DNA Replication + DNA Expression + DNA Regulation

Selena Yang

July 2023

Contents

1 Introduction **3**

2 DNA **3**
2.1 History . . . 3
2.1.1 Thomas Hunt Morgan's Fly experiments . . . 3
2.1.2 **Griffith/Avery experiment** . . . 3
2.1.3 Hershey and Chase experiment . . . 4
2.1.4 Chargaff's rules . . . 5
2.1.5 Watson and Crick . . . 5
2.2 Structure . . . 6

3 Replication **7**
3.1 Theory . . . 7
3.1.1 Messelsohn and Stahl Experiment . . . 8
3.2 Process . . . 9
3.3 Repair . . . 10
3.3.1 Polymerase Exonuclease Repair . . . 11
3.3.2 Base excision repair . . . 11
3.3.3 Nucleotide excision repair . . . 11
3.3.4 Direct Reversal . . . 12
3.3.5 Non Homologous End Joining . . . 12
3.3.6 Homologous End Joining . . . 12

4 Expression **13**
4.1 Transcription . . . 13
4.1.1 initiation . . . 13
4.1.2 elongation . . . 13
4.1.3 termination . . . 14
4.2 Post transcriptional modification . . . 14
4.3 Translation . . . 15
4.3.1 initiation . . . 15
4.3.2 elongation . . . 15
4.3.3 termination . . . 16
4.4 Protein folding + migration . . . 16

5 Regulation **17**
5.1 Prokaryotic Operons 17
5.1.1 Repressible (*trp*) operon 18
5.1.2 inducible (*lac*) operon 18
5.2 Eukaryotic Regulation 19
5.2.1 Epigenetics 19
5.2.2 RNAi 20
5.2.3 Protein Degradation 21

6 Conclusion **22**

1 Introduction

We'll be exploring everything about DNA in this handout, starting from its history to replication, expression, and regulation. These topics are very important to the USABO (being covered in both cell biology and genetics), but we might go on several tangents that are interesting but less relevant for the USABO. I will mention the (ir)relevance of a section to the USABO, but generally most of these topics are very commonly tested.

2 DNA

DNA (deoxyribonucleic acid) is how our cells store their genetic information, providing instructions for how every part of the cell should be produced and used.

2.1 History

To understand how DNA is structured, we should first retrace how it was discovered by following some major experiments.

2.1.1 Thomas Hunt Morgan's Fly experiments

Recall from the Genetics handout that Morgan performed various experiments with the fruit fly *Drosophilia melanogaster*. Through these experiments, Morgan was able to show that **chromoses** contained the hereditary material. The question was which of its components (proteins or DNA) were the source of genes. Originally, most biologists thought that proteins had to be the hereditary material, due to its vast diversity, however several experiments came out that showed DNA was in fact the hereditary material.

2.1.2 Griffith/Avery experiment

Griffith's experiment focused on the idea of **transformation**, which is that bacteria can take up genes from their environment. In this experiment, Griffth used two strains of *Streptococcus pneumoniae*: a deadly *S* (smooth) strain and a harmless *R* (rough) strain. When he injected a mixture of heat-killed S cells and alive R cells in a mouse, the mouse died! This showed that the R cells picked up the pathogenic DNA that S cells released upon dying, which provided the first evidence for bacterial transformation.

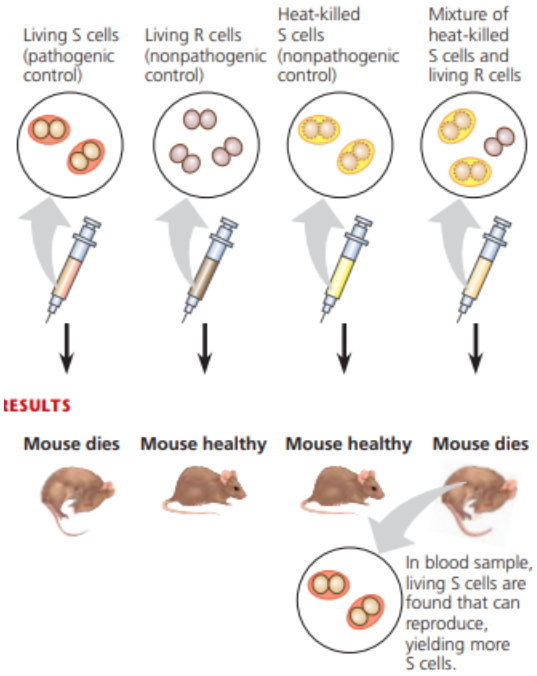

Figure 1: Griffith's experiment (Source: Campbell's 9th Edition)

However, we are interested in **Avery's experiment**, which took this idea one step further. Avery used the mixture of heat-killed S cell and alive R cells again, but this time added an enzyme that would destroy proteins to one group and an enzyme that would destroy DNA to the other group. The mice injected with the DNA-destroyed batch lived while the mice injected with the protein-destroyed batch died! This shows that DNA contained the deadly genes (since the mice only stayed alive when the DNA was destroyed).

2.1.3 Hershey and Chase experiment

The **Hershey and Chase experiment** relied on **transduction**, which is when a *bacteriophage* (a virus for bacteria) injects its genetic material into a bacterium. For this experiment, they used a **T2** phage to infect *E. coli*. To see what the genetic material in viruses was made of, they used special radioisotopes to label the protein and DNA. **P32** was used to tag DNA, while **S35** was used to tag protein (I remember it as P is the 16^{th} letter in the alphabet so $16 \times 2 = 32$. Then, S is 3 letters after P in the alphabet so $32 + 3 = 35$). They saw that the radioactive P32 entered the bacteria, but S35 did not, showing that DNA was the genetic material of bacteriophages.

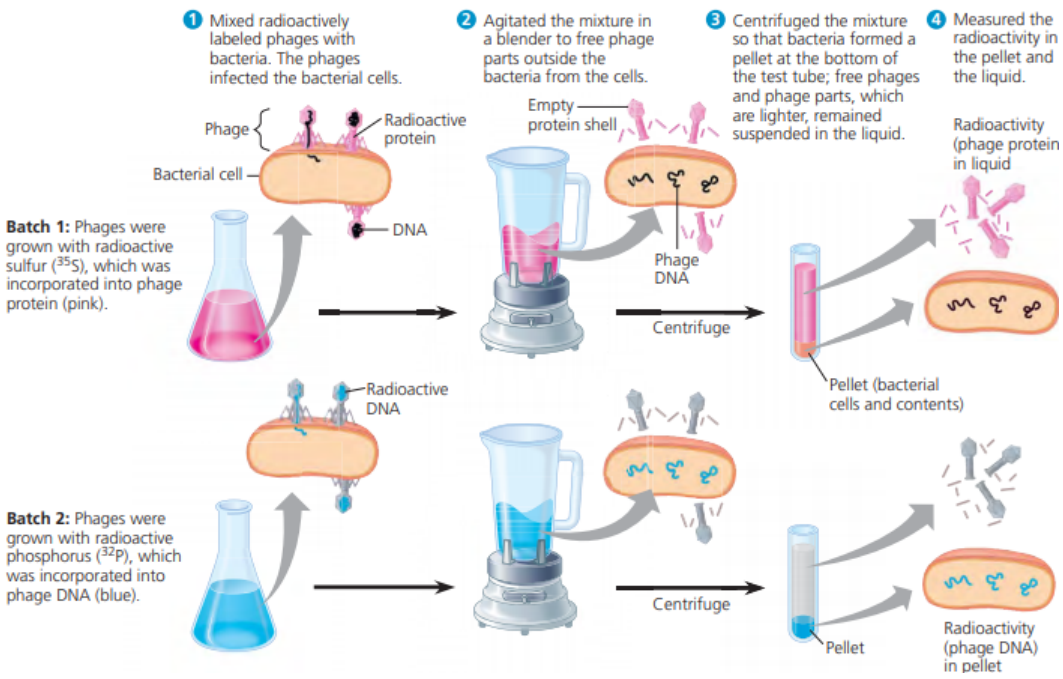

Figure 2: The Hershey and Chase experiment (Source: Campbell's 9th Edition)

2.1.4 Chargaff's rules

By this point it was clear that DNA was the genetic material and it was made of special nitrogenous bases. After studying the base composition of many species, **Chargaff** came up with 2 important rules:

1. The base composition of DNA varied between species
2. A and T always made up the same percent of bases in DNA. Similarly, C and G always made up the same percent of bases in DNA.

2.1.5 Watson and Crick

By this point, biologists knew that DNA was the genetic material and they knew what it was made of. However, no one was able to explain how it was structured or how it could explain the vast differences across all the species. The american **James Watson** and british **Francis Crick** were working together to uncover DNA's structure. When they visited the lab of **Maurice Wilkins**, they saw the work of another researcher, **Rosalind Franklin**, who was using **X-ray crystallography** to image the structure of DNA. Upon seeing her work, Watson and Crick realized that DNA had a *helical* structure, allowing them to piece together the structure of DNA, which we will discuss next. Unfortunately, Franklin had no idea her work was being used and never received recognition while she was alive. In addition, X-ray crystallography exposed her to large amounts of radiation, causing her to die young :(

2.2 Structure

DNA is a polymer made up of **nucleotides**, which consist of a phosphate, 5-carbon sugar, and a nitrogenous base. The 5-carbon sugar has some important features you should know:

- The **phosphate** group is bound to the 5th carbon, so we call this the **5'** (5 prime) end. The negative charge on phosphates makes DNA *negatively charged*.
- An OH **(hydroxyl)** group is bound to the 3rd carbon, so we call this the **3'** (3 prime) end.
- Sometimes the sugar has another OH group on the 2nd carbon. If so, the sugar is called **ribose**. However, sometimes we remove the OH (de-oxy), resulting in the sugar **deoxyribose**.
- The nitrogen group is bound to the 1st carbon

To hold these monomers together, two important types of bonds are found in DNA. One of them occurs when an O from the 5' phosphate and the 3' OH combine. This releases water (**dehydration synthesis**), and forms a **phosphodiester bond** in the process. This bond links the phosphate to the sugar, and thus creates the **sugar-phosphate backbone** of DNA.

Phosphodiester diester bonds in the sugar-phosphate backbone hold a single strand of DNA together. From this backbone jut out the nitrogen bases, which can form **hydrogen bonds** that hold two *complementary* strands together. Due to the way the nitrogen bases come together, complementary strands are **antiparallel**, or go in opposite directions. Therefore, if one strand is 5' - 3' (5 prime to 3 prime), the other strand will be 3' - 5'

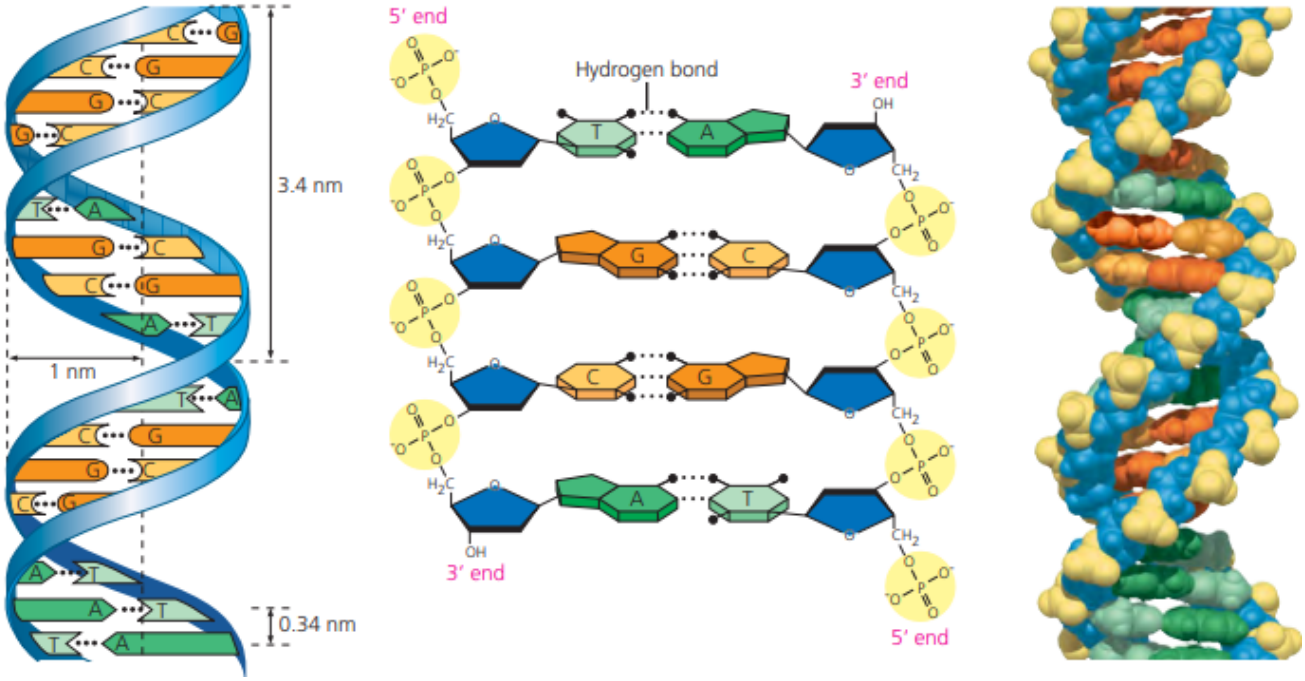

Figure 3: The **double helix** structure of DNA. The distance between two adjacent bases is .34nm and is known as **rise**. Roughly every 10 nucleotides, the helix makes a full turn. This distance ($.34 \times 10 = 3.4$nm) is known as **pitch**. (Source: Campbell's 9th Edition)

The nitrogenous bases come in two types:

- **Pyramidine:** These bases have a single ring and include **C (cytosine)** and **T (thymine)**. In place of T, RNA will have **U (uracil)**.
- **Purine:** These bases have 2 rings and include **G (guanine)** and **A (adenine)**.

Since pyramidines are bigger than purines, we must always have a purine bonded to a pyramidine so that DNA has a constant width (minimizing strain).
A always bonds with **T**, forming **2** hydrogen bonds.
G always bonds with **C**, forming **3** hydrogen bonds. This is why Chargaff's 2nd law is true. The diversity in life (Chargaff's 1st Law) comes from the *order* of nucleotides in DNA.

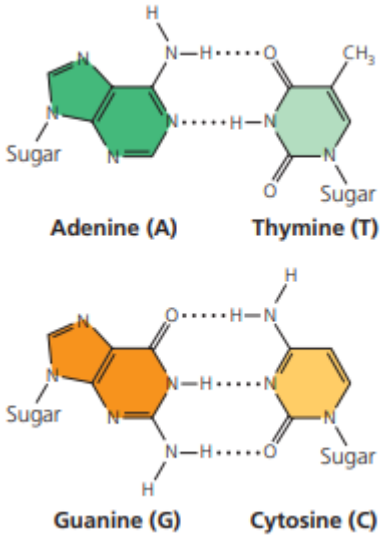

Figure 4: A goes with T and C goes with G
(Source: Campbell's 9th Edition)

3 Replication

DNA Replication, as the name implies, is the process of duplicating DNA. When a cell divides during mitosis, it splits all of its components in half, including its genetic material. In order to preserve the quantity of genetic material in each cell, they must replicate their DNA.

3.1 Theory

Early experiments to determine the mechanism of DNA replication had settled on three hypotheses:

1. **The conservative model:** One copy of DNA will consist entirely of original material, while the other copy will consist entirely of new material.
2. **The semiconservative model:** Both copies of DNA will contain one original strand and one new strand of DNA. This model was proven to be the correct one by the **Messelsohn and Stahl experiment**.
3. **The dispersive model:** All strands of DNA have original pieces and new pieces of DNA mixed together.

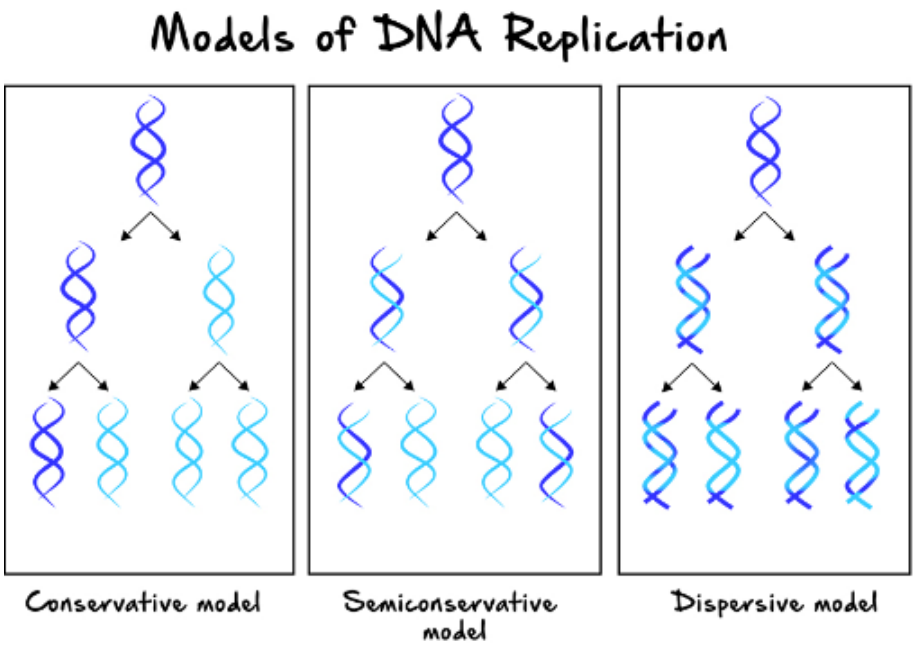

Figure 5: The three proposed mechanisms of DNA replication. (Source: MedSchoolCoach)

3.1.1 Messelsohn and Stahl Experiment

Bacteria were grown in a special growth medium with **N-15**, a heavy nitrogen isotope (nitrogen is a component of DNA), so all the original DNA was heavier than regular DNA. Then, the bacteria were transferred to the regular growth medium with "normal" N-14 nitrogen and collected after one generation (one round of DNA replication) and analyzed. To analyze the DNA, they used equilibrium sedimentation. Basically, the heavier the molecule, the lower it would be on the column.

After the first generation, conservative would give one light band and one heavy band, while semiconservative and dispersive would give one medium band. After running the experiment, they got one medium band, meaning that replication was either semiconservative or dispersive. To differentiate between the two models, they grew the bacteria for another generation (two rounds of DNA replication) and found that the results matched the semiconservative model.

> **Exercise 1.1:**
> Draw out the expected results for each replication model after the 2nd generation.
>
> *Hint*: use the bottom row of Figure 5. There are three possible locations for the bands in equilibrium sedimentation: 1) at the top, made only of N-14 DNA 2) in the middle, a band made of a mix of N-15 and N-14 DNA, 3) at the bottom, a band made of only N-15 DNA.

Solution: Conservative would have a thin band at the very bottom and a thicker band at the very top. Semiconservative would have two equally sized bands, one in the middle and one at the top. Dispersive would have one band at the very top.

Example 3.1: (USABO Opens 2018)

5. Which scientist(s) is incorrectly matched with his/her/their discoveries?

A. Frederick Griffith – discovery of transformation through bacteriophages.
B. Hershey and Chase – discovery of DNA as the hereditary material of cells.
C. Chargaff – discovery that relative amounts of nucleotide bases varies among organisms.
D. Watson and Crick – discovery of the double-helix structure of DNA.
E. Meselson and Stahl – discovery of the semi-conservative replication of DNA.

Solution: The **answer is A**. Transduction uses bacteriophages, while transformation takes in DNA from the environment.

3.2 Process

Before we begin exploring the process of DNA replication, it is good to familiarize yourself with the enzymes involved:

- **DNA Polymerase III:** Synthesizes DNA from 5' to 3'. The reason polymerases go from 5' to 3' is that the nucleotides they use are dNTPs (deoxyribose nucleotide tri-phosphate). The energy needed to continue the chain comes from breaking the phosphate bond, releasing a pyrophosphate. Then, the remaining phosphate binds to the 3' end. This method makes DNA repair easier as the energy we rely on comes from the new nucleotide, not the existing strand.

  If a ddNTP is used (**dideoxyribose** means that the 3' OH is also used), then the phosphate has no where to be added to and the polymerase stops.

- **Primase:** Synthesizes RNA primer

- **DNA Polymerase I:** Replaces RNA primers with DNA.

  – DNA Polymerase I, II, and III are found in *bacteria*. Eukaryotes on the other hand have 5 DNA Polymerases ($\alpha, \beta, \gamma, \epsilon, \delta$) that you do not need to know.

- **Helicase:** Unwinds the DNA.

- **Topoisomerase:** Relieves stress from supercoiling by unwinding DNA.

- **Single Strand Binding Proteins:** Prevents strands from going back togethe

- **Ligase:** Joins together adjacent segments of DNA.

Let's start from the beginning. As mentioned before, DNA Helicase needs to unravel the strands. However, like trying to untwist a very tightly braided rope, some parts will become supercoiled. DNA relies on topoisomerases, enzymes that basically cleave the DNA to release the supercoil. After helicase separates the two strands, single strand binding proteins bind to the separated strands to prevent them from going back together. Since the two separated strands run in different directions, they need to be replicated differently.

DNA Polymerase can only synthesize from 5' to 3', meaning that it reads from 3' to 5' (since the parent strand and the new strand being synthesized are antiparallel). Thus, the strand with the 5' end closest to the helicase is the **leading strand** because polymerase can replicate continuously towards the helicase as it unwraps the DNA. However, the other **lagging strand**, requires discontinuous replication away from the helicase.

The lagging strand considerably more complex than the leading strand. DNA Polymerase needs a preexisting 3' OH in order to start elongating. However, there is no preexisting 3' OH because the replication is discontinuous. Therefore, another enzynme, Primase, comes in and puts down an RNA primer, giving Polymerase the 3' OH it needs to replicate. Each discontinuous strand is called an **Okazaki fragment**.

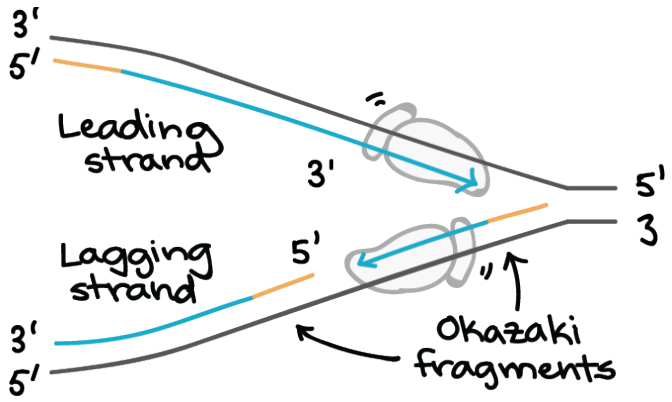

Figure 6: Another nice diagram of the lagging strand. (Source: Khan Academy)

However, it would be very bad to keep the RNA primers inside the newly replicated DNA. Thus, DNA Polymerase I goes in afterwards and replaces the RNA primer with DNA. The discontinous Okazaki fragments are then connected by DNA Ligase. In eukaryotes, the cell is able to recognize which strand is newly synthesized and which strand is the parent by the nicks in the DNA not yet sealed up by Ligase. This is particularly important when there is a mismatched base that needs repair.

3.3 Repair

Why do cells need DNA Repair? Sometimes, DNA Polymerase makes a mistake. It's only human after all. Other times, mutagens such as UV radiation, superoxide radicals, and benzene (found in sunscreen which is ironically designed to prevent DNA damage) may damage the DNA.

> **Exercise 1.2:** There are many mechanisms for repairing DNA damage. Why aren't there are as many for transcription or translation mistakes?

Solution: Because DNA damage is passed on to future cells, but incorrect mRNA and proteins do not ultimately affect the fate of the cell (unless we are talking about prions, which are their own headache), DNA needs to be preserved much more accurately.

DNA Repair is a fascinating topic, but it is not very commonly tested in the USABO, much to the author's dismay. If you are studying for the USABO, it is the most important to know

about the function of these processes instead of the mechanisms. You can always search up these mechanisms on Wikipedia to learn more.

3.3.1 Polymerase Exonuclease Repair

DNA Polymerase has its own repair system. When replicating, it may detect that it added the wrong base. Polymerase will then move back and use its special 3' to 5' exonuclease activity to remove the base and add the correct one. This exonuclease activity is the main reason why polymerase needs to synthesize from 5' to 3'. If it was the other way, after cleaving the incorrect base, there would be no way for DNA Polymerase to keep on replicating (since there are no phopshate bonds that can be broken for energy).

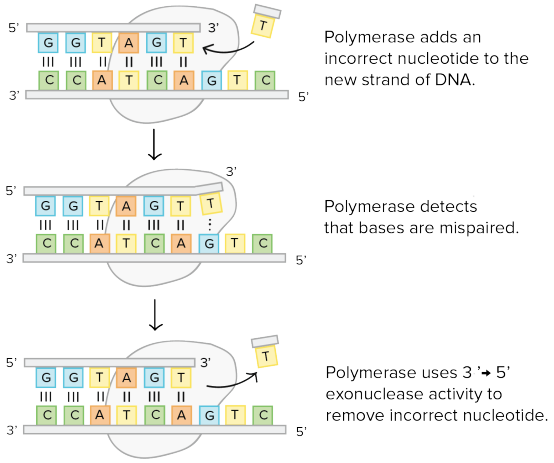

Figure 7: DNA Polymerase catalyzing repair. (Source: KhanAcademy)

3.3.2 Base excision repair

Base excision repair (BER) is used for non bulky lesions, or DNA damage that generally doesn't affect the structure of the helix. For example, chemical modification of one nucleotide (e.g. deamination) would be a non bulky lesion. This basic rule is sufficient for most of the questions about BER.

As a brief overview, BER functions by having glycosylases cleave out the base, forming an AP (apurinic/apyrimidinic) site (thankfully not associated with CollegeBoard), that is then repaired by Polymerases.

3.3.3 Nucleotide excision repair

Nucleotide excision repair (NER) repairs "bulky lesions", or lesions that alter the helix structure. You can often go by vibe here: mutations that "feel" structural are generally repaired by NER. Most notably, thymine dimers cause by UV radiation is repaired by NER.
Contrary to BER, NER excises out the entire nucleotide (or segment of nucleotides) and is generally more complex.

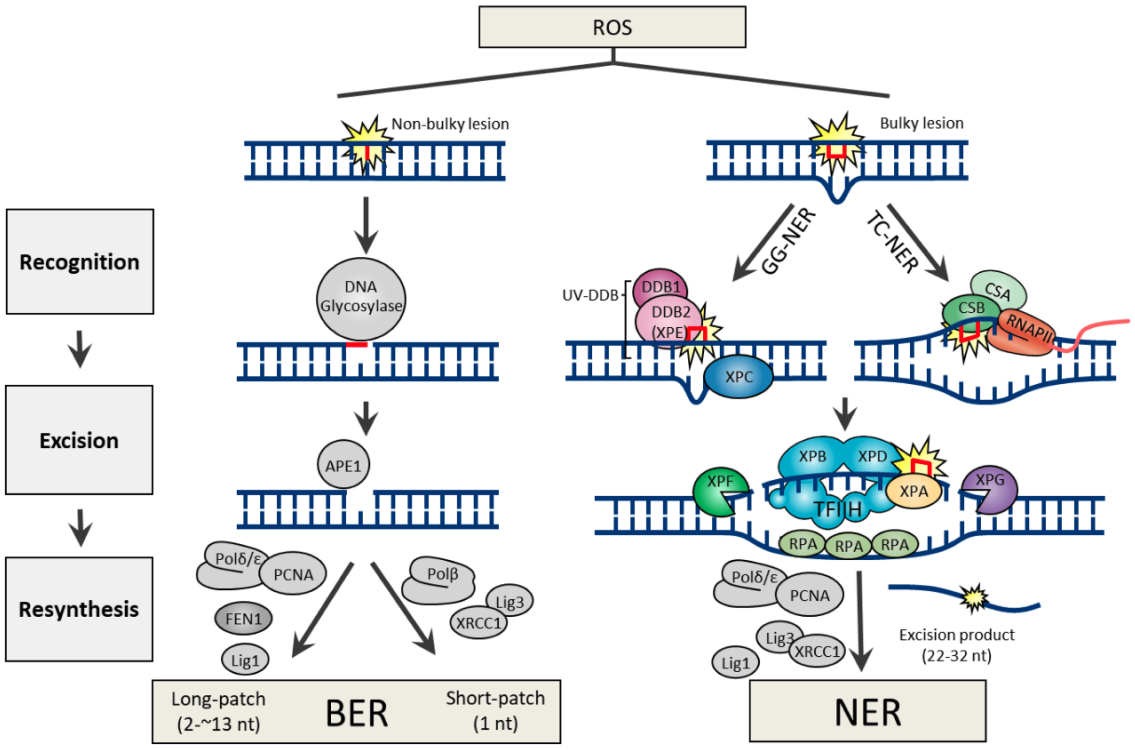

Figure 8: Base Excision vs Nucleotide Excision Repair. (Source: Wikipedia)

3.3.4 Direct Reversal

Direct Reversal is an evolutionarily rudimentary version of Nucleotide Excision Repair. The enzymes involved in Direct Reversal cannot be used repeatedly (ie, after one use they must be degraded), which makes the process costly.
However, it is still important in repair after exposure to sunlight, and there are some cancers caused partially by the malfunction of the Direct Reversal System.

3.3.5 Non Homologous End Joining

Non Homologous End Joining (NHEJ) repairs double strand breaks. While very inaccurate, it is used most of the time because it is not as costly in terms of energy.
If a double strand break occurs, NHEJ does the obvious solution: it just anneals the two ends back together. The downside to this is that some nucleotides may have been lost, leading to a potentially harmful deletion.

3.3.6 Homologous End Joining

Homologous End Joining (HEJ) also repairs double strand breaks, but only after the DNA has replicated. Similar to the process of crossing over, HEJ uses the homologous strand as a template to repair the broken strand in, forming a **Holliday Junction**. The details are not important, but Albert's *Molecular Biology of the Cell* has some amazing diagrams if you're interested.
Some CRISPR editing techniques take advantage of this to insert a mutant protein into organisms. In addition, the famous tumor suppressor genes BRCA1/2, which are most commonly mutated in breast cancer, both play a role in Homologous End Joining.

4 Expression

DNA replication is all about how the genetic material propagates through generations of cells. However, we have still not learned anything about how DNA actually commands the cell. That is done through proteins, which are made through the processes of **transcription** (DNA → RNA) and **translation** (RNA → proteins).

4.1 Transcription

Transcription converts DNA to RNA through the use of the enzyme **RNA polymerase** (eukaryotes have more than 1 polymerase so they use **RNA polymerase II**), which synthesizes RNA using the complementary DNA strand. Since we are still using the same "language" of nucleotides, we are simply "transcribing" from one text (DNA) to another (RNA).

4.1.1 initiation

Initiation of transcription requires a wide array of proteins, including enhancers, transcription factors, mediators, but most importantly, transcription requires the binding of RNA Polymerase. The sequence of DNA where RNA Polymerase binds is called the **promoter**. Various transcription factors also bind to the **TATA box** (or **Pribnow box** in prokaryotes), which is just before the promoter.

RNA Polymerase also undergoes a complex "scrunching" of the DNA where it synthesizes 10-25 nucleotides then releases the DNA to start over. Other proteins are required to break this cycle and allow RNA Polymerase to elongate, but the details are unnecessary (check out *Molecular Biology of the Cell*, one of the author's favorite books, if you are interested.)

4.1.2 elongation

Elongation of RNA Polymerase is quite simple. It traverses down the DNA strand, unwinding it and synthesizing mRNA. The DNA strand which is complementary to the mRNA (and thus used for synthesis) is called **template DNA**, while the other strand is called **coding DNA**, though unlike the author, it does not need to spend its time debugging C++ code.

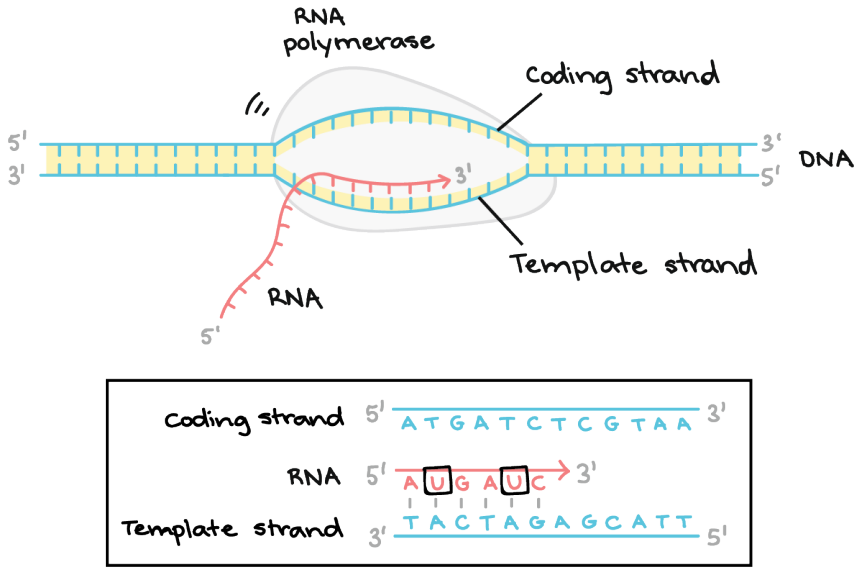

Figure 9: DNA Transcription. (Source: KhanAcademy)

4.1.3 termination

Termination in prokaryotes and eukaryotes are actually quite different.
In prokaryotes, there are two methods of termination: Rho-dependent and Rho-independent, one of which depends on the protein Rho (can you guess which one?) The other requires **attenuation**, which is the formation of a hairpin-like structure by the mRNA, leading to stalling and dissociation from the RNA Polymerase. (Do not confuse attenuation with the *other* attenuation relating to vaccines.)

In eukaryotes, termination is not very well understood (both by the scientific community and by the author). Basically, once the RNA Polymerase transcribes a certain sequence, it recruits proteins that cleave the mRNA off. However, due to the uncontrollable nature of the RNA Polymerase, it keeps on transcribing. The proteins need to chase down the RNA Polymerase and stop it, which makes for an amusing mental image.

4.2 Post transcriptional modification

Prokaryotic mRNA typically does not undergo any modification. Thus, some ribosomes begin translating the mRNA before it's done being transcribed. However, due to the contrasting environments of the nucleus and the cytoplasm, post transcriptional modification is necessary in eukaryotes.

Eukaryotes also have **introns**, long segments of noncoding DNA in between **exons**, which code for genes. In order to get rid of these introns, a RNA-protein complex called the **spliceosome** pinches out the two ends of the introns and ligates the exons in a process called **splicing**.

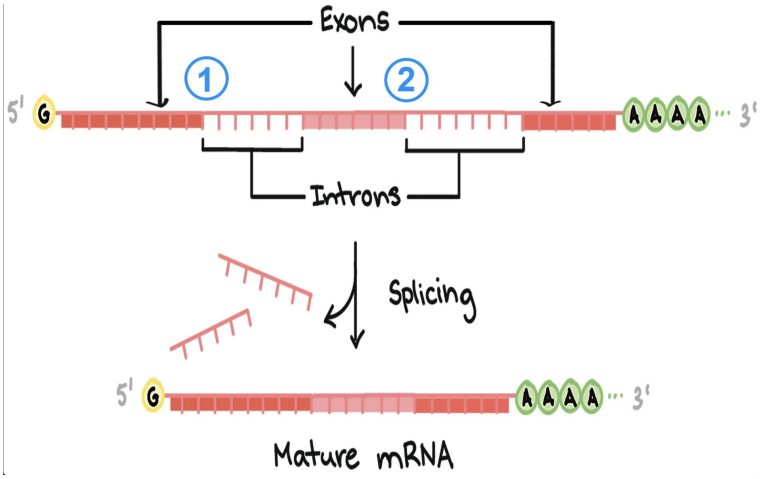

Figure 10: The three post-transcriptional modifications present in eukaryotes. (Source: KhanAcademy)

In addition to splicing, eukaryotic mRNA also receives two other modifications– a 5' modified guanine cap, and a 3' poly-A tail. Both of these modifications protect the mRNA against degradation once it is transported into the cytoplasm. Once a mRNA has gone through these three modifications, it is considered "mature" and ready to be exported through the nuclear pore complexes.

4.3 Translation

Before we start talking about translation, it is important to know that not all DNA codes for proteins. In fact, only three percent of the genome actually codes for protein!
Some DNA codes for RNAs that don't become protein, and they have a wide range of functions. For example, the RNA can be in the spliceosome (snRNA), it can regulate gene expression (miRNA, siRNA, and lncRNA), and it can even repress transposons (piRNA)! However, the USABO mainly focuses on mRNAs that end up coding for proteins, and we will too. However, I will mention some of the regulatory RNAs later.

4.3.1 initiation

Translation starts when the smaller portion of the ribosome binds to the mRNA. There are various sequences that make binding more favorable, including the **Shine Dalgarno sequence** in prokaryotes and the **Kozak sequence** in eukaryotes.
Then, the initiator tRNA, which is **methionine (AUG)** (**formyl-methionine** in prokaryotes), binds to the mRNA. This attracts the large ribosomal subunit, and translation starts.

4.3.2 elongation

The ribosome has three sites: E, P, and A.

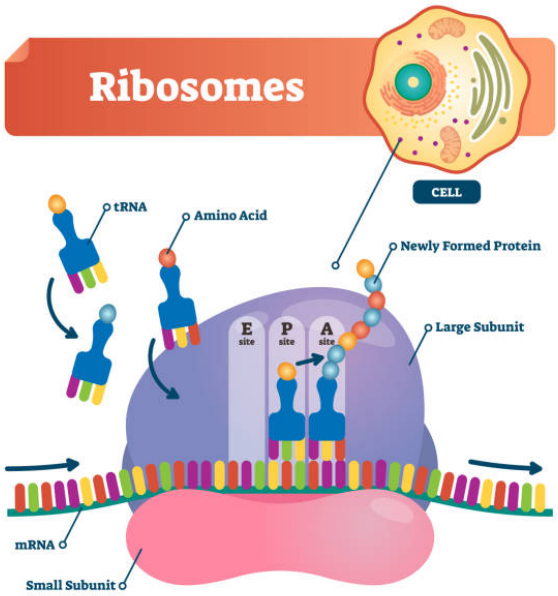

Figure 11: A diagram of the ribosome. (Source: iStock)

E is the entry site, where the tRNAs (which bring the amino acids to the ribosome) arrive. The P site is where the tRNA transfers its amino acid to the growing polypeptide. The A site is the exit. When a new tRNA arrives, it enters at the E site and binds to the complementary mRNA codon (three bases). The ribosome then shifts over so the tRNA is positioned at the P site, and the ribosomal RNAs catalyze the tranfer of the amino acid from the tRNA to the polypeptide chain. The ribosome then shifts over again to allow the tRNA to exit. This process continues until a stop codon is read.

4.3.3 termination

When a stop codon **(UGA, UAA, UAG)** is read, instead of a tRNA bringing over an amino acid, termination factors arrive at the E site and separate the ribosome from the mRNA, terminating translation.

4.4 Protein folding + migration

Typically, the USABO doesn't test too much about the specifics of protein folding and migration. However, it's good to know the basics, as they have come up before. If you are interested in the details, please read *Molecular Biology of the Cell*.

Protein migration begins even before the protein is completely translated. Some mRNAs have a **signal sequence (signal peptide)** at their N-terminus (the 5' end), which, when translated, binds to **SRP (signal receptor protein)** on the ER and causes the ribosome to move to the Rough ER, becoming a *bound* instead of *free* ribosome. Thus, the polypeptide is translated and folds in the Rough ER, and eventually moves to the Golgi for packaging and transport to the right place. The Golgi sorts the proteins using their signal sequence and eventually cleaves it off.

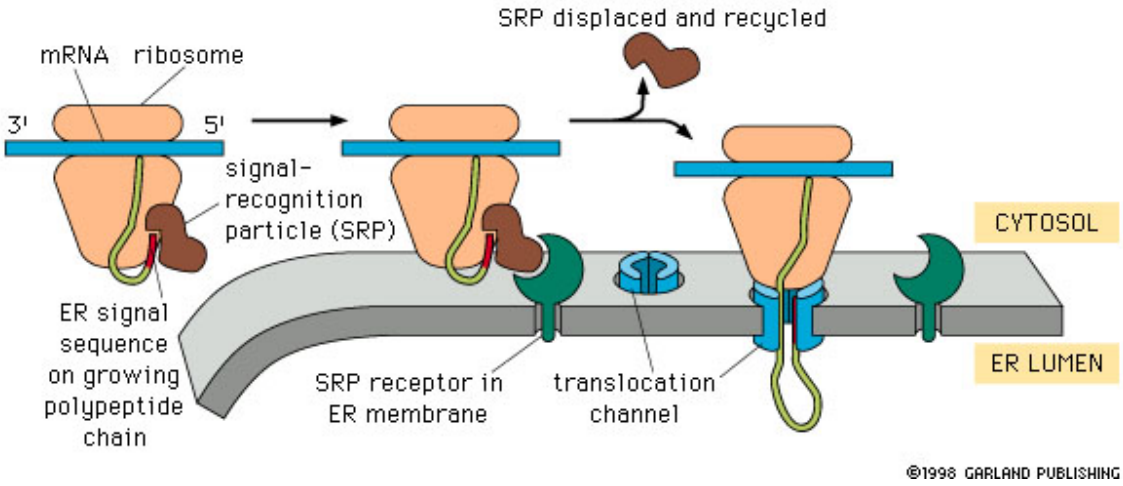

Figure 12: Translocation of an mRNA with a signal sequence into the ER. (Source: Garland Publishing)

Quick Quiz: Identify all of the proteins that would be translated by a bound ribosome: 1. ATP Synthase 2. Antibodies 3. Hexokinase

Solution: 1 and 2. ATP Synthase is transported into the mitochondria, Antibodies are either secreted or on the membrane, so they both have signal sequences and are both translated by a bound ribosome. Hexokinase is an enzyme in glycolysis, which takes place in the cytoplasm, so it would be translated by a free ribosome. A good rule of thumb is that unless the protein functions in the cytoplasm, it's probably synthesized by bound ribosomes in the ER.

Protein folding is generally considered "spontaneous", however, some proteins, especially in times of stress (e.g. altered temperature/pH), may need a special environment to fold properly. **Chaperonins** are special proteins resembling a barrel that provide a specialized environment for proteins. Some chaperonins, like hsp70, are upregulated in response to stress and also prevent aggregation of misfolded proteins until the conditions return to normal.

5 Regulation

5.1 Prokaryotic Operons

Prokaryotes generally have a smaller size, giving them less space for their genomic DNA. Thus, their genome size is generally very tightly regulated. One way in which prokaryotes conserve space is by putting very similar genes together in an **operon**.

These genes are regulated by an operator, which is located a few bases after the promoter. When a repressor binds to the operator, it prevents RNA Polymerase from moving forward and transcribing the genes. Other proteins can interact with the repressor and allow it or block it from binding, thus regulating transcription. Operons are generally regulated in two manners: repressible and inducible.

5.1.1 Repressible (*trp*) operon

Repressible operons are generally "on", unless something turns them off. For example, the *trp* operon, which has genes necessary for synthesizing tryptophan, turns off in the prescence of tryptophan (which is a **corepressor**). Tryptophan works by activating the repressor, so it binds and inhibits transcription. This is a great example of negative feedback. If you make too much of something, you will want to stop making it. Thus, most genes associated with anabolic processes are in repressible operons.

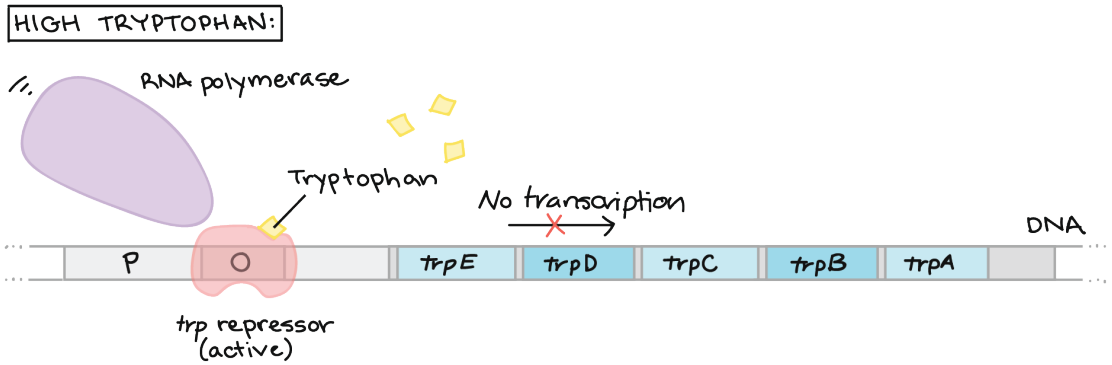

Figure 13: The *trp* operon. (Source: Khan Academy)

5.1.2 inducible (*lac*) operon

On the contrary, **inducible operons** are generally "off", unless something turns them on. The classic example is the *lac* operon, which codes for proteins that allow for lactose metabolism, turns on when allolactose (the **coinducer** and an isoform of lactose) is present. Allolactose works by inactivating the repressor, so it is unable to bind. This makes sense because you don't want lactose metabolism genes to be expressed when you don't have lactose.

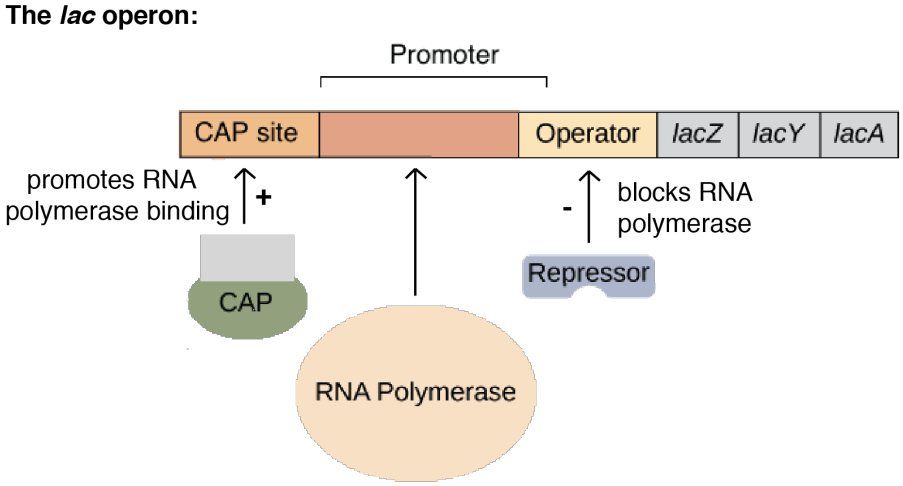

Figure 14: The *lac* operon. (Source: Khan Academy)

There is another detail I need to mention about the *lac* operon. Bacteria very much prefer to metabolize glucose instead of lactose. Thus, if glucose is present, even if allolactose is inactivating the repressor, the operon will still be off. This is because of the **CAP** site.

When glucose is low, a molecule called cyclic AMP (cAMP) has a high concentration. cAMP binds to the CAP protein, which allows the CAP to bind to the CAP site just before the promoter, which increases RNA Polymerase binding. Thus, CAP is known as an **activator**. When glucose levels rise, cAMP is degraded and the CAP goes away, meaning that RNA Polymerase doesn't bind to the promoter as much. This type of regulation is known as **positive gene regulation** since the additional protein activates the promoter instead of repressing it.

Example 5.1: (USABO Opens 2018)

37. Which of the following is incorrect about operons?

A. Operons can be switched off with a repressor.
B. The lac operon is an inducible operon.
C. In the presence of lactose and glucose, the lac operon will not be repressed.
D. CAP is an activator of transcription when not bound to cAMP.
E. A mutation in the operator of the trp operon can lead to overproduction of tryptophan.

Solution: As explained earlier, the **answer is D**.

5.2 Eukaryotic Regulation

Eukaryotic regulation of DNA Expression is extremely complex. Luckily, the USABO doesn't test about it in too much detail.

5.2.1 Epigenetics

Epigenetics is the structural modification of DNA or DNA associated proteins (histones). Epigenetic modification can be passed down from cell to cell, and plays a significant role in genetic inheritance and genome expression.

DNA methylation is the most commonly mentioned modification, and it consists of adding methyl (CH_3) groups to DNA. Generally, DNA methylation *decreases* transcription, though it can also increase it in some cases. The methyl groups cause DNA to be more tightly wound, forming **heterochromatin** and reducing the chance of RNA Polymerase binding. Methylation is also implicated in **imprinting**, which you will learn about later in genetics.

One other interesting thing about methylation is that it is uncommon on cysteines. The reason for this is that methylated cysteines deaminate to thymines, which make them indistinguishable from regular thymines. So, over time, the methylated cysteines turn into thymines.

Other epigenetic modifications include **Histone Acetylation**, which widens the space between histones and forms **euchromatin**, increasing RNA Polymerase binding.

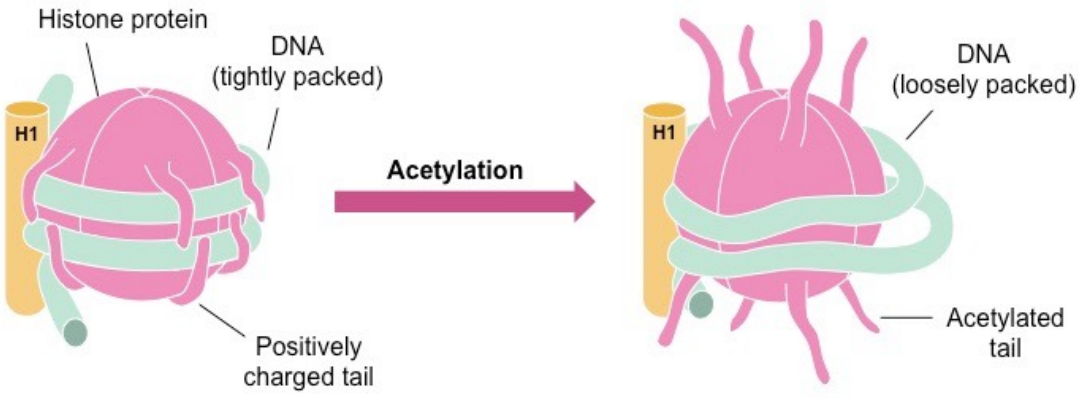

Figure 15: Histone acetylation affecting DNA structure. (Source: BioNinja)

Example 5.2: (USABO Opens 2018)

16. The *Arabidopsis* plant utilizes a range of posttranslational modifications to the genome as a means of regulating gene expression. These include activation of genes by histone acetylation, and methylation of DNA. Which of the following statements does NOT correctly describe the mechanism or impacts of these modifications in the plant cell?

A. Addition of DNA and histone modifications in the cell is performed by protein-based enzymes.
B. Lysine acetylation leads to a tighter association of histone proteins with DNA in the promoter regions.
C. DNA methylation patterns can be propagated to daughter strands after DNA replication.
D. The covalent addition of methyl groups to plant DNA is reversible in the cell.
E. More than one of the above does NOT correctly describe the impacts of these modifications.

Solution: The **answer is B**, since acetylation forms the looser euchromatin. A, C, and D are all correct statements about how DNA methylation works.

5.2.2 RNAi

The relevant parts of RNA interference can be divided into three categories: siRNAs, miRNAs, and lncRNAs. While not commonly tested, RNAi has many lab applications so it is useful to know and also very interesting.

siRNA is a double stranded RNA molecule that silences transcription and cleaves complementary mRNA. It is made in response to dsDNA or RNA hairpins in the cytoplasm, characteristics of viral entry. To silence viral genes, *Dicer* forms siRNA, which then combines with other enzymes like *Argonaute* to form the **RISC** (RNA induced silencing complex). The RISC can then identify complementary mRNA and halt translation or cleave the mRNA. siRNA can also go into the nucleus and induce transcriptional silencing.

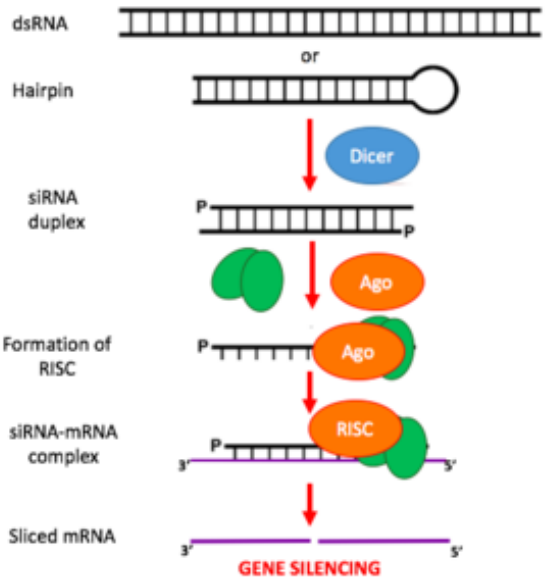

Figure 16: Mechanism of siRNA. (Source: Wikipedia)

miRNA works in a similar mechanism to siRNA, except it is single stranded RNA and it doesn't affect transcription, only translation. Both siRNAs and miRNAs are used for **gene knockdown**, where translation of a gene is greatly reduced (but not fully eliminated, unlike in **gene knockout**)

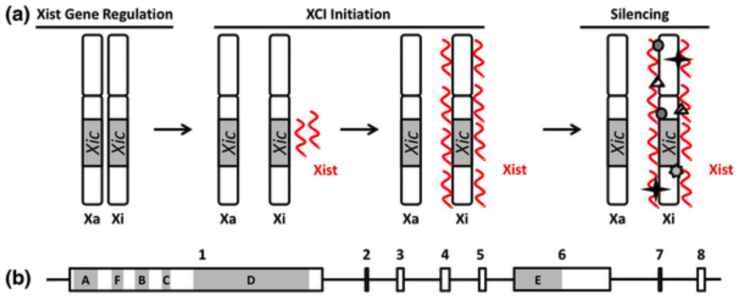

Figure 17: Xist mediated X chromosome inactivation. (Source: Springer)

Finally, **lncRNA**, which stands for long noncoding RNA, is best exemplified by *Xist* (X inactive specific transcript). Since females have two X chromosomes, one needs to become a **Barr body** so there isn't double the amount of X chromosome genes being transcribed (this is called *dosage compensation*). Formation of a Barr body is partially carried out by Xist, which remodels the X chromosome into a very tightly packed heterochromatin structure by crawling around the chromosome.

5.2.3 Protein Degradation

Protein degradation is mediated by a small polypeptide called **ubiquitin**. While having one ubiquitin (monoubiquitylation) is a normal modification, polyubiquitylation leads to degredation by a protease. Ubiquitin ligases add ubiquitin to proteins in many different scenarios, including misfolding and general recycling.

> **Question:** Generally, there are many more regulation processes that target transcription instead of translation. Why?

Solution: Before transcription, there is nothing synthesized so no materials are wasted by preventing transcription. On the contrary, regulating translation requires the degradation of previously made product, which wastes energy.

6 Conclusion

Ultimately, DNA Replication, Expression and Regulation is an extremely complex topic and cannot be covered in one handout or one USABO question. Though this is a guide to the USABO, I would encourage you to explore topics which might not come up, but are fascinating anyway.
Contact me at selenastraceae on Discord if you found a mistake/have questions. Good luck!

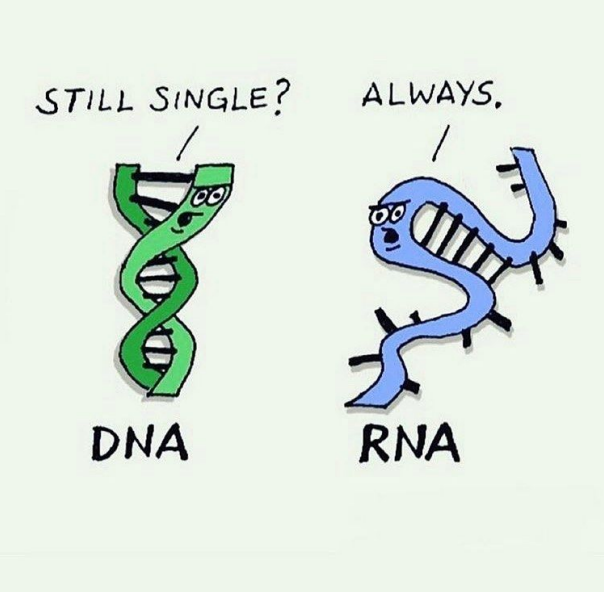

Figure 18: A classic biology meme.