DNA:

Structure and Replication

DNA: the stuff of life. Well, not really, despite the hype. DNA does contain the instructions to make a lot of the stuff of life (proteins), although again, not all the stuff of life. At least not directly. Deoxyribonucleic acid (and its very close cousin ribonucleic acid, or RNA) is a very long chain polymer. You may recall that a polymer is just a really big molecule made by connecting many small similar molecules together). In this polymer, the small (monomer) molecules are known as nucleotides, and are composed of a pentose (5-carbon sugar either deoxyribose or ribose), a nitrogenous base, and a phosphate group. As you can see in the figure, the nucleotides only vary slightly, and only in the nitrogenous base. In the case of DNA, those bases are adenine, guanine, cytosine, and thymine. Note the similarity of the shapes of adenine and guanine, and also the similarity between cytosine and thymine. A and G are classified as purines, while C and T are classified as pyrimidines. As long as we're naming things, notice "deoxyribose" and "ribose". As the name implies, deoxyribose is just a ribose without an oxygen. More specifically, where there is a hydroxyl group attached to the 2-carbon of ribose, there is only a hydrogen attached to the 2-carbon of deoxyribose. That is the only difference between the two sugars.

In randomly constructing a single strand of nucleic acid *in vitro*, there are no particular rules regarding the ordering of the nucleotides with respect to their bases. The identities of their nitrogenous bases are irrelevant because the nucleotides are attached by phosphodiester bonds through the phosphate group and the pentose. It is therefore often referred to as the sugar-phosphate backbone. If we break down the word "phosphodiester", we see that it quite handily describes the connection: the sugars are connected by two ester bonds (–O–) with a phosphorous in between. One of the ideas that often confuses students, is the directionality of this bond, and therefore, of nucleic acids in general. For example, when we talk about DNA polymerase, the enzyme that catalyzes the addition of nucleotides in living cells, we say that it works in a 5-prime (5') to 3-prime (3') direction. This may seem like arcane molecular-biologist-speak, but it is actually very simple. Take another look at two of the nucleotides joined together by the phosphodiester bond (fig. 1, bottom left). An adenine nucleotide is joined to a

Using this book: This book is designed to be used in both introductory and advanced cell biology courses. The primary text is generally on the left side of the vertical divider, and printed in black. Details that are usually left to an advanced course are printed in blue and found on the right side of the divider. Finally, additional biomedically relevant information can be found in red print on either side of the divider.

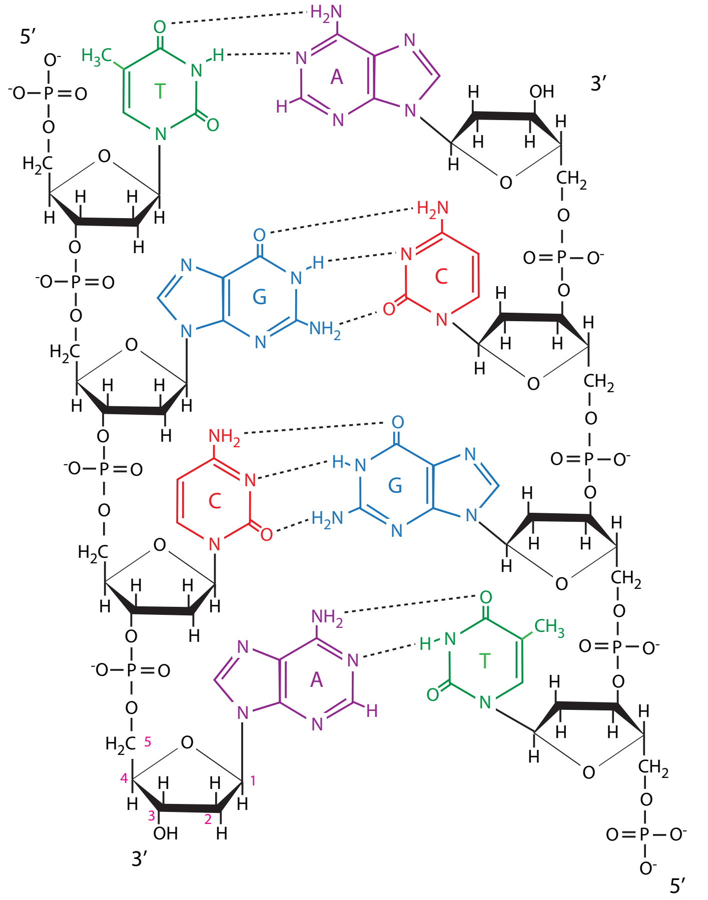


Figure 1. DNA. Deoxyribonucleic acid is a polymer chain of nucleotides connected by 5' to 3' phosphodiester bonds. DNA normally exists as a two antiparallel complementary strands held together by hydrogen bonds between adenines (A) and thymines (T), and between guanines (G) and cytosines (C).

cytosine nucleotide. The phosphodiester bond will always link the 5-carbon of one deoxyribose (or ribose in RNA) to the 3-carbon of the next sugar. This also means that on one end of a chain of linked nucleotides, there will be a free 5' phosphate (-PO4) group, and on the other end, a free 3' hydroxyl (-OH). These define the directionality of a strand of DNA or RNA.

DNA is normally found as a double-stranded molecule in the cell whereas RNA is mostly single-stranded. It is important to understand though, that under the appropriate conditions, DNA could be made single-stranded, and RNA can be double-stranded. In fact, the molecules are so similar that it is even possible to create double-stranded hybrid molecules with one strand of DNA and one of RNA. Interestingly, RNA-RNA double helices and RNA-DNA double helices are actually slightly more stable than the more conventional DNA-DNA double helix.

The basis of the double-stranded nature of DNA, and in fact the basis of nucleic acids as the medium for storage and transfer of genetic information, is *base-pairing*. Base-pairing refers to the formation of hydrogen bonds between adenines and thymines, and between guanines and cytosines. These pairs are significantly more stable than any association formed with the other possible bases. Furthermore, when these base-pair associations form in the context of two strands of nucleic acids, their spacing is also uniform and highly stable. You may recall that hydrogen bonds are relatively weak bonds. However, in the context of DNA, the hydrogen bonding is what makes DNA extremely stable and therefore well suited as a long-term storage medium for genetic information. Since even in simple prokaryotes, DNA double helices are at least thousands of nucleotides long, this means that there are several thousand hydrogen bonds holding the two strands together. Although any individual nucleotide-to-nucleotide hydrogen bonding interaction could easily be temporarily disrupted by a slight increase in temperature, or a miniscule change in the ionic strength of the solution, a full double-helix of DNA requires very high temperatures (generally over 90 ºC) to completely denature the double helix into individual strands.

Because there is an exact one-to-one pairing of nucleotides, it turns out that the two strands are essentially backup copies of each other - a safety net in the event that nucleotides are lost from one strand. In fact, even if parts of both strands are damaged, as long as the other strand is intact in the area of damage, then the essential information is still there in the complementary sequence of the opposite strand and can be written into place. Keep in mind though, that while one strand of DNA can thus act as a "backup" of the other, the two strands are not identical - they are complementary. An intererresting consequence of this system of complementary and antiparallel strands is that the two strands can each carry unique information.

Bi-directional gene pairs are two genes on opposite strands of DNA, but sharing a promoter, which lies in between them. Since DNA can only be made in one direction, 5' to 3', this bi-directional promoter, often a CpG island (see next chapter), thus sends the RNA polymerase for each gene in opposite physical directions. This has been shown for a number of genes involved in cancers (breast, ovarian), and is a mechanism for coordinating the expression of networks of gene products.

The strands of a DNA double-helix are antiparallel. This means that if we looked at a double-helix of DNA from left to right, one strand would be constructed in the 5' to 3' direction, while the complementary strand is constructed in the 3' to 5' direction. This is important to the function of enzymes that create and repair DNA, as we will be discussing soon. In fig. 1, the left strand is 5' to 3' from top to bottom, and the other is 5' to 3' from bottom to top.

From a physical standpoint, DNA molecules are negatively charged (all those phosphates), and normally a double-helix with a right-handed twist. In this normal (also called the "B" conformation) state, one full twist of the molecule encompasses 11 base pairs, with 0.34 nm between each nucleotide base. Each of the nitrogenous bases are planar, and when paired with the complementary base, forms a flat planar "rung" on the "ladder" of DNA. These are perpendicular to the longitudinal axis of the DNA. Most of the free-floating DNA in a cell, and most DNA in any aqueous solution of near-physiological osmolarity and pH, is found in this B conformation. However, other conformations have been found, usually under very specific environmental circumstances. A compressed conformation, A-DNA, was observed as an artifact of *in vitro* crystallization, with slightly more bases per turn, shorter turn length, and base-pairs that are not perpendicular to the longitudinal axis. Another, Z-DNA, appears to form transiently in GC-rich stretches of DNA in which, interestingly, the DNA twists the opposite direction.

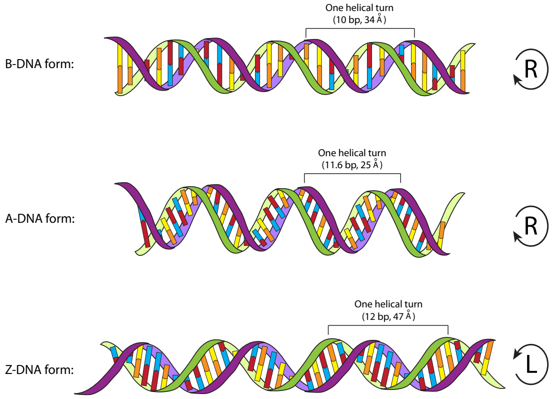

One helical turn (10 bp, 34 Å)
B-DNA form:
R
One helical turn (11.6 bp, 25 Å)
A-DNA form:
R
One helical turn (12 bp, 47 Å)
Z-DNA form:
L

Figure 2. Three conformations of DNA. B-DNA is most common, A-DNA is likely an artifact of crystallization in vitro, and Z-DNA may form transiently in parts of the chromosome.

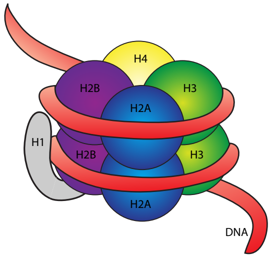
In prokaryotes, the DNA is found in the cytoplasm (rather obvious since there is no other choice in those simple organisms), while in eukaryotes, the DNA is found inside the nucleus. Despite the differences in their locations, the level of protection from external forces, and most of all, their sizes, both prokaryotic and eukaryotic DNA is packaged with proteins that help to organize and stabilize the overall chromosome structure. Relatively little is understood with regard to prokaryotic chromosomal pack-

It has been suggested that both the A and Z forms of DNA are, in fact, physiologically relevant. There is evidence to suggest that the A form may occur in RNA-DNA hybrid double helices as well as when DNA is complexed to some enzymes. The Z conformation may occur in response to methylation of the DNA. Furthermore, the "normal" B-DNA conformation is something of a idealized structure based on being fully hydrated, as is certainly very likely inside a cell. However, that hydration state is constantly changing, albeit minutely, so the DNA conformation will often vary slightly from the B-conformation parameters in figure 2.

aging although there are structural similarities between some of the proteins found in prokaryotic and eukaryotic chromosomes. Therefore, most introductory cell biology courses stick to eukaryotic chromosomal packaging.

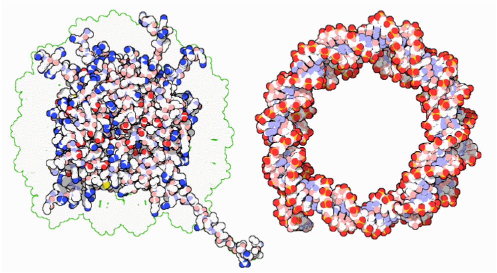
Naked DNA, whether prokaryotic or eukaryotic, is an extremely thin strand of material, roughly 11 nm in diameter. However, given the size of eukaryotic genomes, if the DNA was stored that way inside the nucleus, it would become unmanageably tangled. Picture a bucket into which you have tossed a hundred meters of yarn without any attempt whatsoever to organize it by coiling it or bunching it. Now consider whether you would be able to reach into that bucket pull on one strand, and expect to pull up only one strand, or if instead you are likely to pull up at least a small tangle of yarn. The cell does essentially what you would do with the yarn to keep it organized: it is packaged neatly into smaller, manageable skeins. In the case of DNA, each chromosome is looped around a histone complex to form the first order of chromosomal organization: the nucleosome.

Figure 4. The nucleosome is composed of slightly over two turns of DNA around a histone core containing two copies each of H2A, H2B, H3, and H4 histones. The H1 histone is not part of the core unit and functions in coordinating interaction between nucleosomes.



Histones are a family of basic (positively-charged) proteins. They all function primarily in organizing DNA, and the nucleosome is formed when DNA wraps (a little over 2 times) around a core of eight histones - two each of H2A, H2B, H3, and H4. The number and position of the positive charges (mostly from lysines and arginines) are crucial to their ability to tightly bind DNA, which as previously pointed out, is very negatively charged. That "opposites attract" idea is not just a dating tip from the advice columns.

Figure 5. Nucleosomes. (left) Core histone complex - the nitrogens indicative of positively-charged side chains are in blue. (right) DNA wraps around the histone. The red oxygens surrounding the yellow phosphorus atoms are the negatively-charged phosphate groups. Figure from RCSB Protein Data Bank (http://www.rcsb.org).



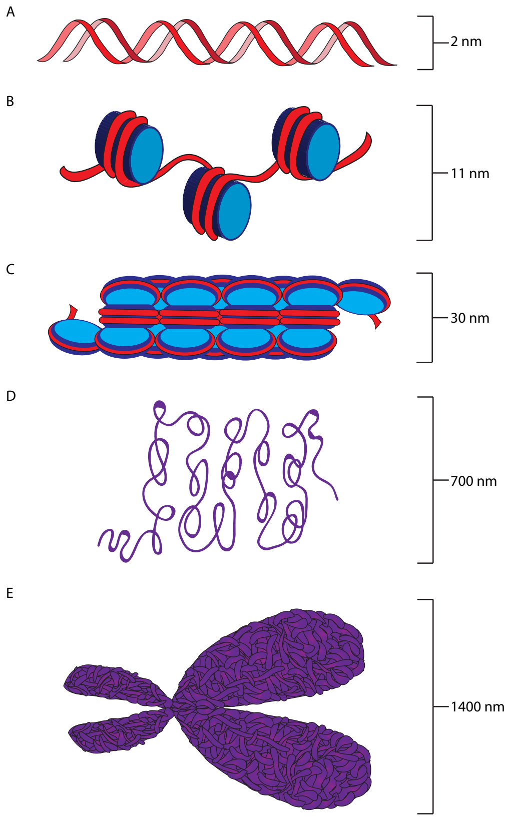


Figure 3. DNA packaging. (A) A naked strand of DNA is approximately 2nm in diameter. (B) Histones, which are octameric proteins depicted here as a roughly cylindrical protein, have positive charges distributed on the outer surface to interact with the negatively-charged DNA backbone. (C) Even the organization afforded by histone binding can leave an unmanageable tangle of DNA, especially with longer eukaryotic genomes, and therefore the histone-bound DNA is packaged into the "30-nm strand". This is held together, in part, by histone interactions. (D) The 30-nm fibers are looped into 700-nm fibers, which are themselves formed into the typical eukaryotic chromosome (E).

The 30-nm fiber is held together by two sets of interactions. First, the linker histone, H1, brings the nucleosomes together into an approximate 30-nm structure. This structure is then stabilized by disulfide bonds that form between the H2A histone of one nucleosome and the H4 histone of its neighbor.

Upon examination of the 3D structure of the histone core complex, we see that while relatively uncharged protein interaction domains hold the histones together in the center, the positively charged residues are found around the outside of the complex, available to interact with the negatively charged phosphates of DNA.

In a later chapter, we will discuss how enzymes read the DNA to transcribe its information onto smaller, more manageable pieces of RNA. For now, we only need to be aware that at any given time, much of the DNA is packaged tightly away, while some parts of the DNA are not. Because the parts that are available for use can vary depending on what is happening to/in the cell at any given time, the packaging of DNA must be dynamic. There must be a mechanism to quickly loosen the binding of DNA to histones when that DNA is needed for gene expression, and to tighten the binding when it is not. As it turns out, this process involves *acetylation* and *deacetylation* of the histones.

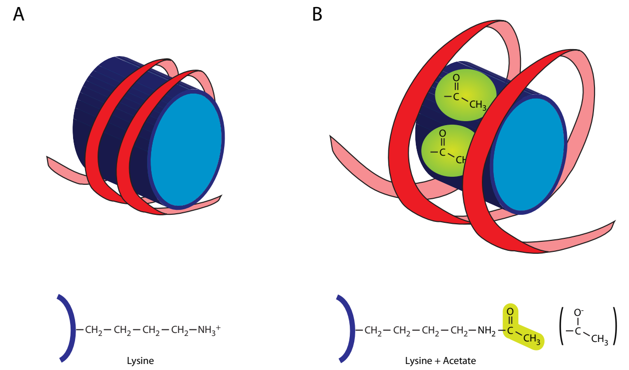


Figure 6. (A) Deacetylated histone allows interaction between the negatively charged phosphates of the DNA and the positively charged lysines of the histone. (B) When the histone is acetylated, not only is the positive charge on the lysine lost, the acetyl group also imparts a negative charge, repelling the DNA phosphates.

Histone Acetyltransferases (HATs) are enzymes that place an acetyl group on a lysine of a histone protein. The acetyl groups are negatively charged, and the acetylation not only adds a negatively charged group, it also removes the positive charge from the lysine. This has the effect of not only neutralizing a point of attraction between the protein and the DNA, but even slightly repelling it (with like charges). On the other side of the mechanism, Histone Deactylases (HDACs) are enzymes that remove the acetylation, and thereby restore the interaction between histone protein and DNA. Since these are such important enzymes, it stands to reason that they are not allowed to operate willy-nilly on any available histone, and in fact, they are often found in a complex with other proteins that control and coordinate their activation with other processes such as activation of transcription.

Semi-Conservative DNA Replication

DNA replication is similar to transcription in its most general idea: a polymerase enzyme reads a strand of DNA one nucleotide at a time, it takes a random nucleotide from the nucleoplasm, and if it is complementary to the nucleotide in the DNA, the polymerase adds it to the new strand it is creating. Of course, there are significant differences between replication and transcription too, not the least of which is that both strands of DNA are being read simultaneously in order to create two new complementary strands that will eventually result in a complete and nearly perfect copy of an entire organismal genome.

One of the most important concepts of DNA replication is that it is a *semi-conservative* process (fig. 7). This means that every double helix in the new generation of an organism consists of one complete "old" strand and one complete "new" strand wrapped around each other. This is in contrast to the two other possible models of DNA replication, the conservative model, and the dispersive model. A conservative mechanism of replication proposes that the old DNA is used as a template only and is not incorporated into the new double-helix. Thus the new cell has one completely new double-helix and one completely old double-helix. The dispersive model of replication posits a final product in which each double helix of DNA is a mixture of fragments of old and new DNA. In light of current knowledge, it is difficult to imagine a dispersive mechanism, but at the time, there were no mechanistic models at all. The Meselson-Stahl experiments (1958) clearly demonstrated that the mechanism must be semi-conservative, and this was confirmed once the key enzymes were discovered and their mechanisms elucidated.

Prokaryotic Replication

DNA replication begins at an origin of replication. There is only one origin in prokaryotes (in E. coli, oriC) and it is characterized by arrays of repeated sequences. These sequences wrap around a DNA-binding protein, and in doing so, exert pressure on the H-bonds between the strands of DNA, and the chromosome begins to unzip in an AT-rich area wrapped around this protein. Remember that A-T pairs are 33% weaker than G-C pairs due to fewer hydrogen bonds. The use of AT-rich stretches of DNA as points of strand separation is a recurring theme through a variety of DNA operations. The separation of the two strands is bidirectional, and DNA polymerases will act in both directions in order to finish the process as quickly as possible. Speed is important here because while replication is happening, the DNA is vulnerable to breakage, and most metabolic processes are shut down to devote the energy to the replication. Even in prokaryotes, where DNA molecules are orders of magnitude smaller than in eukaryotes,

In the Meselson-Stahl experiments, E. coli were first incubated with ^{15}N, a heavy isotope of nitrogen. Although it is only a difference in mass of one neutron per atom, there is a great enough difference in mass between heavy nitrogen-containing DNA (in the purine and pyrimidine bases) and light/normal nitrogen-containing DNA that they can be separated from one another by ultracentrifugation through a CsCl concentration gradient (fig. 7).

Over 14 generations, this led to a population of E. coli that had heavy nitrogen incorporated into all of the DNA (shown in blue below). Then, the bacteria are grown for one or two divisions in "light" nitrogen, ^{14}N. When the DNA from the bacterial populations was examined by centrifugation, it was found that instead of light DNA and heavy DNA, as would be expected if DNA replications was conservative, there was a single band in and intermediate position on the gradient. This supports a semi-conservative model in which each strand of original DNA not only acts as a template for making new DNA, it is itself incorporated into the new double-helix.

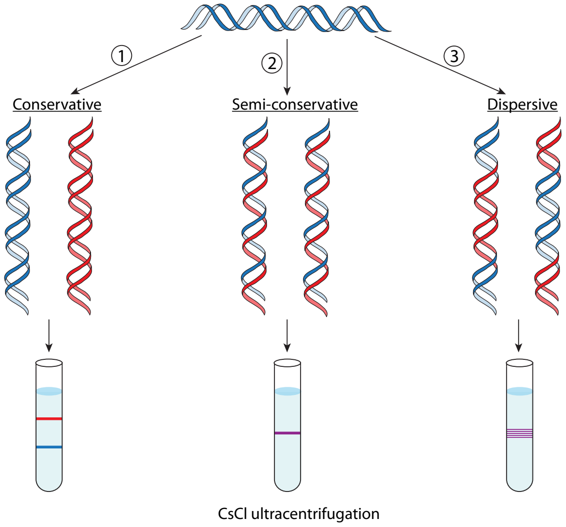


Figure 7. DNA replication. Prior to the discovery of the enzymes involved in replication, three general mechanisms were proposed. In conservative replication, the original DNA strands stay associated with each other, while the newly made DNA forms its own double-helix. Semi-conservative replication posits the creation of hybrid old-new double helices. Dispersive replication proposed molecules composed of randomized fragments of double-old and double-new DNA.

the size of the DNA molecule when it is unraveled from protective packaging proteins makes it highly susceptible to physical damage just from movements of the cell.

The first OriC binding protein, DnaA, binds to DnaA boxes, which are 9 base pair segments with a consensus sequence of TTATCCACA. OriC has five of these repeats, and one DnaA protein binds to each of them. HU and IHF are histone-like proteins that associate with DnaA and together bend that part of the DNA into a circular loop, situating it just over the other major feature of oriC, the 13-bp AT-rich repeats (GATCTNTTNTTTT). DnaA hydrolyzes ATP and breaks the H-bonds between strands in the 13mer repeats, also known as melting the DNA. This allows complexes of DnaB [and DnaC, which is a loading protein that helps attach DnaB(6) to the strand with accompanying hydrolysis of ATP. Also, five more DnaA are recruited to stabilize the loop.] to bind to each single-stranded region of the DNA on opposite sides of the newly opened replication bubble.

DnaB is a helicase; its enzymatic activity is to unzip/unwind the DNA ahead of the DNA polymerase, to give it single-stranded DNA to read and copy. It does so in association with single-stranded-DNA binding proteins (SSBs), and DNA gyrase. The function of SSB is nearly self-explanatory: single-stranded DNA is like RNA in its ability to form complex secondary structures by internal base-pairing, so SSB prevents that. DNA gyrase is a type II topoisomerase, and is tasked with introducing negative supercoiling to the DNA. This is necessary because the unzipping of the DNA by helicase also unwinds it (since it is a double helix) and causes the introduction of positive supercoiling. This means that the entire circular molecule twists on itself: imagine holding a rubber band in two hands and twisting it. As the supercoiling accumulates, the DNA becomes more tightly coiled, to the point that it would be impossible for helicase to unzip it. DnaB/gyrase can relieve this stress by temporarily cutting the double-stranded DNA, passing a loop of the molecule through the gap, and resealing it. This (hopefully) makes a lot more sense

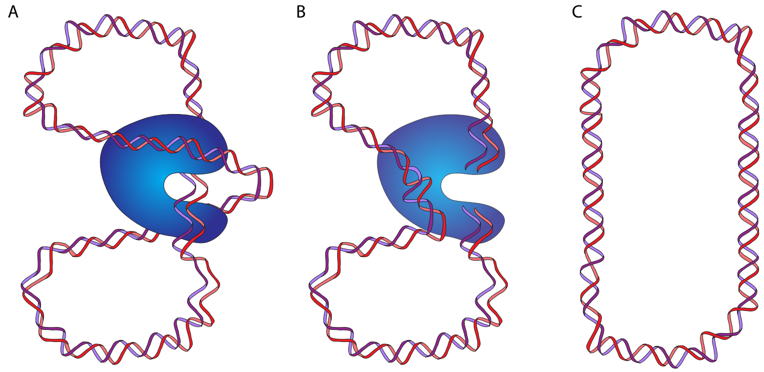


Figure 8. Type II DNA topoisomerases like DNA gyrase relieve supercoiling by making temporary double-strand cut.

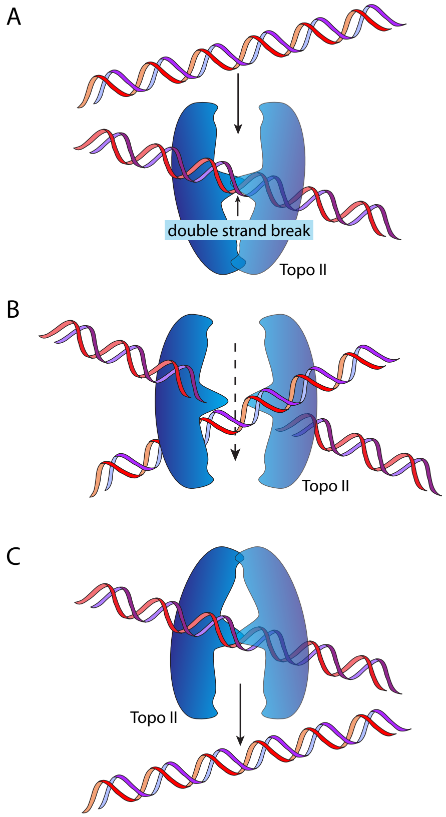


Figure 9. Detail of DNA topoisomerase type II action. (A) First, the enzyme binds to the DNA and initiates an endonuclease activity, cutting both strands of the DNA at that point. (B) This complete transection of the DNA (as opposed to the single-strand cut by type I topoisomerases) allows another part of the same DNA molecule to slip through the gap. (C) Finally, the two temporarily broken ends of the DNA, which had been held closely in place by the enzyme.

looking at the diagram. Or, going back to our rubber band, give the rubber band a twist or two, then tape down the two ends. If you snip the rubber band, and pass an adjacent portion of the rubber band through that snip, then reconnect the cut ends, you will find that there is one less twist. Nifty, eh? At this point, some of you are going to say, but if you twist a free-floating rubber band, as one might imagine a free-floating circular DNA chromosome in E. coli, you would expect it to naturally untwist. Technically, yes, but due to the large mass of the chromosome, its association with various proteins and the cell membrane, and the viscosity of its environment, it does not behave as though it were completely free.

Once oriC has been opened and the helicases have attached to the two sides of the replication fork, the replication machine, aka the replisome can begin to form. However, before the DNA polymerases take positions, they need to be primed. *DNA polymerases are unable to join two individual free nucleotides* together to begin forming a nucleic acid; they can only add onto a pre-existing strand of at least two nucleotides. Therefore, a specialized RNA polymerase (RNAP's do not have this limitation) known as primase is a part of the replisome, and reads creates a short RNA strand termed the *primer* for the DNA polymerase to add onto. Although only a few nucleotides are needed, the prokaryotic primers may be as long as 60 nt depending on the species.

At least five prokaryotic DNA polymerases have been discovered to date. The primary DNA polymerase for replication in E. coli is DNA Polymerase III (Pol III). Pol I is also involved in the basic mechanism of DNA replication, primarily to fill in gaps created during lagging strand synthesis (defined 3 pages ahead) or through error-correcting mechanisms. DNA polymerase II and the recently discovered Pol IV and Pol V do not participate in chromosomal replication, but rather are used to synthesize DNA when certain types of repair is needed at other times in the cellular life cycle.

DNA polymerase III is a multi-subunit holoenzyme, with α, ε, and θ subunits comprising the core polymerase, and τ, γ, δ, δ', χ, ψ, and β coming together to form the complete holoenzyme. The core polymerase has two activities: the α subunit is the polymerase function, reading a strand of DNA and synthesizing a complementary strand with great speed, around 150 nt/sec; the ε subunit is a 3'-5' "proofreading" exonuclease and acts as an immediate proofreader, removing the last nucleotide if it is incorrect. This proofreading does not reach any further back: it only acts on the most recently added nucleotide to correct misincorporation. Other mechanisms and enzymes are used to correct DNA lesions that arise at other times. [As a matter of nomenclature, *exo*nucleases only cut off nucleotides from DNA or RNA from either end, but not in the middle. *Endo*nucleases cleave phosphodiester bonds located deeper within a nucleic acid strand.] The θ subunit has no enzymatic activity and regulates the exonuclease function. Although it has polymerase activity, the Pol III core polymerase has poor processivity - that is, it can only add up to 15 nucleotides before dissociating from the template DNA. Since genomes of E. coli strains average near 5 million base pairs, replication in little 15 nt segments would be extraordinarily inefficient.

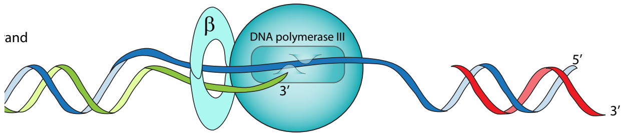


Figure 10. The dimeric β clamp holds DNA Polymerase III on the template, allowing it polymerize more nucleotides before dissociating.

This is where the β subunit is needed. Also known as the β clamp, it is a dimer of semicircular subunits that has a central hole through which the DNA is threaded. The core polymerase, via an α-β interaction, is attached to this β clamp so that it stays on the DNA longer, increasing the processivity of Pol III to over 5000nt. The β clamp is loaded onto (and unloaded off of) the DNA by a clamp loader complex (also called γ complex) consisting of γ (x3), δ, δ', χ, and ψ subunits.

The replication bubble has two replication forks - once the DNA is opened up (unzipped) at the origin, a replication machine can form on each end, with the helicases heading in opposite directions. For simplification, we will consider just one fork – opening left to right – in this discussion with the understanding that the same thing is happening with the other fork, but in the opposite direction.

The first thing to notice when looking at a diagram of a replication fork (fig. 11) is that the two single-stranded portions of template DNA are anti-parallel. This should come as no surprise at this point in the course, but it does introduce an interesting mechanical problem. Helicase opens up the double stranded DNA and leads the rest of the replication machine along. So, in the single-stranded region trailing the helicase, if we look left to right, one template strand is 3' to 5' (in blue), while the other is 5' to 3' (in red). Since we know that nucleic acids are polymerized by adding the 5' phosphate of a new nucleotide to the 3' hydroxyl of the previous nucleotide (5' to 3', in green), this means that one of the strands, called the *leading strand*, is being synthesized in the same direction that the replication machine moves. No problem there.

The other strand is problematic: looked at linearly, the newly synthesized strand would be going 3' to 5' from left to right but DNA polymerases cannot add nucleotides that way. How do cells resolve this problem? A number of possibilities have been proposed,

The clamp loader complex is an ATPase assembly that binds to the β-clamp unit upon binding of ATP (but the ATPase activity is not turned on). When the complex then binds to DNA, it activates the ATPase, and the resulting hydrolysis of ATP leads to conformational changes that open up the clamp temporarily (to encircle or to move off of the DNA strand), and then dissociation of the clamp loader from the clamp assembly.

Understanding the mechanics of DNA replication is a highly visual process, and it is recommended that students frequently flip back and forth between Figure 11 and the text description of replication. In fact, with the extra space around Figure 11, we recommend writing you own description of the process to help understand the mechanism step by step.

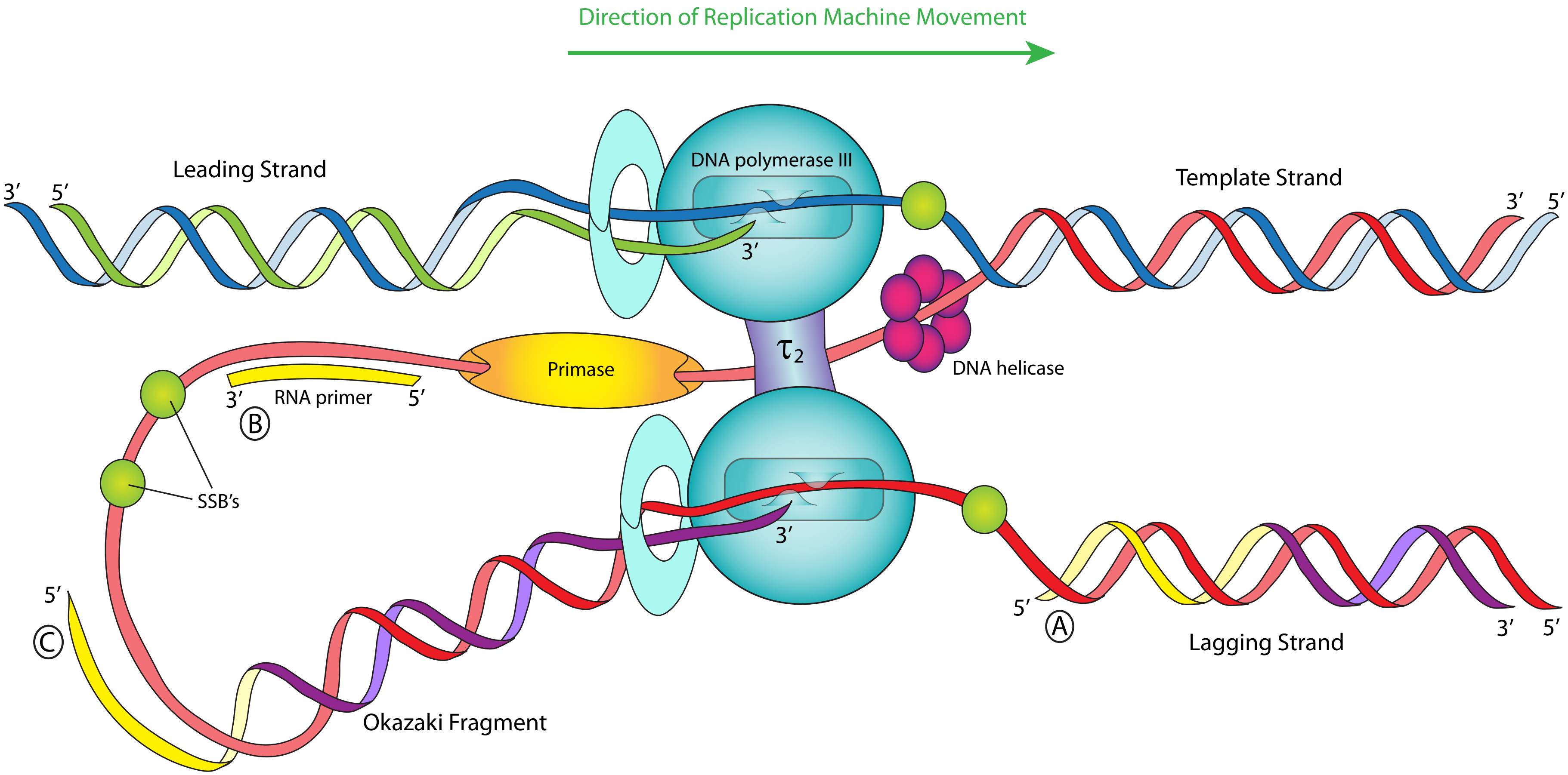

Direction of Replication Machine Movement
Leading Strand
3′
5′
DNA polymerase III
3′
Template Strand
3′
5′
τ_2
DNA helicase
Primase
3′
RNA primer
5′
Ⓑ
SSB's
5′
Ⓒ
3′
5′
Ⓐ
Lagging Strand
3′
5′
Okazaki Fragment

Figure 11. DNA Replication in prokaryotes.

but the current model is depicted here. The replication machine consists of the helicase, primases, and two DNA polymerase III holoenzymes moving in the same physical direction (following the helicase). In fact, the pol III complexes are physically linked through τ subunits.

In order for the template strand that is 5' to 3' from left to right to be replicated, the strand must be fed into the polymerase backwards. This can be accomplished either by turning the polymerase around or by looping the DNA around. As the figure shows, the current model is that the primase is also moving along left to right, so it has just a short time to quickly synthesize a short primer before having to move forward with the replisome and starting up again, leaving intermittent primers in its wake. Because of this, Pol III is forced to synthesize only short fragments of the chromosome at a time, called Okazaki fragments after their discoverer. Pol III begins synthesizing by adding nucleotides onto the 3' end of a primer and continues until it hits the 5' end of the next primer. It does not (and *can not*) connect the strand it is synthesizing with the 5' primer end.

DNA replication is called a semi-discontinuous process because while the leading strand is being synthesized continuously, the lagging strand is synthesized in fragments. This leads to two major problems: first, there are little bits of RNA left behind in the newly made strands (just at the 5' end for the leading strand, in many places for the lagging); and second, Pol III can only add free nucleotides to a fragment of single stranded DNA; it cannot connect another fragment. Therefore, the new "strand" is not whole, but riddled with missing phosphodiester bonds.

The first problem is resolved by DNA polymerase I. Unlike Pol III, Pol I is a monomeric protein and acts alone, without additional proteins. There are also 10-20 times as many Pol I molecules as there are Pol III molecules, since they are needed for so many Okazaki fragments. DNA Polymerase I has three activities: (1) like Pol III, it can synthesize a DNA strand based on a DNA template, (2) also like Pol III, it is a 3'-5' proofreading exonuclease, but unlike Pol III, (3) it is also a 5'-3' exonuclease. The 5'-3' exonuclease activity is crucial in removing the RNA primer (fig. 12). The 5'-3' exonuclease binds to double-stranded DNA that has a single-stranded break in the phosphodiester backbone such as what happens after Okazaki fragments have been synthesized from one primer to the next, but cannot be connected. This 5'-3' exonuclease then removes the RNA primer. The polymerase activity then adds new DNA nucleotides to the upstream Okazaki fragment, filling in the gap created by the removal of the RNA primer. The proofreading exonuclease acts just like it does for Pol III, immediately removing a newly incorporated incorrect nucleotide. After proofreading, the overall error rate of nucleotide incorporation is approximately 1 in 10^7.

Technically, the 5'-3' exonuclease cleaves the DNA at a double-stranded region downstream of the nick, and may then remove anywhere from 1-10nt at a time. Experimentally, the 5'-3' exonuclease activity can be cleaved from the rest of Pol I by the protease trypsin. This generates the "Klenow fragment" containing the polymerase and 3'-5' proofreading exonuclease.

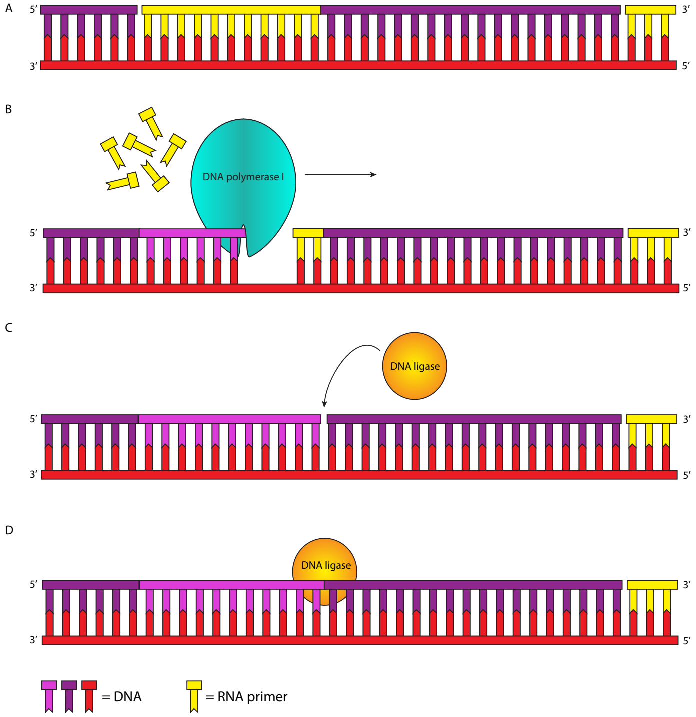


Figure 12. Lagging Strand Synthesis. After DNA polymerase III has extended the primers (yellow), DNA polymerase I removes the primer and replaces it by adding onto the previous fragment. When it finishes removing RNA, and replacing it with DNA, it leaves the DNA with a missing phosphodiester bond between the pol III-synthesized DNA downstream and the pol I-synthesized DNA upstream. This break in the sugar-phosphate backbone is repaired by DNA ligase.

Even though the RNA has been replaced with DNA, this still leaves a fragmented strand. The last major player in the DNA replication story finally appears: DNA ligase. This enzyme has one simple but crucial task: it catalyzes the attack of the 3’-OH from one fragment on the 5’ phosphate of the next fragment, generating a phosphodiester bond. This reaction requires energy in the form of hydrolysis of either ATP or NAD+ depending on the species (E. coli uses NAD$^+$) generating AMP and either PP$_i$ or NMN$^+$.

Eukaryotic Replication

Given the crucial nature of chromosomal replication for life to exist, it is not surprising to find that eukaryotic DNA replication is very similar to the prokaryotic process. This section will highlight some of the differences, which are generally elaborations on the prokaryotic version.

Unlike prokaryotes, eukaryotic chromosomes often have multiple origins of replication. Considering the size of eukaryotic chromosomes, this is necessary to finish complete replication in a timely manner. Each of these origins defines a *replicon*, or the stretch of the DNA that is replicated from a particular origin. The replicons do not replicate at exactly the same time (although all within the same phase of the cell cycle, see chapter 15), so it is important to make sure that replicons are used only once during a cell cycle.

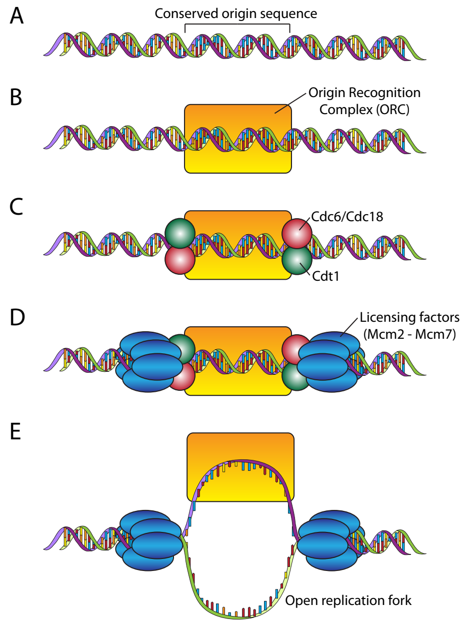

A
Conserved origin sequence
B
Origin Recognition Complex (ORC)
C
Cdc6/Cdc18
Cdt1
D
Licensing factors (Mcm2 - Mcm7)
E
Open replication fork

Figure 13. Eukaryotic Origin of Replication

This requires a "licensing" mechanism. During the cell cycle phase before DNA replication is initiated, a pre-replicative complex is assembled at each origin (Fig. 13). The origins are highly variable in composition and length, ranging from ~100 to well over 10000 base pairs. The pre-RC proteins, on the other hand, are very highly conserved. The pre-RC begins by making the ORC (origin recognition complex, not a creature battling Frodo and Aragorn), which is comprised of six subunits (Orc1-Orc6). Although there is not significant sequence homology, the ORC approximates the function of the DnaA protein in E. coli. To complete the pre-RC, the ORC recruits a pair of proteins, Cdc6 and Cdt1 to each side, and they bind the MCM complex (a hexamer of Mcm2-Mcm6 that has an inactive helicase activity), leading to the fully licensed pre-RC. The origin is now ready for activation.

Activation of the pre-RC at an origin of replication requires first Mcm10, which facilitates protein kinases that phosphorylate the MCM complex, activating the helicase activity, and making the replication fork ready to accept the replication machine.

The components of the eukaryotic replication machine may have different names from the prokaryotic, but the functions should be very familiar (Fig. 14). There is a primase to make RNA primers, and tightly associated with the primase is DNA polymerase α. Pol α has neither a 5'-3' exonuclease nor a 3'-5' proofreading exonuclease, but it can synthesize DNA. However, this is not the primary DNA polymerase. The primase/pol α complex is essentially a fancy primase that begins with a short <10nt RNA primer and then adds another short ~15nt DNA fragment, making a *hybrid primer*.

Replication factor C (RFC), acting like the prokaryotic clamp loader complex, then replaces pol α with PCNA, the eukaryotic version of the β clamp. This then recruits DNA polymerase δ, which is the primary replicative DNA polymerase, equivalent to prokaryotic Pol III in function, and necessary for both leading and lagging strand synthesis. Finally, instead of SSB, eukaryotic cells use replication protein A (RPA) to organize and control single-stranded DNA as it is generated during the replicative process.

You may have noticed that none of the eukaryotic DNA polymerases discussed so far have a 5' to 3' exonuclease activity as used by prokaryotic DNA polymerase I to remove primers. In place of that, RNaseH1 and and FEN-1 remove the RNA primers (all but one ribonucleotide, and the last ribonucleotide, respectively). Interestingly, FEN-1 also excises chunks of lagging strand DNA within about 15 base pairs of the RNA if they contain mistakes. This seems to help alleviate the problem of lower fidelity of replication by pol α, which has no proofreading capability. After RNaseH1 and FEN-1 have removed primers and near-primer mistakes, pol δ fills in the missing nucleotides, and a ligase enzyme joins the fragments. Pol ε does have a 3'-5' exonuclease activity that chops single-stranded DNA into small oligonucleotide fragments and is also associated with the replication machine. The function of Pol ε is not clearly understood.

DNA Lesions

The robust nature of DNA due to its complementary double strands has been noted several times already. We now consider in more detail the repair processes that rescue damaged DNA. DNA is not nearly as robust as popular media makes it out to be. In fact, to take the blockbuster book and film, Jurassic Park, as an example, Although there is unquestionably some DNA to be found either embedded in amber-bound parasites, or pehaps in preserved soft tissue (found deep in a fossilized femur, Schweitzer et al, 2007.). It is likely to be heavily degraded, and accurate reproduction is impossible without many samples to work from.

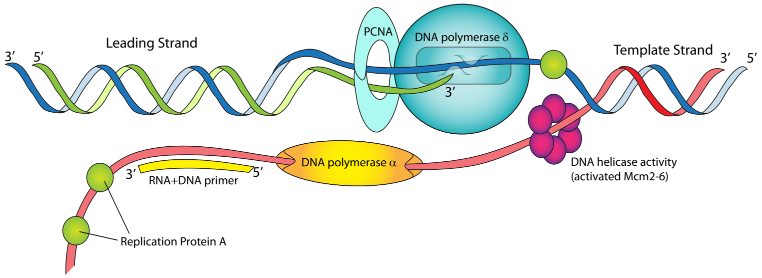


Figure 14. The eukaryotic replication machine is very similar to the prokaryotic machine.

The most common insult to the DNA of living organisms is depurination, in which the β-N-glycosidic bond between an adenine or guanine and the deoxyribose is hydrolyzed. In mammalian cells, it is estimated at nearly 10000 purines per cell generation, and generally, the average rate of loss at physiological pH and ionic strength, and at 37 °C, is approximately 3 x 10^{-11}/sec. Depyrimidination of cytosine and thymine residues can also occur, but do so at a much slower rate than depurination. Despite the high rate of loss of these bases, they are generally remediated easily by base excision repair (BER), which is discussed later in this section. Therefore it is rare for depurination or depyrimidination to lead to mutation.

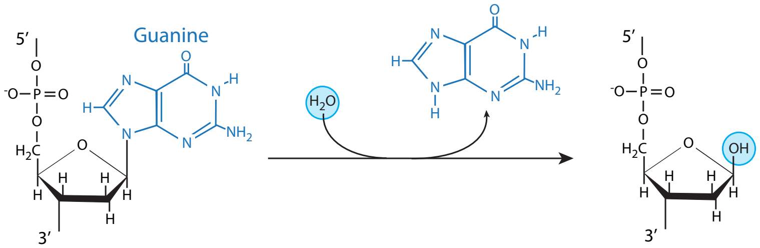


Figure 15. Depurination of guanines (or adenines) is a common DNA lesion.

Three of the four DNA bases, adenine, guanine, and cytosine, contain amine groups that can be lost in a variety of pH and temperature-dependent reactions that convert

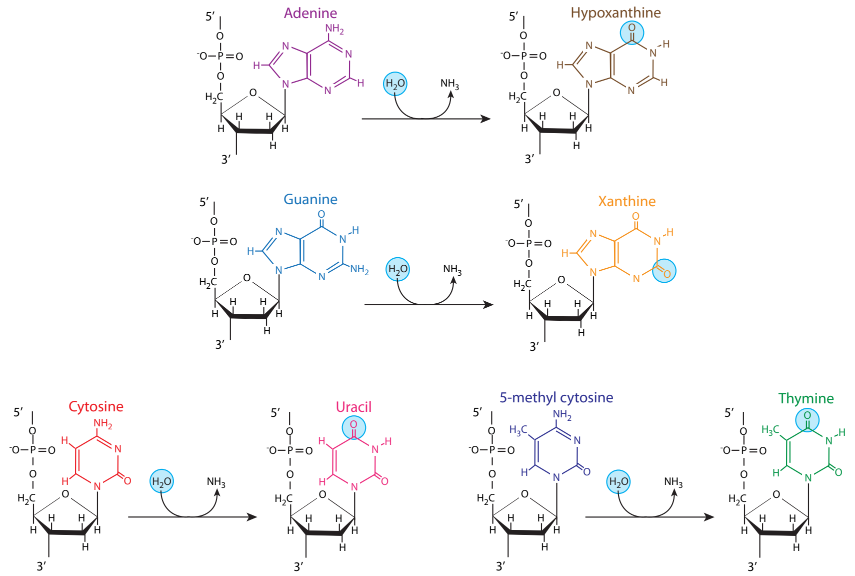


Figure 16. Deamination of adenine, guanine, cytosine, and methylcytosine can lead to mutations upon replication if unrepaired.

Thymine good, Uracil bad. Why is thymine found in DNA rather than uracil? It turns out that the frequency of cytosine deamination may yield a clue as to why cells have gone the extra step (literally, since uracil is a precursor in thymine biosynthesis) to make a new "standard" nucleotide for DNA when uracil worked just fine for RNA, presumably the older genetic molecule. Consider this: if uracil was standard for DNA, then the very frequent deamination conversions of C to U would not be caught by error-checking for non-DNA bases, and the mutation rate would skyrocket. Fortunately, since T has evolved to be the standard base-pairing partner of adenine in DNA, uracil is quickly recognized and removed by multiple uracil DNA glycosylases (more on that later in this chapter), and the integrity of our DNA sequences is much safer.

the bases to hypoxanthine, xanthine, and uracil, respectively. This can sometimes lead to permanent mutations since during replication, they serve as a template for the synthesis of a complementary strand, and where a guanine should go, for example (complementary to cytosine), an adenine may be inserted (because it complements uracil, the deamination product of cytosine). Another deamination, of the modified base methylcytosine, can also lead to a mutation upon replication. Some cytosines may be methylated as part of a regulatory process to inactivate certain genes in eukaryotes, or in prokaryotes as protection against restriction endonucleases. When the methylated cytosine is deaminated, it produces a thymine, which changes the complementary nucleotide (upon replication) from a guanine to an adenine. Deamination of cytosines occurs at nearly the same rate as depurination, but deamination of other bases are not as pervasive: deamination of adenines, for example, is 50 times less likely than deamination of cytosine.

All DNA bases can spontaneously shift to a tautomeric isomer (amino to imino, keto to enol, etc), although equilibrium leans heavily toward one than the other. When a rare tautomer occurs, it base-pairs differently than its more common structural form: guanines with thymines and adenines with cytosines. Here again, a mutation can be propagated during replication of the DNA.

DNA inside a cell must also contend with reactive oxidative species (ROS) generated by the cell's metabolic processes. These include singlet oxygen, peroxide and peroxide radicals, as well as hydroxyl radicals. although it is thought that the hydrogen peroxide and peroxide radicals do not directly attack the DNA but rather generate hydroxyl radicals that do. Most of these ROS are generated in the mitochondria during oxidative phosphorylation and leak out, although some may be generated in peroxisomes, or in some cytosolic reactions. Depending on what part of the DNA is targeted, ROS can cause a range of lesions including strand breaks and removal of bases.

Ionizing radiation (e.g. X-rays) and ultraviolet radiation can each cause DNA lesions. Ionizing radiation is often a cause for double-stranded breaks of the DNA. As described later in the chapter, the repair process for double-stranded breaks necessarily leads to some loss of information, and could potentially knock out a gene. Ultraviolet radiation that hits adjacent thymines can cause them to react and form a cyclobutyl (four carbons bonded in closed loop) thymine dimer. The dimer pulls each thymine towards the other, out of the normal alignment. Depending on the structural form of the dimer, this is sufficient to stymie the replication machine and halt replication. However, some data suggests that normal basepairing to adenine may be possible under some conditions, although, it is likely only one base-pair would result, and the missing base could lead to either random substitution or a deletion in the newly synthesized strand.

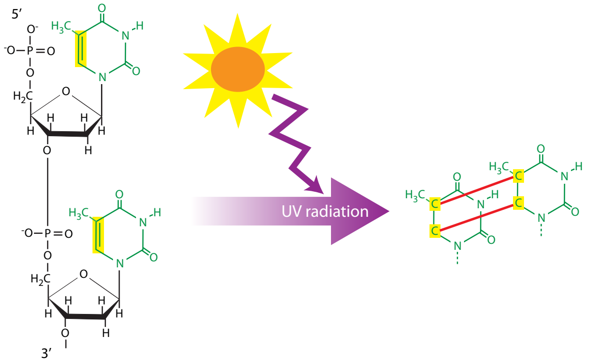


Figure 17. Ultraviolet radiation can be absorbed by some DNA and commonly causes pyrimidine cyclobutyl dimers connecting adjacent nucleotide bases.

Finally, we consider the formation of chemical adducts (covalently attached groups) on DNA. They may come from a variety of sources, including lipid oxidation, cigarette smoke, and fungal toxins. These adducts attach to the DNA in different ways, so there are a variety of different effects from the adducts as well. Some may be very small adducts - many environmental carcinogens are alkylating agents, transferring methyl groups or other small alkyl groups to the DNA. Other adducts are larger, but also attach covalently to a nitrogenous base of DNA. Common examples are benzo(a)pyrene, a major mutagenic component of cigarette smoke, and aflatoxin B1, produced by a variety of Aspergillus-family fungi. Benzo(a)pyrene is converted to benzo(a)pyrene diol epoxide, which can then attack the DNA. When this happens, the flat pyrene ring intercalates between bases, causing steric changes that lead to local deformation of the DNA and disruption of normal DNA replication.

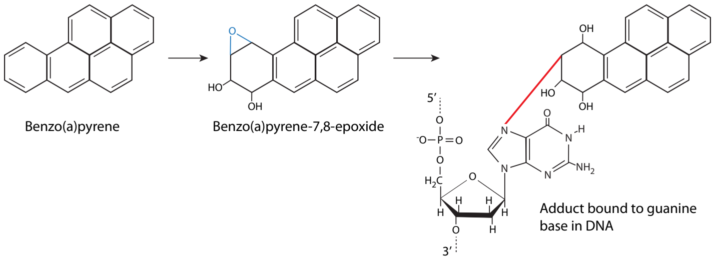


Figure 18. Benzo(a)pyrene is converted to an epoxide form by the cell. The epoxide and form an adduct on DNA.

Aflatoxin B1 is the primary aflatoxin produced by some species (esp. *flavus, parasiticus*) of Aspergillus, a very common mold that grows on stored grain (as well as detritus and other dead or dying plant matter). In addition to infecting grain, it is a common problem with stored peanuts. At high levels, aflatoxin is acutely toxic, but at lower levels, it has the insidious property of being unnoticeably toxic but mutagenic. Like benzo(a) pyrene, it is metabolized into an epoxide and will then react with DNA to form an adduct that can disrupt replication.

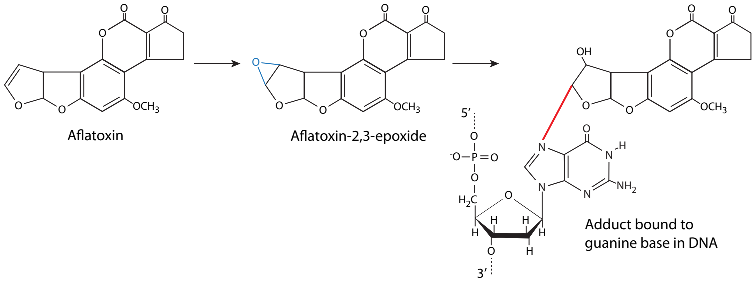


Figure 19. The epoxide form of aflatoxin also forms adducts on DNA.

Some alkylating agents, particularly N-nitroso compounds, are formed in the acidic conditions of the stomach from nitrosation of naturally occurring nitrites produced from food (reduction of nitrates), or environmental nitrites in drinking water. Ironically, while some alkylating agents can cause cancers, others are used therapeutically as anticancer treatments, e.g. mitomycin, melphalan. The idea, as with many cancer treatments, is that although such drugs cause DNA damage to non-cancerous cells as well as cancer cells, the high rate of cancer cell proliferation gives them fewer chances for repair of damaged DNA, and thus greater likelihood that the damage might halt replication and lead to cell death.

In a similar vein, crosslinking chemotherapeutic agents such as cisplatin (a platinum atom bonded to two chloride groups and two amino groups) also bind to DNA. The chloride groups are displaced first by water and then by other groups including sites on DNA. Although sometimes classified as an alkylating agent, it obviously is not, but it acts similarly. Cisplatin goes a step further than a simple alkylating agent though, because it has another reactive site and can thus crosslink (covalently bond) another nucleotide, possibly on another strand of DNA, making a strong obstruction to DNA replication. Cisplatin can also crosslink proteins to DNA.

Benzo(a)pyrene and aflatoxin B1 are not themselves mutagens. Once they are in the cell, the normal metabolism of these compounds leads to diol epoxide formation, which can then attack the DNA. Although the 7-nitrogen (N7) of guanine is more nucleophilic, and is a target for aflatoxin, most benzo(a)pyrene diol epoxide adducts attach to the 2-nitrogen of guanine residues.

There are federal standards (20-300 parts per billion depending on usage) for aflatoxin in various forms of grain-based animal feed, especially corn-based feeds, because the toxin can pass through the animal into milk, as well as linger in the meat. In addition to feed, there are federal maximums for peanuts and peanut products, brazil nuts, pistachios, and other foodstuffs (actionable at 20 ppb).

Well then, what's a poor cell to do when its DNA is being constantly ravaged? As it turns out, there are some very good repair processes that are constantly at work on the DNA, scanning it for defects, and where possible, making repairs. Often the repairs are perfect, if the complementary strand is intact, sometimes mutations must be introduced, and finally there are occasions when repair is impossible, and apoptosis is triggered to kill the cell and prevent propagation of damaged DNA.

DNA Repair

Strictly defined, the simplest repair mechanism does not use an enzyme. Dealkylation, or removal of alkyl groups (like $-CH_3$ or $-C_2H_5$) involves only the transfer of an alkyl group from an O6-methylguanine or O6-ethylguanine onto O6-alkylguanyl-DNA alkyltransferase. Despite the name, the alkyltransferase is not really an enzyme, since it is permanently altered and inactivated by the reaction and therefore does not fit the definition of a catalyst. Note that this does not remediate alkylation at N7 or other sites, just the O6-linked ones.

The next simplest repair mechanism is the uncoupling of pyrimidine cyclobutyl dimers. This can be accomplished through the activity of DNA *photolyases*, also known as photoreactivating enzymes. These are named not just because the formation of the pyrimidine cyclobutyl dimers is usually due to UV light exposure, but because the repair enzymes themselves require exposure to light (300-500nm, near UV to visible blue) to catalyze the dimer-breaking reaction.

While dealkylation and dimer lysis are relatively simple processes that make only a subtle change to the DNA, excision repair mechanisms are more complicated and require multiple enzymatic steps to complete. When a small (not sterically bulky) lesion is limited to a single base, whether missing from depurination or incorrectly formed due to deamination or misincorporation, the process known as *base excision repair* (BER) is engaged. As illustrated in figure 20, if a non-conventional base is recognized, it is then removed by an appropriate *DNA glycosylase*. At present (Genbank search, July 2009), there are at least 8 specific genes encoding human DNA glycosylases, although three encode glycosylases that recognize uracil in various situations. Once the base has been removed by the glycosylase, an endonuclease is enlisted to break the phosphodiester bonds than hold the now-empty phosphodeoxyribose. The resulting gap in the DNA is filled in by a DNA polymerase and finally the strand is reconnected by DNA ligase.

More specifically, DNA photolyases (a ~60kD protein), are non-covalently associated with a chromophore (N^5,N^{10}-methylenyltetrahydrofolate or 5-deazaflavin) and an $FADH^-$. The photolyase binds to the pyrimidine cyclobutyl dimer of either single-stranded or double-stranded DNA in a light-independent and sequence-independent manner. However it does not catalyze any change in the bond until light is absorbed by the chromophore, which then transfers the energy to $FADH^-$, causing it to eject and electron to the dimer, thus breaking it apart.

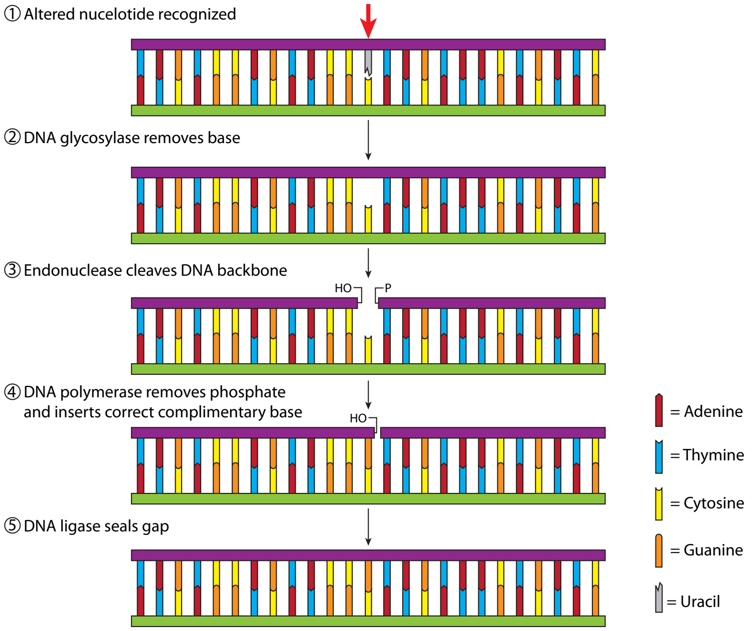

① Altered nucelotide recognized
② DNA glycosylase removes base
③ Endonuclease cleaves DNA backbone
HO
P
④ DNA polymerase removes phosphate and inserts correct complimentary base
HO
⑤ DNA ligase seals gap
= Adenine
= Thymine
= Cytosine
= Guanine
= Uracil

Figure 20. Base Excision Repair requires five separate steps.

In the case of bulky lesions that significantly alter the physical presentation of the DNA to the polymerases and other enzymes that process DNA, a different type of repair process is involved. Nucleotide excision repair (NER), perhaps better named *poly*nucleotide excision repair, involves the removal of the lesion as well as some of the nucleotides in the immediate vicinity. There are two major initiators of NER: either a non-transcriptionally active portion of the DNA is scanned by XPC (fig. 21A), which recognizes a bulky lesion and recruits the repair complex, or as a gene is being transcribed, RNA polymerase runs into a lesion, and then recruits the repair complex via CSA and CSB (fig. 21F and G). If the detection is through XPC, one of the early repair factors recruited to the site is Transcription Factor IIH/XPB/XPD, which is a DNA helicase (fig. 21B). This type of global genome detection is inefficient and relatively slow, but provides a basal level of error-checking for all DNA. In the case of DNA being transcribed, the RNA polymerase complex already includes TFIIH, of which XPB and XPD are a part. This transcriptionally-directed detection is more efficient and targets those parts of the DNA in greatest use in a given cell. In the next step (fig. 21C), XPG, associated with BRCA1/2, and XPF, associated with ERCC1, excise a portion of the affected strand, including but not limited to the lesion itself. DNA polymerase δ or ε can then add onto the free 3'OH to fill in the gap based on the complementary strand sequence (fig. 21D). Finally, the repair is connected on its 3' end to the rest of the strand by DNA ligase (fig. 21E).

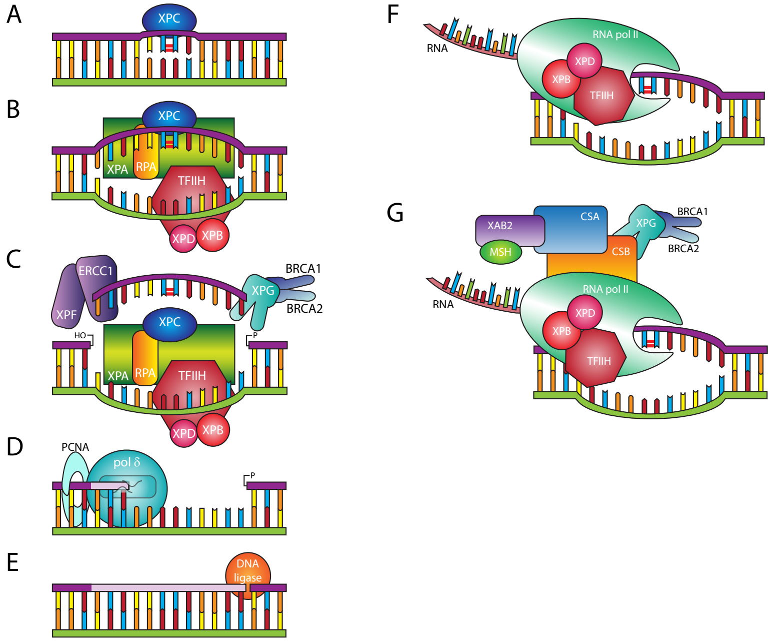

A
XPC
B
XPC
XPA
RPA
TFIIH
XPD
XPB
C
ERCC1
XPF
BRCA1
XPG
BRCA2
XPC
HO
P
XPA
RPA
TFIIH
XPD
XPB
D
PCNA
pol δ
P
E
DNA ligase
F
RNA
RNA pol II
XPD
XPB
TFIIH
G
XAB2
CSA
XPG
BRCA1
BRCA2
MSH
CSB
RNA
RNA pol II
XPD
XPB
TFIIH

Figure 21. Nucleotide Excision Repair in Eukaryotic Systems

The "XP" in XPC, XPB, XPD, and the others in fig. 18 refers to xeroderma pigmentosa, another autosomal recessive disease, of which the primary characteristic is the formation of skin carcinomas at a young age. Because NER is a major form of pyrimidine dimer repair (in addition to photolyases), its disruption by mutations to one or more of the XP genes leads to extreme sensitivity to UV-induced lesions. Affected individuals must minimize exposure to the sun. The name of the disease comes from the characteristic pigmented lesions (keratoses) that often form on the skin when exposed to sun.

CSA and CSB are named for Cockayne syndrome, an autosomal recessive aging disorder. Mutations in either gene can cause the disorder, which is characterized by premature aging, stunted growth, photosensitivity, and developmental defects of the nervous system. Presumably, knocking out the DNA repair capability of CSA or CSB leads to fast accumulation of damage, inability to transcribe needed genes, and eventually cell death.

A sort of variation on NER is the mismatch repair (MMR) system. This is best understood in prokaryotes: in E. coli, MutS is a small protein that forms homodimers at mismatch sites. The MutS dimers recruit two MutL proteins, each of which interacts with one of the MutS units. Each MutS/MutL complex pushes DNA through inwardly, forming a loop with the mismatch in the center of the loop. This continues until one of the MutS/MutL complexes encounters a *hemimethylated* GATC sequence. This causes recruitment of MutH, a highly specialized endonuclease that makes a single-stranded nick in the backbone of the non-methylated strand. This provides an opening for the 3'-5' exonuclease I or the 5'-3' exonuclease VII (or RecJ) to degrade the strand from the nick to the point of mismatch. This is then, as you may have guessed, filled in by DNA polymerase and the backbone connected by ligase. In eukaryotes, multiple homologues to the MutS and MutL proteins have been discovered and the process is similar, but not clearly understood yet, as no homologue to MutH has yet been discovered.

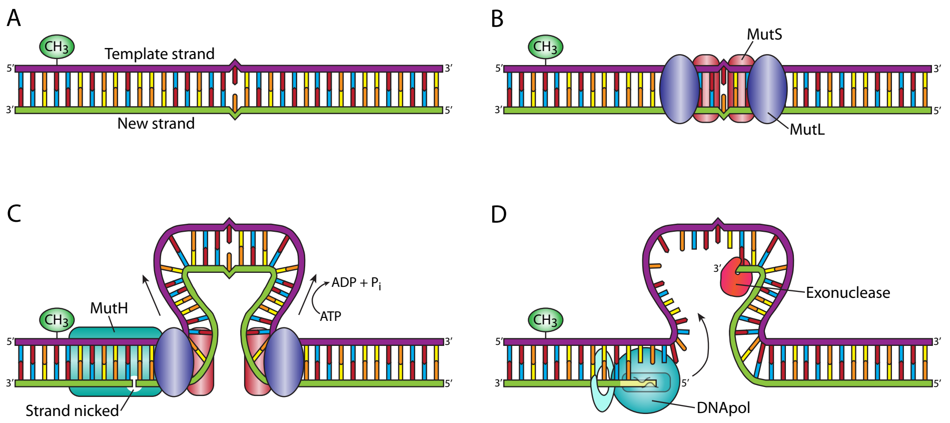


Figure 22. Mismatch Repair System. Mismatched nucleotides do not for normal H-bonds, and cause a minor disturbance (A) in the DNA structure. (B) MutS dimerizes at mismatch sites and each recruits a MutL protein. (C) The MutS/L complex pushes DNA through to form a loop with the mismatch in the center. When a methylated GATC is encountered, the MutS/L stop moving DNA and recruit th the MutH endonuclease. An exonuclease then digests the DNA strand to the mismatch. (D) Finally, a DNA polymerase fills in the missing segment and the backbone is sealed by ligase.

So far, the repairs have been based on the assumption that a lesion affects only one strand, and the other strand can provide a reliable template for effecting repairs without the loss of information. Unfortunately, that is not always the case, and some lesions and repair processes necessarily lead to sequence loss. When a double-strand break occurs, perhaps as the product of ionizing radiation, the most common repair mechanism is known as non-homologous end joining (NHEJ). The double-stranded ends are first recognized by Ku, a heterodimeric circular protein that binds the DNA ends. Ku then recruits the kinase DNA-PK$_{CS}$. The DNA-PK$_{CS}$ acts as a bridge to bring the

Recall that in E. coli, Dam methyltransferases eventually methylate the DNA as a method of protecting its genome, but newly synthesized DNA is not methylated. Thus, the assumption is that the methylated strand contains the original and correct base, while the mismatch is due to misincorporation in the newer strand.

Another prokaryotic DNA repair system is the SOS response. As depicted in fig. 23 below, if there is no damage, RecA is inactive, so LexA protein can repress the production of more SOS repair proteins. However, if there is damage, RecA proteins bind to the single stranded DNA and are activated. They in turn cleave the LexA repressor allowing production from a number of DNA repair genes.

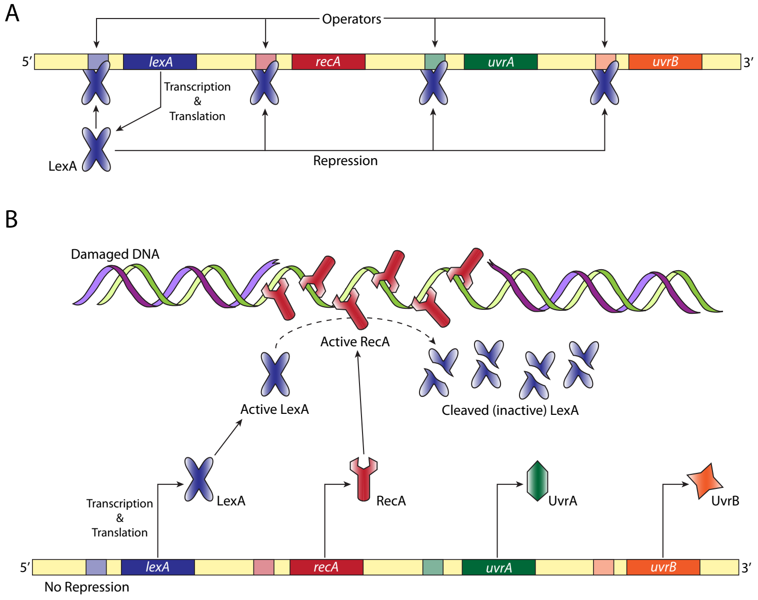


Figure 23. The prokaryotic SOS operon. (A) repressed state when there is no DNA damage, and (B) the activated state in which DNA damage has been detected.

two ends together, and a DNA ligase can then join the ends together. If the strands were broken in different places, resulting in complementary single-stranded overhangs at each end (like those generated by some restriction endonucleases) then the repair is often perfect, since the complementary sequences align the two ends correctly in their original positions. However, if the strand ends have already been acted upon by nucleases and are no longer complementary, then the rejoining of the ends will likely lead to loss of information. In some parts of the DNA, this would have little effect, but if it happened within a gene, the mutated gene product could have abnormal or compromised function.

Telomeres

If there is a mechanism for recognizing loose ends of DNA, what about the ends of every eukaryotic chromosome? They are linear chromosomes, so they have ends, right? What prevents the double-strand-break repair systems from mis-recognizing them all as broken DNA and concatenating all of the chromosomes together? Interestingly, the answer to this question is intimately tied up with the answer to the problem of end-replication, which was very briefly alluded to in our description of replication.

The end-replication problem is one that affects all linear chromosomes. It boils down to one simple fact: an RNA primer is needed to start any DNA replication. So on the 5' end of each strand is an RNA primer (fig. 24 in yellow) that gets removed by the error-correction process. Thus with every round of replication, information is lost from the 5' end of each strand of each chromosome. Eventually, crucial genes are lost and the cell will die; most likely many cellular functions will be compromised long before that happens. The solution to the end-replication problem might be considered more a treatment of symptoms than a cure, to use an analogy to medicine. In short, during the very early stages of an organism's life, a lot of non-coding DNA is added onto the ends of the DNA so that as the cell and its progeny continue to reproduce, the nucleotides do not affect any functional genes. This process is catalyzed by the enzyme, telomerase.

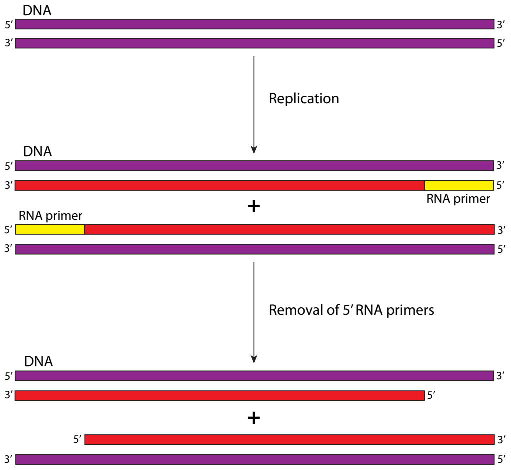


Figure 24. End-replication problem.

Telomerase is a large holoenzyme that acts as a reverse transcriptase, reading a self-contained RNA template to add on telomere sequence to the 3' ends of linear chromosomes. Note that this does not add onto the 5' ends - as mentioned above, there is nothing to be done about the 5' ends directly. However, while telomerases are active, the 3' ends can be extended, and thus when the 5' end primer is removed, the sequence lost (forever) from that strand of DNA is only a telomeric repeat and not something more useful. The repeats are well conserved across eukaryotes, and almost completely conserved across mammalian species (shown in Fig. 25).

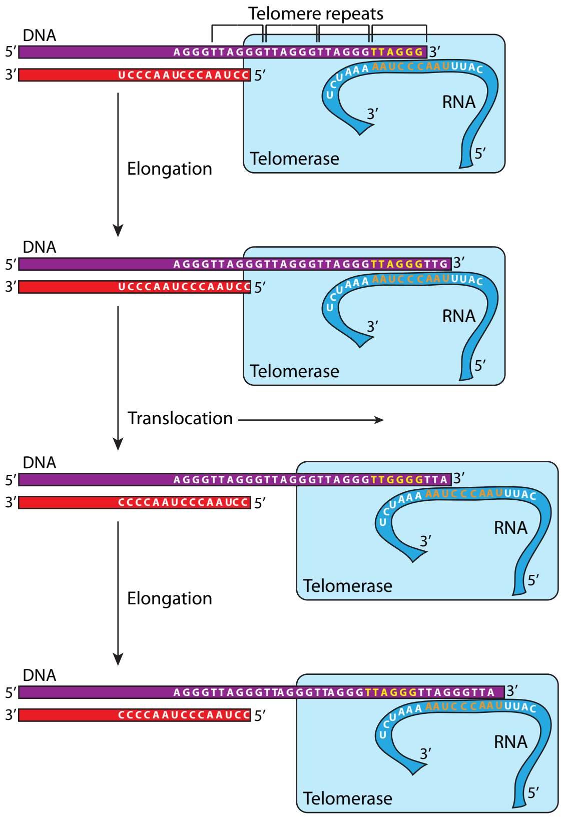

Telomere repeats
DNA
5'
AGGGTTAGGGTTAGGGTTAGGGTTAGGG
3'
3'
UCCCAAUCCCAAUCC
5'
UCUAAAAAUCCCAAUUUAC
3'
RNA
5'
Telomerase
Elongation
DNA
5'
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTG
3'
3'
UCCCAAUCCCAAUCC
5'
UCUAAAAAUCCCAAUUUAC
3'
RNA
5'
Telomerase
Translocation
DNA
5'
AGGGTTAGGGTTAGGGTTAGGGTTGGGGTTA
3'
3'
CCCCAAUCCCAAUCC
5'
UCUAAAAAUCCCAAUUUAC
3'
RNA
5'
Telomerase
Elongation
DNA
5'
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
3'
3'
CCCCAAUCCCAAUCC
5'
UCUAAAAAUCCCAAUUUAC
3'
RNA
5'
Telomerase

Figure 25. Telomerases are reverse transcriptases that generate DNA repeats from a built-in RNA template.

In metazoans, telomerase activity is high in the embryonic stages of life, but is virtually non-existent in adults except in cell types that must constantly proliferate (e.g. blood and epithelial cells). The activity of telomerase is primarily regulated by expression of the TERT gene (telomerase reverse transcriptase) although construction of the full telomerase also requires the expression of the TERC gene (the telomerase RNA, also abbreviated TR), and dyskerin. Roughly speaking, the number of telomeric repeats that are placed on a chromosome in early development determines the number of DNA replications and cell divisions that the cell can undergo before succumbing to apoptosis (programmed cell death). Experiments on cells in culture demonstrate a strong correlation between telomere length and longevity, and it is known that cells taken from people with the premature aging disease, progeria, have relatively short telomeres.

Conversely, cancer cells almost universally have upregulated expression of telomerase. Given that a defining characteristic of cancer cells is the ability to proliferate rapidly and indefinitely, turning telomerase back on is, not surprisingly, an important aspect of carcinogenesis. It is therefore a target for anti-cancer treatments; however, to date no telomerase-targeting therapies have proven effective.

Now that we know about telomeres, the question that started this section becomes even more problematic: with these repeated sequence overhangs, how are chromosomes prevented from connecting end-to-end through a double-strand repair-like process? In part due to their repeated sequences, telomeres are able to form end-caps and protect chromosomal ends. The telomeres protect the ends of each chromosome by binding to protective proteins and by forming complex structures. Telomere end binding proteins (TEBP) bind to the 3' overhanging end of the telomere. Other capping proteins, such as the mammalian TRF1 and TRF2 (telomere repeat binding factors) not only bind the telomere, but help to organize it into large looped structures known as T-loops (Fig. 26).

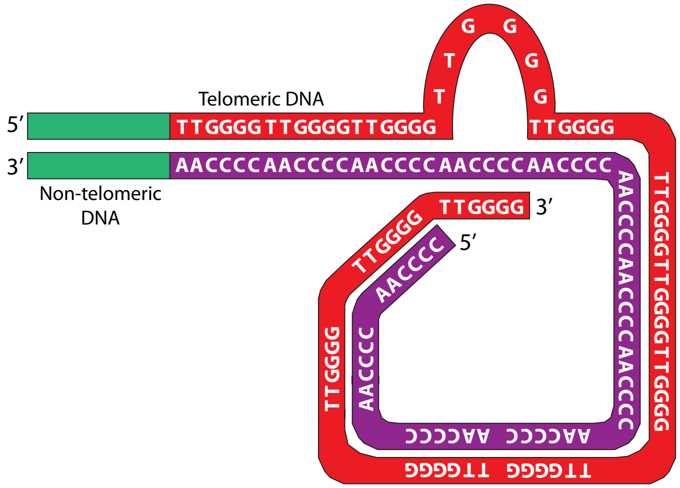


Figure 21. Telomeres form protective T-loop structures on the ends of linear chromosomes.

Finally, the T-loop ends are further stabilized by the formation of G-quartets (fig. 27). G-quartets are a cyclic tetramers that can form in sequences with four consecutive guanine residues, which hydrogen-bond to each other to make a linked square shape stabilized by a metal ion in the center. Furthermore in cases like the telomere, in which such sequences are repeated, the G-quartets can stack and associate three-dimensionally, increasing their stability.

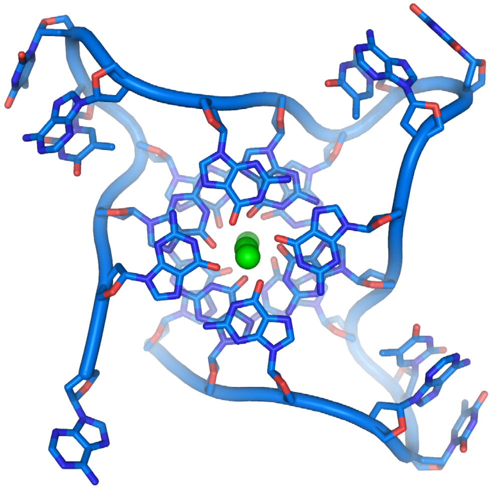


Figure 27. G-quartets form by hydrogen bonding of fours consecutive guanine residues stabilized by a metal ion (green), and can form three-dimensional stacks due to telomeric repeats. Produced from Protein DataBank data by T. Splettstoesser.