

BME 1532-CELL BIOLOGY

DNA, DNA Replication and Repair

by Assist. Prof. Görke Gürel Peközer

Yıldız Technical University
Biomedical Engineering Department
Spring 2020

Last Week on BME 1532

- Membrane Enclosed Cellular Organelles
 - Nucleus
 - Endoplasmic Reticulum
 - Golgi Apparatus
 - Lysosome
 - Mitochondria
 - Chloroplast
- Cytoskeleton
 - Actin Filaments
 - Intermediate Filaments
 - Microtubules
- Microscopical Techniques
 - Light Microscopy (Brightfield, Darkfield, Phase-Contrast)
 - Fluorescence Microscopy
 - Confocal Microscopy
 - Electron Microscopy (SEM and TEM)

- Life depends on the ability of cells to store, retrieve, and translate the genetic instructions required to make and maintain a living organism.
- This hereditary information is passed on from a cell to its daughter cells at cell division, and from generation to generation in multicellular organisms through the reproductive cells—eggs and sperm.
- The information in genes is copied and transmitted from cell to daughter cells millions of times during the life of a multicellular organism, and it survives the process essentially unchanged.

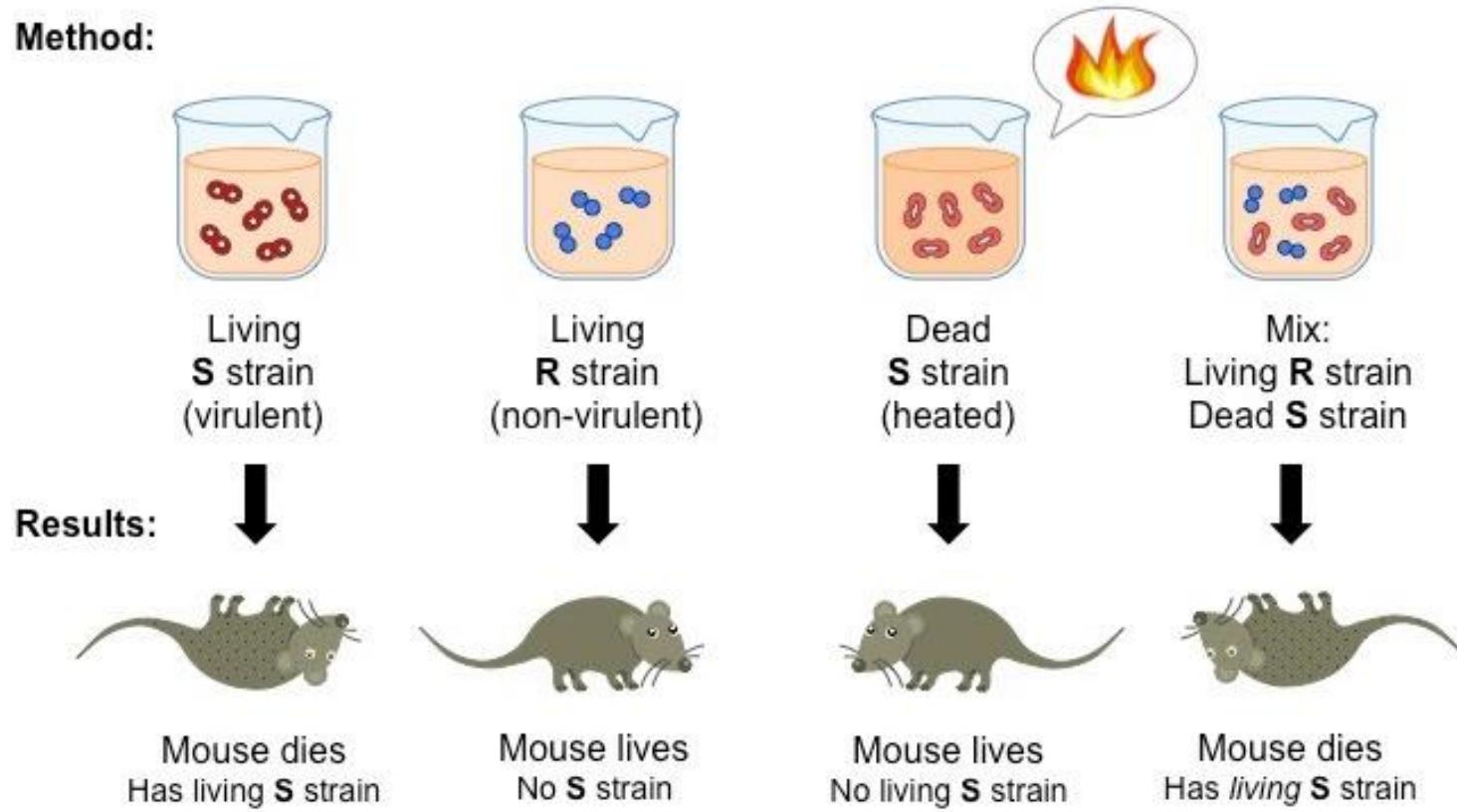
- The biochemical investigation of DNA began with Friedrich Miescher, who carried out the first systematic chemical studies of cell nuclei.
- In 1868 Miescher isolated a phosphorus-containing substance, which he called “nuclein,” from the nuclei of pus cells (leukocytes) obtained from discarded surgical bandages.
- He found nuclein to consist of an acidic portion, which we know today as DNA, and a basic portion, protein.
- Miescher later found a similar acidic substance in the heads of sperm cells from salmon. Although he partially purified nuclein and studied its properties, the covalent (primary structure of DNA) was not known with certainty until the late 1940s.

Griffith Experiment

- Miescher and many others suspected that nuclein (nucleic acid) was associated in some way with cell inheritance.
- In 1928, British bacteriologist Frederick Griffith conducted a series of experiments using *Streptococcus pneumoniae* bacteria and mice.
- Griffith found that DNA extracted from a virulent (disease-causing) strain of the bacterium *Streptococcus pneumoniae*, also known as pneumococcus, genetically transformed a nonvirulent strain of this organism into a virulent form.

Griffith Experiment

Method:

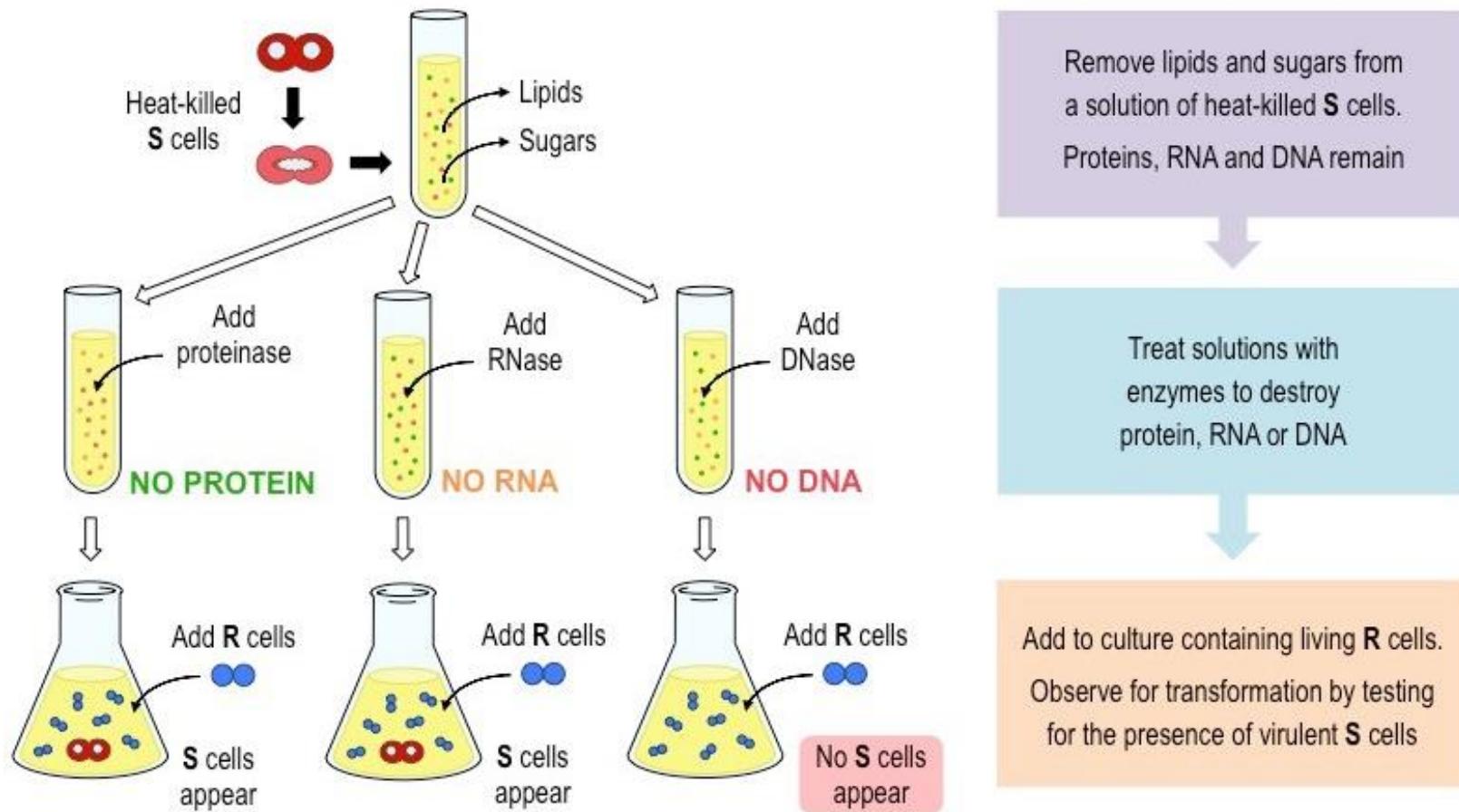


Conclusion: A chemical substance from one cell is genetically transforming another cell

Avery-Macload-McCarty Experiment

- The first direct evidence that DNA is the bearer of genetic information came in 1944 through a discovery made by Oswald T. Avery, Colin MacLeod, and Maclyn McCarty.
- These investigators expanded upon the findings of Griffith and found that DNA extracted from a virulent strain of the bacterium *Streptococcus pneumoniae* genetically transformed a nonvirulent strain of this organism into a virulent form.

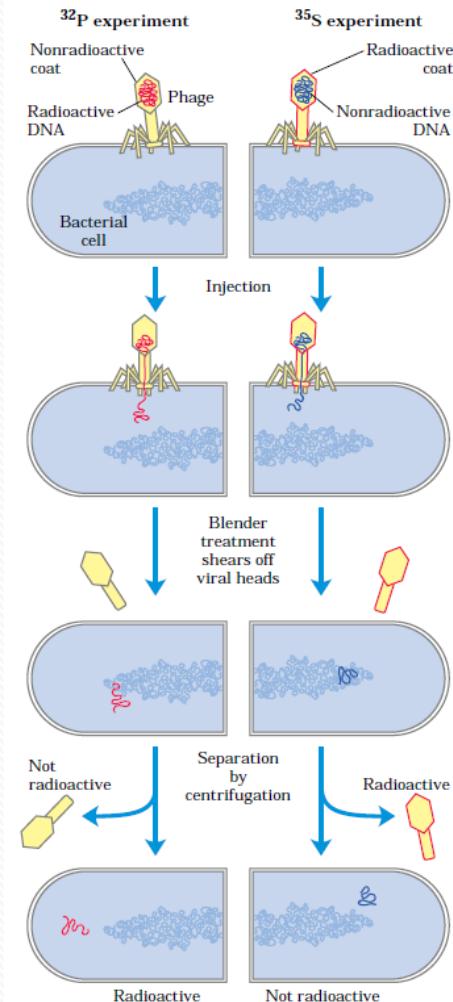
Avery-Macload-McCarty Experiment



Conclusion: Transformation requires DNA, therefore it is the genetic material of the cell

Hershey and Chase Experiment

- A second important experiment provided independent evidence that DNA carries genetic information.
- In 1952 Alfred D. Hershey and Martha Chase used radioactive phosphorus (^{32}P) and radioactive sulfur (^{35}S) tracers to show that when the bacterial virus (bacteriophage) T2 infects its host cell, *Escherichia coli*, it is the phosphorus-containing DNA of the viral particle, not the sulfur-containing protein of the viral coat, that enters the host cell and furnishes the genetic information for viral replication.



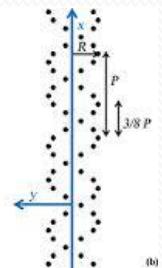
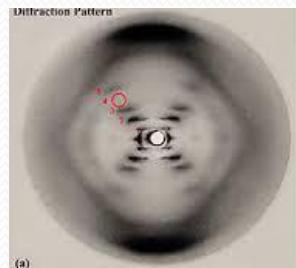
- But the mechanism whereby the hereditary information is copied for transmission from one generation of cells to the next, and how proteins are specified by the instructions in DNA, remained completely mysterious at the time.
- Erwin Chargaff and his colleagues in the late 1940s found that the four nucleotide bases of DNA occur in different ratios in the DNAs of different organisms and that the amounts of certain bases are closely related.

Chargaff rules

1. The base composition of DNA generally varies from one species to another.
 2. DNA specimens isolated from different tissues of the same species have the same base composition.
 3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.
 4. In *all* cellular DNAs, regardless of the species, the number of adenosine residues is equal to the number of thymidine residues (that is, $A = T$), and the number of guanosine residues is equal to the number of cytidine residues ($G = C$). From these relationships it follows that the sum of the purine residues equals the sum of the pyrimidine residues; that is, $A + G = T + C$.
- These rules were a key to establishing the 3D structure of DNA and yielded clues to how genetic information is encoded in DNA and passed from one generation to the next.

Double Helical Structure of DNA

- To shed more light on the structure of DNA, Rosalind Franklin and Maurice Wilkins used the powerful method of x-ray diffraction to analyze DNA fibers.
- They showed in the early 1950s that DNA produces a characteristic x-ray diffraction pattern.
- From this pattern it was deduced that DNA molecules are helical with two periodicities along their long axis, a primary one of 3.4 \AA and a secondary one of 34 \AA .
- The problem then was to formulate a 3D model of the DNA molecule that could account not only for the x-ray diffraction data but also for the specific $A = T$ and $G = C$ base equivalences discovered by Chargaff and for the other chemical properties of DNA.
- In 1953 Watson and Crick postulated a 3D model of DNA structure that accounted for all the available data.



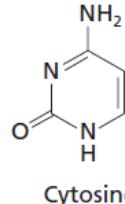
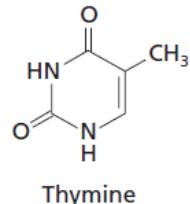
1962 Nobel Prize in
Physiology and
Medicine

Nucleic Acid Structure

- Nucleic acids are the polymers of nucleotides.
- A molecule of deoxyribonucleic acid (DNA) consists of two long polynucleotide chains.
- Each nucleotide consists of a five-carbon sugar (pentose), a weakly basic nitrogenous compound called a base and a phosphate group.
- For the nucleotides in DNA, the sugar is deoxyribose (hence the name deoxyribonucleic acid), and the base can be either *adenine* (A), *cytosine* (C), *guanine* (G), or *thymine* (T).
- The nucleotides are covalently linked together in a chain through the sugars and phosphates, which thus form a backbone of alternating sugar–phosphate–sugar–phosphate.

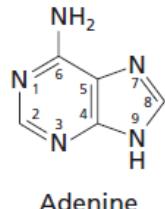


- Each *chain*, or *strand*, is composed of four types of nucleotide subunits, and the two strands are held together by hydrogen bonds between the base portions of the nucleotides.
 - All the bases are therefore on the inside of the double helix, with the sugar–phosphate backbones on the outside
 - A always pairs with T, and G always pairs with C.
 - In each case, a bulkier two-ring base (a purine is paired with a single-ring base (a pyrimidine).
 - Each purine– pyrimidine pair is called a *base pair*, and this is called *complementary base-pairing*.

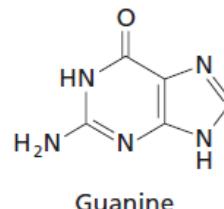


Cytosine

Pyrimidines

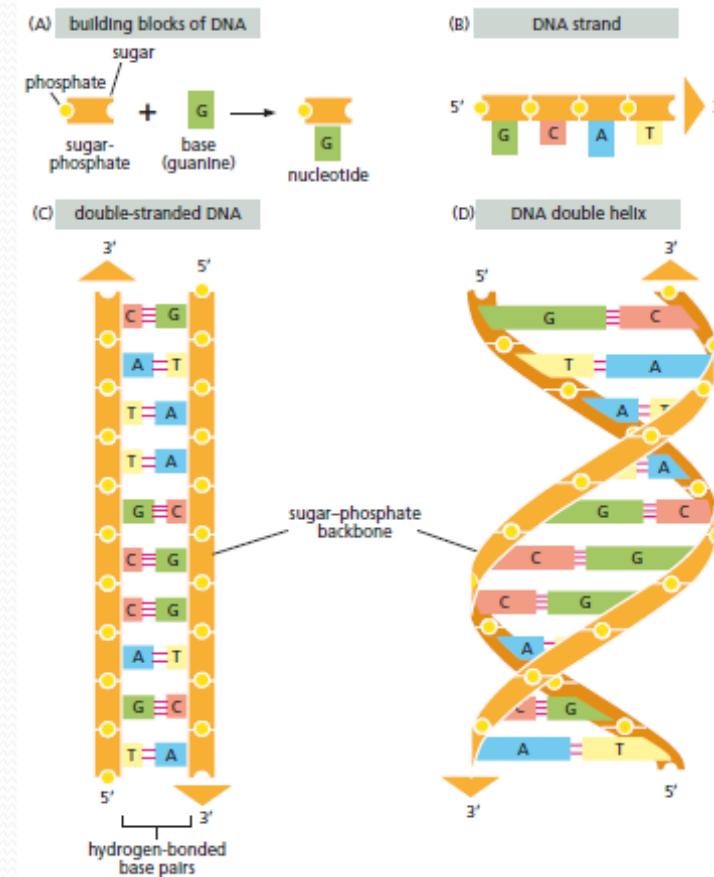


Adenine

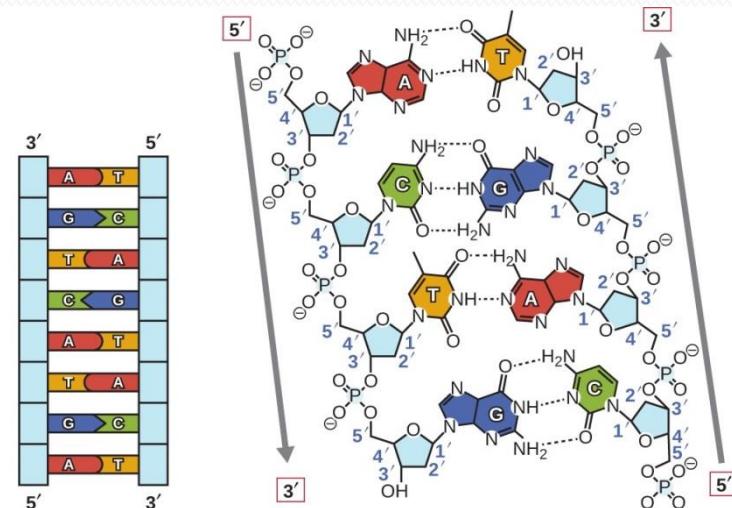
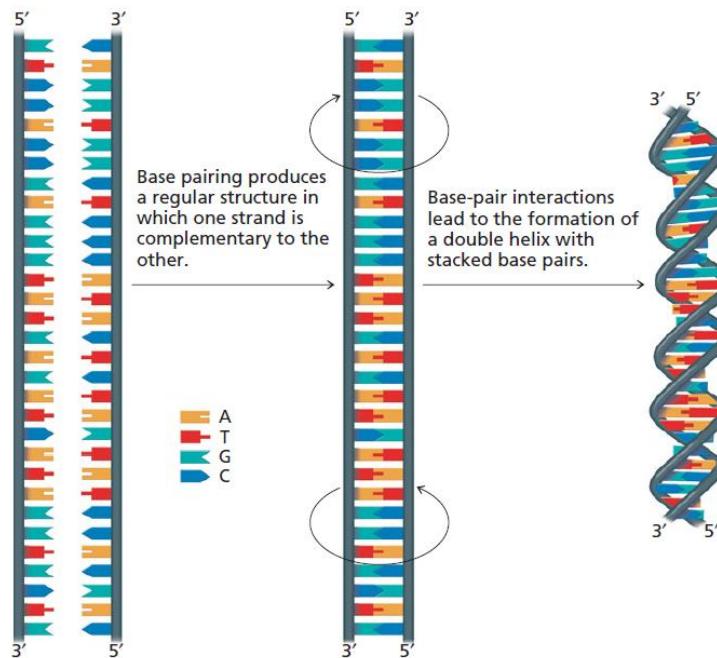


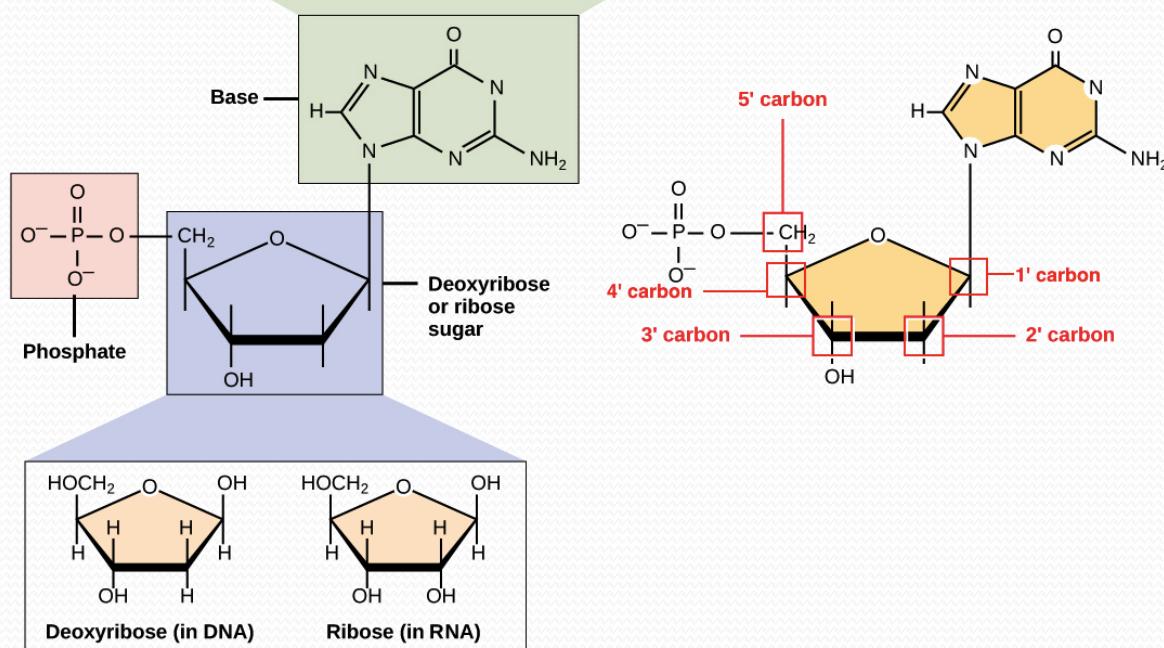
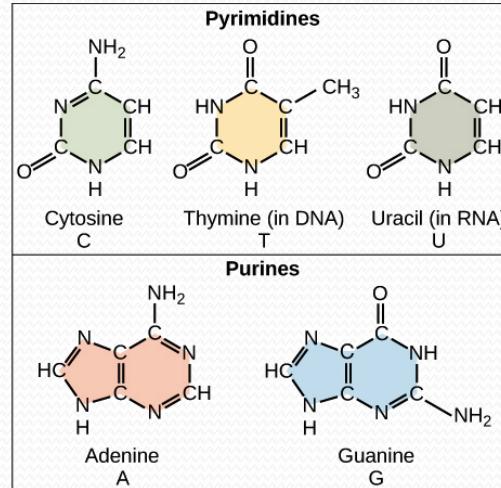
Guanine

Purines



- In this arrangement, each base pair has a similar width, thus holding the sugar–phosphate backbones an equal distance apart along the DNA molecule.
 - The members of each base pair can fit together within the double helix because the two strands of the helix run ***antiparallel*** to each other—that is, they are oriented with opposite polarities.
 - The antiparallel sugar–phosphate strands then twist around each other to form a double helix.





- A consequence of the base-pairing requirements is that each strand of a DNA double helix contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand—an A always matches a T on the opposite strand, and a C always matches a G.
- This complementarity is of crucial importance when it comes to both copying and repairing the DNA because one strand can act as a template for the other strand.
- Each base—A, C, T, or G—can be considered a letter in a four-letter alphabet that is used to spell out biological messages
- Organisms differ from one another because their respective DNA molecules have different *nucleotide sequences* and, consequently, carry different biological messages.

(A) molecular biology is...

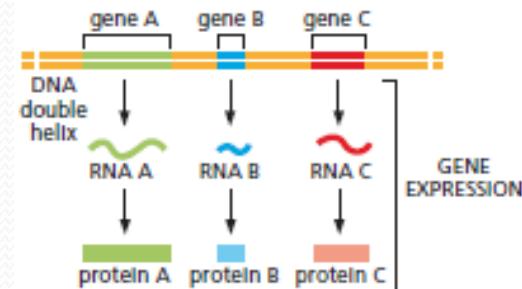
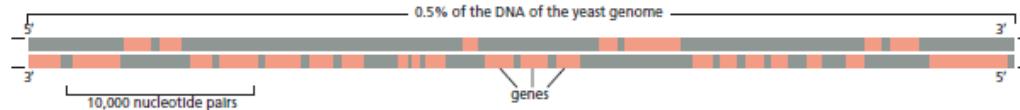
(B)

(C) - - - - -

(D) 细胞生物学 乐趣无穷

(E) TTGAGCGACCTAACCTATAG

- The 4-letter nucleotide alphabet (A, T, G, C) of DNA corresponds to the 20-letter amino acid alphabet (Gly, Ala, Val, Leu, Ile, ...) of proteins and this is called the genetic code.
- **Genome** is the complete set of DNA in an organism.
- **Genes** are the basic functional units of DNA which encodes either protein or RNA.
- **Gene expression**—the process by which the nucleotide sequence of a gene is *transcribed* into the nucleotide sequence of an RNA molecule, which, in most cases, is then *translated* into the amino acid sequence of a protein.



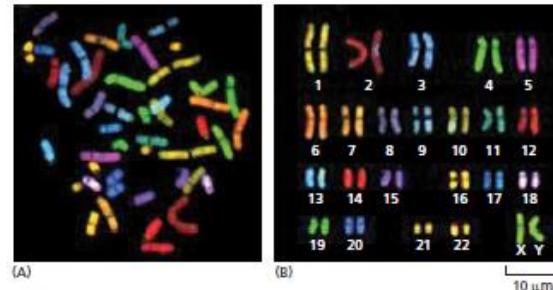
Genome

- Large amounts of DNA are required to encode all the information needed to make even a single-celled bacterium, and far more DNA is needed to encode the information to make a multicellular organism.
- The entire genetic material of an organism is called its *genome*.
- Each human cell contains about 2 meters of DNA; yet the cell nucleus is only 5–8 μm in diameter.
- Tucking all this material into such a small space is the equivalent of trying to fold 40 km of extremely fine thread into a tennis ball.
- The complex task of packaging DNA is accomplished by specialized proteins that bind to and fold the DNA, generating a series of coils and loops that provide increasingly higher levels of organization.
- In eukaryotic cells, very long double-stranded DNA molecules are packaged into chromosomes.
- The DNA is compacted in a way that allows it to remain accessible to all of the enzymes and other proteins that replicate it, repair it, and control the expression of its genes.

- In eukaryotes, the DNA in the nucleus is distributed among a set of different chromosomes.
- The DNA in a human nucleus, for example, contains approximately 3.2×10^9 nucleotides parceled out into 23 or 24 different types of chromosome (males, with their Y chromosome, have an extra type of chromosome that females do not have).
- With the exception of the germ cells (sperm and eggs) and highly specialized cells that lack DNA entirely (such as mature red blood cells), human cells each contain two copies of each chromosome, one inherited from the mother and one from the father.
- The maternal and paternal chromosomes of a pair are called ***homologous chromosomes***.
- The only nonhomologous chromosome pairs are the sex chromosomes in males, where a *Y chromosome* is inherited from the father and an *X chromosome* from the mother. (Females inherit one X chromosome from each parent and have no Y chromosome.)

Karyotype

- An ordered display of the full set of 46 human chromosomes is called the *human karyotype*.
- If parts of a chromosome are lost, or switched between chromosomes, these changes can be detected.
- Cytogeneticists analyze karyotypes to detect chromosomal abnormalities that are associated with some inherited defects and with certain types of cancer.

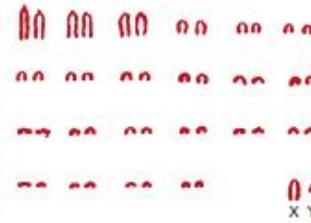


Complexity Relationships

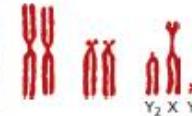
- In general, the more complex an organism, the larger is its genome.
- But this relationship does not always hold true. The human genome, for example, is 200 times larger than that of the yeast *S. cerevisiae*, but 30 times smaller than that of some plants and at least 60 times smaller than some species of amoeba.
- Some correlation exists between the complexity of an organism and the number of genes in its genome. For example, the total number of genes ranges from less than 500 for a simple bacterium to about 30,000 for humans.
- Furthermore, humans have a total of 46 chromosomes (including both maternal and paternal sets), but a species of small deer has only 7, while some carp species have more than 100. A fruit fly has 4 pairs of chromosomes, while a rice plant has 12 and a dog, 39.
- Even closely related species with similar genome sizes can have very different numbers and sizes of chromosomes.
- Thus, although gene number is roughly correlated with species complexity, there is no simple relationship between chromosome number and total genome size.
- Even closely related species with similar genome sizes can have very different numbers and sizes of chromosomes.



Chinese muntjac



Indian muntjac



Higher Level Packing of DNA

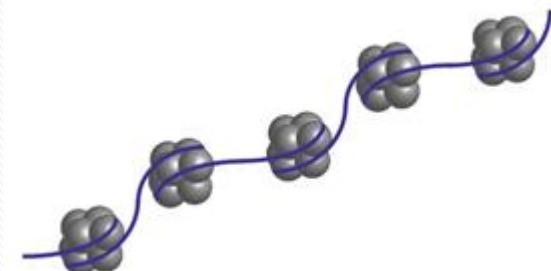
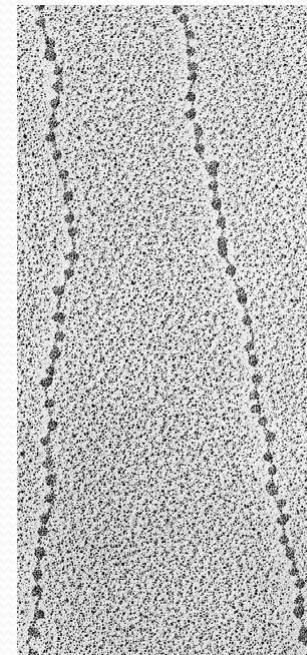
- In a normal resting cell, chromatin exists as 30 nm fiber.
- In humans, the nucleus must accommodate 46 such chromatin fibers, or chromosomes.
- The largest human chromosome (chr. 1) is about 2.4×10^8 bp; it would be about 8 cm long if it were stretched out.
- During metaphase (when chromosomes are most condensed) the largest chromosome is about 10 μm long.
- This remarkable feat of compression is performed by proteins that coil and fold the DNA into higher and higher levels of organization.

Chromatin

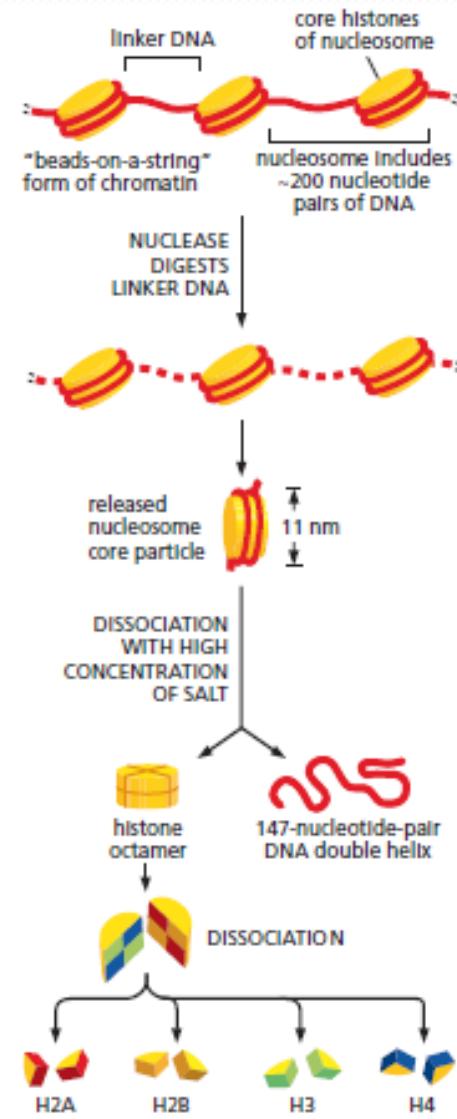
- The proteins that bind to DNA to form eukaryotic chromosomes are traditionally divided into two general classes: the histones and the *nonhistone chromosomal proteins*.
- The complex of both classes of protein with nuclear DNA is called **chromatin**.
- Histones are responsible for the first and most fundamental level of chromatin packing, the **nucleosome**.

Histone Proteins

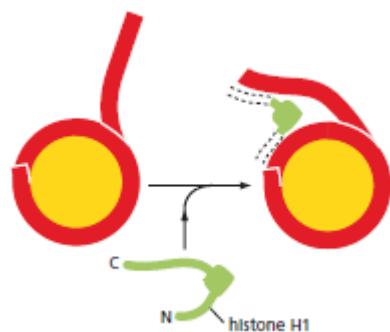
- Histone proteins are involved in the packing of DNA.
- Histones are small, basic proteins containing numerous lysine and arginine residues whose positive charges allow the proteins to bind to the negatively charged sugar-phosphate backbone of DNA.
- DNA is wrapped around the histone proteins.
- DNA-histone complex, which is chromatin fiber, looks like beads on a string in an electron micrograph.
- The “beads” are DNA-histone complexes called **nucleosomes** and the “string” is double-stranded DNA.



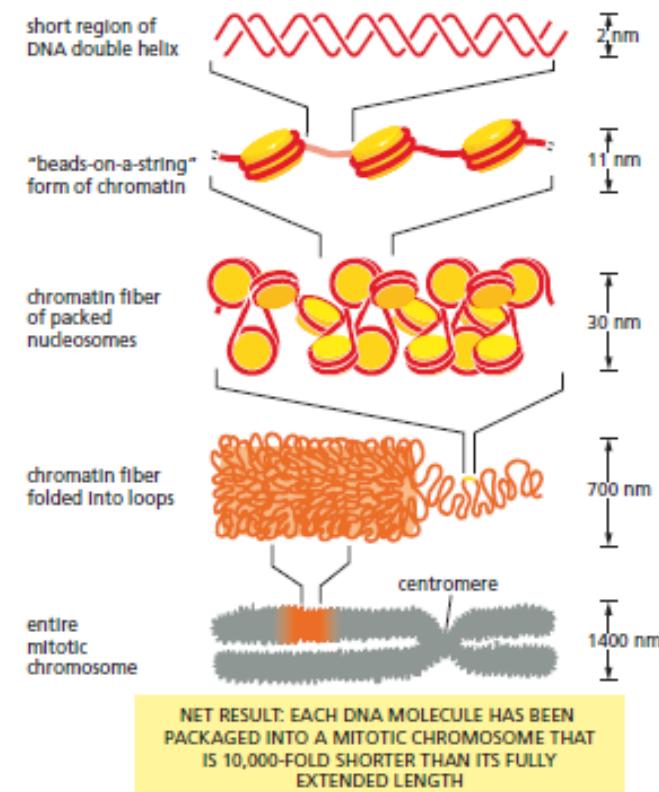
- An individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H₂A, H₂B, H₃, and H₄—and a stretch of double-stranded DNA, 147 nucleotide pairs long, that winds around this *histone octamer*.
- The linker DNA between each nucleosome core particle can vary in length from a few nucleotide pairs up to about 80.
- The formation of nucleosomes converts a DNA molecule into a chromatin thread that is approximately one-third the length of the initial piece of DNA, and it provides the first level of DNA packing.



- The nucleosomes are further packed on top of one another to generate a more compact structures.
- This additional packing of nucleosomes into a chromatin fiber depends on a fifth histone called histone H1, which is thought to pull adjacent nucleosomes together into a regular repeating array.
- This “linker” histone changes the path the DNA takes as it exits the nucleosome core, allowing it to form a more condensed chromatin fiber.

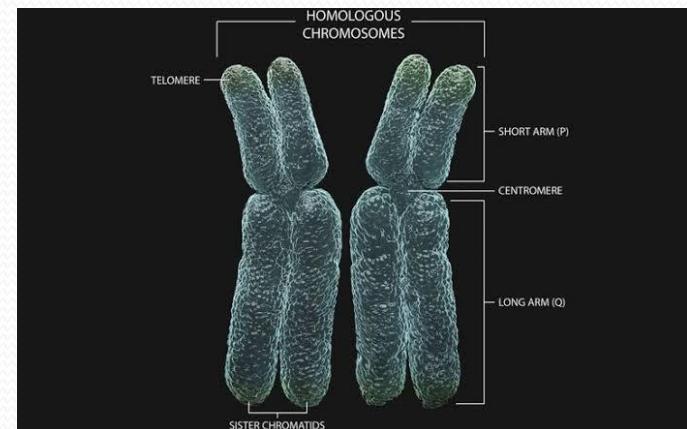
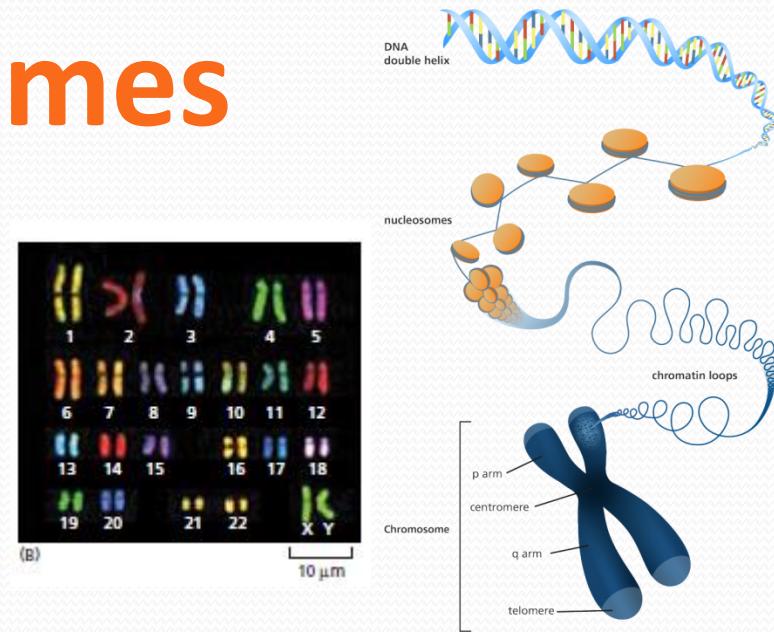


- Chromatin fiber is folded into a series of loops, and that these loops are further condensed to produce the interphase chromosome; finally, this compact string of loops is thought to undergo at least one more level of packing to form the mitotic chromosome.
- Chromosomes appear only after DNA replication in cell division during metaphase.

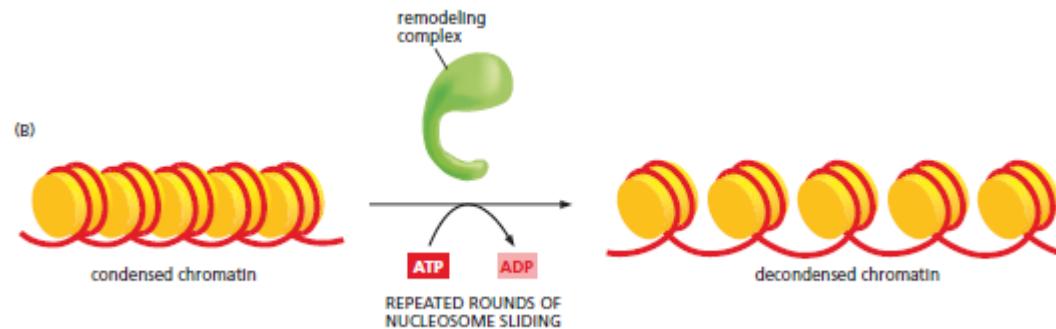


Chromosomes

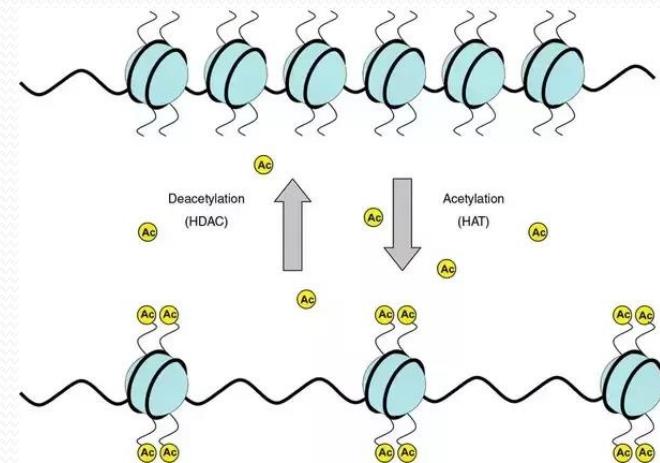
- Each chromosome has a constriction point called the **centromere**, which divides the chromosome into two sections, or “arms.”
- The short arm of the chromosome is labeled the “p arm.” The long arm of the chromosome is labeled the “q arm.”
- The location of the centromere on each chromosome gives the chromosome its characteristic shape, and can be used to help describe the location of specific genes (for example: 3p22.1).
- **Telomeres** are the region of DNA at the end of the linear eukaryotic chromosome that consist of repeat sequences and are required for the replication and stability of the chromosome.



- Amazingly, the DNA is compacted in a way that allows it to remain accessible to all of the enzymes and other proteins that replicate it, repair it, and control the expression of its genes.
- Eukaryotic cells have several ways to adjust the local structure of their chromatin rapidly.
- ***Chromatin-remodeling complexes***, protein machines that use the energy of ATP hydrolysis to change the position of the DNA wrapped around.
- The complexes, which attach to both the histone octamer and the DNA wrapped around it, can locally alter the arrangement of nucleosomes on the DNA, making the DNA either more accessible or less accessible to other proteins in the cell.

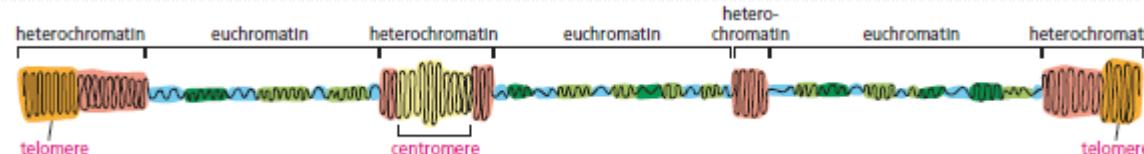


- Another way of altering chromatin structure relies on the reversible chemical modification of the histones. The tails of all four of the core histones are particularly subject to these covalent modifications.
- For example, acetyl, phosphate, or methyl groups can be added to and removed from the tails by enzymes that reside in the nucleus.
- Some of these modifications promote chromatin condensation, whereas others decondense chromatin and facilitate access to the DNA for replication, transcription or repair.
- These processes are reversible, so modified or remodeled chromatin can be returned to its compact state after transcription, replication or repair are complete.

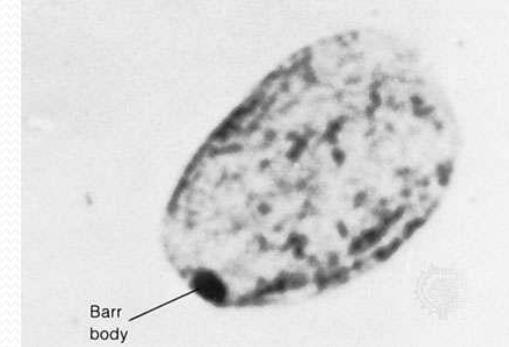


Euchromatin vs. Heterochromatin

- The localized alteration of chromatin packing by remodeling complexes and histone modification has important effects on the large-scale structure of interphase chromosomes.
- Interphase chromatin is not uniformly packed. Instead, regions of the chromosome that contain genes that are being expressed are generally more extended, while those that contain silent genes are more condensed.
- The most highly condensed form of interphase chromatin is called **heterochromatin**, and the rest of the interphase chromatin is called **euchromatin**.
- Most DNA that is permanently folded into heterochromatin in the cell does not contain genes. Because heterochromatin is so compact, genes that accidentally become packaged into heterochromatin usually fail to be expressed.



- When a cell divides, it generally passes on its histone modifications, chromatin structure, and gene expression patterns to the two daughter cells.
- Such “cell memory” is critical for the establishment and maintenance of different cell types during the development of a complex multicellular organism.
- Perhaps the most striking example of the use of heterochromatin to keep genes shut down, or *silenced*, is found in the interphase X chromosomes of female mammals.
- In mammals, female cells contain two X chromosomes, whereas male cells contain one X and one Y. Because a double dose of X-chromosome products would be lethal, female mammals have evolved a mechanism for permanently inactivating one of the two X chromosomes in each cell. At random, one or other of the two X chromosomes in each cell becomes highly condensed into heterochromatin early in embryonic development.

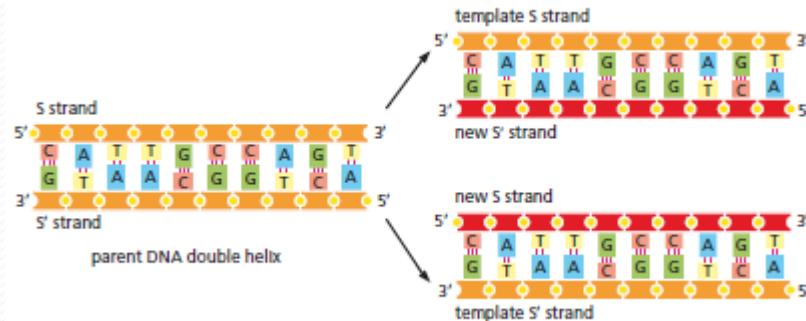


DNA Replication and Repair

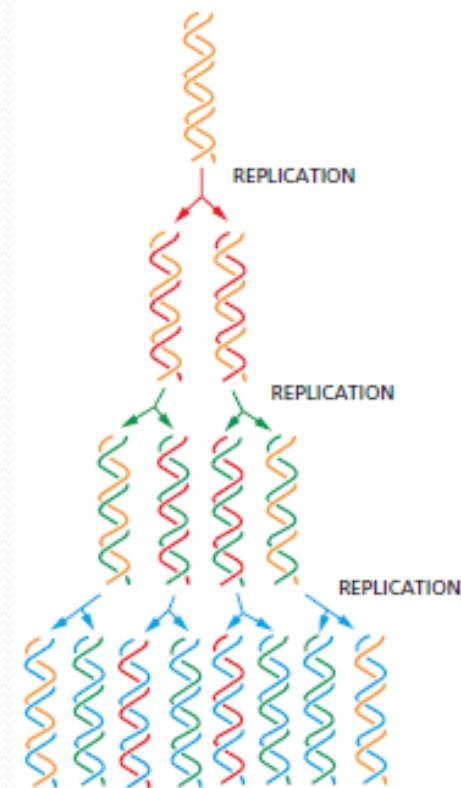
- The ability of a cell to survive and proliferate depends on the accurate duplication of the genetic information carried in its DNA.
- This duplication process, called *DNA replication*, must occur before a cell can divide to produce two genetically identical daughter cells.
- Maintaining order in a cell also requires the continual surveillance and *DNA repair*, as DNA is subjected to unavoidable damage by chemicals and radiation in the environment and by reactive molecules that are generated inside the cell.

DNA Replication

- Each strand of a DNA double helix contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand.
- Each strand can therefore serve as a template, or mold, for the synthesis of a new complementary strand.
- The ability of each strand of a DNA molecule to act as a template for producing a complementary strand enables a cell to copy, or *replicate*, its genes before passing them on to its descendants.



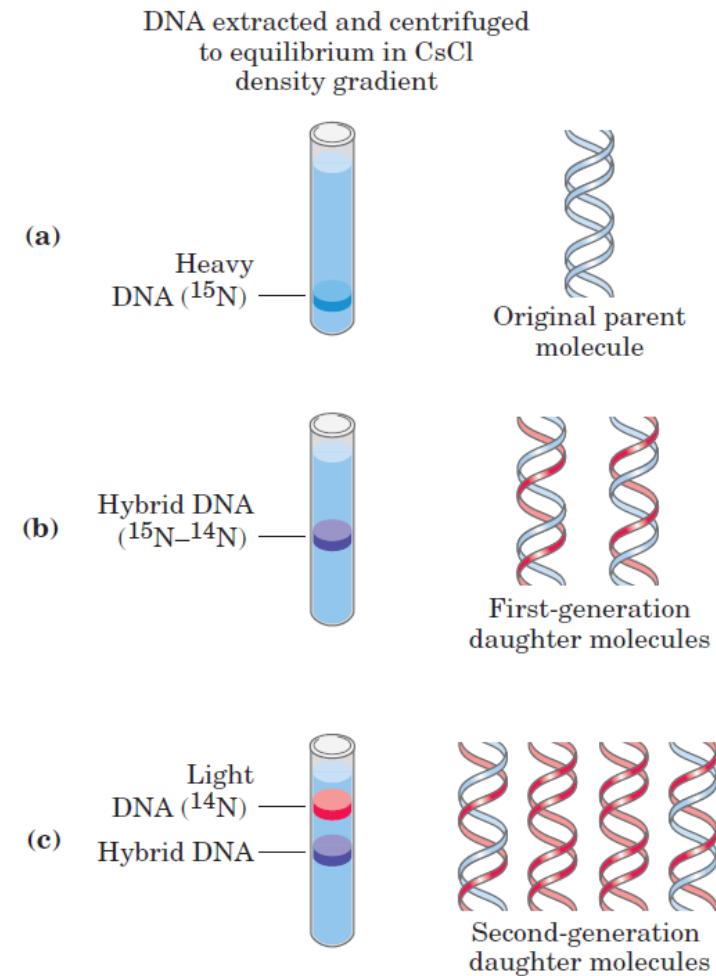
- DNA replication produces two complete double helices from the original DNA molecule, with each new DNA helix being identical in nucleotide sequence to the original DNA double helix.
- Because each parental strand serves as the template for one new strand, each of the daughter DNA double helices ends up with one of the original (old) strands plus one strand that is completely new; this style of replication is called ***semiconservative replication***.



The Meselson-Stahl Experiment

- Although Watson and Crick proposed the hypothesis of semiconservative replication in 1953, the hypothesis was proved by ingeniously designed experiments carried out by Matthew Meselson and Franklin Stahl in 1957.

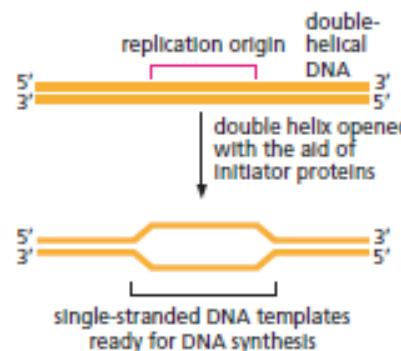
Cells were grown for many generations in a medium containing only heavy nitrogen, ^{15}N , so that all the nitrogen in their DNA was ^{15}N , as shown by a single band (blue) when centrifuged in a CsCl density gradient. (b) Once the cells had been transferred to a medium containing only light nitrogen, ^{14}N , cellular DNA isolated after one generation equilibrated at a higher position in the density gradient (purple band). (c) Continuation of replication for a second generation yielded two hybrid DNA and two light DNAs (red), confirming semiconservative replication.



Origin of Replication

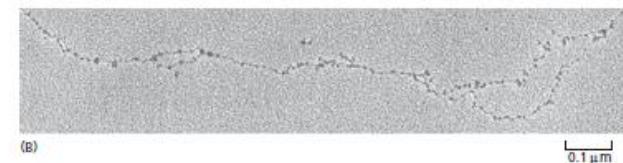
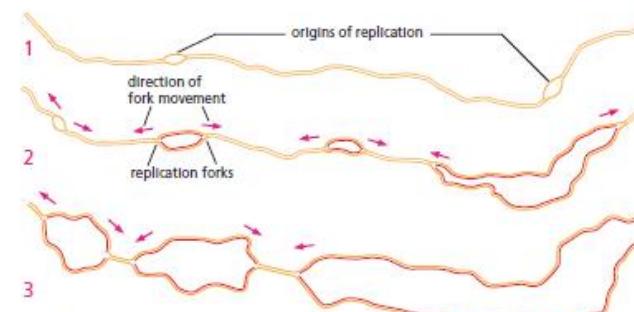
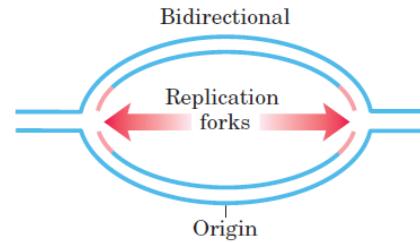
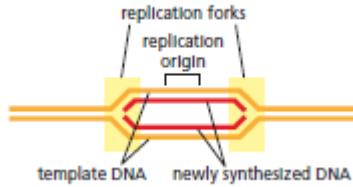
- The DNA double helix is normally very stable: the two DNA strands are locked together firmly by the large numbers of hydrogen bonds between the bases on both strands.
- As a result, only temperatures approaching those of boiling water provide enough thermal energy to separate the two strands.
- To be used as a template, however, the double helix must first be opened up and the two strands separated to expose unpaired bases.
- The process of DNA synthesis is begun by *initiator proteins* that bind to specific DNA sequences called ***replication origins***. Here, the initiator proteins pry the two DNA strands apart, breaking the hydrogen bonds between the bases.

- Origins of replication are composed of DNA sequences that attract the initiator proteins and are especially easy to open.
- Since A-T base pair is held together by fewer hydrogen bonds G-C base pair DNA rich in A-T base pairs is relatively easy to pull apart, and A-T-rich stretches of DNA are typically found at replication origins.
- A bacterial genome, which is typically contained in a circular DNA molecule of several million nucleotide pairs, has a single replication origin.
- The human genome, which is very much larger, has approximately 10,000 such origins—an average of 220 origins per chromosome. Beginning DNA replication at many places at once greatly shortens the time a cell needs to copy its entire genome.
- Once an initiator protein binds to DNA at a replication origin and locally opens up the double helix, it attracts a group of proteins that carry out DNA replication. These proteins form a replication machine, in which each protein carries out a specific function.

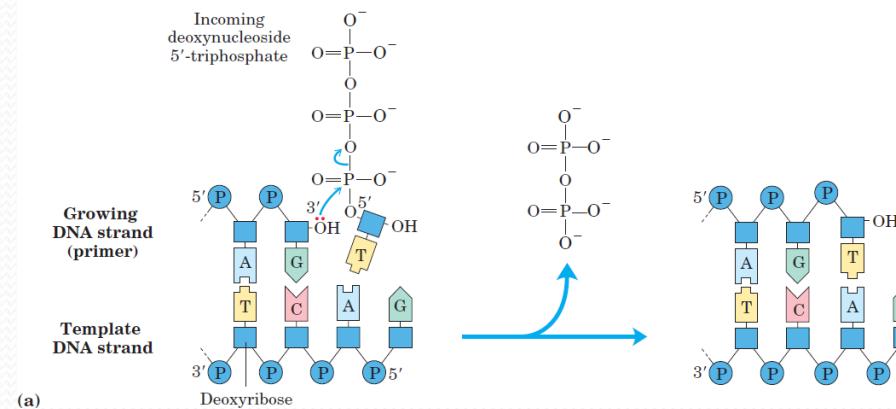
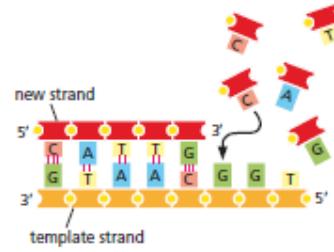


Replication Forks

- DNA molecules in the process of being replicated contain Y-shaped junctions called ***replication forks***. Two replication forks are formed at each replication origin.
- At each fork, a replication machine moves along the DNA, opening up the two strands of the double helix and using each strand as a template to make a new daughter strand.
- The two forks move away from the origin in opposite directions, unzipping the DNA double helix and replicating the DNA as they go.
- DNA replication in bacterial and eukaryotic chromosomes is therefore termed ***bidirectional***.

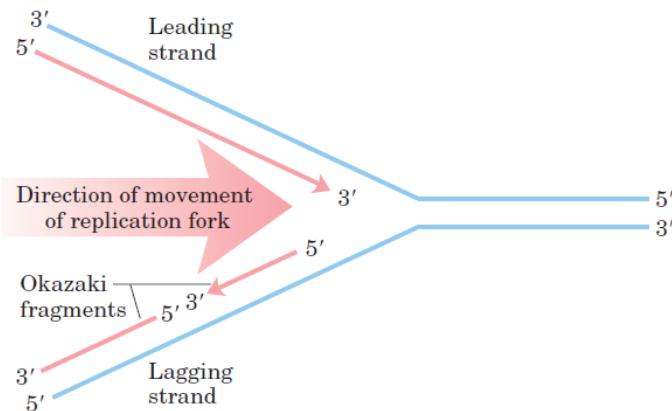


- The movement of a replication fork is driven by the action of the replication machine, at the heart of which is an enzyme called **DNA polymerase**.
- This enzyme catalyzes the addition of nucleotides to the 3' end of a growing DNA strand, using one of the original, parental DNA strands as a template.
- Base pairing between an incoming nucleotide and the template strand determines which of the four nucleotides (A, G, T, or C) will be selected.
- The final product is a new strand of DNA that is complementary in nucleotide sequence to the template.
- **The polymerization reaction involves the formation of a phosphodiester bond between the 3' end of the growing DNA chain and the 5'-phosphate group of the incoming nucleotide**, which enters the reaction as a *deoxyribonucleoside triphosphate*.
- The energy for polymerization is provided by the incoming deoxyribonucleoside triphosphate itself: hydrolysis of one of its high-energy phosphate bonds fuels the reaction that links the nucleotide monomer to the chain.
- DNA polymerase moves along the template strand as it repeatedly adds new nucleotides to the growing strand.



- The 5'-to-3' direction of the DNA polymerization reaction poses a problem at the replication fork.
- The sugar–phosphate backbone of each strand of a DNA double helix has a unique chemical direction, or polarity, determined by the way each sugar residue is linked to the next, and the two strands in the double helix are antiparallel; that is, they run in opposite directions.
- As a consequence, at each replication fork, one new DNA strand is being made on a template that runs in one direction (3' to 5'), whereas the other new strand is being made on a template that runs in the opposite direction (5' to 3').
- The replication fork is therefore asymmetrical.
- However, all DNA polymerases add new subunits only to the 3' end of a DNA strand. As a result, a new DNA chain can be synthesized only in a 5'-to-3' direction.
- Then, how is 3'-to-5' strand is replicated?

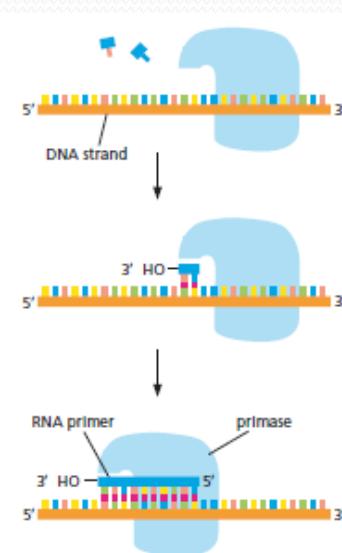
- The DNA strand that appears to grow in 3'-to-5' direction is actually made *discontinuously*, in successive, separate, small pieces—with the DNA polymerase moving backward with respect to the direction of replication-fork movement so that each new DNA fragment can be polymerized in the 5'-to-3' direction.
- The resulting small DNA pieces—called **Okazaki fragments** after the biochemists who discovered them—are later joined together to form a continuous new strand. The DNA strand that is made discontinuously in this way is called **the lagging strand**; the other strand, which is synthesized continuously, is called **the leading strand**.



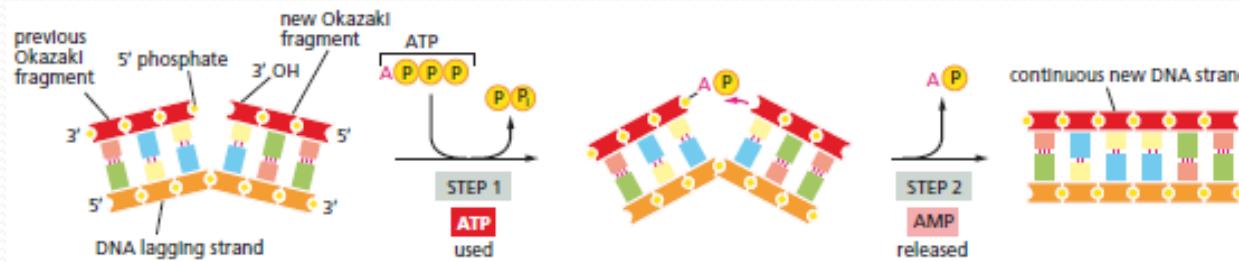
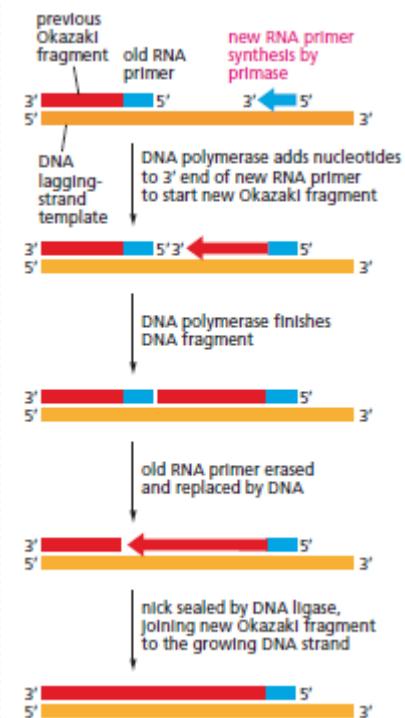
The need for primers

- The accuracy of DNA replication depends on the requirement of the DNA polymerase for a correctly base-paired 3' end before it can add more nucleotides to a growing DNA strand.
- Then, to begin a completely new DNA strand polymerase needs a different enzyme—one that can begin a new polynucleotide strand simply by joining two nucleotides together without the need for a base-paired end.
- This enzyme does not, however, synthesize DNA. It makes a short length of a closely related type of nucleic acid—RNA (ribonucleic acid)—using the DNA strand as a template.
- This short length of RNA, about 10 nucleotides long, is base-paired to the template strand and provides a base-paired 3' end as a starting point for DNA polymerase. It thus serves as a **primer** for DNA synthesis, and the enzyme that synthesizes the RNA primer is known as **primase**.

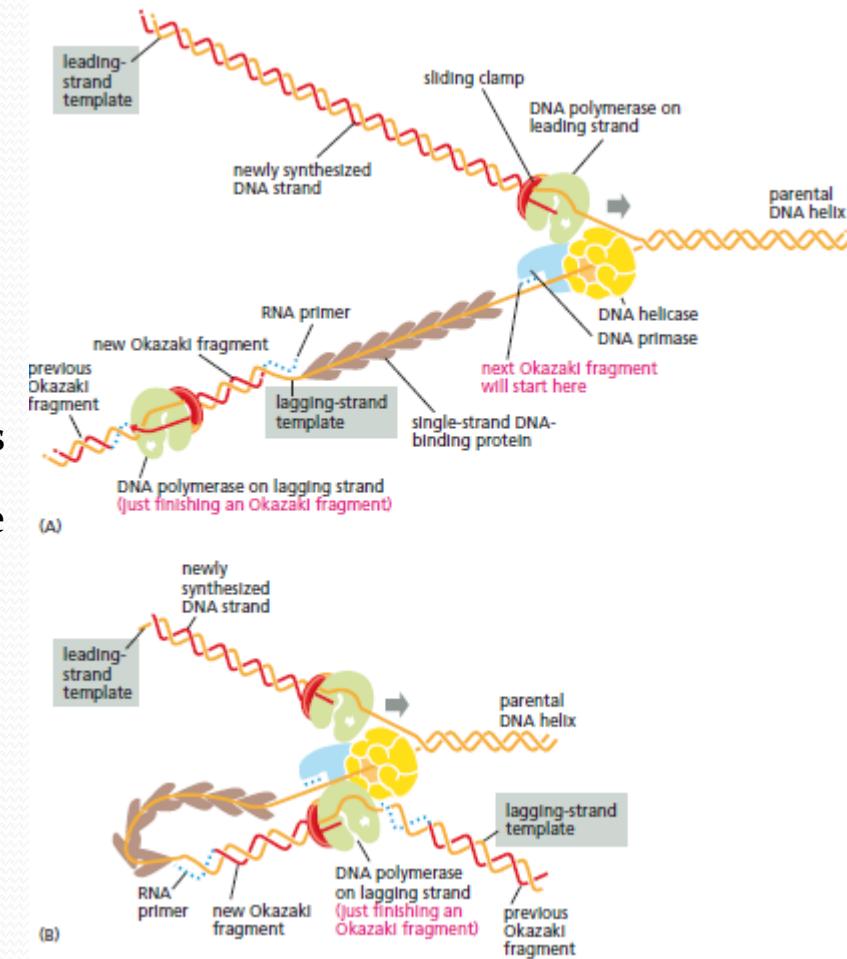
- Primase is a type of *RNA polymerase* that synthesizes RNA using DNA as a template.
- A strand of RNA is very similar chemically to a single strand of DNA except that it is made of ribonucleotide subunits, in which the sugar is ribose, not deoxyribose; RNA also differs from DNA in that it contains the base uracil (U) instead of thymine (T). Because U can form a base pair with A, the RNA primer is synthesized on the DNA strand by complementary base-pairing in exactly the same way as is DNA.
- For the leading strand, an RNA primer is needed only to start replication at a replication origin; once a replication fork has been established, the DNA polymerase is continuously presented with a base-paired 3' end as it tracks along the template strand.



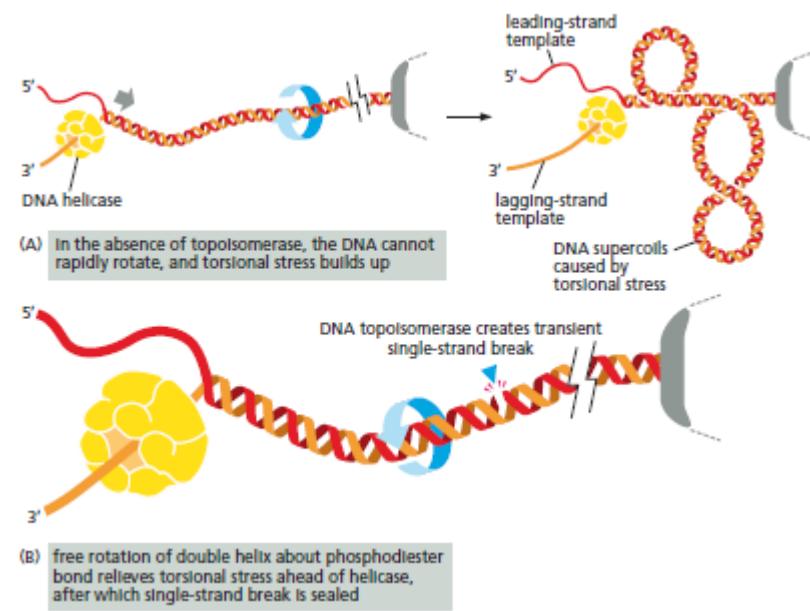
- But on the lagging strand, where DNA synthesis is discontinuous, new primers are needed to keep polymerization going.
- The movement of the replication fork continually exposes unpaired bases on the lagging strand template, and new RNA primers are laid down at intervals along the newly exposed, single-stranded stretch.
- DNA polymerase adds a deoxyribonucleotide to the 3' end of each primer to start a new Okazaki fragment, and it will continue to elongate this fragment until it runs into the next RNA primer.
- To produce a continuous new DNA strand from the many separate pieces of nucleic acid made on the lagging strand, three additional enzymes are needed. These act quickly to remove the RNA primer, replace it with DNA, and join the DNA fragments together.
- Thus, a nuclease degrades the RNA primer, a DNA polymerase called a **repair polymerase** then replaces this RNA with DNA (using the end of the adjacent Okazaki fragment as a primer), and the enzyme **DNA ligase** joins the 5'-phosphate end of one DNA fragment to the adjacent 3'-hydroxyl end of the next.



- DNA replication requires the cooperation of a large number of proteins that act in concert to open up the double helix and synthesize new DNA.
- These proteins form part of a remarkably complex replication machine.
- The first problem faced by the replication machine is accessing the nucleotides that lie at the center of the helix. For DNA replication to occur, the double helix must be unzipped ahead of the replication fork so that the incoming nucleoside triphosphates can form base pairs with each template strand. Two types of replication proteins—*DNA helicases* and *single-strand DNA-binding proteins*—cooperate to carry out this task.
- The **helicase** sits at the very front of the replication machine tearing apart the double helix as it speeds along the DNA.
- ***Single strand DNA-binding proteins*** cling to the single-stranded DNA exposed by the helicase, transiently preventing the strands from re-forming base pairs and keeping them in an elongated form so that they can serve as efficient templates.



- This localized unwinding of the DNA double helix itself presents a problem. As the helicase opens the DNA within the replication fork, DNA on the other side of the fork gets wound more tightly.
- This excess twisting in front of the replication fork creates tension in the DNA that—if allowed to build—makes unwinding the double helix increasingly difficult and impedes the forward movement of the replication machinery.
- Cells use proteins called **DNA topoisomerases** to relieve this tension. These enzymes produce transient nicks in the DNA backbone, which temporarily release the tension; they then reseal the nick before falling off the DNA.



- An additional replication protein, called a ***sliding clamp***, keeps DNA polymerase firmly attached to the template while it is synthesizing new strands of DNA.
- Left on their own, most DNA polymerase molecules will synthesize only a short string of nucleotides before falling off the DNA template strand.
- The sliding clamp forms a ring around the newly formed DNA double helix and, by tightly gripping the polymerase, allows the enzyme to move along the template strand without falling off as it synthesizes new DNA.

- Replicating the very ends of chromosomes also represents a special problem because DNA replication proceeds only in the 5'-to-3' direction, the lagging strand of the replication fork has to be synthesized in the form of discontinuous DNA fragments, each of which is primed with an RNA primer laid down by a primase.
- As the replication fork approaches the end of a chromosome, although the leading strand can be replicated all the way to the chromosome tip, the lagging strand cannot.
- When the final RNA primer on the lagging strand is removed, there is no way to replace it.
- Without a strategy to deal with this problem, the lagging strand would become shorter with each round of DNA replication; after repeated cell divisions, chromosomes would shrink—and eventually lose valuable genetic information.

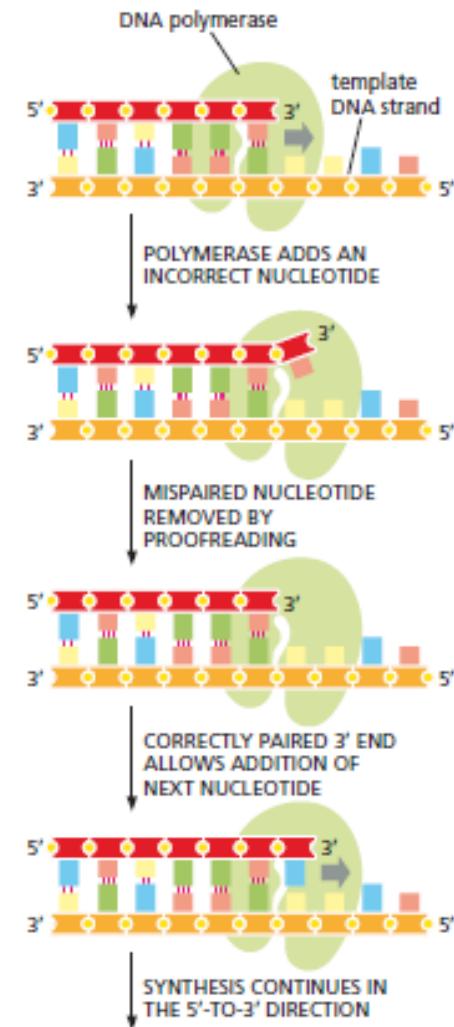
- Bacteria solve this “end-replication” problem by having circular DNA molecules as chromosomes.
- Eukaryotes solve it by having long, repetitive nucleotide sequences at the ends of their chromosomes which are incorporated into structures called **telomeres**.
- These telomeric DNA sequences attract an enzyme called telomerase to the chromosome ends.
- Using an RNA template that is part of the enzyme itself, telomerase extends the ends of the replicating lagging strand by adding multiple copies of the same short DNA sequence to the template strand.
- This extended template allows replication of the lagging strand to be completed by conventional DNA replication.

- Most of the proteins involved in DNA replication are held together in a large multienzyme complex that moves as a unit along the parental DNA double helix, enabling DNA to be synthesized on both strands in a coordinated manner.
- The replication machinery does its work so good and DNA polymerase is so accurate that it makes only about one error in every 10^7 nucleotide pairs it copies.
- This error rate is much lower than can be explained simply by the accuracy of complementary base-pairing.
- Although A-T and C-G are by far the most stable base pairs, other, less stable base pairs—for example, G-T and C-A—can also be formed.
- Such incorrect base pairs are formed much less frequently than correct ones, but, if allowed to remain, they would result in an accumulation of mutations.

DNA Proofreading

- This disaster is avoided because DNA polymerase has two special qualities that greatly increase the accuracy of DNA replication.
- First, the enzyme carefully monitors the base-pairing between each incoming nucleotide and the template strand. Only when the match is correct does DNA polymerase catalyze the nucleotide-addition reaction.
- Second, when DNA polymerase makes a rare mistake and adds the wrong nucleotide, it can correct the error through an activity called *proofreading*.

- Proofreading takes place at the same time as DNA synthesis. Before the enzyme adds the next nucleotide to a growing DNA strand, it checks whether the previously added nucleotide is correctly base-paired to the template strand.
- If so, the polymerase adds the next nucleotide; if not, the polymerase clips off the mispaired nucleotide and tries again.
- This proofreading is carried out by cleaving the phosphodiester backbone using the enzyme's 3'-to-5' exonuclease activity.



DNA Repair

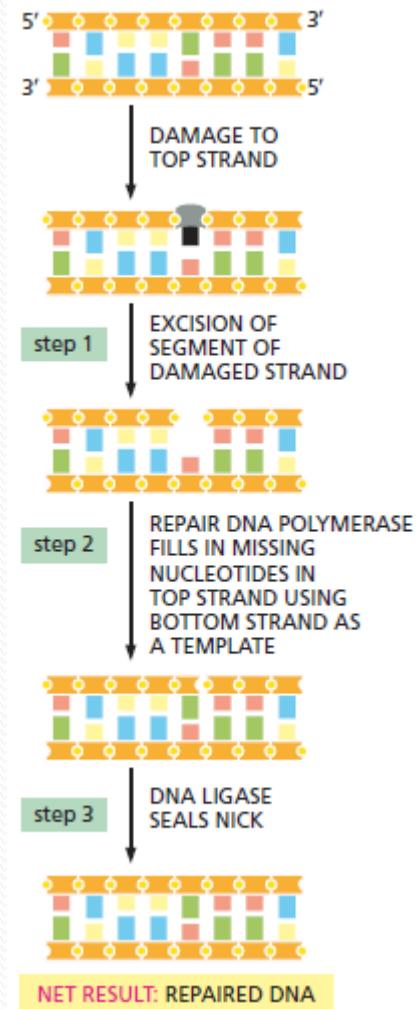
- The diversity of living organisms and their success in colonizing almost every part of the Earth's surface depend on genetic changes accumulated gradually over millions of years.
- Some of these changes allow organisms to adapt to changing conditions and to thrive in new habitats. However, in the short term, and from the perspective of an individual organism, genetic alterations can be detrimental.
- In a multicellular organism, such permanent changes in the DNA—called ***mutations***—can upset the organism's extremely complex and finely tuned development and physiology.
- To survive and reproduce, individuals must be genetically stable. This stability is achieved not only through the extremely accurate mechanism for replicating DNA, but also through the work of a variety of protein machines that continually scan the genome for damage and fix it when it occurs.
- Most DNA damage is only temporary, because it is immediately corrected by processes collectively called DNA repair.

Cause of mutations in DNA

- DNA is continually undergoing thermal collisions with other molecules, often resulting in major chemical changes in the DNA.
- Some chemically reactive by-products of cell metabolism also occasionally react with the bases in DNA, altering them in such a way that their base-pairing properties are changed.
- The ultraviolet radiation in sunlight is also damaging to DNA.
- DNA can also be altered by replication itself. The replication machinery can—quite rarely—incorporate an incorrect nucleotide that it fails to correct via proofreading.

- Thanks to DNA repair mechanism such changes are repaired.
- Repair mechanisms depend on the double-helical structure of DNA, which provides two copies of the genetic information—one in each strand of the double helix. Thus, if the sequence in one strand is accidentally damaged, information is not lost, because a backup version of the altered strand remains in the complementary sequence of nucleotides in the other strand.
- Most DNA damage creates structures that are never encountered in an undamaged DNA strand; thus the good strand is easily distinguished from the bad.

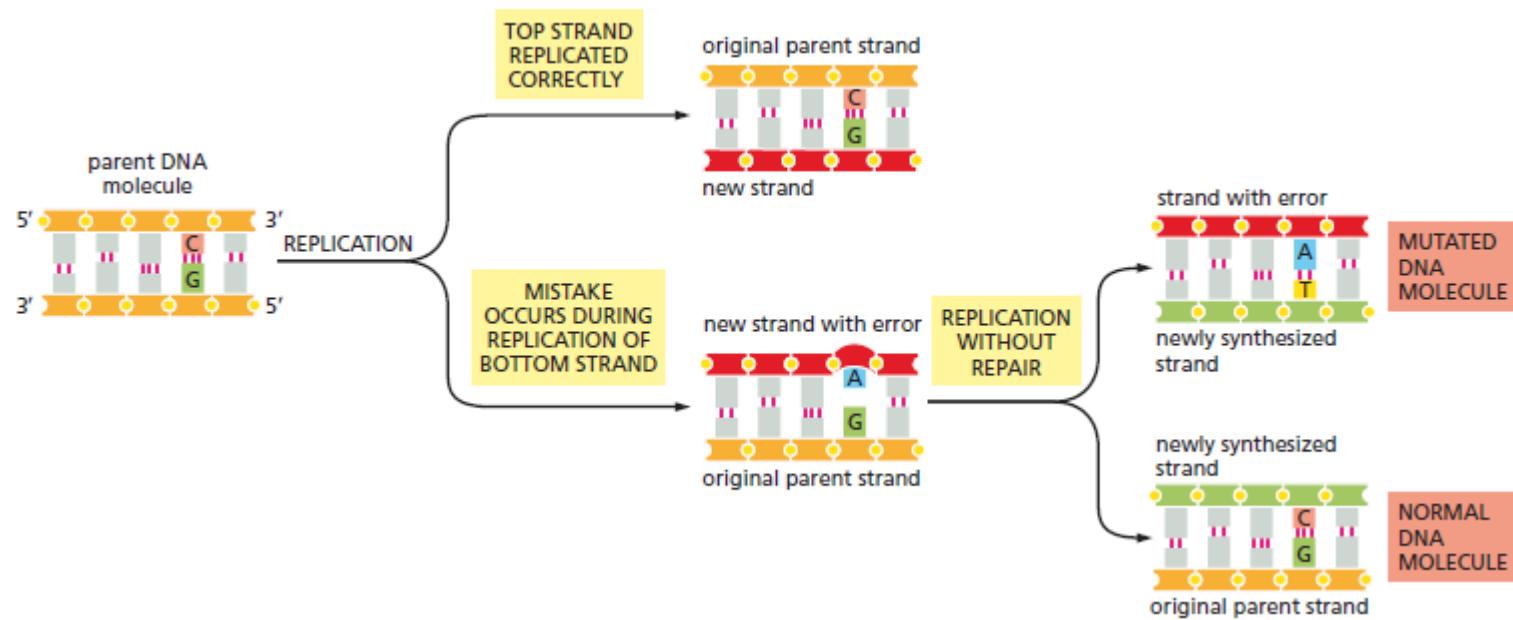
- The basic pathway for repairing damage to DNA involves three basic steps:
 1. The damaged DNA is recognized and removed by one of a variety of mechanisms. These involve nucleases, which cleave the covalent bonds that join the damaged nucleotides to the rest of the DNA strand, leaving a small gap on one strand of the DNA double helix in the region.
 2. A *repair DNA polymerase* binds to the 3'-hydroxyl end of the cut DNA strand. It then fills in the gap by making a complementary copy of the information stored in the undamaged strand.
 3. When the repair DNA polymerase has filled in the gap, a break remains in the sugar–phosphate backbone of the repaired strand. This nick in the helix is sealed by DNA ligase, the same enzyme that joins the Okazaki fragments during replication of the lagging DNA strand.



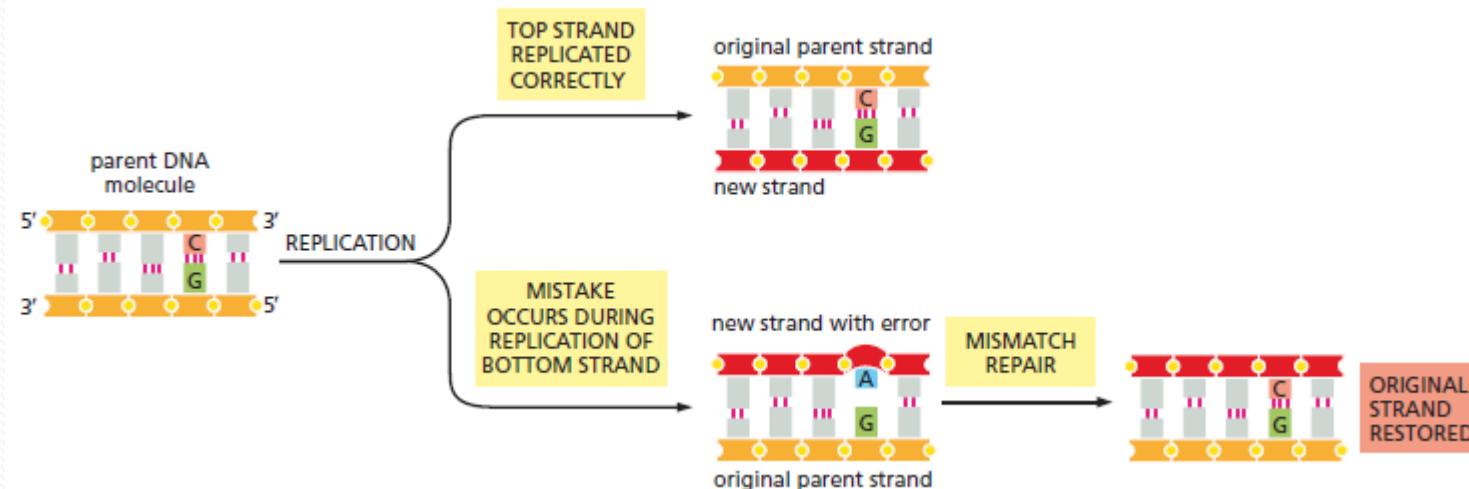
Mismatch Repair

- Although the high fidelity and proofreading abilities of the cell's replication machinery generally prevent replication errors from occurring, rare mistakes do happen. Fortunately, the cell has a backup system—called ***mismatch repair***—which is dedicated to correcting these errors.
- The replication machine makes approximately one mistake per 10^7 nucleotides copied; DNA mismatch repair corrects 99% of these replication errors, increasing the overall accuracy to one mistake in 10^9 nucleotides copied.
- Thus, DNA mismatch repair system removes replication errors that escape proofreading.

- Whenever the replication machinery makes a copying mistake, it leaves behind a mispaired nucleotide (commonly called a *mismatch*).
 - If left uncorrected, the mismatch will result in a permanent mutation in the next round of DNA replication.



- A complex of mismatch repair proteins recognizes such a DNA mismatch, removes a portion of the DNA strand containing the error, and then resynthesizes the missing DNA. This repair mechanism restores the correct sequence.
- To be effective, the mismatch repair system must be able to recognize which of the DNA strands contains the error. Removing a segment from the strand of DNA that contains the correct sequence would only compound the mistake.
- The way the mismatch system solves this problem is by always removing a portion of the newly made DNA strand. In bacteria, newly synthesized DNA lacks a type of chemical modification that is present on the preexisting parent DNA. Other cells use other strategies for distinguishing their parent DNA from a newly replicated strand.



Repair of Double Strand Breaks

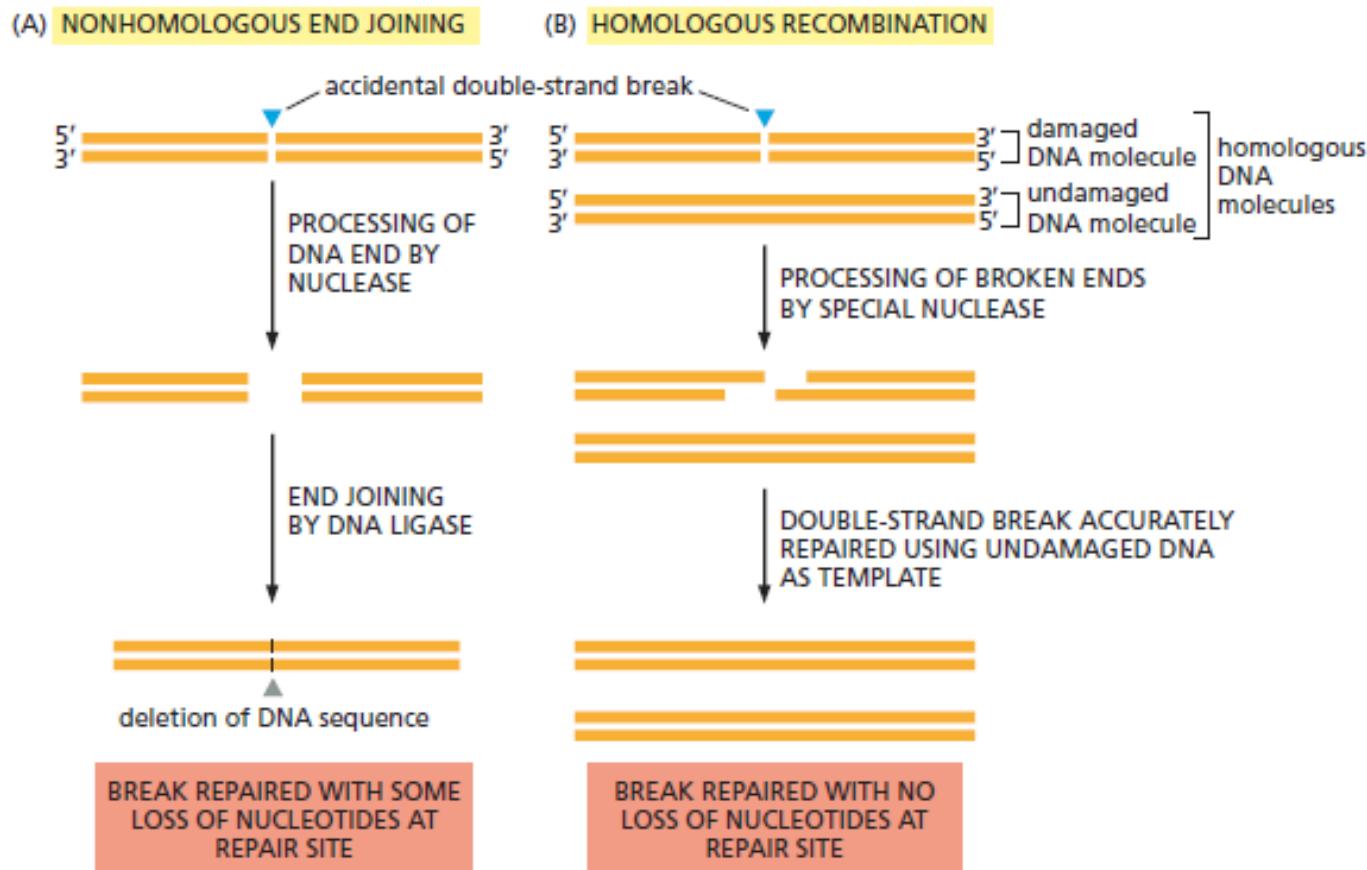
- Radiation, mishaps at the replication fork, and various chemical assaults can all fracture the backbone of DNA, creating a *double-strand break*.
- Such lesions are particularly dangerous, because they can lead to the fragmentation of chromosomes and the subsequent loss of genes.
- This type of damage is especially difficult to repair because both strands of the double helix are damaged at the same time and there is no intact template to guide the repair.

Nonhomologous End Joining

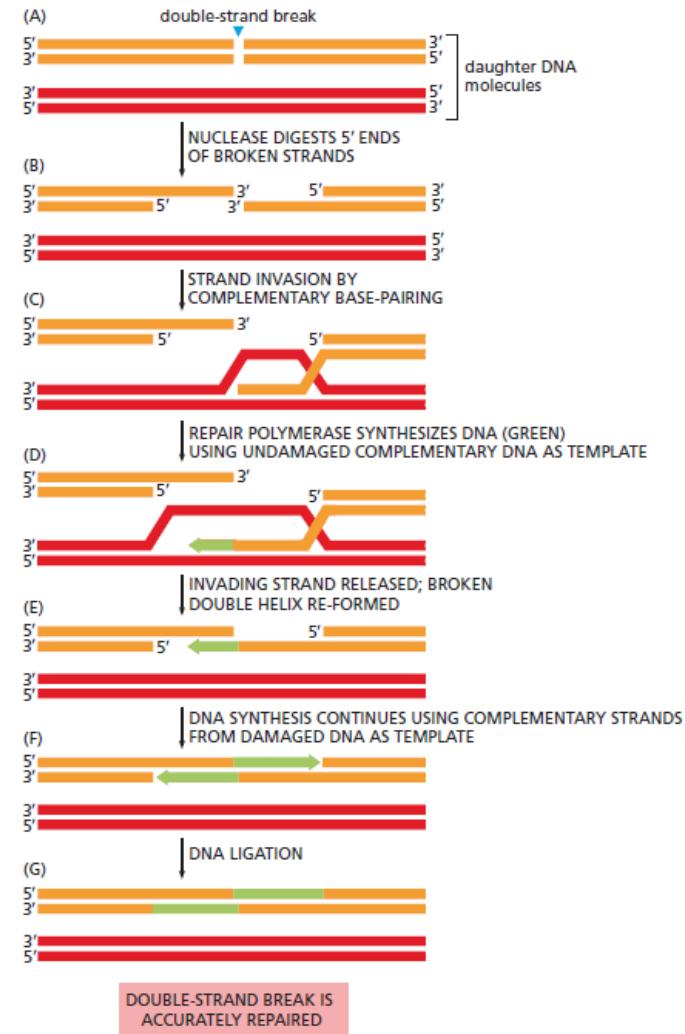
- To handle this potentially disastrous type of DNA damage, cells have evolved two basic strategies: Nonhomologous end joining and homologous recombination.
- The first involves rapidly sticking the broken ends back together, before the DNA fragments drift apart and get lost.
- This repair mechanism, called nonhomologous end joining, occurs in many cell types and is carried out by a specialized group of enzymes that “clean” the broken ends and rejoin them by DNA ligation.
- This “quick and dirty” mechanism rapidly repairs the damage, but it comes with a price: in “cleaning” the break to make it ready for ligation, nucleotides are often lost at the site of repair .
- In most cases, this emergency repair mechanism mends the damage without creating any additional problems.
- But if the imperfect repair disrupts the activity of a gene, the cell could suffer serious consequences. Thus, nonhomologous end joining can be a risky strategy for fixing broken chromosomes.

Homologous Recombination

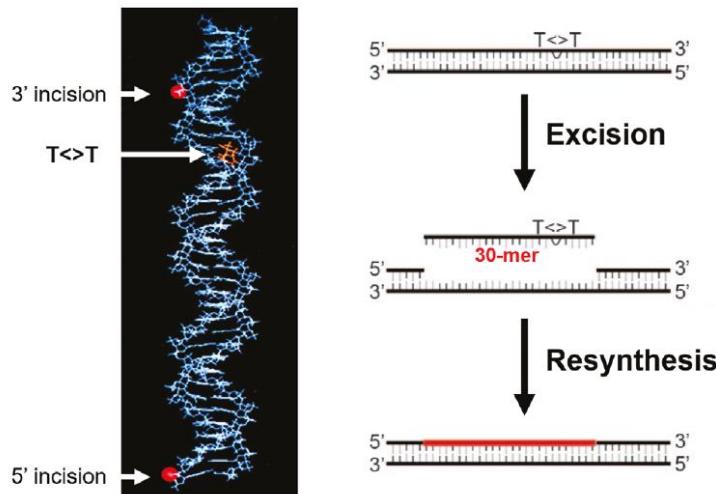
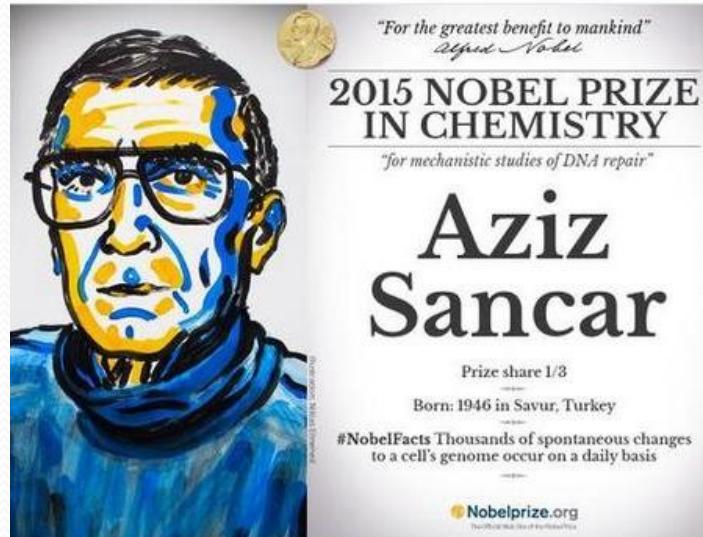
- The problem with repairing a double-strand break, as we mentioned, is finding an intact template to guide the repair.
- However, if a doublestrand break occurs in one double helix shortly after a stretch of DNA has been replicated, the undamaged double helix can readily serve as a template to guide the repair of the broken DNA: information on the undamaged strand of the intact double helix is used to repair the complementary broken strand in the other. Because the two DNA molecules are homologous—they have identical nucleotide sequences outside the broken region—this mechanism is known as homologous recombination.
- It results in a flawless repair of the double-strand break, with no loss of genetic information.



- Homologous recombination most often occurs shortly after a cell's DNA has been replicated before cell division, when the duplicated helices are still physically close to each other.
- To initiate the repair, a nuclease chews back the 5' ends of the two broken strands at the break.
- Then, with the help of specialized enzymes, one of the broken 3' ends "invades" the unbroken homologous DNA duplex and searches for a complementary sequence through base-pairing.
- Once an extensive, accurate match is found, the invading strand is elongated by a repair DNA polymerase, using the complementary strand as a template.
- After the repair polymerase has passed the point where the break occurred, the newly repaired strand rejoins its original partner, forming base pairs that hold the two strands of the broken double helix together. Repair is then completed by additional DNA synthesis at the 3' ends of both strands of the broken double helix, followed by DNA ligation.



- On occasion, the cell’s DNA replication and repair processes fail and give rise to a *mutation*. This permanent change in the DNA sequence can have profound consequences.
- A mutation that affects just a single nucleotide pair can severely compromise an organism’s fitness if the change occurs in a vital position in the DNA sequence.
- Because the structure and activity of each protein depend on its amino acid sequence, a protein with an altered sequence may function poorly or not at all. The many other cells in a multicellular organism (its *somatic cells*) must also be protected against mutation—in this case, against mutations that arise during the life of an individual.
- Nucleotide changes that occur in somatic cells can give rise to variant cells, some of which grow and divide in an uncontrolled fashion at the expense of the other cells in the organism.
- In the extreme case, an unchecked cell proliferation known as cancer results. Thus, the high fidelity with which DNA sequences are replicated and maintained is important both for reproductive cells, which transmit the genes to the next generation, and for somatic cells, which normally function as carefully regulated members of the complex community of cells in a multicellular organism.
- We should therefore not be surprised to find that all cells possess a very sophisticated set of mechanisms to reduce the number of mutations that occur in their DNA, devoting hundreds of genes to these repair processes.



Nobel Prize winner in Chemistry in 2015 by his work on Nucleotide Excision Repair (NER) mechanism of UV damaged DNA.