Chapter 5

Flow of Genetic Information

As the cell's so-called blueprint, DNA must be copied to pass on to new cells and its integrity safeguarded. The information in the DNA must also be accessed and transcribed to make the RNA instructions that direct the synthesis of proteins.

Flow of Genetic Information

DNA Replication

The only way to make new cells is by the division of pre-existing cells. This means that all organisms depend on cell division for their continued existence. DNA, as you know, carries the genetic information that each cell needs. Each time a cell divides, all of its DNA must be copied faithfully so that a copy of this information can be passed on to the daughter cell. This process is called DNA replication.

Before examining the actual process of DNA replication, it is useful to think about what it takes to accomplish this task successfully.

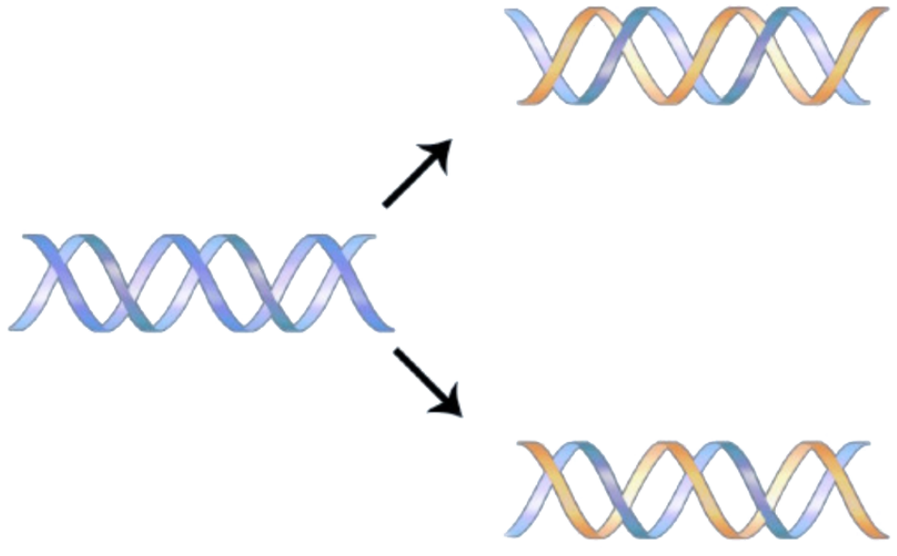
Consider the challenges facing a cell in this process:

•The sheer number of nucleotides to be copied is enormous: e.g., in human cells, on the order of several billion.
• A double-helical parental DNA molecule must be unwound to expose single strands of DNA that can serve as templates for the synthesis of new DNA strands.
• This unwinding must be accomplished without introducing significant topological distortion into the molecule.
• The unwound single strands of DNA must be kept from coming back together long enough for the new strands to be synthesized.

See Kevin's YouTube lectures on DNA Replication and Repair HERE, HERE, HERE, and HERE

• DNA polymerases cannot begin synthesis of a new DNA strand *de novo* and require a free 3' OH to which they can add DNA nucleotides.
• DNA polymerases can only extend a strand in the 5' to 3' direction. The 5' to 3' extension of both new strands at a single replication fork means that one of the strands is made in pieces.
• The use of RNA primers requires that the RNA nucleotides must be removed and replaced with DNA nucleotides and the resulting DNA fragments must be joined.
• Ensuring accuracy in the copying of so much information.

With this in mind, we can begin to examine how cells deal with each of these challenges. Our understanding of the process of DNA replication is derived from studies using bacteria, yeast, and other systems, such as *Xenopus* eggs. These investigations have revealed that DNA replication is carried out by the action of a large number of proteins that act together as a complex protein machine called the replisome. Numerous proteins involved in replication have been identified and characterized, including multiple different DNA polymerases in both prokaryotes and eukaryotes. Although the specific proteins involved are different in bacteria and eukaryotes, it is useful to understand the basic considerations that are relevant in all cells, before attempting to address the details of each system. A generalized account of the steps in DNA replication is presented below, focused on the challenges mentioned above.



Semi-Conservative DNA Replication

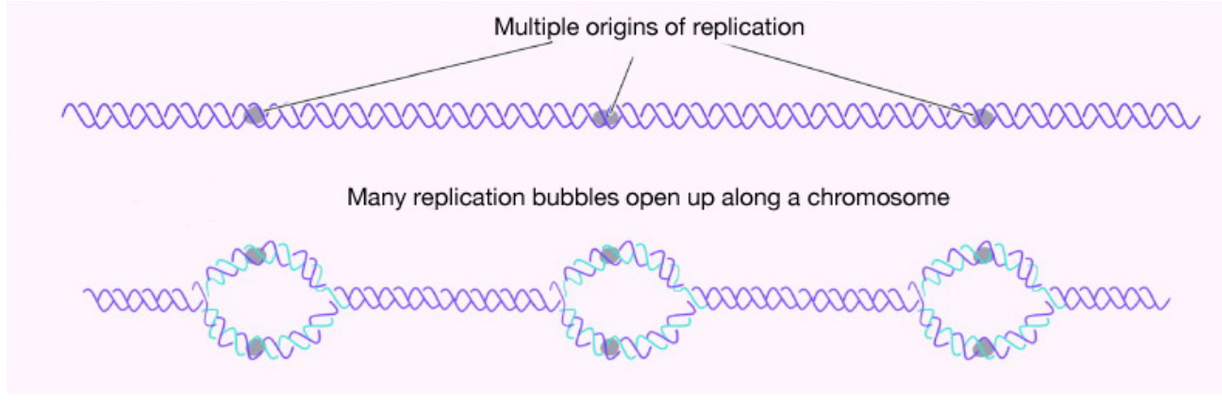
•The sheer number of nucleotides to be copied is enormous.

For example, in human cells, the number of nucleotides to be copied is on the order of several billion. Even in bacteria, the number is in the millions. Cells, whether bacterial or eukaryotic, have to replicate all of their DNA before they can divide. In cells like our own, the vast amount of DNA is broken up into many chromosomes, each of which is composed of a linear strand of DNA. In cells like those of *E. coli*, there is a single circular chromosome.

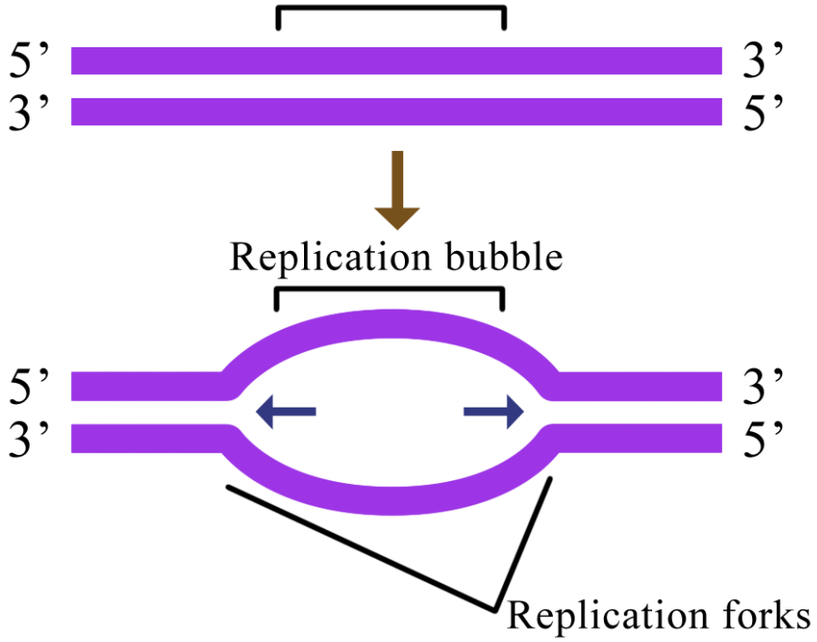
In either situation, DNA replication is initiated at sites called origins of replication. These are regions of the DNA molecule that are recognized by special origin recognition proteins that bind the DNA. The binding of these proteins helps open up a region of single-stranded DNA where the synthesis of new DNA can begin.

In the case of *E. coli*, there is a single origin of replication on its circular chromosome. In eukaryotic cells, there may be many

thousands of origins of replication, with each chromosome having hundreds. DNA replication is thus initiated at multiple points along each chromosome in eukaryotes as shown in the figure at the right. Electron micrographs of replicating DNA from eukaryotic cells show many replication “bubbles” on a single chromosome.



This makes sense in light of the large amount of DNA that there is to be copied in cells like our own, where beginning at one end of each chromosome and replicating all the way through to the other end from a single origin would simply take too long. This is despite the fact that the DNA polymerases in human cells are capable of building new DNA strands at the very respectable rate of about 50 nucleotides per second!



- A double-helical parental molecule must be unwound to expose single strands of DNA that can serve as templates for the synthesis of new DNA strands.

Once a small region of the DNA is opened up at each origin of replication, the DNA helix must be unwound to allow replication to proceed. How are the strands of the parental DNA double helix separated?

The unwinding of the DNA helix requires the action of an enzyme called helicase. Helicase uses the energy released

See Kevin's YouTube lectures on DNA Replication and Repair HERE, HERE, HERE, and HERE

when ATP is hydrolyzed to unwind the DNA helix. Note that each replication bubble is made up of two replication forks that "move" or open up, in opposite directions. At each replication fork, the parental DNA strands must be unwound to expose new sections of single-stranded template.

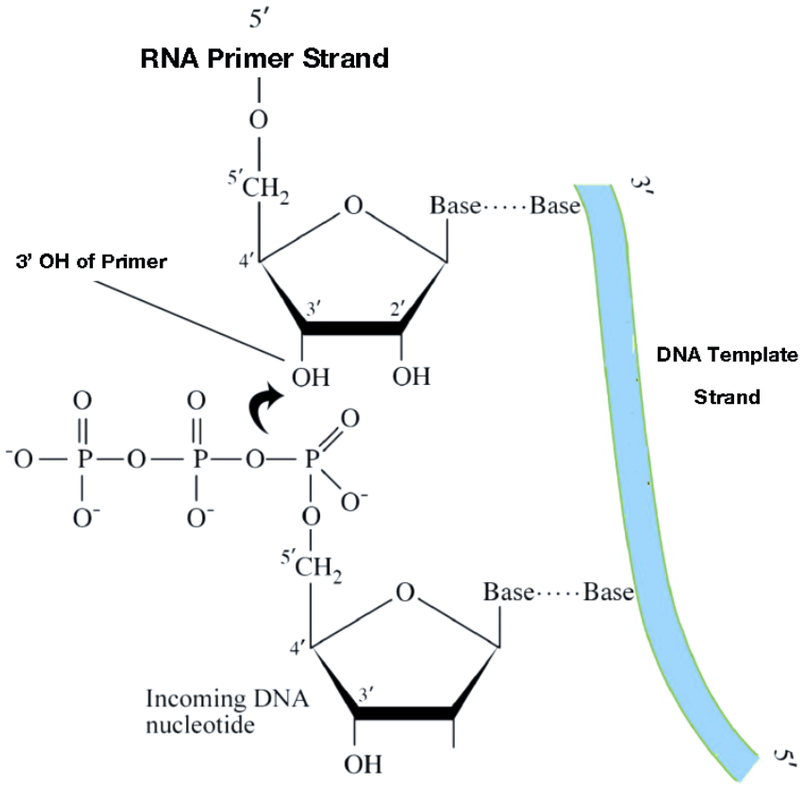
- This unwinding must be accomplished without introducing topological distortion into the molecule.

What is the effect of unwinding one region of the double helix? Unwinding the helix locally causes over-winding or topological distortion of the DNA ahead of the unwound region. The DNA ahead of the unwound helix has to rotate, or it will get twisted on itself.

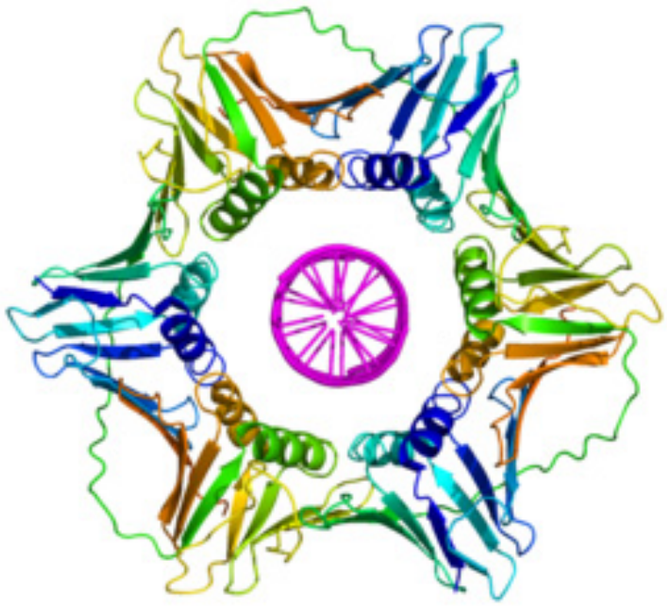
How is this problem solved? Enzymes called topoisomerases can relieve the topological stress caused by local unwinding of the double helix. They do this by cutting the DNA and allowing the strands to swivel around each other to release the tension before rejoining the ends. In *E. coli*, the topoisomerase that performs this function is called gyrase.

- The unwound single strands of DNA must be kept from coming back together long enough for the new strands to be synthesized.

Once the two strands of the parental DNA molecule are separated, they must be prevented from going back together to form double-stranded DNA. To ensure that unwound regions of the parental DNA remain single-stranded and available for copying, the separated strands of the parental DNA are bound by many molecules of a protein called single-strand DNA binding protein (SSB).



Addition of a Nucleotide to a Growing Chain



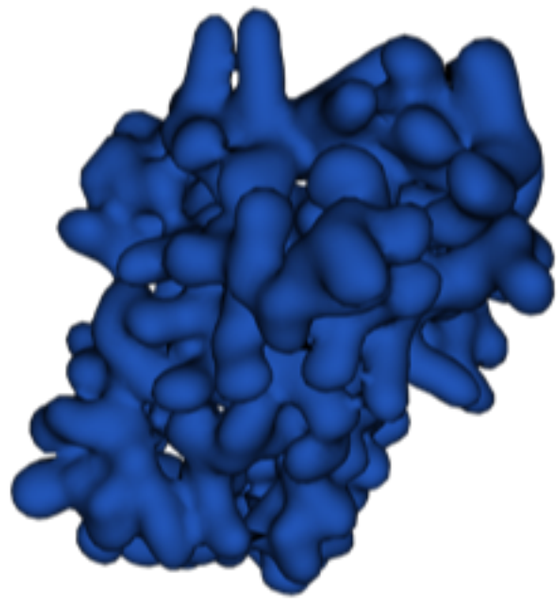
Sliding clamp assembled around DNA (in center, in purple)

- DNA polymerases cannot begin synthesis of a new DNA strand *de novo* and require a free 3' OH to which they can add DNA nucleotides.

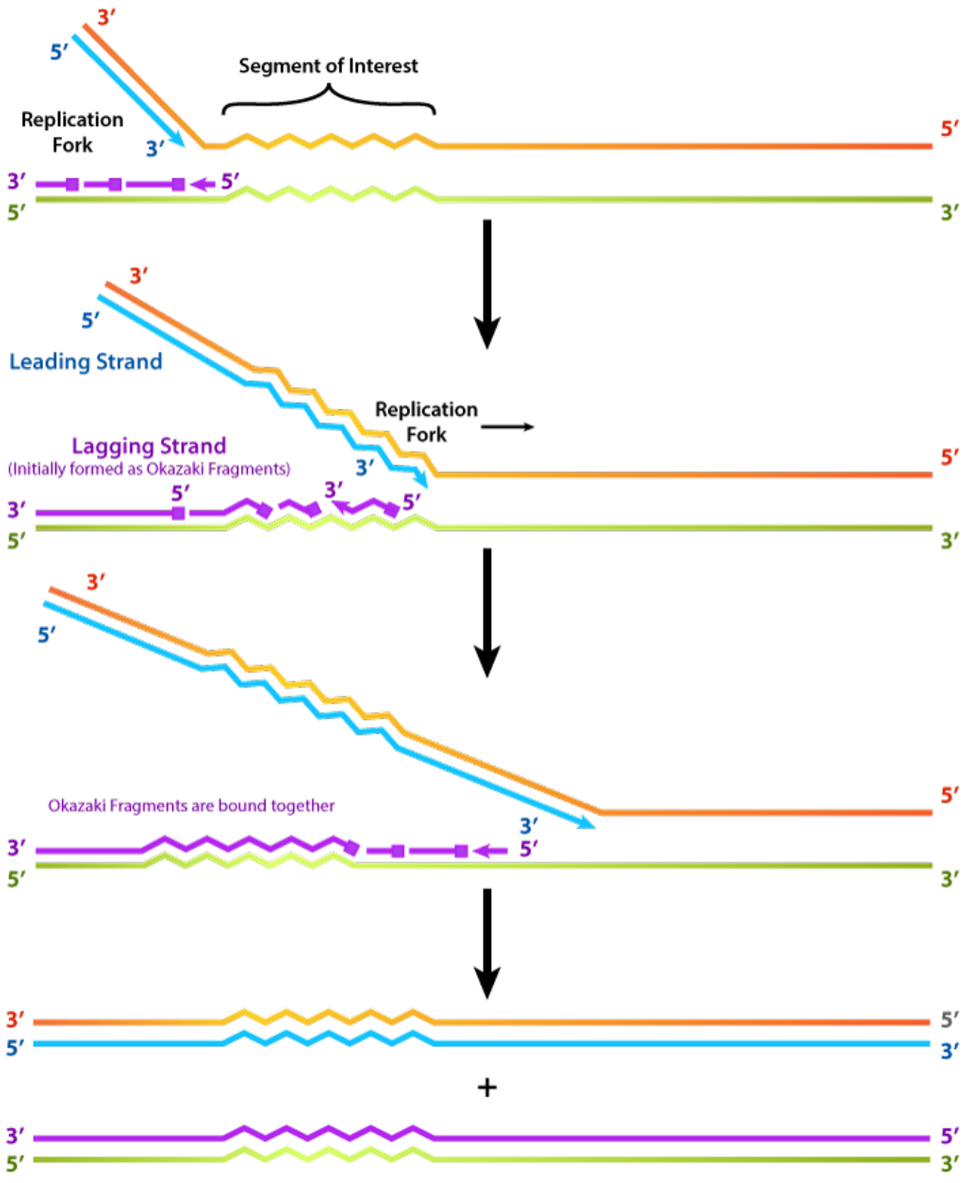
Although single-stranded parental DNA is now available for copying, DNA polymerases cannot begin synthesis of a complementary strand *de novo*. This is because all DNA polymerases can only add new nucleotides to the 3' end of a pre-existing chain. This means that some enzyme other than a DNA polymerase must first make a small region of nucleic acid, complementary to the parental strand, that can provide a free 3' OH to which DNA polymerase can add a deoxyribonucleotide.

This task is accomplished by an enzyme called a primase, which assembles a short stretch of RNA, called the primer, across from the parental DNA template. This provides a short base-paired region with a free 3'OH group to which DNA polymerase can add the first new DNA nucleotide (see figure on previous page)

Interactive 5.1



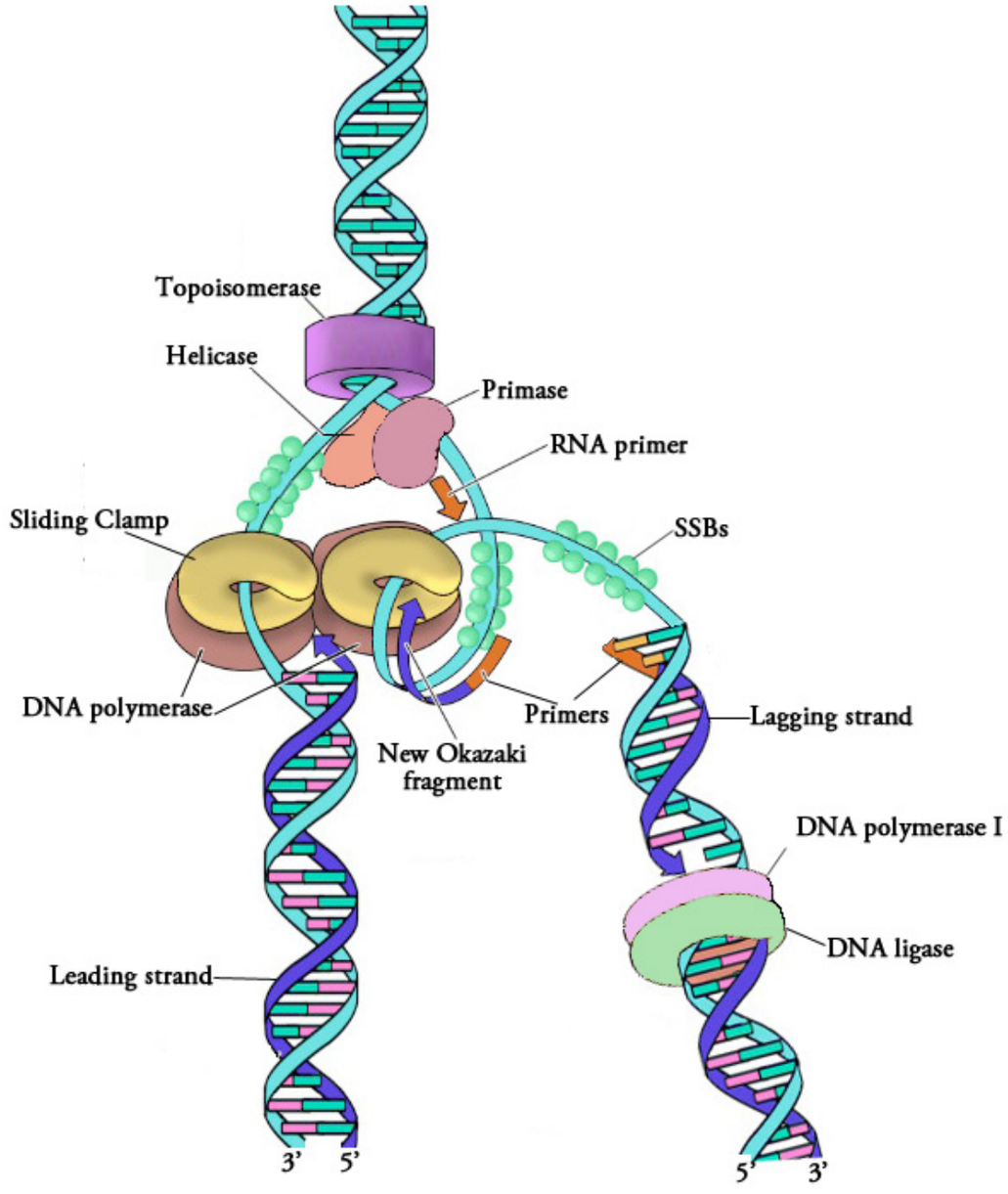
Sliding Clamp Subunit



Leading and Lagging Strand Replication

Once a primer provides a free 3'OH for extension, other proteins get into the act. These proteins are involved in loading the DNA polymerase onto the primed template and help to keep it attached to the DNA once it's on.

The first of these is the clamp loader. As its name suggests, the clamp loader helps to load a protein complex called the sliding

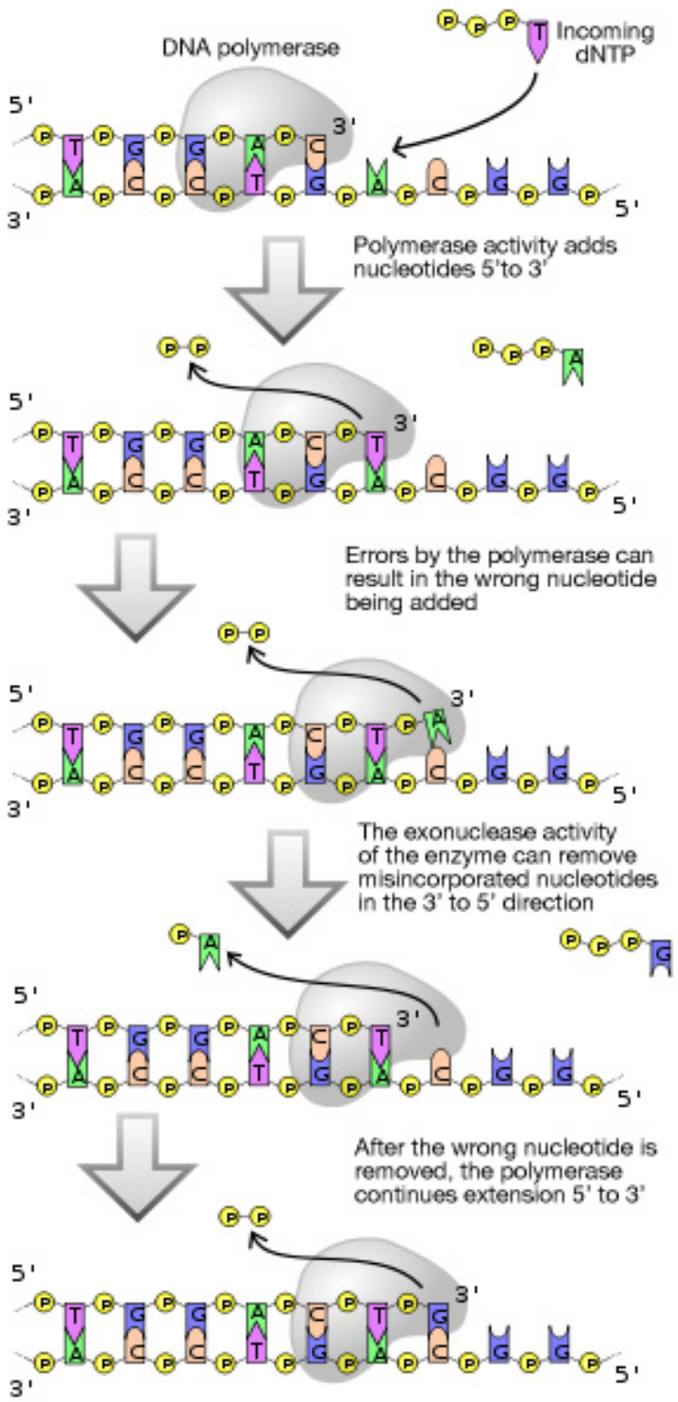


Proteins at a DNA Replication Fork

clamp onto the DNA at the replication fork. The sliding clamp is then joined by the DNA Polymerase. The function of the sliding clamp is to increase the processivity of the DNA polymerase. This is a fancy way of saying that it keeps the polymerase associated with the replication fork by preventing it from falling off- in fact, the sliding clamp has been described as a seat-belt for the DNA polymerase.

See Kevin's YouTube lectures on DNA Replication and Repair HERE, HERE, HERE, and HERE

The DNA polymerase is now poised to start synthesis of the new DNA strand (in *E. coli*, the primary replicative polymerase is called DNA polymerase III). As you already know, the synthesis of new DNA is accomplished by the addition of new nucleotides complementary to those on the parental strand. DNA polymerase catalyzes the reaction by which an incoming deoxyribonucleotide is added onto the 3' end of the previous nucleotide, starting



From Wikimedia Commons

with the 3'OH on the end of the RNA primer.

The 5' phosphate on each incoming nucleotide is joined by the DNA polymerase to the 3' OH on the end of the growing nucleic acid chain. As we already noted, the new DNA strands are synthesized by the addition of DNA nucleotides to the end of an RNA primer. The new DNA molecule thus has a short piece of RNA at the beginning.

•DNA polymerases can only extend a strand in the 5' to 3' direction. The 5' to 3' growth of both new strands means that one of the strands is made in pieces.

We have noted that DNA polymerase can only build a new DNA strand in the 5' to 3' direction. We also know that the two parental strands of DNA are antiparallel. This means that at each replication fork, one new strand, called the leading strand can be synthesized continuously in the 5' to 3' direction because it is being made in the same direction that the replication fork is opening up.

The synthesis of the other new strand, called the lagging strand, requires that multiple RNA

Central Dogma Zen (part 1)

To the tune of "*Those Were the Days*"

Once upon a time a cell decided
The time was ripe for it to split in two
Had to copy cellular instructions
For the daughter cell would need them too.

Bring in a helicase
Unzip the DNAs
To ease the stress a gyrase joins the fray
Strands must be held apart
SSBs do their part
And primase builds a primer RNA.

Sliding clamp comes in behind clamp loader
dNTPs floating all around
In the wings a replicase is waiting
For the chance to start another round.

Polymerase, my friend
Starts at the 3' end
It puts a 'T' across from every 'A'
A 'G' across from 'C'
Perfect simplicity
The leading strand is made in just this way.

The lagging strand is made in little pieces
Okazaki fragments, you recall
Pol I fills the gaps that lie between them
Ligase comes in next and joins them all.

Blueprints can't have mistakes
That's why polymerase
Corrects its work with exonuclease
Proofreading one by one
Till all its work is done
Hurray for D-N-A polymerase!

Recording by Tim Karplus
Lyrics by Indira Rajagopal

primers must be laid down and the new DNA be made in many short pieces that are later joined.These short nucleic acid pieces, each composed of a small stretch of RNA primer and about 1000-2000 DNA nucleotides, are called Okazaki fragments, for Reiji Okazaki, the scientist who first demonstrated their existence.

• The use of RNA primers requires that the RNA nucleotides must be removed and replaced with DNA nucleotides.

We have seen that each newly synthesized piece of DNA starts out with an RNA primer, effectively making a new nucleic acid strand that is part RNA and part DNA. The finished DNA strand cannot be allowed to have pieces of RNA attached.

So the RNA nucleotides are removed and the gaps are filled in with DNA nucleotides (by DNA polymerase I in *E. coli*). The DNA pieces are then joined together by the enzyme DNA ligase.

The steps outlined above essentially complete the process of DNA replication. The figure on the previous page shows a replication fork, complete with the associated proteins that form the replisome.

• Ensuring accuracy in the copying of so much information

How accurate is the copying of information in the DNA by DNA polymerase? As you are aware, changes in DNA sequence (mutations) can change the amino acid sequence of the encoded proteins and that this is often, though not always, deleterious to the functioning of the organism. When billions of bases in DNA are copied during replication, how do cells ensure that the newly synthesized DNA is a faithful copy of the original information?

DNA polymerases, as we have noted earlier, work fast (averaging 50 bases a second in human cells and up to 20 times faster in *E. coli*). Yet, both human and bacterial cells seem to replicate their DNA quite accurately. This is because of the proof-reading function of DNA polymerases. The proof-reading function of a DNA polymerase enables the polymerase to detect when the wrong base has been inserted across from a template strand, back up and remove the mistakenly inserted base. This is possible because the polymerase is a dual-function enzyme. It can extend a DNA chain by virtue of its 5' to 3' polymerase activity but it can also backtrack and remove the last inserted base because it has a 3' to 5' exonuclease activity (an exonuclease is an enzyme that removes bases, one by one, from the ends of nucleic acids). The exonuclease activity of the DNA polymerase allows it to excise a wrongly inserted base, after which the polymerase activity inserts the correct base and proceeds with extending the strand.

In other words, DNA polymerase is monitoring its own accuracy (also termed its fidelity) as it makes new DNA, correcting mistakes immediately before

See Kevin's YouTube lectures on DNA Replication and Repair HERE, HERE, HERE, and HERE

moving on to add the next base. This mechanism, which operates during DNA replication, corrects many errors as they occur, reducing by about 100-fold the mistakes made when DNA is copied.

Maintaining the Integrity of the Cell's Information: DNA Repair

In the last section we considered the ways in which cells deal with the challenges associated with replicating their DNA, a vital process for all cells. It is evident that if DNA is the master copy of instructions for an organism, then it is important not to make mistakes when copying the DNA to pass on to new cells. Although proofreading by DNA polymerases greatly increases the accuracy of replication, there are additional mechanisms in cells to further ensure that newly replicated DNA is a faithful copy of the original, and also to repair damage to DNA during the normal life of a cell.

All DNA suffers damage over time, from exposure to ultraviolet and other radiation, as well as from various chemicals in the environment. Even chemical reactions naturally occurring within cells can give rise to compounds that can damage DNA. As you already know, even minor changes in DNA sequence, such as point mutations can sometimes have far-reaching consequences. Likewise, unrepaired damage caused by radiation, environmental chemicals or even normal cellular chemistry can interfere with the accurate transmission of information in DNA. Maintaining the integrity of the cell's "blueprint" is of vital importance and this is reflected in the numerous mechanisms that exist to repair mistakes and damage in DNA.

Post-Replicative Mismatch Repair

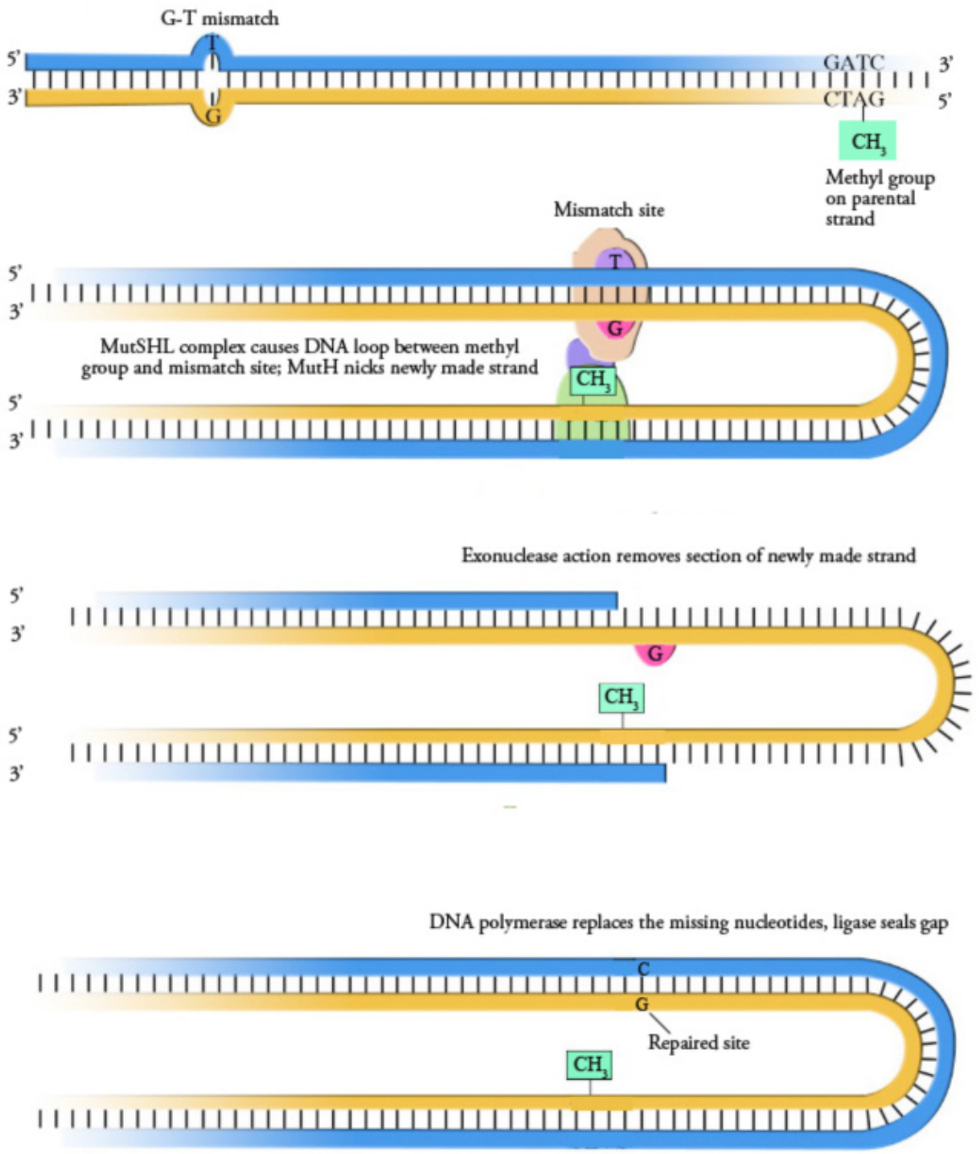
We earlier discussed proof-reading by DNA polymerases during replication. Does proofreading eliminate all errors made during replication? No. While proof-reading significantly reduces the error rate, not all mistakes are fixed on the fly by DNA polymerases. What mechanisms exist to correct the replication errors that are missed by the proof-reading function of DNA polymerases?

Errors that slip by proofreading during replication can be corrected by a mechanism called mismatch repair. While the error rate of DNA replication is about one in 10^7 nucleotides in the absence of mismatch repair, this is further reduced a hundred-fold to one in 10^9 nucleotides when mismatch repair is functional.

What are the tasks that a mismatch repair system faces?

It must:

- Scan newly made DNA to see if there are any mispaired bases (e.g., a G paired to a T)
- Identify and cut out the region of the mismatch.
- Correctly fill in the gap created by the excision of the mismatch region.



Mismatch Repair

Importantly, the mismatch repair system must have a means to distinguish the newly made DNA strand from the template strand, if replication errors are to be fixed correctly. In other words, when the mismatch repair system encounters an A-G mispair, for example, it must know whether the A should be removed and replaced with a C or if the G should be removed and replaced with a T.

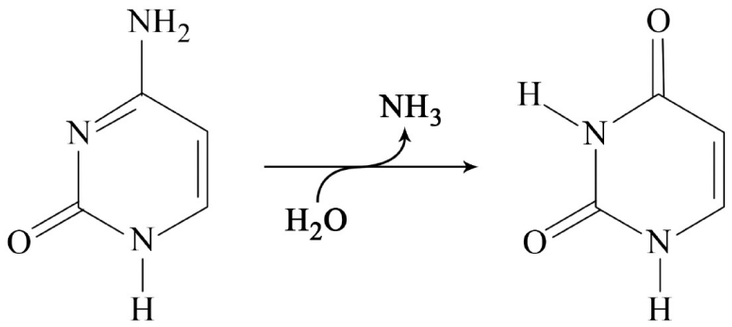
Mismatch repair has been well studied in bacteria, and the proteins involved have been identified. Eukaryotes have a mismatch repair system that repairs not only single base mismatches but also insertions and deletions.

In bacteria, mismatch repair proteins are encoded by a group of genes collectively known as the mut genes. Some of the most important components of the mismatch repair machinery are the proteins MutS, L and H.

MutS acts to recognize the mismatch, while MutL and MutH are recruited to the mismatch site by the binding of Mut S, to help cut out the region containing the mismatch. A DNA polymerase and ligase fill in the gap and join the ends, respectively.

But how does the mismatch repair system distinguish between the original and the new strands of DNA? In bacteria, the existence of a system that methylates the DNA at GATC sequences is the solution to this problem.

E.coli has an enzyme that adds methyl groups on the to adenines in GATC sequences. Newly replicated DNA lacks this methylation and thus, can be distinguished from the template strand, which is methylated. In the figure on the previous page,



Cytosine Uracil

Deamination of Cytosine

the template strand shown in yellow is methylated at GATC sequences.

The mismatch repair proteins selectively replace the strand lacking methylation, shown in blue in the figure, thus ensuring that it is mistakes in the newly made strand that are removed and replaced. Because methylation is the criterion that enables the mismatch repair system to choose the strand that is repaired, the bacterial mismatch repair system is described as being *methyl-directed*.

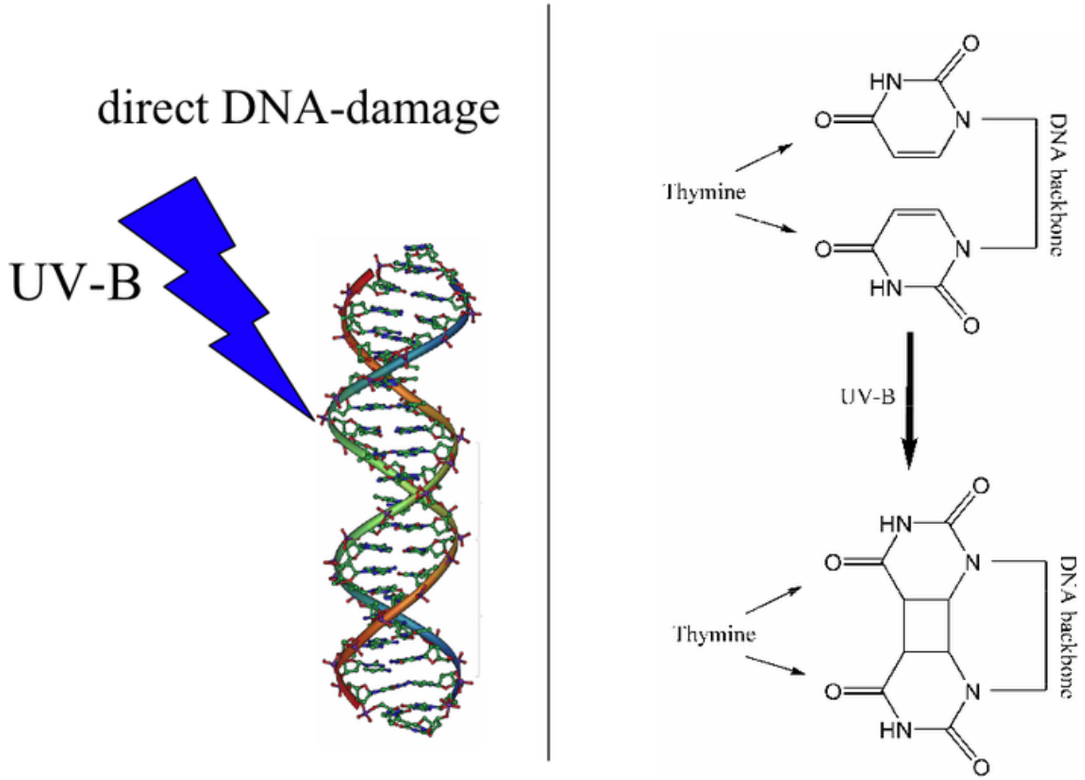
Eukaryotic cells do not use this mechanism to distinguish the new strand from the template, and it is not yet understood how the mismatch repair system in eukaryotes "knows" which strand to repair.

Systems to Repair Damage to DNA

In the preceding section we discussed mistakes made when DNA is copied, where the wrong base is inserted during synthesis of the new strand. But even DNA that is not being replicated can get damaged or mutated. These sorts of damage are not associated with DNA replication, rather they can occur at any time.

What causes damage to DNA?
Some major causes of DNA damage are:

• Radiation (e.g., UV rays in sunlight, in tanning booths)
• Exposure to damaging chemicals (such as benzopyrene in car exhaust and cigarette smoke)
• Chemical reactions within the cell (such as the deamination of cytosine to give uracil).



Thymine Dimer Formation

This means the DNA in your cells is vulnerable to damage simply from normal sorts of actions, such as walking outdoors, being in traffic, or from the chemical transformations occurring in every cell as part of its everyday activities. (Naturally, the damage is much worse in situations where exposure to radiation or damaging chemicals is greater, such as when people repeatedly use tanning beds or smoke.)

What kinds of damage do these agents cause? Radiation can cause different kinds of damage to DNA. Sometimes, as with much of the damage done by UV rays, two adjacent pyrimidine bases in the DNA will be cross-linked to form pyrimidine dimers (note that we are talking about two neighboring pyrimidine bases on the same strand of DNA). This is illustrated in the figure on the previous page where two adjacent thymines on a single DNA strand are cross-linked to form a thymine dimer. Radiation can also cause breaks in the DNA backbone.

The Three R's of DNA

To the tune of "*Dream of Little Dream of Me*"

Base pairs they all provide you
Stair steps to form a helix inside you
A pairs with T and G goes with C
Making DNA for me

Helicases go unwinding
Unzippering at rates almost blindin'
PO-lymerases work night and day
Replicating D-N-A

(Bridge)
Proof-reading - the enzyme's QC path
Chews back from the 3's
I can't have a 'G' paired with 'T' so
Repair it please

Chem damage is concern too
'Cuz it can cause mutations inside you
When dimers stem from sunlight UV
Fix the DNA for me

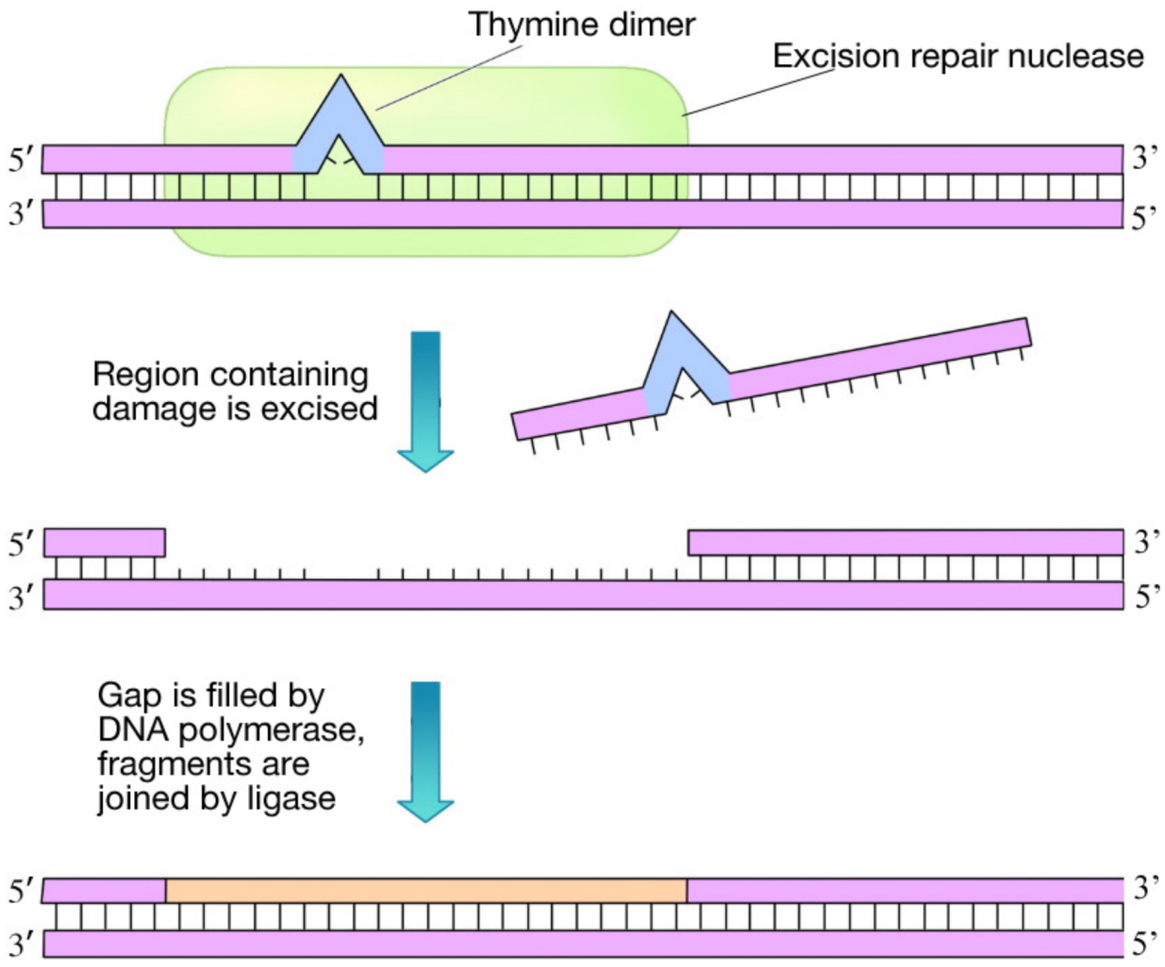
Such pathways of excision
Cause cells to have to make a decision
Should they go straight ahead with repair
Or take themselves right out of there?

(Bridge)
Then lastly, there's recombination
Swap strands readily
Crossover homologous regions
Mix them for me

This story is complete now
The DNA is fit for gametes now
The three R's for the DNA shine
Replicate, repair, recombine

(Oh yeah!)
Replicate, repair, recombine

Recorded by Liz Bacon and David Simmons
Lyrics by Kevin Ahern



Thymine Dimer Removal

Chemicals like benzopyrene can attach themselves to bases, forming bulky DNA adducts in which large chemical groups are linked to bases in the DNA. The formation of chemical adducts can physically distort the DNA helix, making it hard for DNA and RNA polymerases to copy those regions of DNA.

Chemical reactions occurring within cells can cause cytosines in DNA to be deaminated to uracil, as shown in the figure above.

Other sorts of damage in this category include the formation of oxidized bases like 8-oxo-guanine. These do not actually change the physical structure of the DNA helix, but they can cause problems because uracil and 8-oxo-guanine pair with different bases than the original cytosine or guanine, leading to mutations on the next round of replication.

How do cells repair such damage? Cells have several ways to remove the sorts of damage described above, with excision repair being a common strategy. Excision repair is a general term for the cutting out and re-synthesis of the damaged region of the DNA. There are a couple of varieties of excision repair:

Nucleotide Excision Repair (NER)

This system fixes damage by chemicals as well as UV damage. As shown in the figure on the previous page, in nucleotide excision repair, the damage is recognized and a cut is made on either side of the damaged region by an enzyme called an excinuclease (shown in green). A short portion of the DNA strand containing the damage is then removed and a DNA polymerase fills in the gap with the appropriate nucleotides. The newly made DNA is joined to the rest of the DNA backbone by the enzyme DNA ligase. In *E. coli*, NER is carried out by a group of proteins encoded by the uvrABC genes. As you can see, NER is similar, in principle, to mismatch repair. However, in NER, the distortion of

the helix, caused by the DNA damage, clearly indicates which strand of the DNA needs to be removed and replaced.

Base Excision Repair (BER)

BER deals with situations like the deamination of cytosine to uracil. As noted earlier, cytosines in DNA sometimes undergo deamination to form the base uracil.

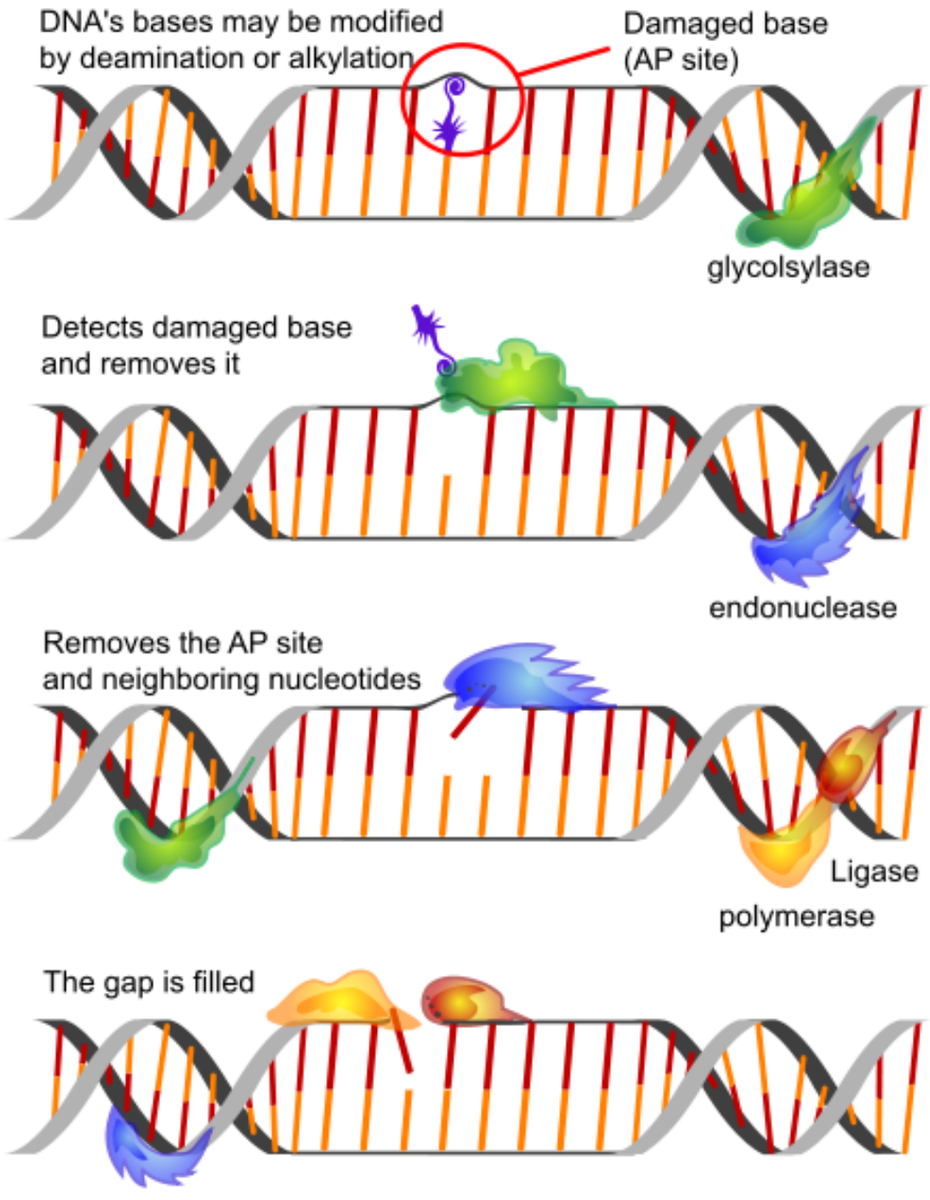
Because cytosines pair with guanines and uracils pair with adenine, the conversion of cytosine to uracil in the DNA would lead to the insertion of an A in the newly replicated strand instead of the G that should have gone in across from a C. To prevent this from happening, uracils are removed from DNA by base excision repair.

In base excision repair, a single base is first removed from the DNA, followed by removal of a region of the DNA surrounding the missing base. The gap is then repaired.

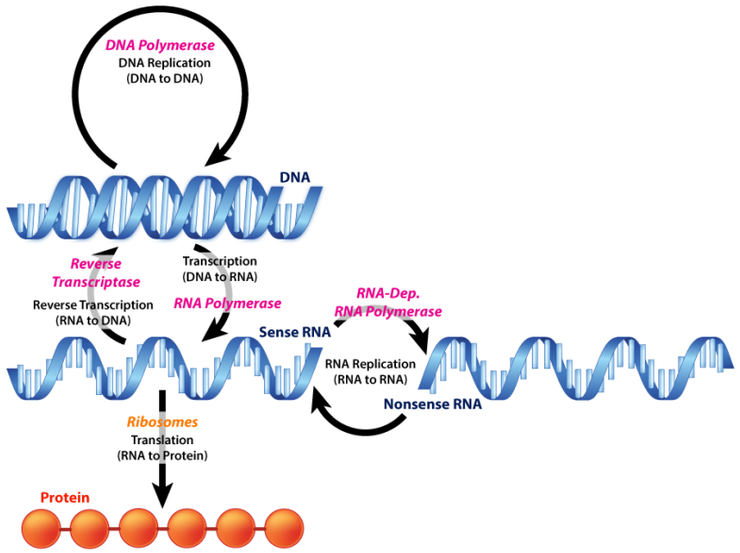
The removal of uracil from DNA is accomplished by the enzyme uracil DNA glycosylase, which breaks the bond between the uracil and the sugar in the nucleotide.

The removal of the uracil base creates a gap called an apyrimidinic site (AP site). The presence of the AP site triggers the activity of an AP endonuclease that cuts the DNA backbone.

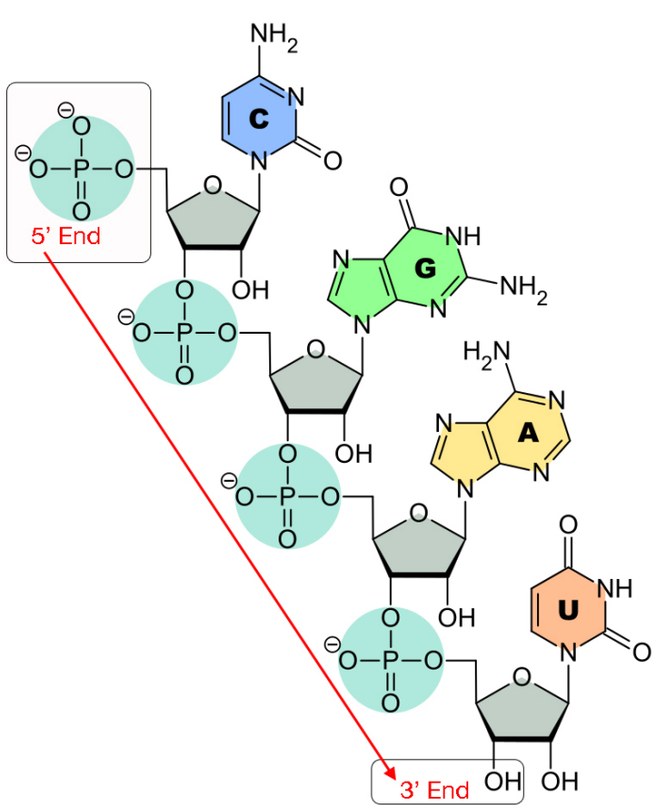
A short region of the DNA surrounding the site of the original uracil is then removed and replaced.



Base Excision Repair



Central Dogma



RNA Structure

Transcription

In the preceding sections, we have discussed the replication of the cell's DNA and the mechanisms by which the integrity of the genetic information is carefully maintained. What do cells do with this information? How does the sequence in DNA control what happens in a cell? If DNA is a giant instruction book containing all of the cell's "knowledge" that is copied and passed down from generation to generation, what are the instructions for? And how do cells use these instructions to make what they need?

Dogma, Central

There's a dogma that's central to cells
That says that the DNA tells
The RNAs to
Make proteins for you
And they'd better get moving, or else.

You have learned in introductory biology courses that genes, which are instructions for making proteins, are made of DNA. You also know that information in genes is copied into temporary instructions called messenger RNAs that direct the synthesis of specific proteins. This description of flow of information from DNA to RNA to protein, shown on the previous page, is often called the Central Dogma of molecular biology and is a good starting point for an examination of how cells use the information in DNA.

Consider that all of the cells in a multicellular organism have

Watch Kevin's YouTube lectures on Transcription HERE, HERE, and HERE

General Features of Transcription

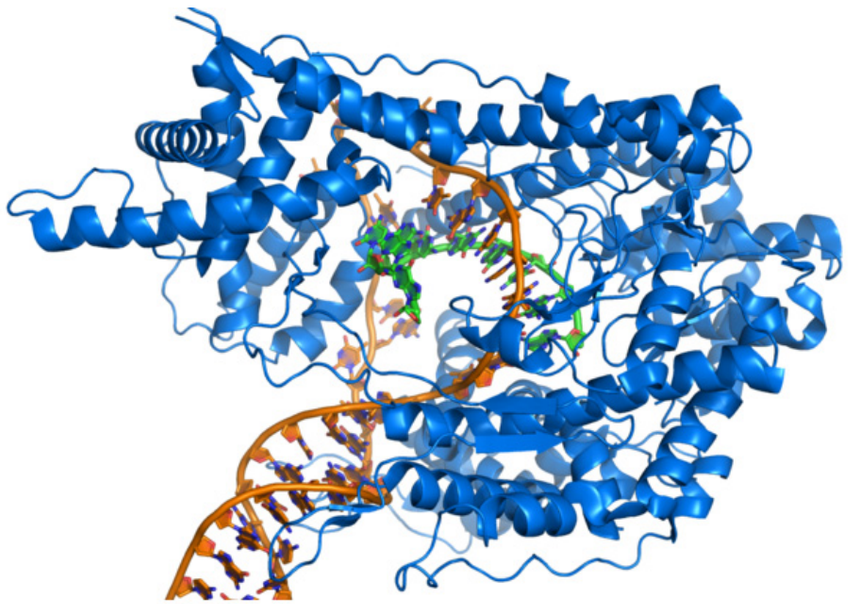
- The process of making RNA copies using a DNA as template is known as transcription
- The process of transcription produces all sorts of RNAs (mRNA, tRNA, rRNA, etc)
- One strand of the DNA serves as a template for the synthesis of RNA
- Enzymes that synthesize RNA are called RNA polymerases
- RNA polymerases synthesize RNA in the 5' to 3' direction
- RNA polymerases do not need a primer
- RNA polymerases uses rNTPs (ATP,GTP,UTP and CTP) to build the new RNA strand
- RNA polymerases bind at specific DNA sequences called promoters to start transcription
- RNA polymerases stop RNA synthesis when they reach sequences called terminators

arisen by division from a single fertilized egg and therefore, all have the same DNA. Division of that original fertilized egg produces, in the case of humans, over a trillion cells, by the time a baby is produced from that egg (that's a lot of DNA replication!). Yet, we also know that a baby is not a giant ball of a trillion identical cells, but has the many different kinds of cells that make up tissues like skin and muscle and bone and nerves. How did cells that have identical DNA turn out so different?

The answer lies in gene expression, which is the process by which the information in DNA is used. Although all the cells in a baby have the same DNA, each different cell type uses a different subset of the genes in that DNA to direct the synthesis of a distinctive set of RNAs and proteins. The first step in gene expression is transcription, which we will examine next.

What is transcription? Transcription is the process of copying information from DNA sequences into RNA sequences.
This process is also known as DNA-dependent RNA synthesis. When a sequence of DNA is transcribed, only one of the two DNA strands is copied into RNA.

But, apart from copying one, rather than both strands of DNA, how is transcription different from replication of DNA? DNA replication serves to copy all the genetic material of the cell and occurs before a cell divides, so that a full copy of the cell's genetic information can be passed on to the daughter cell. Transcription, by contrast, copies short stretches of the coding regions of DNA to make RNA. Different genes may be copied into RNA at different times in the cell's life cycle. RNAs are, so to speak, temporary copies of instructions of the information in DNA and different sets of instructions are copied for use at different times.



T7 RNA Polymerase making RNA (green)
using DNA template (brown)
From Wikimedia Commons

Cells make several different kinds of RNA:
• mRNAs that code for proteins
• rRNAS that form part of ribosomes
• tRNAs that serve as adaptors between mRNA and amino acids during translation
• Micro RNAs that regulate gene expression
• Other small RNAs that have a variety of functions.

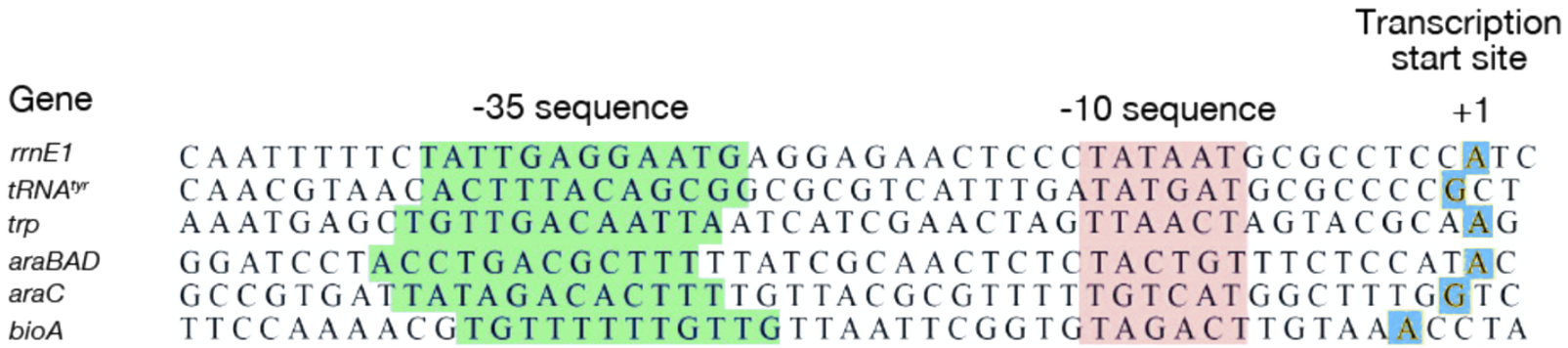
Building an RNA strand is very similar to building a DNA strand. This is not surprising, knowing that DNA and RNA are very similar molecules. What enzyme carries out transcription? Transcription is catalyzed by the enzyme RNA Polymerase. "RNA polymerase" is a general term for an enzyme that makes RNA. There are many different RNA polymerases.

Like DNA polymerases, RNA polymerases synthesize new strands only in the 5' to 3' direction, but because they are making RNA, they use ribonucleotides (i.e., RNA nucleotides) rather than deoxyribonucleotides. Ribonucleotides are joined in exactly the same way as deoxyribonucleotides, which is to say that the 3'OH of the last nucleotide on the growing chain is joined to the 5' phosphate on the incoming nucleotide.

One important difference between DNA polymerases and RNA polymerases is that the latter do not require a primer to start making RNA. Once RNA polymerases are in the right place to start copying DNA, they just begin making RNA by stringing together RNA nucleotides complementary to the DNA template.

This, of course, brings us to an obvious question- how do RNA polymerases "know" where to start copying on the DNA? Unlike the situation in replication, where every nucleotide of the parental DNA must eventually be copied, transcription, as we have already noted, only copies selected genes into RNA at any given time.

Consider the challenge here: in a human cell, there are approximately 6 billion basepairs of DNA. Most of this is non-coding DNA, meaning that it won't need to be transcribed. The small percentage of the genome that is made up of coding sequences still amounts to between 20,000 and 30,000 genes in



Promoter Sequences

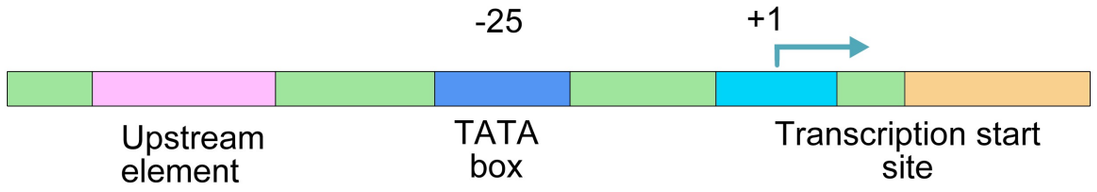
each cell. Of these genes, only a small number will need to be expressed at any given time.

What indicates to an RNA polymerase where to start copying DNA to make a transcript? Signals in DNA indicate to RNA polymerase where it should start (and end) transcription. These signals are special sequences in DNA that are recognized by the RNA polymerase or by proteins that help RNA polymerase determine where it should bind the DNA to start transcription. A DNA sequence at which the RNA polymerase binds to start transcription is called a promoter.

A promoter is generally situated upstream of the gene that it controls. What this means is that on the DNA strand that the gene is on, the promoter sequence is "before" the gene. Remember that, by convention, DNA sequences are read from 5' to 3'. So the promoter lies 5' to the start point of transcription.

Also notice that the promoter is said to "control" the gene it is associated with. This is because expression of the gene is dependent on the binding of RNA polymerase to the promoter sequence to begin transcription. If the RNA polymerase and its helper proteins do not bind the promoter, the gene cannot be transcribed and it will therefore, not be expressed.

What is special about a promoter sequence? In an effort to answer this question, scientists looked at many genes and their surrounding sequences. It makes sense that because the same

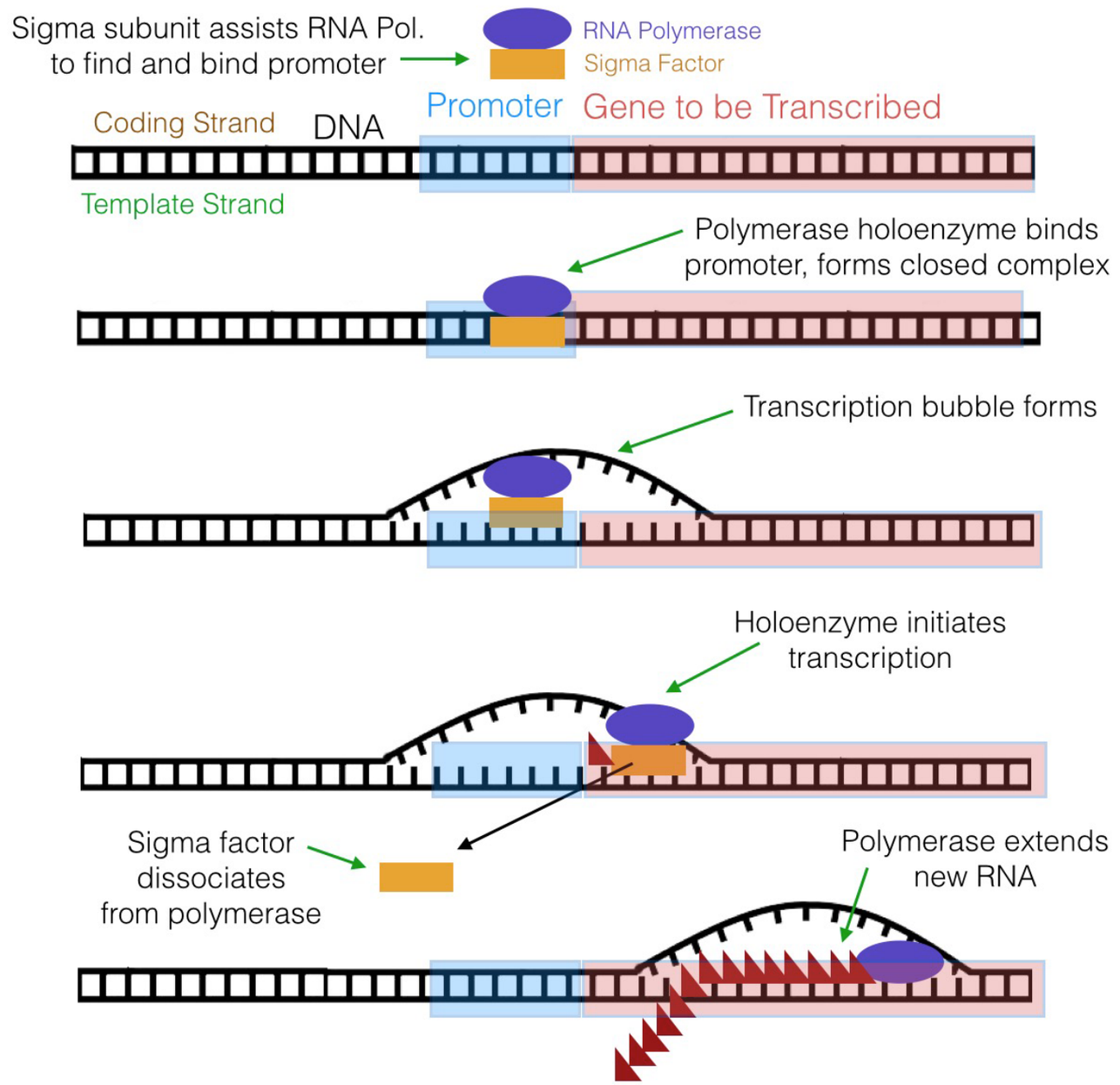


Promoter Sequence Elements

RNA polymerase has to bind to many different promoters, the promoters should have some similarities in their sequences. Sure enough, common sequence patterns were seen to be present in many promoters. We will first take a look at prokaryotic

promoters. When prokaryotic genes were examined, the following features commonly emerged (see figure above):

- A transcription start site (this the base in the DNA across from which the first RNA nucleotide is paired).
- A -10 sequence: this is a 6 bp region centered about 10 bp upstream of the start site. The consensus sequence at this position is TATAAT. In other words, if you count back from the transcription start site, which by convention, is called the +1, the sequence found at -10 in the majority of promoters studied is TATAAT).
- A -35 sequence: this is a sequence at about 35 basepairs upstream from the start of transcription. The consensus sequence at this position is TTGACA.



Transcription Initiation in E. coli

Watch Kevin's YouTube lectures on Transcription HERE, HERE, and HERE

What is the significance of these sequences? It turns out that the sequences at -10 and -35 are recognized and bound by a subunit of prokaryotic RNA polymerase before transcription can begin.

The RNA polymerase of *E. coli*, for example, has a subunit called the sigma subunit (or sigma factor) in addition to the core polymerase, which is the part of the enzyme that actually makes RNA.

Together, the sigma subunit and core polymerase make up what is termed the RNA polymerase holoenzyme. The sigma subunit

of the polymerase (shown in brown in the figure) can recognize and bind to the -10 and -35 sequences in the promoter, thus positioning the RNA polymerase (shown in green) at the right place to initiate transcription. Once transcription begins, the core polymerase and the sigma subunit separate, with the core polymerase continuing RNA synthesis and the sigma subunit wandering off to escort another core polymerase molecule to a promoter. The sigma subunit can be thought of as a sort of usher that leads the polymerase to its "seat" on the promoter.

The Student's Guide to a Perfect Transcript

Pol II's so smart, now let me see
It makes a transcript, one, two, three,
From a billion base pairs can it find
The one promoter it must bind?

I need some help, I hear it plead,
This DNA I cannot read
I cannot see where I must start
'Til TFIIs have done their part.

When TATA's bound by TBP
That sends a signal out, you see
Once DNA has made a bend
More TFIIs will soon attend.

When B arrives, it clears a place
For RNA polymerase
And F and E soon join the fun
TFIIH, the final one.

The last one is a special case
It moonlights as a helicase
And sends Pol II upon its way
To make a brand new RNA.

How does it do that, you may ask
A phosphate does the crucial task
Added on to CTD,
That's how the H sets Pol II free.

And Pol II goes its merry way
There soon will be some RNA
Then introns must be all excised
As RNA is capped and spliced.

Then PolyA polymerase
Will add a couple hundred As
For my own transcript, I can say
That all I want is one more A.

Verse by Indira Rajagopal

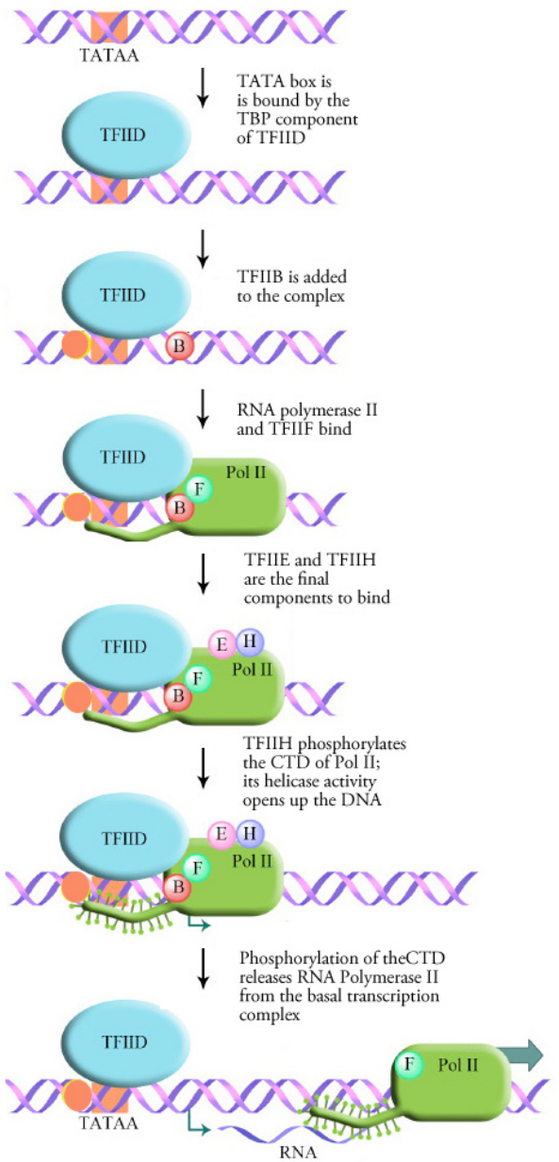
As already mentioned, an RNA chain, complementary to the DNA template, is built by the RNA polymerase by the joining of the 5' phosphate of an incoming ribonucleotide to the 3'OH on the last nucleotide of the growing RNA strand. How does the polymerase know where to stop? A sequence of nucleotides called the terminator is the signal to the RNA polymerase to stop transcription and dissociate from the template.

Although the process of RNA synthesis is the same in eukaryotes as in prokaryotes, there

are some additional issues to keep in mind in eukaryotes. One is that in eukaryotes, the DNA template exists as chromatin, where the DNA is tightly associated with histones and other proteins. The "packaging" of the DNA must therefore be opened up to allow the RNA polymerase access to the template in the region to be transcribed.

A second difference is that eukaryotes have multiple RNA polymerases, not one as in bacterial cells. The different polymerases transcribe different genes. For example, RNA polymerase I transcribes the ribosomal RNA genes, while RNA polymerase III copies tRNA genes. The RNA polymerase we will focus on most is RNA polymerase II, which transcribes protein-coding genes to make mRNAs.

All three eukaryotic RNA polymerases need additional proteins to help them get transcription started. In prokaryotes, RNA polymerase by itself can initiate transcription (remember that the sigma subunit is a subunit of the prokaryotic RNA polymerase). The additional proteins needed by eukaryotic RNA polymerases are referred to as transcription factors. We will see below that there are various categories of transcription factors.



Assembly of the Basal Transcription Complex and Initiation of Transcription

Finally, in eukaryotic cells, transcription is separated in space and time from translation. Transcription happens in the nucleus, and the mRNAs produced are processed further before they are sent into the cytoplasm. Protein synthesis (translation) happens in the cytoplasm. In prokaryotic cells, mRNAs can be translated as they are coming off the DNA template, and because there is no nucleus, transcription and protein synthesis occur in a single cellular compartment.

Like genes in prokaryotes, eukaryotic genes also have promoters. Eukaryotic promoters commonly have a TATA box, a sequence about 25 basepairs upstream of the start of transcription that is recognized and bound by

Watch Kevin's YouTube lectures on Transcription HERE, HERE, and HERE

proteins that help the RNA polymerase to position itself correctly to begin transcription. (Some eukaryotic promoters lack TATA boxes, and have, instead, other recognition sequences to help the RNA polymerase find the spot on the DNA where it spot on the DNA where it binds and initiates transcription.)

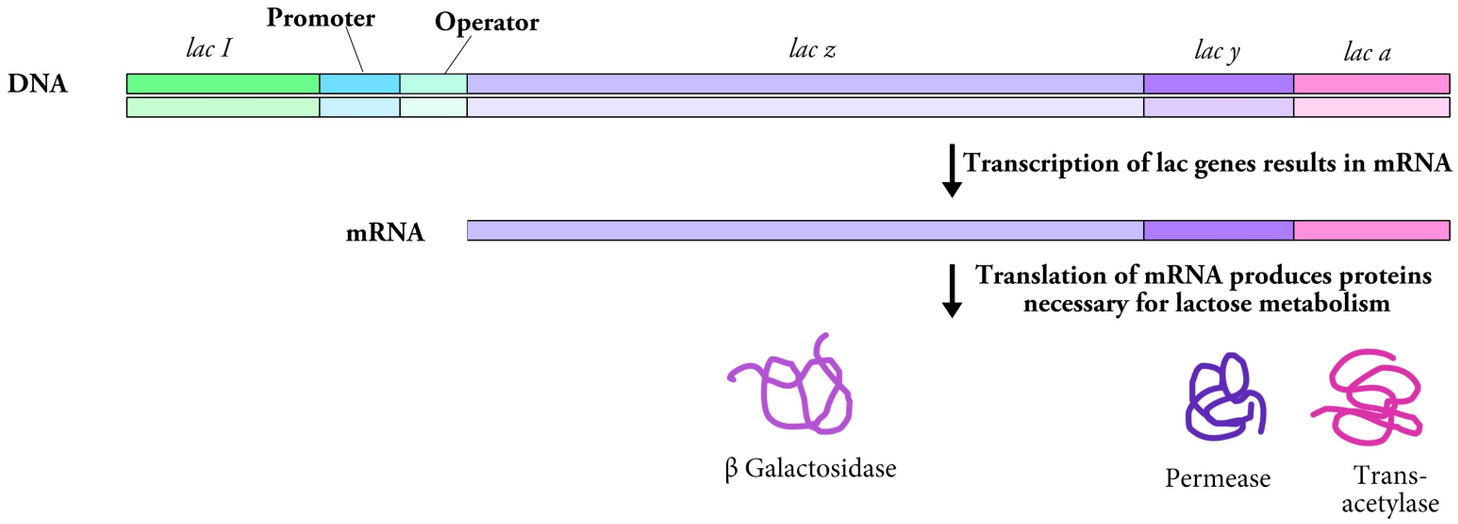
We noted earlier that eukaryotic RNA polymerases need additional proteins to bind promoters and start transcription. What are these additional proteins that are needed to start transcription? General transcription factors are proteins that help eukaryotic RNA polymerases find transcription start sites and initiate RNA synthesis. We will focus on the transcription factors that assist RNA polymerase II. These transcription factors are named TFIIA, TFIIB and so on (TF= transcription factor, II=RNA polymerase II, and the letters distinguish individual transcription factors).

Transcription in eukaryotes requires the general transcription factors and the RNA polymerase to form a complex at the TATA box called the basal transcription complex or transcription initiation complex. This is the minimum requirement for any gene to be transcribed. The first step in the formation of this complex is the binding of the TATA box by a transcription factor called the TATA Binding Protein or TBP. Binding of the TBP causes the DNA to bend at this spot and take on a structure that is suitable for the binding of additional transcription factors and RNA polymerase. As shown in the figure at left, a number of different general transcription factors, together with RNA polymerase (Pol II) form a complex at the TATA box.

The final step in the assembly of the basal transcription complex is the binding of a general transcription factor called TFIIH. TFIIH is a multifunctional protein that has helicase activity (i.e., it is capable of opening up a DNA double helix) as well as kinase activity. The kinase activity of TFIIH adds a phosphate onto the C-terminal domain (CTD) of the RNA polymerase. This phosphorylation appears to be the signal that releases the RNA polymerase from the basal transcription complex and allows it to move forward and begin transcription.

Regulation of Transcription

The processes described above are required whenever any gene is transcribed. But what determines which genes are transcribed at a given time? What are the molecular switches that turn transcription on or off? Although there are entire books written on this one topic, the basic mechanism by which transcription is regulated depends on highly specific interactions between transcription regulating proteins and regulatory sequences on DNA.

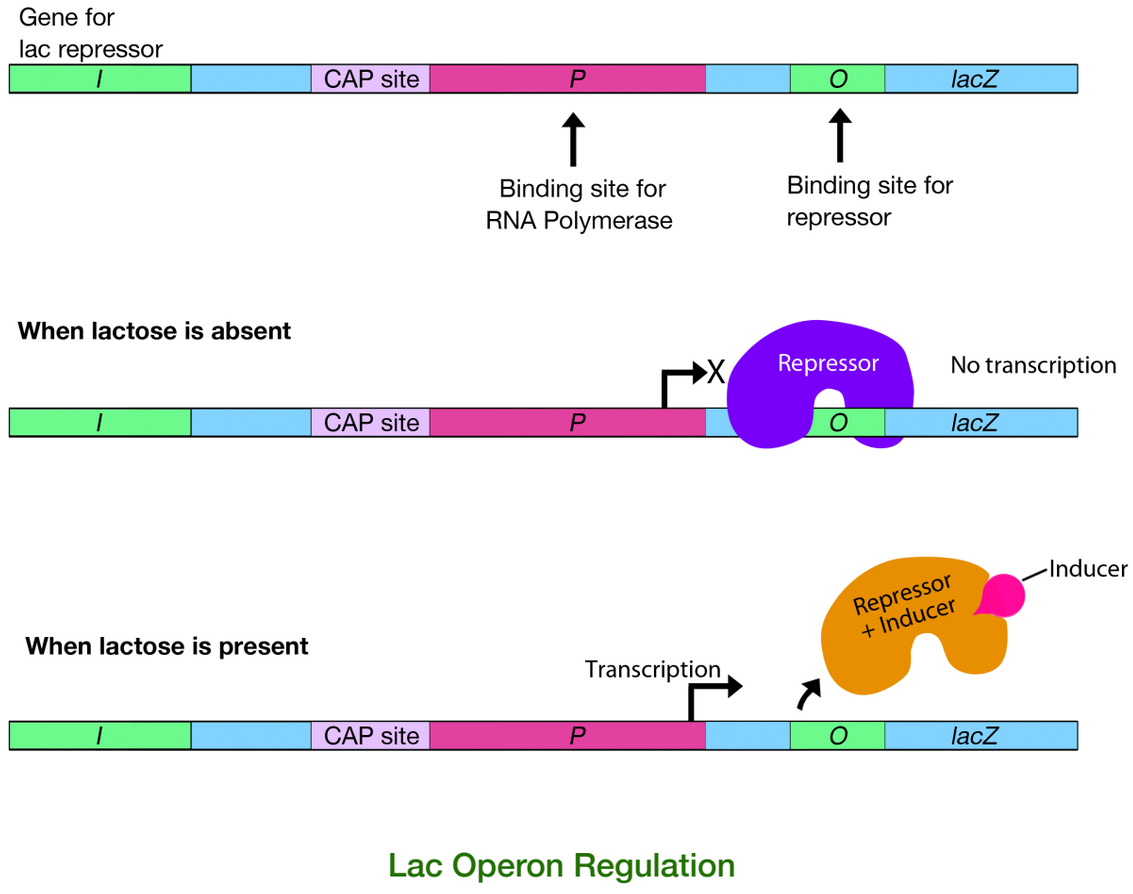


The genes lac z, lac y and lac a are all under the control of a single promoter in the lac operon

We know that promoters indicate where transcription begins, but what determines that a given gene will be transcribed? In addition to the promoter sequences required for transcription initiation, genes have additional regulatory sequences (sequences of DNA on the same DNA molecule as the gene) that control when a gene is transcribed. Regulatory sequences are bound tightly and specifically by transcriptional regulators, proteins that can recognize DNA sequences and bind to them. The binding of such proteins to the DNA can regulate transcription by preventing or increasing transcription from a particular promoter.

Let us first consider an example from prokaryotes. In bacteria, genes are often clustered in groups, such that genes that need to be expressed at the same time are next to each other and all of them are controlled as a single unit by the same promoter. The lac operon, shown below, is one such group of genes that encode proteins needed for the uptake and breakdown of the sugar lactose. The three genes of the lac operon, lac z, lac y and lac a are controlled by a single promoter.

Bacterial cells generally prefer to use glucose for their energy needs, but if glucose is unavailable, and lactose is present, the

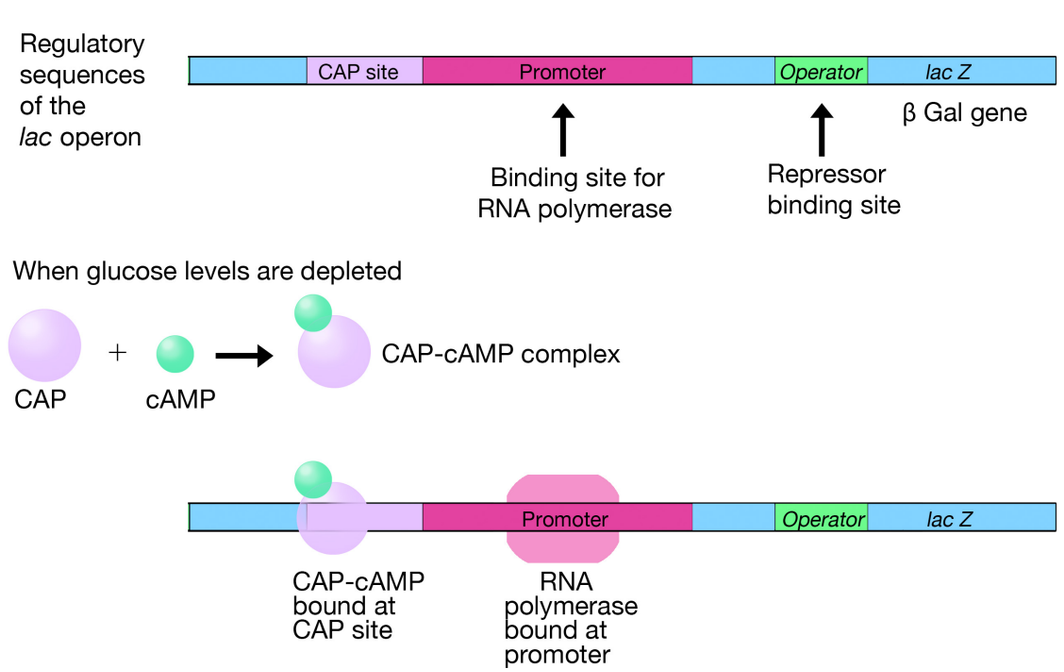


Lac Operon Regulation

bacteria will take up lactose and break it down for energy. Since the proteins for taking up and breaking down lactose are only needed when glucose is absent and lactose is available, the bacterial cells need a way to express the genes of the lac operon only under those conditions. At times when lactose is absent, the cells do not need to express these genes.

Watch Kevin's YouTube lectures on Transcription Regulation HERE, HERE, and HERE

How do bacteria achieve this?
Transcription of the lac cluster of genes is primarily controlled by a repressor protein that binds to a region of the DNA just downstream of the -10 sequence of the lac promoter. Recall that the promoter is where the RNA polymerase must bind to begin transcription. The place where the repressor is bound is called the operator (labeled O in the figure). When the repressor is bound at this position, it physically blocks the RNA polymerase from transcribing the genes, just as a vehicle blocking your driveway would prevent you from pulling out.



Lac Operon Activation

Obviously, if you want to leave, the vehicle that is blocking your path must be removed. Likewise, in order for transcription to occur, the repressor must be removed from the operator to clear the path for RNA polymerase. How is the repressor removed?

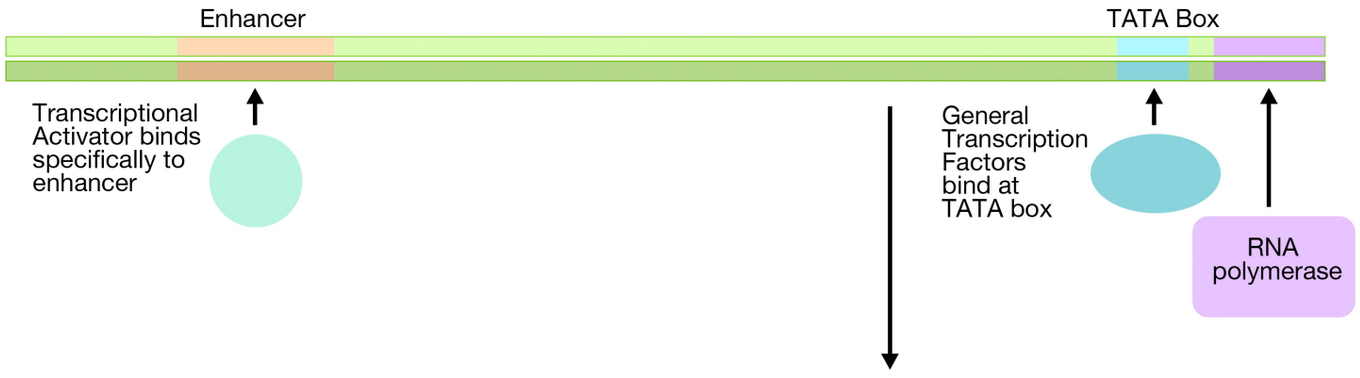
When the sugar lactose is present, it binds to the repressor, changing its conformation so that it no longer binds to the operator. When the repressor is no longer bound at the operator, the "road-block" in front of the RNA polymerase is removed, permitting the transcription of the genes of the lac operon.

Because the binding of the lactose induces the expression of the genes in the lac operon, lactose is called an inducer. (Technically, the inducer is allolactose, a molecule made from lactose by the cell, but the principle is the same.)

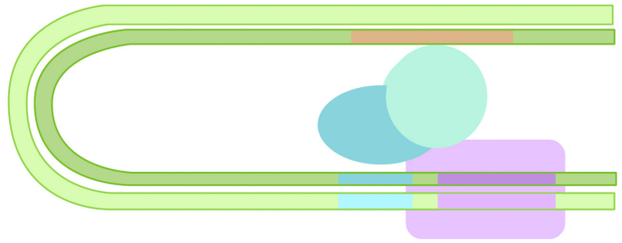
What makes this an especially

effective control system is that the genes of the lac operon encode proteins that break down lactose. Turning on these genes requires lactose to be present. Once the lactose is broken down, the repressor binds to the operator once more and the lac genes are no longer expressed. This allows the genes to be expressed only when they are needed.

But how do glucose levels affect the expression of the lac genes? We noted earlier that if glucose was present, lactose would not be used. A second level of control is exerted by a protein called CAP that binds to a site adjacent to the promoter and recruits RNA polymerase to bind the lac promoter. When glucose is depleted, there is an increase in levels of cAMP which binds to CAP. The CAP-cAMP complex then binds the CAP site, as shown in the figure. The combination of CAP binding and the lac repressor dissociating from the operator when lactose levels are high ensures high levels of transcription of the lac operon just when it is most needed. The CAP protein binding may be thought of as a green light for the RNA polymerase, while the removal of repressor is like the lifting of a barricade in front of it. When both conditions are met, the RNA polymerase transcribes the downstream genes.



DNA looping allows transcriptional activator bound to enhancer to interact with basal transcription complex at TATA box



Enhancer Mechanism

The lac operon we have just described is a set of genes that are expressed only under the specific conditions of glucose depletion and lactose availability. Other genes may be expressed *unless* a particular condition is met. An example of this is the *trp* operon in bacterial cells, which encodes enzymes necessary for the synthesis of the amino acid tryptophan. These genes are expressed at all times, except when tryptophan is available from the cell's surroundings. This means that these genes must be prevented from being expressed in the presence of tryptophan. This is achieved by having a

Watch Kevin's YouTube lectures on Transcription Regulation HERE, HERE, and HERE

repressor protein that will bind to the operator only in the presence of tryptophan.

Transcription in eukaryotes is also regulated by the binding of proteins to specific DNA sequences, but with some differences from the simple schemes outlined above. For most eukaryotic genes, general transcription factors and RNA polymerase (i.e., the basal transcription complex) are necessary, but not sufficient, for high levels of transcription.

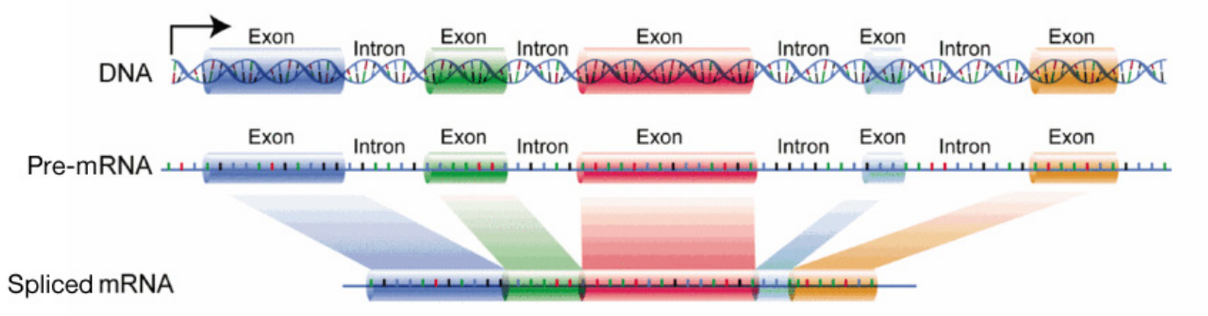
In eukaryotes, additional regulatory sequences called enhancers and the proteins that bind to the enhancers are needed to achieve high levels of transcription. Enhancers are DNA sequences that regulate the transcription of genes. Unlike prokaryotic regulatory sequences, enhancers don't need to be next to the gene they control. Often they are many kilobases away on the DNA. As the name suggests, enhancers can enhance (increase) transcription of a particular gene.

How can a DNA sequence far from the gene being transcribed affect the level of its transcription?

Enhancers work by binding proteins (transcriptional activators) that can, in turn, interact with the proteins bound at the promoter. The enhancer region of the DNA, with its associated transcriptional activator(s) can come in contact with the basal transcription complex that is bound at a distant TATA box by looping of the DNA (previous page). This allows the protein bound at the enhancer to make contact with the proteins in the basal transcription complex.

One way that the transcriptional activator bound to the enhancer increases the transcription from a distant promoter is that it increases the frequency and efficiency with which the basal transcription complex is formed at the promoter.

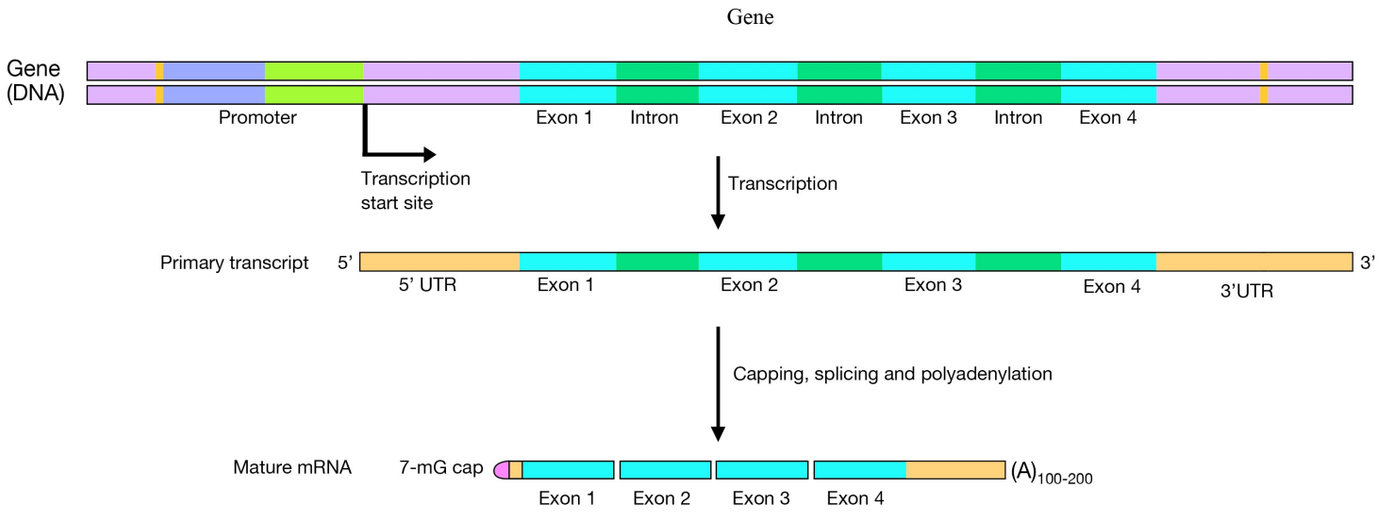
Another mechanism by which proteins bound at the enhancer can affect transcription is by recruiting to the promoter other proteins that can modify the structure of the chromatin in that region. As we noted earlier, in eukaryotes, DNA is packaged with proteins to form chromatin. When the DNA is tightly associated with these proteins, it is difficult to access for transcription. So proteins that can make the DNA more accessible to the transcription



RNA Splicing

machinery can also play a role in the extent to which transcription occurs.

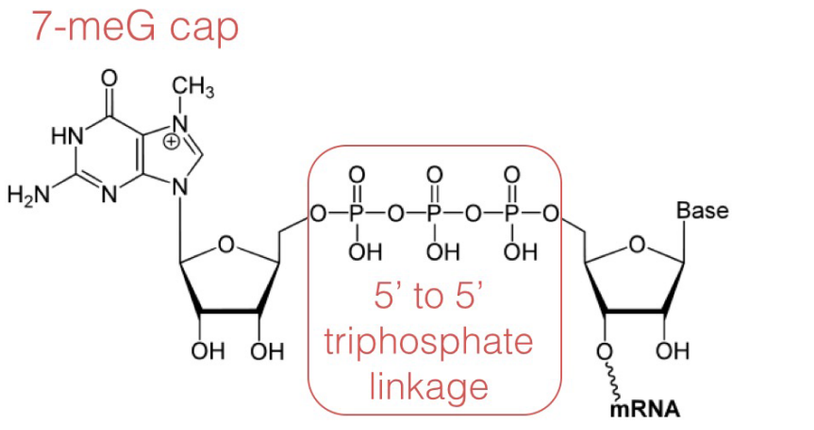
In addition to enhancers, there are also negative regulatory sequences called silencers. Such regulatory sequences bind to transcriptional repressor proteins. Transcriptional activators and repressors are modular proteins- they have a part that binds DNA and a part that activates or represses transcription by interacting with the basal transcription complex.



Steps in Processing Eukaryotic Messenger RNAs

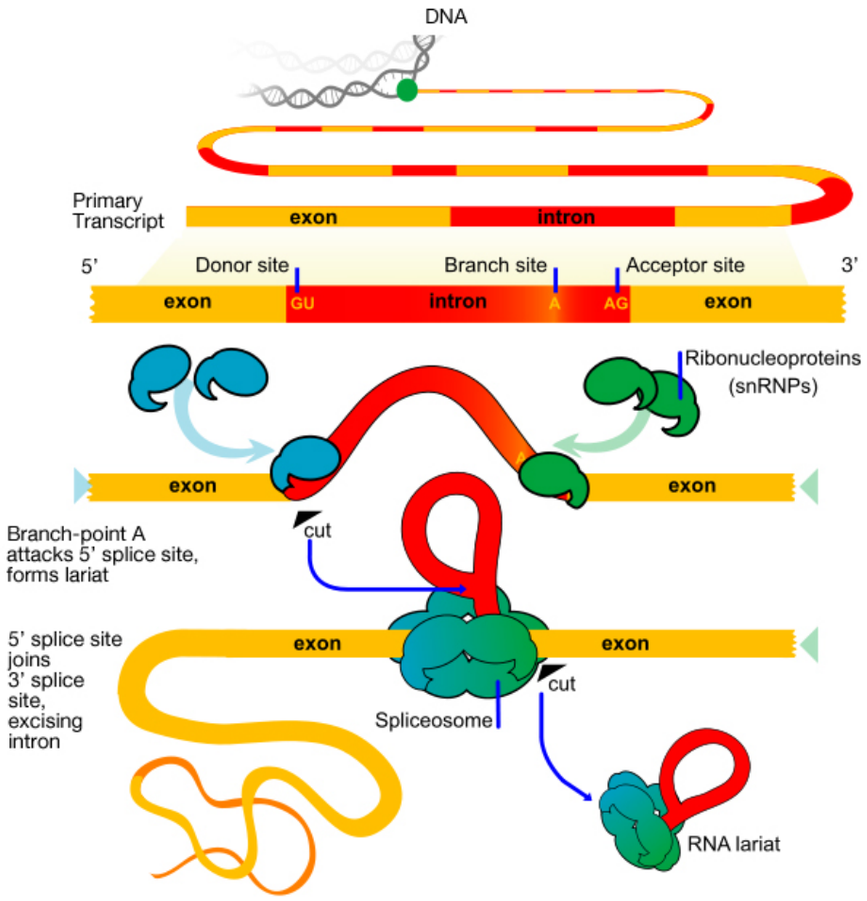
RNA Processing

So far, we have looked at the mechanism by which the information in genes (DNA) is transcribed into RNA. The newly made RNA, also known as the primary transcript (the product of transcription is known as a transcript) is further processed before it is functional. Both prokaryotes and eukaryotes process their ribosomal and transfer RNAs.



mRNA Cap Structure

The major difference in RNA processing, however, between prokaryotes and eukaryotes, is in the processing of messenger RNAs. We will focus on the processing of mRNAs in this discussion. You will recall that in bacterial cells, the mRNA is translated directly as it comes off the DNA template. In eukaryotic cells, RNA synthesis, which occurs in the nucleus, is separated from the protein synthesis machinery, which is in the cytoplasm. In addition, eukaryotic genes have introns, noncoding regions that interrupt the gene's coding sequence. The mRNA copied from genes containing introns will also therefore have regions that interrupt the information in the gene. These regions must be



Splicing

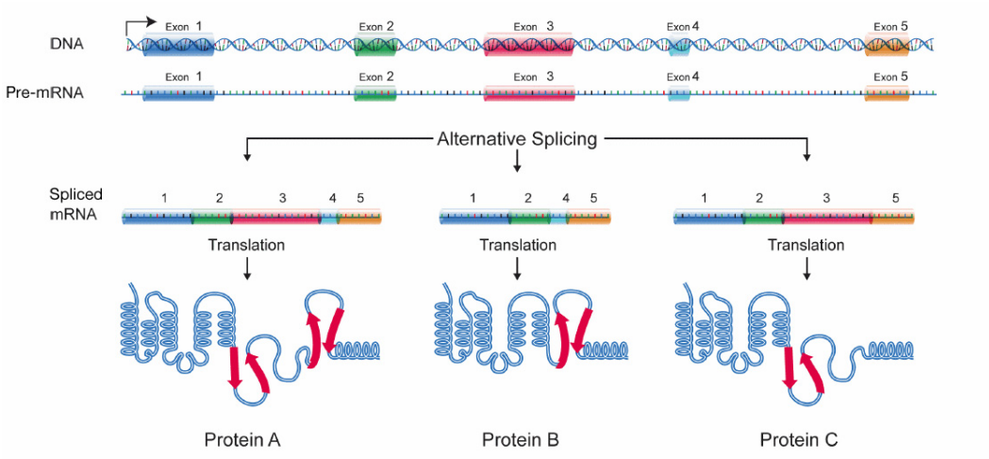
removed before the mRNA is sent out of the nucleus to be used to direct protein synthesis. The process of removing the introns and rejoining the coding sections or exons, of the mRNA, is called splicing. Once the mRNA has been capped, spliced and had a polyA tail added, it is sent from the nucleus into the cytoplasm for translation.

The initial product of transcription of a protein coding gene is called the pre-mRNA (or primary transcript). After it has been processed and is ready to be exported from the nucleus, it is called the mature mRNA or processed mRNA.

What are the processing steps for messenger RNAs?
In eukaryotic cells, pre-mRNAs undergo three main processing steps:

- Capping at the 5' end
- Addition of a polyA tail at the 3' end. and
- Splicing to remove introns

In the capping step of mRNA processing, a 7-methyl guanosine (shown at left) is added at the 5' end of the mRNA. The cap protects the 5' end of the mRNA from degradation by nucleases and also helps to position the mRNA correctly on the ribosomes during protein synthesis.



Splicing and Protein Diversity

The 3' end of a eukaryotic mRNA is first trimmed, then an enzyme called PolyA Polymerase adds a "tail" of about 200 ‘A’ nucleotides to the 3' end. There is evidence that the polyA tail plays a role in efficient translation of the mRNA, as well as in the stability of the mRNA. The cap and the polyA tail on an mRNA are also indications that the mRNA is complete (i.e., not defective).

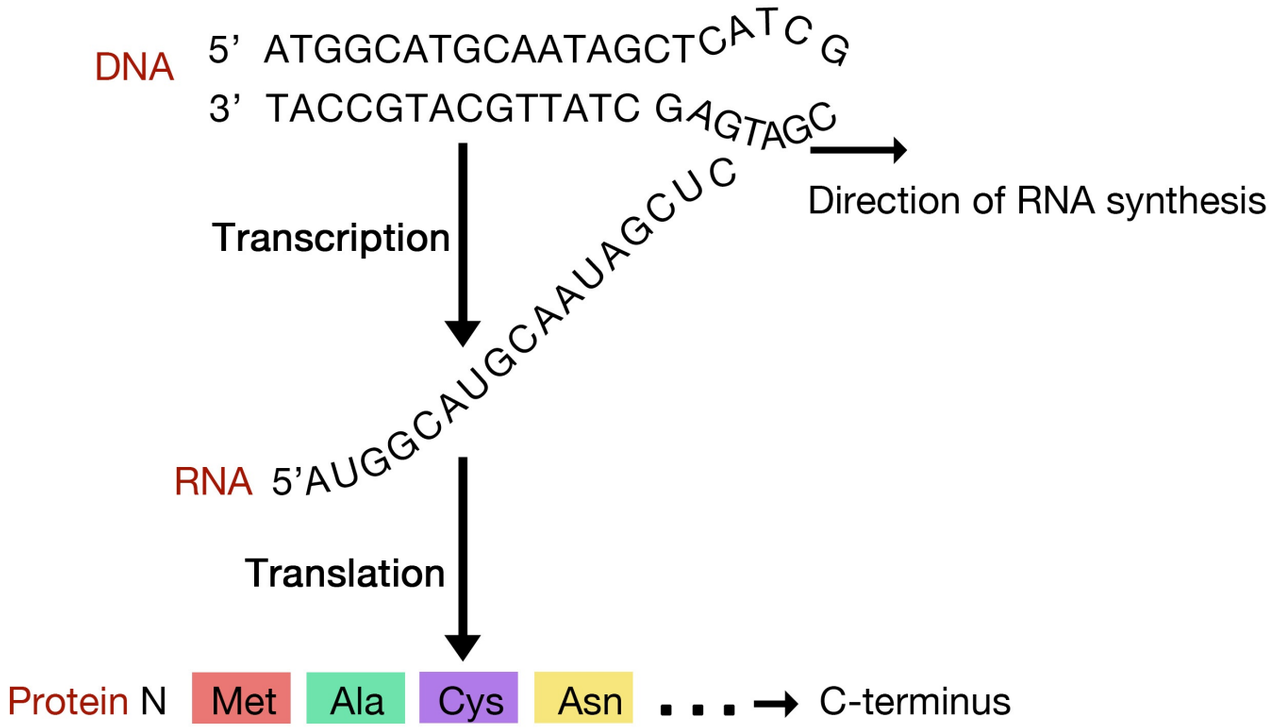
Introns are removed from the pre-mRNA by the activity of a complex called the spliceosome. The spliceosome is made up of proteins and small RNAs that are associated to form protein-RNA enzymes called small nuclear ribonucleoproteins or snRNPs (pronounced SNURPS).
The splicing machinery must be able to recognize splice junctions (i.e., the end of each exon and the start of the next) in order to correctly cut out the introns and join the exons to make the mature, spliced mRNA.

What signals indicate where an intron starts and ends?
The base sequence at the start (5' or left end, also called the donor site) of an intron is GU while the sequence at the 3' or right end (a.k.a. acceptor site) is AG. There is also a third important sequence within the intron, called a branch point, that is important for splicing.

Watch Kevin’s YouTube lectures on Translation HERE, HERE, and HERE

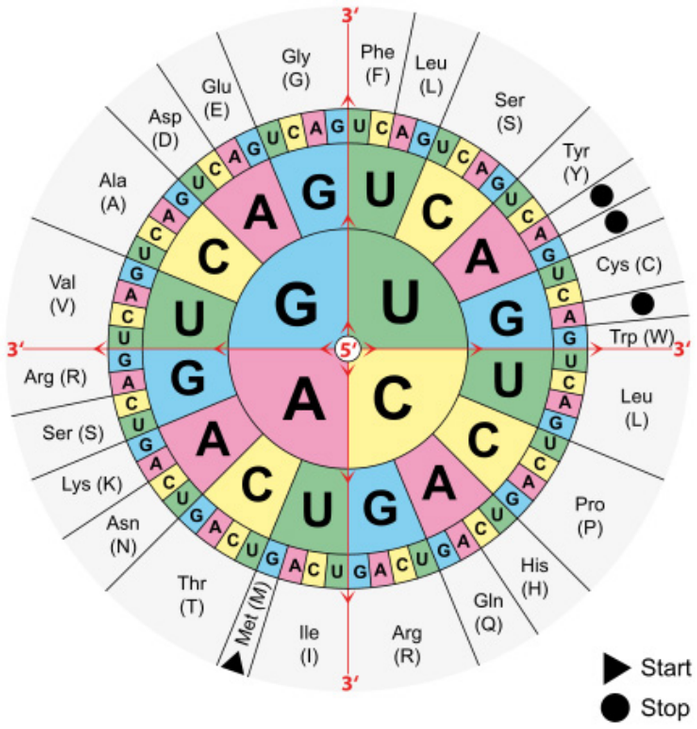
There are two main steps in splicing:

- In the first step, the pre-mRNA is cut at the 5' splice site (the junction of the 5' exon and the intron). The 5' end of the intron then is joined to the branch point within the intron. This generates the lariat-shaped molecule characteristic of the splicing process



Coupled Transcription and Translation in Prokaryotes

- In the second step, the 3' splice site is cut, and the two exons are



The Genetic Code

joined together, and the intron is released.

Many pre-mRNAs have a large number of exons that can be spliced together in different combinations to generate different mature mRNAs. This is called alternative splicing, and allows the production of many different proteins using relatively few genes, since a single RNA can, by combining different exons during splicing, create many different protein coding messages. Because of alternative splicing, each gene in our DNA gives rise, on average, to three different proteins.

Once protein coding messages have been processed by capping, splicing and addition of a poly A tail, they are transported out of the nucleus to be translated in the cytoplasm.

Translation

Translation is the process by which information in mRNAs is used to direct the synthesis of proteins. As you have learned in introductory biology, in eukaryotic cells, this process is carried out in the

The Codon Song

To the tune of "*When I'm Sixty Four*"

Building of proteins, you oughta know
Needs amino A's
Peptide bond catalysis in ribosomes
Triplet bases, three letter codes
Mixing and matching nucleotides
Who is keeping score?
Here is the low down
If you count codons
You'll get sixty four

Got - to - line - up - right
16-S R-N-A and
Shine Dalgarno site

You can make peptides, every size
With the proper code
Start codons positioned
In the P site place
Initiator t-RNAs
UGA stops and AUGs go
Who could ask for more?
You know the low down

Count up the codons

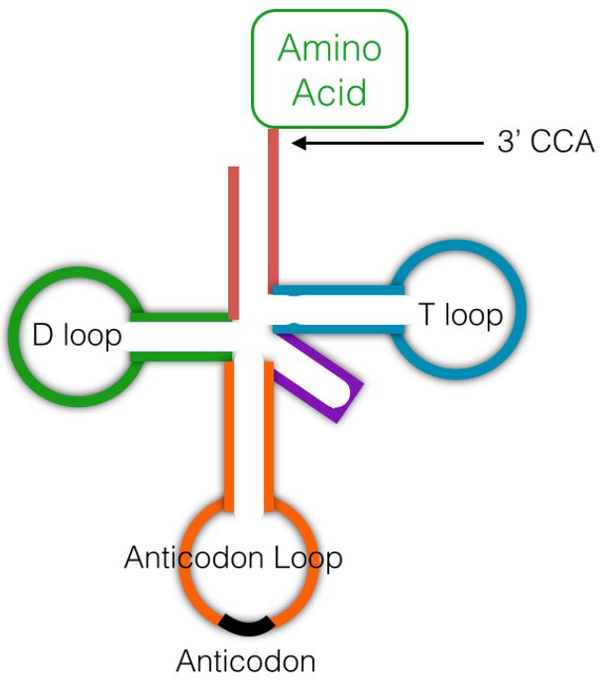
There are sixty four

T

Recorded by Tim Karplus
Lyrics by Kevin Ahern

cytoplasm of the cell, by large RNA-protein machines called ribosomes. Ribosomes contain ribosomal RNAs (rRNAs) and proteins. The proteins and rRNAs are organized into two subunits, a large and a small. The large subunit has an enzymatic activity, known as a peptidyl transferase, that makes the peptide bonds that join amino acids to make a polypeptide. The small and large subunits assemble on the mRNA at its 5'end to initiate translation. Ribosomes function by binding to mRNAs and holding them in a way that allows the amino acids encoded by the RNA to be joined sequentially to form a polypeptide.

Watch Kevin's YouTube lectures on Translation HERE, HERE, and HERE

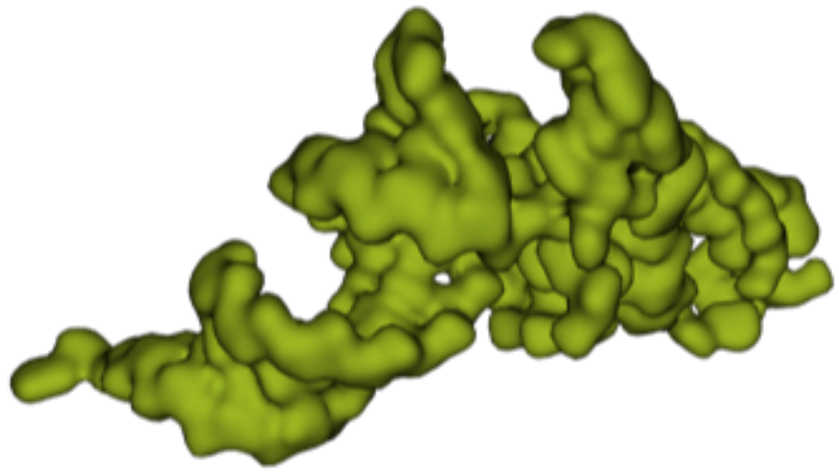


tRNA Structure

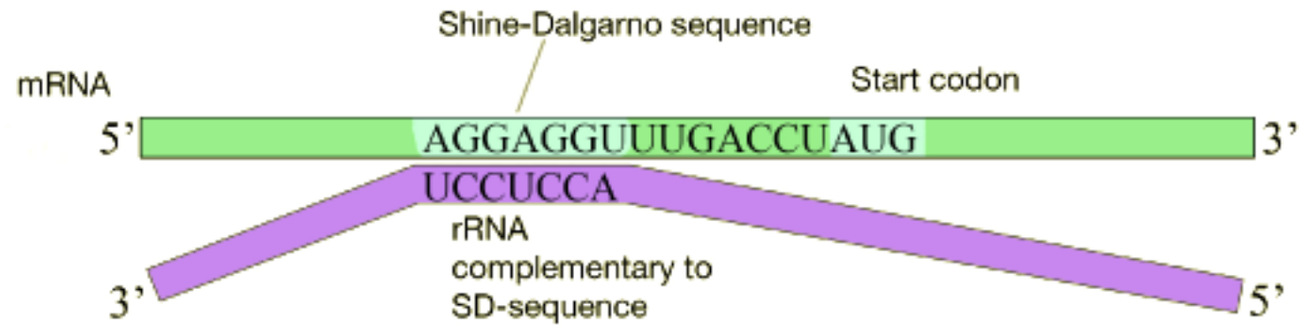
The sequence of an mRNA directly specifies the sequence of amino acids in the protein it encodes. Each amino acid in the protein is specified by a sequence of 3 bases called a codon in the mRNA. For example, the amino acid tryptophan is encoded by the sequence 5'UGG3' on an mRNA.

Given that there are 4 bases in RNA, the number of different 3-base combinations that are possible is 4^3, or 64. There are, however, only 20 amino acids that are used in building proteins. This discrepancy in the number of possible codons and the actual number of amino acids they specify is explained by the fact that the same amino acid may be specified by more than one codon. In fact, with the exception of the amino acids methionine and tryptophan, all the other amino acids are encoded by multiple codons. The

Interactive 5.2



Phenylalanyl-tRNA



mRNA Alignment by Shine-Dalgarno Sequence

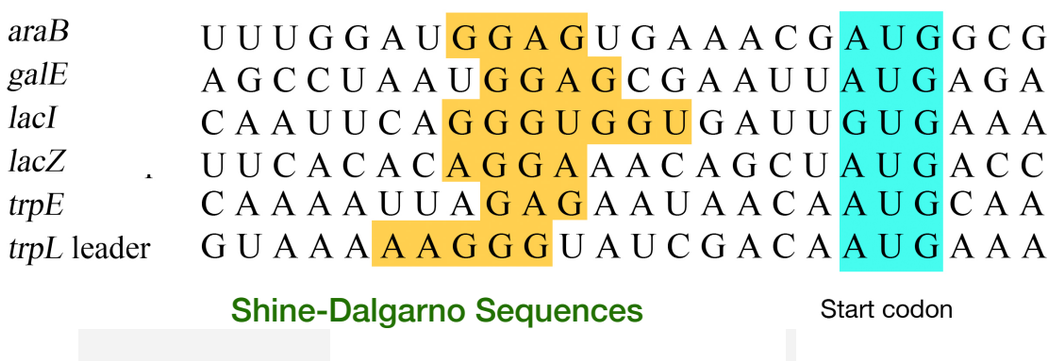
figure above shows the codons that are used for each of the twenty amino acids.

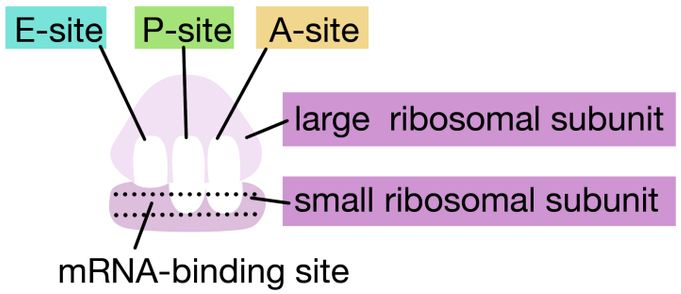
Three of the 64 codons are known as termination or stop codons and as their name suggests, indicate the end of a protein coding sequence. The codon for methionine, AUG, is used as the start, or initiation, codon.

This ingenious system is used to direct the assembly of a protein in the same way that you might string together colored beads in a particular order using instructions that used symbols like 111 for a red bead, followed by 222 for a green bead, 333 for yellow, and so on, till you came to 000, indicating that you should stop stringing beads.

While the ribosomes are literally the protein factories that join amino acids together using the instructions in mRNAs, another class of RNA molecules, the transfer RNAs (tRNAs) are also needed for translation. Transfer RNAs (see figure, left) are small RNA molecules, about 75-80 nucleotides long, that function to 'interpret' the instructions in the mRNA during protein synthesis. In terms of the bead analogy above, someone, or something, has to be able to bring a red bead in when the instructions indicate 111, and a green bead when the instructions say 222. Unlike a human, who can choose a red bead when 111 is present in the instructions, neither ribosomes nor tRNAs can think. The system, therefore, relies, like so many processes in cells, solely on molecular recognition.

A given transfer RNA is specific for a particular amino acid. It is linked covalently to this amino acid at its 3' end by an enzyme called aminoacyl tRNA synthetase. There is an aminoacyl tRNA synthetase specific for each amino acid. A tRNA with an amino acid attached to it is said to be charged. Another region of the tRNA has a sequence of 3 bases, the anticodon, that is



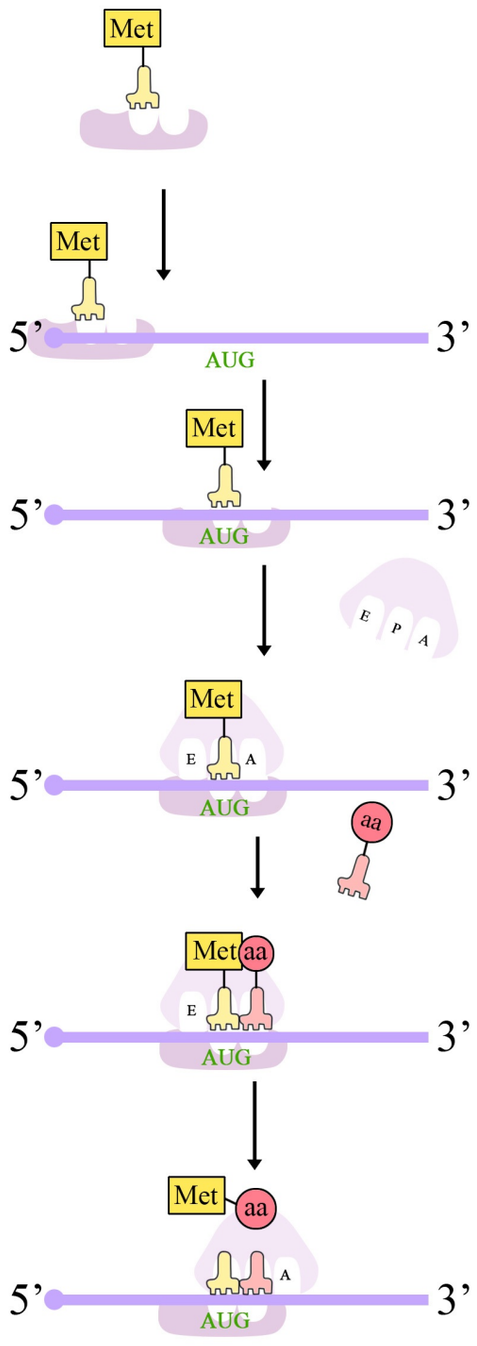


A,P and E Sites in the Ribosome

complementary to the codon for the amino acid it is carrying. When the tRNA encounters the codon for its amino acid on the messenger RNA, the anticodon will base-pair with the codon, and the amino acid attached to it will be brought in to the ribosome to be added on to the growing protein chain.

With an idea of the various components necessary for translation we can now take a look at the process of protein synthesis. The main steps in the process are similar in prokaryotes and eukaryotes. As we already noted, ribosomes bind to mRNAs and facilitate the interaction between the codons in the mRNA and the anticodons on charged tRNAs.

In bacterial cells, translation is coupled with transcription and begins even before the mRNA has been completely synthesized. How does the ribosome recognize and bind to the mRNA? Many



Initiation of Translation

Charged initiator tRNA and small ribosomal subunit

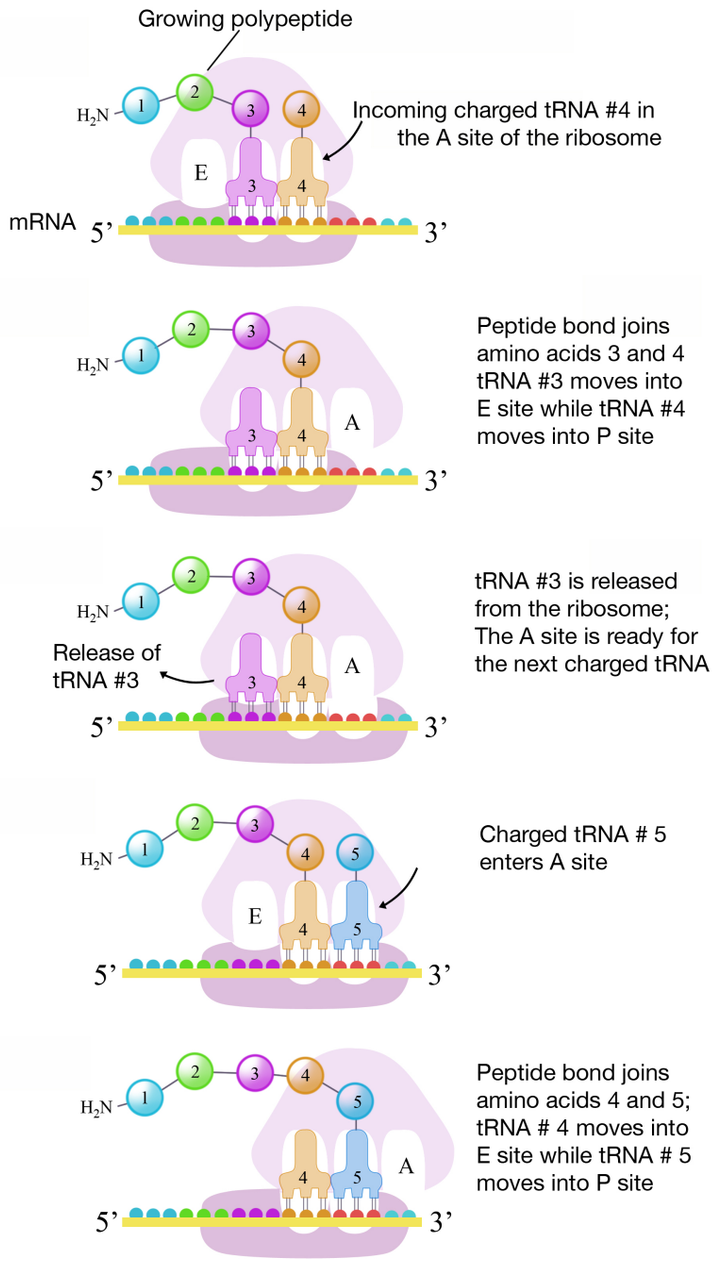
Small ribosomal subunit binds to ribosome binding site on mRNA

Methionine-charged tRNA binds to the start codon (AUG)

Large ribsomal subunit joins initiation complex. Initiator tRNA occupies the P-site

Second tRNA corresponding to the next codon occupies A-site

A peptide bond is made between the methionine and the amino acid on the second tRNA; initiator tRNA moves to E-site, second tRNA moves to P-site



Elongation of Translation

The Ribosome

To the tune of "*America the Beautiful*"

O beautiful with R-N-A
That makes the peptide bonds
You hold t-RNA so it
Can pair up with co-dons

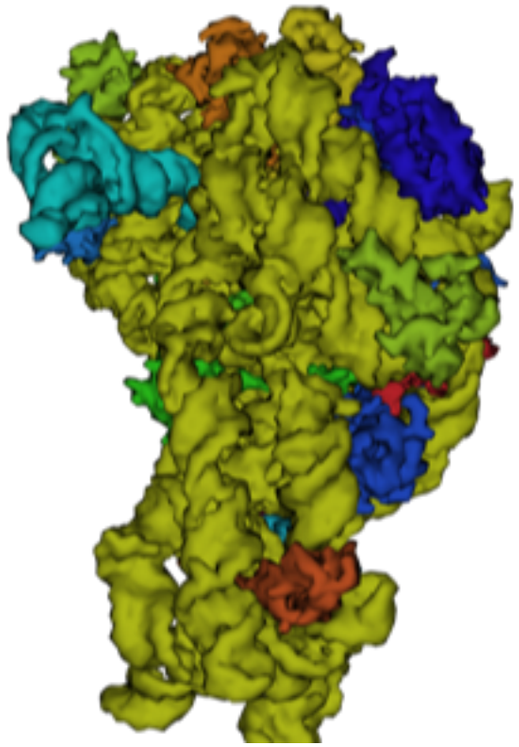
The Ribosome! The Ribosome!
Translate m-RNA
Initiate and translocate
From start to U-G-A

Recorded by Tim Karplus
Lyrics by Kevin Ahern

bacterial mRNAs carry a short purine-rich sequence known as the Shine-Dalgarno site upstream of the AUG start codon, as shown in the figure below. This sequence is recognized and bound by a complementary sequence in the 16S rRNA that is part of the small ribosomal subunit as shown above. Because the Shine-Dalgarno site serves to recruit and bind the ribosome, it is also referred to as the Ribosome Binding Site or RBS.

A variation of this process of ribosome assembly operates in eukaryotic cells. We already know that in eukaryotic cells, processed mRNAs are sent from the nucleus to the cytoplasm.

Interactive 5.3



The Ribosome

The small and large subunits of ribosomes, each composed of characteristic rRNAs and proteins are found in the cytoplasm and assemble on mRNAs to form complete ribosomes that carry out translation.

Protein synthesis in eukaryotes starts with the binding of the small subunit of the ribosome to the 5' end of the mRNA. The assembly of the translation machinery begins with the binding of the small ribosomal subunit to the 7-methyl guanosine cap on the 5'end of an mRNA. Meanwhile, the initiator tRNA pairs with the start codon. (Recall that the start codon is AUG, and codes for methionine. The initiator tRNA carries the amino acid methionine). The large subunit of the ribosome then joins the complex, which is now ready to start protein synthesis.

Ribosomes have two sites for binding charged tRNAs, each of which is positioned to make two adjacent codons on the mRNA available for binding by tRNAs. The initiation codon occupies the first of these ribosomal sites, the P-site. The anticodon complementary to this is on the initiator tRNA, which brings in the first amino acid of the protein. This initial phase of translation is called initiation and requires the help of protein factors called initiation factors.

The second codon of the mRNA is positioned adjacent to the second site on the ribosome, the A site. This is where the tRNA carrying the amino acid specified by the second codon binds. The binding of aminoacyl tRNA to the A-site is mediated by proteins called elongation factors and requires the input of energy. Once the appropriate charged tRNAS have "docked" on the codons by base-pairing between the anticodon on the tRNA and the codon on the mRNA, the ribosome joins the amino acids carried by the two tRNAs by making a peptide bond (see figure at right).

Interestingly, the formation of the peptide bond is catalyzed by a catalytic RNA (the 23S rRNA in prokaryotes) rather than by a protein enzyme.

This and

Central Dogma Zen- Part 2

To the tune of "*Those Were The Days*"

An organism's cellular construction
With blueprints for the things they have to do
Requires converting DNA instructions
To ribopolymers, oh yes it's true

Because they've been bestowed
With a genetic code
The RNAs provide the cell with means
To link amino A's
In most directed ways
Inside the protein-making cell machines

If "coli" cells don't have galactosidase
And lactose should appear inside its food
The lac repressor leaves the operator
'Cause otherwise metabolism's screwed

Polymerase unwinds
The DNAs it binds
Adjacent to the start site where it docks
Unravels A's and T's
With such amazing ease
At the promoter's little TATA box

The process moves along without much trouble
While making RNA inside the cell
It all occurs inside transcription bubbles
Where bases get linked anti-parallel

mRNA then roams
To find some ribosomes
Subunits large and small bind near the end
The A-U-G's in place
Inside the P site space
Initiation you can comprehend

The mechanism shifts to elongation
Proceeding by three bases at a stretch
A GTP's required for translocation
Advancing 5 to 3 the whole complex

The process moves anon
Until a stop codon
Arrives and causes movement to suspend
Translation has to cease
A peptide gets released
And we have reached the central dogma's end

Recorded by Tim Karplus
Lyrics by Kevin Ahern

subsequent steps in the synthesis of the polypeptide are called the elongation phase of translation. Once the first two amino acids are linked , the first tRNA dissociates, and moves out of the P-site and into the E, or Exit site. The second tRNA then moves into the P-site, vacating the A-site for the tRNA corresponding to the next codon.

The process repeats till the stop codon is in the A-site. At this point, a release factor binds at the A-site, adds a water molecule to the polypeptide at the P-site, and releases the completed polypeptide from the ribosome, which itself, then dissociates into subunits.

As described in Chapter 3, polypeptides made in this way are then folded into their three dimensional shapes, post-translationally modified and delivered to the appropriate cellular compartments to carry out their functions.

Jump to Chapter

1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10 / 11 / 12