulated form of cell death. During these events, the structural components of the nucleus—the envelope and lamina—are systematically degraded.

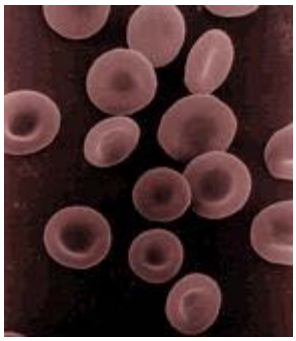
During the cell cycle the cell divides to form two cells. In order for this process to be possible, each of the new daughter cells must have a full set of genes, a process requiring replication of the chromosomes as well as segregation of the separate sets. This occurs by the replicated chromosomes, the sister chromatids, attaching to microtubules, which in turn are attached to different centrosomes. The sister chromatids can then be pulled to separate locations in the cell. In many cells the centrosome is located in the cytoplasm, outside the nucleus, the microtubules would be unable to attach to the chromatids in the presence of the nuclear envelope. Therefore the early stages in the cell cycle, beginning in prophase and until around prometaphase, the nuclear membrane is dismantled. Likewise, during the same period, the nuclear lamina is also disassembled, a process regulated by phosphorylation of the lamins. Towards the end of the cell cycle, the nuclear membrane is reformed, and around the same time, the nuclear lamina are reassembled by dephosphorylating the lamins.

However, in dinoflagellates the nuclear envelope remains intact, the centrosomes are located in the cytoplasm, and the microtubules come in contact with chromosomes, whose centromeric regions are incorporated into the nuclear envelope (the so-called closed mitosis with extranuclear spindle). In many other protists (e.g. ciliates, sporozoans) and fungi the centrosomes are intranuclear, and their nuclear envelope also does not disassemle during cell division.

Apoptosis is a controlled process in which the cell's structural components are destroyed, resulting in death of the cell. Changes associated with apoptosis directly affect the nucleus and its contents, for example in the condensation of chromatin and the disintegration of the nuclear envelope and lamina. The destruction of the lamin networks is controlled by specialized apoptotic proteases called caspases, which cleave the lamin proteins and thus degrade the nucleus' structural integrity. Lamin cleavage is sometimes used as a laboratory indicator of caspase activity in assays for early apoptotic activity. Cells that express mutant caspase-resistant lamins are deficient in nuclear changes related to apoptosis, suggesting that lamins play a role in initiating the events that lead to apoptotic degradation of the nucleus. Inhibition of lamin assembly itself is an inducer of apoptosis.

The nuclear envelope acts as a barrier that prevents both DNA and RNA viruses from entering the nucleus. Some viruses require access to proteins inside the nucleus in order to replicate and/or assemble. DNA viruses, such as herpesvirus replicate and assemble in the cell nucleus, and exit by budding through the inner nuclear membrane. This process is accompanied by disassembly of the lamina on the nuclear face of the inner membrane.

Anucleated and polynucleated cells



Human red blood cells, like those of other mammals, lack nuclei. This occurs as a normal part of the cells' development.

Although most cells have a single nucleus, some eukaryotic cell types have no nucleus, and others have many nuclei. This can be a normal process, as in the maturation of mammalian red blood cells, or a result of faulty cell division.

Anucleated cells contain no nucleus and are therefore incapable of dividing to produce daughter cells. The best-known anucleated cell is the mammalian red blood cell, or

erythrocyte, which also lacks other organelles such as mitochondria and serves primarily as a transport vessel to ferry oxygen from the lungs to the body's tissues. Erythrocytes mature through erythropoiesis in the bone marrow, where they lose their nuclei, organelles, and ribosomes. The nucleus is expelled during the process of differentiation from an erythroblast to a reticulocyte, which is the immediate precursor of the mature erythrocyte. The presence of mutagens may induce the release of some immature "micronucleated" erythrocytes into the bloodstream. Anucleated cells can also arise from flawed cell division in which one daughter lacks a nucleus and the other has two nuclei.

Polynucleated cells contain multiple nuclei. Most Acantharean species of protozoa and some fungi in mycorrhizae have naturally polynucleated cells. Other examples include the intestinal parasites in the genus *Giardia*, which have two nuclei per cell. In humans, skeletal muscle cells, called myocytes, become polynucleated during development; the resulting arrangement of nuclei near the periphery of the cells allows maximal intracellular space for myofibrils. Multinucleated cells can also be abnormal in humans; for example, cells arising from the fusion of monocytes and macrophages, known as giant multinucleated cells, sometimes accompany inflammation and are also implicated in tumor formation.

Evolution

As the major defining characteristic of the eukaryotic cell, the nucleus' evolutionary origin has been the subject of much speculation. Four major theories have been proposed to explain the existence of the nucleus, although none have yet earned widespread support.

The theory known as the "syntrophic model" proposes that a symbiotic relationship between the archaea and bacteria created the nucleus-containing eukaryotic cell. (Organisms of the Archaea domain have no cell nucleus.) It is hypothesized that the symbiosis originated when ancient archaea, similar to modern methanogenic archaea, invaded and lived within bacteria similar to modern myxobacteria, eventually forming the early nucleus. This theory is analogous to the accepted theory for the origin of eukaryotic mitochondria and chloroplasts, which are thought to have developed from a similar endosymbiotic relationship between proto-eukaryotes and aerobic bacteria. The archaeal origin of the nucleus is supported by observations that archaea and eukarya have similar genes for certain proteins, including histones. Observations that myxobacteria are motile, can form multicellular complexes, and possess kinases and G proteins similar to eukarya, support a bacterial origin for the eukaryotic cell.

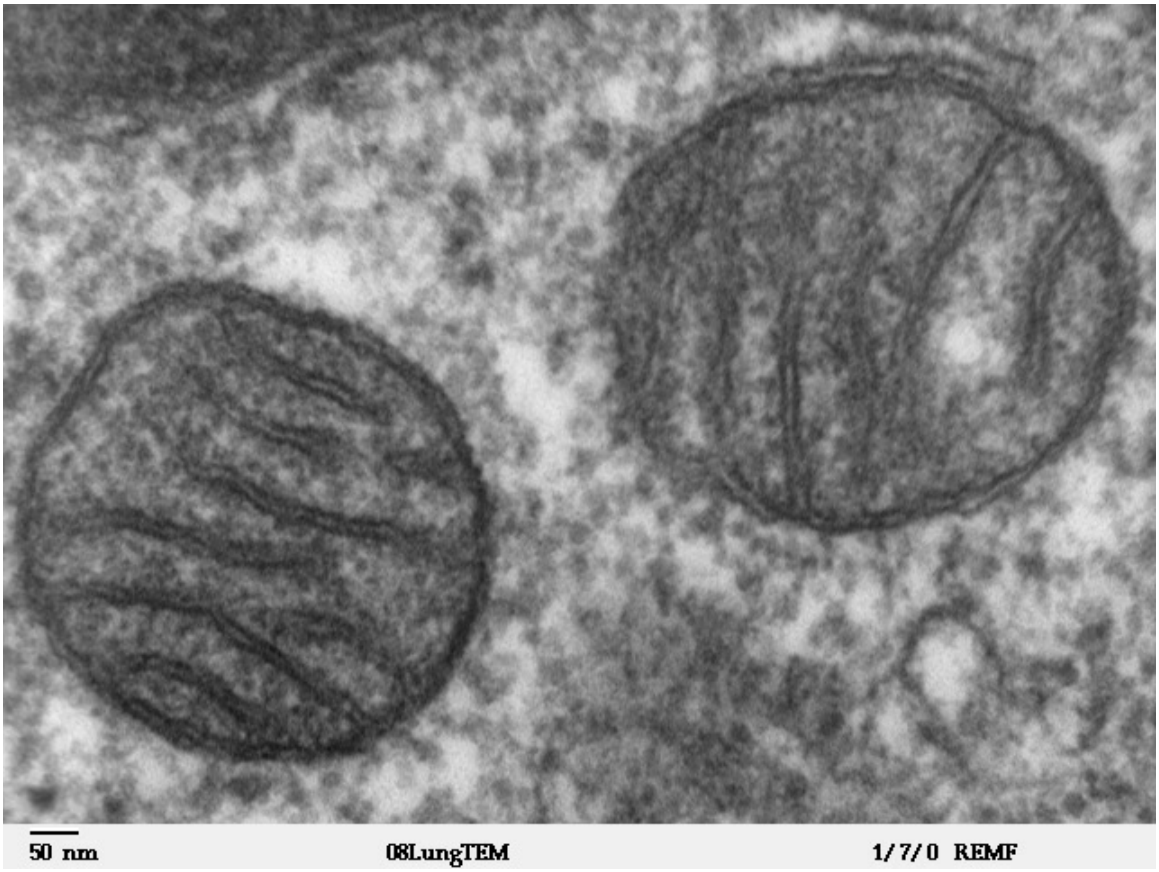
A second model proposes that proto-eukaryotic cells evolved from bacteria without an endosymbiotic stage. This model is based on the existence of modern planctomycetes bacteria that possess a nuclear structure with primitive pores and other compartmentalized membrane structures. A similar proposal states that a eukaryote-like cell, the chronocyte, evolved first and phagocytosed archaea and bacteria to generate the nucleus and the eukaryotic cell.

The most controversial model, known as *viral eukaryogenesis*, posits that the membrane-bound nucleus, along with other eukaryotic features, originated from the infection of a prokaryote by a virus. The suggestion is based on similarities between eukaryotes and viruses such as linear DNA strands, mRNA capping, and tight binding to proteins (analogizing histones to viral envelopes). One version of the proposal suggests that the nucleus evolved in concert with phagocytosis to form an early cellular "predator". Another variant proposes that eukaryotes originated from early archaea infected by poxviruses, on the basis of observed similarity between the DNA polymerases in modern poxviruses and eukaryotes. It has been suggested that the unresolved question of the evolution of sex could be related to the viral eukaryogenesis hypothesis.

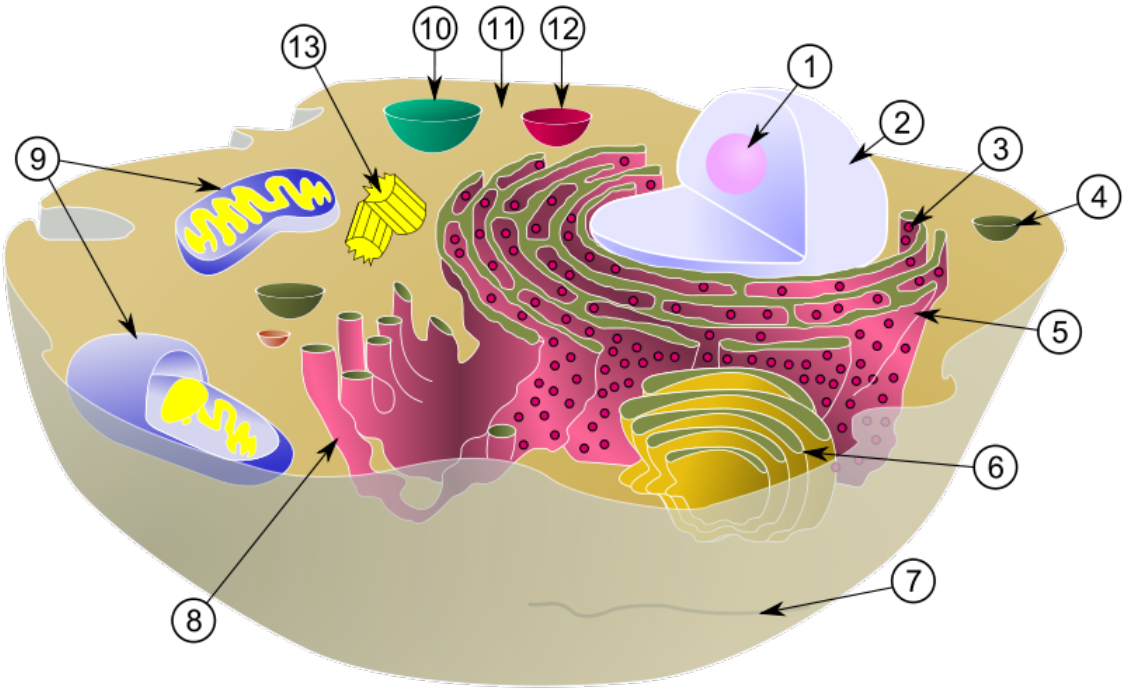
Finally, a very recent proposal suggests that traditional variants of the endosymbiont theory are insufficiently powerful to explain the origin of the eukaryotic nucleus. This model, termed the *exomembrane hypothesis*, suggests that the nucleus instead originated from a single ancestral cell that evolved a second exterior cell membrane; the interior membrane enclosing the original cell then became the nuclear membrane and evolved increasingly elaborate pore structures for passage of internally synthesized cellular components such as ribosomal subunits.

Chapter- 3

Mitochondrion



Two mitochondria from mammalian lung tissue displaying their matrix and membranes as shown by electron microscopy.



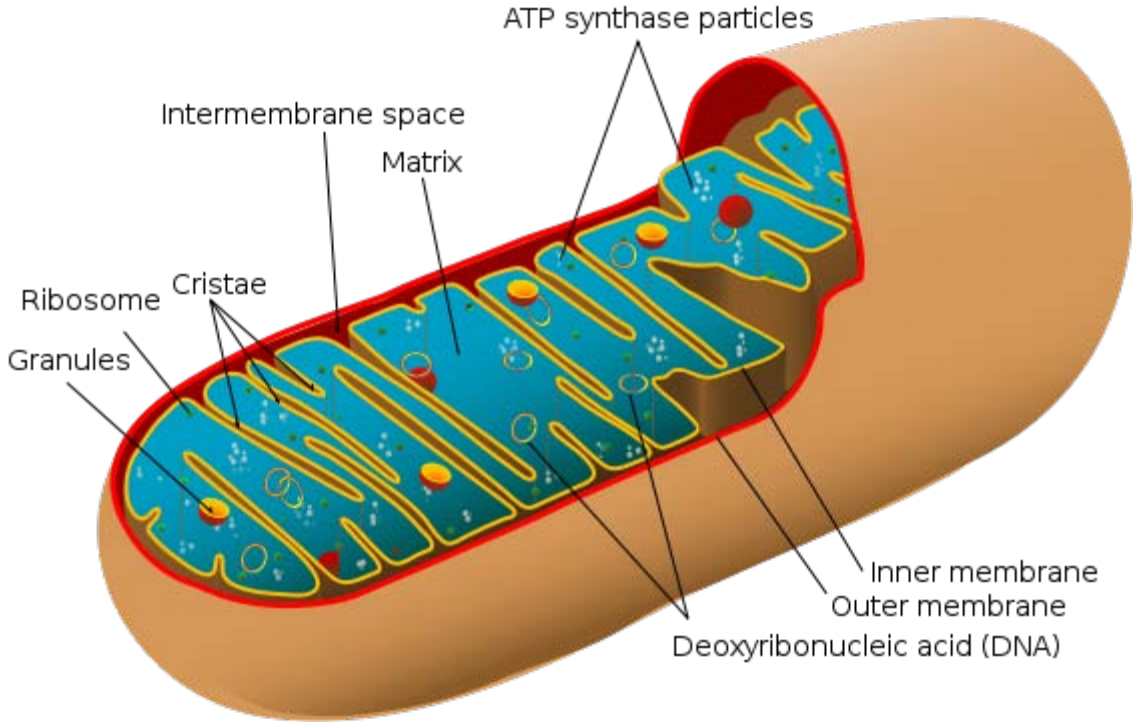
Schematic of typical animal cell, showing subcellular components. Organelles:
(1) nucleolus
(2) nuclear membrane
(3) Ribosomes
(4) Vesicle
(5) Rough endoplasmic reticulum (ER)
(6) Golgi body
(7) Cytoskeleton
(8) Smooth ER
(9) Mitochondria
(13) Centrioles within centrosome

In cell biology, a **mitochondrion** (plural **mitochondria**) is a membrane-enclosed organelle found in most eukaryotic cells. These organelles range from 0.5 to 10 micrometers (μm) in diameter. Mitochondria are sometimes described as "cellular power plants" because they generate most of the cell's supply of adenosine triphosphate (ATP), used as a source of chemical energy. In addition to supplying cellular energy, mitochondria are involved in a range of other processes, such as signaling, cellular differentiation, cell death, as well as the control of the cell cycle and cell growth. Mitochondria have been implicated in several human diseases, including mitochondrial disorders and cardiac dysfunction, and may play a role in the aging process. The word mitochondrion comes from the Greek *μίτος* or *mitos*, thread + *χονδρίον* or *chondrion*, granule.

Several characteristics make mitochondria unique. The number of mitochondria in a cell varies widely by organism and tissue type. Many cells have only a single mitochondrion, whereas others can contain several thousand mitochondria. The organelle is composed of

compartments that carry out specialized functions. These compartments or regions include the outer membrane, the intermembrane space, the inner membrane, and the cristae and matrix. Mitochondrial proteins vary depending on the tissue and the species. In humans, 615 distinct types of proteins have been identified from cardiac mitochondria, whereas in Murinae (rats), 940 proteins encoded by distinct genes have been reported. The mitochondrial proteome is thought to be dynamically regulated. Although most of a cell's DNA is contained in the cell nucleus, the mitochondrion has its own independent genome. Further, its DNA shows substantial similarity to bacterial genomes.

Structure



A mitochondrion contains outer and inner membranes composed of phospholipid bilayers and proteins. The two membranes, however, have different properties. Because of this double-membraned organization, there are five distinct compartments within the mitochondrion. There is the outer mitochondrial membrane, the intermembrane space (the space between the outer and inner membranes), the inner mitochondrial membrane, the cristae space (formed by infoldings of the inner membrane), and the matrix (space within the inner membrane).

Outer membrane

The outer mitochondrial membrane, which encloses the entire organelle, has a protein-to-phospholipid ratio similar to that of the eukaryotic plasma membrane (about 1:1 by weight). It contains large numbers of integral proteins called *porins*. These porins form channels that allow molecules 5000 Daltons or less in molecular weight to freely diffuse

from one side of the membrane to the other. Larger proteins can enter the mitochondrion if a signaling sequence at their N-terminus binds to a large multisubunit protein called translocase of the outer membrane, which then actively moves them across the membrane. Disruption of the outer membrane permits proteins in the intermembrane space to leak into the cytosol, leading to certain cell death. The mitochondrial outer membrane can associate with the endoplasmic reticulum (ER) membrane, in a structure called MAM (mitochondria-associated ER-membrane). This is important in ER-mitochondria calcium signaling and involved in the transfer of lipids between the ER and mitochondria.

Intermembrane space

The intermembrane space is the space between the outer membrane and the inner membrane. Because the outer membrane is freely permeable to small molecules, the concentrations of small molecules such as ions and sugars in the intermembrane space is the same as the cytosol. However, large proteins must have a specific signaling sequence to be transported across the outer membrane, so the protein composition of this space is different from the protein composition of the cytosol. One protein that is localized to the intermembrane space in this way is cytochrome c.

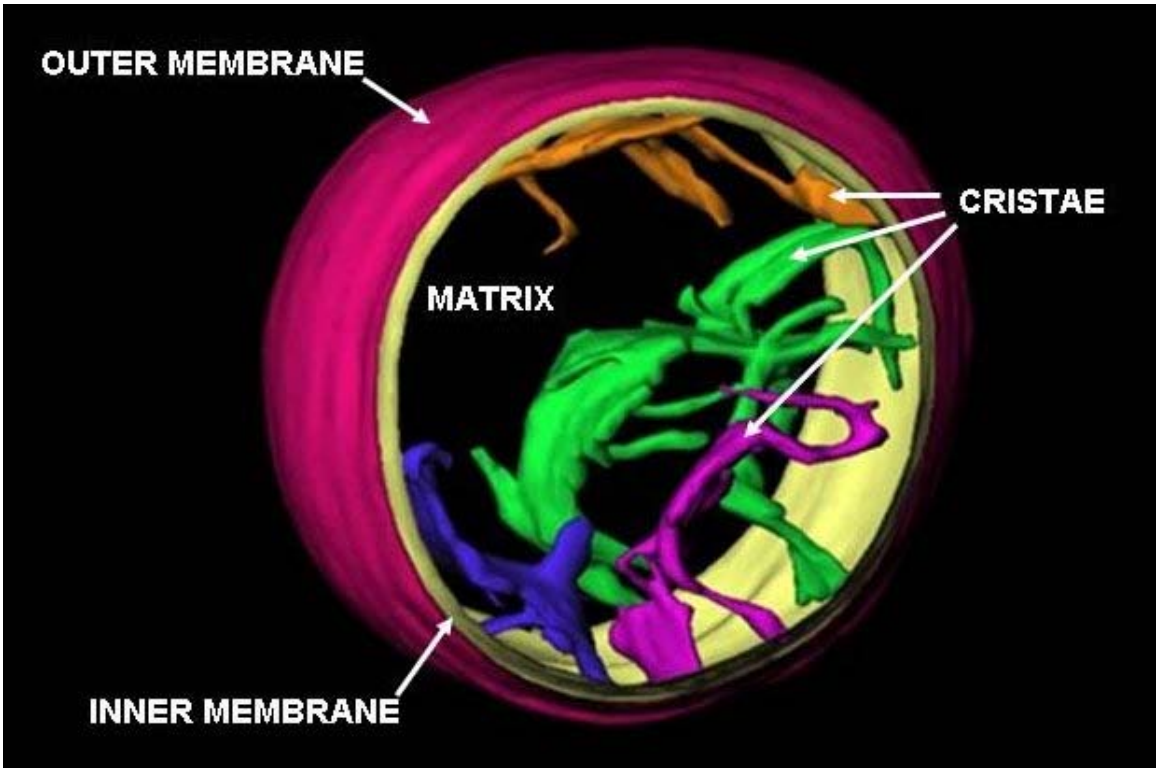
Inner membrane

The inner mitochondrial membrane contains proteins with five types of functions:

1. Those that perform the redox reactions of oxidative phosphorylation
2. ATP synthase, which generates ATP in the matrix
3. Specific transport proteins that regulate metabolite passage into and out of the matrix
4. Protein import machinery.
5. Mitochondria fusion and fission protein

It contains more than 151 different polypeptides, and has a very high protein-to-phospholipid ratio (more than 3:1 by weight, which is about 1 protein for 15 phospholipids). The inner membrane is home to around 1/5 of the total protein in a mitochondrion. In addition, the inner membrane is rich in an unusual phospholipid, cardiolipin. This phospholipid was originally discovered in cow hearts in 1942, and is usually characteristic of mitochondrial and bacterial plasma membranes. Cardiolipin contains four fatty acids rather than two and may help to make the inner membrane impermeable. Unlike the outer membrane, the inner membrane doesn't contain porins and is highly impermeable to all molecules. Almost all ions and molecules require special membrane transporters to enter or exit the matrix. Proteins are ferried into the matrix via the translocase of the inner membrane (TIM) complex or via Oxa1. In addition, there is a membrane potential across the inner membrane formed by the action of the enzymes of the electron transport chain.

Cristae



Cross-sectional image of cristae in rat liver mitochondrion to demonstrate the likely 3D structure and relationship to the inner membrane.

The inner mitochondrial membrane is compartmentalized into numerous cristae, which expand the surface area of the inner mitochondrial membrane, enhancing its ability to produce ATP. For typical liver mitochondria the area of the inner membrane is about five times greater than the outer membrane. This ratio is variable and mitochondria from cells that have a greater demand for ATP, such as muscle cells, contain even more cristae. These folds are studded with small round bodies known as F_1 particles or oxysomes. These are not simple random folds but rather invaginations of the inner membrane, which can affect overall chemiosmotic function.

One recent mathematical modeling study has suggested that the optical properties of the christae in filamentous mitochondria may affect the generation and propogation of light within the tissue.

Matrix

The matrix is the space enclosed by the inner membrane. It contains about 2/3 of the total protein in a mitochondrion. The matrix is important in the production of ATP with the aid of the ATP synthase contained in the inner membrane. The matrix contains a highly-concentrated mixture of hundreds of enzymes, special mitochondrial ribosomes, tRNA, and several copies of the mitochondrial DNA genome. Of the enzymes, the major functions include oxidation of pyruvate and fatty acids, and the citric acid cycle.

Mitochondria have their own genetic material, and the machinery to manufacture their own RNAs and proteins (*see: protein biosynthesis*). A published human mitochondrial DNA sequence revealed 16,569 base pairs encoding 37 total genes: 22 tRNA, 2 rRNA, and 13 peptide genes. The 13 mitochondrial peptides in humans are integrated into the inner mitochondrial membrane, along with proteins encoded by genes that reside in the host cell's nucleus.

Organization and distribution

Mitochondria are found in nearly all eukaryotes. They vary in number and location according to cell type. A single mitochondrion is often found in unicellular organisms. Conversely, numerous mitochondria are found in human liver cells, with about 1000–2000 mitochondria per cell making up 1/5th of the cell volume. The mitochondria can be found nestled between myofibrils of muscle or wrapped around the sperm flagellum. Often they form a complex 3D branching network inside the cell with the cytoskeleton. The association with the cytoskeleton determines mitochondrial shape, which can affect the function as well. Recent evidence suggests vimentin, one of the components of the cytoskeleton, is critical to the association with the cytoskeleton.

Function

The most prominent roles of mitochondria are to produce ATP (i.e., phosphorylation of ADP) through respiration, and to regulate cellular metabolism. The central set of reactions involved in ATP production are collectively known as the citric acid cycle, or the Krebs Cycle. However, the mitochondrion has many other functions in addition to the production of ATP.

Energy conversion

A dominant role for the mitochondria is the production of ATP, as reflected by the large number of proteins in the inner membrane for this task. This is done by oxidizing the major products of glucose, pyruvate, and NADH, which are produced in the cytosol. This process of cellular respiration, also known as aerobic respiration, is dependent on the presence of oxygen. When oxygen is limited, the glycolytic products will be metabolized by anaerobic respiration, a process that is independent of the mitochondria. The production of ATP from glucose has an approximately 13-fold higher yield during aerobic respiration compared to anaerobic respiration. Recently it has been shown that plant mitochondria can produce a limited amount of ATP without oxygen by using the alternate substrate nitrite.

Pyruvate and the citric acid cycle

Each pyruvate molecule produced by glycolysis is actively transported across the inner mitochondrial membrane, and into the matrix where it is oxidized and combined with coenzyme A to form CO_2, acetyl-CoA, and NADH.

The acetyl-CoA is the primary substrate to enter the *citric acid cycle*, also known as the *tricarboxylic acid (TCA) cycle* or *Krebs cycle*. The enzymes of the citric acid cycle are located in the mitochondrial matrix, with the exception of succinate dehydrogenase, which is bound to the inner mitochondrial membrane as part of Complex II. The citric acid cycle oxidizes the acetyl-CoA to carbon dioxide, and, in the process, produces reduced cofactors (three molecules of NADH and one molecule of $FADH_2$) that are a source of electrons for the *electron transport chain*, and a molecule of GTP (that is readily converted to an ATP).

NADH and $FADH_2$: the electron transport chain

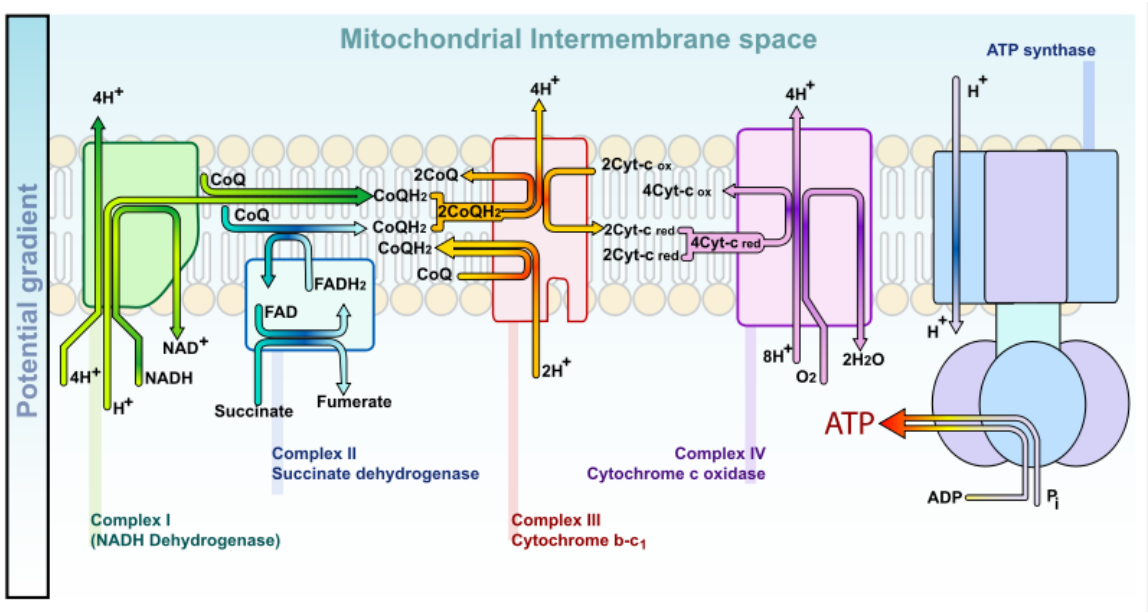


Diagram of the electron transport chain in the mitonchondrial intermembrane space

The redox energy from NADH and $FADH_2$ is transferred to oxygen (O_2) in several steps via the electron transport chain. These energy-rich molecules are produced within the matrix via the citric acid cycle but are also produced in the cytoplasm by glycolysis. Reducing equivalents from the cytoplasm can be imported via the malate-aspartate shuttle system of antiporter proteins or feed into the electron transport chain using a glycerol phosphate shuttle. Protein complexes in the inner membrane (NADH dehydrogenase, cytochrome c reductase, and cytochrome c oxidase) perform the transfer and the incremental release of energy is used to pump protons (H^+) into the intermembrane space. This process is efficient, but a small percentage of electrons may prematurely reduce oxygen, forming reactive oxygen species such as superoxide. This can cause oxidative stress in the mitochondria and may contribute to the decline in mitochondrial function associated with the aging process.

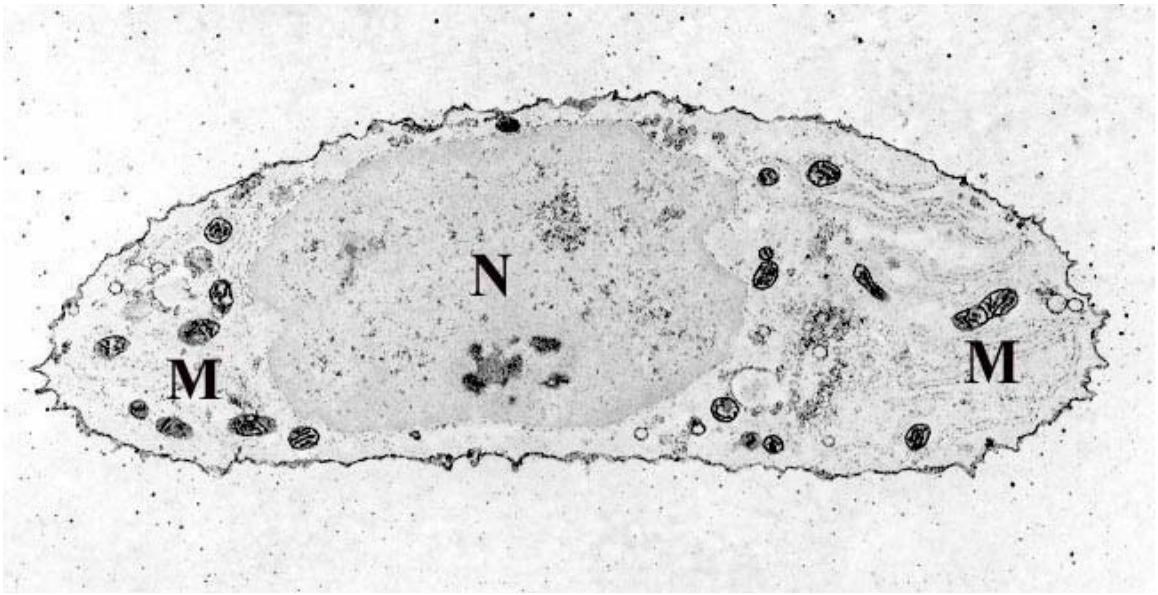
As the proton concentration increases in the intermembrane space, a strong electrochemical gradient is established across the inner membrane. The protons can return to the matrix through the ATP synthase complex, and their potential energy is used

to synthesize ATP from ADP and inorganic phosphate (P_i). This process is called chemiosmosis, and was first described by Peter Mitchell who was awarded the 1978 Nobel Prize in Chemistry for his work. Later, part of the 1997 Nobel Prize in Chemistry was awarded to Paul D. Boyer and John E. Walker for their clarification of the working mechanism of ATP synthase.

Heat production

Under certain conditions, protons can re-enter the mitochondrial matrix without contributing to ATP synthesis. This process is known as *proton leak* or *mitochondrial uncoupling* and is due to the facilitated diffusion of protons into the matrix. The process results in the unharnessed potential energy of the proton electrochemical gradient being released as heat. The process is mediated by a proton channel called thermogenin, or UCP1. Thermogenin is a 33kDa protein first discovered in 1973. Thermogenin is primarily found in brown adipose tissue, or brown fat, and is responsible for non-shivering thermogenesis. Brown adipose tissue is found in mammals, and is at its highest levels in early life and in hibernating animals. In humans, brown adipose tissue is present at birth and decreases with age.

Storage of calcium ions



Mitochondria (M) within a chondrocyte stained for calcium as shown by electron microscopy.

The concentrations of free calcium in the cell can regulate an array of reactions and is important for signal transduction in the cell. Mitochondria can transiently store calcium, a contributing process for the cell's homeostasis of calcium. In fact, their ability to rapidly take in calcium for later release makes them very good "cytosolic buffers" for calcium. The endoplasmic reticulum (ER) is the most significant storage site of calcium, and there is a significant interplay between the mitochondrion and ER with regard to calcium. The

calcium is taken up into the matrix by a calcium uniporter on the inner mitochondrial membrane. It is primarily driven by the mitochondrial membrane potential. Release of this calcium back into the cell's interior can occur via a sodium-calcium exchange protein or via "calcium-induced-calcium-release" pathways. This can initiate calcium spikes or calcium waves with large changes in the membrane potential. These can activate a series of second messenger system proteins that can coordinate processes such as neurotransmitter release in nerve cells and release of hormones in endocrine cells.

Additional functions

Mitochondria play a central role in many other metabolic tasks, such as:

- Regulation of the membrane potential
- Apoptosis-programmed cell death
- Calcium signaling (including calcium-evoked apoptosis)
- Cellular proliferation regulation
- Regulation of cellular metabolism
- Certain heme synthesis reactions
- Steroid synthesis.

Some mitochondrial functions are performed only in specific types of cells. For example, mitochondria in liver cells contain enzymes that allow them to detoxify ammonia, a waste product of protein metabolism. A mutation in the genes regulating any of these functions can result in mitochondrial diseases.

Origin

Mitochondria have many features in common with prokaryotes. As a result, they are believed to be originally derived from endosymbiotic prokaryotes.

A mitochondrion contains DNA, which is organized as several copies of a single, circular chromosome. This mitochondrial chromosome contains genes for redox proteins such as those of the respiratory chain. The CoRR hypothesis proposes that this **co**-location is required for **r**edox **r**egulation. The mitochondrial genome codes for some RNAs of ribosomes, and the twenty-two tRNAs necessary for the translation of messenger RNAs into protein. The circular structure is also found in prokaryotes, and the similarity is extended by the fact that mitochondrial DNA is organized with a variant genetic code similar to that of Proteobacteria. This suggests that their ancestor, the so-called proto-mitochondrion, was a member of the Proteobacteria. In particular, the proto-mitochondrion was probably closely related to the rickettsia. However, the exact relationship of the ancestor of mitochondria to the alpha-proteobacteria and whether the mitochondria was formed at the same time or after the nucleus, remains controversial.

The ribosomes coded for by the mitochondrial DNA are similar to those from bacteria in size and structure. They closely resemble the bacterial 70S ribosome and not the 80S cytoplasmic ribosomes, which are coded for by nuclear DNA.

The endosymbiotic relationship of mitochondria with their host cells was popularized by Lynn Margulis. The endosymbiotic hypothesis suggests that mitochondria descended from bacteria that somehow survived endocytosis by another cell, and became incorporated into the cytoplasm. The ability of these bacteria to conduct respiration in host cells that had relied on glycolysis and fermentation would have provided a considerable evolutionary advantage. In a similar manner, host cells with symbiotic bacteria capable of photosynthesis would have had an advantage. The incorporation of symbiotes would have increased the number of environments in which the cells could survive. This symbiotic relationship probably developed 1.7-2 billion years ago.

A few groups of unicellular eukaryotes lack mitochondria: the microsporidians, metamonads, and archamoebae. These groups appear as the most primitive eukaryotes on phylogenetic trees constructed using rRNA information, which once suggested that they appeared before the origin of mitochondria. However, this is now known to be an artifact of long-branch attraction—they are derived groups and retain genes or organelles derived from mitochondria (e.g., mitosomes and hydrogenosomes).

Genome

The human mitochondrial genome is a circular DNA molecule of about 16 kilobases. It encodes 37 genes: 13 for subunits of respiratory complexes I, III, IV and V, 22 for mitochondrial tRNA (for the 20 standard amino acids, plus an extra gene for leucine and serine), and 2 for rRNA. One mitochondrion can contain two to ten copies of its DNA.

As in prokaryotes, there is a very high proportion of coding DNA and an absence of repeats. Mitochondrial genes are transcribed as multigenic transcripts, which are cleaved and polyadenylated to yield mature mRNAs. Not all proteins necessary for mitochondrial function are encoded by the mitochondrial genome; most are coded by genes in the cell nucleus and the corresponding proteins are imported into the mitochondrion. The exact number of genes encoded by the nucleus and the mitochondrial genome differs between species. In general, mitochondrial genomes are circular, although exceptions have been reported. In general, mitochondrial DNA lacks introns, as is the case in the human mitochondrial genome; however, introns have been observed in some eukaryotic mitochondrial DNA, such as that of yeast and protists, including *Dictyostelium discoideum*.

In animals the mitochondrial genome is typically a single circular chromosome that is approximately 16-kb long and has 37 genes. The genes while highly conserved may vary in location. Curiously this pattern is not found in the human body louse (*Pediculus humanus*). Instead this mitochondrial genome is arranged in 18 minicircular chromosomes each of which is 3–4 kb long and has one to three genes. This pattern is also found in other sucking lice but not in chewing lice. Recombination has been shown to occur between the minichromosomes. The reason for this difference is not known.

While slight variations on the standard code had been predicted earlier, none was discovered until 1979, when researchers studying human mitochondrial genes determined

that they used an alternative code. Many slight variants have been discovered since, including various alternative mitochondrial codes. Further, the AUA, AUC, and AUU codons are all allowable start codons.

Exceptions to the universal genetic code (UGC) in mitochondria

Organism	Codon	Standard	Novel
Mammalian	AGA, AGG	Arginine	Stop codon
	AUA	Isoleucine	Methionine
	UGA	Stop codon	Tryptophan
Invertebrates	AGA, AGG	Arginine	Serine
	AUA	Isoleucine	Methionine
	UGA	Stop codon	Tryptophan
Yeast	AUA	Isoleucine	Methionine
	UGA	Stop codon	Tryptophan
	CUA	Leucine	Threonine

Some of these differences should be regarded as pseudo-changes in the genetic code due to the phenomenon of RNA editing, which is common in mitochondria. In higher plants, it was thought that CGG encoded for tryptophan and not arginine; however, the codon in the processed RNA was discovered to be the UGG codon, consistent with the universal genetic code for tryptophan. Of note, the arthropod mitochondrial genetic code has undergone parallel evolution within a phylum, with some organisms uniquely translating AGG to lysine.

Mitochondrial genomes have far fewer genes than the bacteria from which they are thought to be descended. Although some have been lost altogether, many have been transferred to the nucleus, such as the respiratory complex II protein subunits. This is thought to be relatively common over evolutionary time. A few organisms, such as the *Cryptosporidium*, actually have mitochondria that lack any DNA, presumably because all their genes have been lost or transferred. In *Cryptosporidium*, the mitochondria have an altered ATP generation system that renders the parasite resistant to many classical mitochondrial inhibitors such as cyanide, azide, and atovaquone.

Replication and inheritance

Mitochondria divide by binary fission similar to bacterial cell division; unlike bacteria, however, mitochondria can also fuse with other mitochondria. The regulation of this division differs between eukaryotes. In many single-celled eukaryotes, their growth and division is linked to the cell cycle. For example, a single mitochondrion may divide synchronously with the nucleus. This division and segregation process must be tightly controlled so that each daughter cell receives at least one mitochondrion. In other eukaryotes (in mammals for example), mitochondria may replicate their DNA and divide

mainly in response to the energy needs of the cell, rather than in phase with the cell cycle. When the energy needs of a cell are high, mitochondria grow and divide. When the energy use is low, mitochondria are destroyed or become inactive. In such examples, and in contrast to the situation in many single celled eukaryotes, mitochondria are apparently randomly distributed to the daughter cells during the division of the cytoplasm.

An individual's mitochondrial genes are not inherited by the same mechanism as nuclear genes. At fertilization of an egg cell by a sperm, the egg nucleus and sperm nucleus each contribute equally to the genetic makeup of the zygote nucleus. In contrast, the mitochondria, and therefore the mitochondrial DNA, usually comes from the egg only. The sperm's mitochondria enter the egg but do not contribute genetic information to the embryo. Instead, paternal mitochondria are marked with ubiquitin to select them for later destruction inside the embryo. The egg cell contains relatively few mitochondria, but it is these mitochondria that survive and divide to populate the cells of the adult organism. Mitochondria are, therefore, in most cases inherited down the female line, known as maternal inheritance. This mode is seen in most organisms including all animals. However, mitochondria in some species can sometimes be inherited paternally. This is the norm among certain coniferous plants, although not in pine trees and yew trees. It has been suggested that it occurs at a very low level in humans.

Uniparental inheritance leads to little opportunity for genetic recombination between different lineages of mitochondria, although a single mitochondrion can contain 2–10 copies of its DNA. For this reason, mitochondrial DNA usually is thought to reproduce by binary fission. What recombination does take place maintains genetic integrity rather than maintaining diversity. However, there are studies showing evidence of recombination in mitochondrial DNA. It is clear that the enzymes necessary for recombination are present in mammalian cells. Further, evidence suggests that animal mitochondria can undergo recombination. The data are a bit more controversial in humans, although indirect evidence of recombination exists. If recombination does not occur, the whole mitochondrial DNA sequence represents a single haplotype, which makes it useful for studying the evolutionary history of populations.

Population genetic studies

The near-absence of genetic recombination in mitochondrial DNA makes it a useful source of information for scientists involved in population genetics and evolutionary biology. Because all the mitochondrial DNA is inherited as a single unit, or haplotype, the relationships between mitochondrial DNA from different individuals can be represented as a gene tree. Patterns in these gene trees can be used to infer the evolutionary history of populations. The classic example of this is in human evolutionary genetics, where the molecular clock can be used to provide a recent date for mitochondrial Eve. This is often interpreted as strong support for a recent modern human expansion out of Africa. Another human example is the sequencing of mitochondrial DNA from Neanderthal bones. The relatively large evolutionary distance between the mitochondrial DNA sequences of Neanderthals and living humans has been interpreted as

evidence for lack of interbreeding between Neanderthals and anatomically-modern humans.

However, mitochondrial DNA reflects the history of only females in a population and so may not represent the history of the population as a whole. This can be partially overcome by the use of paternal genetic sequences, such as the non-recombining region of the Y-chromosome. In a broader sense, only studies that also include nuclear DNA can provide a comprehensive evolutionary history of a population.

Dysfunction and disease

Mitochondrial diseases

With their central place in cell metabolism, damage — and subsequent dysfunction — in mitochondria is an important factor in a wide range of human diseases. Mitochondrial disorders often present as neurological disorders, but can manifest as myopathy, diabetes, multiple endocrinopathy, or a variety of other systemic manifestations. Diseases caused by mutation in the mtDNA include Kearns-Sayre syndrome, MELAS syndrome and Leber's hereditary optic neuropathy. In the vast majority of cases, these diseases are transmitted by a female to her children, as the zygote derives its mitochondria and hence its mtDNA from the ovum. Diseases such as Kearns-Sayre syndrome, Pearson's syndrome, and progressive external ophthalmoplegia are thought to be due to large-scale mtDNA rearrangements, whereas other diseases such as MELAS syndrome, Leber's hereditary optic neuropathy, myoclonic epilepsy with ragged red fibers (MERRF), and others are due to point mutations in mtDNA.

In other diseases, defects in nuclear genes lead to dysfunction of mitochondrial proteins. This is the case in Friedreich's ataxia, hereditary spastic paraplegia, and Wilson's disease. These diseases are inherited in a dominance relationship, as applies to most other genetic diseases. A variety of disorders can be caused by nuclear mutations of oxidative phosphorylation enzymes, such as coenzyme Q10 deficiency and Barth syndrome. Environmental influences may interact with hereditary predispositions and cause mitochondrial disease. For example, there may be a link between pesticide exposure and the later onset of Parkinson's disease.

Other pathologies with etiology involving mitochondrial dysfunction include schizophrenia, bipolar disorder, dementia, Alzheimer's disease, Parkinson's disease, epilepsy, stroke, cardiovascular disease, retinitis pigmentosa, and diabetes mellitus. A common thread thought to link these seemingly-unrelated conditions is cellular damage causing oxidative stress. How exactly mitochondrial dysfunction fits into the etiology of these pathologies is yet to be elucidated.

Possible relationships to aging

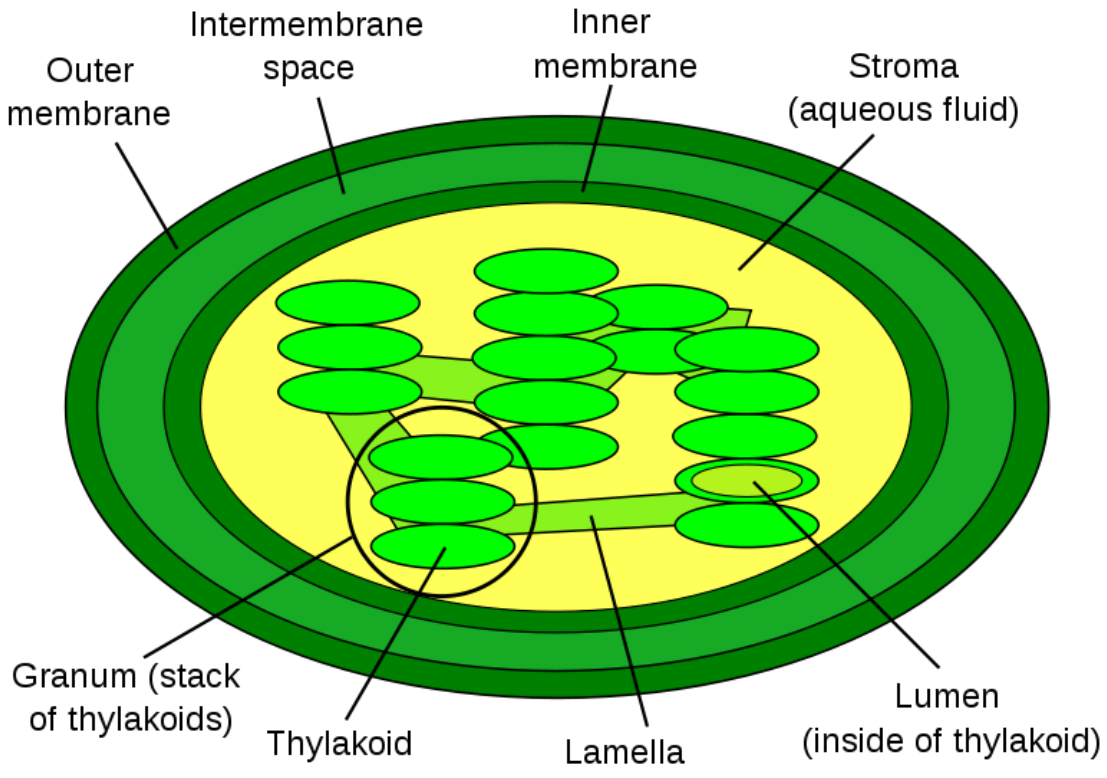
Given the role of mitochondria as the cell's powerhouse, there may be some leakage of the high-energy electrons in the respiratory chain to form reactive oxygen species. This

can result in significant oxidative stress in the mitochondria with high mutation rates of mitochondrial DNA. A vicious cycle is thought to occur, as oxidative stress leads to mitochondrial DNA mutations, which can lead to enzymatic abnormalities and further oxidative stress. A number of changes occur to mitochondria during the aging process. Tissues from elderly patients show a decrease in enzymatic activity of the proteins of the respiratory chain. Large deletions in the mitochondrial genome can lead to high levels of oxidative stress and neuronal death in Parkinson's disease. Hypothesized links between aging and oxidative stress are not new and were proposed over 50 years ago; however, there is much debate over whether mitochondrial changes are causes of aging or merely characteristics of aging. One notable study in mice demonstrated shortened lifespan but no increase in reactive oxygen species despite increasing mitochondrial DNA mutations, suggesting that mitochondrial DNA mutations can cause lifespan shortening by other mechanisms. As a result, the exact relationships between mitochondria, oxidative stress, and aging have not yet been settled.

Chapter- 4

Chloroplast and Lysosome

Chloroplast

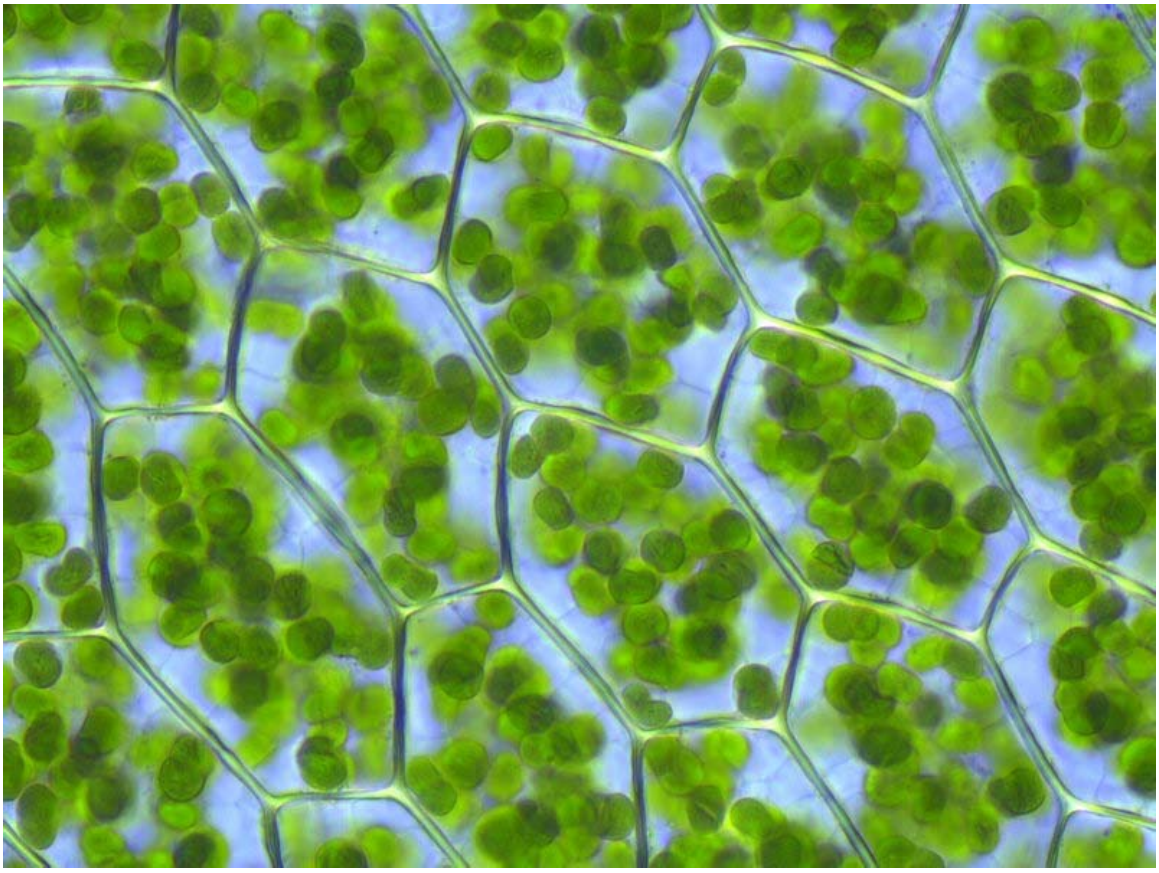


The simplified internal structure of a chloroplast

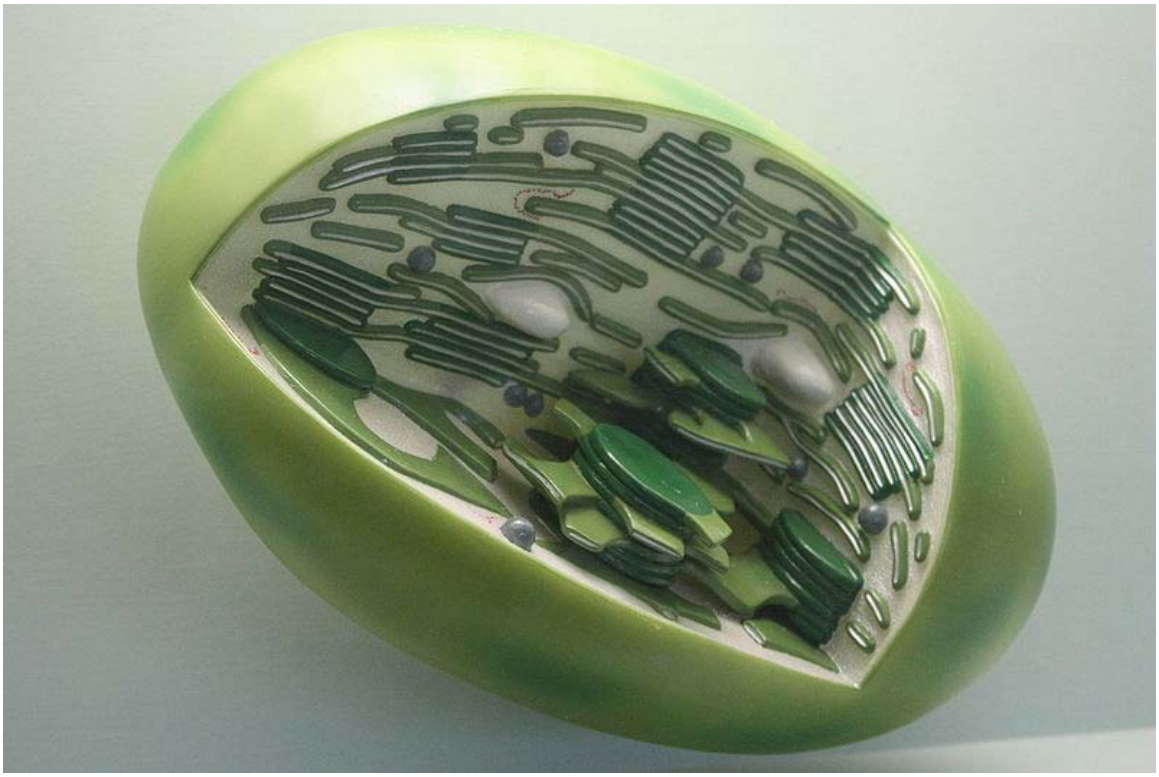
Chloroplasts are organelles found in plant cells and other eukaryotic organisms that conduct photosynthesis. Chloroplasts capture light energy to conserve free energy in the form of ATP and reduce NADP to NADPH through a complex set of processes called photosynthesis.

The word chloroplast (χλωροπλάστης) is derived from the Greek words *chloros* (χλωρός), which means green, and *plastis* (πλάστης), which means "the one who forms". Chloroplasts are members of a class of organelles known as plastids.

Evolutionary origin



Chloroplasts visible in the cells of *Plagiomnium affine* — Many-fruited Thyme-moss



A model chloroplast

Chloroplasts are one of the many different types of organelles in the plant cell. In general, they are considered to have originated from cyanobacteria through endosymbiosis. This was first suggested by Mereschkowsky in 1905 after an observation by Schimper in 1883 that chloroplasts closely resemble cyanobacteria. All chloroplasts are thought to derive directly or indirectly from a single endosymbiotic event (in the Archaeplastida), except for *Paulinella chromatophora*, which has recently acquired a photosynthetic cyanobacterial endosymbiont which is not closely related to chloroplasts of other eukaryotes. In that they derive from an endosymbiotic event, chloroplasts are similar to mitochondria, but chloroplasts are found only in plants and protista. The chloroplast is surrounded by a double-layered composite membrane with an intermembrane space; further, it has reticulations, or many infoldings, filling the inner spaces. The chloroplast has its own DNA, which codes for redox proteins involved in electron transport in photosynthesis; this is termed the plastome.

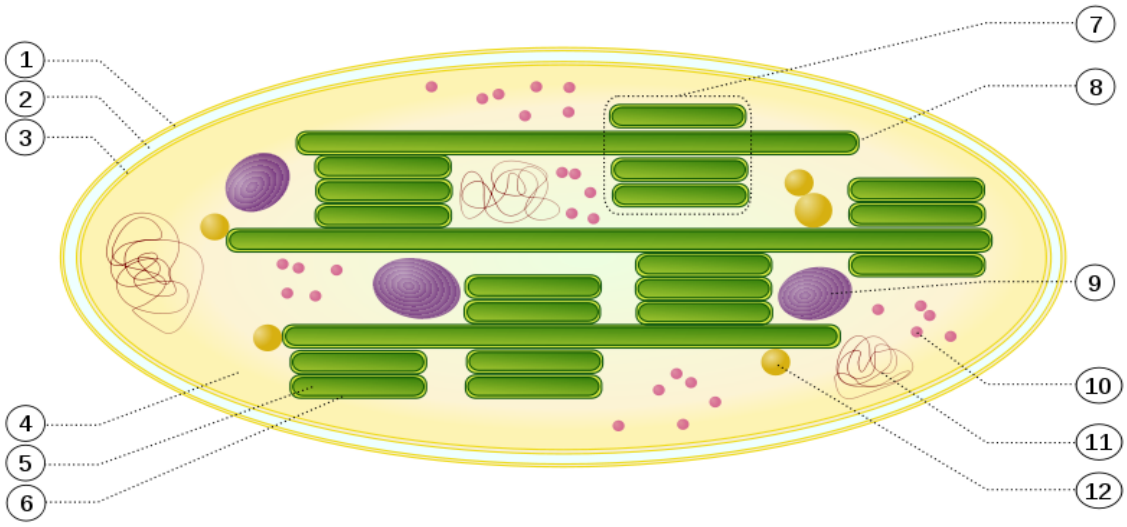
In green plants, chloroplasts are surrounded by two lipid-bilayer membranes. They are believed to correspond to the outer and inner membranes of the ancestral cyanobacterium. Chloroplasts have their own genome, which is considerably reduced compared to that of free-living cyanobacteria, but the parts that are still present show clear similarities with the cyanobacterial genome. Plastids may contain 60-100 genes whereas cyanobacteria often contain more than 1500 genes. Many of the missing genes are encoded in the nuclear genome of the host. The transfer of nuclear information has been estimated in tobacco plants at one gene for every 16000 pollen grains.

In some algae (such as the heterokonts and other protists such as Euglenozoa and Cercozoa), chloroplasts seem to have evolved through a secondary event of endosymbiosis, in which a eukaryotic cell engulfed a second eukaryotic cell containing chloroplasts, forming chloroplasts with three or four membrane layers. In some cases, such secondary endosymbionts may have themselves been engulfed by still other eukaryotes, thus forming tertiary endosymbionts. In the alga *Chlorella*, there is only one chloroplast, which is bell-shaped.

In some groups of mixotrophic protists such as the dinoflagellates, chloroplasts are separated from a captured alga or diatom and used temporarily. These klepto chloroplasts may only have a lifetime of a few days and are then replaced.

Structure

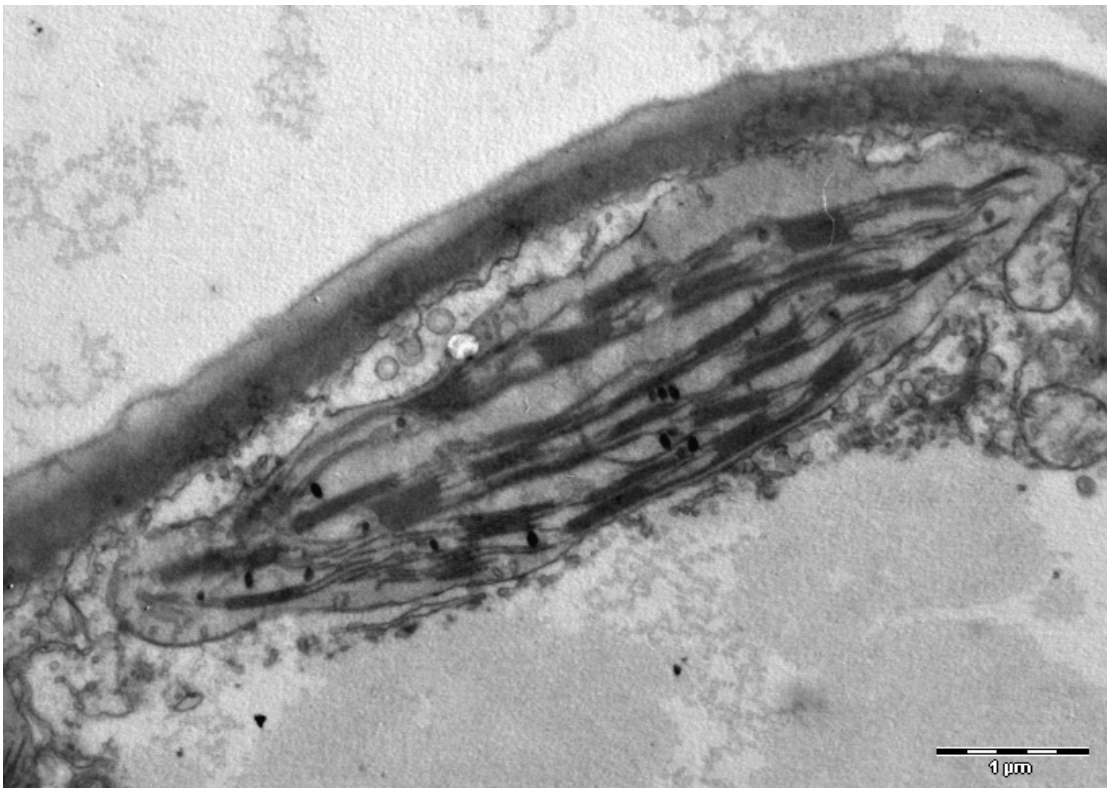
Chloroplasts are observable as flat discs usually 2 to 10 micrometers in diameter and 1 micrometer thick. In land plants, they are, in general, 5 µm in diameter and 2.3 µm thick. The chloroplast is contained by an envelope that consists of an inner and an outer phospholipid membrane. Between these two layers is the intermembrane space. A typical parenchyma cell contains about 10 to 100 chloroplasts.



Chloroplast ultrastructure:
1. outer membrane
2. intermembrane space
3. inner membrane (1+2+3: envelope)
4. stroma (aqueous fluid)
5. thylakoid lumen (inside of thylakoid)
6. thylakoid membrane
7. granum (stack of thylakoids)
8. thylakoid (lamella)
9. starch

10. ribosome
11. plastidial DNA
12. plastoglobule (drop of lipids)

The material within the chloroplast is called the stroma, corresponding to the cytosol of the original bacterium, and contains one or more molecules of small circular DNA. It also contains ribosomes; however most of its proteins are encoded by genes contained in the host cell nucleus, with the protein products transported to the chloroplast.



TEM image of a chloroplast

Within the stroma are stacks of thylakoids, the sub-organelles, which are the site of photosynthesis. The thylakoids are arranged in stacks called grana (singular: granum). A thylakoid has a flattened disk shape. Inside it is an empty area called the thylakoid space or lumen. Photosynthesis takes place on the thylakoid membrane; as in mitochondrial oxidative phosphorylation, it involves the coupling of cross-membrane fluxes with biosynthesis via the dissipation of a proton electrochemical gradient.

In the electron microscope, thylakoid membranes appear as alternating light-and-dark bands, each 0.01 µm thick. Embedded in the thylakoid membrane are antenna complexes, each of which consists of the light-absorbing pigments, including chlorophyll and carotenoids, as well as proteins that bind the pigments. This complex both increases the surface area for light capture, and allows capture of photons with a wider range of wavelengths. The energy of the incident photons is absorbed by the pigments and

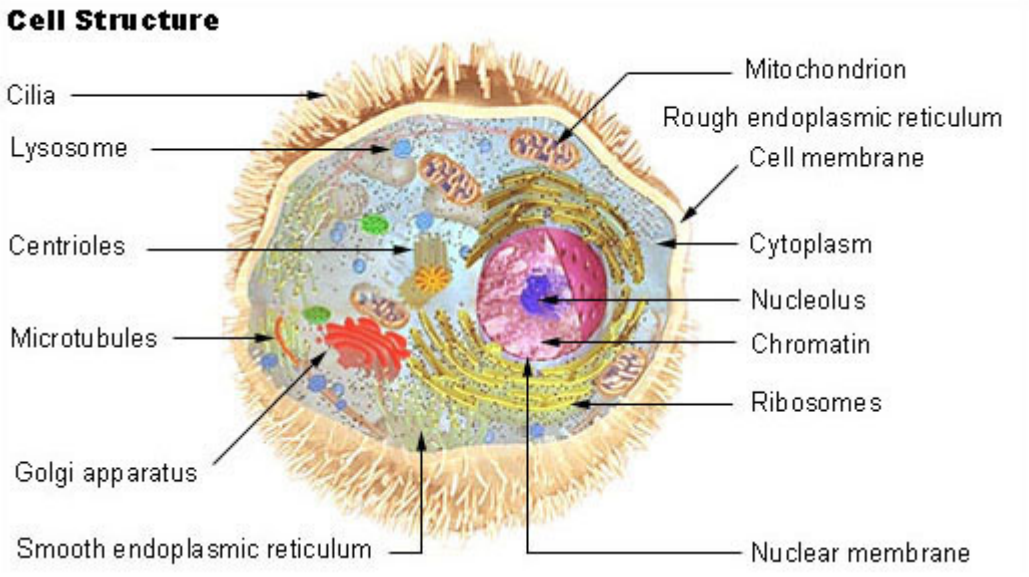
funneled to the reaction centre of this complex through resonance energy transfer. Two chlorophyll molecules are then ionised, producing an excited electron, which then passes onto the photochemical reaction centre.

Recent studies have shown that chloroplasts can be interconnected by tubular bridges called stromules, formed as extensions of their outer membranes. Chloroplasts appear to be able to exchange proteins via stromules, and thus function as a network.

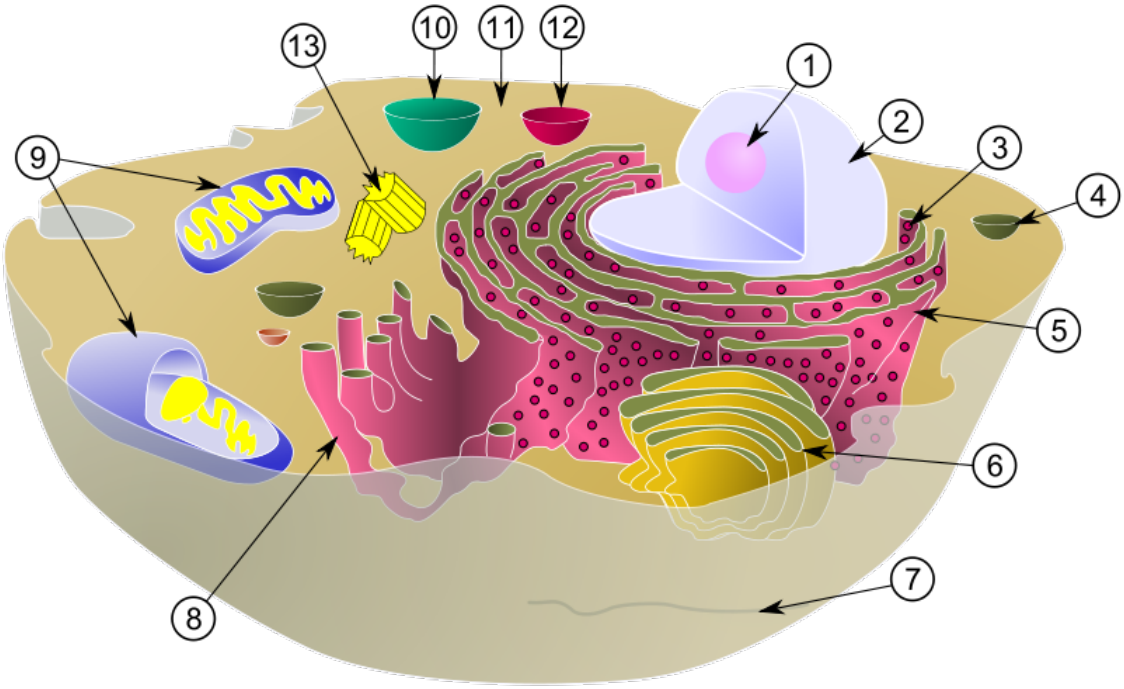
Transplastomic plants

Recently, chloroplasts have caught attention by developers of genetically modified plants. In most flowering plants, chloroplasts are not inherited from the male parent, although in plants such as pines, chloroplasts are inherited from males. Where chloroplasts are inherited only from the female, transgenes in these plastids cannot be disseminated by pollen. This makes plastid transformation a valuable tool for the creation and cultivation of genetically modified plants that are biologically contained, thus posing significantly lower environmental risks. This biological containment strategy is therefore suitable for establishing the coexistence of conventional and organic agriculture. While the reliability of this mechanism has not yet been studied for all relevant crop species, recent results in tobacco plants are promising, showing a failed containment rate of transplastomic plants at 3 in 1,000,000.

Lysosome



Various organelles labeled. The **lysosome** is labeled in the upper left.



Schematic of typical animal cell, showing subcellular components. Organelles:
(1) nucleolus
(2) nucleus
(3) ribosomes (little dots)
(4) vesicle
(5) rough endoplasmic reticulum (ER)
(6) Golgi apparatus
(7) Cytoskeleton
(8) smooth endoplasmic reticulum
(9) mitochondria
(10) vacuole
(11) cytosol
(12) lysosome
(13) centrioles within centrosome

Lysosomes are cellular organelles that contain acid hydrolase enzymes to break up waste materials and cellular debris. They are found in animal cells, while in yeast and plants the same roles are performed by lytic vacuoles. Lysosomes digest excess or worn-out organelles, food particles, and engulfed viruses or bacteria. The membrane around a lysosome allows the digestive enzymes to work at the 4.5 pH they require. Lysosomes fuse with vacuoles and dispense their enzymes into the vacuoles, digesting their contents. They are created by the addition of hydrolytic enzymes to early endosomes from the Golgi apparatus. The name *lysosome* derives from the Greek words **lysis**, *to separate*, and **soma**, *body*. They are frequently nicknamed "suicide-bags" or "suicide-sacs" by cell biologists due to their role in autolysis. Lysosomes were discovered by the Belgian cytologist Christian de Duve in the 1950s.

The size of lysosomes varies from 0.1–1.2 μm. At pH 4.8, the interior of the lysosomes is acidic compared to the slightly alkaline cytosol (pH 7.2). The lysosome maintains this pH differential by pumping protons (H^+ ions) from the cytosol across the membrane via proton pumps and chloride ion channels. The lysosomal membrane protects the cytosol, and therefore the rest of the cell, from the degradative enzymes within the lysosome. The cell is additionally protected from any lysosomal acid hydrolases that leak into the cytosol, as these enzymes are pH-sensitive and do not function as well in the alkaline environment of the cytosol.

Enzymes

Some important enzymes found within lysosomes include:

- Lipase, which digests lipids
- Amylase, which digests amylose, starch, and maltodextrins
- Proteases, which digest proteins
- Nucleases, which digest nucleic acids
- phosphoric acid monoesters.

Lysosomal enzymes are synthesized in the cytosol and the endoplasmic reticulum, where they receive a mannose-6-phosphate tag that targets them for the lysosome . Aberrant lysosomal targeting causes inclusion-cell disease, whereby enzymes do not properly reach the lysosome, resulting in accumulation of waste within these organelles.

Functions

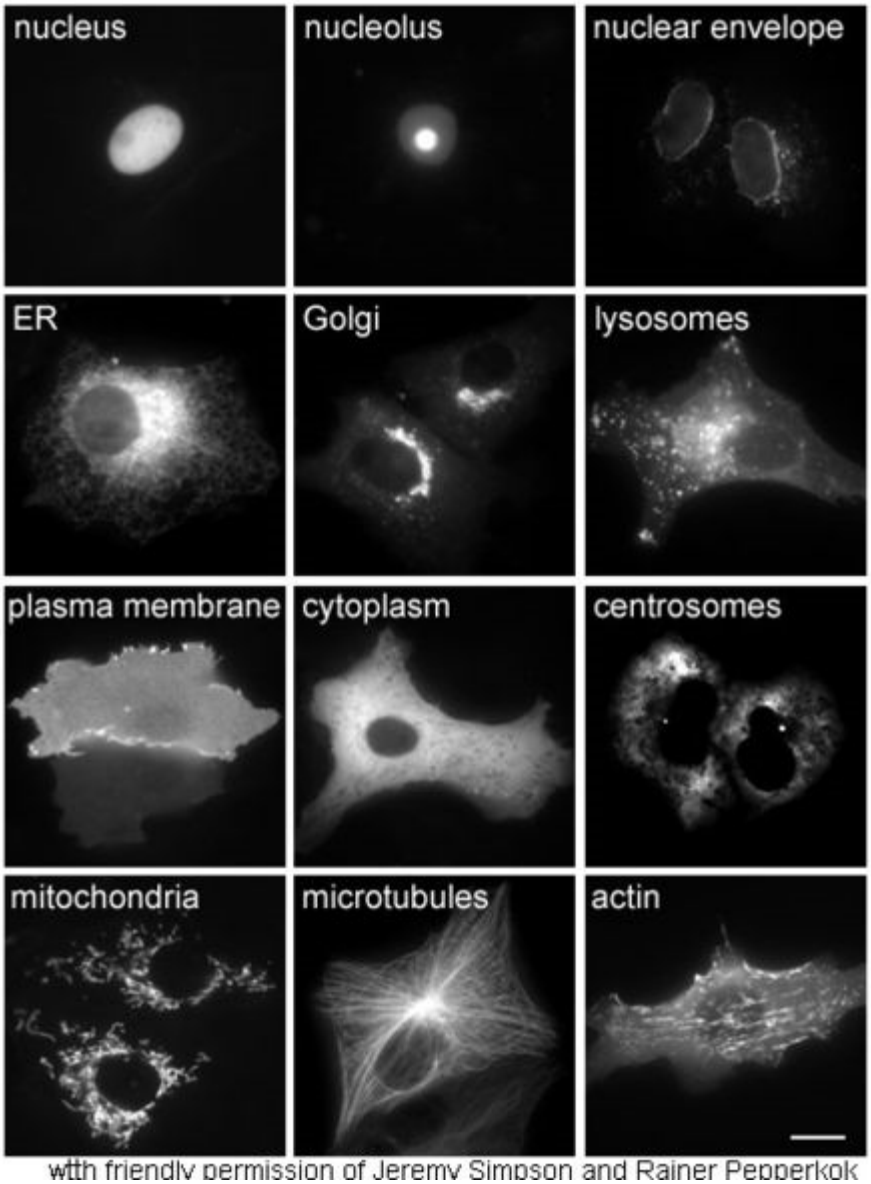
Lysosomes are the cell's waste disposal system and can break up anything. They digest almost everything. They are used for the digestion of macromolecules from phagocytosis (ingestion of other dying cells or larger extracellular material, like foreign invading microbes), endocytosis (where receptor proteins are recycled from the cell surface), and autophagy (where in old or unneeded organelles or proteins, or microbes that have invaded the cytoplasm are delivered to the lysosome). Autophagy may also lead to autophagic cell death, a form of programmed self-destruction, or autolysis, of the cell, which means that the cell is digesting itself.

Other functions include digesting foreign bacteria (or other forms of waste) that invade a cell and helping repair damage to the plasma membrane by serving as a membrane patch, sealing the wound. In the past, lysosomes were thought to kill cells that are no longer wanted, such as those in the tails of tadpoles or in the web from the fingers of a 3- to 6-month-old fetus. While lysosomes digest some materials in this process, it is actually accomplished through programmed cell death, called apoptosis.

Clinical relevance

There are a number of **lysosomal storage diseases** that are caused by the malfunction of the lysosomes or one of their digestive proteins; examples include Tay-Sachs disease and Pompe's disease. These diseases are caused by a defective or missing digestive protein, which leads to the accumulation of substrates within the cell, impairing metabolism.

In the broad sense, these can be classified as mucopolysaccharidoses, GM_2 gangliosidoses, lipid storage disorders, glycoproteinoses, mucolipidoses, or leukodystrophies.

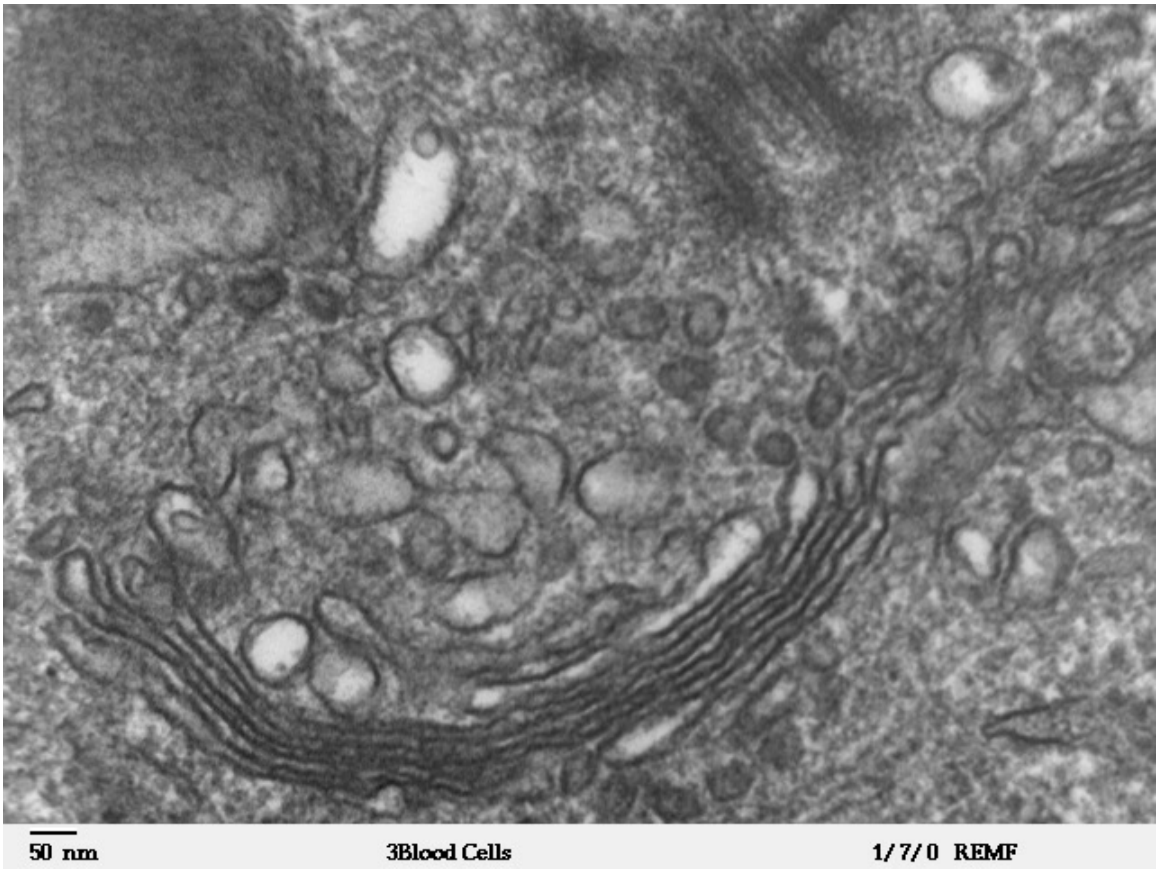


With friendly permission of Jeremy Simpson and Rainer Pepperkok

Proteins in different cellular compartments and structures tagged with green fluorescent protein.

Chapter- 5

Golgi Apparatus



Micrograph of Golgi apparatus, visible as a stack of semicircular black rings near the bottom. Numerous circular vesicles can be seen in proximity to the organelle

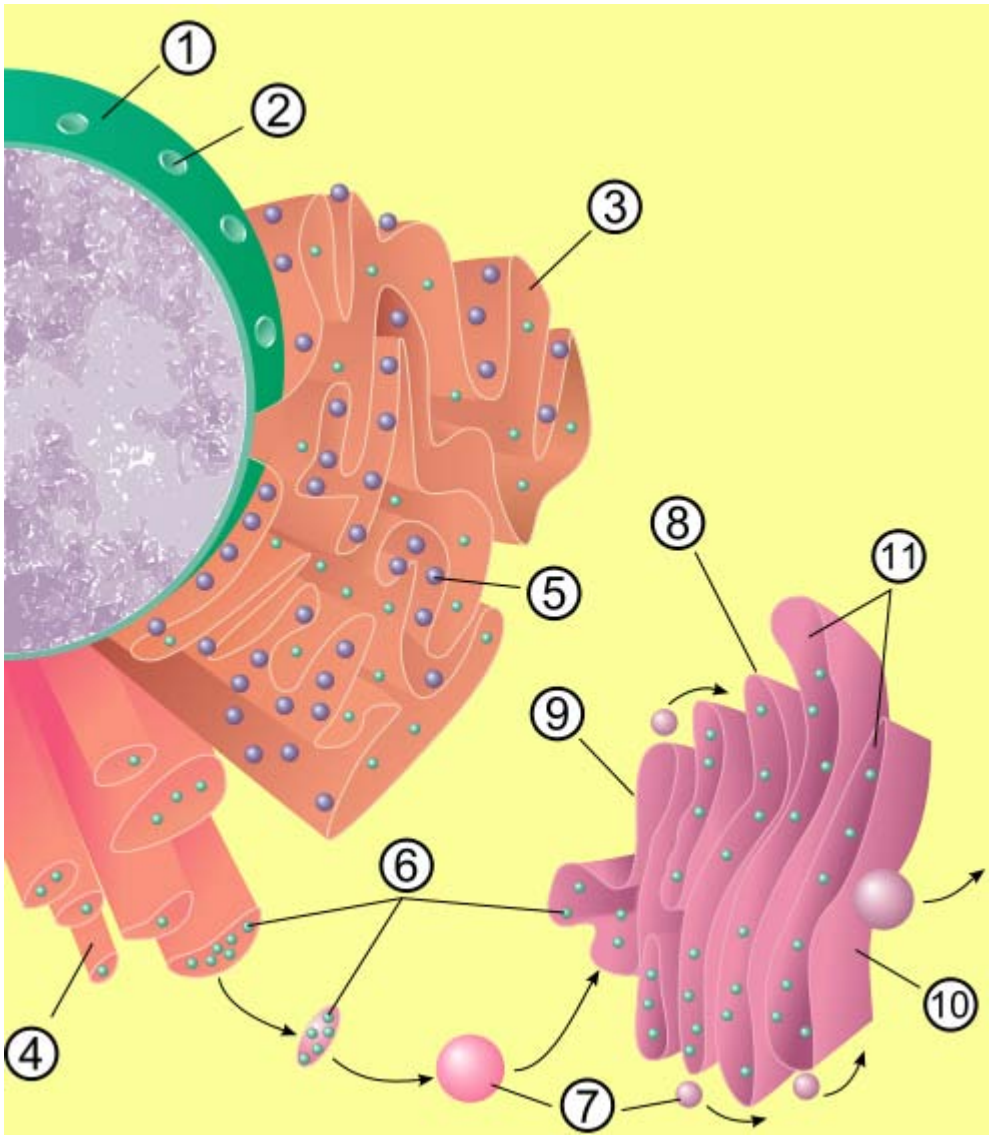


Diagram of secretory process from endoplasmic reticulum (orange) to Golgi apparatus (pink). 1. Nuclear membrane; 2. Nuclear pore; 3. Rough endoplasmic reticulum (RER); 4. Smooth endoplasmic reticulum (SER); 5. Ribosome attached to RER; 6. Macromolecules; 7. Transport vesicles; 8. Golgi apparatus; 9. *Cis* face of Golgi apparatus; 10. *Trans* face of Golgi apparatus; 11. Cisternae of lipids

The **Golgi apparatus** (also **Golgi body** or the **Golgi complex**) is an organelle found in most eukaryotic cells. It was identified in 1897 by the Italian physician Camillo Golgi, after whom the Golgi apparatus is named.

The Golgi apparatus processes and packages macromolecules, such as proteins and lipids, after their synthesis and before they make their way to their destination; it is particularly important in the processing of proteins for secretion. The Golgi apparatus forms a part of the cellular endomembrane system.

Evolution

The Golgi apparatus appears to have existed even in the "ancestral eukaryote" from which all modern eukaryotes evolved, even though some no longer have it.

Discovery

Due to its fairly large size, the Golgi apparatus was one of the first organelles to be discovered and observed in detail. The apparatus was discovered in 1897 by Italian physician Camillo Golgi during an investigation of the nervous system. After first observing it under his microscope, he termed the structure the *internal reticular apparatus*. The structure was then renamed after Golgi not long after the announcement of his discovery in 1898. However, some doubted the discovery at first, arguing that the appearance of the structure was merely an optical illusion created by the observation technique used by Golgi. With the development of modern microscopes in the 20th century, the discovery was confirmed.

Structure

Found in both plant and animal cells, the Golgi is composed of stacks of membrane-bound structures known as cisternae (singular: *cisterna*). An individual stack is sometimes called a dictyosome (from Greek *dictyon*, net + *soma*, body), especially in plant cells. A mammalian cell typically contains 40 to 100 stacks. Between four and eight cisternae are usually present in a stack; however, in some protists as many as sixty have been observed. Each cisterna comprises a flat, membrane enclosed disc that includes special Golgi enzymes which modify or help to modify cargo proteins that travel through it.

The cisternae stack has four functional regions: the cis-Golgi network, medial-Golgi, endo-Golgi, and trans-Golgi network. Vesicles from the endoplasmic reticulum (via the vesicular-tubular clusters) fuse with the network and subsequently progress through the stack to the trans Golgi network, where they are packaged and sent to the required destination. Each region contains different enzymes which selectively modify the contents depending on where they reside. The cisternae also carry structural proteins important for their maintenance as flattened membranes which stack upon each other.

Function

Cells synthesize a large number of different macromolecules. The Golgi apparatus is integral in modifying, sorting, and packaging these macromolecules for cell secretion (exocytosis) or use within the cell. It primarily modifies proteins delivered from the rough endoplasmic reticulum but is also involved in the transport of lipids around the cell, and the creation of lysosomes. In this respect it can be thought of as similar to a post office; it packages and labels items which it then sends to different parts of the cell.

Enzymes within the cisternae are able to modify the proteins by addition of carbohydrates (glycosylation) and phosphates (phosphorylation). In order to do so, the Golgi imports substances such as nucleotide sugars from the cytosol. These modifications may also form a signal sequence which determines the final destination of the protein. For example, the Golgi apparatus adds a mannose-6-phosphate label to proteins destined for lysosomes.

The Golgi plays an important role in the synthesis of proteoglycans, which are molecules present in the extracellular matrix of animals. It is also a major site of carbohydrate synthesis. This includes the production of glycosaminoglycans (GAGs), long unbranched polysaccharides which the Golgi then attaches to a protein synthesised in the endoplasmic reticulum to form proteoglycans. Enzymes in the Golgi polymerize several of these GAGs via a xylose link onto the core protein. Another task of the Golgi involves the sulfation of certain molecules passing through its lumen via sulphotranferases that gain their sulphur molecule from a donor called PAPs. This process occurs on the GAGs of proteoglycans as well as on the core protein. The level of sulfation is very important to the proteoglycans' signalling abilities as well as giving the proteoglycan its overall negative charge.

The phosphorylation of molecules requires that ATP is imported into the lumen of the Golgi and then utilised by resident kinases such as casein kinase 1 and casein kinase 2. One molecule that is phosphorylated in the Golgi is Apolipoprotein, which forms a molecule known as VLDL that is a constitute of blood serum. It is thought that the phosphorylation of these molecules is important to help aid in their sorting for secretion into the blood serum.

The Golgi has a putative role in apoptosis, with several Bcl-2 family members localised there, as well as to the mitochondria. A newly characterized protein, GAAP (Golgi anti-apoptotic protein), almost exclusively resides in the Golgi and protects cells from apoptosis by an as-yet undefined mechanism.

Vesicular transport

The vesicles that leave the rough endoplasmic reticulum are transported to the *cis* face of the Golgi apparatus, where they fuse with the Golgi membrane and empty their contents into the lumen. Once inside the lumen, the molecules are modified, sorted and shipped towards their final destination. The Golgi apparatus tends to be larger and more numerous in cells that synthesise and secrete large amounts of substances, for example, the plasma B cells and the antibody-secreting cells of the immune system have prominent Golgi complexes.

Those proteins destined for areas of the cell other than either the endoplasmic reticulum or Golgi apparatus are moved towards the *trans* face, to a complex network of membranes and associated vesicles known as the *trans-Golgi network* (TGN). This area of the Golgi is the point at which proteins are sorted and shipped to their intended

destinations by their placement into one of at least three different types of vesicles, depending upon the molecular marker they carry:

Type	Description	Example
Exocytotic vesicles *(continuous)*	Vesicle contains proteins destined for extracellular release. After packaging the vesicles bud off and immediately move towards the plasma membrane, where they fuse and release the contents into the extracellular space in a process known as *constitutive secretion*.	Antibody release by activated plasma B cells
Secretory vesicles *(regulated)*	Vesicle contains proteins destined for extracellular release. After packaging, the vesicles bud off and are stored in the cell until a signal is given for their release. When the appropriate signal is received they move towards the membrane and fuse to release their contents. This process is known as *regulated secretion*.	Neurotransmitter release from neurons
Lysosomal vesicles	Vesicle contains proteins destined for the lysosome, an organelle of degradation containing many acid hydrolases, or to lysosome-like storage organelles. These proteins include both digestive enzymes and membrane proteins. The vesicle first fuses with the late endosome, and the contents are then transferred to the lysosome via unknown mechanisms.	Digestive proteases destined for the lysosome

Transport mechanism

The transport mechanism which proteins use to progress through the Golgi apparatus is not yet clear; however a number of hypotheses currently exist. Until recently, the vesicular transport mechanism was favoured but now more evidence is coming to light to support cisternal maturation. The two proposed models may actually work in conjunction with each other, rather than being mutually exclusive. This is sometimes referred to as the *combined* model.

- ***Cisternal maturation model***: the cisternae of the Golgi apparatus move by being built at the *cis* face and destroyed at the *trans* face. Vesicles from the endoplasmic reticulum fuse with each other to form a cisterna at the *cis* face, consequently this cisterna would appear to move through the Golgi stack when a new cisterna is formed at the *cis* face. This model is supported by the fact that structures larger than the transport vesicles, such as collagen rods, were observed microscopically to progress through the Golgi apparatus. This was initially a popular hypothesis, but lost favour in the 1980s. Recently it has made a comeback, as laboratories at the University of Chicago and the University of Tokyo have been able to use new technology to directly observe Golgi compartments maturing. Additional evidence comes from the fact that COPI vesicles move in the retrograde direction,

transporting Endoplasmic Reticulum proteins back to where they belong by recognizing a signal peptide.

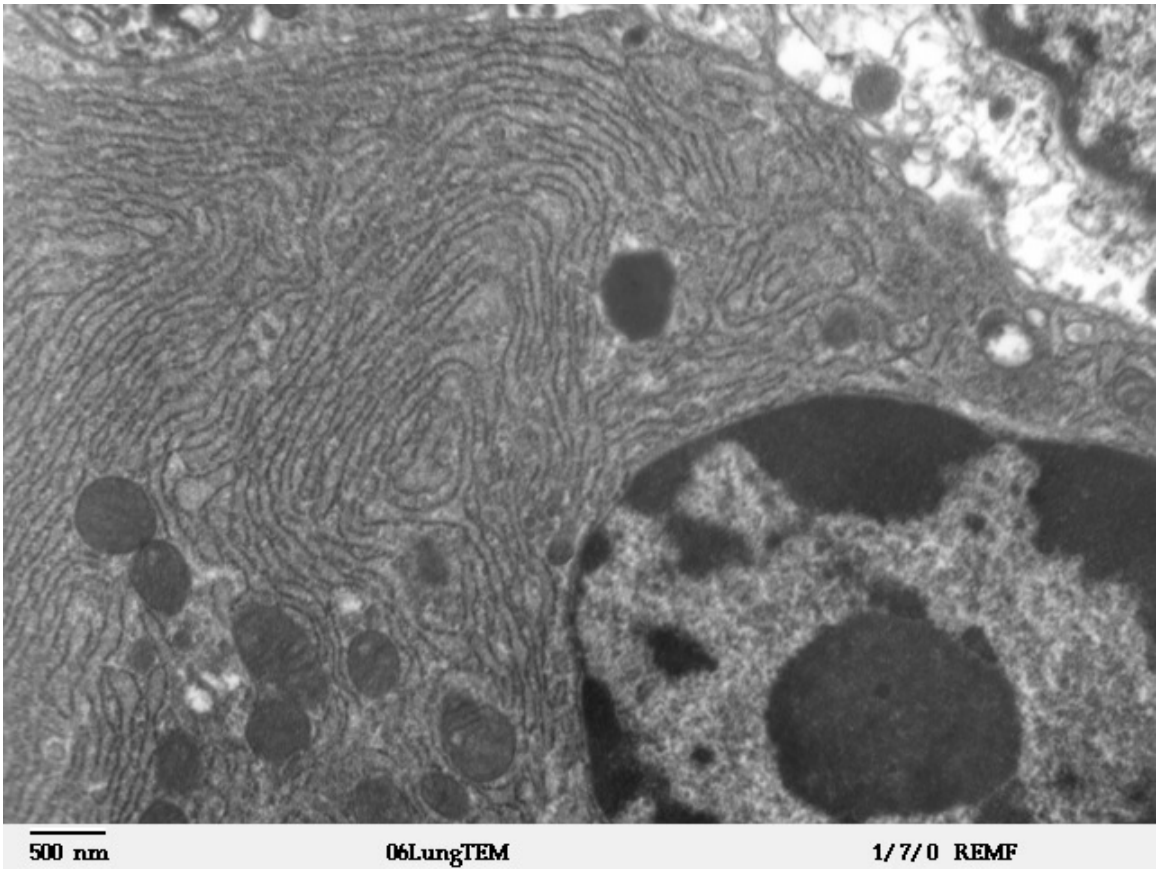
- ***Vesicular transport model***: Vesicular transport views the Golgi as a very stable organelle, divided into compartments in the cis to trans direction. Membrane bound carriers transport material between the ER and the different compartments of the Golgi. Experimental evidence includes the abundance of small vesicles (known technically as shuttle vesicles) in proximity to the Golgi apparatus. To direct the vesicles, actin filaments connect packaging proteins to the membrane to ensure that they fuse with the correct compartment.

Golgi apparatus during mitosis

The Golgi apparatus will break up and disappear following the onset of mitosis, or cellular division. During the telophase of mitosis, the Golgi apparatus reappears; however, it is still uncertain how this occurs.

Chapter- 6

Endoplasmic Reticulum

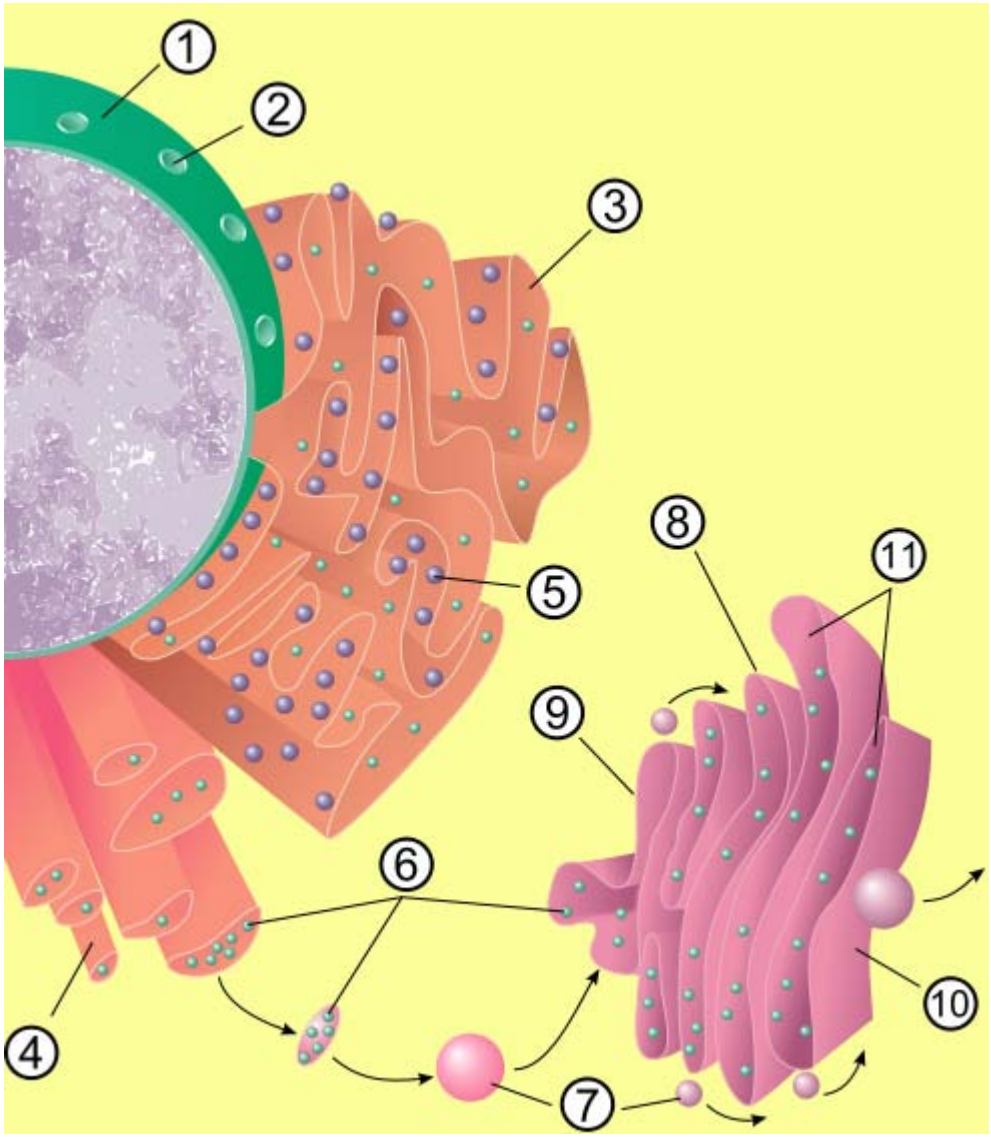


Micrograph of rough endoplasmic reticulum network around the nucleus (shown in lower right-hand side of the picture). Dark small circles in the network are mitochondria.

The **endoplasmic reticulum** (**ER**) is an eukaryotic organelle that forms an interconnected network of tubules, vesicles, and cisternae within cells. **Rough endoplasmic reticulua** synthesize proteins, while **smooth endoplasmic reticulua** synthesize lipids and steroids, metabolize carbohydrates and steroids, and regulate calcium concentration, drug detoxification, and attachment of receptors on cell membrane proteins. **Sarcoplasmic reticulua** solely regulate calcium levels.

The lacey membranes of the endoplasmic reticulum were first seen by Keith R. Porter, Albert Claude, and Ernest F. Fullam in 1945.

Structure



1 Nucleus **2** Nuclear pore **3** Rough endoplasmic reticulum (RER) **4** Smooth endoplasmic reticulum (SER) **5** Ribosome on the rough ER **6** Proteins that are transported **7** Transport vesicle **8** Golgi apparatus **9** Cis face of the Golgi apparatus **10** Trans face of the Golgi apparatus **11** Cisternae of the Golgi apparatus

The general structure of the endoplasmic reticulum is an extensive membrane network of cisternae (sac-like structures) held together by the cytoskeleton. The phospholipid membrane encloses a space, the cisternal space (or lumen), from the cytosol, which is continuous with the perinuclear space. The functions of the endoplasmic reticulum vary

greatly depending on the exact type of endoplasmic reticulum and the type of cell in which it resides. The three varieties are called *rough endoplasmic reticulum*, *smooth endoplasmic reticulum* and *sarcoplasmic reticulum*.

The quantity of RER and SER in a cell can quickly interchange from one type to the other, depending on changing metabolic needs: one type will undergo numerous changes including new proteins embedded in the membranes in order to transform. Also, massive changes in the protein content can occur without any noticeable structural changes, depending on the enzymatic needs of the cell (as per the functions listed below).

Rough endoplasmic reticulum

The surface of the rough endoplasmic reticulum (RER) is studded with protein-manufacturing ribosomes giving it a "rough" appearance (hence its name). However, the ribosomes bound to the RER at any one time are not a stable part of this organelle's structure as ribosomes are constantly being bound and released from the membrane. A ribosome only binds to the ER once it begins to synthesize a protein destined for the secretory pathway. Here, a ribosome in the cytosol begins synthesizing a protein until a signal recognition particle recognizes the pre-piece of 5-15 hydrophobic amino acids preceded by a positively charged amino acid. This signal sequence allows the recognition particle to bind to the ribosome, causing the ribosome to bind to the RER and pass the new protein through the ER membrane. The pre-piece is then cleaved off within the lumen of the ER and the ribosome released back into the cytosol.

The membrane of the RER is continuous with the outer layer of the nuclear envelope. Although there is no continuous membrane between the RER and the Golgi apparatus, membrane-bound vesicles shuttle proteins between these two compartments. Vesicles are surrounded by coating proteins called COPI and COPII. COPII targets vesicles to the golgi and COPI marks them to be brought back to the RER. The RER works in concert with the Golgi complex to target new proteins to their proper destinations. A second method of transport out of the ER are areas called membrane contact sites, where the membranes of the ER and other organelles are held closely together, allowing the transfer of lipids and other small molecules.

The RER is key in multiple functions:

- lysosomal enzymes with a mannose-6-phosphate marker added in the *cis*-Golgi network
- Secreted proteins, either secreted constitutively with no tag, or regulated secretion involving clathrin and paired basic amino acids in the signal peptide.
- integral membrane proteins that stay imbedded in the membrane as vesicles exit and bind to new membranes. Rab proteins are key in targeting the membrane, SNAP and SNARE proteins are key in the fusion event.
- initial glycosylation as assembly continues. This is either N-linked (O-linking occur in the golgi).

- N-linked glycosylation: if the protein is properly folded, glycosyltransferase recognizes the AA sequence NXS or NXT (with the S/T residue phosphorylated) and adds a 14 sugar backbone (2 *N*-acetylglucosamine, 9 branching mannose, and 3 glucose at the end) to the side chain nitrogen of Asn.

Smooth endoplasmic reticulum

The smooth endoplasmic reticulum (SER) has functions in several metabolic processes, including synthesis of lipids and steroids, metabolism of carbohydrates, regulation of calcium concentration, drug detoxification, attachment of receptors on cell membrane proteins, and steroid metabolism. It is connected to the nuclear envelope. Smooth endoplasmic reticulum is found in a variety of cell types (both animal and plant) and it serves different functions in each. The Smooth ER also contains the enzyme glucose-6-phosphatase which converts glucose-6-phosphate to glucose, a step in gluconeogenesis. The SER consists of tubules and vesicles that branch forming a network. In some cells there are dilated areas like the sacs of RER. The network of SER allows increased surface area for the action or storage of key enzymes and the products of these enzymes.

Sarcoplasmic reticulum

The sarcoplasmic reticulum (SR), from the Greek *sarx*, ("flesh"), is a special type of smooth ER found in smooth and striated muscle. The only structural difference between this organelle and the SER is the medley of proteins they have, both bound to their membranes and drifting within the confines of their lumens. This fundamental difference is indicative of their functions: the SER synthesizes molecules while the SR stores and pumps calcium ions. The SR contains large stores of calcium, which it sequesters and then releases when the muscle cell is stimulated. The SR's release of calcium upon electrical stimulation of the cell plays a major role in excitation-contraction coupling.

Functions

The endoplasmic reticulum serves many general functions, including the facilitation of protein folding and the transport of synthesized proteins in sacs called cisternae.

Correct folding of newly-made proteins is made possible by several endoplasmic reticulum chaperone proteins, including protein disulfide isomerase (PDI), ERp29, the Hsp70 family member Grp78, calnexin, calreticulin, and the peptidylpropyl isomerase family. Only properly-folded proteins are transported from the rough ER to the Golgi complex.

Transport of proteins

Secretory proteins, mostly glycoproteins, are moved across the endoplasmic reticulum membrane. Proteins that are transported by the endoplasmic reticulum and from there throughout the cell are marked with an address tag called a signal sequence. The N-

terminus (one end) of a polypeptide chain (i.e., a protein) contains a few amino acids that work as an address tag, which are removed when the polypeptide reaches its destination. Proteins that are destined for places outside the endoplasmic reticulum are packed into transport vesicles and moved along the cytoskeleton toward their destination.

The endoplasmic reticulum is also part of a protein sorting pathway. It is, in essence, the transportation system of the eukaryotic cell. The majority of endoplasmic reticulum resident proteins are retained in the endoplasmic reticulum through a retention motif. This motif is composed of four amino acids at the end of the protein sequence. The most common retention sequence is KDEL (*lys-asp-glu-leu*). However, variation on KDEL does occur and other sequences can also give rise to endoplasmic reticulum retention. It is not known if such variation can lead to sub-endoplasmic reticulum localizations. There are three KDEL receptors in mammalian cells, and they have a very high degree of sequence identity. The functional differences between these receptors remain to be established.

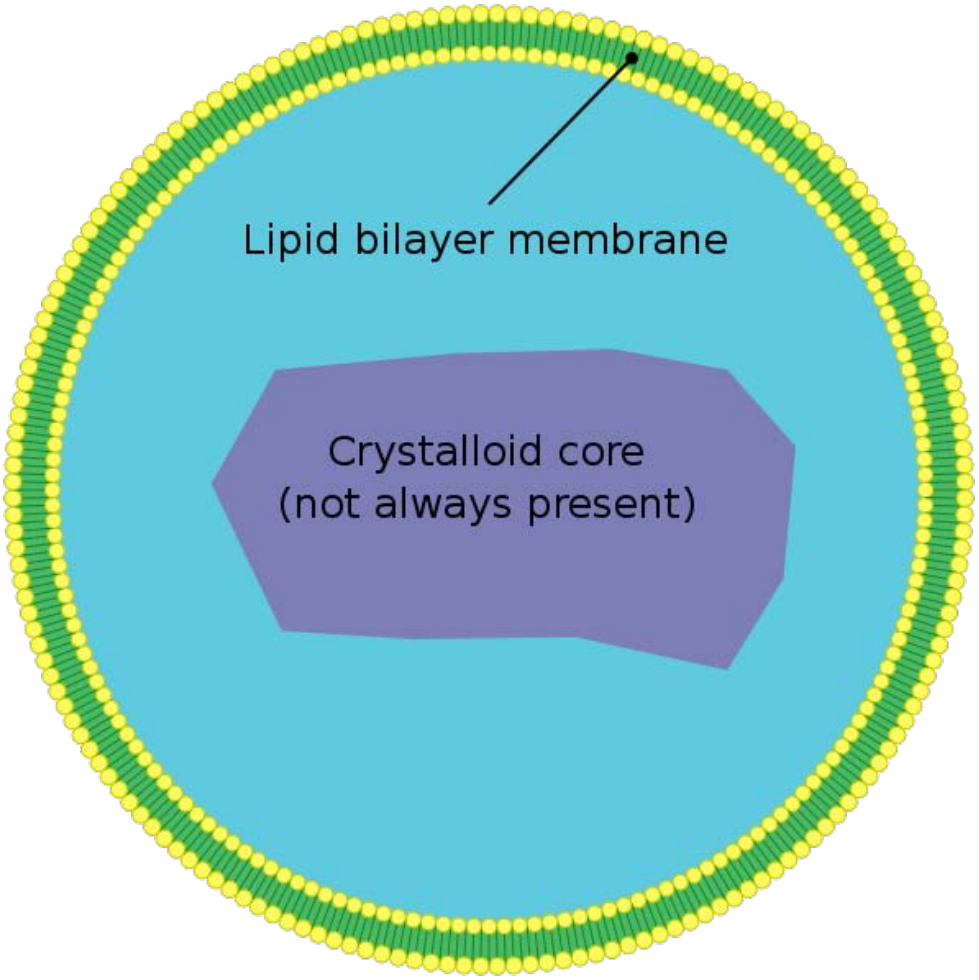
Other functions

- **Insertion of proteins into the endoplasmic reticulum membrane:** Integral membrane proteins are inserted into the endoplasmic reticulum membrane as they are being synthesized (co-translational translocation). Insertion into the endoplasmic reticulum membrane requires the correct topogenic signal sequences in the protein.
- **Glycosylation**: Glycosylation involves the attachment of oligosaccharides.
- **Disulfide bond formation and rearrangement**: Disulfide bonds stabilize the tertiary and quaternary structure of many proteins.
- **Drug metabolism**: The smooth ER is the site at which some drugs are modified by microsomal enzymes which include the cytochrome P450 enzymes.

Chapter- 7

Peroxisome and Vacuole

Peroxisome



Basic structure of a peroxisome

Peroxisomes'also called microbodies' are organelles found in virtually all eukaryotic cells. They are involved in the catabolism of very long chain fatty acids, branched chain fatty acids, D-amino acids, polyamines, and biosynthesis of plasmalogens, etherphospholipids critical for the normal function of mammalian brains and lungs. They also contain approximately 10% of the total activity of two enzymes in the pentose phosphate pathway, which is important for energy metabolism. It is rigorously debated if peroxisomes are involved in isoprenoid and cholesterol synthesis in animals. Other known peroxisomal functions include the glyoxylate cycle in germinating seeds ("glyoxysomes"), photorespiration in leaves, glycolysis in trypanosomes ("glycosomes"), and methanol and/or amine oxidation and assimilation in some yeasts.

Peroxisomes were identified as organelles by the Belgian cytologist Christian de Duve in 1967 after they had been first described in a PhD thesis of Rhodin a decade earlier.

Metabolic functions

A major function of the peroxisome is the breakdown of very long chain fatty acids through beta-oxidation. In animal cells, the very long fatty acids are converted to medium chain fatty acids, which are subsequently shuttled to mitochondria where they are eventually broken down to carbon dioxide and water. In yeast and plant cells, this process is exclusive for the peroxisome.

The first reactions in the formation of plasmalogen in animal cells also occur in peroxisomes. Plasmalogen is the most abundant phospholipid in myelin. Deficiency of plasmalogens causes profound abnormalities in the myelination of nerve cells, which is one reason why many peroxisomal disorders affect the nervous system. Peroxisomes also play a role in the production of bile acids important for the absorption of fats and fat-soluble vitamins, such as vitamin K.

Peroxisomes contain oxidative enzymes, such as catalase, D-amino acid oxidase, and uric acid oxidase. However the last enzyme is absent in humans, explaining the disease known as gout, caused by the accumulation of uric acid. Certain enzymes within the peroxisome, by using molecular oxygen, remove hydrogen atoms from specific organic substrates (labeled as R), in an oxidative reaction, producing hydrogen peroxide (H_2O_2, itself toxic):

$$\mathrm{RH_2 + O_2 \rightarrow R + H_2O_2}$$

Catalase, another peroxisomal enzyme, uses this H_2O_2 to oxidize other substrates, including phenols, formic acid, formaldehyde, and alcohol, by means of the peroxidation reaction:

$\mathrm{H_2O_2 + R'H_2 \rightarrow R' + 2H_2O}$, thus eliminating the poisonous hydrogen peroxide in the process.

This reaction is important in liver and kidney cells, where the peroxisomes detoxify various toxic substances that enter the blood. About 25% of the ethanol humans drink is oxidized to acetaldehyde in this way. In addition, when excess H_2O_2 accumulates in the cell, catalase converts it to H_2O through this reaction:

$$2H_2O_2 \rightarrow 2H_2O + O_2$$

In higher plants, peroxisomes contain also a complex battery of antioxidative enzymes such as superoxide dismutase, the components of the ascorbate-glutathione cycle, and the NADP-dehydrogenases of the pentose-phosphate pathway. It has been demonstrated the generation of superoxide ($O_2^{\bullet -}$) and nitric oxide ($^{\bullet}NO$) radicals.

The peroxisome of plant cells is polarised when fighting fungal penetration. Infection causes a glucosinolate molecule to play an antifungal role to be made and delivered to the outside of the cell through the action of the peroxisomal proteins (PEN2 and PEN3).

Peroxisome assembly

Peroxisomes can be derived from the endoplasmic reticulum and replicate by fission. Peroxisome matrix proteins are translated in the cytoplasm prior to import. Specific amino acid sequences (PTS or peroxisomal targeting signal) at the *C-terminus* (PTS1) or *N-terminus* (PTS2) of peroxisomal matrix proteins signals them to be imported into the organelle. There are at least 32 known peroxisomal proteins, called peroxins, which participate in the process of peroxisome assembly. Proteins do not have to unfold to be imported into the peroxisome. The protein receptors, the peroxins *PEX5* and *PEX7*, accompany their cargoes (containing a PTS1 or a PTS2 amino acid sequence, respectively) all the way into the peroxisome where they release the cargo and then return to the cytosol - a step named *recycling*. Overall, the import cycle is referred to as the *extended shuttle mechanism*. There is now evidence that ATP hydrolysis is required for the recycling of receptors to the cytosol. Also, ubiquitination appears to be crucial for the export of PEX5 from the peroxisome, to the cytosol.

Associated medical conditions

Peroxisomal disorders are a class of medical conditions that typically affect the human nervous system as well as many other organ systems. Two common examples are X-linked adrenoleukodystrophy and peroxisome biogenesis disorders.

Genes

PEX genes encode the protein machinery ("peroxins") required for proper peroxisome assembly, as described above. Membrane assembly and maintenance requires three of these (peroxins 3, 16, and 19) and may occur without the import of the matrix (lumen) enzymes. Proliferation of the organelle is regulated by Pex11p.

Genes that encode peroxin proteins include: PEX1, PEX2 - PXMP3, PEX3, PEX5, PEX6, PEX7, PEX10, PEX11A, PEX11B, PEX11G, PEX12, PEX13, PEX14, PEX16, PEX19, PEX26, PEX28, PEX30, and PEX31

Evolutionary origins

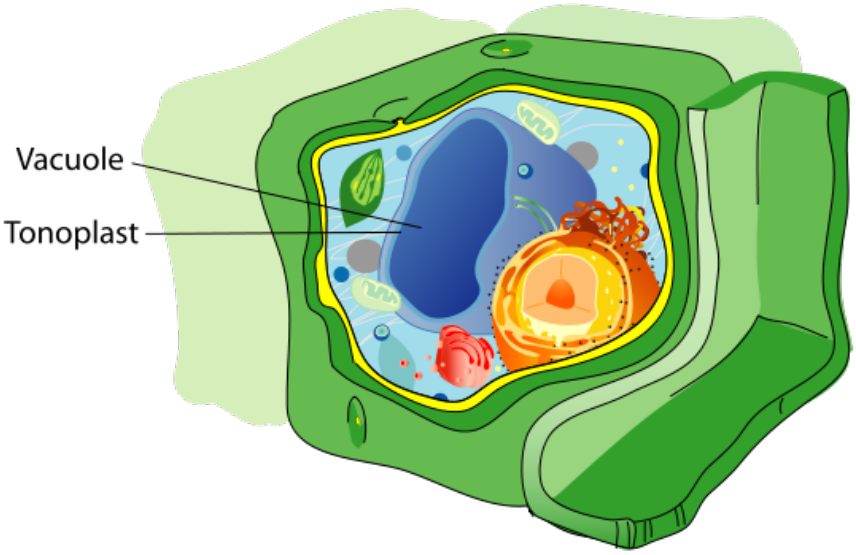
The protein content of peroxisomes varies across species, but the presence of proteins common to many species has been used to suggest an endosymbiotic origin; that is, peroxisomes evolved from bacteria that invaded larger cells as parasites, and very gradually evolved a symbiotic relationship. However, this view has been challenged by recent discoveries. For example, peroxisome-less mutants can restore peroxisomes upon introduction of the wild-type gene.

Two independent evolutionary analyses of the peroxisomal proteome found homologies between the peroxisomal import machinery and the ERAD pathway in the endoplasmic reticulum, along with a number of metabolic enzymes that were likely recruited from the mitochondria. Recently, it has been suggested that the peroxisome may have had an actinobacterial origin, however, this is controversial.

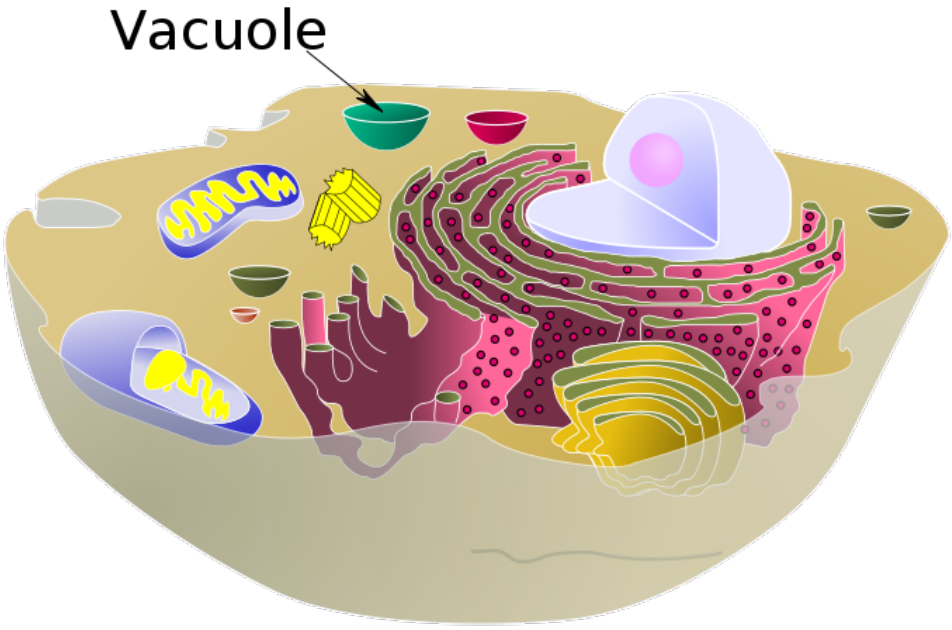
Other related organelles

Other organelles of the microbody family related to peroxisomes include glyoxysomes of plants and filamentous fungi, glycosomes of kinetoplastids and Woronin bodies of filamentous fungi.

Vacuole



Plant cell structure



Animal cell structure

A **vacuole** is a membrane-bound organelle which is present in all plant and fungal cells and some protist, animal and bacterial cells. Vacuoles are essentially enclosed compartments which are filled with water containing inorganic and organic molecules including enzymes in solution, though in certain cases they may contain solids which have been engulfed. Vacuoles are formed by the fusion of multiple membrane vesicles and are effectively just larger forms of these. The organelle has no basic shape or size, its structure varies according to the needs of the cell.

The function and importance of vacuoles varies greatly according to the type of cell in which they are present, having much greater prominence in the cells of plants, fungi and certain protists than those of animals and bacteria. In general, the functions of the vacuole include:

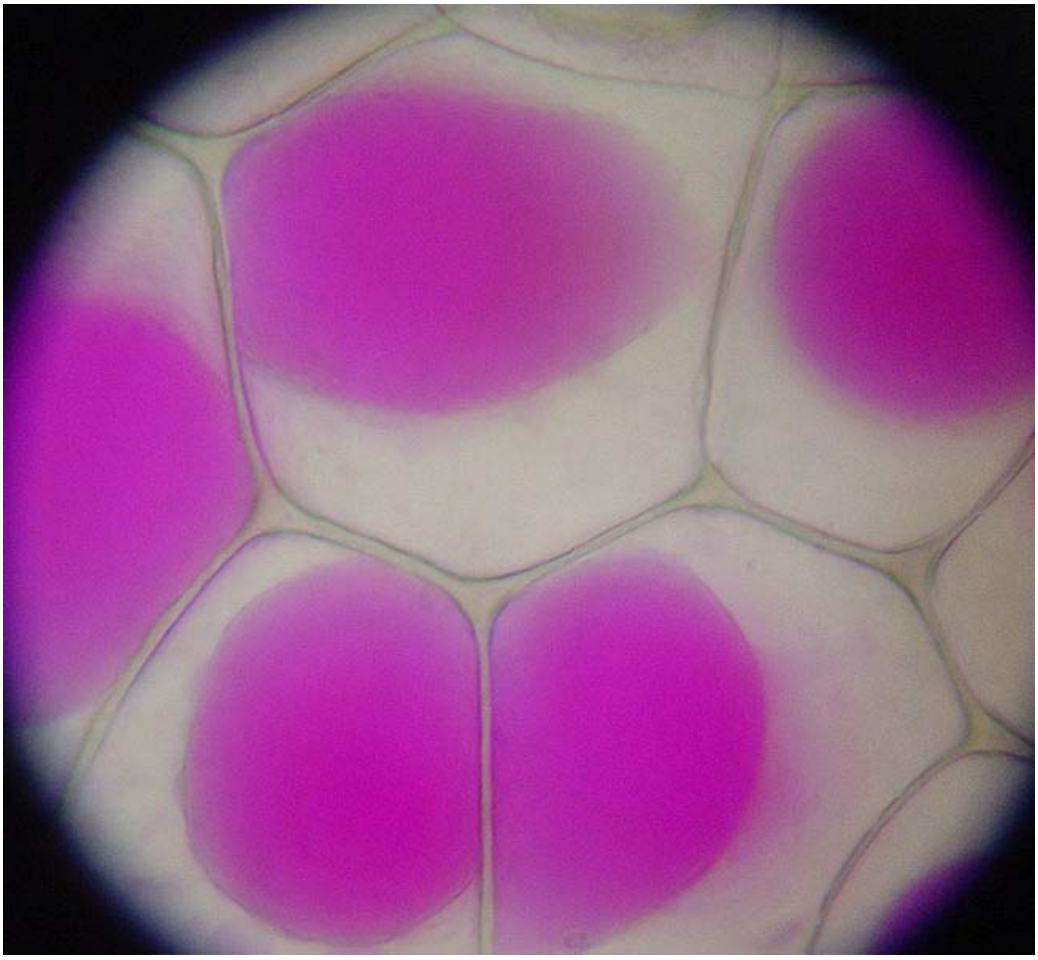
- Isolating materials that might be harmful or a threat to the cell
- Containing waste products
- Containing water in plant cells
- Maintaining internal hydrostatic pressure or turgor within the cell
- Maintaining an acidic internal pH
- Containing small molecules
- Exporting unwanted substances from the cell
- Allows plants to support structures such as leaves and flowers due to the pressure of the central vacuole
- In seeds, stored proteins needed for germination are kept in 'protein bodies', which are modified vacuoles.

Vacuoles also play a major role in autophagy, maintaining a balance between biogenesis (production) and degradation (or turnover), of many substances and cell structures in certain organisms. They also aid in the lysis and recycling of misfolded proteins that have begun to build up within the cell. Thomas Boller and others proposed that the vacuole participates in the destruction of invading bacteria and Robert B Mellor proposed organ-specific forms have a role in 'housing' symbiotic bacteria. In protists, vacuoles have the additional function of storing food which has been absorbed by the organism and assisting in the digestive and waste management process for the cell.

Bacteria

Large vacuoles are found in three genera of filamentous sulfur bacteria, the *Thioploca*, *Beggiatoa* and *Thiomargarita*. The cytosol is extremely reduced in these genera and the vacuole can occupy between 40-98% of the cell. The vacuoles contain high concentrations of nitrate ions and is therefore thought to be a storage organelle.

Plants



The vacuoles of spiderwort cells, stained in pink

Most mature plant cells have one large central vacuole that typically occupies more than 30% of the cell's volume, and that can occupy as much as 80% of the volume for certain cell types and conditions. Strands of cytoplasm often run through the vacuole.

A vacuole is surrounded by a membrane called the **tonoplast** (word origin: Gk tón(os) + -o-, meaning “stretching”, “tension”, “tone” + comb. form repr. Gk plastós formed, molded). Also called the **vacuolar membrane**, the tonoplast is the cytoplasmic membrane surrounding a vacuole, separating the vacuolar contents from the cell's cytoplasm. As a membrane, it is mainly involved in regulating the movements of ions around the cell, and isolating materials that might be harmful or a threat to the cell.

Transport of protons from the cytosol to the vacuole stabilises cytoplasmic pH, while making the vacuolar interior more acidic creating a proton motive force which the cell can use to transport nutrients into or out of the vacuole. The low pH of the vacuole also allows degradative enzymes to act. Although single large central vacuoles are most

common, the size and number of vacuoles may vary in different tissues and stages of development. For example, developing cells in the meristems contain small provacuoles and cells of the vascular cambium have many small vacuoles in the winter and one large one in the summer.

Aside from storage, the main role of the central vacuole is to maintain turgor pressure against the cell wall. Proteins found in the tonoplast (aquaporins) control the flow of water into and out of the vacuole through active transport, pumping potassium (K^+) ions into and out of the vacuolar interior. Due to osmosis, water will diffuse into the vacuole, placing pressure on the cell wall. If water loss leads to a significant decline in turgor pressure, the cell will plasmolyse. Turgor pressure exerted by vacuoles is also required for cellular elongation: as the cell wall is partially degraded by the action of expansins, the less rigid wall is expanded by the pressure coming from within the vacuole. Turgor pressure exerted by the vacuole is also essential in supporting plants in an upright position. Another function of a central vacuole is that it pushes all contents of the cell's cytoplasm against the cellular membrane, and thus keeps the chloroplasts closer to light.

Most plants store chemicals in the vacuole that react with chemicals in the cytosol. If the cell is broken, for example by a herbivore, then the two chemicals can react forming toxic chemicals. In garlic, alliin and the enzyme alliinase are normally separated but form allicin if the vacuole is broken. A similar reaction is responsible for the production of syn-propanethial-S-oxide when onions are cut.

Fungi

Vacuoles in fungal cells perform similar functions to those in plants and there can be more than one vacuole per cell. In yeast cells the vacuole is a dynamic structure that can rapidly modify its morphology. They are involved in many processes including the homeostasis of cell pH and the concentration of ions, osmoregulation, storing amino acids and polyphosphate and degradative processes. Toxic ions, such as strontium (Sr^{2+}), cobalt(II) (Co^{2+}), and lead(II) (Pb^{2+}) are transported into the vacuole to isolate them from the rest of the cell.

Animals

In animal cells, vacuoles perform mostly subordinate roles, assisting in larger processes of exocytosis and endocytosis.

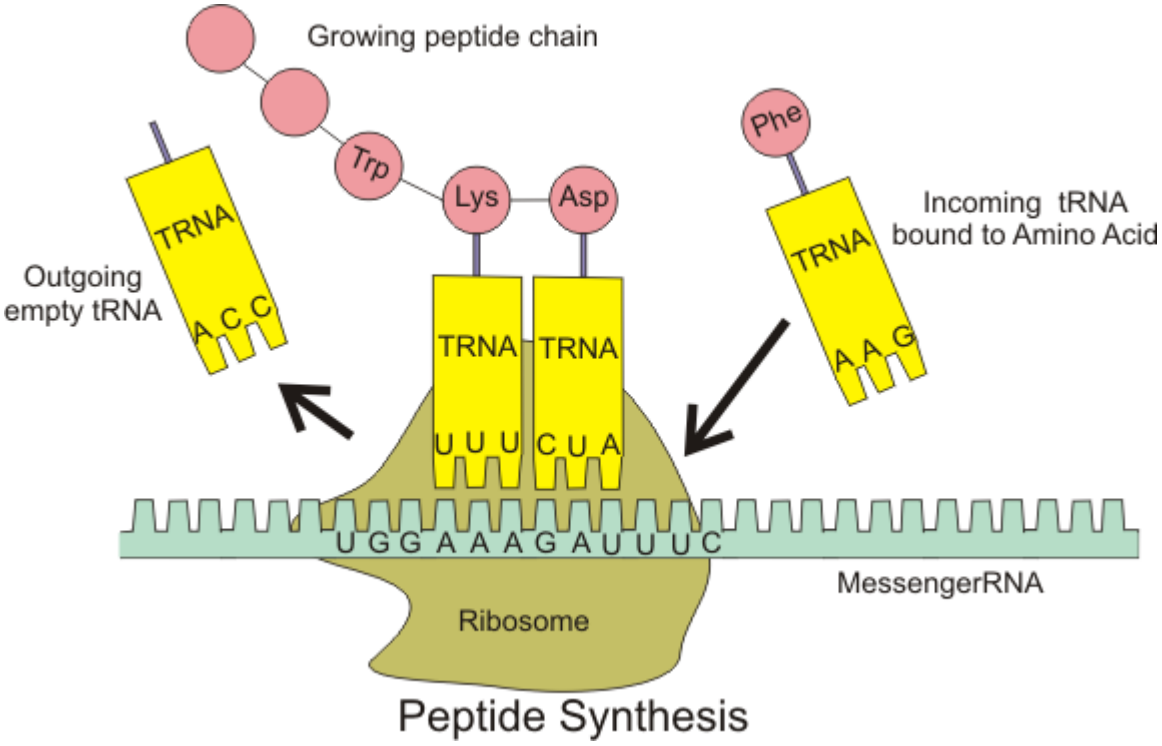
Exocytosis is the extrusion process of proteins and lipids from the cell. These materials are absorbed into secretory granules within the Golgi apparatus before being transported to the cell membrane and secreted into the extracellular environment. In this capacity, vacuoles are simply storage vesicles which allow for the containment, transport and disposal of selected proteins and lipids to the extracellular environment.

Endocytosis is the reverse of exocytosis and can occur in a variety of forms. Phagocytosis ("cell eating") is the process by which bacteria, dead tissue, or other bits of material visible under the microscope are engulfed by cells. The material makes contact with the cell membrane, which then invaginates. The invagination is pinched off, leaving the engulfed material in the membrane-enclosed vacuole and the cell membrane intact. Pinocytosis ("cell drinking") is essentially the same process, the difference being that the substances ingested are in solution and not visible under the microscope. Phagocytosis and Pinocytosis are both undertaken in association with lysosomes which complete the breakdown of the material which has been engulfed.

Chapter- 8

Ribosome and Centrosome

Ribosome



Ribosomes read the sequence of messenger RNAs and assemble proteins out of amino acids bound to transfer RNAs.

Ribosomes are the components of cells that make proteins from all amino acids. One of the central tenets of biology, often referred to as the "central dogma," is that DNA is used to make RNA, which, in turn, is used to make protein. The DNA sequence in genes is copied into a messenger RNA (mRNA). Ribosomes then read the information in this RNA and use it to create proteins. This process is known as translation; i.e., the ribosome "translates" the genetic information from RNA into proteins. Ribosomes do this by

binding to an mRNA and using it as a template for the correct sequence of amino acids in a particular protein. The amino acids are attached to transfer RNA (tRNA) molecules, which enter one part of the ribosome and bind to the messenger RNA sequence. The attached amino acids are then joined together by another part of the ribosome. The ribosome moves along the mRNA, "reading" its sequence and producing a chain of amino acids.

Ribosomes are made from complexes of RNAs and proteins. Ribosomes are divided into two subunits, one larger than the other. The smaller subunit binds to the mRNA, while the larger subunit binds to the tRNA and the amino acids. When a ribosome finishes reading a mRNA, these two subunits split apart. Ribosomes have been classified as ribozymes, since the ribosomal RNA seems to be most important for the peptidyl transferase activity that links amino acids together.

Ribosomes from bacteria, archaea and eukaryotes (the three domains of life on Earth), have significantly different structures and RNA sequences. These differences in structure allow some antibiotics to kill bacteria by inhibiting their ribosomes, while leaving human ribosomes unaffected. The ribosomes in the mitochondria of eukaryotic cells resemble those in bacteria, reflecting the likely evolutionary origin of this organelle. The word ribosome comes from *ribo*nucleic acid and the Greek: *soma* (meaning body).

Description

Archaeal, eubacterial and eukaryotic ribosomes differ in their size, composition and the ratio of protein to RNA. Because they are formed from two subunits of non-equal size, they are slightly longer in the axis than in diameter. **Prokaryotic ribosomes** are around 20 nm (200 ångströms) in diameter and are composed of 65% ribosomal RNA and 35% ribosomal proteins (known as a ribonucleoprotein or RNP). **Eukaryotic ribosomes** are between 25 and 30 nm (250-300 ångströms) in diameter and the ratio of rRNA to protein is close to 1. Ribosomes translate messenger RNA (mRNA) and build polypeptide chains (e.g., proteins) using amino acids delivered by transfer RNA (tRNA). Their active sites are made of RNA, so ribosomes are now classified as "ribozymes".

Ribosomes build proteins from the genetic instructions held within messenger RNA. Free ribosomes are suspended in the cytosol (the semi-fluid portion of the cytoplasm); others are bound to the rough endoplasmic reticulum, giving it the appearance of roughness and thus its name, or to the nuclear envelope. As ribozymes are partly constituted from RNA, it is thought that they might be remnants of the RNA world. Although catalysis of the peptide bond involves the C2 hydroxyl of RNA's P-site adenosine in a protein shuttle mechanism, other steps in protein synthesis (such as translocation) are caused by changes in protein conformations.

Ribosomes are sometimes referred to as organelles, but the use of the term *organelle* is often restricted to describing sub-cellular components that include a phospholipid membrane, which ribosomes, being entirely particulate, do not. For this reason, ribosomes may sometimes be described as "non-membranous organelles".

Ribosomes were first observed in the mid-1950s by Romanian cell biologist George Palade using an electron microscope as dense particles or granules for which he would win the Nobel Prize. The term "ribosome" was proposed by scientist Richard B. Roberts in 1958:

During the course of the symposium a semantic difficulty became apparent. To some of the participants, "microsomes" mean the ribonucleoprotein particles of the microsome fraction contaminated by other protein and lipid material; to others, the microsomes consist of protein and lipid contaminated by particles. The phrase “microsomal particles” does not seem adequate, and “ribonucleoprotein particles of the microsome fraction” is much too awkward. During the meeting, the word "ribosome" was suggested, which has a very satisfactory name and a pleasant sound. The present confusion would be eliminated if “ribosome” were adopted to designate ribonucleoprotein particles in sizes ranging from 35 to 100S.

– Roberts, R. B., Microsomal Particles and Protein Synthesis

The structure and function of the ribosomes and associated molecules, known as the *translational apparatus*, has been of research interest since the mid-twentieth century and is a very active field of study today.

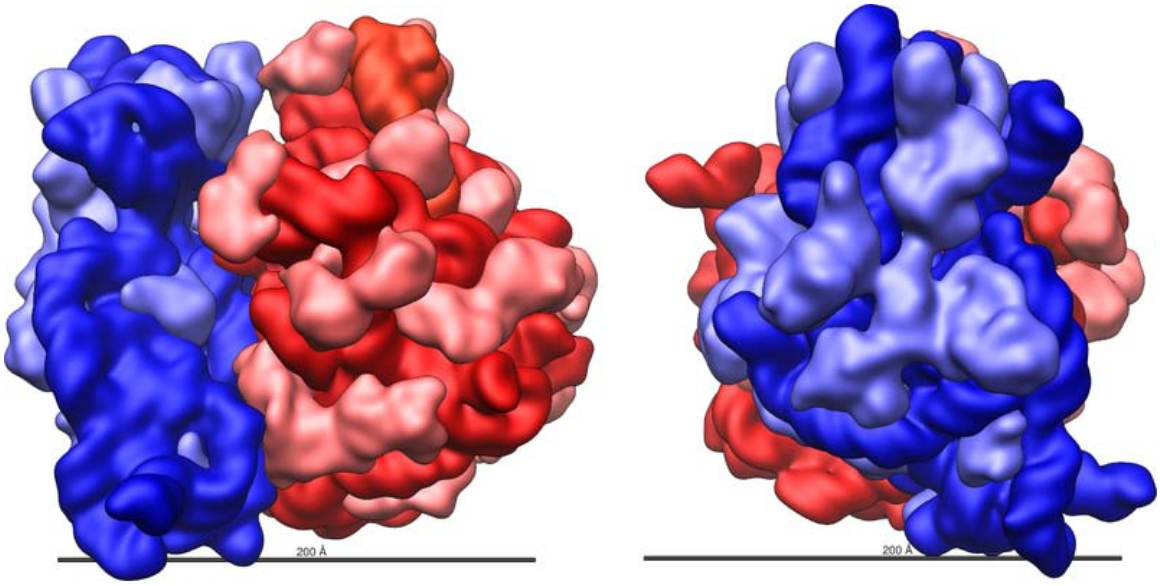


Figure 2 : Large (red) and small (blue) subunit fit together

Ribosomes consist of two subunits (Figure 1) that fit together (Figure 2) and work as one to translate the mRNA into a polypeptide chain during protein synthesis (Figure 3). Bacterial subunits consist of one or two and eukaryotic of one or three very large RNA molecules (known as ribosomal RNA or rRNA) and multiple smaller protein molecules. Crystallographic work has shown that there are no ribosomal proteins close to the reaction site for polypeptide synthesis. This suggests that the protein components of

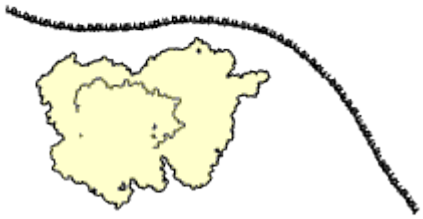
ribosomes act as a scaffold that may enhance the ability of rRNA to synthesize protein rather than directly participating in catalysis (See: Ribozyme).

Biogenesis

In bacterial cells, ribosomes are synthesized in the cytoplasm through the transcription of multiple ribosome gene operons. In eukaryotes, the process takes place both in the cell cytoplasm and in the nucleolus, which is a region within the cell nucleus. The assembly process involves the coordinated function of over 200 proteins in the synthesis and processing of the four rRNAs, as well as assembly of those rRNAs with the ribosomal proteins.

Ribosome locations

Ribosomes are classified as being either "free" or "membrane-bound".



A ribosome translating a protein that is secreted into the endoplasmic reticulum.

Free and membrane-bound ribosomes differ only in their spatial distribution; they are identical in structure. Whether the ribosome exists in a free or membrane-bound state depends on the presence of an ER-targeting signal sequence on the protein being synthesized, so an individual ribosome might be membrane-bound when it is making one protein, but free in the cytosol when it makes another protein.

Free ribosomes

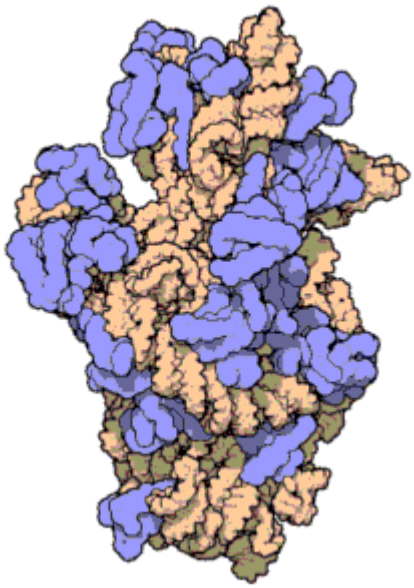
Free ribosomes can move about anywhere in the cytosol, but are excluded from the cell nucleus and other organelles. Proteins that are formed from free ribosomes are released into the cytosol and used within the cell. Since the cytosol contains high concentrations of glutathione and is, therefore, a reducing environment, proteins containing disulfide

bonds, which are formed from oxidized cysteine residues, cannot be produced in this compartment.

Membrane-bound ribosomes

When a ribosome begins to synthesize proteins that are needed in some organelles, the ribosome making this protein can become "membrane-bound". In eukaryotic cells this happens in a region of the endoplasmic reticulum (ER) called the "rough ER". The newly produced polypeptide chains are inserted directly into the ER by the ribosome and are then transported to their destinations, through the secretory pathway. Bound ribosomes usually produce proteins that are used within the plasma membrane or are expelled from the cell via *exocytosis*.

Structure



Atomic structure of the 30S Subunit from *Thermus thermophilus*. Proteins are shown in blue and the single RNA strand in orange.

The ribosomal subunits of prokaryotes and eukaryotes are quite similar.

The unit of measurement is the Svedberg unit, a measure of the rate of sedimentation in centrifugation rather than size and accounts for why fragment names do not add up (70S is made of 50S and 30S).

Prokaryotes have 70S ribosomes, each consisting of a small (30S) and a large (50S) subunit. Their large subunit is composed of a 5S RNA subunit (consisting of 120

nucleotides), a 23S RNA subunit (2900 nucleotides) and 34 proteins. The 30S subunit has a 1540 nucleotide RNA subunit (16S) bound to 21 proteins.

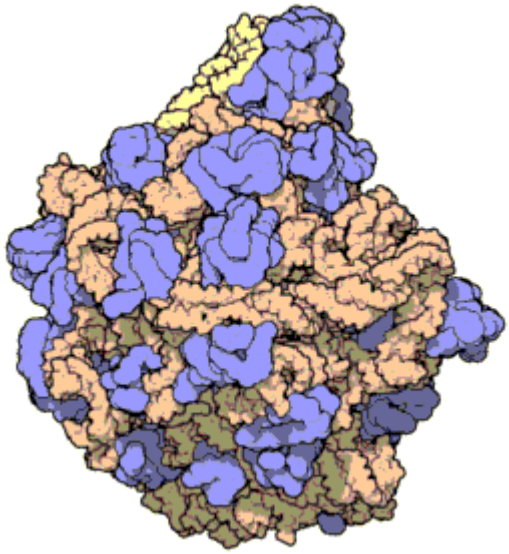
Eukaryotes have 80S ribosomes, each consisting of a small (40S) and large (60S) subunit. Their large subunit is composed of a 5S RNA (120 nucleotides), a 28S RNA (4700 nucleotides), a 5.8S subunit (160 nucleotides) and ~49 proteins. The 40S subunit has a 1900 nucleotide (18S) RNA and ~33 proteins.

The ribosomes found in chloroplasts and mitochondria of eukaryotes also consist of large and small subunits bound together with proteins into one 70S particle. These organelles are believed to be descendants of bacteria and as such their ribosomes are similar to those of bacteria.

The various ribosomes share a core structure, which is quite similar despite the large differences in size. Much of the RNA is highly organized into various tertiary structural motifs, for example pseudoknots that exhibit coaxial stacking. The extra RNA in the larger ribosomes is in several long continuous insertions, such that they form loops out of the core structure without disrupting or changing it. All of the catalytic activity of the ribosome is carried out by the RNA; the proteins reside on the surface and seem to stabilize the structure.

The differences between the bacterial and eukaryotic ribosomes are exploited by pharmaceutical chemists to create antibiotics that can destroy a bacterial infection without harming the cells of the infected person. Due to the differences in their structures, the bacterial 70S ribosomes are vulnerable to these antibiotics while the eukaryotic 80S ribosomes are not. Even though mitochondria possess ribosomes similar to the bacterial ones, mitochondria are not affected by these antibiotics because they are surrounded by a double membrane that does not easily admit these antibiotics into the organelle.

High-resolution structure



Atomic structure of the 50S Subunit from *Haloarcula marismortui*. Proteins are shown in blue and the two RNA strands in orange and yellow. The small patch of green in the center of the subunit is the active site.

The general molecular structure of the ribosome has been known since the early 1970s. In the early 2000s the structure has been achieved at high resolutions, on the order of a few ångströms.

The first papers giving the structure of the ribosome at atomic resolution were published in rapid succession in late 2000. First, the 50S (large prokaryotic) subunit from the archaeon *Haloarcula marismortui* was published. Soon after, the structure of the 30S subunit from *Thermus thermophilus* was published. Shortly thereafter, a more detailed structure was published. These structural studies were awarded the Nobel Prize in Chemistry in 2009. Early the next year (May 2001) these coordinates were used to reconstruct the entire *T. thermophilus* 70S particle at 5.5 ångström resolution.

Two papers were published in November 2005 with structures of the *Escherichia coli* 70S ribosome. The structures of a vacant ribosome were determined at 3.5-ångström resolution using x-ray crystallography. Then, two weeks later, a structure based on cryo-electron microscopy was published, which depicts the ribosome at 11-15 ångström resolution in the act of passing a newly synthesized protein strand into the protein-conducting channel.

First atomic structures of the ribosome complexed with tRNA and mRNA molecules were solved by using X-ray crystallography by two groups independently, at 2.8 ångström and at 3.7 ångström. These structures allow one to see the details of interactions of the *Thermus thermophilus* ribosome with mRNA and with tRNAs bound at classical

ribosomal sites. Interactions of the ribosome with long mRNAs containing Shine-Dalgarno sequences were visualized soon after that at 4.5- to 5.5-ångström resolution.

Function

Ribosomes are the workhorses of protein biosynthesis, the process of translating mRNA into protein. The mRNA comprises a series of codons that dictate to the ribosome the sequence of the amino acids needed to make the protein. Using the mRNA as a template, the ribosome traverses each codon (3 nucleotides) of the mRNA, pairing it with the appropriate amino acid provided by a tRNA. Molecules of transfer RNA (tRNA) contain a complementary anticodon on one end and the appropriate amino acid on the other. The small ribosomal subunit, typically bound to a tRNA containing the amino acid methionine, binds to an AUG codon on the mRNA and recruits the large ribosomal subunit. The ribosome then contains three RNA binding sites, designated A, P and E. The A site binds an aminoacyl-tRNA (a tRNA bound to an amino acid); the P site binds a peptidyl-tRNA (a tRNA bound to the peptide being synthesized); and the E site binds a free tRNA before it exits the ribosome. Protein synthesis begins at a start codon AUG near the 5' end of the mRNA. mRNA binds to the P site of the ribosome first. The ribosome is able to identify the start codon by use of the Shine-Dalgarno sequence of the mRNA in prokaryotes and Kozak box in eukaryotes.

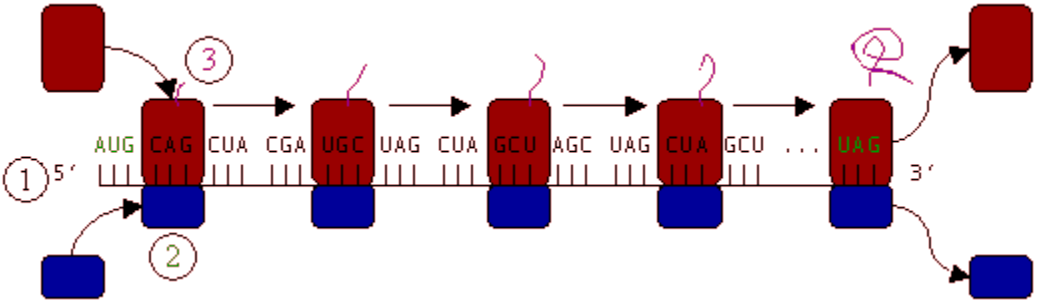
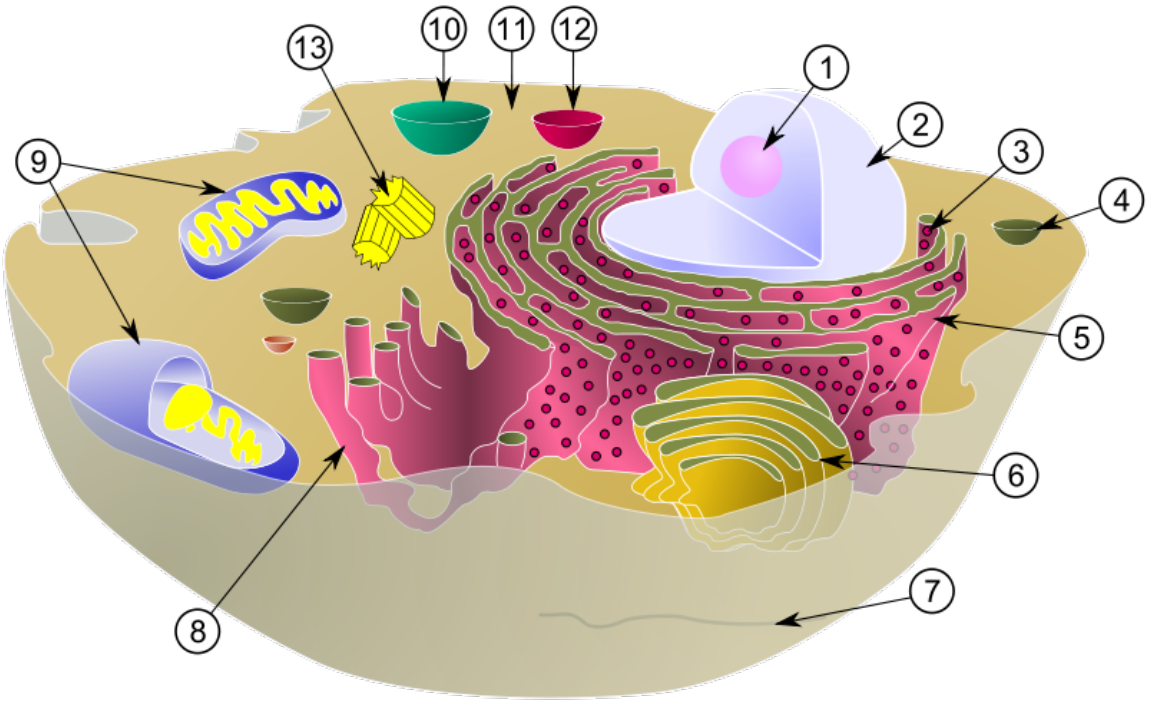


Figure 3 : Translation of mRNA (1) by a ribosome (2)(shown as small and large subunits) into a polypeptide chain (3). The ribosome begins at the start codon of mRNA (AUG) and ends at the stop codon (UAG).

In Figure 3, both ribosomal subunits (small and large) assemble at the start codon (towards the 5' end of the mRNA). The ribosome uses tRNA that matches the current codon (triplet) on the mRNA to append an amino acid to the polypeptide chain. This is done for each triplet on the mRNA, while the ribosome moves towards the 3' end of the mRNA. Usually in bacterial cells, several ribosomes are working parallel on a single mRNA, forming what is called a *polyribosome* or *polysome*.

Centrosome

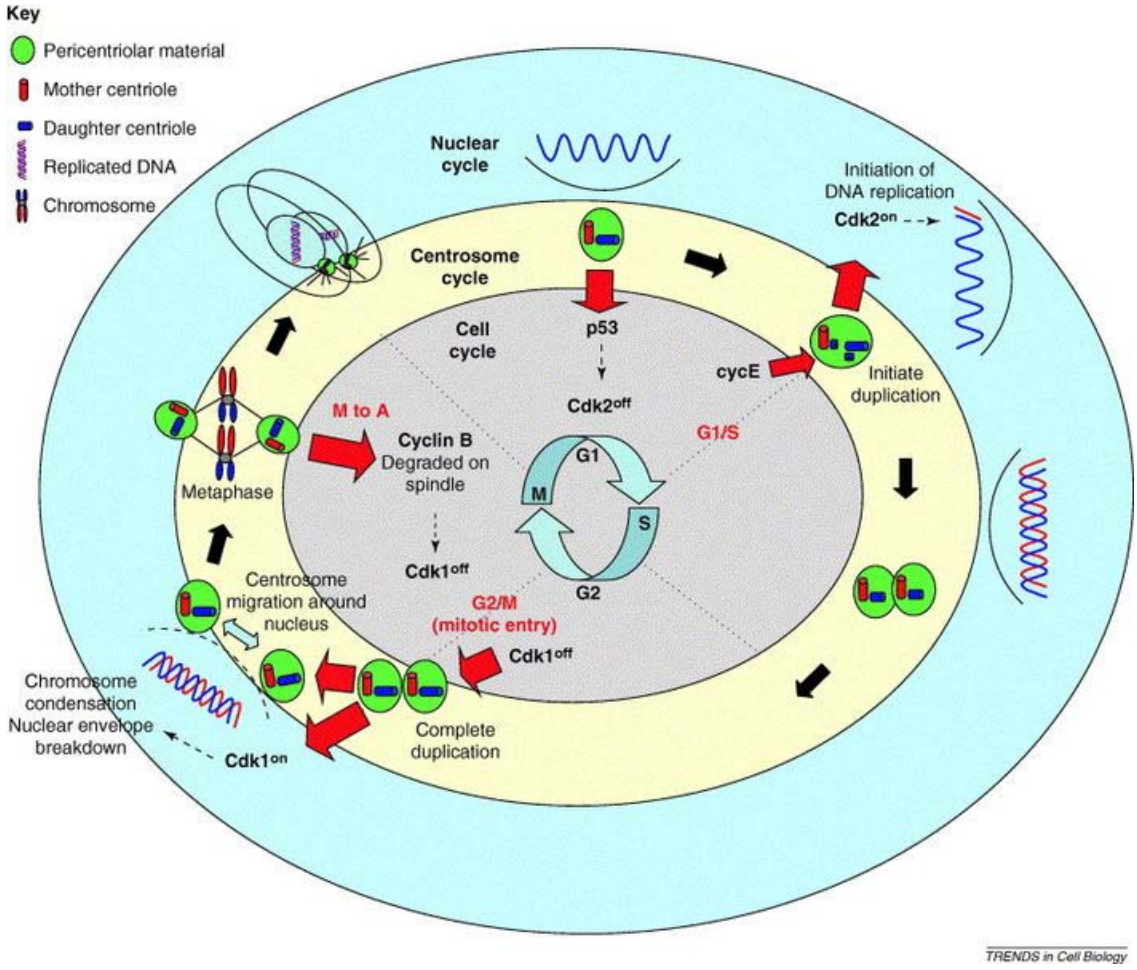


Schematic of typical animal cell, showing subcellular components. Organelles:
(1) Nucleolus
(2) Nucleus
(3) Ribosomes (little dots)
(4) Vesicle
(5) Rough endoplasmic reticulum (ER)
(6) Golgi apparatus
(7) Cytoskeleton
(8) Smooth ER
(9) Mitochondria
(10) Vacuole
(11) Cytoplasm
(12) Lysosome
(13) Centrioles within **Centrosome**

In cell biology, the **centrosome** is an organelle that serves as the main microtubule organizing center (MTOC) of the animal cell as well as a regulator of cell-cycle progression. It was discovered by Edouard Van Beneden in 1883 and was described and named in 1888 by Theodor Boveri. The centrosome is thought to have evolved only in the metazoan lineage of eukaryotic cells. Fungi and plants use other MTOC structures to organize their microtubules. Although the centrosome has a key role in efficient mitosis in animal cells, it is not necessary.

Centrosomes are composed of two orthogonally arranged centrioles surrounded by an amorphous mass of protein termed the pericentriolar material (PCM). The PCM contains proteins responsible for microtubule nucleation and anchoring including γ-tubulin, pericentrin and ninein. In general, each centriole of the centrosome is based on a nine triplet microtubule assembled in a cartwheel structure, and contains centrin, cenexin and tektin.

Roles of the centrosome

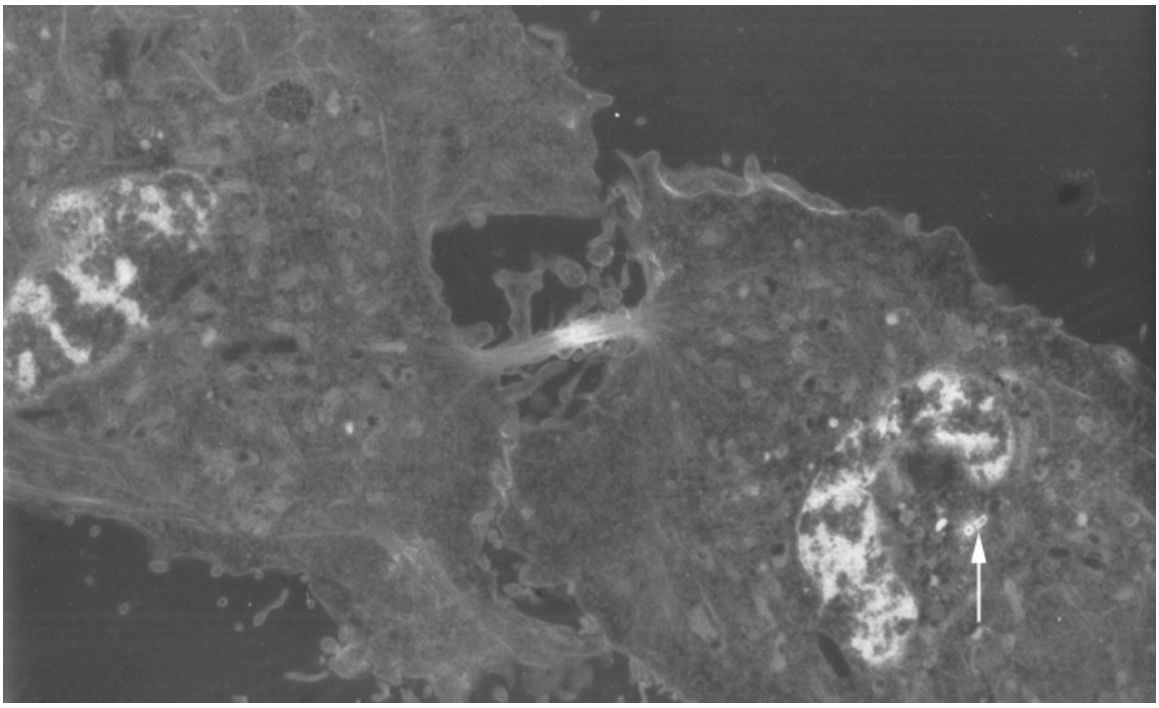


Role of the centrosome in cell cycle progression

Centrosomes are often associated with the nuclear membrane during interphase of the cell cycle. In mitosis the nuclear membrane breaks down and the centrosome nucleated microtubules can interact with the chromosomes to build the mitotic spindle.

The mother centriole, the one that was inherited from the mother cell, also has a central role in making cilia and flagella.

The centrosome is copied only once per cell cycle so that each daughter cell inherits one centrosome, containing two centrioles. The centrosome replicates during the S phase of the cell cycle. During the prophase in the process of cell division called mitosis, the centrosomes migrate to opposite poles of the cell. The mitotic spindle then forms between the two centrosomes. Upon division, each daughter cell receives one centrosome. Aberrant numbers of centrosomes in a cell have been associated with cancer. Doubling of a centrosome is similar to DNA replication in two respects: the semiconservative nature of the process and the action of cdk2 as a regulator of the process. But the processes are essentially different in that centrosome doubling does not occur by template reading and assembly. The mother centriole just aids in the accumulation of materials required for the assembly of the daughter centriole.



Centrosome (shown by arrow) next to nucleus

In animal cells, centrosomes contain two structures called centrioles. Interestingly, centrioles are not required for the progression of mitosis. When the centrioles are irradiated by a laser, mitosis proceeds normally with a morphologically normal spindle. Moreover, development of the fruit fly *Drosophila* is largely normal when centrioles are absent due to a mutation in a gene required for their duplication. In the absence of the centrioles the microtubules of the spindle are focused by motors allowing the formation of a bipolar spindle. Many cells can completely undergo interphase without centrioles. In fact, these structures are not even present in the centrosomes of plant cells.

Unlike centrioles, centrosomes are required for survival of the organism. Acentrosomal cells lack radial arrays of astral microtubules. They are also defective in spindle positioning and in ability to establish a central localization site in cytokinesis. The

function of centrosome in this context is hypothesized to ensure the fidelity of cell division because it greatly increases the efficacy. Some cell types arrest in the following cell cycle when centrosomes are absent. This is not a universal phenomenon.

When the nematode *C. elegans* egg is fertilized the sperm delivers a pair of centrioles. These centrioles will form the centrosomes which will direct the first cell division of the zygote and this will determine its polarity. It is not yet clear whether the role of the centrosome in polarity determination is microtubule dependent or independent.

Evolution of the centrosome

The evolutionary history of the centrosome and the centriole has been traced for some of the signature genes, e.g. the centrins. Centrins participate in calcium signaling and are required for centriole duplication. There exist two main subfamilies of centrins, both of which are present in the early-branching eukaryote Giardia intestinalis. Centrins have therefore been present in the common ancestor of eukaryotes. Conversely, they have no recognizable homologs in archea and bacteria and are thus part of the "eukaryotic signature genes." Although there are studies on the evolution of the centrins and centrioles, no studies have been published on the evolution of the pericentriolar material.

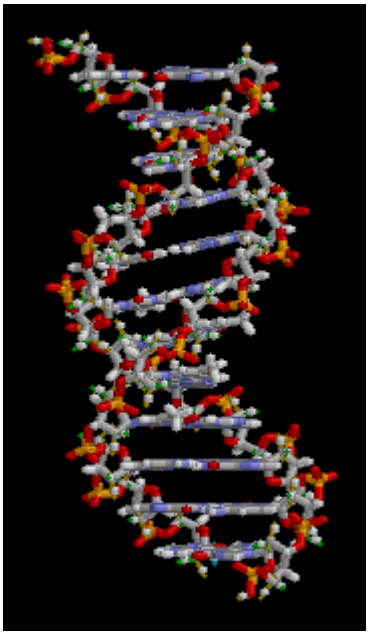
It is evident that some parts of the centrosome are highly diverged in the model species Drosophila melanogaster and Caenorhabditis elegans. For example, both species have lost one of the centrin subfamilies that are usually associated with centriole duplication. Drosophila melanogaster mutants that lack centrosomes can even develop to morphologically normal adult flies, which then die shortly after birth because their sensory neurons lack cilia. Thus, these flies have evolved functionally redundant machinery, which is independent of the centrosomes.

Centrosome associated nucleotides

Research in 2006 indicated that centrosomes from Surf clam eggs contain RNA sequences. The sequences identified were found in "few to no" other places in the cell, and do not appear in existing genome databases. One identified RNA sequence contains a putative RNA polymerase, leading to the hypothesis of an RNA based genome within the centrosome. However, subsequent research has shown that centrosome do not contain their own DNA-based genomes. While it was confirmed that RNA molecules associate with centrosomes, the sequences have still been found within the nucleus. Furthermore, centrosomes can form *de novo* after having been removed (e.g. by laser irradiation) from normal cells.

Chapter- 9

DNA



The structure of part of a DNA double helix

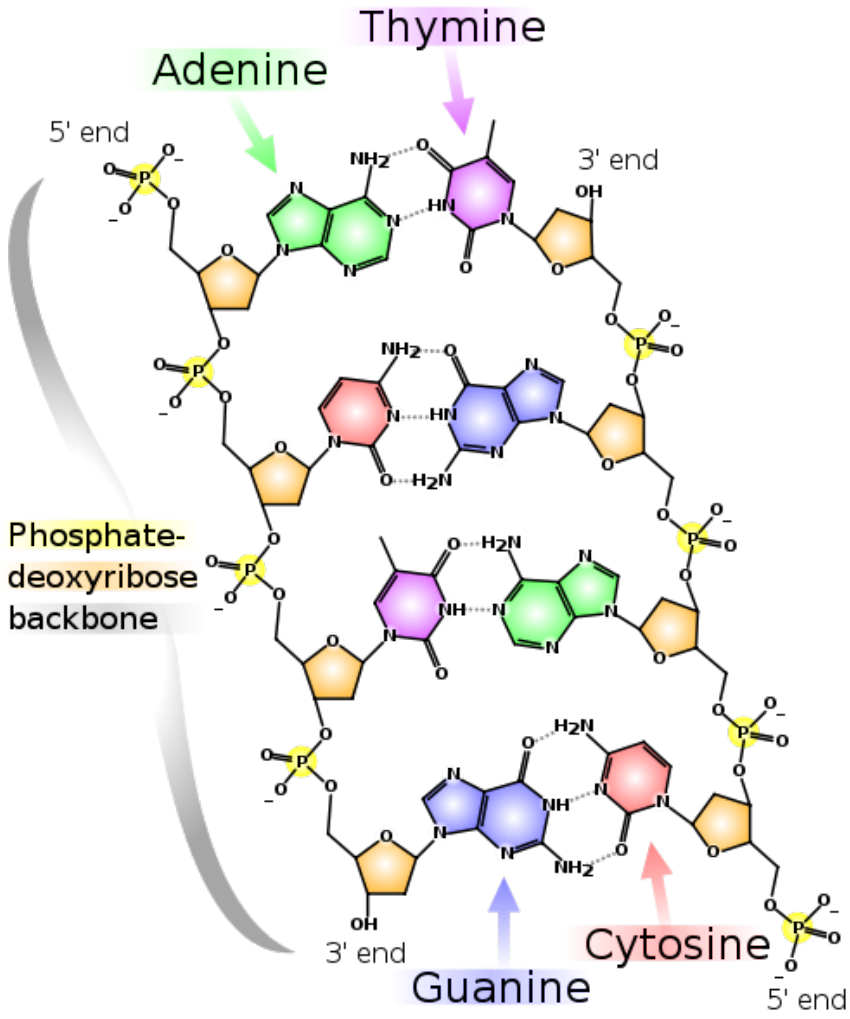
Deoxyribonucleic acid , or **DNA**, is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms (with the exception of RNA viruses). The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set 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

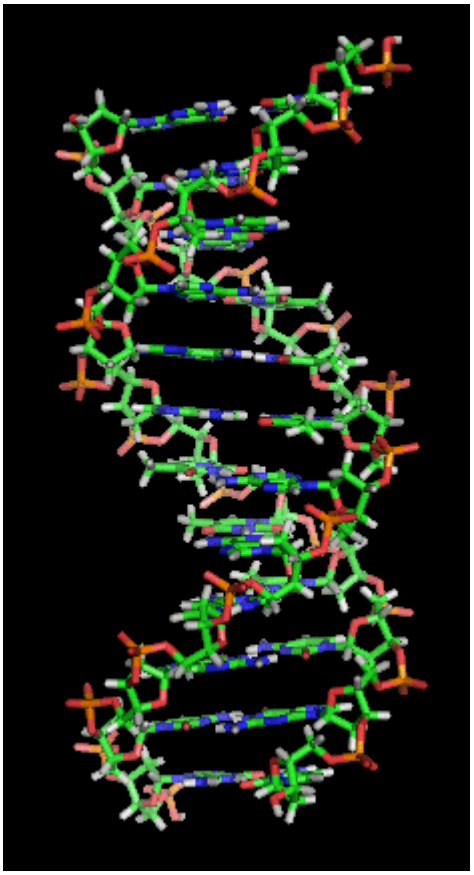


Chemical structure of DNA. Hydrogen bonds shown as dotted lines.

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.

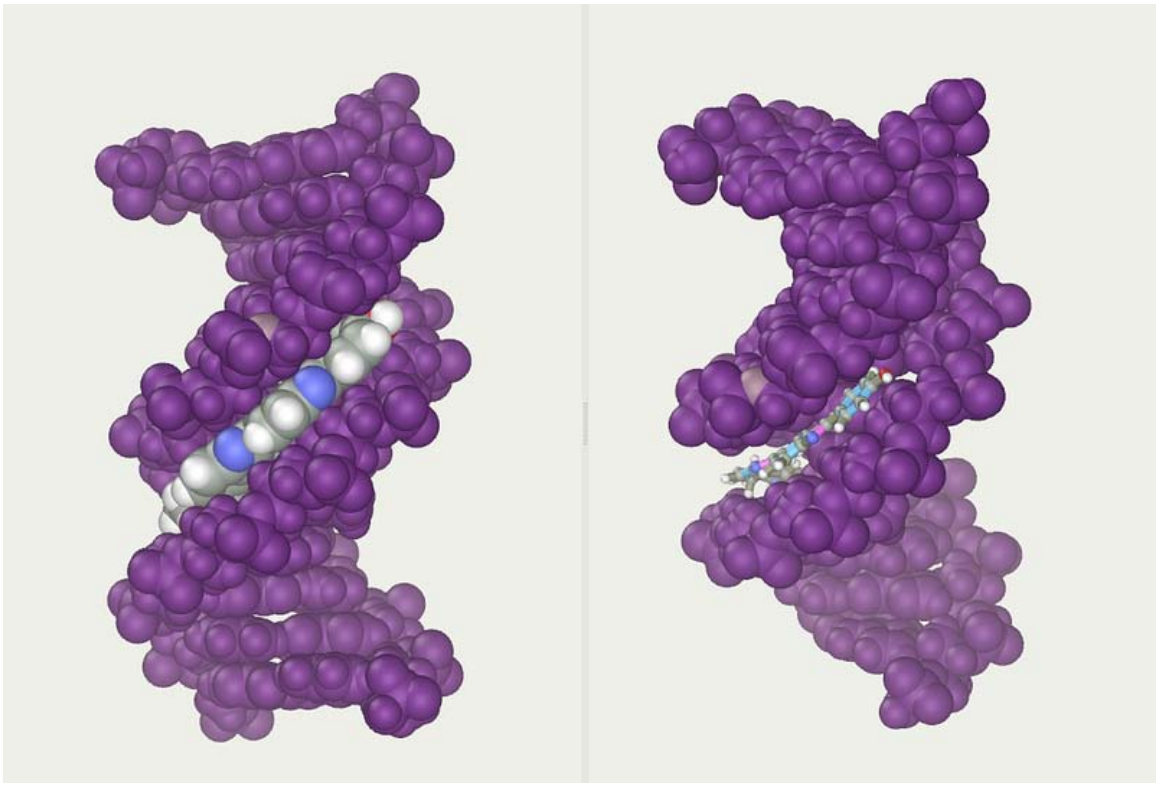
The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar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

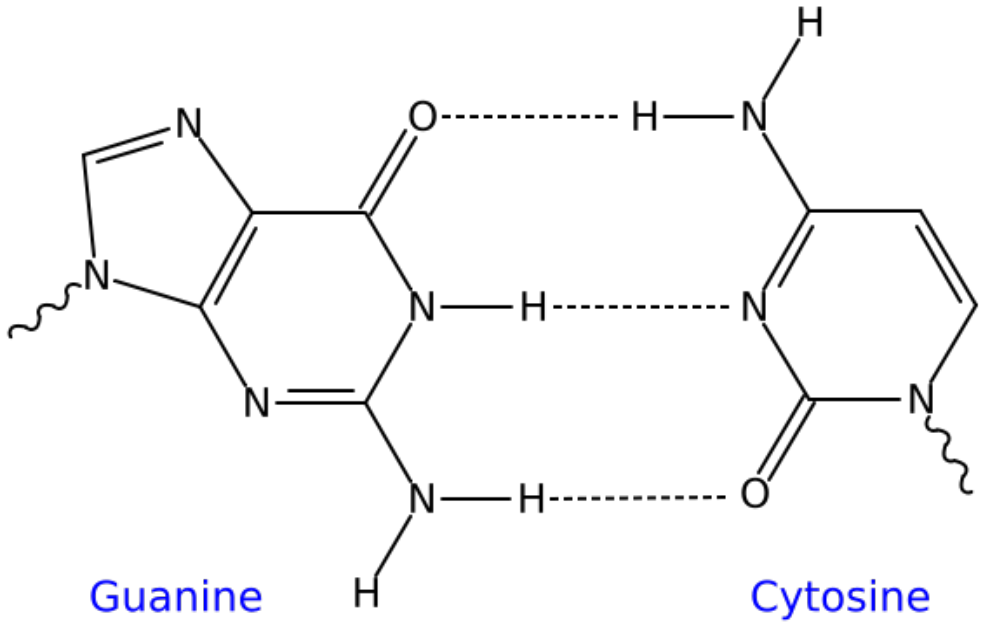
Grooves

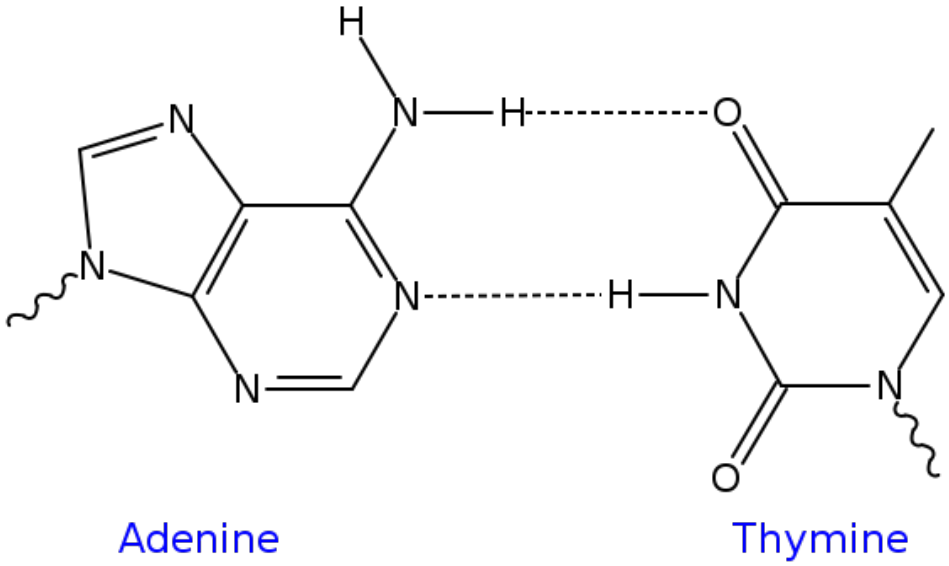
Twin helical strands form the DNA backbone. Another double helix may be found by tracing the spaces, or grooves, between the strands. These voids are adjacent to the base pairs and may provide a binding site. As the strands are not directly opposite each other, the grooves are unequally sized. One groove, 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, 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

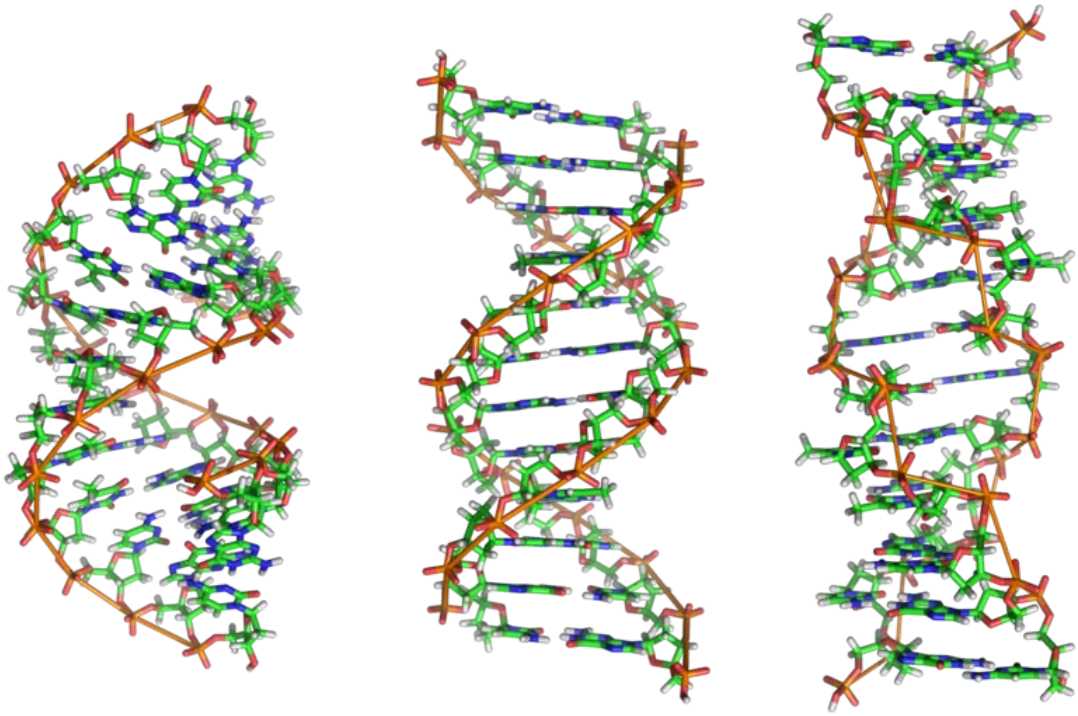
A DNA sequence is called "sense" if its sequence is the same as that of a messenger RNA copy that is translated into protein. The sequence on the opposite strand is called the "antisense" sequence. Both sense and antisense sequences can exist on different parts of the same strand of DNA (i.e. both strands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

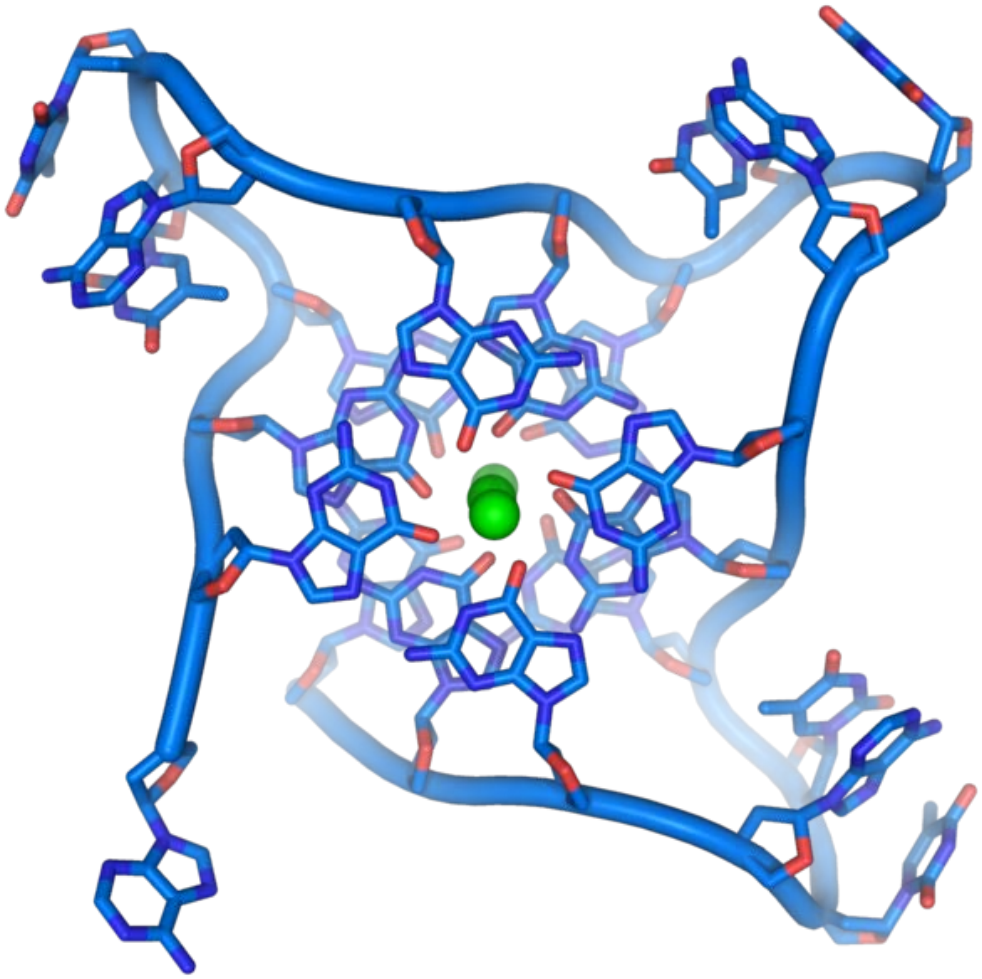
Alternate DNA chemistry

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

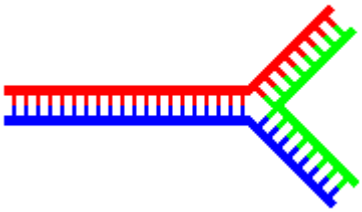
At the ends of the linear chromosomes are specialized regions of DNA called telomeres. The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, as the enzymes that normally replicate DNA cannot copy the extreme 3′ ends of chromosomes. These specialized 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.



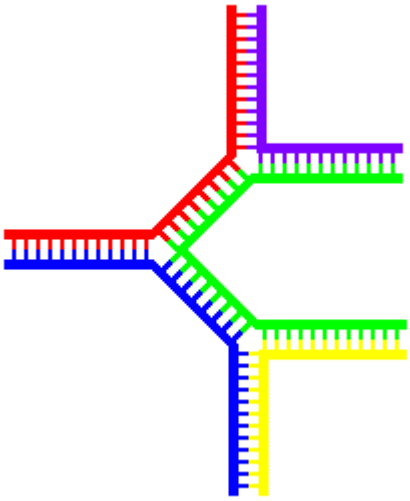
DNA quadruplex formed by telomere repeats. The looped conformation of the DNA backbone is very different from the typical DNA helix.

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.

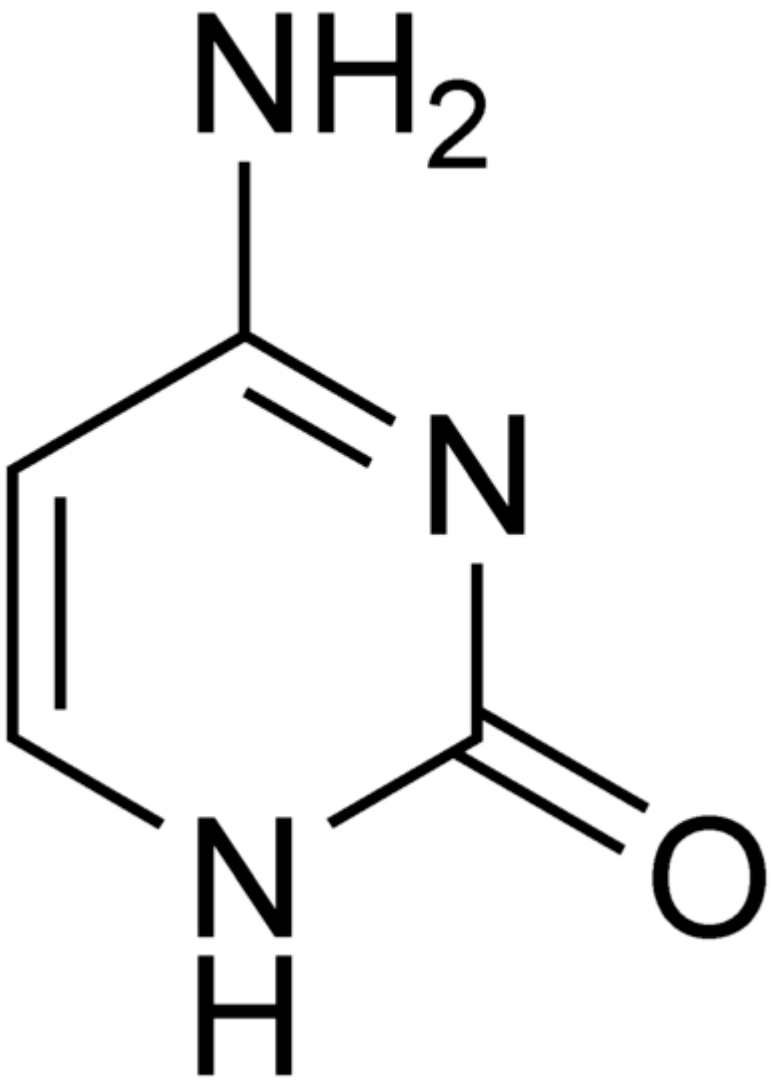
Branched DNA

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.

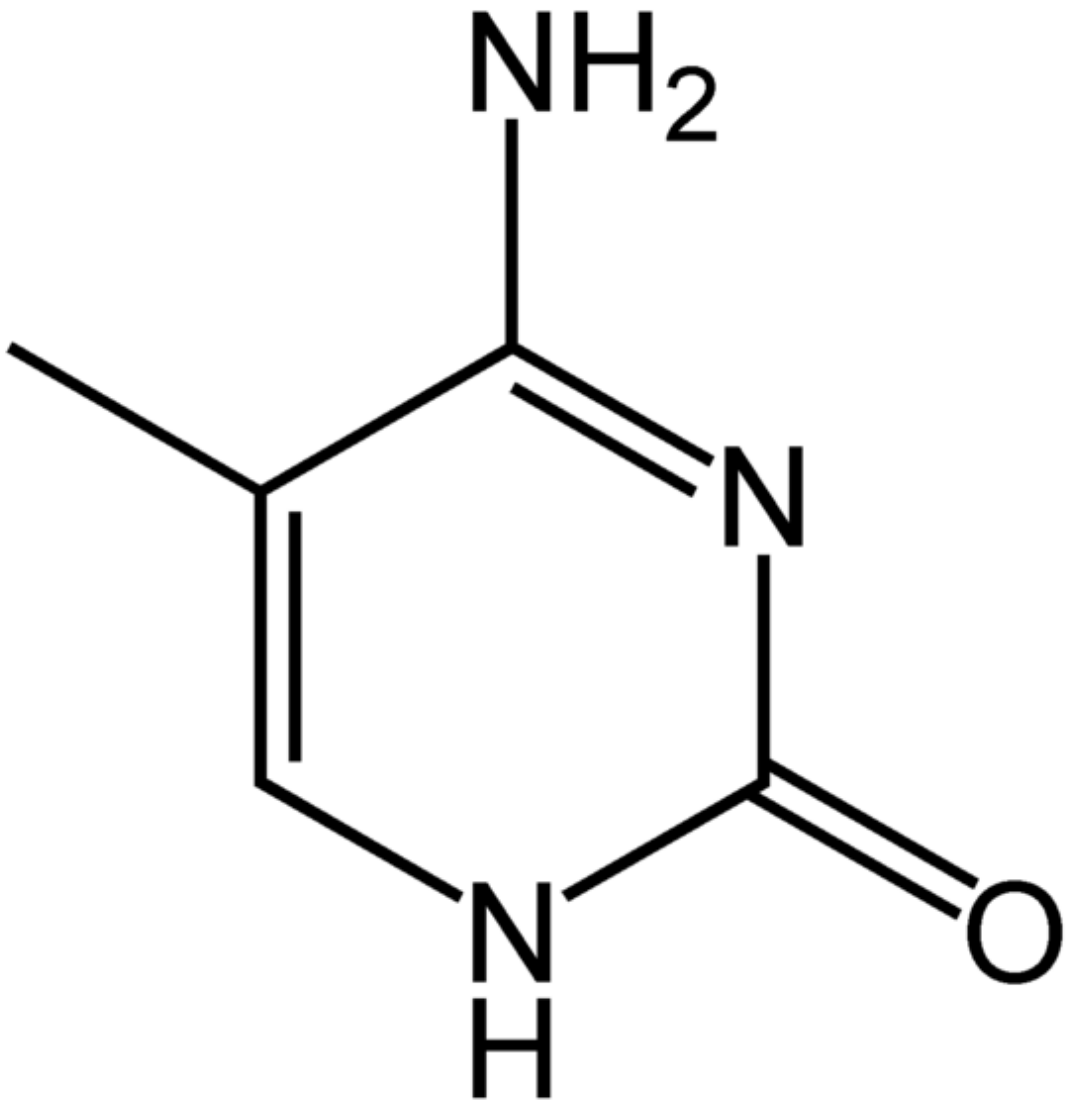
Vibration

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

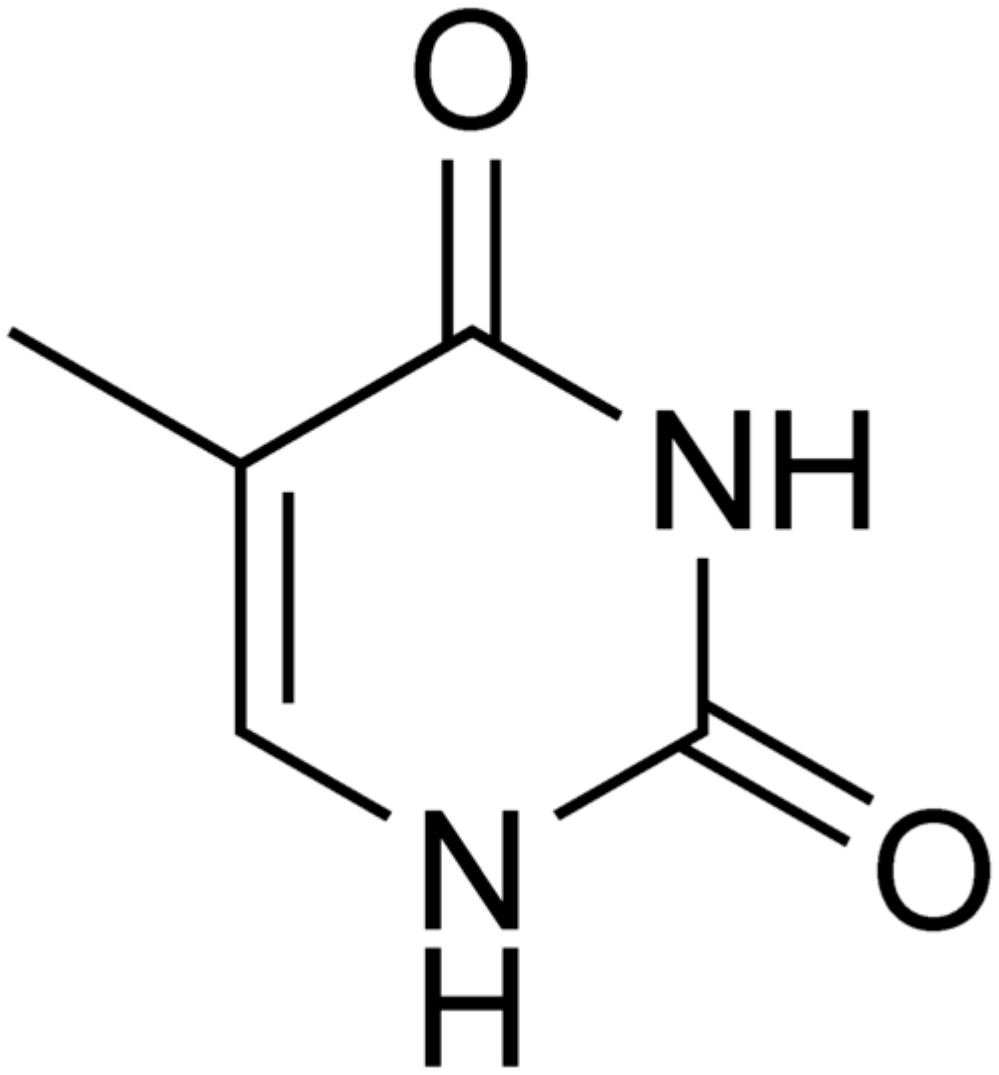
Chemical modifications



cytosine



5-methylcytosine



thymine

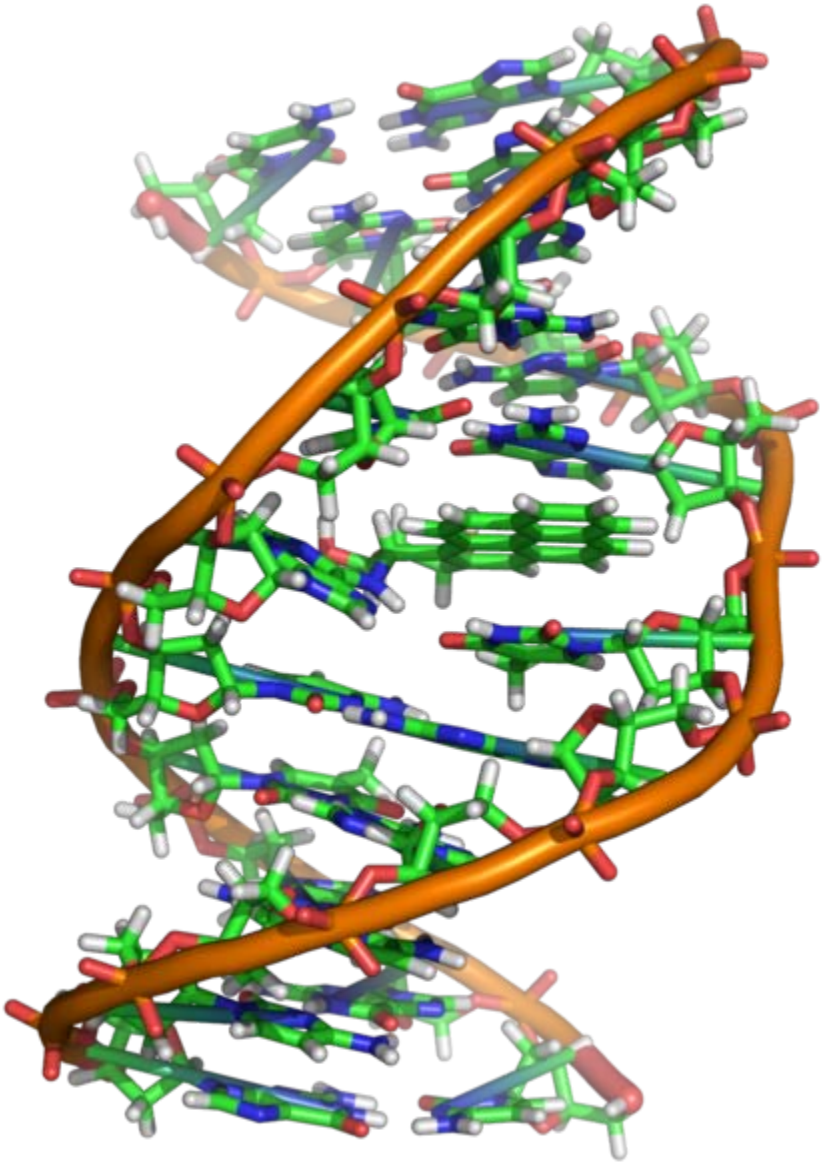
Structure of cytosine with and without the 5-methyl group. Deamination converts 5-methylcytosine into thymine.

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 kinetoplastids.

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

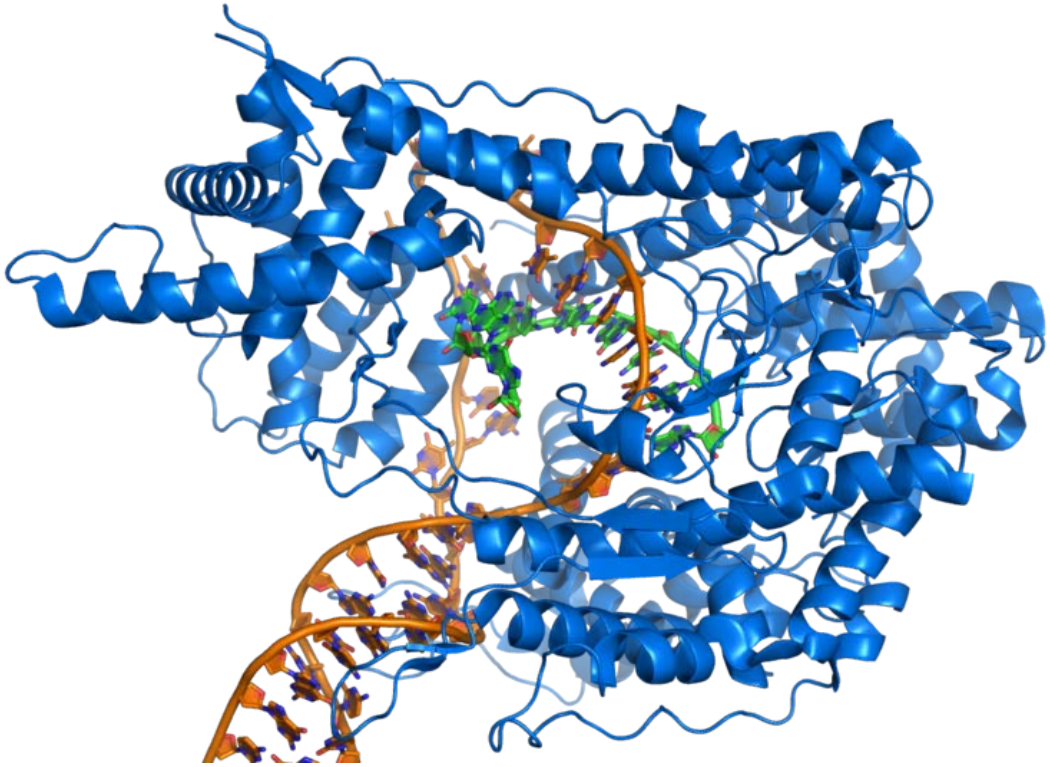
DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes. The information carried by DNA is held in the sequence of 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 non-coding 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.



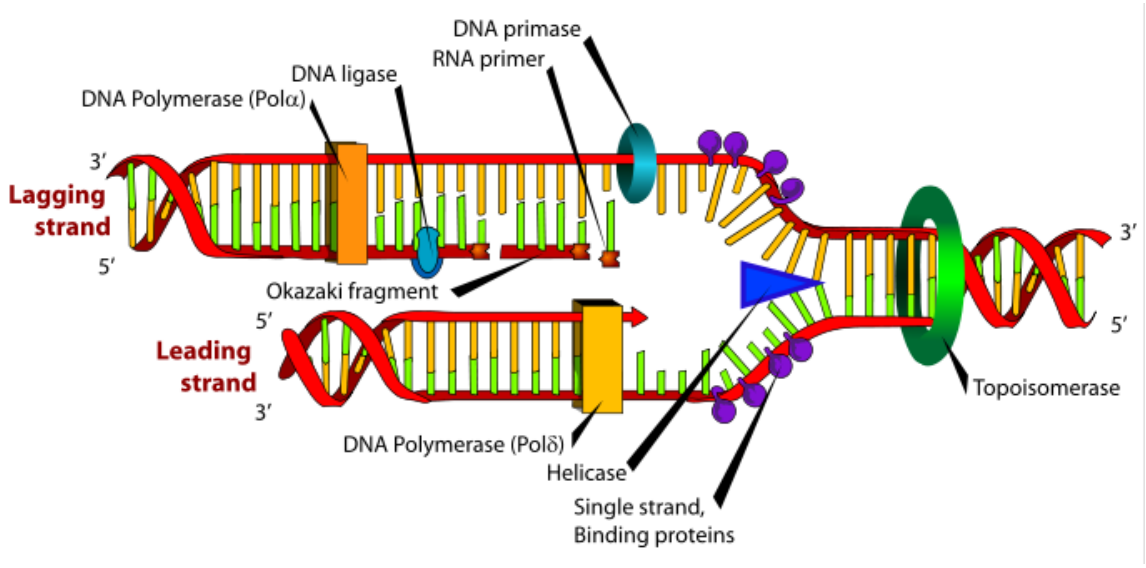
T7 RNA polymerase (blue) producing a mRNA (green) from a DNA template (orange).

Some non-coding 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 non-coding 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

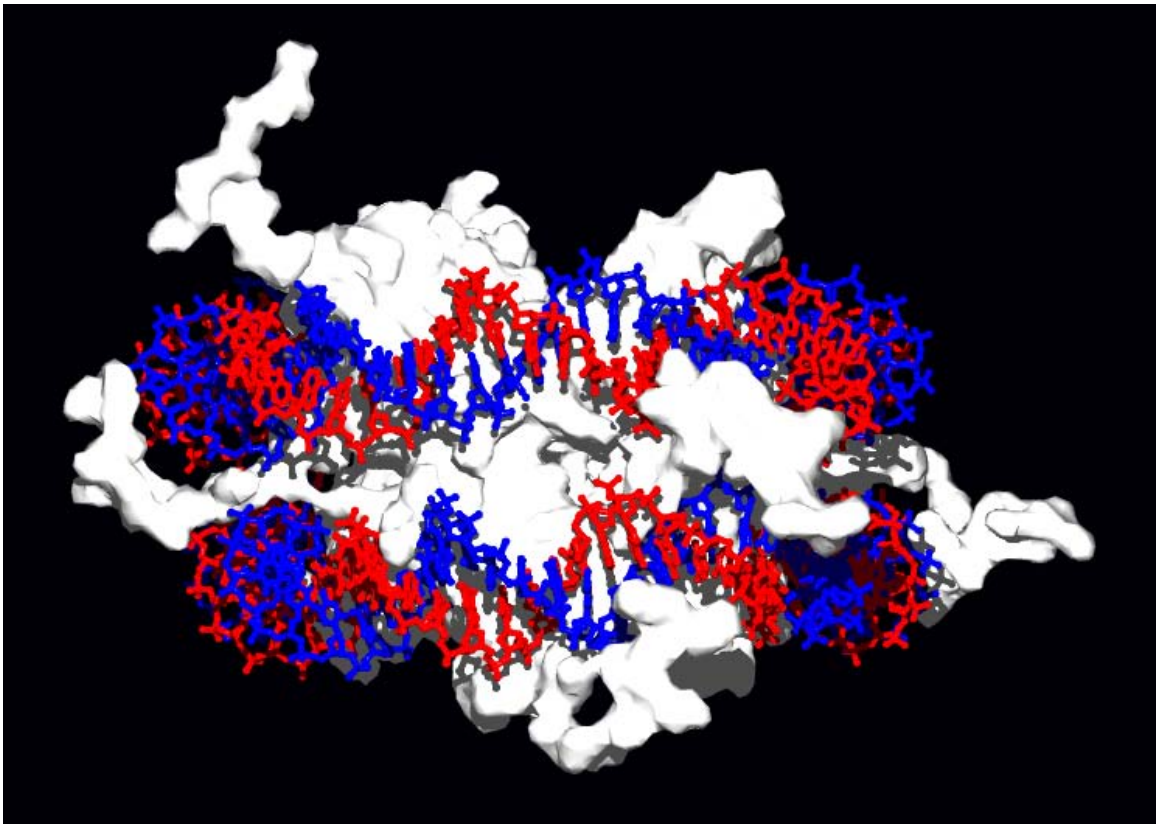
Cell division is essential for an organism to grow, but, when a cell divides, it must replicate the DNA in its genome so that the two daughter cells have the same genetic information as their parent. The double-stranded structure of DNA provides a simple mechanism for DNA replication. Here, the two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA

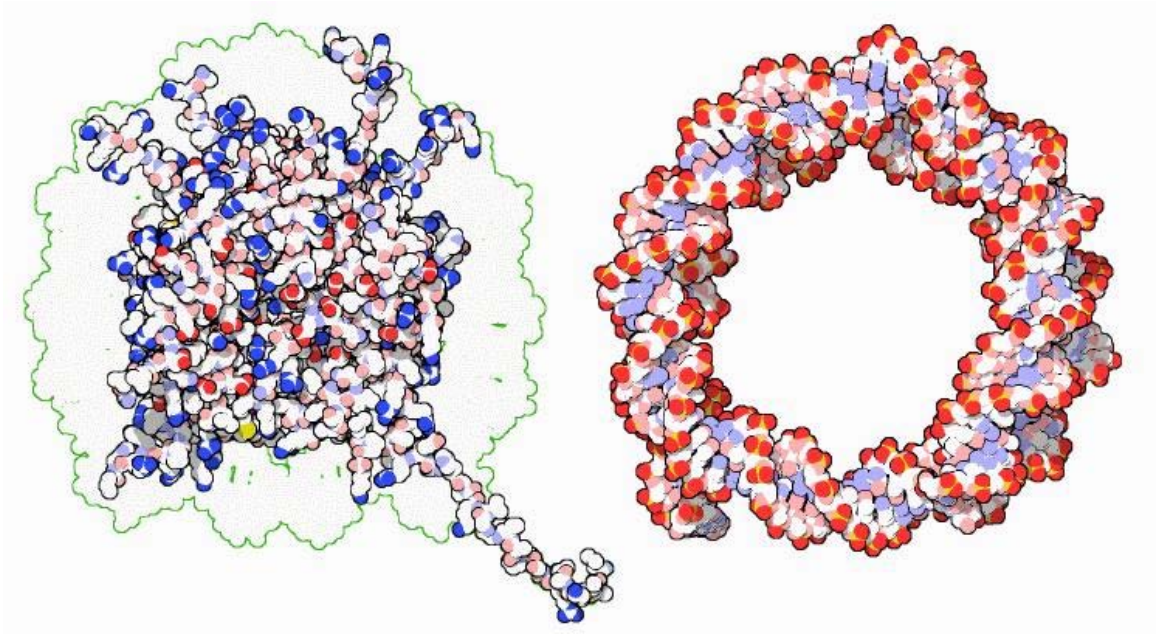
polymerase. This enzyme makes the complementary strand by finding the correct base through complementary base pairing, and bonding it onto the original strand. As DNA polymerases can only extend a DNA strand in a 5′ to 3′ direction, different mechanisms are used to copy the antiparallel strands of the 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

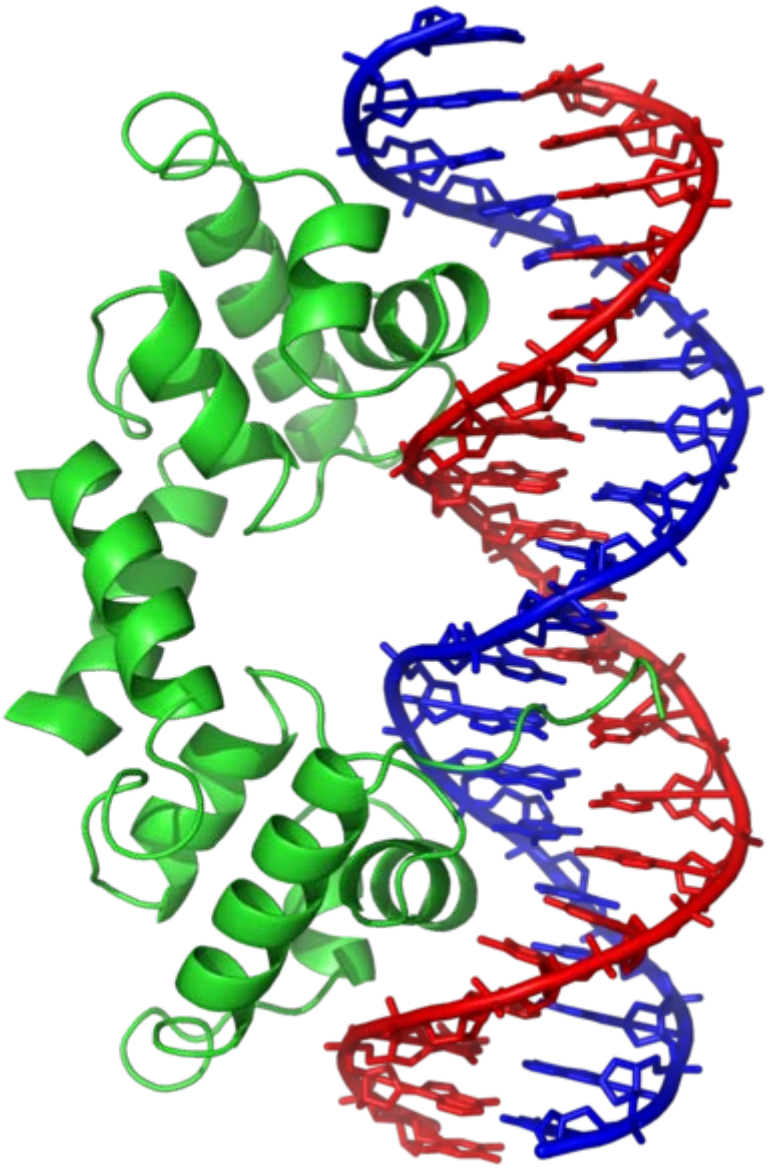




Interaction of DNA with histones (shown in white, top). These proteins' basic amino acids (below left, blue) bind to the acidic phosphate groups on DNA (below right, red).

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.

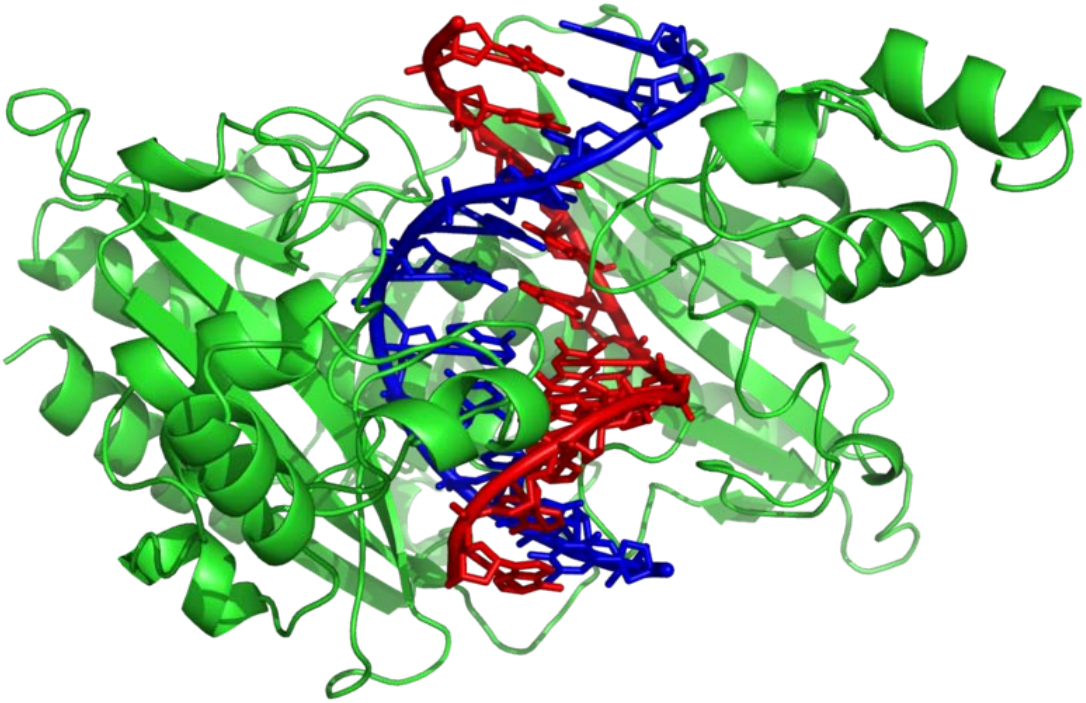
A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA 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.

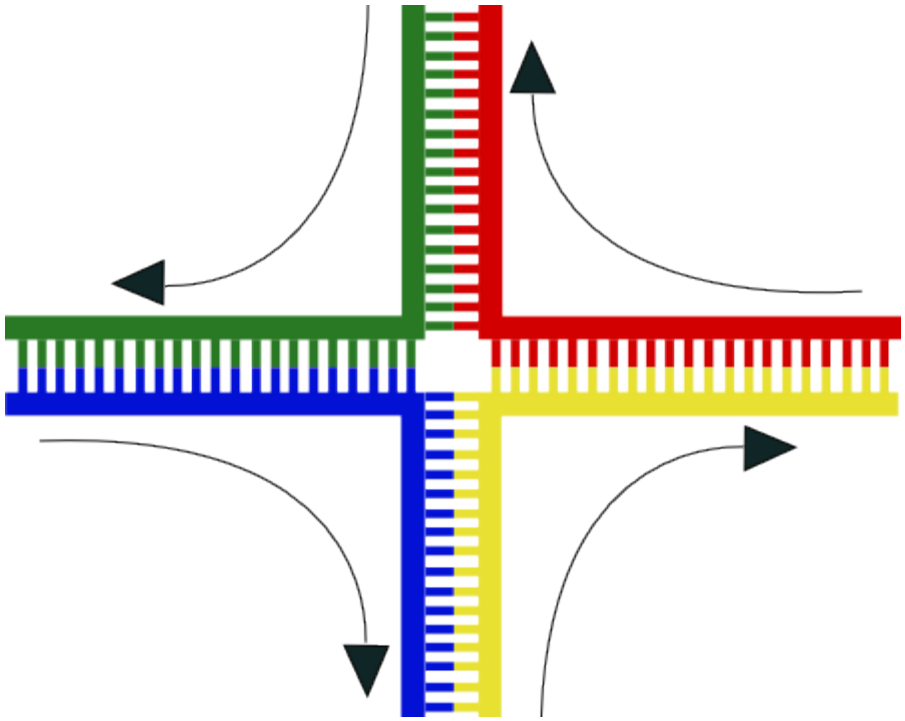
In DNA replication, a DNA-dependent DNA polymerase makes a copy of a DNA sequence. Accuracy is vital in this process, so many of these polymerases have a proofreading activity. Here, the polymerase recognizes the occasional mistakes in the synthesis reaction by the lack of base pairing between the mismatched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

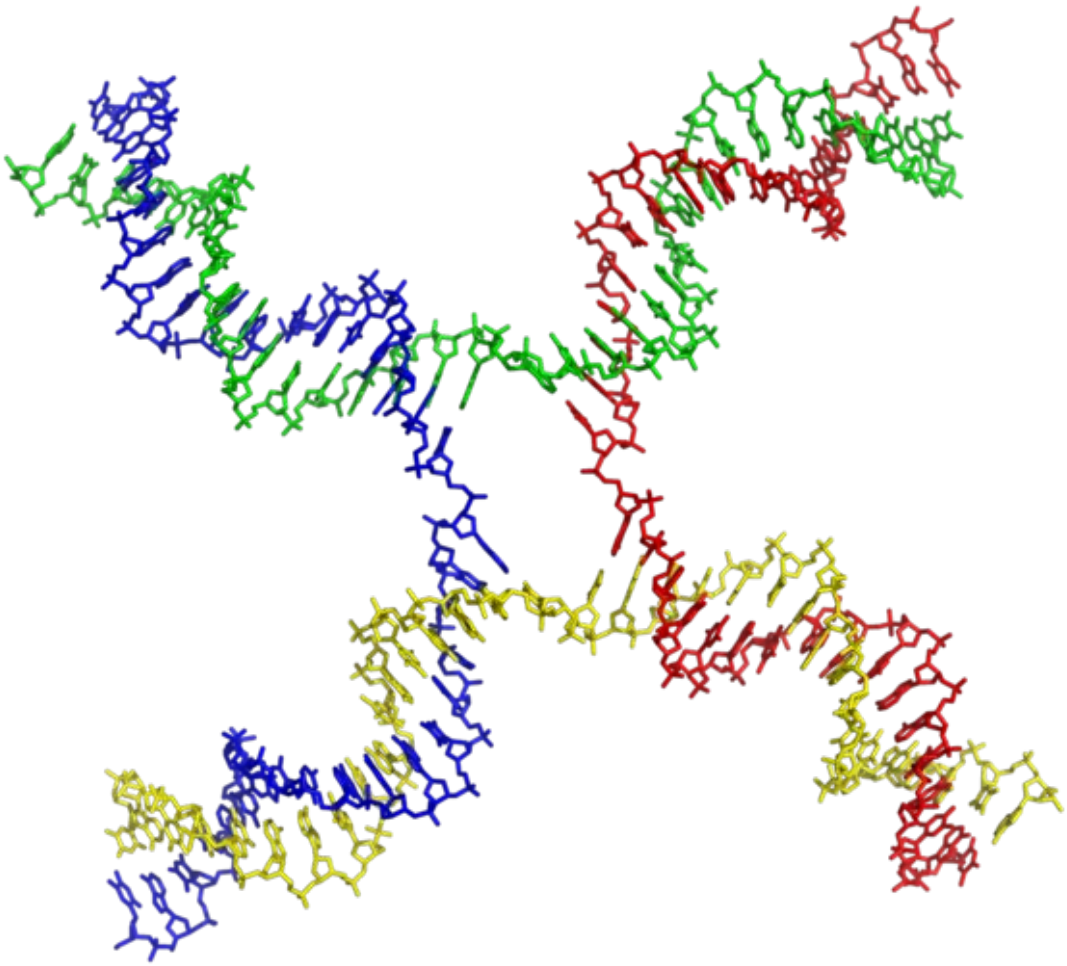
RNA-dependent DNA polymerases are a specialized class of polymerases that copy the sequence of an RNA strand into DNA. They include reverse transcriptase, which is a viral enzyme involved in the infection of cells by retroviruses, and telomerase, which is

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

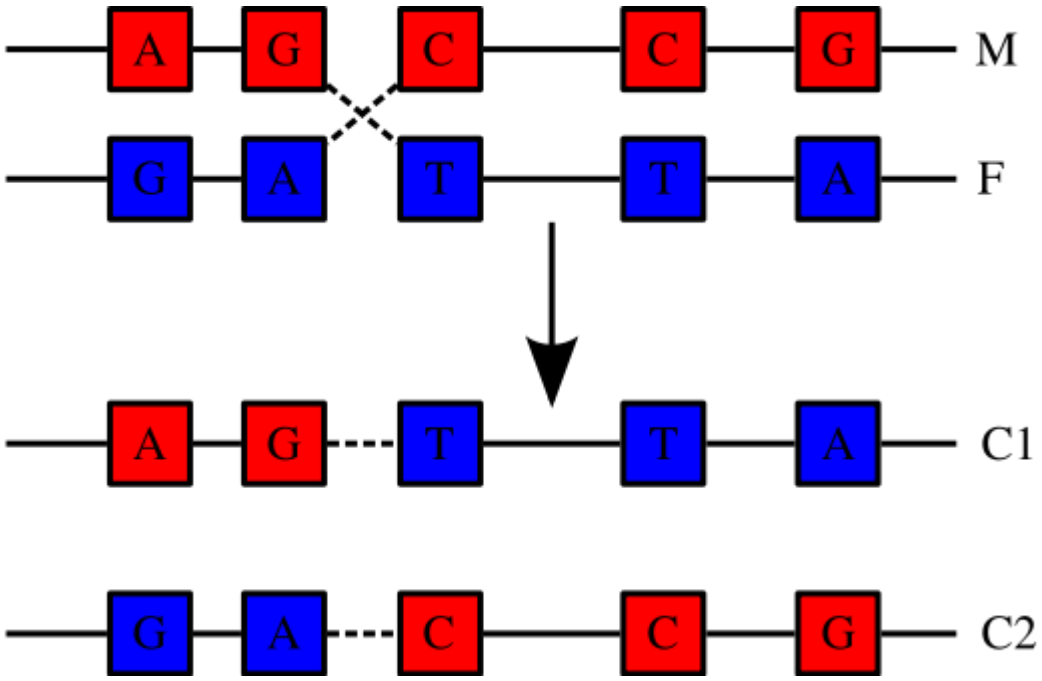
Transcription is carried out by a DNA-dependent RNA polymerase that copies the sequence of a DNA strand into RNA. To begin transcribing a gene, the RNA polymerase binds to a sequence of DNA called a promoter and separates the DNA strands. It then copies the gene sequence into a messenger RNA transcript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

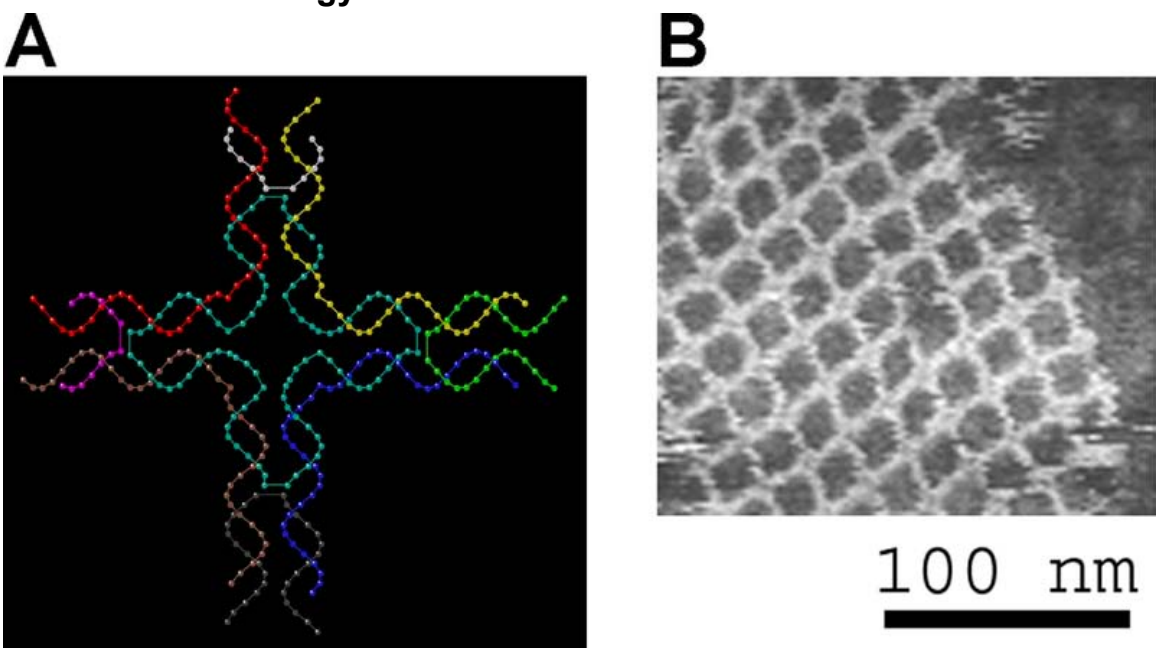
Forensic scientists can use DNA in blood, semen, skin, saliva or hair found at a crime scene to identify a matching DNA of an individual, such as a perpetrator. This process is formally termed DNA profiling, but may also be called "genetic fingerprinting". In DNA profiling, the lengths of variable sections 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



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. Image from Strong, 2004.

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).

Rosalind Franklin, co-creator of the single X-ray diffraction image

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

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.

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.

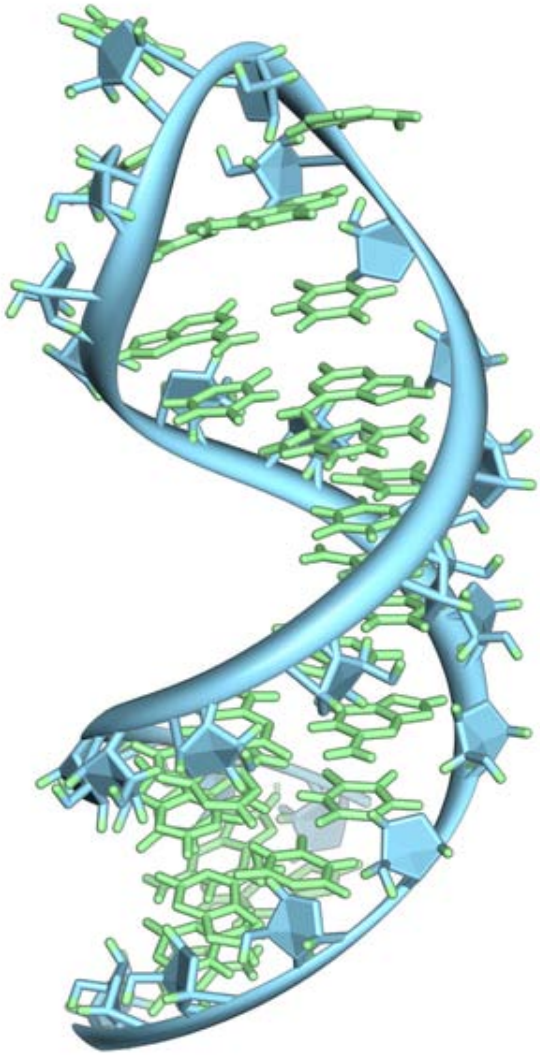
In 1953, James D. Watson and Francis Crick suggested what is now accepted as the first correct double-helix model of DNA structure in the journal *Nature*. Their double-helix, molecular model of DNA was then based on a single X-ray diffraction image (labeled as "Photo 51") taken by Rosalind Franklin 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- 10

RNA



A hairpin loop from a pre-mRNA. Highlighted are the nucleobases (green) and the ribose-phosphate backbone (blue).

Ribonucleic acid (**RNA**) is one of the three major macromolecules (along with DNA and proteins) that are essential for all known forms of life.

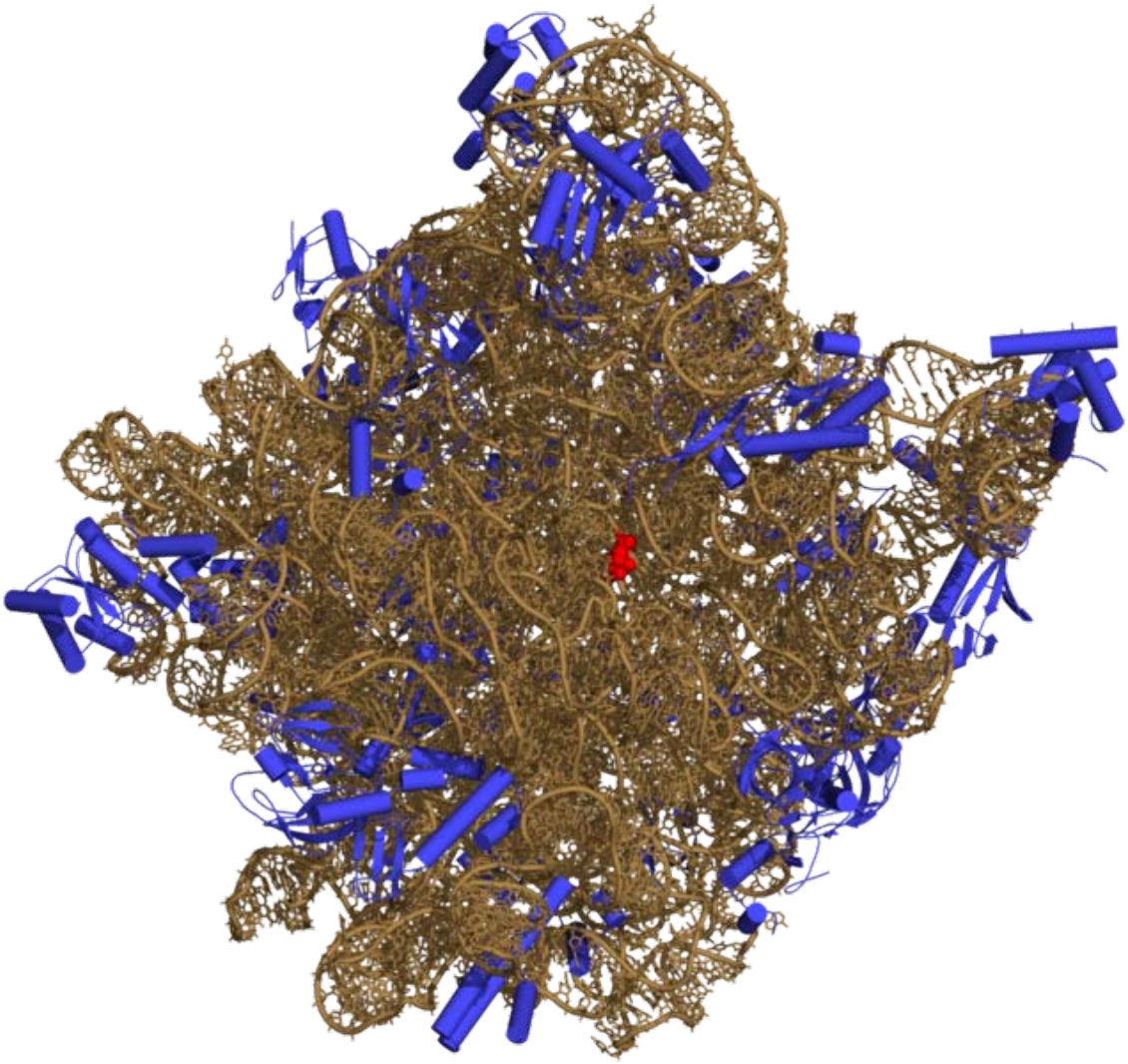
Like DNA, RNA is made up of a long chain of components called nucleotides. Each nucleotide consists of a nucleobase (sometimes called a nitrogenous base), a ribose sugar, and a phosphate group. The sequence of nucleotides allows RNA to encode genetic information. For example, some viruses use RNA instead of DNA as their genetic material, and all organisms use messenger RNA (mRNA) to carry the genetic information that directs the synthesis of proteins.

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

The chemical structure of RNA is very similar to that of DNA, with two differences--(a) RNA contains the sugar ribose while DNA contains the slightly different sugar deoxyribose (a type of ribose that lacks one oxygen atom), and (b) RNA has the nucleobase uracil while DNA contains thymine (uracil and thymine have similar base-pairing properties).

Unlike DNA, most RNA molecules are single-stranded. Single-stranded RNA molecules adopt very complex three-dimensional structures, since they are not restricted to the repetitive double-helical form of double-stranded DNA. RNA is made within living cells by RNA polymerases, enzymes which act to copy a DNA or RNA template into a new RNA strand through processes known as transcription or RNA replication, respectively.

Comparison with DNA



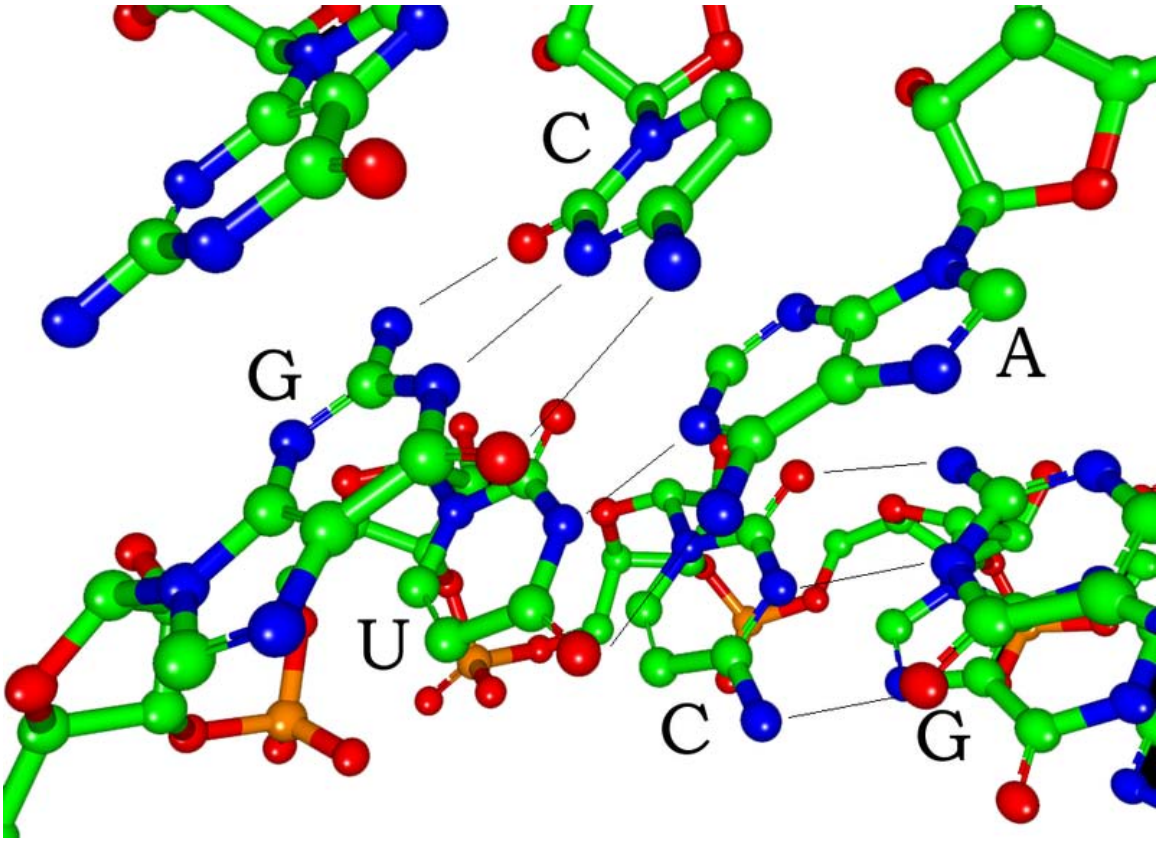
Three-dimensional representation of the 50S ribosomal subunit. RNA is in ochre, protein in blue. The active site is in the middle (red).

RNA and DNA are both nucleic acids, but differ in three main ways. First, unlike DNA which is generally double-stranded, RNA is a single-stranded molecule in many of its biological roles and has a much shorter chain of nucleotides. Second, while DNA contains *deoxyribose*, RNA contains *ribose* (in deoxyribose there is no hydroxyl group attached to the pentose ring in the 2' position). These hydroxyl groups make RNA less stable than DNA because it is more prone to hydrolysis. Third, the complementary base to adenine is not thymine, as it is in DNA, but rather uracil, which is an unmethylated form of thymine.

Like DNA, most biologically active RNAs, including mRNA, tRNA, rRNA, snRNAs and other non-coding RNAs, contain self-complementary sequences that allow parts of the

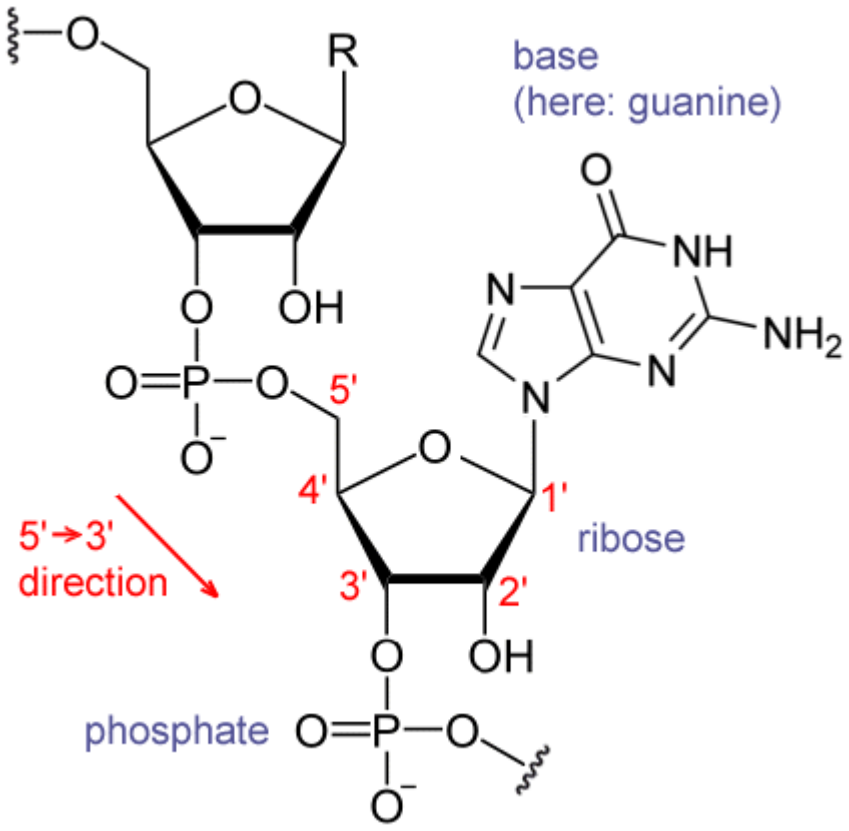
RNA to fold and pair with itself to form double helices. Structural analysis of these RNAs has revealed that they are highly structured. Unlike DNA, their structures do not consist of long double helices but rather collections of short helices packed together into structures akin to proteins. In this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

Structure



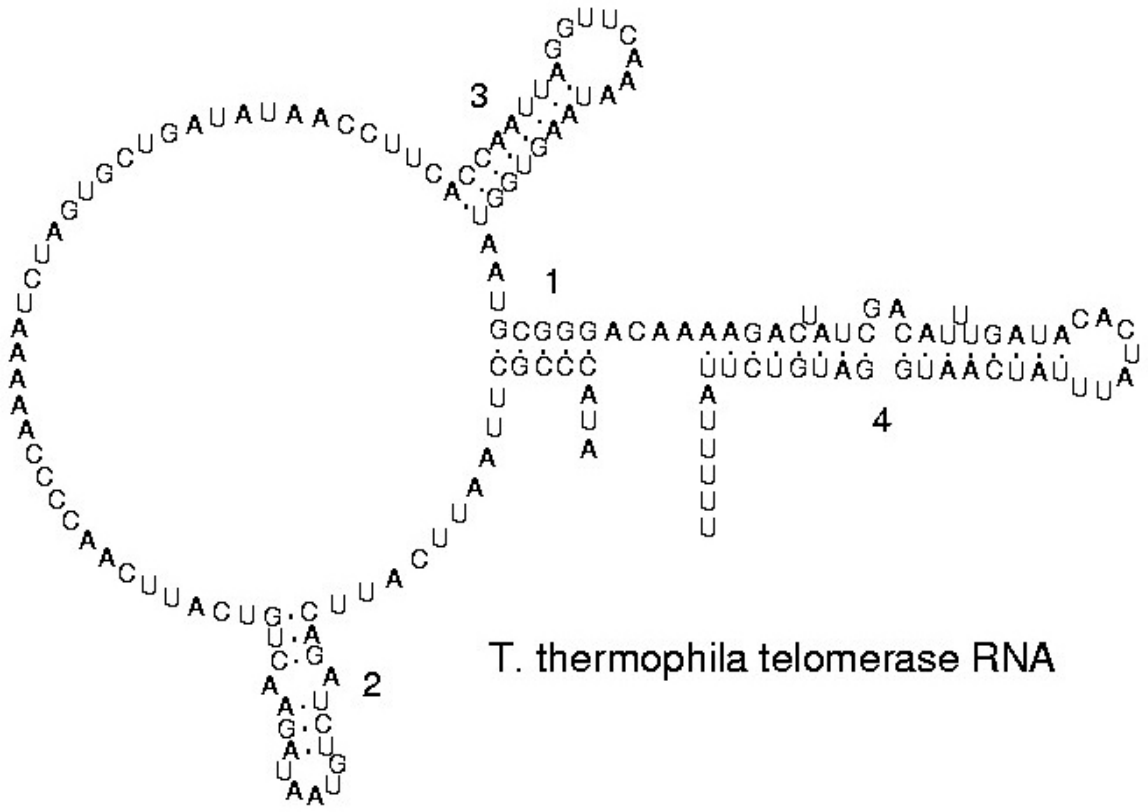
Watson-Crick base pairs in a siRNA (hydrogen atoms are not shown)

Each nucleotide in RNA contains a ribose sugar, with carbons numbered 1' through 5'. A base is attached to the 1' position, generally adenine (A), cytosine (C), guanine (G) or uracil (U). Adenine and guanine are purines, cytosine and uracil are pyrimidines. A phosphate group is attached to the 3' position of one ribose and the 5' position of the next. The phosphate groups have a negative charge each at physiological pH, making RNA a charged molecule (polyanion). The bases may form hydrogen bonds between cytosine and guanine, between adenine and uracil and between guanine and uracil. However other interactions are possible, such as a group of adenine bases binding to each other in a bulge, or the GNRA tetraloop that has a guanine–adenine base-pair.



Chemical structure of RNA

An important structural feature of RNA that distinguishes it from DNA is the presence of a hydroxyl group at the 2' position of the ribose sugar. The presence of this functional group causes the helix to adopt the A-form geometry rather than the B-form most commonly observed in DNA. This results in a very deep and narrow major groove and a shallow and wide minor groove. A second consequence of the presence of the 2'-hydroxyl group is that in conformationally flexible regions of an RNA molecule (that is, not involved in formation of a double helix), it can chemically attack the adjacent phosphodiester bond to cleave the backbone.



Secondary structure of a telomerase RNA.

RNA is transcribed with only four bases (adenine, cytosine, guanine and uracil), but these bases and attached sugars can be modified in numerous ways as the RNAs mature. Pseudouridine (Ψ), in which the linkage between uracil and ribose is changed from a C–N bond to a C–C bond, and ribothymidine (T), are found in various places (most notably in the TΨC loop of tRNA). Another notable modified base is hypoxanthine, a deaminated adenine base whose nucleoside is called inosine (I). Inosine plays a key role in the wobble hypothesis of the genetic code.

There are nearly 100 other naturally occurring modified nucleosides, of which pseudouridine and nucleosides with 2'-O-methylribose are the most common. The specific roles of many of these modifications in RNA are not fully understood. However, it is notable that in ribosomal RNA, many of the post-transcriptional modifications occur in highly functional regions, such as the peptidyl transferase center and the subunit interface, implying that they are important for normal function.

The functional form of single stranded RNA molecules, just like proteins, frequently requires a specific tertiary structure. The scaffold for this structure is provided by secondary structural elements which are hydrogen bonds within the molecule. This leads to several recognizable "domains" of secondary structure like hairpin loops, bulges and

internal loops. Since RNA is charged, metal ions such as Mg^{2+} are needed to stabilise many secondary and tertiary structures.

Synthesis

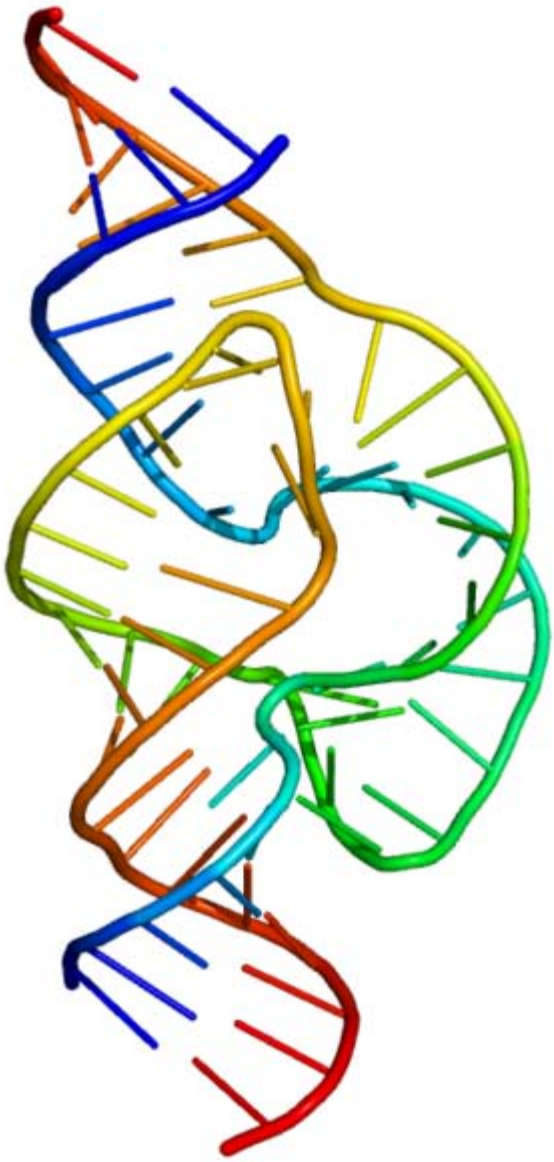
Synthesis of RNA is usually catalyzed by an enzyme—RNA polymerase—using DNA as a template, a process known as transcription. Initiation of transcription begins with the binding of the enzyme to a promoter sequence in the DNA (usually found "upstream" of a gene). The DNA double helix is unwound by the helicase activity of the enzyme. The enzyme then progresses along the template strand in the 3’ to 5’ direction, synthesizing a complementary RNA molecule with elongation occurring in the 5’ to 3’ direction. The DNA sequence also dictates where termination of RNA synthesis will occur.

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

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

Types of RNA

Overview



Structure of a hammerhead ribozyme, a ribozyme that cuts RNA

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

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

In translation

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

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

Ribosomal RNA (rRNA) is the catalytic component of the ribosomes. Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA. Three of the rRNA molecules are synthesized in the nucleolus, and one is synthesized elsewhere. In the cytoplasm, ribosomal RNA and protein combine to form a nucleoprotein called a ribosome. The ribosome binds mRNA and carries out protein synthesis. Several ribosomes may be attached to a single mRNA at any time. rRNA is extremely abundant and makes up 80% of the 10 mg/ml RNA found in a typical eukaryotic cytoplasm.

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

Regulatory RNAs

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

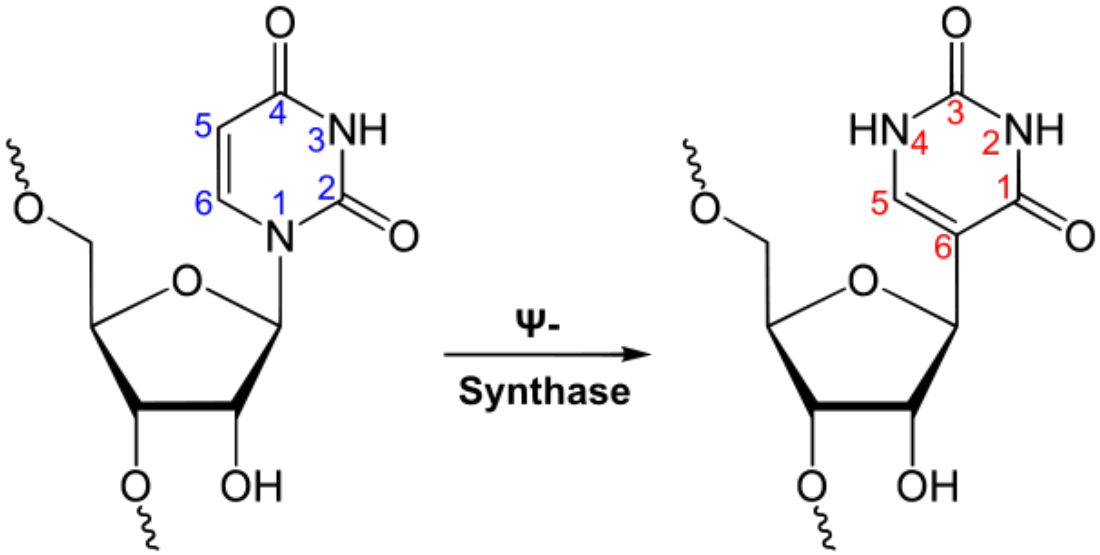
interfering RNAs (siRNA; 20-25 nt) are often produced by breakdown of viral RNA, there are also endogenous sources of siRNAs.

siRNAs act through RNA interference in a fashion similar to miRNAs. Some miRNAs and siRNAs can cause genes they target to be methylated, thereby decreasing or increasing transcription of those genes. Animals have Piwi-interacting RNAs (piRNA; 29-30 nt) which are active in germline cells and are thought to be a defense against transposons and play a role in gametogenesis.

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

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

In RNA processing



Uridine to pseudouridine is a common RNA modification.

Many RNAs are involved in modifying other RNAs. Introns are spliced out of pre-mRNA by spliceosomes, which contain several small nuclear RNAs (snRNA), or the introns can be ribozymes that are spliced by themselves. RNA can also be altered by having its nucleotides modified to other nucleotides than A, C, G and U. In eukaryotes, modifications of RNA nucleotides are generally directed by small nucleolar RNAs

(snoRNA; 60-300 nt), found in the nucleolus and cajal bodies. snoRNAs associate with enzymes and guide them to a spot on an RNA by basepairing to that RNA. These enzymes then perform the nucleotide modification. rRNAs and tRNAs are extensively modified, but snRNAs and mRNAs can also be the target of base modification.

RNA genomes

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

In reverse transcription

Reverse transcribing viruses replicate their genomes by reverse transcribing DNA copies from their RNA; these DNA copies are then transcribed to new RNA. Retrotransposons also spread by copying DNA and RNA from one another, and telomerase contains an RNA that is used as template for building the ends of eukaryotic chromosomes.

Double-stranded RNA

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

Key discoveries in RNA biology

Research on RNA has led to many important biological discoveries and numerous Nobel Prizes. Nucleic acids were discovered in 1868 by Friedrich Miescher, who called the material 'nuclein' since it was found in the nucleus. It was later discovered that prokaryotic cells, which do not have a nucleus, also contain nucleic acids. The role of RNA in protein synthesis was suspected already in 1939. Severo Ochoa won the 1959 Nobel Prize in Medicine (shared with Arthur Kornberg) after he discovered an enzyme that can synthesize RNA in the laboratory. Ironically, the enzyme discovered by Ochoa (polynucleotide phosphorylase) was later shown to be responsible for RNA degradation, not RNA synthesis.

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

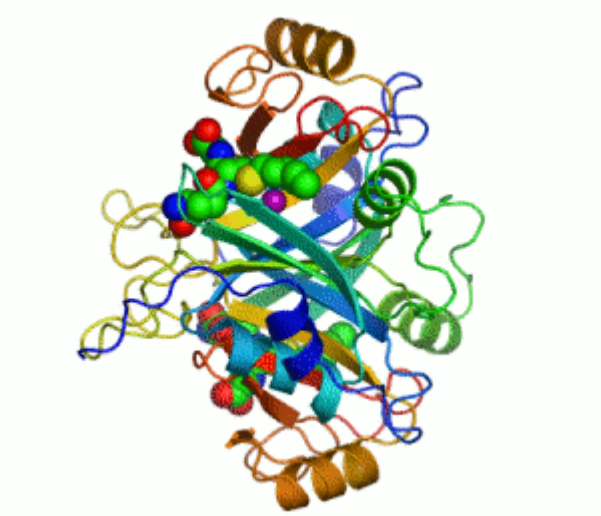
During the early 1970s, retroviruses and reverse transcriptase were discovered, showing for the first time that enzymes could copy RNA into DNA (the opposite of the usual route for transmission of genetic information). For this work, David Baltimore, Renato Dulbecco and Howard Temin were awarded a Nobel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

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

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

Chapter- 11

Enzyme



Human glyoxalase I. Two zinc ions that are needed for the enzyme to catalyze its reaction are shown as purple spheres, and an enzyme inhibitor called *S*-hexylglutathione is shown as a space-filling model, filling the two active sites.

Enzymes are proteins that catalyze (*i.e.*, increase or decrease the rates of) chemical reactions. In enzymatic reactions, the molecules at the beginning of the process are called substrates, and they are converted into different molecules, called the products. Almost all processes in a biological cell need enzymes to occur at significant rates. Since enzymes are selective for their substrates and speed up only a few reactions from among many possibilities, the set of enzymes made in a cell determines which metabolic pathways occur in that cell.

Like all catalysts, enzymes work by lowering the activation energy ($E_a^{\ddagger}$) for a reaction, thus dramatically increasing the rate of the reaction. As a result, products are formed faster and reactions reach their equilibrium state more rapidly. Most enzyme reaction rates are millions of times faster than those of comparable un-catalyzed reactions. As with all catalysts, enzymes are not consumed by the reactions they catalyze, nor do they alter the equilibrium of these reactions. However, enzymes do differ from most other catalysts by being much more specific. Enzymes are known to catalyze about 4,000

biochemical reactions. A few RNA molecules called ribozymes also catalyze reactions, with an important example being some parts of the ribosome. Synthetic molecules called artificial enzymes also display enzyme-like catalysis.

Enzyme activity can be affected by other molecules. Inhibitors are molecules that decrease enzyme activity; activators are molecules that increase activity. Many drugs and poisons are enzyme inhibitors. Activity is also affected by temperature, chemical environment (*e.g.*, pH), and the concentration of substrate. Some enzymes are used commercially, for example, in the synthesis of antibiotics. In addition, some household products use enzymes to speed up biochemical reactions (*e.g.*, enzymes in biological washing powders break down protein or fat stains on clothes; enzymes in meat tenderizers break down proteins into smaller molecules, making the meat easier to chew).

Etymology and history

Eduard Buchner

As early as the late 18th and early 19th centuries, the digestion of meat by stomach secretions and the conversion of starch to sugars by plant extracts and saliva were known. However, the mechanism by which this occurred had not been identified.

In the 19th century, when studying the fermentation of sugar to alcohol by yeast, Louis Pasteur came to the conclusion that this fermentation was catalyzed by a vital force contained within the yeast cells called "ferments", which were thought to function only within living organisms. He wrote that "alcoholic fermentation is an act correlated with the life and organization of the yeast cells, not with the death or putrefaction of the cells."

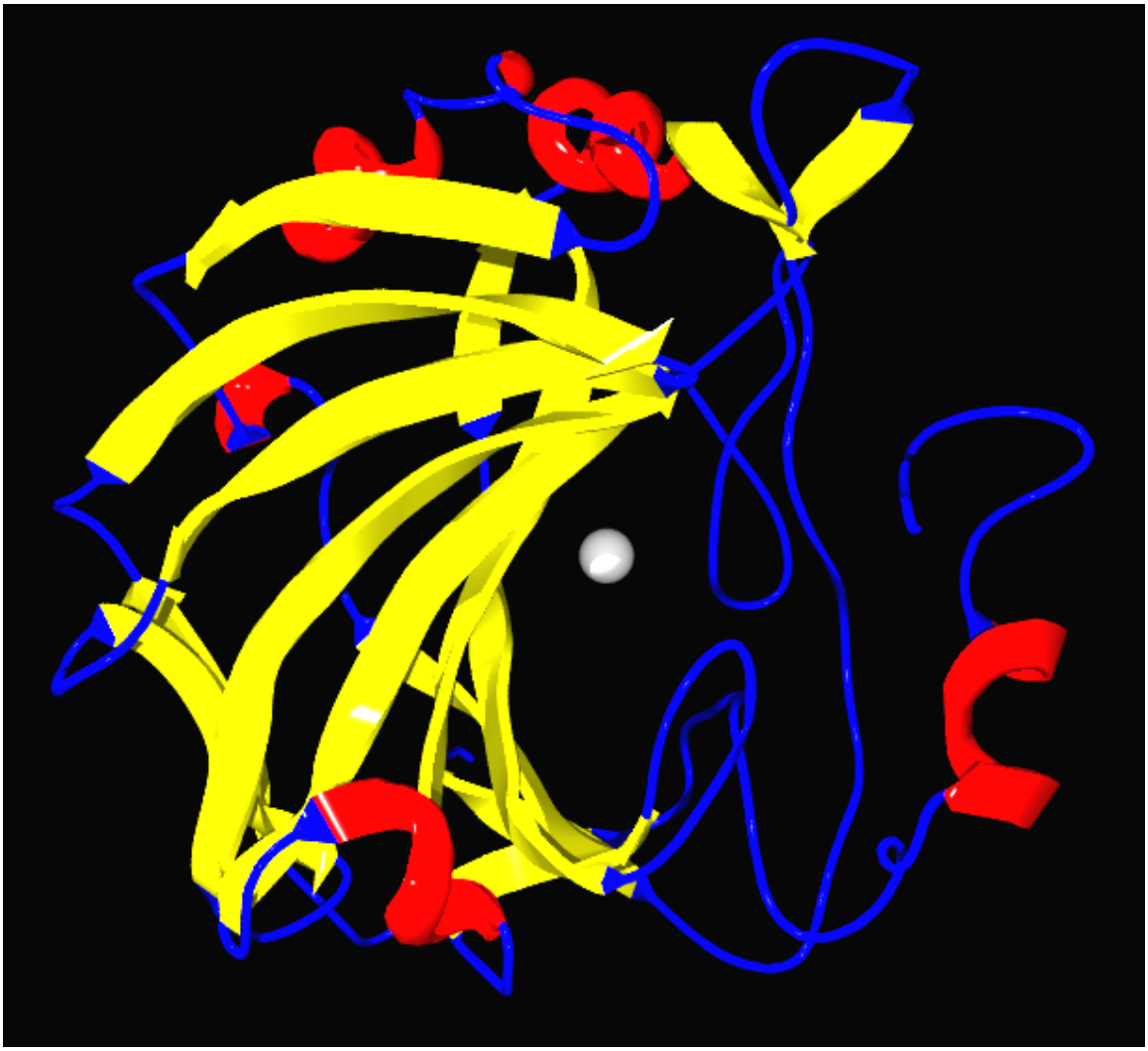
In 1877, German physiologist Wilhelm Kühne (1837–1900) first used the term *enzyme*, which comes from Greek *ενζυμον*, "in leaven", to describe this process. The word *enzyme* was used later to refer to nonliving substances such as pepsin, and the word *ferment* was used to refer to chemical activity produced by living organisms.

In 1897, Eduard Buchner submitted his first paper on the ability of yeast extracts that lacked any living yeast cells to ferment sugar. In a series of experiments at the University of Berlin, he found that the sugar was fermented even when there were no living yeast cells in the mixture. He named the enzyme that brought about the fermentation of sucrose "zymase". In 1907, he received the Nobel Prize in Chemistry "for his biochemical research and his discovery of cell-free fermentation". Following Buchner's example, enzymes are usually named according to the reaction they carry out. Typically, to generate the name of an enzyme, the suffix *-ase* is added to the name of its substrate (*e.g.*, lactase is the enzyme that cleaves lactose) or the type of reaction (*e.g.*, DNA polymerase forms DNA polymers).

Having shown that enzymes could function outside a living cell, the next step was to determine their biochemical nature. Many early workers noted that enzymatic activity was associated with proteins, but several scientists (such as Nobel laureate Richard Willstätter) argued that proteins were merely carriers for the true enzymes and that proteins *per se* were incapable of catalysis. However, in 1926, James B. Sumner showed that the enzyme urease was a pure protein and crystallized it; Sumner did likewise for the enzyme catalase in 1937. The conclusion that pure proteins can be enzymes was definitively proved by Northrop and Stanley, who worked on the digestive enzymes pepsin (1930), trypsin and chymotrypsin. These three scientists were awarded the 1946 Nobel Prize in Chemistry.

This discovery that enzymes could be crystallized eventually allowed their structures to be solved by x-ray crystallography. This was first done for lysozyme, an enzyme found in tears, saliva and egg whites that digests the coating of some bacteria; the structure was solved by a group led by David Chilton Phillips and published in 1965. This high-resolution structure of lysozyme marked the beginning of the field of structural biology and the effort to understand how enzymes work at an atomic level of detail.

Structures and mechanisms



Ribbon diagram showing human carbonic anhydrase II. The grey sphere is the zinc cofactor in the active site. Diagram drawn from PDB 1MOO.

Enzymes are generally globular proteins and range from just 62 amino acid residues in size, for the monomer of 4-oxalocrotonate tautomerase, to over 2,500 residues in the animal fatty acid synthase. A small number of RNA-based biological catalysts exist, with the most common being the ribosome; these are referred to as either RNA-enzymes or ribozymes. The activities of enzymes are determined by their three-dimensional structure. However, although structure does determine function, predicting a novel enzyme's activity just from its structure is a very difficult problem that has not yet been solved.

Most enzymes are much larger than the substrates they act on, and only a small portion of the enzyme (around 3–4 amino acids) is directly involved in catalysis. The region that contains these catalytic residues, binds the substrate, and then carries out the reaction is

known as the active site. Enzymes can also contain sites that bind cofactors, which are needed for catalysis. Some enzymes also have binding sites for small molecules, which are often direct or indirect products or substrates of the reaction catalyzed. This binding can serve to increase or decrease the enzyme's activity, providing a means for feedback regulation.

Like all proteins, enzymes are long, linear chains of amino acids that fold to produce a three-dimensional product. Each unique amino acid sequence produces a specific structure, which has unique properties. Individual protein chains may sometimes group together to form a protein complex. Most enzymes can be denatured—that is, unfolded and inactivated—by heating or chemical denaturants, which disrupt the three-dimensional structure of the protein. Depending on the enzyme, denaturation may be reversible or irreversible.

Structures of enzymes in complex with substrates or substrate analogs during a reaction may be obtained using Time resolved crystallography methods.

Specificity

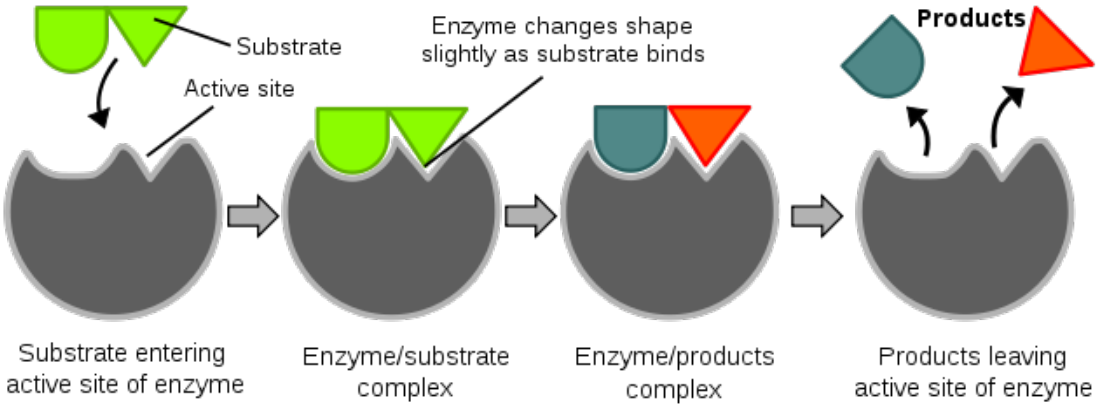
Enzymes are usually very specific as to which reactions they catalyze and the substrates that are involved in these reactions. Complementary shape, charge and hydrophilic/hydrophobic characteristics of enzymes and substrates are responsible for this specificity. Enzymes can also show impressive levels of stereospecificity, regioselectivity and chemoselectivity.

Some of the enzymes showing the highest specificity and accuracy are involved in the copying and expression of the genome. These enzymes have "proof-reading" mechanisms. Here, an enzyme such as DNA polymerase catalyzes a reaction in a first step and then checks that the product is correct in a second step. This two-step process results in average error rates of less than 1 error in 100 million reactions in high-fidelity mammalian polymerases. Similar proofreading mechanisms are also found in RNA polymerase, aminoacyl tRNA synthetases and ribosomes.

Some enzymes that produce secondary metabolites are described as promiscuous, as they can act on a relatively broad range of different substrates. It has been suggested that this broad substrate specificity is important for the evolution of new biosynthetic pathways.

"Lock and key" model

Enzymes are very specific, and it was suggested by the Nobel laureate organic chemist Emil Fischer in 1894 that this was because both the enzyme and the substrate possess specific complementary geometric shapes that fit exactly into one another. This is often referred to as "the lock and key" model. However, while this model explains enzyme specificity, it fails to explain the stabilization of the transition state that enzymes achieve.



Diagrams to show the induced fit hypothesis of enzyme action

In 1958, Daniel Koshland suggested a modification to the lock and key model: since enzymes are rather flexible structures, the active site is continually reshaped by interactions with the substrate as the substrate interacts with the enzyme. As a result, the substrate does not simply bind to a rigid active site; the amino acid side chains which make up the active site are molded into the precise positions that enable the enzyme to perform its catalytic function. In some cases, such as glycosidases, the substrate molecule also changes shape slightly as it enters the active site. The active site continues to change until the substrate is completely bound, at which point the final shape and charge is determined. Induced fit may enhance the fidelity of molecular recognition in the presence of competition and noise via the conformational proofreading mechanism .

Mechanisms

Enzymes can act in several ways, all of which lower $\Delta G^{\ddagger}$:

- Lowering the activation energy by creating an environment in which the transition state is stabilized (e.g. straining the shape of a substrate—by binding the transition-state conformation of the substrate/product molecules, the enzyme distorts the bound substrate(s) into their transition state form, thereby reducing the amount of energy required to complete the transition).
- Lowering the energy of the transition state, but without distorting the substrate, by creating an environment with the opposite charge distribution to that of the transition state.
- Providing an alternative pathway. For example, temporarily reacting with the substrate to form an intermediate ES complex, which would be impossible in the absence of the enzyme.
- Reducing the reaction entropy change by bringing substrates together in the correct orientation to react. Considering $\Delta H^{\ddagger}$ alone overlooks this effect.
- Increases in temperatures speed up reactions. Thus, temperature increases help the enzyme function and develop the end product even faster. However, if heated too

much, the enzyme’s shape deteriorates and the enzyme becomes denatured. Some enzymes like thermolabile enzymes work best at low temperatures.

Interestingly, this entropic effect involves destabilization of the ground state, and its contribution to catalysis is relatively small.

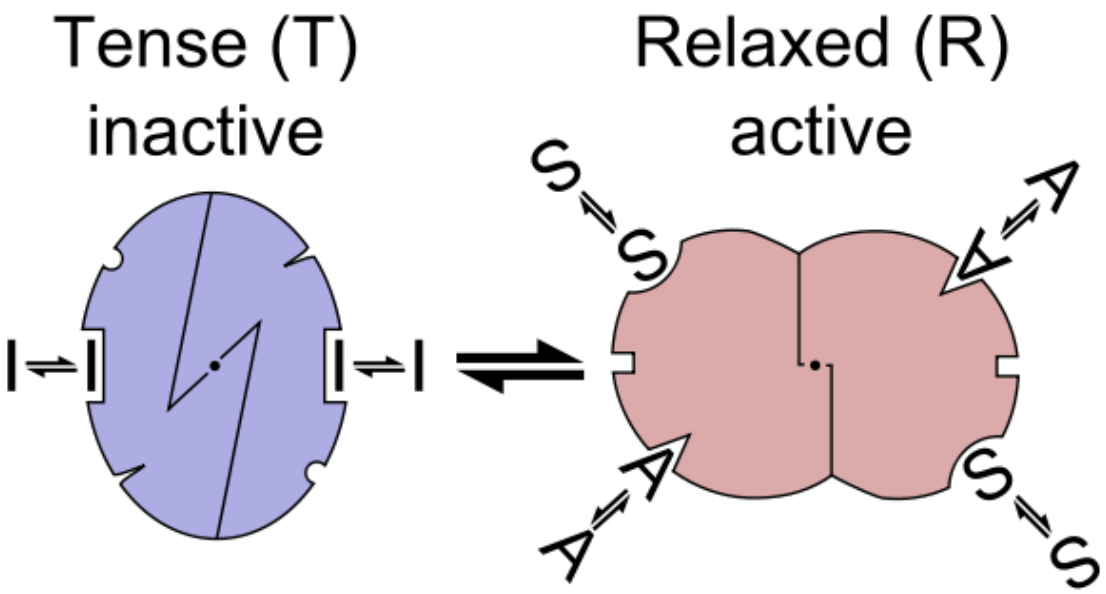
Transition State Stabilization

The understanding of the origin of the reduction of $\Delta G^{\ddagger}$ requires one to find out how the enzymes can stabilize its transition state more than the transition state of the uncatalyzed reaction. Apparently, the most effective way for reaching large stabilization is the use of electrostatic effects, in particular, by having a relatively fixed polar environment that is oriented toward the charge distribution of the transition state. Such an environment does not exist in the uncatalyzed reaction in water.

Dynamics and function

The internal dynamics of enzymes is linked to their mechanism of catalysis. Internal dynamics are the movement of parts of the enzyme's structure, such as individual amino acid residues, a group of amino acids, or even an entire protein domain. These movements occur at various time-scales ranging from femtoseconds to seconds. Networks of protein residues throughout an enzyme's structure can contribute to catalysis through dynamic motions. Protein motions are vital to many enzymes, but whether small and fast vibrations, or larger and slower conformational movements are more important depends on the type of reaction involved. However, although these movements are important in binding and releasing substrates and products, it is not clear if protein movements help to accelerate the chemical steps in enzymatic reactions. These new insights also have implications in understanding allosteric effects and developing new drugs.

Allosteric modulation



Allosteric transition of an enzyme between R and T states, stabilized by an agonist, an inhibitor and a substrate (the MWC model)

Allosteric sites are sites on the enzyme that bind to molecules in the cellular environment. The sites form weak, noncovalent bonds with these molecules, causing a change in the conformation of the enzyme. This change in conformation translates to the active site, which then affects the reaction rate of the enzyme. Allosteric interactions can both inhibit and activate enzymes and are a common way that enzymes are controlled in the body.

Cofactors and coenzymes

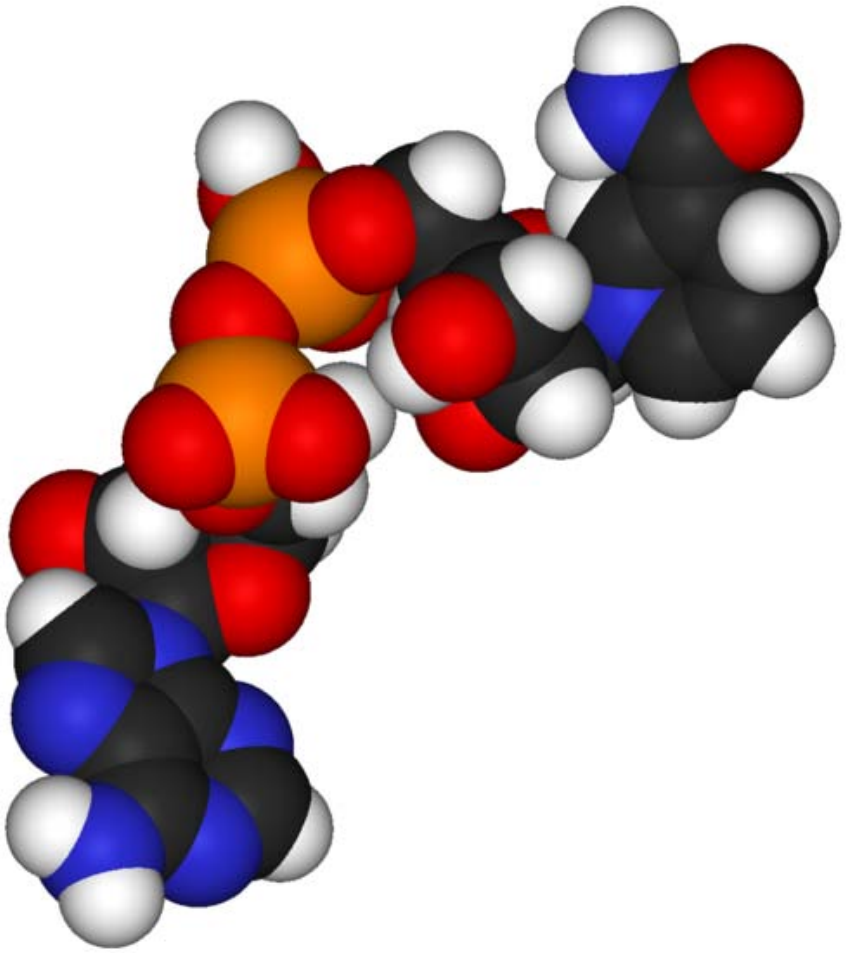
Cofactors

Some enzymes do not need any additional components to show full activity. However, others require non-protein molecules called cofactors to be bound for activity. Cofactors can be either inorganic (*e.g.*, metal ions and iron-sulfur clusters) or organic compounds (e.g., flavin and heme). Organic cofactors can be either prosthetic groups, which are tightly bound to an enzyme, or coenzymes, which are released from the enzyme's active site during the reaction. Coenzymes include NADH, NADPH and adenosine triphosphate. These molecules transfer chemical groups between enzymes.

An example of an enzyme that contains a cofactor is carbonic anhydrase, and is shown in the ribbon diagram above with a zinc cofactor bound as part of its active site. These tightly bound molecules are usually found in the active site and are involved in catalysis. For example, flavin and heme cofactors are often involved in redox reactions.

Enzymes that require a cofactor but do not have one bound are called *apoenzymes* or *apoproteins*. An apoenzyme together with its cofactor(s) is called a *holoenzyme* (this is the active form). Most cofactors are not covalently attached to an enzyme, but are very tightly bound. However, organic prosthetic groups can be covalently bound (*e.g.*, thiamine pyrophosphate in the enzyme pyruvate dehydrogenase). The term "holoenzyme" can also be applied to enzymes that contain multiple protein subunits, such as the DNA polymerases; here the holoenzyme is the complete complex containing all the subunits needed for activity.

Coenzymes



Space-filling model of the coenzyme NADH

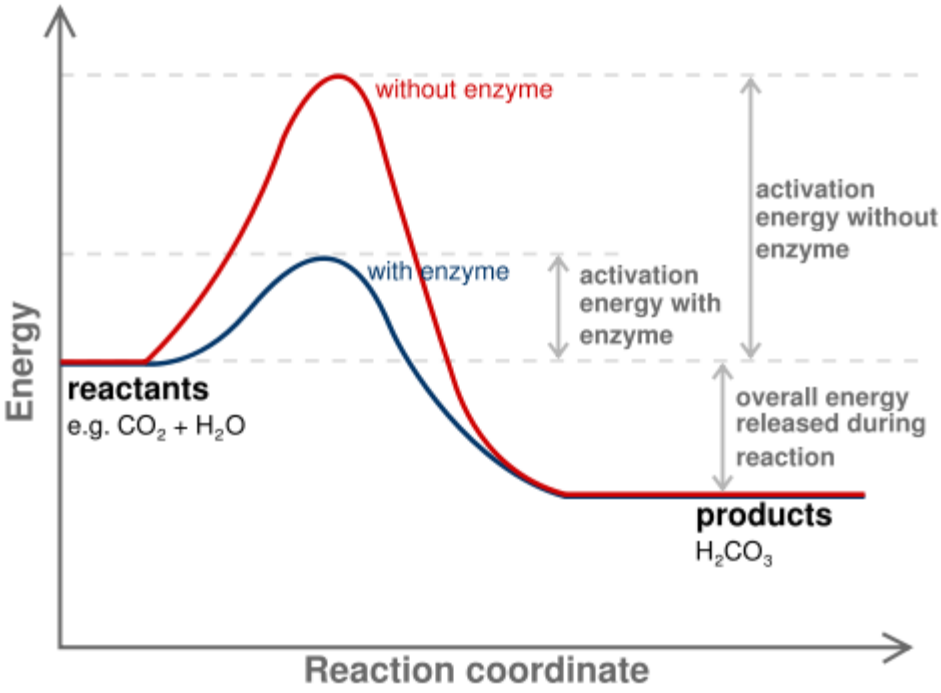
Coenzymes are small organic molecules that can be loosely or tightly bound to an enzyme. Tightly bound coenzymes can be called allosteric groups. Coenzymes are transport chemical groups from one enzyme to another. Some of these chemicals such as riboflavin, thiamine and folic acid are vitamins (compounds which cannot be synthesized by the body and must be acquired from the diet). The chemical groups carried include the

hydride ion (H^-) carried by NAD or $NADP^+$, the phosphate group carried by adenosine triphosphate, the acetyl group carried by coenzyme A, formyl, methenyl or methyl groups carried by folic acid and the methyl group carried by S-adenosylmethionine.

Since coenzymes are chemically changed as a consequence of enzyme action, it is useful to consider coenzymes to be a special class of substrates, or second substrates, which are common to many different enzymes. For example, about 700 enzymes are known to use the coenzyme NADH.

Coenzymes are usually continuously regenerated and their concentrations maintained at a steady level inside the cell: for example, NADPH is regenerated through the pentose phosphate pathway and *S*-adenosylmethionine by methionine adenosyltransferase. This continuous regeneration means that even small amounts of coenzymes are used very intensively. For example, the human body turns over its own weight in ATP each day.

Thermodynamics



The energies of the stages of a chemical reaction. Substrates need a lot of energy to reach a transition state, which then decays into products. The enzyme stabilizes the transition state, reducing the energy needed to form products.

As all catalysts, enzymes do not alter the position of the chemical equilibrium of the reaction. Usually, in the presence of an enzyme, the reaction runs in the same direction as

it would without the enzyme, just more quickly. However, in the absence of the enzyme, other possible uncatalyzed, "spontaneous" reactions might lead to different products, because in those conditions this different product is formed faster.

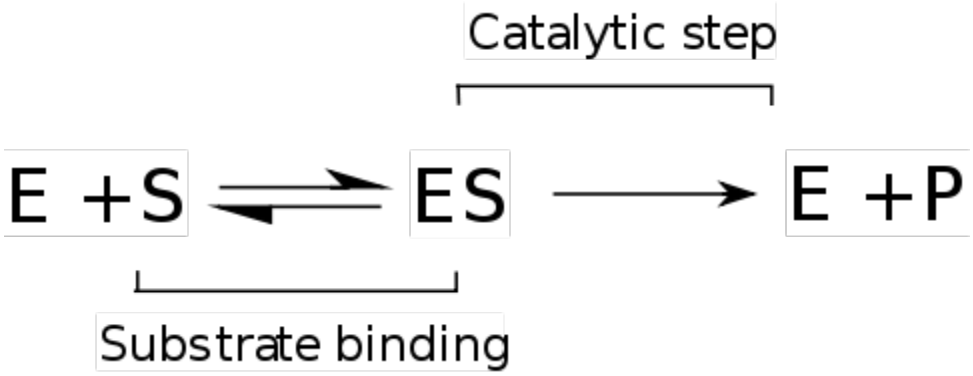
Furthermore, enzymes can couple two or more reactions, so that a thermodynamically favorable reaction can be used to "drive" a thermodynamically unfavorable one. For example, the hydrolysis of ATP is often used to drive other chemical reactions.

Enzymes catalyze the forward and backward reactions equally. They do not alter the equilibrium itself, but only the speed at which it is reached. For example, carbonic anhydrase catalyzes its reaction in either direction depending on the concentration of its reactants.

$$CO_2 + H_2O \xrightarrow{\text{Carbonic anhydrase}} H_2CO_3 \text{(in tissues; high } CO_2 \text{ concentration)}$$
$$H_2CO_3 \xrightarrow{\text{Carbonic anhydrase}} CO_2 + H_2O \text{(in lungs; low } CO_2 \text{ concentration)}$$

Nevertheless, if the equilibrium is greatly displaced in one direction, that is, in a very exergonic reaction, the reaction is *effectively* irreversible. Under these conditions the enzyme will, in fact, only catalyze the reaction in the thermodynamically allowed direction.

Kinetics



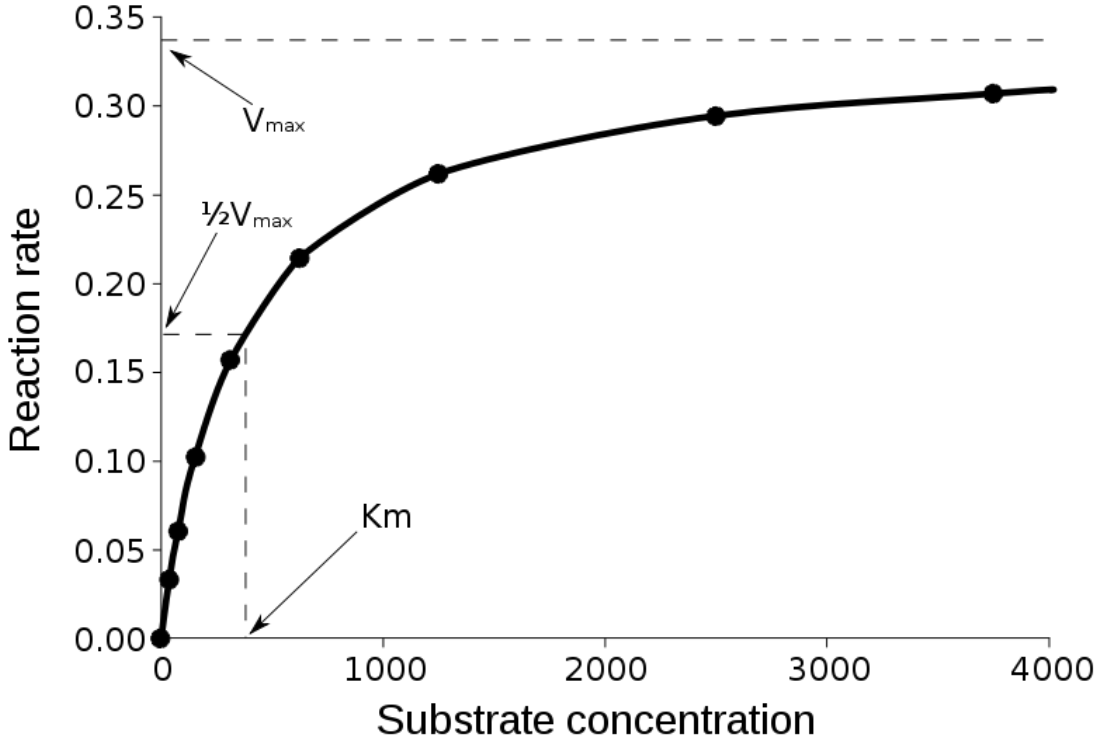
Mechanism for a single substrate enzyme catalyzed reaction. The enzyme (E) binds a substrate (S) and produces a product (P).

Enzyme kinetics is the investigation of how enzymes bind substrates and turn them into products. The rate data used in kinetic analyses are obtained from enzyme assays.

In 1902 Victor Henri proposed a quantitative theory of enzyme kinetics, but his experimental data were not useful because the significance of the hydrogen ion concentration was not yet appreciated. After Peter Lauritz Sørensen had defined the logarithmic pH-scale and introduced the concept of buffering in 1909 the German

chemist Leonor Michaelis and his Canadian postdoc Maud Leonora Menten repeated Henri's experiments and confirmed his equation which is referred to as Henri-Michaelis-Menten kinetics (sometimes also Michaelis-Menten kinetics). Their work was further developed by G. E. Briggs and J. B. S. Haldane, who derived kinetic equations that are still widely used today.

The major contribution of Henri was to think of enzyme reactions in two stages. In the first, the substrate binds reversibly to the enzyme, forming the enzyme-substrate complex. This is sometimes called the Michaelis complex. The enzyme then catalyzes the chemical step in the reaction and releases the product.



Saturation curve for an enzyme reaction showing the relation between the substrate concentration (S) and rate (v)

Enzymes can catalyze up to several million reactions per second. For example, the uncatalyzed decarboxylation of orotidine 5'-monophosphate has a half life of 78 million years. However, when the enzyme orotidine 5'-phosphate decarboxylase is added, the same process takes just 25 milliseconds. Enzyme rates depend on solution conditions and substrate concentration. Conditions that denature the protein abolish enzyme activity, such as high temperatures, extremes of pH or high salt concentrations, while raising substrate concentration tends to increase activity. To find the maximum speed of an enzymatic reaction, the substrate concentration is increased until a constant rate of product formation is seen. This is shown in the saturation curve on the right. Saturation happens because, as substrate concentration increases, more and more of the free enzyme

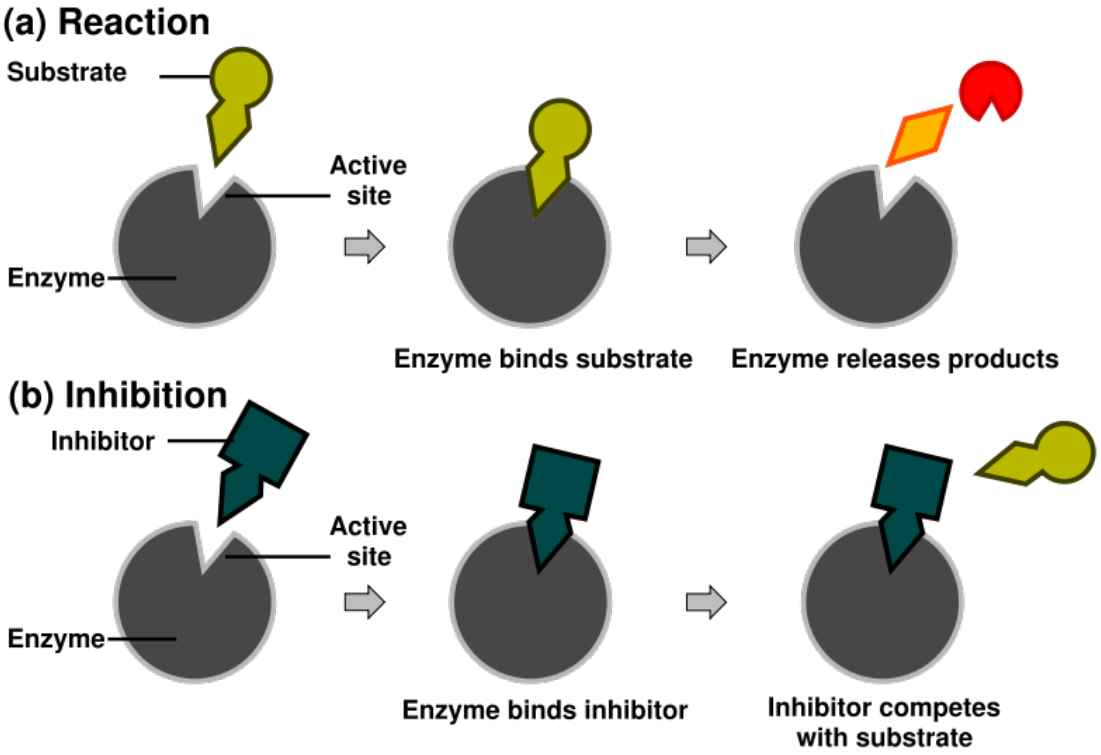
is converted into the substrate-bound ES form. At the maximum velocity (V_{max}) of the enzyme, all the enzyme active sites are bound to substrate, and the amount of ES complex is the same as the total amount of enzyme. However, V_{max} is only one kinetic constant of enzymes. The amount of substrate needed to achieve a given rate of reaction is also important. This is given by the Michaelis-Menten constant (K_m), which is the substrate concentration required for an enzyme to reach one-half its maximum velocity. Each enzyme has a characteristic K_m for a given substrate, and this can show how tight the binding of the substrate is to the enzyme. Another useful constant is k_{cat}, which is the number of substrate molecules handled by one active site per second.

The efficiency of an enzyme can be expressed in terms of k_{cat}/K_m. This is also called the specificity constant and incorporates the rate constants for all steps in the reaction. Because the specificity constant reflects both affinity and catalytic ability, it is useful for comparing different enzymes against each other, or the same enzyme with different substrates. The theoretical maximum for the specificity constant is called the diffusion limit and is about 10^8 to 10^9 (M^{-1} s^{-1}). At this point every collision of the enzyme with its substrate will result in catalysis, and the rate of product formation is not limited by the reaction rate but by the diffusion rate. Enzymes with this property are called *catalytically perfect* or *kinetically perfect*. Example of such enzymes are triose-phosphate isomerase, carbonic anhydrase, acetylcholinesterase, catalase, fumarase, β-lactamase, and superoxide dismutase.

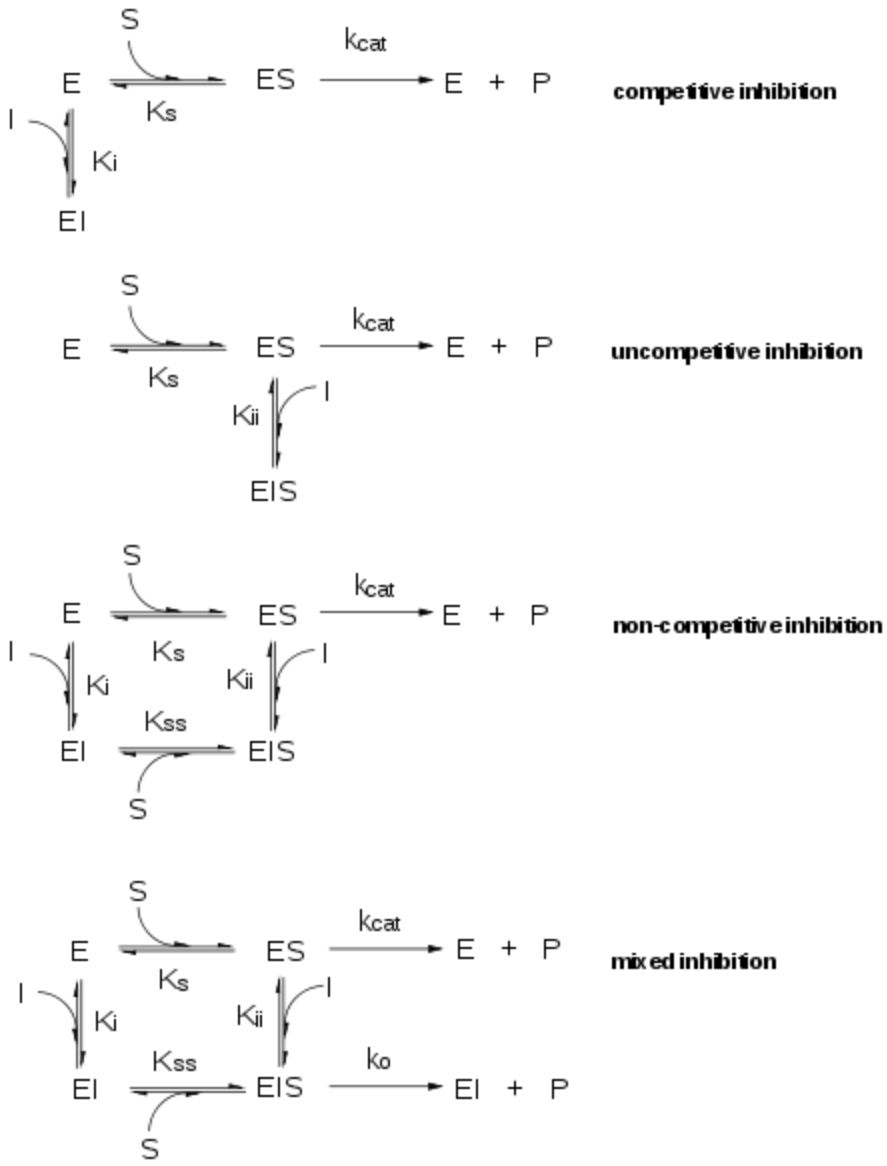
Michaelis-Menten kinetics relies on the law of mass action, which is derived from the assumptions of free diffusion and thermodynamically driven random collision. However, many biochemical or cellular processes deviate significantly from these conditions, because of macromolecular crowding, phase-separation of the enzyme/substrate/product, or one or two-dimensional molecular movement. In these situations, a fractal Michaelis-Menten kinetics may be applied.

Some enzymes operate with kinetics which are faster than diffusion rates, which would seem to be impossible. Several mechanisms have been invoked to explain this phenomenon. Some proteins are believed to accelerate catalysis by drawing their substrate in and pre-orienting them by using dipolar electric fields. Other models invoke a quantum-mechanical tunneling explanation, whereby a proton or an electron can tunnel through activation barriers, although for proton tunneling this model remains somewhat controversial. Quantum tunneling for protons has been observed in tryptamine. This suggests that enzyme catalysis may be more accurately characterized as "through the barrier" rather than the traditional model, which requires substrates to go "over" a lowered energy barrier.

Inhibition



Competitive inhibitors bind reversibly to the enzyme, preventing the binding of substrate. On the other hand, binding of substrate prevents binding of the inhibitor. Substrate and inhibitor compete for the enzyme.



Types of inhibition. This classification was introduced by W.W. Cleland.

Enzyme reaction rates can be decreased by various types of enzyme inhibitors.

Competitive inhibition

In competitive inhibition, the inhibitor and substrate compete for the enzyme (i.e., they can not bind at the same time). Often competitive inhibitors strongly resemble the real substrate of the enzyme. For example, methotrexate is a competitive inhibitor of the enzyme dihydrofolate reductase, which catalyzes the reduction of dihydrofolate to tetrahydrofolate. The similarity between the structures of folic acid and this drug are shown in the figure to the *right* bottom. Note that binding of the inhibitor need *not* be to the substrate binding site (as frequently stated), if binding of the inhibitor changes the

conformation of the enzyme to prevent substrate binding and *vice versa*. In competitive inhibition the maximal velocity of the reaction is not changed, but higher substrate concentrations are required to reach a given velocity, increasing the apparent K_m.

Uncompetitive inhibition

In uncompetitive inhibition the inhibitor can not bind to the free enzyme, but only to the ES-complex. The EIS-complex thus formed is enzymatically inactive. This type of inhibition is rare, but may occur in multimeric enzymes.

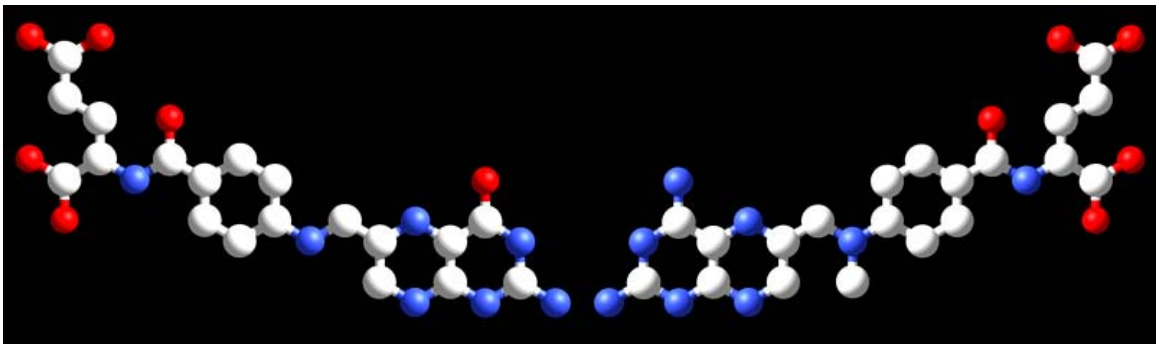
Non-competitive inhibition

Non-competitive inhibitors can bind to the enzyme at the binding site at the same time as the substrate,but not to the active site. Both the EI and EIS complexes are enzymatically inactive. Because the inhibitor can not be driven from the enzyme by higher substrate concentration (in contrast to competitive inhibition), the apparent V_{max} changes. But because the substrate can still bind to the enzyme, the K_m stays the same.

Mixed inhibition

This type of inhibition resembles the non-competitive, except that the EIS-complex has residual enzymatic activity.This type of inhibitor does not follow Michaelis-Menten equation.

In many organisms inhibitors may act as part of a feedback mechanism. If an enzyme produces too much of one substance in the organism, that substance may act as an inhibitor for the enzyme at the beginning of the pathway that produces it, causing production of the substance to slow down or stop when there is sufficient amount. This is a form of negative feedback. Enzymes which are subject to this form of regulation are often multimeric and have allosteric binding sites for regulatory substances. Their substrate/velocity plots are not hyperbolar, but sigmoidal (S-shaped).



The coenzyme folic acid (left) and the anti-cancer drug methotrexate (right) are very similar in structure. As a result, methotrexate is a competitive inhibitor of many enzymes that use folates.

Irreversible inhibitors react with the enzyme and form a covalent adduct with the protein. The inactivation is irreversible. These compounds include eflornithine a drug used to treat the parasitic disease sleeping sickness. Penicillin and Aspirin also act in this manner. With these drugs, the compound is bound in the active site and the enzyme then converts the inhibitor into an activated form that reacts irreversibly with one or more amino acid residues.

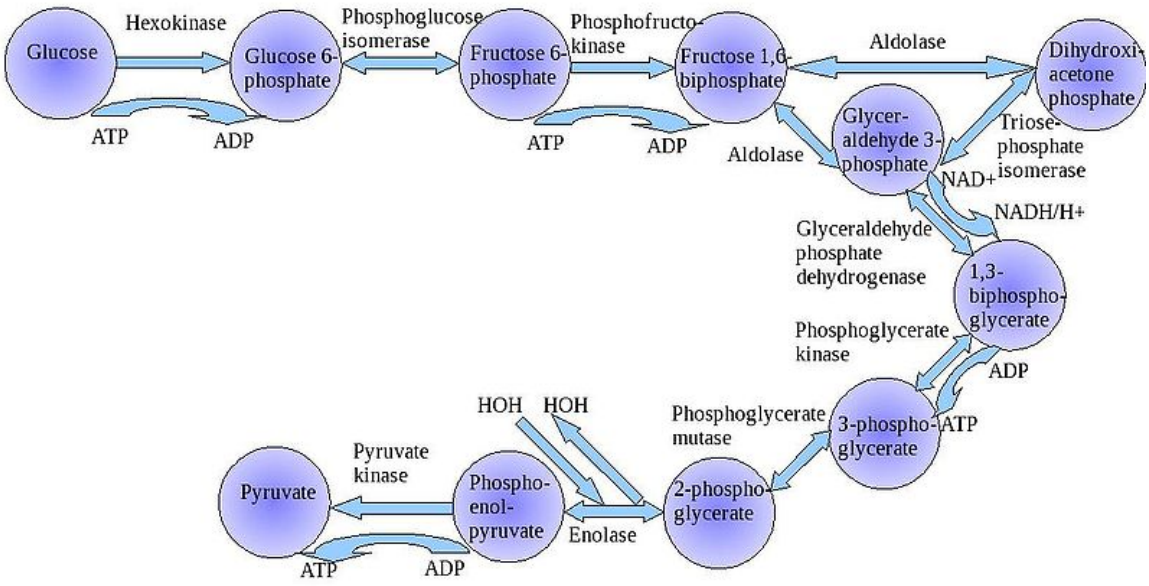
Uses of inhibitors

Since inhibitors modulate the function of enzymes they are often used as drugs. An common example of an inhibitor that is used as a drug is aspirin, which inhibits the COX-1 and COX-2 enzymes that produce the inflammation messenger prostaglandin, thus suppressing pain and inflammation. However, other enzyme inhibitors are poisons. For example, the poison cyanide is an irreversible enzyme inhibitor that combines with the copper and iron in the active site of the enzyme cytochrome c oxidase and blocks cellular respiration.

Biological function

Enzymes serve a wide variety of functions inside living organisms. They are indispensable for signal transduction and cell regulation, often via kinases and phosphatases. They also generate movement, with myosin hydrolysing ATP to generate muscle contraction and also moving cargo around the cell as part of the cytoskeleton. Other ATPases in the cell membrane are ion pumps involved in active transport. Enzymes are also involved in more exotic functions, such as luciferase generating light in fireflies. Viruses can also contain enzymes for infecting cells, such as the HIV integrase and reverse transcriptase, or for viral release from cells, like the influenza virus neuraminidase.

An important function of enzymes is in the digestive systems of animals. Enzymes such as amylases and proteases break down large molecules (starch or proteins, respectively) into smaller ones, so they can be absorbed by the intestines. Starch molecules, for example, are too large to be absorbed from the intestine, but enzymes hydrolyse the starch chains into smaller molecules such as maltose and eventually glucose, which can then be absorbed. Different enzymes digest different food substances. In ruminants which have herbivorous diets, microorganisms in the gut produce another enzyme, cellulase to break down the cellulose cell walls of plant fiber.



Glycolytic enzymes and their functions in the metabolic pathway of glycolysis

Several enzymes can work together in a specific order, creating metabolic pathways. In a metabolic pathway, one enzyme takes the product of another enzyme as a substrate. After the catalytic reaction, the product is then passed on to another enzyme. Sometimes more than one enzyme can catalyze the same reaction in parallel, this can allow more complex regulation: with for example a low constant activity being provided by one enzyme but an inducible high activity from a second enzyme.

Enzymes determine what steps occur in these pathways. Without enzymes, metabolism would neither progress through the same steps, nor be fast enough to serve the needs of the cell. Indeed, a metabolic pathway such as glycolysis could not exist independently of enzymes. Glucose, for example, can react directly with ATP to become phosphorylated at one or more of its carbons. In the absence of enzymes, this occurs so slowly as to be insignificant. However, if hexokinase is added, these slow reactions continue to take place except that phosphorylation at carbon 6 occurs so rapidly that if the mixture is tested a short time later, glucose-6-phosphate is found to be the only significant product. Consequently, the network of metabolic pathways within each cell depends on the set of functional enzymes that are present.

Control of activity

There are five main ways that enzyme activity is controlled in the cell.

1. **Enzyme production** (transcription and translation of enzyme genes) can be enhanced or diminished by a cell in response to changes in the cell's environment. This form of gene regulation is called enzyme induction and inhibition. For example, bacteria may become resistant to antibiotics such as penicillin because enzymes called beta-lactamases are induced that hydrolyse the crucial beta-lactam

ring within the penicillin molecule. Another example are enzymes in the liver called cytochrome P450 oxidases, which are important in drug metabolism. Induction or inhibition of these enzymes can cause drug interactions.

2. Enzymes can be **compartmentalized**, with different metabolic pathways occurring in different cellular compartments. For example, fatty acids are synthesized by one set of enzymes in the cytosol, endoplasmic reticulum and the Golgi apparatus and used by a different set of enzymes as a source of energy in the mitochondrion, through β-oxidation.
3. Enzymes can be regulated by **inhibitors and activators**. For example, the end product(s) of a metabolic pathway are often inhibitors for one of the first enzymes of the pathway (usually the first irreversible step, called *committed step*), thus regulating the amount of end product made by the pathways. Such a regulatory mechanism is called a negative feedback mechanism, because the amount of the end product produced is regulated by its own concentration. Negative feedback mechanism can effectively adjust the rate of synthesis of intermediate metabolites according to the demands of the cells. This helps allocate materials and energy economically, and prevents the manufacture of excess end products. The control of enzymatic action helps to maintain a stable internal environment in living organisms.
4. Enzymes can be regulated through **post-translational modification**. This can include phosphorylation, myristoylation and glycosylation. For example, in the response to insulin, the phosphorylation of multiple enzymes, including glycogen synthase, helps control the synthesis or degradation of glycogen and allows the cell to respond to changes in blood sugar. Another example of post-translational modification is the cleavage of the polypeptide chain. Chymotrypsin, a digestive protease, is produced in inactive form as chymotrypsinogen in the pancreas and transported in this form to the stomach where it is activated. This stops the enzyme from digesting the pancreas or other tissues before it enters the gut. This type of inactive precursor to an enzyme is known as a zymogen.
5. Some enzymes may become **activated when localized to a different environment** (e.g. from a reducing (cytoplasm) to an oxidizing (periplasm) environment, high pH to low pH etc.). For example, hemagglutinin in the influenza virus is activated by a conformational change caused by the acidic conditions, these occur when it is taken up inside its host cell and enters the lysosome.

Involvement in disease

Phenylalanine hydroxylase. Created from PDB 1KW0

Since the tight control of enzyme activity is essential for homeostasis, any malfunction (mutation, overproduction, underproduction or deletion) of a single critical enzyme can lead to a genetic disease. The importance of enzymes is shown by the fact that a lethal illness can be caused by the malfunction of just one type of enzyme out of the thousands of types present in our bodies.

One example is the most common type of phenylketonuria. A mutation of a single amino acid in the enzyme phenylalanine hydroxylase, which catalyzes the first step in the

degradation of phenylalanine, results in build-up of phenylalanine and related products. This can lead to mental retardation if the disease is untreated.

Another example is when germline mutations in genes coding for DNA repair enzymes cause hereditary cancer syndromes such as xeroderma pigmentosum. Defects in these enzymes cause cancer since the body is less able to repair mutations in the genome. This causes a slow accumulation of mutations and results in the development of many types of cancer in the sufferer.

Naming conventions

An enzyme's name is often derived from its substrate or the chemical reaction it catalyzes, with the word ending in ***-ase***. Examples are lactase, alcohol dehydrogenase and DNA polymerase. This may result in different enzymes, called isozymes, with the same function having the same basic name. Isoenzymes have a different amino acid sequence and might be distinguished by their optimal pH, kinetic properties or immunologically. Isoenzyme and isozyme are homologous proteins. Furthermore, the normal physiological reaction an enzyme catalyzes may not be the same as under artificial conditions. This can result in the same enzyme being identified with two different names. *E.g.* Glucose isomerase, used industrially to convert glucose into the sweetener fructose, is a xylose isomerase *in vivo*.

The International Union of Biochemistry and Molecular Biology have developed a nomenclature for enzymes, the **EC numbers**; each enzyme is described by a sequence of four numbers preceded by "EC". The first number broadly classifies the enzyme based on its mechanism.

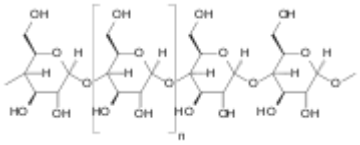
The top-level classification is

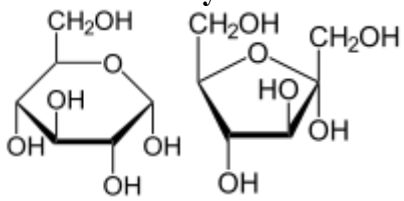
- EC 1 *Oxidoreductases*: catalyze oxidation/reduction reactions
- EC 2 *Transferases*: transfer a functional group (*e.g.* a methyl or phosphate group)
- EC 3 *Hydrolases*: catalyze the hydrolysis of various bonds
- EC 4 *Lyases*: cleave various bonds by means other than hydrolysis and oxidation
- EC 5 *Isomerases*: catalyze isomerization changes within a single molecule
- EC 6 *Ligases*: join two molecules with covalent bonds.

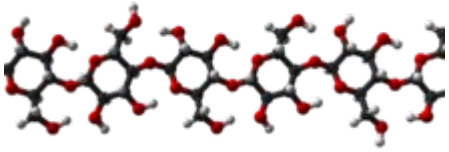
According to the naming conventions, enzymes are generally classified into six main family classes and many sub-family classes. Some web-servers, e.g., EzyPred and bioinformatics tools have been developed to predict which main family class and sub-family class an enzyme molecule belongs to according to its sequence information alone via the pseudo amino acid composition.

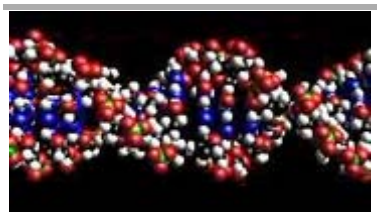
Industrial applications

Enzymes are used in the chemical industry and other industrial applications when extremely specific catalysts are required. However, enzymes in general are limited in the number of reactions they have evolved to catalyze and also by their lack of stability in organic solvents and at high temperatures. Consequently, protein engineering is an active area of research and involves attempts to create new enzymes with novel properties, either through rational design or *in vitro* evolution. These efforts have begun to be successful, and a few enzymes have now been desiged "from scratch" to catalyze reactions that do not occur in nature.

Application	Enzymes used	Uses
Food processing  Amylases catalyze the release of simple sugars from starch.	Amylases from fungi and plants	Production of sugars from starch, such as in making high-fructose corn syrup. In baking, catalyze breakdown of starch in the flour to sugar. Yeast fermentation of sugar produces the carbon dioxide that raises the dough.
	Proteases	Biscuit manufacturers use them to lower the protein level of flour.
Baby foods	Trypsin	To predigest baby foods
Brewing industry  Germinating barley used for malt	Enzymes from barley are released during the mashing stage of beer production.	They degrade starch and proteins to produce simple sugar, amino acids and peptides that are used by yeast for fermentation.
	Industrially produced barley enzymes	Widely used in the brewing process to substitute for the natural enzymes found in barley.
	Amylase, glucanases, proteases	Split polysaccharides and proteins in the malt.
	Betaglucanases and arabinoxylanases	Improve the wort and beer filtration characteristics.

	Amyloglucosidase and pullulanases	Low-calorie beer and adjustment of fermentability.
	Proteases	Remove cloudiness produced during storage of beers.
	Acetolactatedecarboxylase (ALDC)	Increases fermentation efficiency by reducing diacetyl formation.
Fruit juices	Cellulases, pectinases	Clarify fruit juices.
Dairy industry  Roquefort cheese	Rennin, derived from the stomachs of young ruminant animals (like calves and lambs)	Manufacture of cheese, used to hydrolyze protein
	Microbially produced enzyme	Now finding increasing use in the dairy industry
	Lipases	Is implemented during the production of Roquefort cheese to enhance the ripening of the blue-mould cheese.
	Lactases	Break down lactose to glucose and galactose.
Meat tenderizers	Papain	To soften meat for cooking
Starch industry  Glucose Fructose	Amylases, amyloglucosideases and glucoamylases	Converts starch into glucose and various syrups.
	Glucose isomerase	Converts glucose into fructose in production of high fructose syrups from starchy materials. These syrups have enhanced sweetening properties and lower calorific values than sucrose for the same level of sweetness.
Paper industry	Amylases, Xylanases, Cellulases and ligninases	Degrade starch to lower viscosity, aiding sizing and coating paper. Xylanases reduce

 A paper mill in South Carolina		bleach required for decolorising; cellulases smooth fibers, enhance water drainage, and promote ink removal; lipases reduce pitch and lignin-degrading enzymes remove lignin to soften paper.
Biofuel industry  Cellulose in 3D	Cellulases	Used to break down cellulose into sugars that can be fermented
	Ligninases	Use of lignin waste
Biological detergent	Primarily proteases, produced in an extracellular form from bacteria	Used for presoak conditions and direct liquid applications helping with removal of protein stains from clothes
	Amylases	Detergents for machine dish washing to remove resistant starch residues
	Lipases	Used to assist in the removal of fatty and oily stains
	Cellulases	Used in biological fabric conditioners
Contact lens cleaners	Proteases	To remove proteins on contact lens to prevent infections
Rubber industry	Catalase	To generate oxygen from peroxide to convert latex into foam rubber
Photographic industry	Protease (ficin)	Dissolve gelatin off scrap film, allowing recovery of its silver content.
Molecular biology	Restriction enzymes,	Used to manipulate

 Part of the DNA double helix	DNA ligase and polymerases	DNA in genetic engineering, important in pharmacology, agriculture and medicine. Essential for restriction digestion and the polymerase chain reaction. Molecular biology is also important in forensic science.