

Chem 419/511

Biomolecular Chemistry

1st half of 419 and 511: Biomolecular chemistry
2nd half of 419 and 519: Bioanalytical methods

Biomolecular chemistry

1. DNA structure and replication

Primary Source Material

- Molecular Cell Biology (Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E.) (courtesy of the NCBI bookshelf)
- Biochemistry (Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert), courtesy of the NCBI bookshelf
- Many figures and the descriptions for the figures are from the educational resources provided at the Protein Data Bank (<http://www.pdb.org/>)
- Most of these figures and accompanying legends have been written by David S. Goodsell of the Scripps Research Institute and are being used with permission. I highly recommend browsing the Molecule of the Month series at the PDB (http://www.pdb.org/pdb/101/motm_archive.do)

A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus

Felisa Wolfe-Simon,^{1,2*} Jodi Switzer Blum,² Thomas R. Kulp,² Gwyneth W. Gordon,³ Shelley E. Hoefft,² Jennifer Pett-Ridge,⁴ John F. Stolz,⁵ Samuel M. Webb,⁶ Peter K. Weber,⁴ Paul C. W. Davies,^{1,7} Ariel D. Anbar,^{1,3,8} Ronald S. Oremland²

Life is mostly composed of the elements carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus. Although these six elements make up nucleic acids, proteins, and lipids and thus the bulk of living matter, it is theoretically possible that some other elements in the periodic table could serve the same functions. Here, we describe a bacterium, strain GFAJ-1 of the Halomonadaceae, isolated from Mono Lake, California, that is able to substitute arsenic for phosphorus to sustain its growth. Our data show evidence for arsenate in macromolecules that normally contain phosphate, most notably nucleic acids and proteins. Exchange of one of the major bio-elements may have profound evolutionary and geochemical importance.

470

27 JULY 2012 VOL 337 SCIENCE

Absence of Detectable Arsenate in DNA from Arsenate-Grown GFAJ-1 Cells

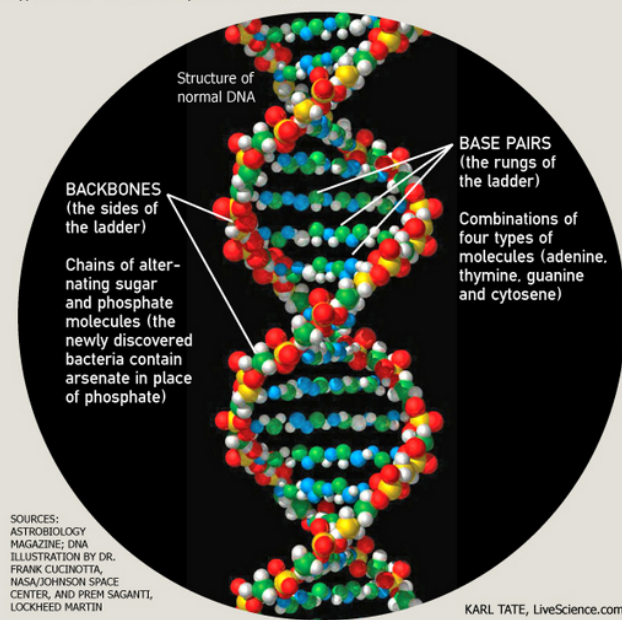
Marshall Louis Reaves,^{1,2} Sunita Sinha,³ Joshua D. Rabinowitz,^{1,4} Leonid Kruglyak,^{1,5,6} Rosemary J. Redfield^{3*}

A strain of *Halomonas* bacteria, GFAJ-1, has been claimed to be able to use arsenate as a nutrient when phosphate is limiting and to specifically incorporate arsenic into its DNA in place of phosphorus. However, we have found that arsenate does not contribute to growth of GFAJ-1 when phosphate is limiting and that DNA purified from cells grown with limiting phosphate and abundant arsenate does not exhibit the spontaneous hydrolysis expected of arsenate ester bonds. Furthermore, mass spectrometry showed that this DNA contains only trace amounts of free arsenate and no detectable covalently bound arsenate.



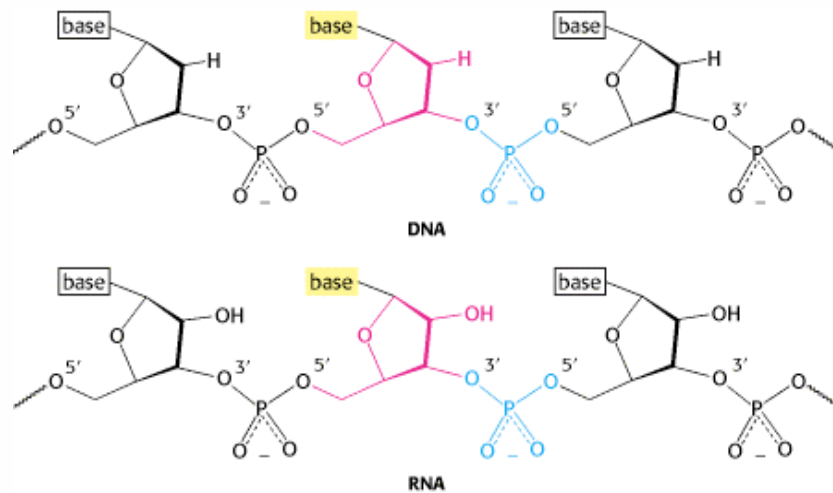
DNA, but not as we knew it

Scientists have discovered bacteria based on a previously unknown kind of DNA molecule. In place of phosphorus, these molecules contain **arsenic**, which is a deadly poison to normal types of life – but the newly discovered bacteria thrive on it.



The polymeric structure of nucleic acids (DNA and RNA)

3



Why is RNA more prone to hydrolysis?

- DNA = deoxyribonucleic acid, RNA = ribonucleic acid
- The sugars in nucleic acids are linked to one another by phosphodiester bridges. Specifically, the 3'-hydroxyl group of the sugar moiety of one nucleotide is linked to a phosphate group via an ester bond, which is, in turn, joined to the 5'-hydroxyl group of the adjacent sugar via an ester bond. The chain of sugars linked by phosphodiester bridges is referred to as the backbone of the nucleic acid.
- Both RNA and DNA are long unbranched polymers consisting of nucleotides joined by 3'-phosphodiester bonds. The covalent structure of RNA differs from that of DNA in two respects. As stated earlier and as indicated by its name, the sugar units in RNA are riboses rather than deoxyriboses. Ribose contains a 2'-hydroxyl group not present in deoxyribose. As a consequence, in addition to the standard 3'-linkage, a 2'-linkage is possible for RNA. This later linkage is important in the removal of introns and the joining of exons for the formation of mature RNA. The other difference between RNA and DNA is that one of the four major bases in RNA is uracil (U) instead of thymine (T).
- Note that each phosphodiester bridge has a negative charge. This negative charge repels nucleophilic species such as hydroxide ion; consequently, phosphodiester linkages are much less susceptible to hydrolytic attack than are other esters such as carboxylic acid esters. This resistance is crucial for maintaining the integrity of information stored in nucleic acids. The absence of the 2'-hydroxyl group in DNA further increases its resistance to hydrolysis. The greater stability of DNA probably accounts for its use (rather than RNA) as the hereditary material in all modern cells and in many viruses.
- *Q: I am wondering what is the 5' end and 3' end of an oligonucleotide? Is the 3' end remaining -OH group but 5' end phosphorylated ?*
- *A: There could, in principle, be either a hydroxyl or a phosphate at either end. The name 3' and 5' refers to the position within the terminal sugar that is not connected to the rest of the chain. That is, it is the 'free' hydroxyl (or phosphate) of the terminal sugar.*

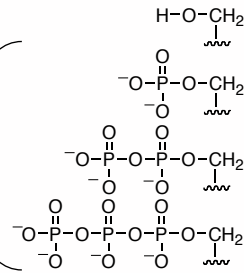
The structure of nucleosides and nucleotides

4

5' modifications

Nucleoside

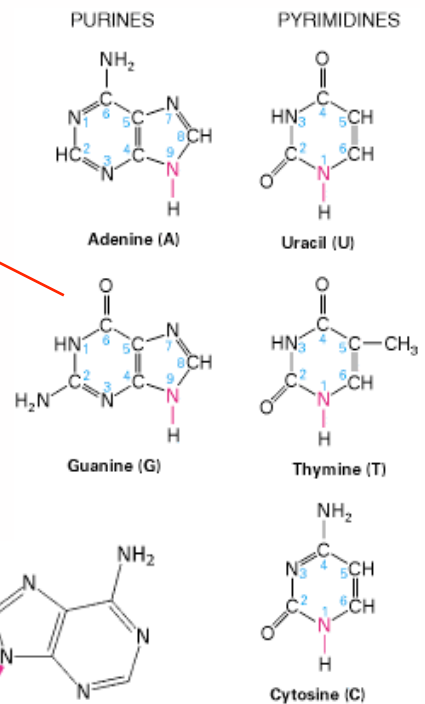
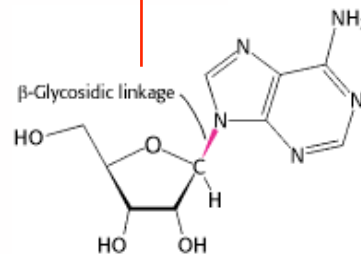
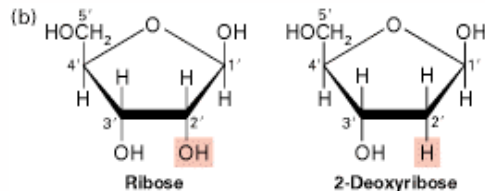
Nucleotides
(or nucleoside
mono/di/tri-
phosphates)



3' modifications

3'-PO₄⁻²

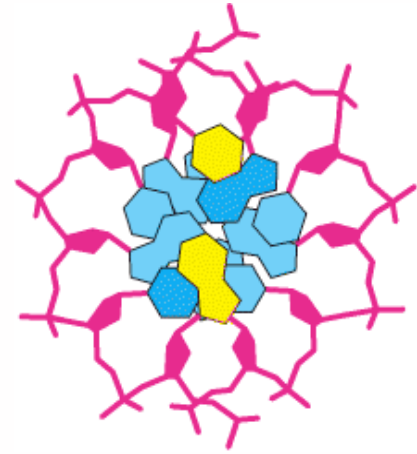
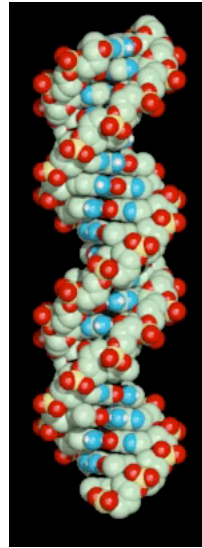
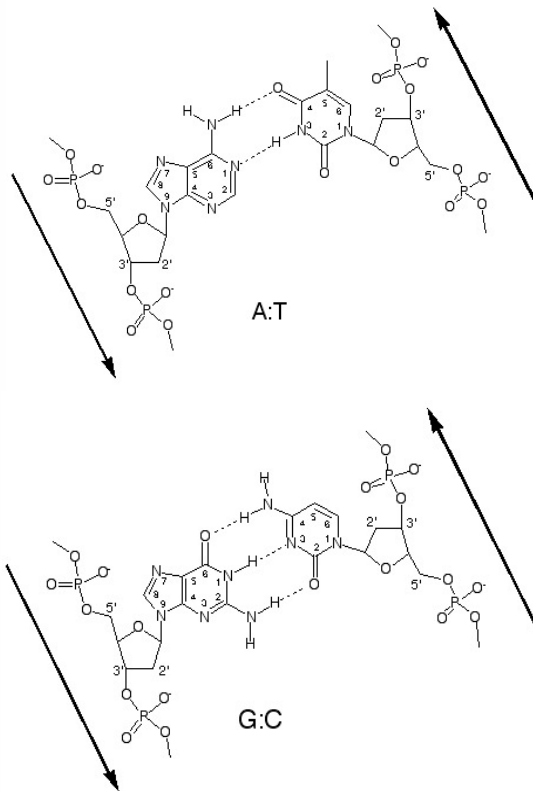
3',5' cyclic phosphate (i.e. cAMP)



- A unit consisting of a base bonded to a sugar is referred to as a nucleoside. The four nucleoside units in RNA are called adenosine, guanosine, cytidine, and uridine, whereas those in DNA are called deoxyadenosine, deoxyguanosine, deoxycytidine, and thymidine (we often leave out the 'deoxy' portion of these names, since it is redundant). In each case, N-9 of a purine or N-1 of a pyrimidine is attached to C-1' of the sugar. The base lies above the plane of sugar when the structure is written in the standard orientation; that is, the configuration of the N-glycosidic linkage is β .
- A nucleotide is a nucleoside joined to one or more phosphate groups by an ester linkage. The most common site of esterification in naturally occurring nucleotides is the hydroxyl group attached to C-5' of the sugar. A compound formed by the attachment of a phosphate group to the C-5' of a nucleoside sugar is called a nucleoside 5'-phosphate or a 5'-nucleotide. For example, ATP is adenosine 5'-triphosphate.
- Another nucleotide is deoxyguanosine 3'-monophosphate (3'dGMP; draw this molecule). This nucleotide differs from ATP in that it contains guanine rather than adenine, contains deoxyribose rather than ribose (indicated by the prefix "d"), contains one rather than three phosphates, and has the phosphate esterified to the hydroxyl group in the 3' rather than the 5' position.
- The four nucleotide units in DNA are called deoxyadenylate, deoxyguanylate, deoxycytidylate, and thymidylate (Why is it not called deoxythymidylate?).
- Draw 3',5'-cAMP - an important second messenger in cells
- Q: Do we have to memorize the structure of nucleic acids or will that be given to us?*
- A: The structure of the nucleic acids will be provided on the tests.*
- Q: Why nucleotide in DNA not called deoxythymidylate? Is it due to T only exist in DNA so it won't cause any confusion when without deoxy-?*
- A: Yes. Since thymine is not normally found attached to a ribose sugar, and is only ever attached to a deoxyribose sugar, there is no need to specify that it is deoxythymidylate.*
- Q: Why the absence of 2-OH in DNA increase its resistance to hydrolysis when compared to RNA? Is it because OH group will cause intramolecular attack that result in hydrolysis?*
- A: Yes. RNA is susceptible to hydrolysis due to intramolecular attack of the 2' OH on the phosphate attached to the 3' OH.*

The double stranded nature of DNA

5

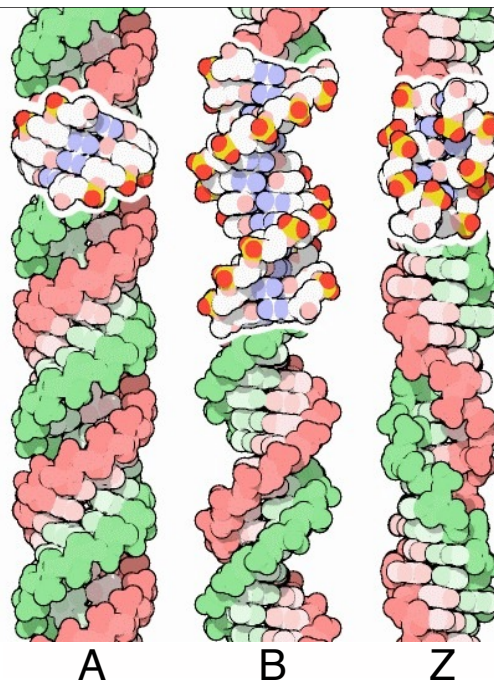


What is the information content of the human genome in bytes?

http://www.neb.com/neb/sitemap/sitemap_5-1-10.html
<http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/lgdna.html>

- The existence of specific base-pairing interactions was discovered in the course of studies directed at determining the three-dimensional structure of DNA. Maurice Wilkins and Rosalind Franklin obtained x-ray diffraction photographs of fibers of DNA. The characteristics of these diffraction patterns indicated that DNA was formed of two chains that wound in a regular helical structure. From these and other data, James Watson and Francis Crick inferred a structural model for DNA that accounted for the diffraction pattern and was also the source of some remarkable insights into the functional properties of nucleic acids. The features of the Watson-Crick model of DNA deduced from the diffraction patterns are:
 - Two helical polynucleotide chains are coiled around a common axis. The chains run in opposite directions.
 - The sugar-phosphate backbones are on the outside and, therefore, the purine and pyrimidine bases lie on the inside of the helix.
 - The bases are nearly perpendicular to the helix axis, and adjacent bases are separated by 3.4 Å. The helical structure repeats every 35.4 Å, so there are 10.4 base pairs (= 35.4 Å per repeat/3.4 Å per base) per turn of helix. There is a rotation of 34.6 degrees per base (360 degrees per full turn/10.4 bases per turn).
 - The diameter of the helix is 23.7 Å.
 - It is commonly said that the human genome is approximately 3,000,000,000 base pairs
- *Q: If the DNA is ionic, what is the counter ion (cation) of this structure?*
- *A: The solution inside of a cell (the cytosol), contains approximately 140 mM K⁺ and 10 mM Na⁺. These two monovalent ions are the predominant counter ions of DNA.*
- *Q: With regards to the question on the counter ion of the DNA, is it also possible that histone protein serves as the counter ion? From what I know, histone is rich in arginine and lysine that both have positive side chains.*
- *A: It is true that histones, the proteins that help to package DNA into the chromosomes, have many positively charged residues. Some of these residues must be making contacts with the phosphate groups and acting as the counter ions. Many of the lysine residues in histones can be subjected to post-translational modification (i.e., acetylation or methylation). It was once thought that this probably changed the interaction of the proteins with the DNA, but now it is generally thought that these modified sites serve as binding sites to recruit other proteins that help to make the DNA more or less accessible for transcription.*

3 helical forms of DNA



A

B

Z

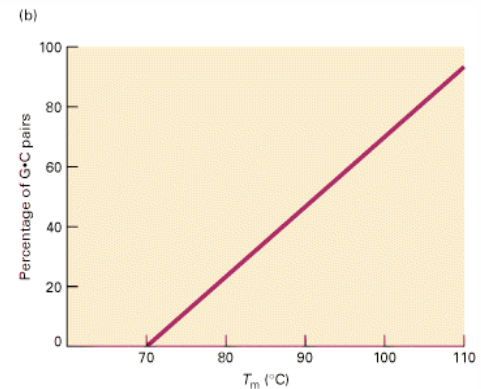
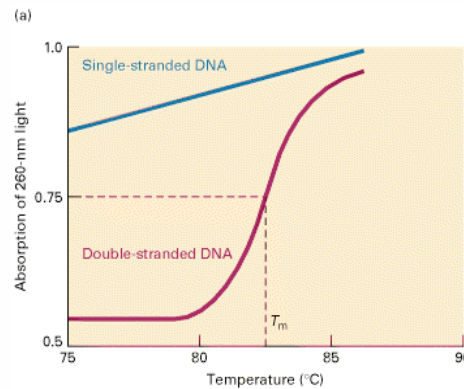
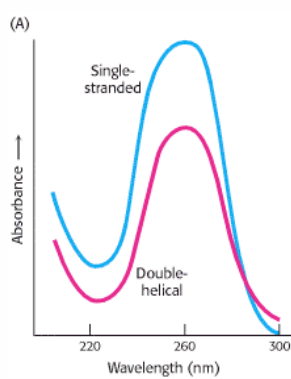
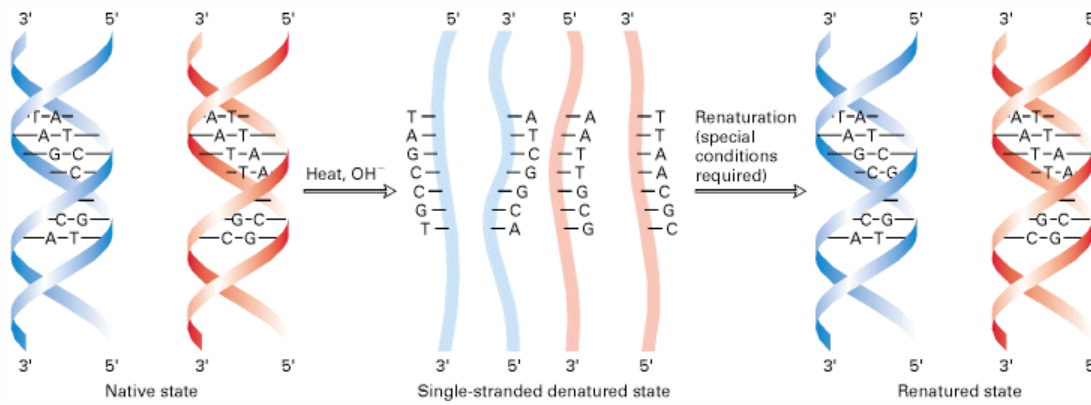
Shape	Broadest	Intermediate	Narrowest
Rise per base pair	2.3 Å	3.4 Å	3.8 Å
Helix diameter	25.5 Å	23.7 Å	18.4 Å
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond	<i>anti</i>	<i>anti</i>	alternating <i>anti</i> and <i>syn</i>
Base pairs per turn of helix	11	10.4	12
Pitch per turn of helix	25.3 Å	35.4 Å	45.6 Å
Tilt of base pairs from normal to helix axis	19°	1°	9°
Major groove	Narrow and very deep	Wide and quite deep	Flat
Minor groove	Very broad and shallow	Narrow and quite deep	Very narrow and deep

David S. Goodsell: The [Molecule of the Month](#) appearing at the [PDB](#)

- DNA is a structurally dynamic molecule that can exist in a variety of helical forms: A-DNA, B-DNA (the classic Watson-Crick helix), and Z-DNA. DNA can be bent, kinked, and unwound. In A-, B-, and Z-DNA, two antiparallel chains are held together by Watson-Crick base pairs and stacking interactions between bases in the same strand.
- DNA adopts the familiar smooth double helix, termed a B-helix, under the typical conditions found in living cells. An example is shown in the center, exemplified by the crystal structure in PDB entry 1bna, shown at the top superimposed over the idealized version of the B-helix.
- Under other conditions, however, DNA can form other structures, as revealed in two early crystal structures: PDB entries 1ana on the left and 2dcg on the right. The one on the left, with tipped bases and a deep major groove, is termed A-DNA. It is formed under dehydrating conditions. Also, RNA most often shows this form, because its extra hydroxyl group on the sugar gets in the way, making the B-form unstable (for instance in the A-helical structure of transfer RNA).
- The form on the right, which winds in the opposite direction from A-DNA and B-DNA, is termed Z-DNA. It is found under high salt conditions and requires a special type of base sequence, with many alternating cytosine-guanine and guanine-cytosine base pairs. This conformation is called Z-DNA because the phosphate groups zigzag along the backbone.
- As more and more structures of DNA are studied, it is becoming clear that DNA is a dynamic molecule, quite flexible on its own, which is bent, kinked, knotted and unknotted, unwound and rewound by the proteins that interact with it.
- *Q: If I have a single DNA strand in solution like 5'-AGCTGAT-3', then I add another two ssDNA like 3'-TCGACTA-5' and 3'-ACGACTA. The latter one has only one base mismatched to the single stand and the former one is perfectly matched. I wonder what percentage of the latter one that will be linked to the single strand in the solution compared to the former one? Is there any paper indicating this interaction trend? Thank you for your time.*
- *A: First thing: don't write DNA 3' to 5'!!!! The mismatched ssDNA can still bind to the first sequence, but with a lower melting temperature. Just like in the practice problem, it should be possible to calculate the concentration of each free and bound species at equilibrium, if you know the deltaG for each binding interaction. There are modified versions of the melting temperature formula that take mismatches into account. Here is one example: <http://depts.washington.edu/bakerpg/primertemp/primertemp.html>*
- *Q: Should I remember the parameters for A/B/Z form DNA or only B-form is enough?*
- *A: Generally speaking, you should learn everything in the notes except for those things that I say will be provided as extra information with the test. Everything is testable.*

DNA melting

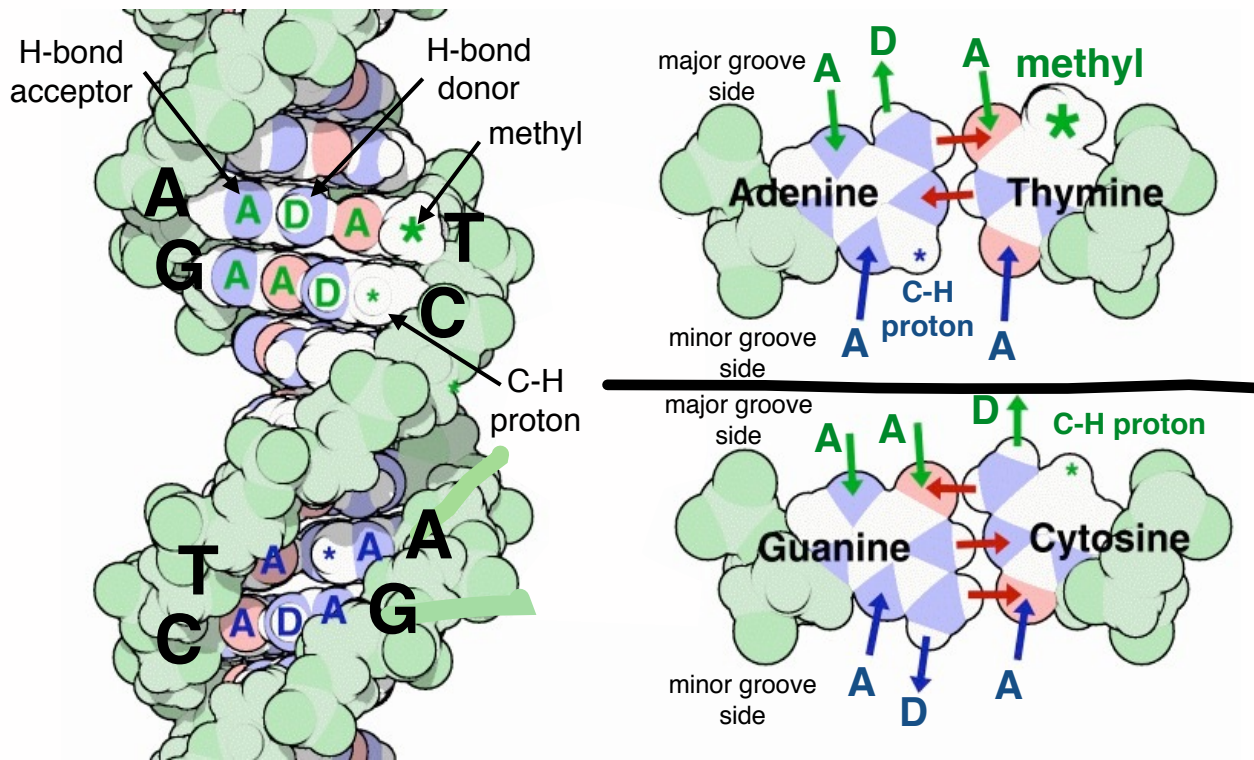
7



- In DNA replication and in the copying of RNA from DNA, the strands of the helix must separate - at least temporarily.
- The unwinding and separation of DNA strands, referred to as denaturation, or “melting,” can be induced experimentally. For example, if a solution of DNA is heated, the thermal energy increases molecular motion, eventually breaking the hydrogen bonds and other forces that stabilize the double helix, and the strands separate.
- This melting of DNA changes its absorption of ultraviolet (UV) light (in the 260-nm range), which is routinely used to measure DNA concentration because of the high absorbance of UV light by nucleic acid bases. Native double-stranded DNA absorbs about one-half as much light at 260 nm as does the equivalent amount of single-stranded DNA.
- Near the denaturation temperature, a small increase in temperature causes an abrupt, near simultaneous, loss of the multiple, weak, cooperative interactions holding the two strands together, so that denaturation rapidly occurs throughout the entire length of the DNA. The melting temperature, T_m , at which the strands of DNA will separate depends on several factors. Molecules that contain a greater proportion of G·C pairs require higher temperatures to denature because the three hydrogen bonds in G·C pairs make them more stable than A·T pairs with two hydrogen bonds. Indeed, the percentage of G·C base pairs in a DNA sample can be estimated from its T_m .

The grooves of DNA show the sequence

8



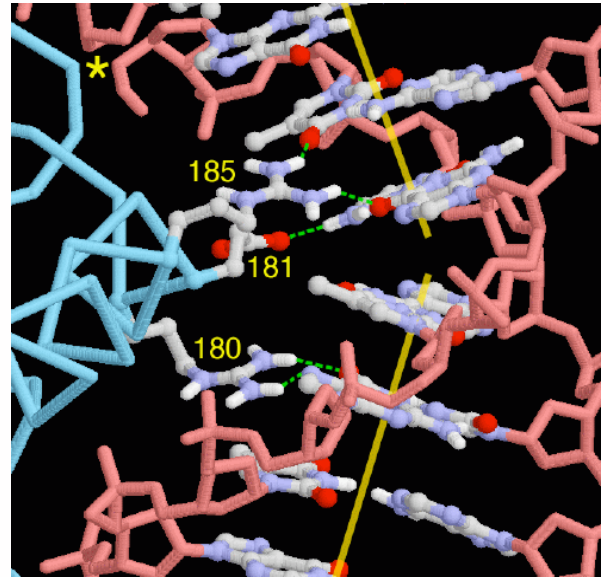
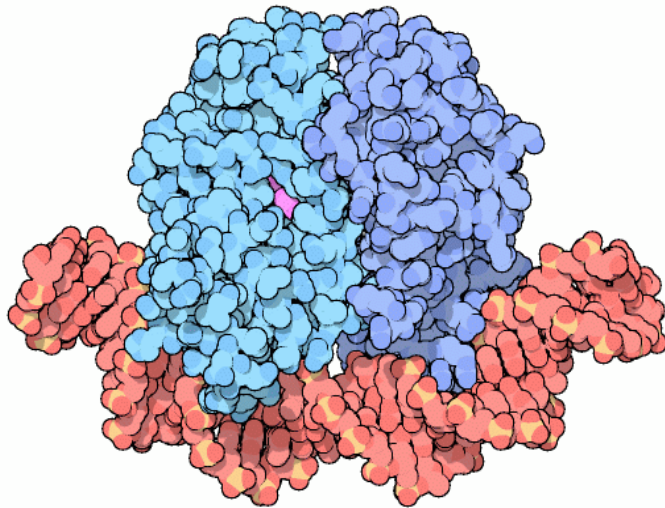
David S. Goodsell: The [Molecule of the Month](#) appearing at the [PDB](#)

- The information content of DNA is actually multiplied by the fact that there are multiple ways to 'read' DNA sequence.
- Additional 'extragenetic' information is read from the surfaces that are left exposed in the double helix.
- In the major groove (the wider of the two grooves in the structure on the left), the different base pairs have a characteristic pattern of chemical groups that carry information, shown by green arrows in the close-up diagrams on the right. These include hydrogen bond donors (D) and acceptors (A) as well as a site with a large, bulky group in adenine-thymine base pairs (large asterisk) or a small group in guanine-cytosine base pairs (small asterisk).
- In the minor groove, there is a different arrangement of chemical groups that carry additional information, indicated with blue arrows in the diagram on the right and the blue letters in the structure on the left.
- As revealed in hundreds of structures in the PDB, this extragenetic information is used by proteins to read the genetic code in DNA without unwinding the double helix. It is also targeted by a number of toxins and drugs that attack DNA.

DNA major groove-protein interactions

Specific DNA sequences or individual bases can be recognized through a variety of mechanism including:

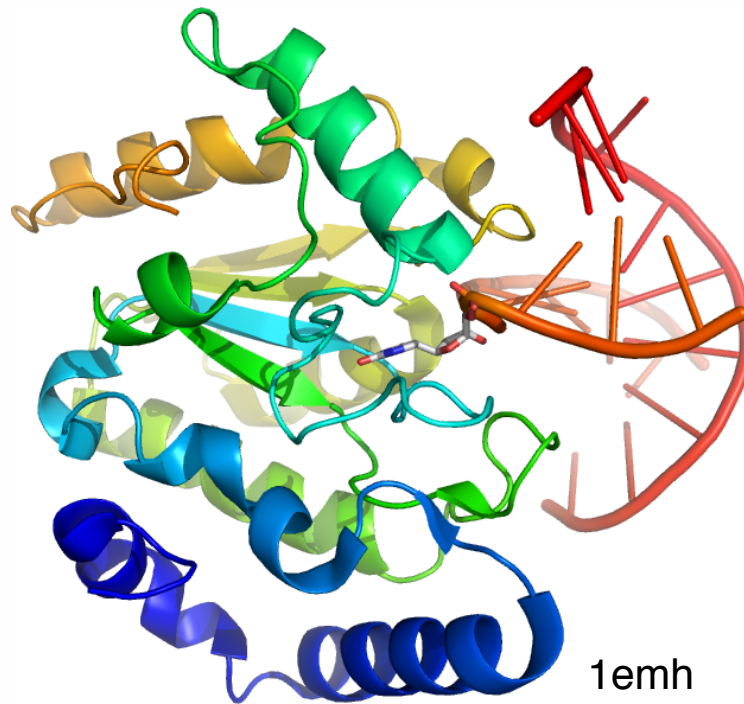
- Major groove interactions
- Minor groove interactions
- Local unwinding to reveal bases
- Individual base flipping out of helix



[David S. Goodsell](#): The [Molecule of the Month](#) appearing at the [PDB](#)

- When glucose is plentiful, bacteria ignore other nutrients in their environment, feasting on their favored source. But, when glucose is rare, they shift gears and mobilize the machinery needed to use other sources of energy. As glucose levels drop, the cell-surface enzyme adenyl cyclase is activated. It grabs ATP molecules and clips off two phosphates as it cyclizes it to form cAMP (a molecule that we saw earlier). cAMP spreads through the cell, stimulating production of the enzymes that process other food molecules. Because of its role in delivering messages from the primary glucose sensor (adenyl cyclase) to the synthetic machinery, cAMP is often known as a second messenger.
- Catabolite activator protein (CAP), also known as cyclic AMP receptor protein (CRP), is activated by cAMP and stimulates synthesis of the enzymes that break down non-glucose food molecules. It is composed of two identical subunits, shown here in blue from PDB entry 1cgp. When cAMP (shown in purple) binds, it changes the conformation of the protein slightly, making it perfect for binding to DNA (shown here in red). CAP binds to a specific DNA sequence, which is found next to the genes that are activated.
- CAP is not gentle when it binds to DNA: it grabs the DNA and bends it by nearly 90 degrees. This is not a smooth bend, however. Since DNA is a double helix with two grooves that spiral up along the strand, it bends more easily in some directions.

Base flipping by uracil-DNA glycosylase (UDG)¹⁰

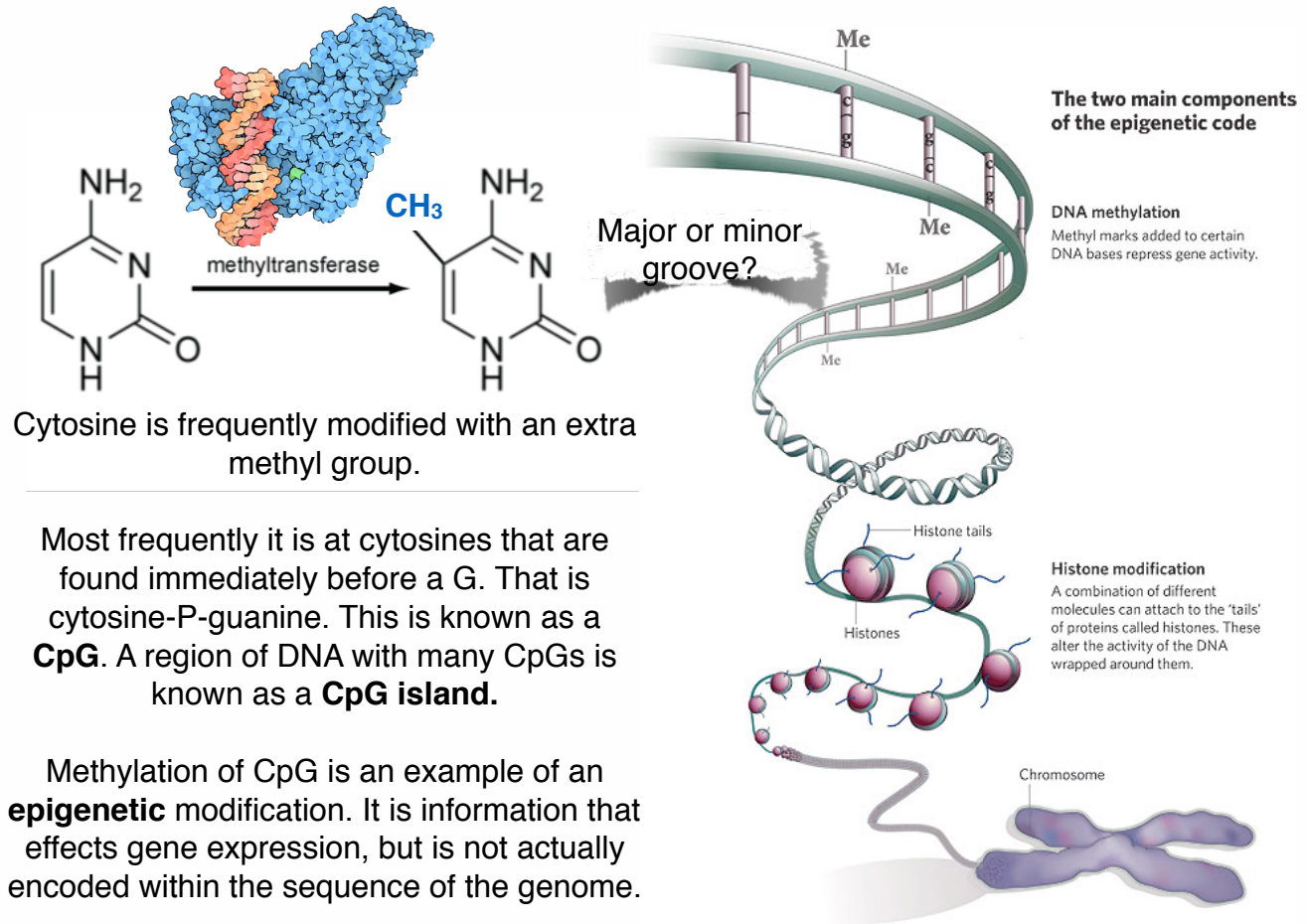


David S. Goodsell:
The Molecular Perspective
appearing in [The Oncologist](#)

- An even more aggressive approach to DNA recognition is taken by certain DNA-repair enzymes. Human uracil-DNA glycosylase (UDG) is an archetypical DNA glycosylase that initiates the multienzymatic DNA base-excision repair pathway. Both misincorporation and cytosine deamination put the RNA base uracil into DNA. UDG excises uracil from both single- and double-stranded DNA by cleaving the N-C1' glycosylic bond between the base and deoxyribose. UDG undergoes a global conformational change from an "open" unbound state to a "closed" DNA-bound state in the UDG product complex, which evidently creates the catalytically competent active centre. In the product UDG-DNA complex, UDG has flipped its target uridine nucleotide out of the DNA base stack and into this active centre, where the glycosylic bond is cleaved. Some text and pdb file from Parikh et al. PNAS 2000, 97 5083-5088.
- *Q: Is UDG a DNA repair enzyme or it is a complex that initiate the repair enzymes? (How UDG works exactly? It just flips the bases until it reaches a U and once it found a U, it just hydrolysis the U?)*
- *A: You are correct that UDG just does the hydrolysis of U, so it is best described as initiating a repair pathway. Other enzymes are required to repair the site with the missing base (an abasic site). We will learn more about how this done a bit later in the course.*

DNA can be covalently modified

11

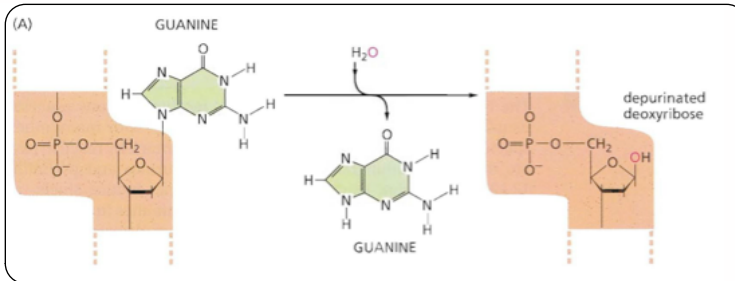


- CpG methylation plays an important role in regulating gene expression. If there is a lot of methylation, it will tend to cause decreased expression of associated genes in the DNA. One mechanism can be by interfering with the binding of proteins (i.e., transcription factors) to the DNA, which would normally lead to transcription.
- Regions with lots of methylation are also more prone to mutation, as well will see in a following slide.
- <http://www.ks.uiuc.edu/Research/methylation/>
- <http://www.rcsb.org/pdb/101/motm.do?momID=139>
- <http://cerch.org/research-programs/chamacos/chamacos-epigenetics/>

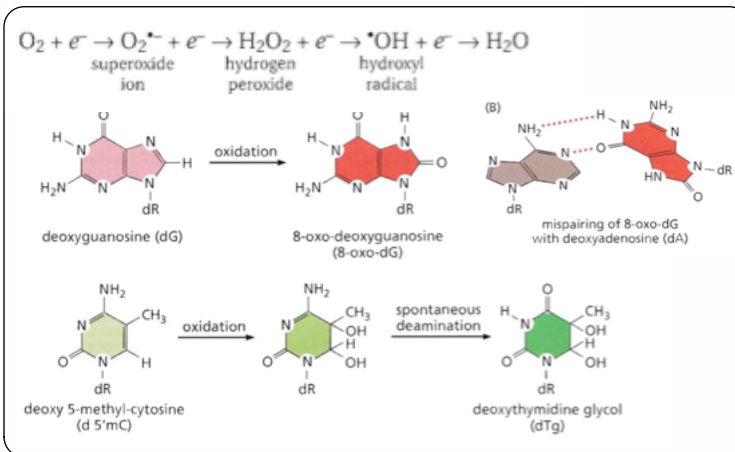
DNA is constantly under attack from within

12

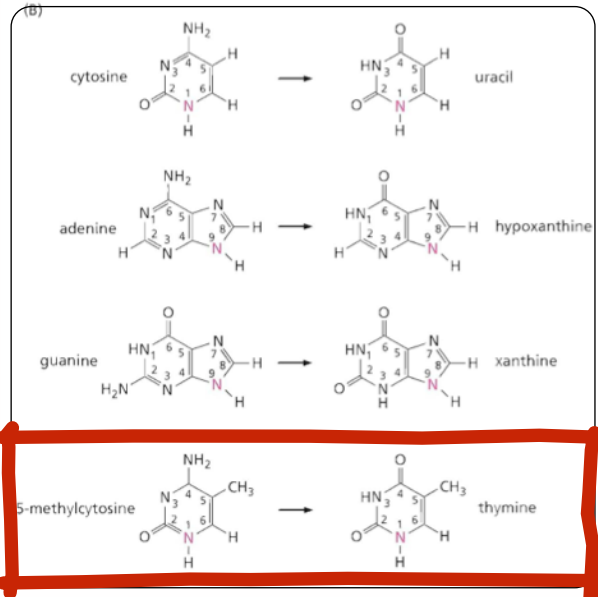
Depurination (Gs and As)



Oxidation by reactive oxygen species



Deamination



This is the worst. Since T is a normal nucleotide present in DNA, there is no way for the cell to "know" that a mutation has occurred.

Chapter 12 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

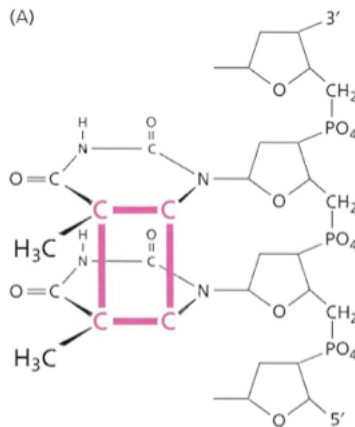
- Depurination seems to affect G most frequently. It is estimated that this might happen 10,000 times per cell per day! Depyrimidination (Ts and Cs) happens at a much lower rate (20-100x less frequent).
- Reactive oxygen species (ROS) are generated in the mitochondria and peroxisomes of a cell, and some can leak out and into the nucleus. ROS are highly reactive and can react with the bases of DNA. 8-oxo-dG is a troublesome product, since it can mispair with A during DNA replication, resulting in a mutation in the newly formed DNA.
- Deamination is the loss of the amine group from a base (replaced with O from water).
- The conversion of C to U by deamination is troublesome, since U can be read as T. This is why we have specific enzymes for removing U from DNA (since it should not be there normally).
- Deamination of 5-methylcytosine is even worse, since it produces thymine (since it is part of normal DNA).
 - It is estimated that this is the main source of mutations in internal organs (those shielded from UV light).
- A mutation of a purine to purine (A to G or G to A) or pyrimidine to pyrimidine (T to C or C to T) is called a transition.
- A mutation from a purine to a pyrimidine (or vice versa) is called a transversion.

DNA is constantly under attack from without ¹³

UV radiation (~254 nm) produces crosslinks between pyrimidine bases that are next to each other

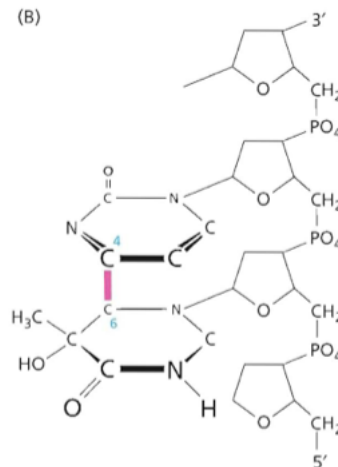
cyclobutane pyrimidine dimers (CPD) formed from T-T

(A)



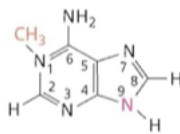
pyrimidine-pyrimidinone photoproducts (6-4 PP)

(B)

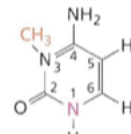


(C)

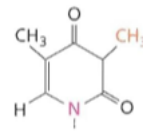
Some of the most potent mutagens are DNA alkylating chemicals



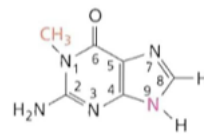
1-methyladenine



3-methylcytosine



3-methylthymine



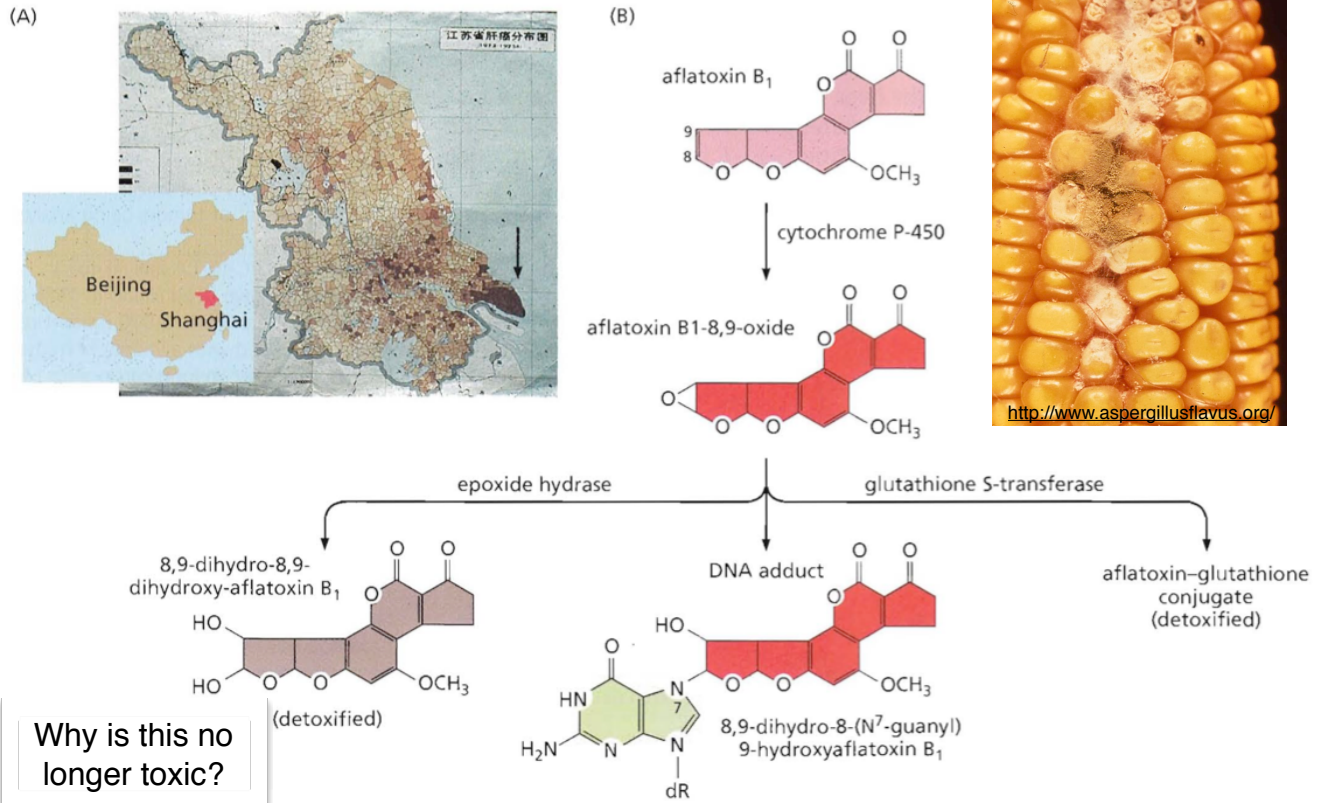
1-methylguanine

5-methylcytosine is normal, 3-methylcytosine is not.

Chapter 12 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

- Examples of DNA alkylating reagents polyaromatic hydrocarbons such as benzo[a]pyrene.
- You might wonder how an aromatic molecule could cause alkylation of DNA. It turns out that it is typically the metabolic products that are the actual carcinogen. The body has enzymes that oxidize the molecule to make it more soluble so it can be excreted in the urine. However, these oxidized forms also become better electrophiles and so they can react with nucleophilic groups in the cell.
- Benzo[a]pyrene is one of the main carcinogens in tobacco smoke. Cancer is typically introduced in lab animals by exposing them to such compounds.
- It is not necessarily just a methyl group that gets added, it could be a larger molecule

Aflatoxin B₁: one of the most potent carcinogens



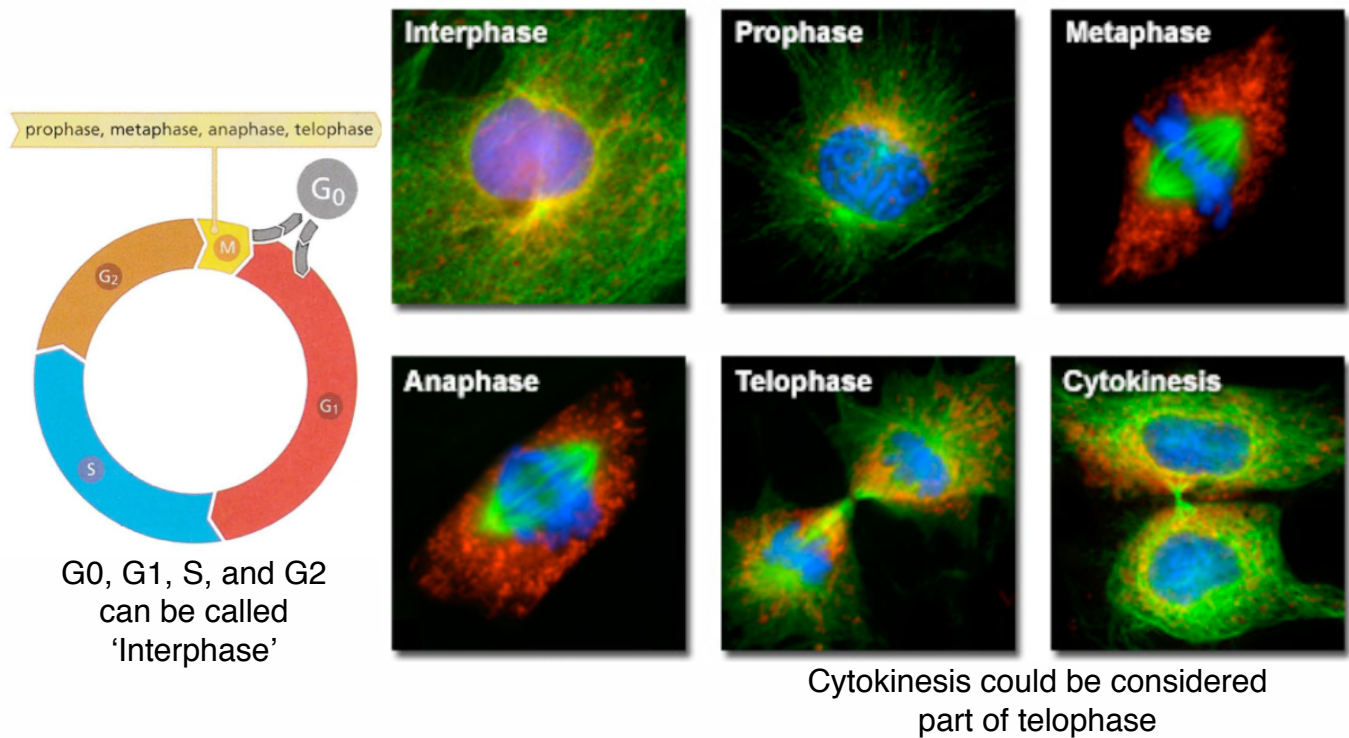
Chapter 12 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

- Aflatoxin is a molecule produced by *Aspergillus* fungus, which grows on grains that have been stored in areas that are too moist, due to high rainfall and humidity.
- This toxin is known to cause higher rates of liver cancer (hepatocellular carcinoma). The rate of this type of cancer is 8x higher in the humid portions of Jiangsu province than in the drier areas of the province.
- The people that get this cancer have a higher rate of G to T transversion at codon 249 of the p53 gene (a key tumor suppressor).
- For aflatoxin to react with DNA, it first inserts itself into the helix between two adjacent base pairs. It is able to do this due to its flat and hydrophobic structure.

Cell division: mitosis

15

Mitosis in Rat Kangaroo Epithelial Kidney Cells

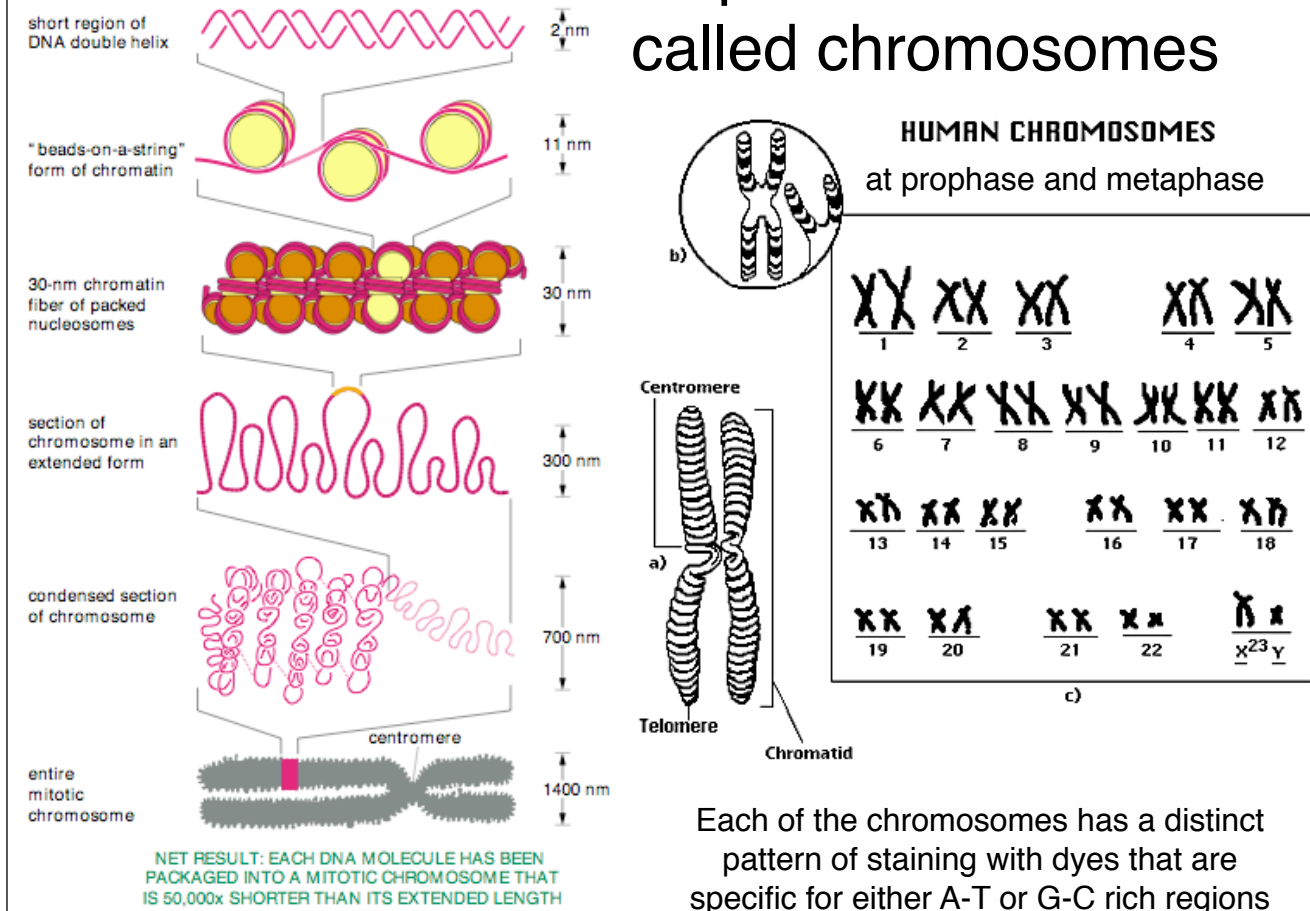


Chapter 8 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

<http://micro.magnet.fsu.edu/cells/fluorescencemitosis/index.html>

- G₀ = resting, non-replicative state of a cell
- G₁ = first gap phase, not much happens here. Cell is basically making decisions about whether it should continue through the cycle or go back to G₀ or differentiate. 12-15 h.
- S = synthetic phase. This is where the DNA is being copied. 6-8 h.
- G₂ = second gap phase, not much happens here either. 3-5 h.
- Together, G₀, G₁, S, and G₂ can be called 'Interphase'
- M = mitosis phase, this is where all the action happens. Divided into additional subphases. Takes about 1 h total.
 - Prophase = DNA chromatin condenses into discrete 'mitotic' chromosomes, organized as 46 pair. 23 chromosomes, maternal and paternal copies of each, all duplicated. A chromosome and its identical copy (sister chromatids) are connected near the middle by a structure called the centromere
 - Metaphase = Chromosome align along a plane in the middle of the cell and become attached to microtubules (protein filaments) which are in turn attached to one of two spindle poles located at either end of the cell.
 - Anaphase = Microtubules begin to shorten, separating the sister chromatids and pulling one copy towards one spindle pole and the second copy towards the other spindle pole. This happens very fast (a few minutes)
 - Telophase (includes cytokinesis) = The chromosomes decondense and go back to their normal state, which is known as chromatin

DNA is packed into structures¹⁶ called chromosomes



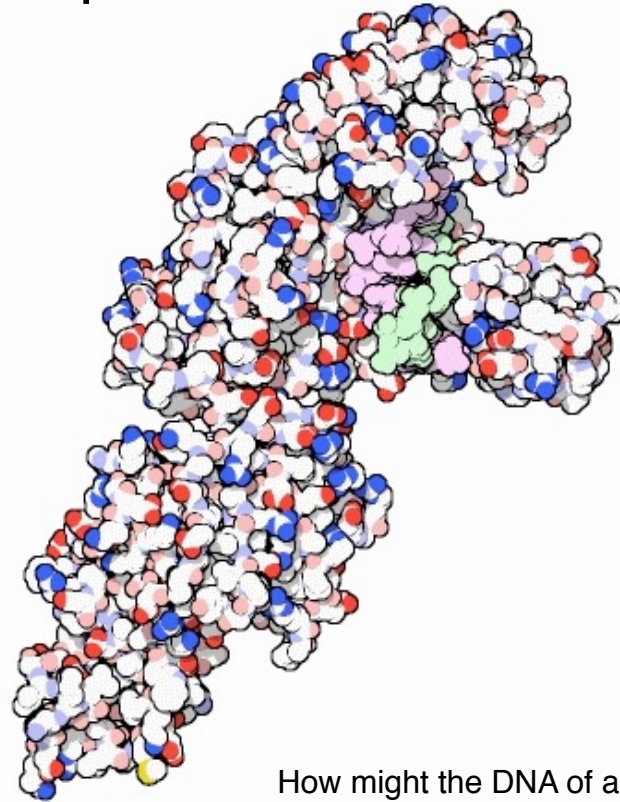
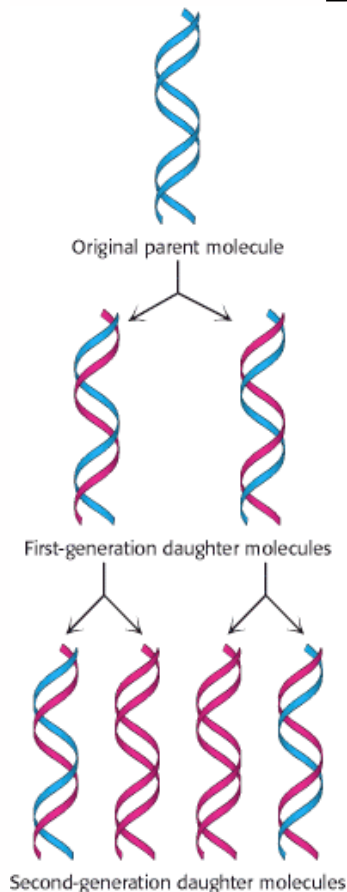
<http://www.accessexcellence.org/>

Each of the chromosomes has a distinct pattern of staining with dyes that are specific for either A-T or G-C rich regions

- Chromosome: a very long dsDNA molecule together with associated proteins (histones) that are responsible for packing the molecule into a dense structure. A nucleosome is a protein octamer composed of two copies each of 4 different histone proteins.
- The average number of genes per chromosome is about 1000. Generally speaking, larger chromosomes contain more genes (chromosome 1 with ~3000 versus Y chromosome with ~100).
- A human cell contains about 2 meters of DNA (1 m per chromosome set). The human body consists of approximately 10^{13} cells, and each cell is diploid; therefore, the body contains a total of about 2×10^{13} m of DNA. The distance from Earth to the sun is 1.4×10^{11} m, so the DNA in your body could stretch to the sun and back almost 70 times.
- If DNA from a large number of cells is extracted, purified with great care to avoid breaking the molecules, and analyzed by pulsed-field electrophoresis, all the DNA of one chromosomal type ends up in one band. The number of bands of DNA is always equal to the number of chromosomes. Human chromosomes are too big to be separated by pulsed field electrophoresis, but other eukaryotic organisms have smaller chromosomes that can be analyzed in this way.
- A chromosome is formed from a single DNA molecule that contains many genes. A chromosomal DNA molecule contains three specific nucleotide sequences which are required for replication: a DNA replication origin; a centromere to attach the DNA to the mitotic spindle.; a telomere located at each end of the linear chromosome.
- Chromosomes can be stained with dyes that bind either to regions that are rich in A-T (G bands) or G-C (R bands) base pairs. When they are stained, the mitotic chromosomes have a banded structure that unambiguously identifies each chromosome of a karyotype. Each band contains millions of DNA nucleotide pairs which do not correspond to any functional structure.
- Note that mice have 19 chromosomes plus XY plus mtDNA. Chimpanzees have 23 chromosomes plus XY plus mtDNA. Interestingly, our second largest chromosome (#2) is composed of a fusion of two smaller chimpanzee chromosomes (designated 2a and 2b). It's not clear if there is evolutionary significance to this difference, or if it is just a coincidental difference.
- *Q: what is mitotic chromosomes?*
- *A: During mitosis (cell division) each chromosome is duplicated and compacted and takes on the characteristic 'X' shape that we have seen in a number of slides. This is what we refer to as the mitotic chromosomes.*

DNA replication

17

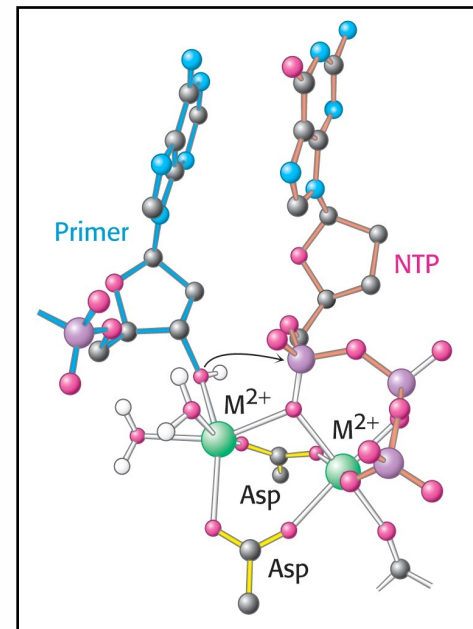
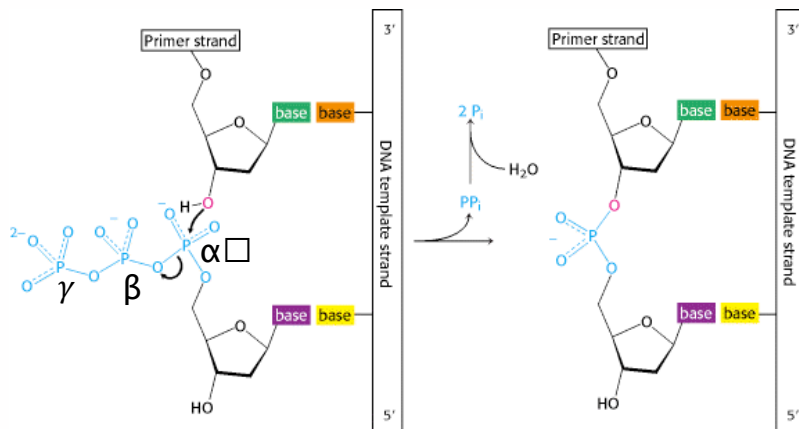
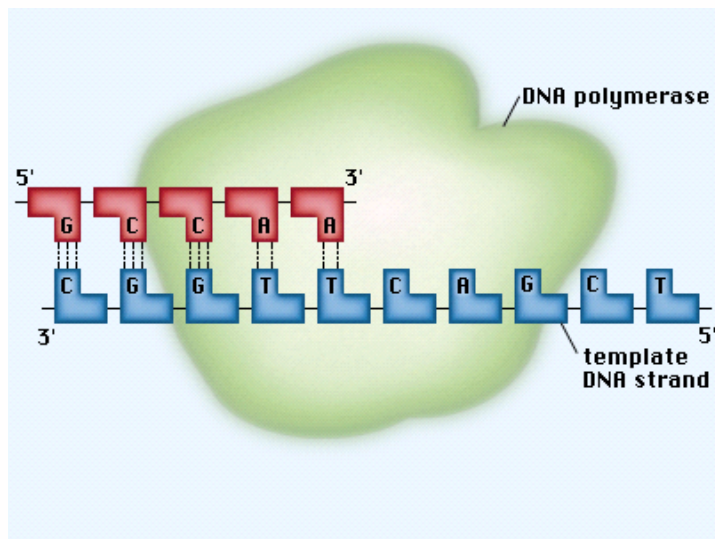


How might the DNA of a thermophile differ from a non-thermophilic species?

David S. Goodsell: The [Molecule of the Month](#) appearing at the [PDB](#)

- The regular pairing of bases in the double-helical DNA structure suggested to Watson and Crick a mechanism of DNA synthesis. Their proposal that new strands of DNA are synthesized by copying of parental strands of DNA has proved to be correct. The DNA strand that is copied to form a new strand is called a template. The information in the template is preserved: although the first copy has a complementary sequence, not an identical one, a copy of the copy produces the original (template) sequence again.
- In the replication of a double-stranded, or duplex, DNA molecule, both original (parental) DNA strands are copied. When copying is finished, the two new duplexes, each consisting of one of the two original strands plus its copy, separate from each other. In some viruses, single-stranded RNA molecules function as templates for synthesis of complementary RNA or DNA chains. However, the vast majority of RNA and DNA in cells is synthesized from preexisting duplex DNA.
- The process of DNA replication is catalyzed by an enzyme called DNA polymerase. Shown in the figure is the DNA polymerase from *Thermus aquaticus*, a bacterium that lives in hot springs. This polymerase, often just referred to as Taq, is perfectly happy at 70 degrees centigrade. This enzyme may be found in the PDB in the file 1tau

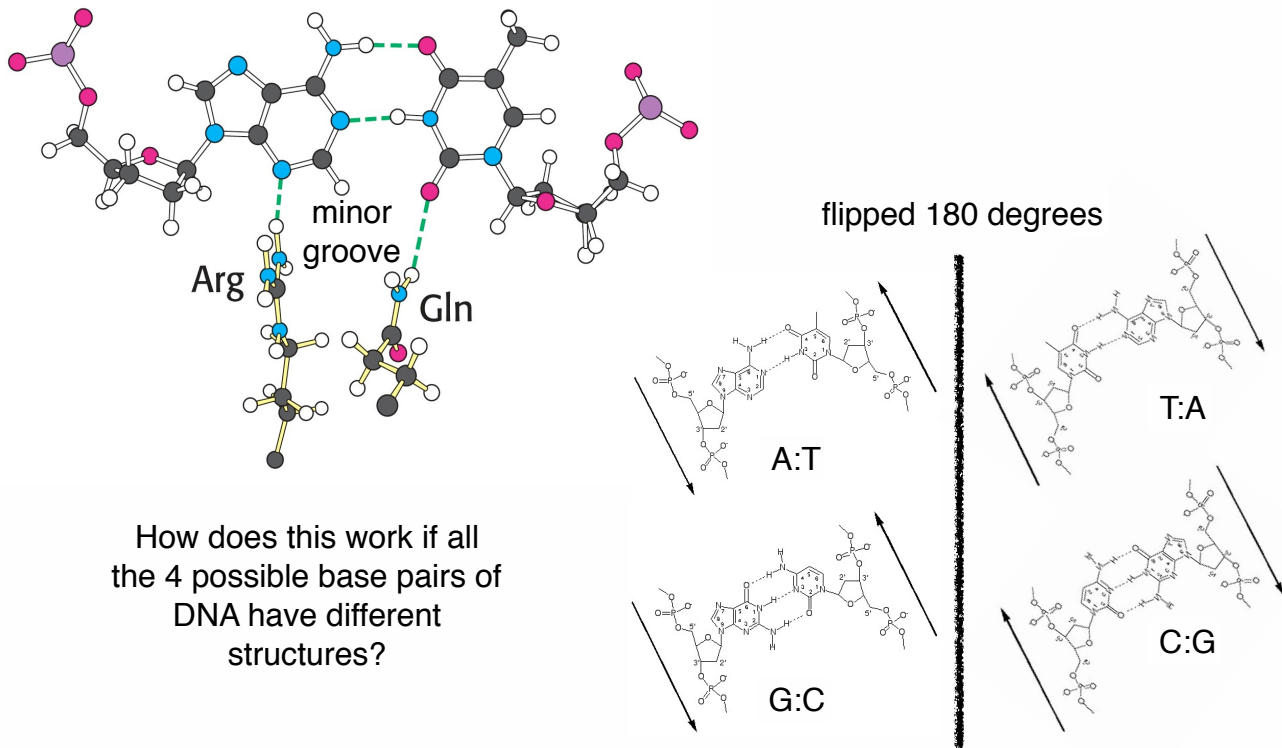
DNA replication



- Like all enzymes with nucleoside triphosphate substrates, DNA polymerases require metal ions for activity. Examination of the structures of DNA polymerases with bound substrates and substrate analogs reveals the presence of two magnesium ions (Mg^{2+}) in the active site. One metal ion binds the deoxynucleoside triphosphate (dNTP) whereas the other interacts with the 3'-hydroxyl group of the primer strand. The 'primer' is the growing strand of newly synthesized (or at least previously formed) DNA.
- The two Mg^{2+} ions are bridged by the carboxylate groups of two aspartate residues in the active site of the polymerase. These side chains hold the metal ions in the proper position and orientation. The metal ion bound to the primer activates the 3'-hydroxyl group of the primer, facilitating its attack on the α -phosphate group of the dNTP substrate in the active site. The two metal ions together help stabilize the negative charge that accumulates on the pentacoordinate transition state. The metal ion initially bound to dNTP stabilizes the negative charge on the pyrophosphate product.
- *Q: When replication is started, the nucleotides come and join to the chain one by one, but where do they come from? Does the DNA polymerase produce them?!*
- *A: The cell is full of the nucleotide triphosphates which are randomly diffusing. These building blocks are made by other biochemical pathways in the cell, but that is beyond the scope of this course.*

Structural basis for the high-fidelity of DNA replication

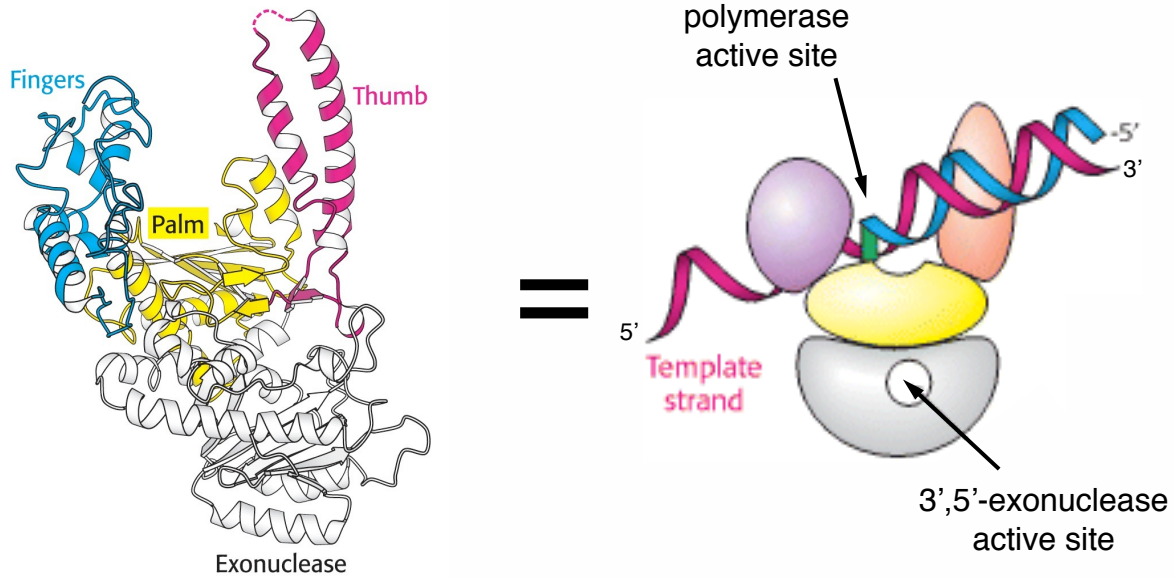
19



- Each base added to the growing chain should with high probability be the Watson-Crick complement of the base in the corresponding position in the template strand. The binding of the NTP containing the proper base is favoured by the formation of a base pair, which is stabilized by specific hydrogen bonds. The binding of a noncomplementary base is unlikely, because the interactions are unfavourable. The hydrogen bonds linking two complementary bases make a significant contribution to the fidelity of DNA replication. However, DNA polymerases replicate DNA more faithfully than these interactions alone can account for. The examination of the crystal structures of various DNA polymerases indicated several additional mechanisms by which replication fidelity is improved.
- First, residues of the enzyme form hydrogen bonds with the minor-groove side of the base pair in the active site. In the minor groove, hydrogen-bond acceptors are present in the same positions for all Watson-Crick base pairs. These interactions act as a “ruler” that measures whether a properly spaced base pair has formed in the active site.
- Second, DNA polymerases close down around the incoming NTP. The binding of a nucleoside triphosphate into the active site of a DNA polymerase triggers a conformational change: the finger domain rotates to form a tight pocket into which only a properly shaped base pair will readily fit. The mutation of a conserved tyrosine residue at the top of the pocket results in a polymerase that is approximately 40 times as error prone as the parent polymerase

DNA replication: the Klenow fragment from *E. coli* DNA polymerase 1

20

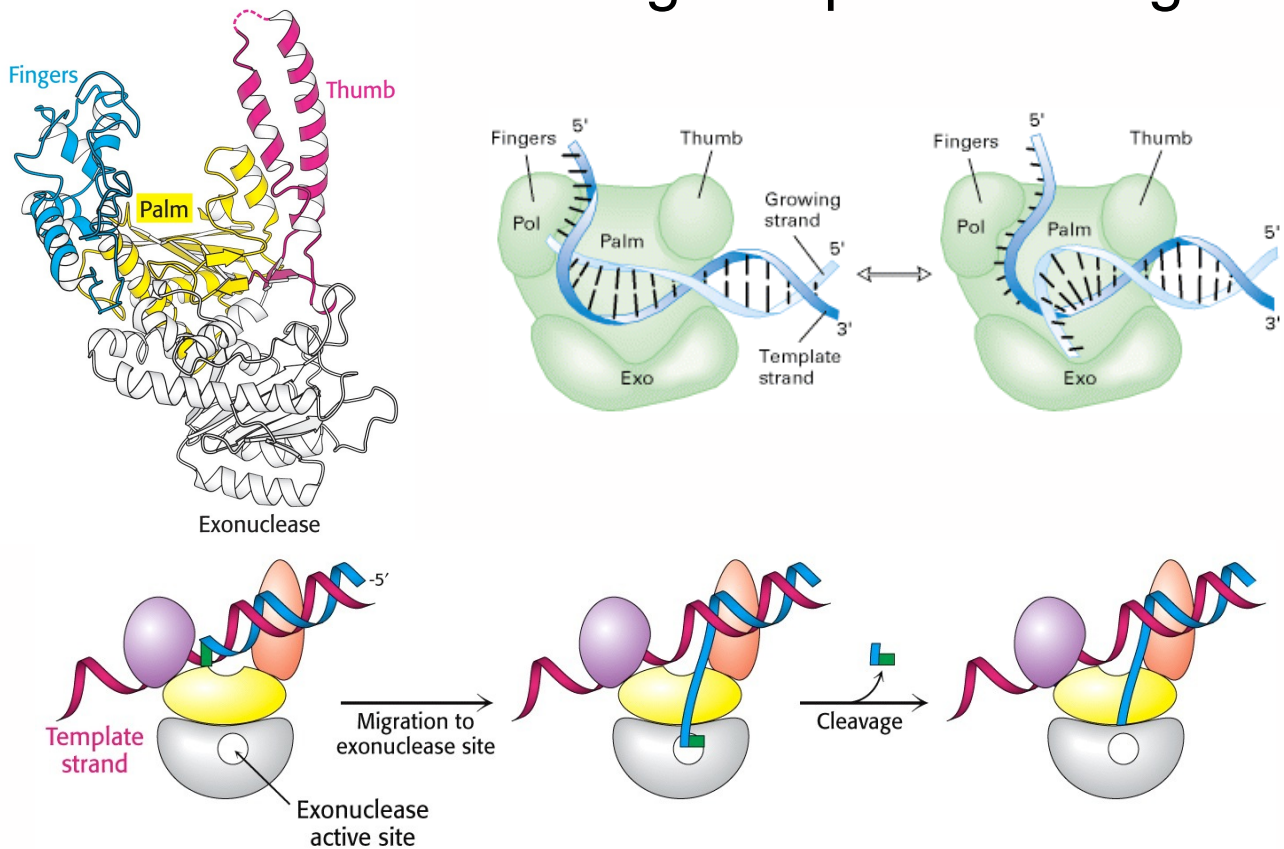


What is the difference between an endo- and an exonuclease?

- In addition to Taq, the three-dimensional structures of a number of other DNA polymerase enzymes are known. The first such structure to be determined was that of the so-called Klenow fragment of DNA polymerase I from *E. coli*. Like other DNA polymerases, the polymerase domain resembles a right hand with fingers (blue), palm (yellow), and thumb (red).
- In addition to the polymerase, the Klenow fragment includes a domain with 3'-5'-exonuclease activity that participates in proofreading and correcting the polynucleotide product.
- Endo- From Greek endon, within
- Exo- From Greek exo, outside,

DNA error checking and proofreading

21



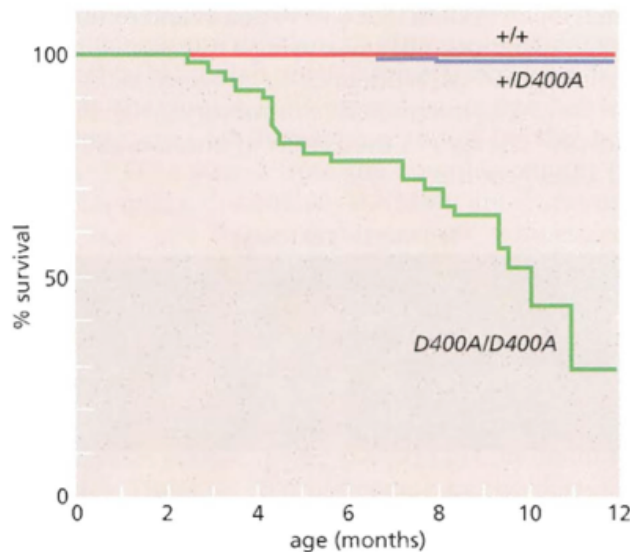
Is there an indirect benefit to having some errors during DNA replication?

- The wrong base occasionally is inserted during DNA synthesis. Indeed, the polymerase active site of *E. coli* DNA polymerase III introduces about 1 incorrect base in 10,000 internucleotide linkages during replication in vitro. Since an average *E. coli* gene is about 1,000 bases long, an error frequency of 1 in 10,000 base pairs would cause a potentially harmful mutation in every tenth gene during each replication. The measured mutation rate in bacterial cells is much less, about 1 mistake in a billion nucleotide polymerization events. This increased accuracy in vivo is largely due to the proofreading function of *E. coli* DNA polymerases.
- The 3'-5'-exonuclease activity of *E. coli* DNA polymerase I can remove a mismatched base at the 3'-growing end of a synthetic primer-template complex. When an incorrect base is incorporated during DNA synthesis, the polymerase pauses, then transfers the 3'-end of the growing chain to the exonuclease site where the mispaired base is removed. Then the 3' end is transferred back to the polymerase site, where this region is copied correctly. Proofreading is a property of almost all bacterial DNA polymerases. It seems likely that this function is indispensable for all cells to avoid excessive genetic damage.
- *Q: You mentioned that chain termination mutations lead to inactive proteins. whereas substitutions of one amino acid for another can sometimes be harmless and sometimes be beneficial. Can you please explain how it can be beneficial ?*
- *A: Random genetic changes that lead to chain termination or amino acid substitutions are almost always harmful to the organism. Very rarely an amino acid substitution (for example) in a protein sequence can provide an organism with a survival advantage relative to the other members of its species. This is the basis of evolution by natural selection. That is what I had meant by beneficial.*

The importance of 3'-5' proofreading

22

Researchers have created mice that have a point mutation (D400A) in the active site of the 3'-5' exonuclease domain of mouse DNA polymerase- δ .



note: wild-type vs.
heterozygous vs.
homozygous



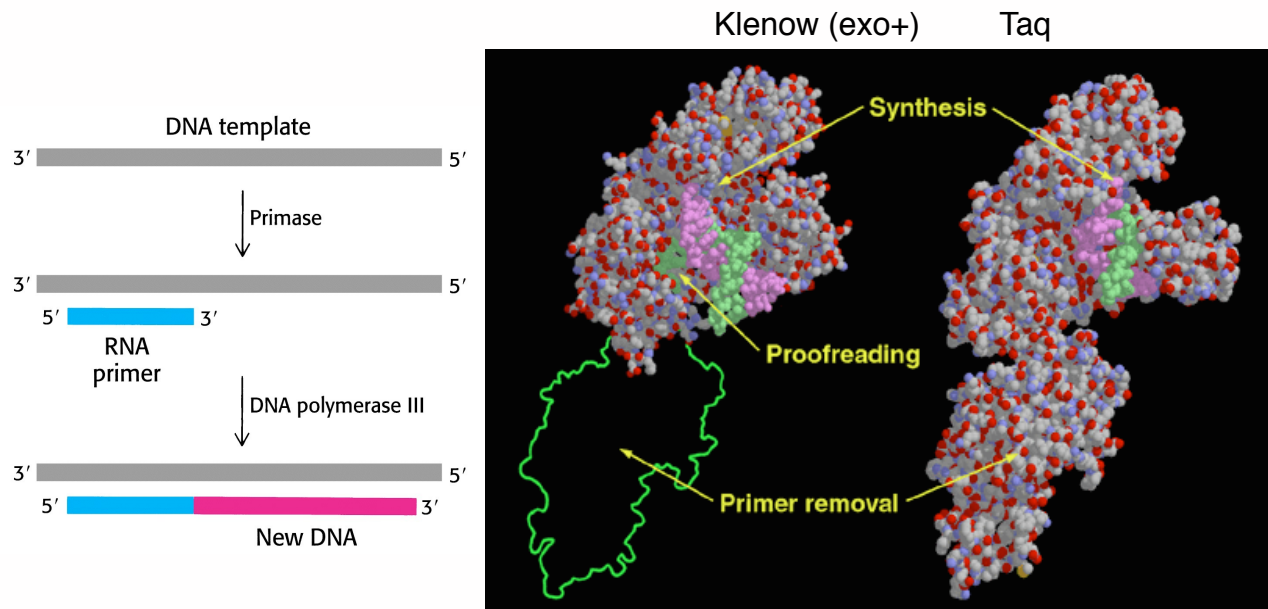
These mice are unable to accomplish proofreading of newly synthesized DNA following replication.

Chapter 12 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

- Data is shown for mice that are either wild-type, homozygous, or heterozygous.
- Mice are dying due to the development of tumors.
- This indicates that the proofreading function of DNA polymerase is key to protecting us from cancer. Furthermore, it shows that mutations lead to cancer.

DNA priming for replication

23



Klenow fragment has no 5',3' exonuclease activity.

Klenow fragment with no 3',5' exonuclease activity (exo-) is also available.

How do you think the proofreading activity was abolished in the exo- mutant?

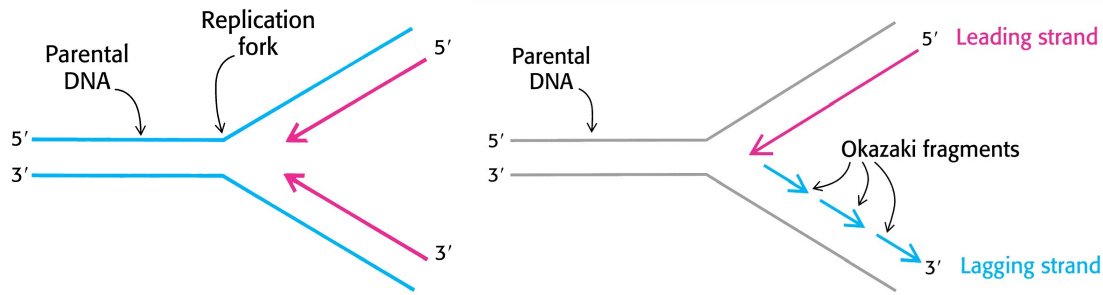
David S. Goodsell: The [Molecule of the Month](#) appearing at the [PDB](#)

- To initiate the polymerization reaction, DNA polymerases require a primer with a free 3'-hydroxyl group already base-paired to the template. They cannot start from scratch by adding nucleotides to a free single-stranded DNA template. RNA polymerase, in contrast, can initiate RNA synthesis without a primer.
- Recall that, to ensure fidelity, DNA polymerase tests the correctness of the preceding base pair before forming a new phosphodiester bond. RNA polymerases can start chains de novo because they do not examine the preceding base pair. Consequently, their error rates are orders of magnitude higher than those of DNA polymerases. The ingenious solution is to start DNA synthesis with a low-fidelity stretch of polynucleotide but mark it "temporary" by placing ribonucleotides in it.
- The RNA primer is removed by hydrolysis by a 5',3'-exonuclease; in *E. coli*, the exonuclease is present as an additional domain of DNA polymerase. Thus, the complete polymerase has three distinct active sites: a 3',5'-exonuclease proofreading activity, a polymerase activity, and a 5',3'-exonuclease activity.
- *Q: In DNA replication, two DNA polymerase are exist in one place? as I found Polymerase I can edit the incorrect base with its exonuclease, but polymerase III do polymerization.*
- *A: For the sake of this course, we can just consider that there is one type of DNA polymerase in a cell. In reality, there are several different types of polymerases in bacteria and mammalian cells, each of which has a special role to play. I think it is too complicated for this course to start thinking about all of the different polymerases.*
- *Q: I don't know where the RNA Primers come from at first. I can see them just at the beginning of the template, but where does they have come from?*
- *A: The RNA primers are made by another enzyme called simply, primase. This is an RNA polymerase that can make small sections of RNA, that are complementary to a DNA template, without the need of a preexisting primer.*

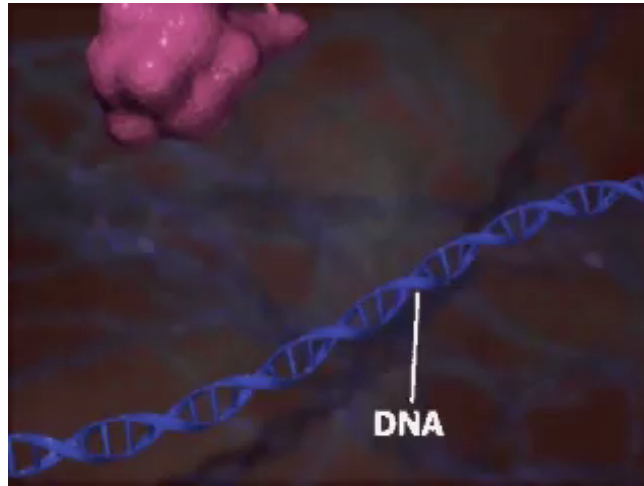
DNA priming for replication

Question: How is this possible if DNA polymerase only synthesizes DNA in the 5' to 3' direction?

Answer: It synthesizes small disconnected fragments (Okazaki fragments) in the 5' to 3' direction which are later joined together by a DNA ligase.



This animation schematically represents the overall process of DNA replication in mammalian cells. You can see that there are more processes involved than what we have discussed in this course.

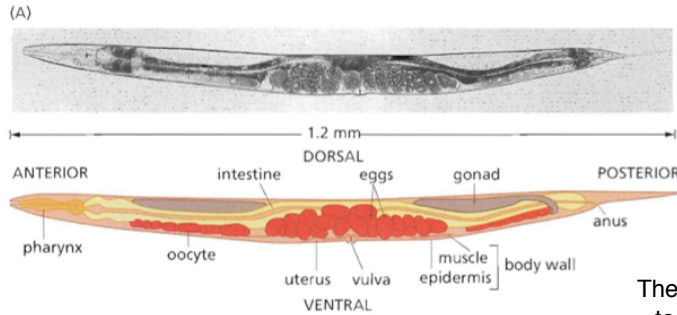


The primase is doing something impossible (and wrong) in this movie. What is it?

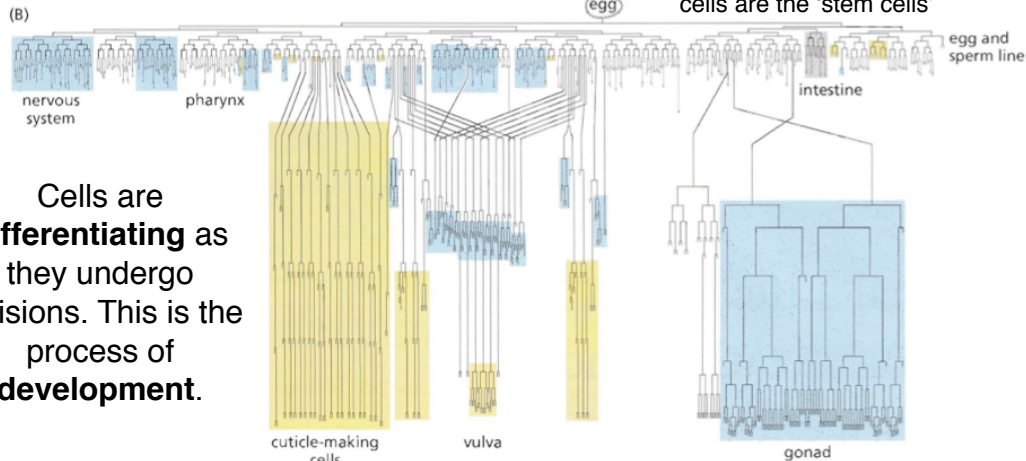
bubblare.se/movie/dna_replication_process

- Both strands of parental DNA serve as templates for the synthesis of new DNA. The site of DNA synthesis is called the replication fork because the complex formed by the newly synthesized daughter strands arising from the parental duplex resembles a two-pronged fork. Recall that the two strands are antiparallel; that is, they run in opposite directions.
- At a 'low resolution' view of the process, both daughter strands appear to grow in the same direction. However, all known DNA polymerases synthesize DNA in the 5'-3' direction and not in the 3'-5' direction. How then does one of the daughter DNA strands appear to grow in the 3'-5' direction?
- This dilemma was resolved by Reiji Okazaki, who found that a significant proportion of newly synthesized DNA exists as small fragments. These units of about a thousand nucleotides (called Okazaki fragments) are present briefly in the vicinity of the replication fork. Each Okazaki fragment had to have been primed by a section of RNA which much be digested by the 5'-3' exonuclease activity of the DNA polymerase that is synthesizing the next Okazaki fragment.
- As replication proceeds, these fragments become covalently joined through the action of DNA ligase to form one of the daughter strands. The other new strand is synthesized continuously. The strand formed from Okazaki fragments is termed the lagging strand, whereas the one synthesized without interruption is the leading strand. Both the Okazaki fragments and the leading strand are synthesized in the 5'-3' direction.
- A DNA ligase is an enzyme that can join together two extended pieces of DNA, provided that their 5' and 3' ends are in close proximity. We will learn more about these enzyme later in the course.
- *Q: On the lagging strand primase places some RNA primers. The DNA polymerase then synthesizes the Okazaki fragments and another DNA polymerase follows and digests the RNA primer with its 5'-3'-exonuclease site. Does this not result in loss of genetic information? I thought the ligase just stuck the 5' end of one nucleotide and the 3' of the other together or does it also replace the nucleotides digested by the 5'-3'-exonuclease?*
- *A: DNA polymerase synthesizes the Okazaki fragments starting from an RNA primer. When it bumps into the 5' end of the next RNA primer it digests it in the 5' to 3' direction. This allows DNA polymerase to continue synthesizing DNA through the region where the primer was. Eventually it will bump into the DNA itself and then it will stop and dissociate itself from the DNA. A ligase can then come along and connect the two pieces. No information should be lost in this process, since the new DNA is always synthesized to be complementary to the template.*
- *Q: Which one is the enzyme responsible for chewing up the Okazaki fragments in the lagging strand? Is it the DNA polymerase who removes them and then synthesizes the DNA fragment?*
- *A: It is only the RNA primer portions of the Okazaki fragments that get chewed up by the 5'-3' exonuclease site of DNA polymerase. It is very likely the same copy of the polymerase that is synthesizing the 'next' Okazaki fragment along the strand.*

DNA must be replicated many times to make to²⁵ go from an egg to an adult



The worm *Caenorhabditis elegans* is composed of 956 cells. The lineage from the egg has been traced for every one of these cells!



Cells are **differentiating** as they undergo divisions. This is the process of **development**.

The cells that have the potential to become many other types cells are the 'stem cells'



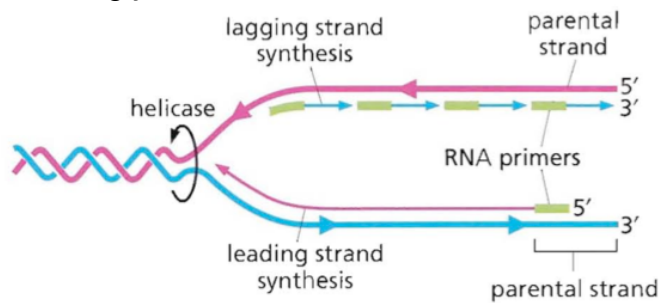
Chapter 10 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

- A cell lineage is the sequence of divisions that led from a previous cell (ultimately the egg and sperm) to a descendent cell.
- All cells must eventually stop undergoing divisions at some point, since they have reached their 'final form'
- Differentiation: a less specialized cell becoming a more specialized cell.
- In *C. elegans*, cell division effectively stops in the adult animal. In contrast, cells continue to divide through the whole life a mammal as old cells are replaced with newer ones.
- A human cell typically undergoes 50-80 cycles of division.
- An adult human has 10^{14} cells that may undergo as many as 10^{16} divisions in a lifetime.
- One hallmark of cancer is that a cell keeps dividing long after it should have stopped and undergone programmed cell death (apoptosis). Cells that keep dividing for (effectively) forever are said to be immortal.
- One of the most famous examples of an immortal **cell line** are those taken from an aggressive cervical cancer (cervical adenocarcinoma) of a woman named Henrietta Lacks in 1951. These cells (widely known as HeLa cells) continue to replicate to this day and are used in labs all around the world. They seem to have no limits to their ability to divide.
- lvlofdetail.com and bulbapedia.bulbagarden.net

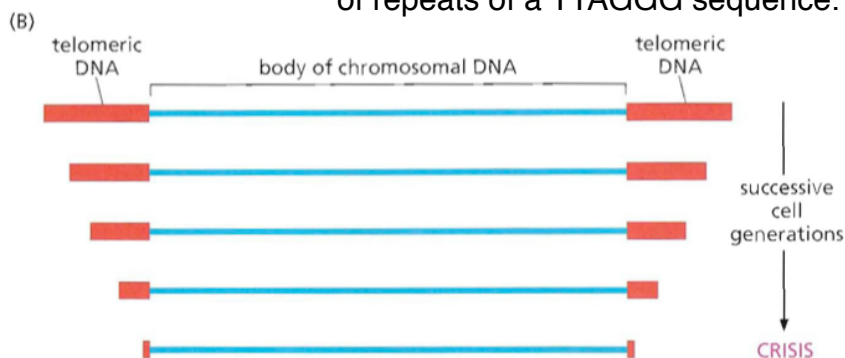
What limits the ability of DNA to be copied? ²⁶

As we saw a few slides ago, a primer is needed to start DNA replication. One problem is that this primer never starts exactly at the end of each chromosome.

Accordingly, chromosomes shorten with each replication!



For this reason, each chromosome has a built in 'buffer zone' at the ends of each chromosome known as the **telomeres**. Telomeres are made of thousands of repeats of a TTAGGG sequence.



When the telomeres get very short, the cell reaches a 'crisis' point and undergoes apoptosis (programed cell death)

Chapter 10 of Robert A. Weinberg, The Biology Of Cancer, Volume 1; Garland Pub, 2007

- Telomeres were discovered by Barbara McClintock in 1941 and for this she was awarded the 1983 Nobel Prize in Physiology or Medicine.
- The shortening of telomeres is a sort of 'molecular clock' that tells the age of the cells (number of generations).
- The primase enzyme creates a primer some distance from the 3' end of the template DNA strand, and thus the new strand is shorter. Even if the primase enzyme was perfect, the RNA would be degraded and the new strand would be about 10 nucleotides shorter. This is known as the 'end-replication problem'. Telomeres generally shorten by about 50-100 base pairs per cell division.
- When the telomeres are so short that the chromosomal DNA is no longer protected, the cell undergoes apoptosis.
- Stem cells seem to be better at maintaining their telomeres through multiple divisions, but there is still some shortening over time.