Chapter 8

Molecular Genetics

8.1 Lesson 8.1: DNA and RNA

Lesson Objectives

- Discuss how the work of Griffith, Avery, Hershey, and Chase demonstrated that DNA is the genetic material.
- Define transformation and explain that transformation is the change in genotype and phenotype due to the assimilation of the external DNA by a cell.
- Discuss the findings of Chargaff. Describe the importance of the finding that in DNA, the amount of adenine and thymine were about the same and that the amount of guanine and cytosine were about the same. This finding lead to the base pairing rules.
- Explain Watson and Crick's double helix model of DNA.
- Describe how DNA is replicated.
- Explain the importance of the fact that during DNA replication, each strand serves as a template to make a complementary DNA strand.
- Describe the structure and function of RNA.
- Discuss the role of the three types of RNA: mRNA, rRNA, and tRNA.

Introduction

What tells the first cell of an organism what to do? How does that first cell know to become two cells, then four cells, and so on? Does this cell have instructions? What are those instructions and what do they really do? What happens when those instructions don't work properly? Are the "instructions" the genetic material? Though today it seems completely obvious that Deoxyribonucleic acid, or **DNA**, is the genetic material, this was not always known.

DNA and RNA

Practically everything a cell does, be it a liver cell, a skin cell, or a bone cell, it does because of proteins. It is your proteins that make a bone cell act like a bone cell, a liver cell act like a liver cell, or a skin cell act like a skin cell. In other words, it is the proteins that give an organism its traits. We know that it is your proteins that that make you tall or short, have light or dark skin, or have brown or blue eyes. But what tells those proteins how to act? It is the structure of the protein that determines what it does. And it is the order and type of **amino acids** that determine the structure of the protein. And that order and type of amino acids that make up the protein are determined by your DNA sequence.

The relatively large chromosomes that never leave the nucleus are made of DNA. And, as proteins are made on the ribosomes in the cytoplasm, how does the information encoded in the DNA get to the site of protein synthesis? That's where RNA comes into this three-player act.

DNA → RNA → Protein

That's known as the central dogma of molecular biology. It states that "DNA makes RNA makes protein." This process starts with DNA. And first DNA had to be identified as the genetic material.

The Hereditary Material

For almost 100 years, scientists have known plenty about proteins. They have known that proteins of all different shapes, sizes, and functions exist. For this reason, many scientists believed that proteins were the heredity material. It wasn't until 1928, when Frederick Griffith identified the process of transformation, that individuals started to question this concept. Griffith demonstrated that transformation occurs, but what was the material that caused the transforming process?

Griffith, Avery, Hershey and Chase

Griffith was studying *Streptococcus pneumoniae*, a bacterium that infects mammals. He used two strains, a virulent S (smooth) strain and a harmless R (rough) strain to demonstrate the transfer of genetic material. The S strain is surrounded by a polysaccharide capsule, which protects it from the host's immune system, resulting in the death of the host, while the R strain, which does not have the protective capsule, is defeated by the host's immune system. Hence, when mammalian cells are infected with the R strain bacteria, the host does not die (**Figure** 8.1).

Griffith infected mice with heat-killed S strain bacteria. As expected, the heat-killed bacteria, as they were dead, had no effect on the mice (Figure 1). But then he tried something different.

He mixed the remains of heat-killed S strain bacteria with live R strain bacteria and injected the mixture into mice. Remember, separately both of these bacteria are harmless to the mice. And yet the mice died (**Figure** 8.1). Why? These mice had both live R and live S strain bacteria in their blood. How? Griffith concluded that the R strain had changed, or transformed, into the lethal S strain. Something, such as the "instructions" from the remains of the S strain, had to move into the R strain in order to turn the harmless R strain into the lethal S strain. This material that was transferred between strains had to be the heredity material. But the transforming material had yet to be identified. **Transformation** is now known as the change in genotype and phenotype due to the assimilation of external DNA (heredity material) by a cell.

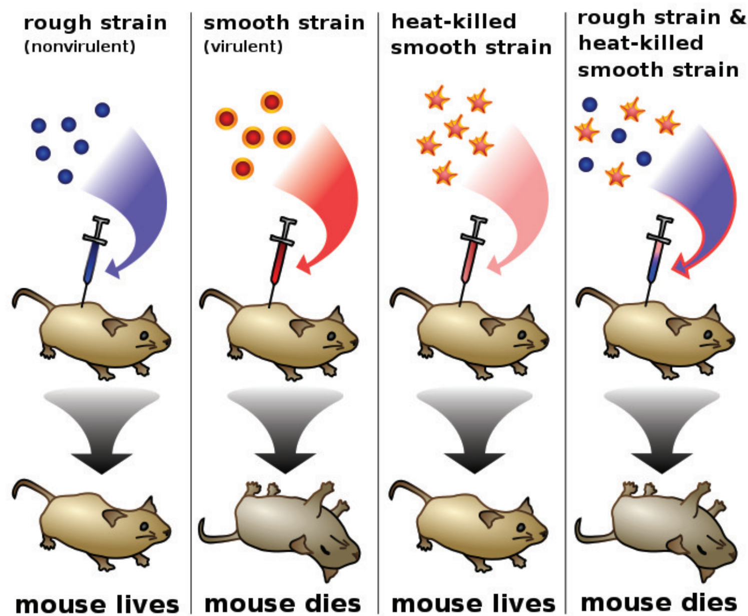

Figure 8.1: The transformation experiments of Griffith. The rough (R) strain has no effect on the mouse, whereas the smooth (S) strain is harmful to the mouse. Heat-killed S strain also has no effect on the mouse, but the mixture of heat-killed S strain and the R strain is harmful to the mouse. (2)

Over the next decade, scientists, led by Oswald Avery, tried to identify the material involved in transformation. Avery, together with his colleagues Maclyn McCarty and Colin MacLeod, removed various organic compounds from bacteria and tested the remaining compounds for the ability to cause transformation. If the remaining material did not cause transformation, than that material could not be the heredity material. Avery treated S strain bacteria with protease enzymes, which remove proteins from cells, and then mixed the remainder with R strain bacteria. The R strain bacteria transformed, meaning that proteins did not carry the genes for causing the disease. Then the remnants of the S strain bacteria were treated with **deoxyribonuclease**, an enzyme which degrades DNA. After this treatment, the R strain

bacteria no longer transformed. This indicated that DNA was the heredity material. The year was 1944.

However, this finding was not widely accepted, partly because so little was known about DNA. It was still thought that proteins were better candidates to be the heredity material. The structure of DNA was still unknown, and many scientists were not convinced that genes from bacteria and more complex organisms could be similar.

In 1952, Alfred Hershey and Martha Chase put this skepticism to rest. They conclusively demonstrated that DNA is the genetic material. Hershey and Chase used the T2 **bacteriophage**, a virus that infects bacteria, to prove this point. A virus is essentially DNA (or RNA) surrounded by a protein coat (**Figure** 8.2). To reproduce, a virus must infect a cell and use that host cell's machinery to make more viruses. The T2 bacteriophage can quickly turn an *Escherichia coli (E. coli)* bacteria into a T2 producing system. But to do that, the genetic material from T2, which could only be protein or DNA, must be transferred to the bacteria. Which one was it?

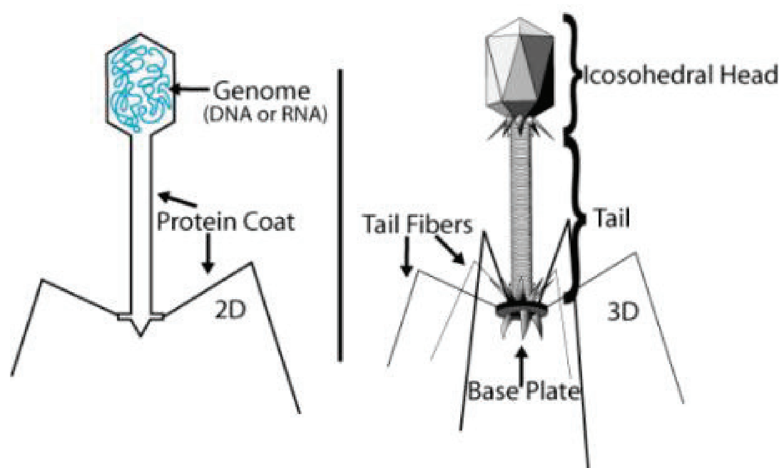

Figure 8.2: Structural overview of T2 phage. A 2-dimensional representation is on the left, and a 3-dimensional representation is on the right. The phage is essentially nucleic acid surrounded by a protein coat. (9)

Hershey and Chase performed a series of classic experiments, taking advantage of the fact that T2 is essentially just DNA and protein. In the experiments, T2 phages with either radioactive ^{32}P-labeled DNA or radioactive ^{35}S-labeled protein were used to infect bacteria. Either the radioactive proteins or radioactive DNA would be transferred to the bacteria. Identifying which one is transferred would identify the genetic material. In both experiments, bacteria were separated from the phage coats by blending, followed by centrifugation. Only the radioactively labeled DNA was found inside the bacteria, whereas the radioactive proteins stayed in the solution (**Figure** 8.3). These experiments demonstrated that DNA is the genetic material and that protein does not transmit genetic information.

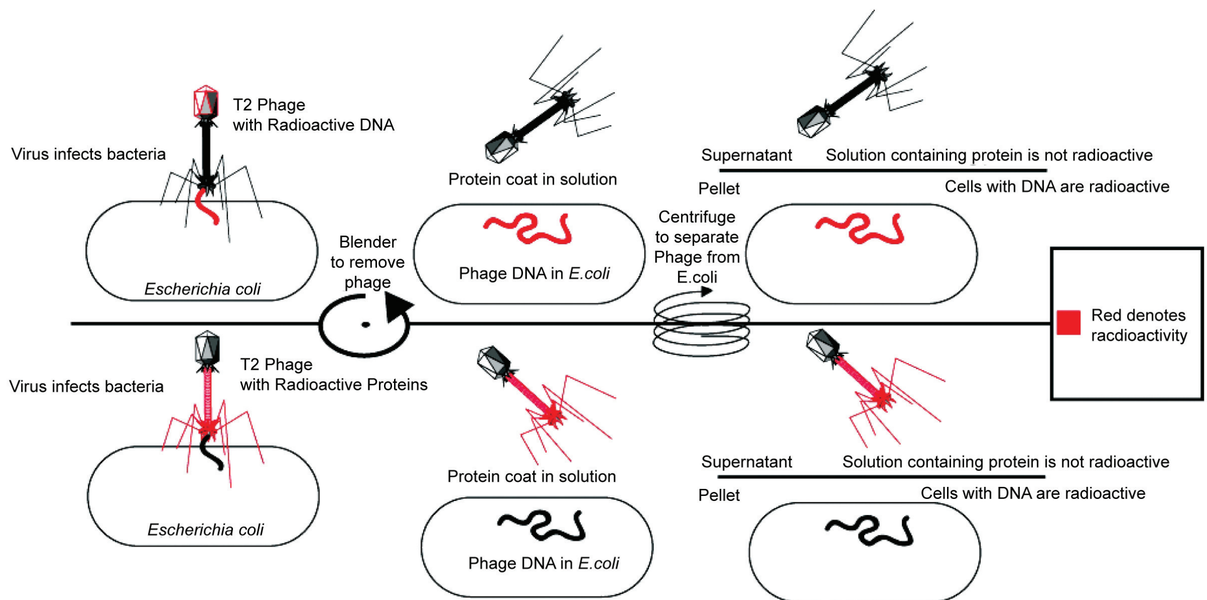

Figure 8.3: The Hershey and Chase experiment. T2 virus with either radioactive DNA (upper section) or radioactive protein (lower section) were used to infect bacteria. A blender was used to remove the phage from the bacteria followed by centrifugation. The radioactive DNA was found inside the bacteria (upper section), demonstrating that DNA is the genetic material. (12)

Chargaff's Rules

It was known that DNA is composed of **nucleotides**, each of which contains a nitrogen-containing base, a five-carbon sugar (deoxyribose), and a phosphate group. In these nucleotides, there is one of the four possible bases: adenine (A), guanine (G), cytosine (C), or thymine (T) (**Figure** 8.4).

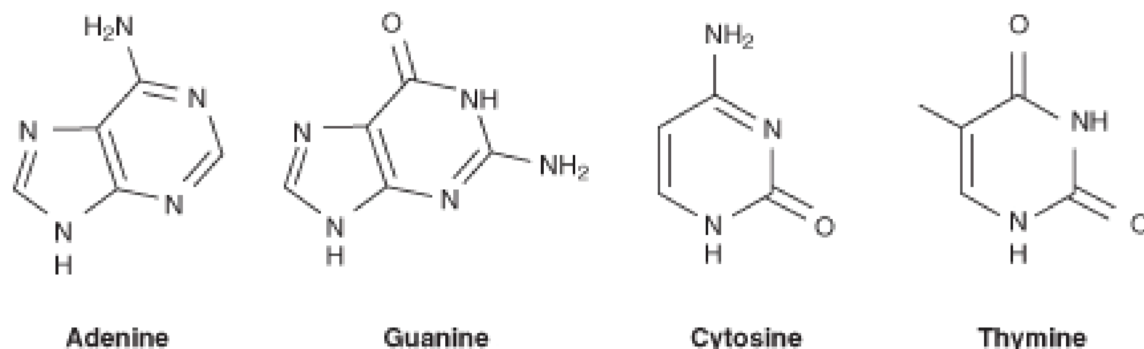

Figure 8.4: Chemical structure of the four nitrogenous bases in DNA. (13)

Erwin Chargaff proposed two main rules that have been appropriately named Chargaff's rules. In 1947 he showed that the composition of DNA varied from one species to another. This molecular diversity added evidence that DNA could be the genetic material. Chargaff also determined that in DNA, the amount of one base always approximately equals the amount of a particular second base. For example, the number of guanines equals the number of cytosines, and the number of adenines equals the number of thymines. Human DNA is 30.9% A and 29.4% T, 19.9% G and 19.8% C. This finding, together with that of the DNA structure, led to the base-pairing rules of DNA.

The Double Helix

In the early 1950s, Rosalind Franklin started working on understanding the structure of DNA fibers. Franklin, together with Maurice Wilkins, used her expertise in x-ray diffraction photographic techniques to analyze the structure of DNA. In February 1953, Francis Crick and James D. Watson of the Cavendish Laboratory in Cambridge University had started to build a model of DNA. Watson and Crick indirectly obtained Franklin's DNA X-ray diffraction data demonstrating crucial information into the DNA structure. Francis Crick and James Watson (**Figure** 8.5) then published their double helical model of DNA in *Nature* on April 25th, 1953.

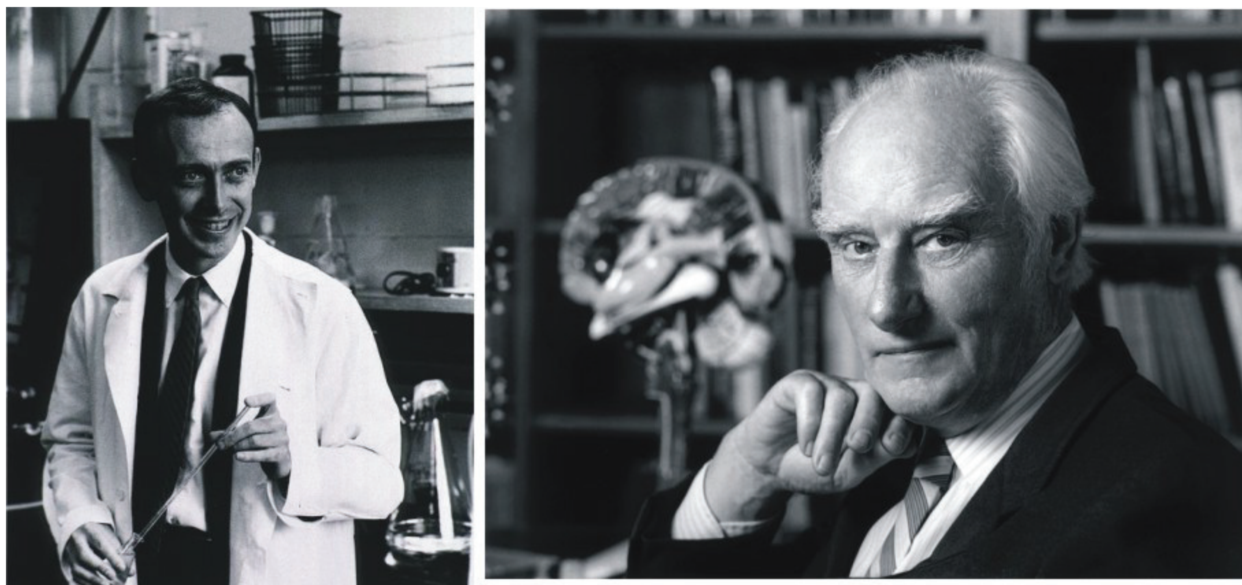

Figure 8.5: James Watson (left) and Francis Crick (right). (14)

DNA has the shape of a **double helix**, just like a spiral staircase (**Figure** 8.6). There are two sides, called the **sugar-phosphate backbone**, because they are made from alternating phosphate groups and deoxyribose sugars. The "steps" of the double helix are made from the base pairs formed between the nitrogenous bases. The DNA double helix is held together by hydrogen bonds between the bases attached to the two strands.

The double helical nature of DNA, together with the findings of Chargaff, demonstrated the base-pairing nature of the bases. Adenine always pairs with thymine, and guanine always pairs with cytosine (**Figure** 8.7). Because of this complementary nature of DNA, the bases on one strand determine the bases on the other strand. These complementary base pairs explain why the amounts of guanine and cytosine are present in equal amounts, as are the amounts of adenine and thymine. Adenine and guanine are known as **purines**. These bases consist of two ring structures. Purines make up one of the two groups of nitrogenous bases. Thymine and cytosine are **pyrimidines**, which have just one ring structure. By having a purine always combine with a pyrimidine in the DNA double helix, the distance between the two sugar-phosphate backbones is constant, maintaining the uniform shape of the DNA molecule.

The two strands in the DNA backbone run in **anti-parallel** directions to each other. That is, one of the DNA strands is built in the 5' → 3' direction, while the complementary strand is built in the 3' → 5' direction. In the DNA backbone, the sugars are joined together by phosphate groups that form bonds between the third and fifth carbon atoms of adjacent sugars. In a double helix, the direction of the nucleotides in one strand is opposite to their direction in the other strand. 5' and 3' each mark one end of a strand. A strand running in the 5'→ 3' direction that has adenine will pair with base thymine on the complementary

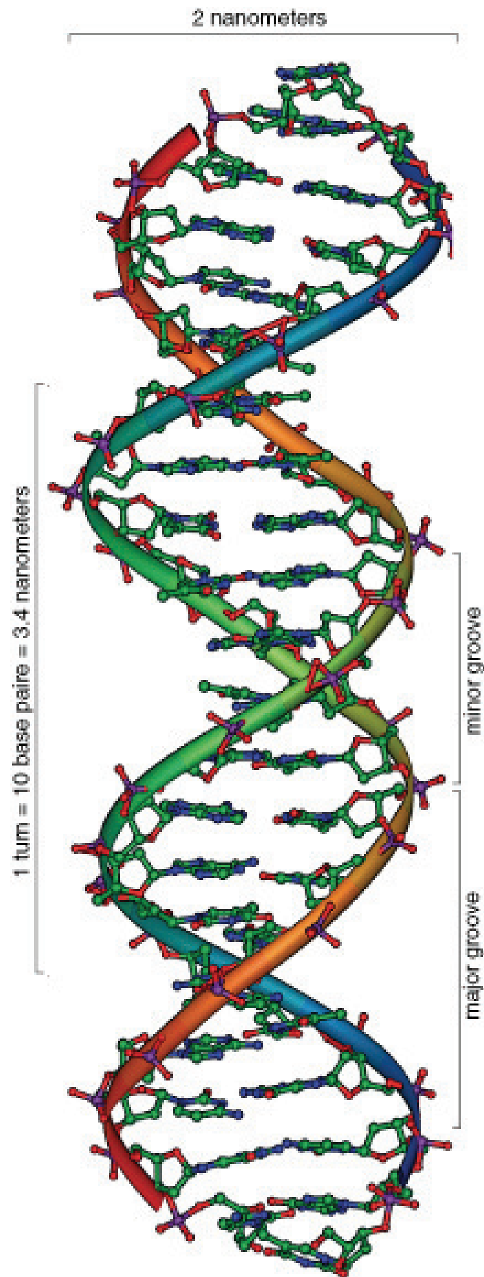

Figure 8.6: The DNA double helix. The two sides are the sugar-phosphate backbones, composed of alternating phosphate groups and deoxyribose sugars. The nitrogenous bases face the center of the double helix. (20)

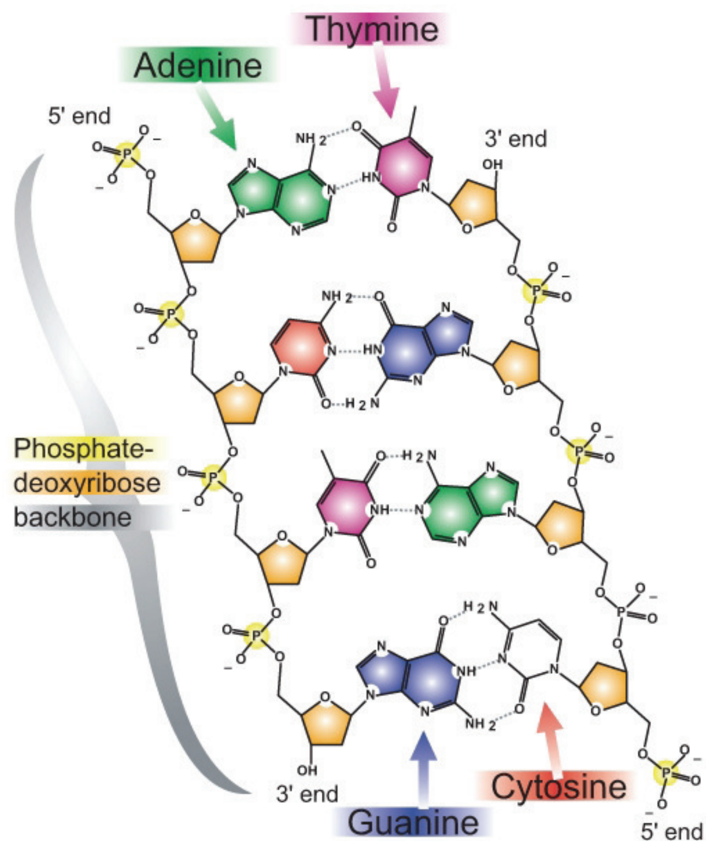

Figure 8.7: The base-pairing nature of DNA. Adenine always pairs with thymine, and they are held together with two hydrogen bonds. The guanine-cytosine base pair is held together with three hydrogen bonds. Note that one sugar-phosphate backbone is in the 5'→ 3' direction, with the other strand in the opposite 3'→ 5' orientation. (24)

strand running in 3'→ 5' direction.

So it is this four letter code, made of just A, C, G, and T, that determines what the organism will become and what it will look like. How can these four bases carry so much information? This information results from the order of these four bases in the chromosomes. This sequence carries the unique genetic information for each species and each individual. Humans have about 3,000,000,000 bits of this information in each cell. A gorilla may also have close to that amount of information, but a slightly different sequence. For example, the sequence AGGTTTACCA will have different information than CAAGGGATTA. The closer the evolutionary relationship is between two species, the more similar their DNA sequences will be. For example, the DNA sequences between two species of reptiles will be more similar than between a reptile and an elm tree.

DNA sequences can be used for scientific, medical, and forensic purposes. DNA sequences can be used to establish evolutionary relationships between species, to determine a person's susceptibility to inherit or develop a certain disease, or to identify crime suspects or victims. Of course, DNA analysis can be used for other purposes as well. So why is DNA so useful for these purposes? It is useful because every cell in an organism has the same DNA sequence. For this to occur, each cell must have a mechanism to copy its entire DNA. How can so much information be exactly copied in such a small amount of time?

DNA Replication

DNA replication is the process in which a cell's entire DNA is copied, or replicated. This process occurs during the Synthesis (S) phase of the eukaryotic cell cycle. As each DNA strand has the same genetic information, both strands of the double helix can serve as templates for the reproduction of a new strand. The two resulting double helices are identical to the initial double helix. For an animation of DNA replication, see http://www.hhmi.org/biointeractive/media/DNAi_replication_vo1-sm.mov.

Helicase and Polymerase

DNA replication begins as an enzyme, **DNA helicase**, breaks the hydrogen bonds holding the two strands together and forms a replication fork. The resulting structure has two branching strands of DNA backbone with exposed bases. These exposed bases allow the DNA to be "read" by another enzyme, **DNA polymerase**, which then builds the complementary DNA strand. As DNA helicase continues to open the double helix, the replication fork grows.

5' → 3'

The two new strands of DNA are "built" in opposite directions, through either a **leading strand** or a **lagging strand**. The leading strand is the DNA strand that DNA polymerase constructs in the 5' → 3' direction. This strand of DNA is made in a continuous manner,

moving as the replication fork grows. The lagging strand is the DNA strand at the opposite side of the replication fork from the leading strand. It goes in the opposite direction, from 3' to 5'. DNA polymerase cannot build a strand in the 3' → 5' direction. Thus, this "lagging" strand is synthesized in short segments known as **Okazaki fragments**. On the lagging strand, an enzyme known as **primase** builds a short RNA primer. DNA polymerase is then able to use the free 3' OH group on the RNA primer to make DNA in the 5' → 3' direction. The RNA fragments are then degraded and new DNA nucleotides are added to fill the gaps where the RNA was present. Another enzyme, **DNA ligase**, is then able to attach (ligate) the DNA nucleotides together, completing the synthesis of the lagging strand (**Figure** 8.8).

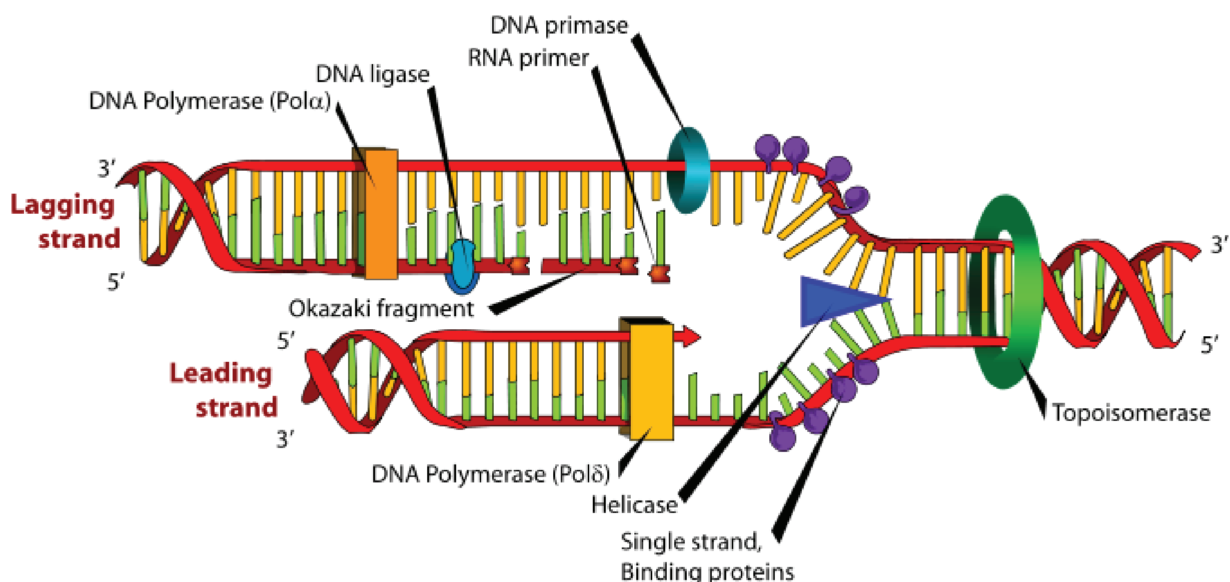

Figure 8.8: DNA replication. The two DNA strands are opened by helicase. The strands are held open by a single strand of binding proteins, preventing premature reannealing. Topoisomerase solves the problem caused by tension generated by winding/unwinding of DNA. This enzyme wraps around DNA and makes a cut permitting the helix to spin and relax. Once DNA is relaxed, topoisomerase reconnects broken strands. DNA primase synthesizes a short RNA primer which initiates the Okazaki fragment. Okazaki fragments are attached by DNA ligase. (17)

Many replication forks develop along a chromosome. This process continues until the replication forks meet, and the all of the DNA in a chromosome has been copied. Each new strand that has formed is complementary to the strand used as the template. Each resulting DNA molecule is identical to the original DNA molecule. During prophase of mitosis or prophase I of meiosis, these molecules of DNA condense into a chromosome made of two identical chromatids. This process ensures that cells that result from cell division have identical sets of genetic material, and that the DNA is an exact copy of the parent cell's DNA.

RNA

DNA → RNA → Protein

"DNA makes RNA makes protein." So what exactly is RNA? Ribonucleic acid, or **RNA**, is the other important nucleic acid in the three player act. When we say that "DNA makes RNA makes protein," what do we mean? We mean that the information in DNA is somehow transferred into RNA, and that the information in RNA is then used to make the protein.

To understand this, it helps to first understand RNA. If you remember from the chapter titled *Cell Division and Reproduction*, a **gene** is a segment of DNA that contains the information necessary to encode an RNA molecule or a protein. Keep in mind that even though you have many thousands of genes, not all are used in every cell type. In fact, probably only a few thousand are used in a particular type of cell, with different cell types using different genes. However, while these genes are embedded in the large chromosomes that never leave the nucleus, the RNA is relatively small and is able to carry information out of the nucleus.

RNA Structure

RNA structure differs from DNA in three specific ways. Both are nucleic acids and made out of nucleotides; however, RNA is single stranded while DNA is double stranded. RNA contains the 5-carbon sugar ribose, whereas in DNA, the sugar is deoxyribose. Though both RNA and DNA contain the nitrogenous bases adenine, guanine and cytosine, RNA contains the nitrogenous base uracil instead of thymine. Uracil pairs with adenine in RNA, just as thymine pairs with adenine in DNA. A comparison of RNA and DNA is shown in **Table** 8.1 and **Figure** 8.9.

Table 8.1:

	RNA	DNA
	single stranded	double stranded
Specific Base	contains uracil	contains thymine
Sugar	ribose	deoxyribose
Size	relatively small	big (chromosomes)
Location	moves to cytoplasm	stays in nucleus
Types	3 types: mRNA, tRNA, rRNA	generally 1 type

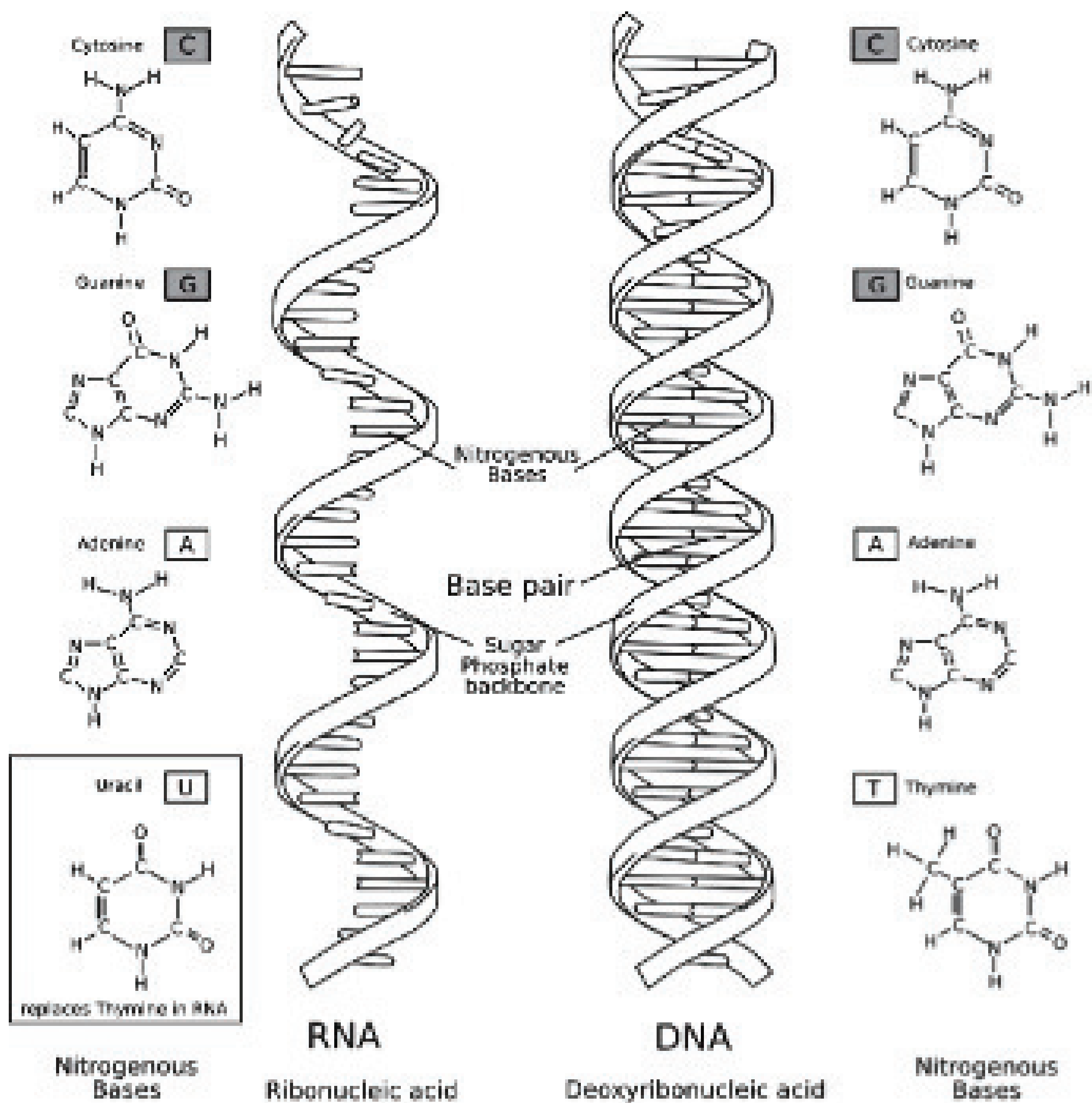

Figure 8.9: A comparison of RNA and DNA. RNA is single stranded and contains the base uracil, which replaces thymine. (11)

Three Types of RNAs

So what does RNA do? There are three types of RNA: messenger RNA (mRNA), transfer RNA (tRNA), and ribisomal RNA (rRNA). All three of these nucleic acids work together to produce a protein. The **mRNA** takes the instructions from the nucleus to the cytoplasm, where the ribosomes are located. **Ribosomes** are where the proteins are made. The ribosomes themselves are made out of **rRNA** and other proteins. The mRNA binds to the ribosome, bringing the instructions to order the amino acids to the site of protein synthesis. Finally, the **tRNA** brings the correct amino acid to the site of protein synthesis (**Figure** 8.10 and **Figure** 8.11). In mRNA, the four nucleotides (A, C, G, and U) are arranged into **codons** of three bases each. Each codon encodes for a specific amino acid, except for the stop codons, which terminate protein synthesis. tRNA, which has a specific "3-leaf clover structure," contains a three base region called the **anticodon**, which can base pair to the corresponding three-base codon region on mRNA. More will be discussed on these processes during the lesson on translation that follows.

Remember, proteins are made out of amino acids, so how does the information get converted from the language of nucleotides to the language of amino acids? The process is called **translation**.

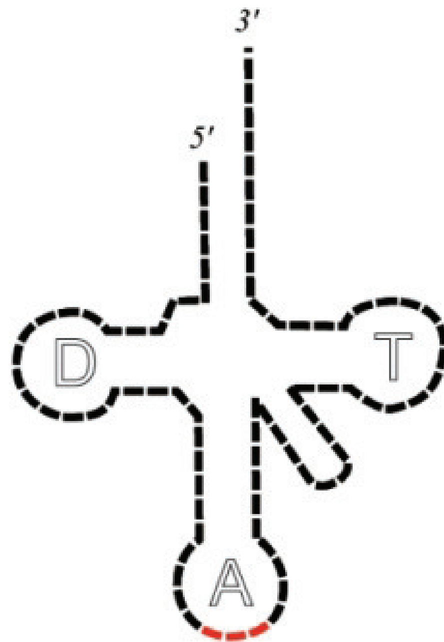

Figure 8.10: 2-Dimensional tRNA structure depicting the 3-leaf clover structure. The D arm (D) is one stem ending in a loop. The anticodon arm (A) is a second stem whose loop contains the anticodon on the bottom of the tRNA. The T arm (T) is the third stem opposite the D arm. (5)

Figure 8.11: 3-Dimensional representation of a tRNA. Coloring: CCA tail in orange, Acceptor stem in purple, D arm in red, Anticodon arm in blue with Anticodon in black, and T arm in green. The acceptor stem is made by the base pairing of the 5'-terminal nucleotide with the 3'-terminal nucleotide. The CCA tail is a CCA sequence at the 3' end of the tRNA molecule, used to attach the amino acid. This sequence is important for the recognition of tRNA by enzymes critical in translation. (16)

Small interfering RNA (siRNA), microRNA (miRNA) and small nuclear RNA (snRNA): siRNA and miRNA are revolutionizing molecular biology, developmental biology, and even medicine. The 2006 Nobel prize in Physiology and Medicine was awarded to Dr. Andrew Fire and Dr. Craig Mello for their discovery of siRNA, which is a type of double-stranded RNA that inhibits gene expression at the mRNA level. Specifically, siRNA acts on homologous processed mRNA by targeting it for degradation. siRNA is responsible for RNA interference (RNAi). RNAi has a natural role in that it is used by plants in defense against plant viral RNAs. miRNAs are also involved in the regulation of gene expression. They are transcribed but not translated into proteins. snRNAs are found within the nucleus of eukaryotic cells. They are involved in a variety of important processes such as RNA splicing (removal of introns), and regulation of **transcription** factors (discussed in Lesson 8.2: Protein Synthesis).

Lesson Summary

- Griffith demonstrated the process of transformation, which is the change in genotype and phenotype due to the assimilation of the external DNA by a cell.
- Avery and colleagues demonstrated that DNA was the transforming material.
- The Hershey and Chase experiments conclusively demonstrated that DNA is the genetic material.
- Watson and Crick demonstrated the double helix model of DNA.
- The Base paring rules state that A always pairs with T and G always pairs with C.
- DNA replication is the process by which a cell's entire DNA is copied, or replicated.
- During DNA replication, the two new strands of DNA are "built" in opposite directions, starting at replication forks.
- RNA is a single-stranded nucleic acid.
- RNA contains the nitrogenous base uracil.
- There are three types of RNA: mRNA, tRNA, and rRNA.
- mRNA is the intermediary between the nucleus, where the DNA lives, and the cytoplasm, where proteins are made.

Review Questions

1. Discuss how DNA was identified as the genetic material.
2. Define transformation.
3. In DNA, why does the amount of adenine approximately equal the amount of thymine?
4. What are the base pairing rules?
5. Explain Watson and Crick's double helix model of DNA.
6. How is DNA replicated?
7. Discuss the importance of mRNA.
8. Explain the main differences between mRNA, rRNA,and tRNA.

Further Reading / Supplemental Links

- Campbell, N.A. and Reece, J.B. *Biology*, Seventh Edition, Benjamin Cummings, San Francisco, CA, 2005.
- Biggs, A., Hagins, W.C., Kapicka, C., Lundgren, L., Rillero, P., Tallman, K.G., and Zike, D., *Biology: The Dynamics of Life*, California Edition, Glencoe Science, Columbus, OH, 2005.
- Nowicki S., *Biology*, McDougal Littell, Evanston, IL, 2008. The Dolan DNA Learning Center.
- http://www.dnalc.org/home_alternate.html
- DNA Interactive:
- http://www.dnai.org/
- A Science Odyssey:
- http://www.pbs.org/wgbh/aso/tryit/dna/
- National Human Genome Research Institute:
- http://www.genome.gov
- The RNA Modification Database:
- http://library.med.utah.edu/RNAmods/
- The RNA World:
- http://nobelprize.org/nobel_prizes/chemistry/articles/altman/index.html
- http://en.wikipedia.org

Vocabulary

amino acid The monomers that combine to form a polypeptide (protein).

anticodon A 3 base sequence on the tRNA that base pairs with the codon on the mRNA.

anti-parallel Describes the orientation of the two DNA strands; one of the DNA strands is built in the 5' → 3' direction, while the complementary strand is built in the 3' → 5' direction.

bacteriophage A virus that infects bacteria.

codon A sequence of three nucleotides within mRNA; encodes for a specific amino acid or termination (stop) sequence.

deoxyribonuclease An enzyme which degrades DNA.

DNA Deoxyribonucleic acid, the genetic (heredity) material.

DNA polymerase The enzyme that builds a new DNA strand during DNA replication.

DNA helicase The enzyme that breaks the hydrogen bonds holding the two DNA strands together during DNA replication.

DNA ligase An enzyme that joins broken nucleotides together by catalyzing the formation of a bond between the phosphate group and deoxyribose sugar of adjacent nucleotides in the DNA backbone.

DNA replication The process in which a cell's entire DNA is copied.

double helix The shape of DNA, resembling a spiral staircase.

gene A segment of DNA that contains the information necessary to encode an RNA molecule or a protein.

lagging strand The DNA strand at the opposite side of the replication fork from the leading strand.

leading strand The DNA strand that DNA polymerase constructs in the 5' $\rightarrow$ 3' direction.

mRNA Messenger RNA; serves as a nucleic acid intermediate between the nucleus and the ribosomes.

nucleotide Monomer of nucleic acids, composed of a nitrogen-containing base, a five-carbon sugar, and a phosphate group.

Okazaki fragments Short fragments of DNA that comprise the lagging strand.

primase An enzyme that builds a short RNA primer on the lagging strand during DNA replication.

purines Nitrogenous bases consisting of two ring structures; adenine and guanine.

pyrimidines Nitrogenous bases consisting of one ring structure; thymine and cytosine.

ribosome Non-membrane bound organelle; site of protein synthesis.

RNA Ribonucleic acid; single-stranded nucleic acid.

rRNA Ribosomal RNA; together with proteins, forms ribosomes.

sugar-phosphate backbone The sides of the DNA double helix; composed of alternating phosphate groups and deoxyribose sugars.

transcription The process of making an mRNA from the information in the DNA sequence.

translation The process of making a protein from the information in a mRNA sequence.

transformation The change in genotype and phenotype due to the assimilation of external DNA (heredity material) by a cell.

tRNA Transfer RNA; brings amino acids to the ribosome.

Points to Consider

- ”DNA → RNA” Can you think of a method in which the information in DNA is transferred to an RNA molecule?
- Can you hypothesize on how the As, Cs, Gs and Us of RNA can code for the 20 amino acids of proteins?
- Can you develop a model in which the three types of RNAs interact to make a protein?

8.2 Lesson 8.2: Protein Synthesis

Lesson Objectives

- Discuss the meaning of DNA → RNA → Protein.
- Describe how transcription makes RNA from a DNA template.
- Explain the various types of modification mRNA undergoes before translation.
- Discuss mRNA splicing and define introns and exons.
- Explain how the Genetic Code is a three letter code, and describe its role in translating nucleotides into amino acids.
- Explain that a reading frame is the group of three bases in which the mRNA is read, and describe how interrupting the reading frame may have severe consequences on the protein.
- Discuss what is meant by the universal genetic code.

- Describe translation. Explain that translation is the process of ordering the amino acids into a polypeptide; translation involves changing the language of nucleotides into the language of amino acids.
- Illustrate the process of translation, describing how mRNA, rRNA, and tRNA all work together to complete the process.
- Discuss what happens to the polypeptide after translation.

DNA → RNA → Protein

The central dogma of molecular biology describes the fundamental process that makes us all different. We all have different proteins. That is, though they may be the same types of proteins, such as we have the protein collagen found in bones, many of our proteins are slightly different and thus work slightly differently. If all our proteins acted the same way, we would all be exactly the same. But because we all have different DNA sequences, and DNA contains genes, and **genes** contain the information to encode an RNA molecule or a protein, we are all different.

So how does "DNA makes RNA makes protein" actually happen? The two processes necessary to make a protein from the information in DNA are transcription and translation. Transcription, which happens in the nucleus, uses the DNA sequence to make an RNA molecule. The RNA then leaves the nucleus and goes to the cytoplasm where translation occurs on a ribosome and produces a protein.

Transcription

Transcription is "DNA → RNA." In other words, transcription is the transfer of the genetic "instructions" from DNA to RNA. During transcription, a complementary copy of RNA is made. Whereas in DNA replication both strands of the DNA double helix are used as templates, in transcription only one strand is needed. RNA polymerase enzymatically "reads" a template strand of DNA, known as the coding strand, to synthesize the complementary RNA strand. Transcription is divided into 3 stages, appropriately named initiation, elongation and termination. For an animation of transcription, see http://www.hhmi.org/biointeractive/media/DNAi_transcription_vo1-sm.mov.

Initiation

Transcription begins with the binding of RNA polymerase to the promoter of a gene. An eukaryotic promoter usually includes specific sequences that are recognized by transcription factors, which are proteins that aid in the binding of RNA polymerase to the correct place on the DNA. The transcription initiation complex formed by the promoter, transcription factors, and RNA polymerase signals the start, or **initiation**, of transcription. The DNA

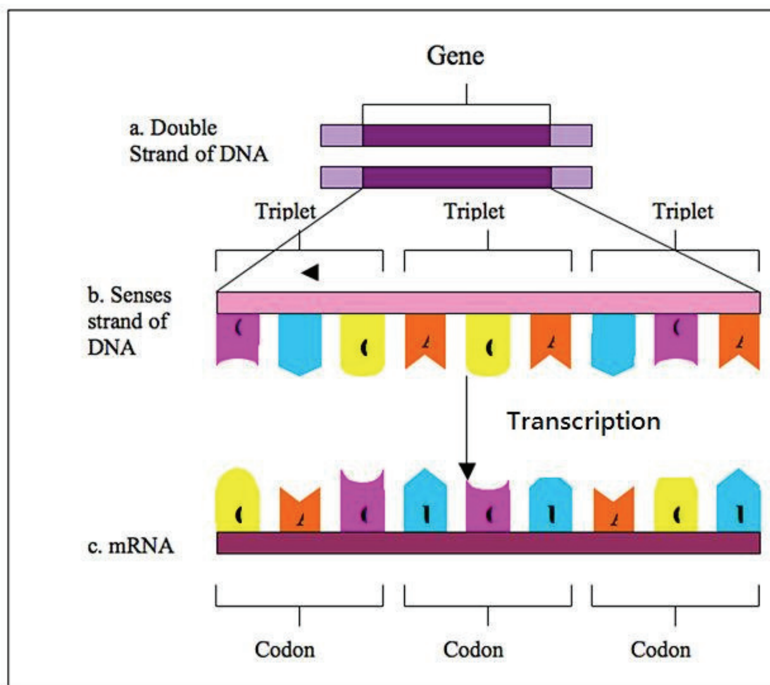

Figure 8.12: Each gene (a) contains triplets of bases (b) that are transcribed into RNA (c). Every triplet, or codon, encodes for a unique amino acid. (4)

unwinds and produces a small open complex, which allows **RNA polymerase** to "read" the DNA template and begin the synthesis of RNA.

Elongation

Transcription **elongation** involves the further addition of RNA nucleotides and the change of the open complex to a transcriptional complex. As the RNA transcript is assembled, DNA in front of RNA polymerase unwinds and transcription continues. As transcription progresses, RNA nucleotides are added to the 3' end of the growing RNA transcript. The transcriptional complex has a short DNA-RNA hybrid, an 8 base-pair stretch in which the newly made RNA is temporarily hydrogen bonded to the DNA template strand. Unlike DNA replication, mRNA transcription can involve multiple RNA polymerases, allowing numerous mRNAs to be produced from a single copy of the gene. This step also involves a proofreading mechanism that can replace an incorrectly added RNA nucleotide.

Termination

The termination of transcription in prokaryotes and eukaryotes is very different. Though both involve the detachment of the RNA from the DNA template, how this occurs is surprisingly distinct. Bacteria use two different strategies for transcription termination, Rho-dependent and Rho-independent termination. In **Rho-dependent termination**, a protein factor called "Rho" destabilizes the RNA-DNA hybrid, releasing the newly synthesized mRNA from the elongation complex. In Rho-independent termination, RNA transcription stops when the newly synthesized RNA molecule forms a hairpin loop followed by a run of uracils. This structure is the signal for the detachment of the RNA from the DNA. The DNA is now ready for translation.

The **termination** of transcription in eukaryotes is less well understood. The RNA polymerase transcribes a polyadenylation signal. Polyadenylation is the addition of a string of A's to the mRNA's 3' end and will be discussed in the next section. However, soon after the transcription of this signal, proteins cut the RNA transcript free from the polymerase and the polymerase eventually falls off the DNA. This process produces a pre-mRNA, an mRNA that is not quite ready to be translated.

Eukaryotic mRNA Processing

Newly transcribed eukaryotic mRNA is not ready for translation. This mRNA requires extensive processing, and so is known as pre-mRNA. The modification processes include splicing, the addition of a 5' cap, editing, and polyadenylation. Once these process have occurred, the mature mRNA can be exported through the nuclear pore.

Splicing

Humans have approximately 22,000 genes, yet make many more proteins. How? A process called **alternative splicing** allows one mRNA to produce many polypeptides. To understand this concept, the structure of the pre-mRNA must be discussed.

Eukaryotic pre-mRNA contains introns and exons. An **exon** is the region of a gene that contains the code for producing a protein. Most genes contain many exons, with each exon containing the information for a specific portion of a complete protein. In many species, a gene's exons are separated by long regions of DNA that have no identified function. These long regions are introns, and must be removed prior to translation. **Splicing** is the process by which introns are removed (**Figure** 8.13). Sometimes a process called alternative splicing allows **pre-mRNA** messages to be spliced in several different configurations, allowing a single gene to encode multiple proteins. Splicing is usually performed by an RNA-protein complex called the spliceosome, but some RNA molecules have their own catalytic activity and are capable of acting like enzymes to catalyze their own splicing. For an animation of RNA splicing, see http://vcell.ndsu.edu/animations/mrnasplicing/first.htm.

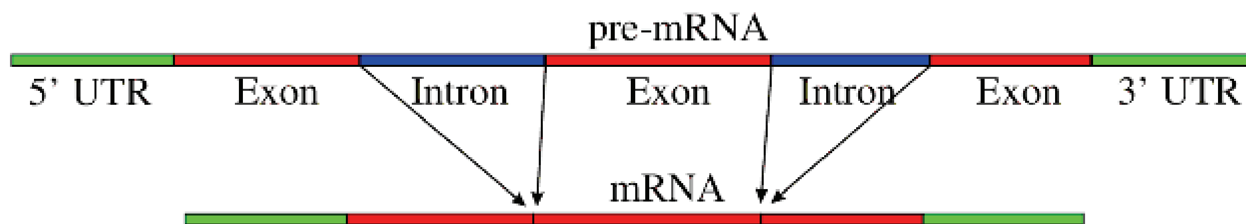

Figure 8.13: Splicing, the process by which introns (blue) are removed from pre-mRNA. Exons (red) contain the information used to produce the polypeptide. There are untranslated regions (UTR) at both the beginning and end of the pre-mRNA. (3)

5' cap Addition

How does the mRNA know it is time to leave the nucleus? Once the mRNA leaves the nucleus, how does it find a ribosome? A signal on the front, 5'-end of the mRNA helps with both jobs. This signal is the 5' cap. The 5' cap is a modified guanine nucleotide added to the 5'-end of the pre-mRNA. This 5' cap is crucial for recognition and proper attachment of the mRNA to the ribosome, as well as protection from exonucleases, enzymes that degrade nucleic acids.

Editing

In certain instances, the nucleotide sequence of an mRNA will be changed to allow the mRNA to produce multiple proteins. This process is called **editing**. The classic example is editing

of the apolipoprotein B (APOB) mRNA in humans. The APOB protein occurs in the plasma in two main forms, APOB48 and APOB100. The first is synthesized exclusively by the small intestine, the second by the liver. Both proteins are coded for by the same gene, which is transcribed into a single pre-mRNA. Editing creates a premature (early) stop codon, which upon translation produces a smaller protein. As a result of the RNA editing, APOB48 and APOB100 share a common N-terminal sequence, but APOB48 lacks APOB100's C-terminal region.

Polyadenylation

In eukaryotic cells, the transcription of the polyadenylation signals indicates the termination of the process. The mRNA transcript is then cut off of the RNA polymerase and freed from the DNA. The cleavage site is characterized by the presence of the sequence AAUAAA near the end of the transcribed message. Polyadenylation then occurs. **Polyadenylation** is the addition of a poly(A) tail to the 3'-end of the mRNA. The poly(A) tail may consist of as many as 80 to 250 adenosine residues. The poly(A) tail protects the mRNA from degradation by exonucleases. Poladenylation is also important for transcription termination, export of the mRNA from the nucleus, and translation. For an animation on RNA polyadenylation, see http://vcell.ndsu.edu/animations/mrnaprocessing/first.htm.

The Genetic Code

So how exactly is the language of nucleotides used to code for the language of amino acids? How can a code of only As, Cs, Gs, and Us carry information for 20 different amino acids? The **genetic code** is the code in which the language of nucleotides is used to create the language of amino acids.

Cracking the Code

A code of at least three letters has to be the answer. A one letter code would only be able to code for four amino acids. A two letter code could only code for 16 amino acids. With a three letter code, there are 64 possibilities. As there are 20 amino acids, the answer must be a code of at least three letters.

In 1961, Francis Crick and Sydney Brenner demonstrated the presence of codons, that is, three bases of RNA that code for one amino acid (**Figure** 8.14). Also in 1961, Marshall Nirenberg and Heinrich Matthaei at the National Institutes of Health (NIH) demonstrated that a poly(U) RNA sequence was translated into a polypeptide consisting of only the amino acid phenylalanine. This proved that the codon UUU coded for the amino acid phenylalanine. Extending this work, Nirenberg and his coworkers were able to determine the nucleotide makeup of 54 of the 64 codons. Others determined the remainder of the genetic code (**Figure**

8.15).

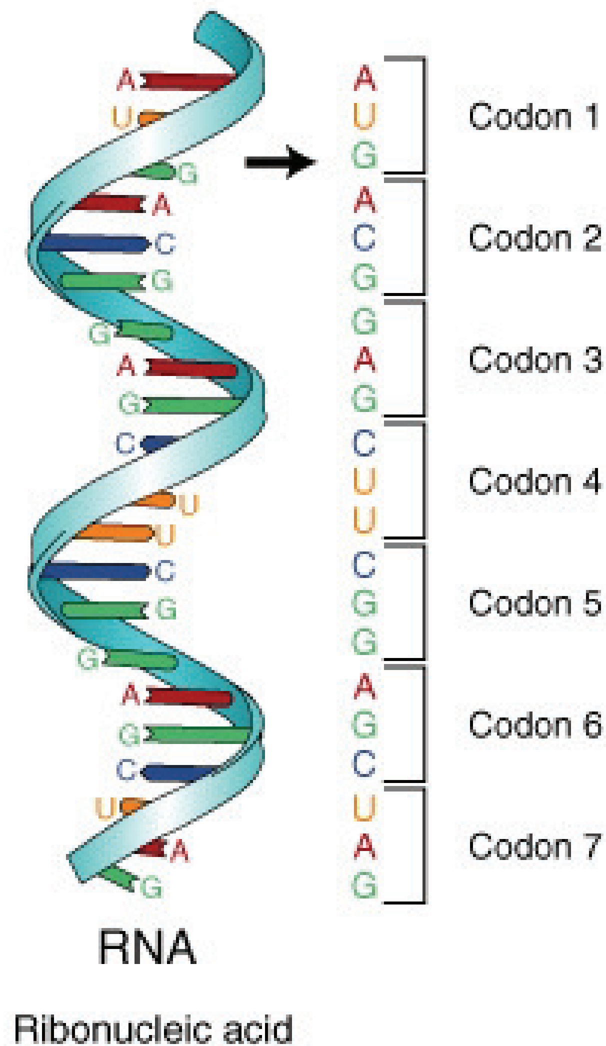

Figure 8.14: The mRNA is divided into three-base segments called codons. A codon is the segment of nucleotides that codes for an amino acid, or for a start or stop signal. There are 64 codons. (21)

Start and Stop Codons

The codon AUG codes for the amino acid methionine. This codon is also the start codon which begins every translation of every amino acid chain. The translational machinery "reads" the mRNA codon by codon until it reaches a stop, or termination, codon. Stop

		2nd base			
		U	C	A	G
1st base	U	UUU (Phe/F) Phenylalanine	UCU (Ser/S) Serine	UAU (Tyr/Y) Tyrosine	UGU (Cys/C) Cysteine
		UUC (Phe/F) Phenylalanine	UCC (Ser/S) Serine	UAC (Tyr/Y) Tyrosine	UGC (Cys/C) Cysteine
		UUA (Leu/L) Leucine	UCA (Ser/S) Serine	UAA Ochre (*Stop*)	UGA Opal (*Stop*)
		UUG (Leu/L) Leucine	UCG (Ser/S) Serine	UAG Amber (*Stop*)	UGG (Trp/W) Tryptophan
	C	CUU (Leu/L) Leucine	CCU (Pro/P) Proline	CAU (His/H) Histidine	CGU (Arg/R) Arginine
		CUC (Leu/L) Leucine	CCC (Pro/P) Proline	CAC (His/H) Histidine	CGC (Arg/R) Arginine
		CUA (Leu/L) Leucine	CCA (Pro/P) Proline	CAA (Gln/Q) Glutamine	CGA (Arg/R) Arginine
		CUG (Leu/L) Leucine	CCG (Pro/P) Proline	CAG (Gln/Q) Glutamine	CGG (Arg/R) Arginine
	A	AUU (Ile/I) Isoleucine	ACU (Thr/T) Threonine	AAU (Asn/N) Asparagine	AGU (Ser/S) Serine
		AUC (Ile/I) Isoleucine	ACC (Thr/T) Threonine	AAC (Asn/N) Asparagine	AGC (Ser/S) Serine
		AUA (Ile/I) Isoleucine	ACA (Thr/T) Threonine	AAA (Lys/K) Lysine	AGA (Arg/R) Arginine
		AUG[A] (Met/M) Methionine	ACG (Thr/T) Threonine	AAG (Lys/K) Lysine	AGG (Arg/R) Arginine
	G	GUU (Val/V) Valine	GCU (Ala/A) Alanine	GAU (Asp/D) Aspartic acid	GGU (Gly/G) Glycine
		GUC (Val/V) Valine	GCC (Ala/A) Alanine	GAC (Asp/D) Aspartic acid	GGC (Gly/G) Glycine
		GUA (Val/V) Valine	GCA (Ala/A) Alanine	GAA (Glu/E) Glutamic acid	GGA (Gly/G) Glycine
		GUG (Val/V) Valine	GCG (Ala/A) Alanine	GAG (Glu/E) Glutamic acid	GGG (Gly/G) Glycine

nonpolar	polar	basic	acidic	(stop codon)

Figure 8.15: The Genetic Code. The Genetic Code: Codons are in the mRNA sequence. The three letter and one letter code for the amino acids are shown. To read the code, find the first base on the left, the second base at the top, and the third base in the center of the table. For example, the codon GGG codes for the amino acid glycine (as does GGC, GGA, and GGU), CCG codes for Proline, UUA codes for Leucine, and AAG codes for Lysine. There are 64 codons that code for 20 amino acids and three stop codons, so an amino acid may have more than one corresponding codon. (23)

codons are not associated with a tRNA or amino acid. Three are three stop codons: UAG, UGA, UAA.

The Reading Frame

The **reading frame** is the frame of three bases in which the mRNA is read or translated. Every sequence can be read in three reading frames, each of which will produce a different amino acid sequence. For example, in the sequence GCAUGGGGGUCUAG, the reading frame can begin with either the first G, the first C, or the first A. As stated above, **translation** starts with the start codon which consists of the three bases AUG. Each subsequent codon is translated until an in-frame stop codon is reached. In the example above, the polypeptide sequence would be: methionine – glycine – valine – stop.

Mutations that disrupt the reading frame by insertions or deletions of a non-multiple of 3 nucleotide bases are known as **frameshift mutations**. Take the example:

THE BIG FAT CAT ATE THE RED RAT

A deletion mutation that disrupts the reading frame, results in a message that does not make any sense. If the 'B' is deleted:

THE IGF ATC ATA TET HER EDR AT

Once the reading frame is disrupted, the mRNA may not be translated properly. These mutations may impair the function of the resulting protein, if the protein is even formed. Many frameshift mutations result in a premature stop codon, in other words, a stop codon that come earlier than normal during translation. This would result in a smaller protein, most likely without normal function.

Degeneracy of the Universal Genetic Code

When there are 64 codon combinations for 20 amino acids (and stop codons), there is going to be some overlap. Within the genetic code there is redundancy but no ambiguity. For example, the codons GGG, GGA, GGC, and GGU all encode the amino acid glycine, but none encode another amino acid. Degenerate codons often differ in the third position.

The genetic code is said to be universal. That is, the same code is utilized by the simplest prokaryotic organism and by humans. This universality is a tremendous benefit to mankind. If a human gene is placed in a bacteria, it looks just like a piece of DNA to the bacteria. The human As, Cs, Gs, and Ts look just like the bacteria's As, Cs, Gs, and Ts. So, the bacterial proteins will transcribe and translate this DNA, making a human protein.

But how exactly are these proteins made? We have been referring to mRNA's, tRNA's, ribosomes, codons and the genetic code throughout this chapter. How do they all come together to make a protein? The process is called translation.

Translation

Translation is "RNA → protein." In other words, translation is the transfer of the instructions in RNA to a protein made of amino acids. Translation uses the products of transcription, mRNA, tRNA, and rRNA, and converts the mRNA sequence into a polypeptide according to the genetic code. The mRNA moves to the cytoplasm and interacts with a **ribosome**, which serves as the site of translation. Translation proceeds in three phases: initiation, elongation and termination.

To understand translation, first we need to understand the ribosome. Ribosomes are composed of two subunits, a small subunit and a larger subunit. Prokaryotic subunits are named the 30S and 50S subunits; eukaryotic subunits are named the 40S and 60S subunits. During translation the tRNA molecules are literally "inside" the ribosomal subunits, as they sit on the mRNA strand. When tRNAs come to the ribosome, adjacent amino acids are brought together, allowing the ribosome to catalyze the formation of the peptide bond between amino acids. The ribosome has three tRNA binding sites: the A site, the P site, and the E site (**Figure** 8.16). The **A site** binds a tRNA with an attached amino acid. The P site contains the tRNA with the growing polypeptide chain attached, and the **E site** contains the tRNA that no longer has an attached amino acid. This tRNA is preparing to exit the ribosome. A single mRNA can be translated simultaneously by multiple ribosomes. For an animation of translation, see http://www.hhmi.org/biointeractive/media/DNAi_translation_vo1-sm.mov.

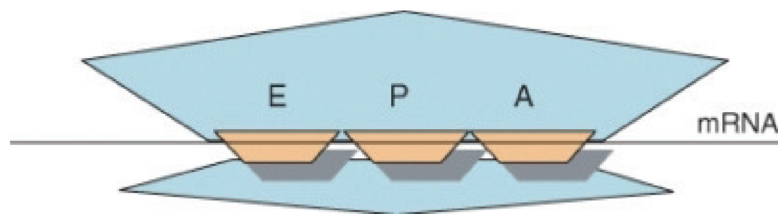

Figure 8.16: This cartoon depicts the relative location of the E, P, and A sites within the ribosome. The A site binds a tRNA bound to an amino acid, the **P site** binds a tRNA bound to the polypeptide being synthesized, and the E site binds a tRNA without an attached amino acid before the tRNA exits the ribosome. (1)

Initiation in Prokaryotes

The initiation of translation in prokaryotes involves the assembly of the ribosome and addition of the first amino acid, methionine. The 30S ribosomal subunit attaches to the mRNA. Next, the specific methionine tRNA is brought into the P. The anticodon of this tRNA will bind to the AUG codon on the mRNA. This is the only time a tRNA will be brought into the P site; all successive tRNA's will be brought to the A site as translation continues. The 50S ribosomal subunit then binds to the 30S subunit, completing the ribosome.

Initiation in Eukaryotes

The initiation of protein translation in eukaryotes is similar to that of prokaryotes with some minor modifications. The 5' cap and 3' poly(A) tail are involved in the recruitment of the ribosome. In eukaryotes the ribosome scans along the mRNA for the first methionine codon. Translation may begin at all AUG codons, however only an in-frame AUG will produce a functional polypeptide. For an animation of translation, see http://vcell.ndsu.edu/animations/translation/first.htm.

Elongation

Elongation is fairly similar between prokaryotes and eukaryotes. As translation begins, the start tRNA is sitting on the AUG codon in the P site of the ribosome, so the next codon available to accept a tRNA is at the A site. Elongation proceeds after initiation with the binding of an tRNA to the A site. The next tRNA binds to the codon, bringing the appropriate amino acid to the ribosome, and a peptide bond joins between the start methionine and the next amino acid. The entire ribosome complex moves along the mRNA, sending the first tRNA into the E site and the tRNA with the growing polypeptide into the P site. The A site is now empty and ready to accept another tRNA. The first tRNA now leaves the ribosome. The A site accepts a tRNA with an attached amino acid, a peptide bond forms between the two adjacent amino acids, and the process continues.

Termination

Termination of translation occurs when the ribosome comes to one of the three stop codons, for which there is no tRNA. At this point, a protein called a release factor binds to the A site. The release factor causes the addition of a water molecule to the polypeptide chain, resulting in the release of the completed chain from the tRNA and ribosome. The ribosome, release factor, and tRNAs then dissociate and translation is complete. The process of translation is summarized in **Figure** 8.17.

Post-Translational Modification and Protein Folding

The events following protein synthesis often include post-translational modification of the peptide chain and folding of the protein into its functional conformation. During and after synthesis, polypeptide chains often fold into secondary and then tertiary structures. These levels of organization were discussed in the chapter titled *Chemical Basis of Life*. Briefly, the primary structure of the protein is determined by the gene. The secondary and tertiary structures are determined by interactions between the amino acids within the protein (**Figure** 8.18).

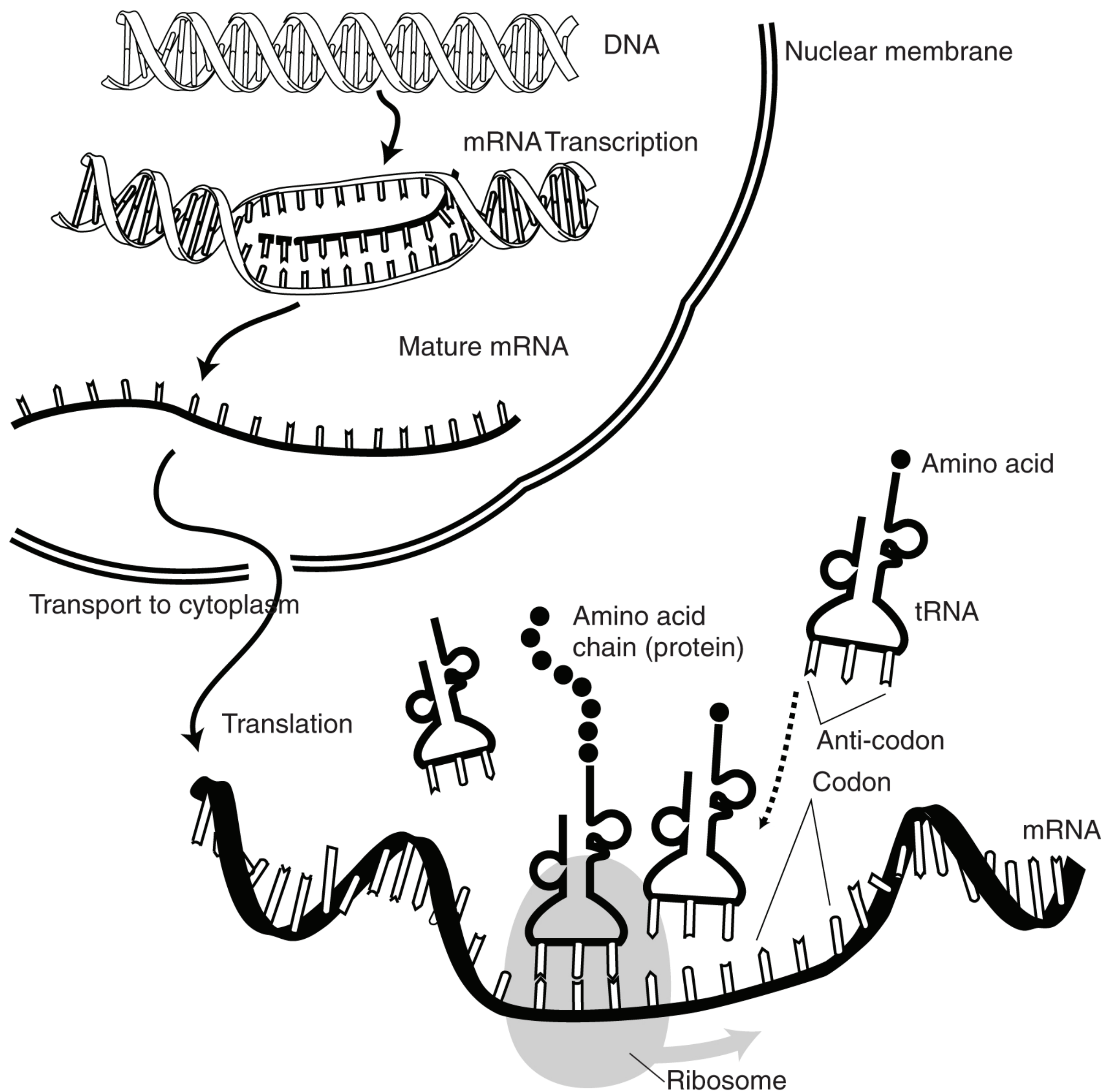

Figure 8.17: Summary of translation. Notice the mRNA segment within the ribosome. A tRNA anticodon binds to the appropriate codon, bringing the corresponding amino acid into the ribosome where it can be added to the growing polypeptide chain. (6)

Many proteins undergo post-translational modification, allowing them to then perform their specific function. This may include the formation of disulfide bridges or attachment of any of a number of biochemical functional groups, such as phosphate groups, carbohydrates or lipids. Certain amino acids may be removed, or the polypeptide chain may be cut into two pieces. Lastly, two or more polypeptides may interact with each other, forming a functional protein with a quaternary structure.

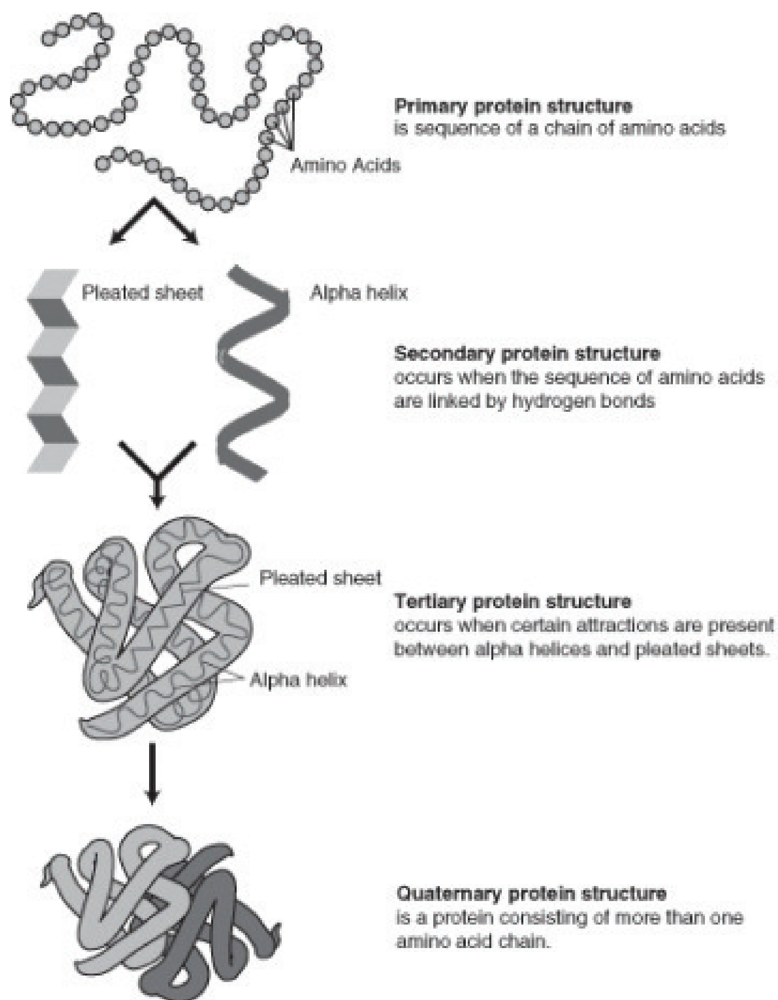

Figure 8.18: The four stages of protein folding. (19)

Lesson Summary

- "DNA makes RNA makes protein" is the central dogma of molecular biology.
- Transcription is the transfer of the genetic instructions from DNA to RNA.
- Translation is the process of using the information in the mRNA to order amino acids into a polypeptide.

- Transcription begins with the binding of RNA polymerase to the promoter of a gene.
- Newly transcribed eukaryotic mRNA is not ready for translation; this mRNA requires extensive processing, including splicing and polyadenylation.
- A codon is a three base code for one amino acid.
- Start and stop codons signal the beginning and end of translation.
- There are three stop codons.
- The reading frame is the frame of three bases in which the mRNA is read.
- The genetic code is universal.
- Translation involves the interactions of the three types of RNA.
- After the protein is made, it must fold into its functional conformation.

Review Questions

1. What is meant by “DNA → RNA → Protein?”
2. Describe transcription.
3. List some of the modification mRNAs undergoes before translation.
4. Define introns and exons.
5. Describe mRNA splicing.
6. Describe the role of the Genetic Code in translation.
7. What is a reading frame?
8. Discuss what is meant by the universal genetic code.
9. Explain translation.
10. What is protein folding?

Further Reading / Supplemental Links

- Transcription Animation

http://www-class.unl.edu/biochem/gp2/m_biology/animation/gene/gene_a2.html

- Campbell, N.A. and Reece, J.B. *Biology*, Seventh Edition, Benjamin Cummings, San Francisco, CA, 2005.
- Biggs, A., Hagins, W.C., Kapicka, C., Lundgren, L., Rillero, P., Tallman, K.G., and Zike, D., *Biology: The Dynamics of Life*, California Edition, Glencoe Science, Columbus, OH, 2005.
- Nowicki S., *Biology*, McDougal Littell, Evanston, IL, 2008.
- The National Health Museum:

http://www.accessexcellence.org/RC/VL/GG/rna_synth.html

- Genetic Science Learning Center:

http://learn.genetics.utah.edu/units/basics/transcribe/

- Kimball's Biology Pages:

http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Codons.html

- The National Health Museum:

http://www.accessexcellence.org/RC/VL/GG/genetic.html

- The National Health Museum:

http://www.accessexcellence.org/RC/VL/GG/protein_synthesis.html

- http://en.wikipedia.org

Vocabulary

5' cap A modified guanine nucleotide added to the 5'-end of the pre-mRNA; crucial for recognition and proper attachment of the mRNA to the ribosome.

A site Within the ribosome; binds a tRNA with an attached amino acid.

alternative splicing Process by which pre-mRNA messages can be spliced in several different configurations, allowing a single gene to encode multiple proteins.

E site Within the ribosome; contains the tRNA that no longer has an attached amino acid.

editing The process of changing the nucleotide sequence of an mRNA to allow the mRNA to produce multiple proteins.

elongation The segment of transcription that the further addition of RNA nucleotides; also refers to the process during translation of adding additional amino acids to the growing polypeptide chain.

exon The region of a gene that contains the code for producing a protein.

frameshift mutation Mutation that disrupt the reading frame by insertions or deletions of a non-multiple of three nucleotide bases.

gene A segment of DNA that contains the information necessary to encode an RNA molecule or a protein.

genetic code The code in which the language of nucleotides is used to create the language of amino acids.

initiation The start of transcription; signaled by the transcription initiation complex formed by the promoter, transcription factors, and RNA polymerase; also refers to the start of translation.

intron Long region of DNA that has no identified function; separates exons.

P site Within the ribosome; contains the tRNA with the growing polypeptide chain attached.

pre-mRNA Newly transcribed eukaryotic mRNA; not yet ready for translation.

polyadenylation The addition of a string of adenines to the mRNAs 3'-end; signals the termination of translation in eukaryotes.

reading frame The frame of three bases (codon) in which the mRNA is translated.

Rho-dependent termination Involves a protein factor (Rho), which destabilizes the RNA-DNA hybrid, releasing the newly synthesized mRNA from the elongation complex.

ribosome Non-membrane bound organelle; site of protein synthesis.

RNA polymerase The enzyme that reads a template strand of DNA during transcription to synthesize the complementary RNA strand.

splicing The process by which introns are removed from pre-mRNA.

spliceosome RNA-protein complex that usually performs splicing of the pre-mRNA.

termination The end of transcription; involves the detachment of the RNA from the DNA template; also refers to the end of translation when the ribosome comes to one of the three stop codons, for which there is no tRNA.

transcription The process that uses the DNA sequence to make an mRNA molecule.

translation The process of converting the language of nucleotides (in the mRNA) into the language of amino acids.

Points to Consider

We know what happens when everything goes right. The result is a correctly made protein that functions properly and maintains homeostasis.

- However, what happens when things do not go right? What can lead to a protein not being made correctly or not functioning correctly?
- Can you think of possible mechanisms that may that can interfere with protein synthesis?
- How can a change in the DNA sequence lead to a different protein?

8.3 Lesson 8.3: Mutation

Lesson Objectives

- Define mutation.
- Describe common causes of mutation.
- Describe common types of mutation.
- Illustrate common chromosomal alterations.
- Discuss potential outcomes of point mutations.
- List and describe three common types of point mutations.
- Discuss consequences of effect-on-function mutations.
- Discuss the significance of germline and somatic mutations.
- Explain why some mutations are harmful and some beneficial.
- Discuss the saying, "Without beneficial mutations, evolution can not occur."

Introduction

You have learned that an **allele** is an alternative form of a gene. Most, if not all genes have alternative forms causing the resulting protein to function slightly differently. But are there alleles that cause proteins to function dramatically differently or not function at all? A **mutation** is a change in the DNA or RNA sequence, and many mutations result in new alleles. Some of these changes can be beneficial. In fact, evolution could not take place without the genetic variation that results from mutations. But some mutations are harmful. There are also chromosomal mutations, large changes with dramatic effects.

Causes of Mutation

Is it possible for mutations to occur spontaneously, or does there have to be a cause of the mutation? Well, the answer is that both are possible. A spontaneous mutation can just

happen, possibly due to a mistake during DNA replication or transcription. Mutations may also occur during mitosis and meiosis. A mutation caused by an environmental factor, or **mutagen**, is known as an induced mutation. Typical mutagens include chemicals, like those inhaled while smoking, and radiation, such as X-rays, ultraviolet light, and nuclear radiation. **Table** 8.2 lists some spontaneous mutations that are common.

Table 8.2: **Common Spontaneous Mutations**

Tautomerism	a base is changed by the repositioning of a hydrogen atom
Depurination	loss of a purine base (A or G)
Deamination	spontaneous deamination of 5-methycytosine
Transition	a purine to purine, or a pyrimidine to pyrimidine change
Transversion	a purine becomes a pyrimidine, or vice versa

Types of Mutations

In multicellular organisms, mutations can be subdivided into germline mutations, which can be passed on to descendants, and somatic mutations, which cannot be transmitted to the next generation. Germline mutations change the DNA sequence within a sperm or egg cell, and therefore can be inherited. This inherited mutation may result in a class of diseases known as a genetic disease. The mutation may lead to a nonfunctional protein, and the embryo may not develop properly or survive. Somatic mutations may affect the proper functioning of the cell with the mutation. During DNA replication, the mutation will be copied. The two daughter cells formed after cell division will both carry the mutation. This may lead to the development of many cells that do not function optimally, resulting a less than optimal phenotype. Various types of mutations can all have severe effects on the individual. These include point mutations, framehift mutations and chromosomal alterations.

Keep in mind, some mutations may be beneficial or have no effect. Mutations that have no effect will not affect the expression of the gene or the sequence of amino acids in an encoded protein.

Chromosomal Alterations

Chromosomal alterations are large changes in the chromosome structure. They occur when a section of a chromosome breaks and rejoins incorrectly, or does not rejoin at all. Sometimes the segment may join backwards or reattach to another chromosome altogether. These mutations are very serious and usually lethal to the zygote or embryo. If the embryo

does survive, the resulting organism is usually sterile and thus, unable to pass along the mutation.

The five types of chromosomal alterations are deletions, duplications, insertions, inversions, and translocations (**Figure** 8.19).

1. **Deletions**: removal of a large chromosomal region, leading to loss of the genes within that region.
2. **Duplications** (or **amplifications**): lead to multiple copies of a chromosomal region, increasing the number of the genes located within that region. Some genes may be duplicated in their entirety.
3. **Insertions**: the addition of material from one chromosome to a nonhomologous chromosome.
4. **Inversions**: reversing the orientation of a chromosomal segment.
5. **Translocations**: interchange of genetic material between nonhomologous chromosomes.

Point Mutations

As the name implies, **point mutations** occur at a single site within the DNA. Lets go back to our earlier example from lesson 8.2:

THE BIG FAT CAT ATE THE RED RAT.

A change at any one position could result in a sequence that does not make sense. Such as:

THE BIG FAT SAT ATE THE RED RAT.

As shown above, point mutations exchange one nucleotide for another and are known as base substitution mutations. These mutations are often caused either by chemicals or by a mistake during DNA replication. A transition exchanges a purine for a purine (A G) or a pyrimidine for a pyrimidine, (C T), and is the most common point mutation. Less common is a transversion, which exchanges a purine for a pyrimidine or a pyrimidine for a purine (C/T A/G). Point mutations that occur within the protein coding region of a gene are classified by the effect on the resulting protein:

1. **Silent mutations**: which code for the same amino acid.
2. **Missense mutations**: which code for a different amino acid.
3. **Nonsense mutations**: which code for a premature stop codon.

These mutations may result in a protein with the same function, with altered function, or with no function. Silent mutations, as they code for the same amino acid, will have no altered effect on the protein. Missense mutations may have a minor effect or a dramatic

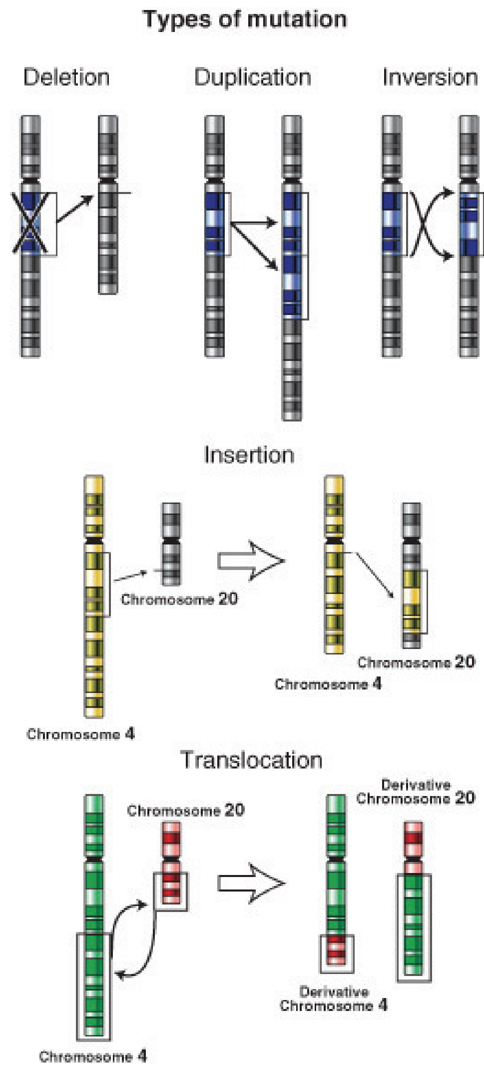

Figure 8.19: Chromosomal alterations. Deletion: the blue segment has been removed; Duplication: the blue segment has been duplicated; Inversions: the blue segment has been reversed; Insertion: the yellow segment has been removed from chromosome 4 and placed into chromosome 20; Translocation: a green segment from chromosome 4 has been exchanged with a red segment from chromosome 20. (8)

effect on the protein. Nonsense mutations usually have the most dramatic effet. Depending on the position of the premature stop codon, nonsense mutations may result in an unstable mRNA that cannot be translated, or in a much ”smaller” protein without any activity.

Deletions and Insertions

As pointed out earlier, a deletion or insertion in the DNA can alter the reading frame. **Deletions** remove one or more nucleotides from the DNA, whereas **insertions** add one or more nucleotides into the DNA. These mutations in the coding region of a gene may also alter splicing of the mRNA (splice site mutation). Mutations which alter the reading frame are known as **frameshift mutations**. **Splice site mutations** and frameshift mutations both can dramatically change the mRNA, altering the final protein product.

Effect-on-Function Mutations

For a cell or organism to maintain homeostasis, the proteins work in a highly defined and regulated manner. It may take just one protein not working correctly to interrupt homeostasis. A protein having more or less activity than normal, or a different activity or function, may be enough to interrupt homeostasis. Mutations that may result in altered function of the gene product or protein are loss-of-function and gain-of-function mutations, as well as dominant negative mutations.

Loss-of-function mutations result in a gene product or protein having less or no function. **Gain-of-function mutations** result in the gene product or protein having a new and abnormal function. **Dominant negative mutations** have an altered gene product that acts in a dominant manner to the wild-type gene product.

Significance of Mutation

Imagine the coding sequence (broken up into codons) TAC CCC GGG. This is a fairly generic coding sequence. It transcribes into the following mRNA: AUG GGG CCC, which would translate into start-glycine-proline. As glycine is encoded by four codons (GGG, GGA, GGC, GGU), any change in the third position of that codon will have no effect. The same is true for the codon for proline. But what about changes in the other nucleotides in the sequence? They could have potentially dramatic effects. The effects depend on the outcome of the mutation. Obviously any change to the start codon will interrupt the start of translation. Turning the simple glycine into the nonpolar (and relatively large) tryptophan (UGG codon) could have dramatic effects on the function of the protein.

Once again, a mutation is the change in the DNA or RNA sequence. As discussed earlier, in multicellular organisms, mutations can be subdivided into germline mutations and somatic mutations. **Germline mutations** occur in the DNA of sex cells, or gametes, and are

therefore potentially very serious. These mutations can be passed to the next generation. **Somatic mutations**, which occur in somatic, or body, cells, cannot be passed to the next generation (offspring). Mutations can be harmful, beneficial, or have no effect. If a mutation does not change the amino acid sequence in a protein, the mutation will have no effect. In fact, the overwhelming majority of mutations have no significant effect, since DNA repair mechanisms are able to mend most of the changes before they become permanent. Furthermore, many organisms have mechanisms for eliminating otherwise permanently mutated somatic cells.

A **gene pool** is the complete set of unique alleles in a species or population. Mutations create variation in the gene pool. Populations with a large gene pool are said to be genetically diverse and very robust. They are able to survive intense times of natural selection against certain phenotypes. During these times of selection, individuals with less favorable phenotypes resulting from deleterious alleles (due to mutations) may be selected against and removed from the population. Concurrently, the more favorable mutations that cause beneficial or advantageous phenotypes tend to accumulate in that population, resulting, over time, in evolution. We will discuss evolution and the genetic effects on evolution in much more detail in a later chapter.

Harmful Mutations

Mutations can result in errors in protein sequence, creating partially or completely non-functional proteins. These can obviously result in harm to the cell and organism. As discussed in the previous lesson, to function correctly and maintain homeostasis, each cell depends on thousands of proteins to all work together to perform the functions of the cell. When a mutation alters a protein that plays a critical role in the cell, the tissue, organ, or organ system may not function properly, resulting in a medical condition. A condition caused by mutations in one or more genes is called a **genetic disorder**, which will be discussed in the next chapter. However, only a small percentage of mutations cause genetic disorders; most have no impact on health. If a mutation does not change the protein sequence or structure, resulting in the same function, it will have no effect on the cell. Often, these mutations are repaired by the DNA repair system of the cell. Each cell has a number of pathways through which enzymes recognize and repair mistakes in DNA (**Figure** 8.20). Because DNA can be damaged or mutated in many ways, the process of DNA repair is an important way in which the cell protects itself to maintain proper function.

A mutation present in a germ cell can be passed to the next generation. If the zygote contains the mutation, every cell in the resulting organism will have that mutation. If the mutation results in a disease phenotype, the mutation causes what is called a hereditary disease. These will be discussed in the next chapter. On the other hand, a mutation that is present in a somatic cell of an organism will be present (by DNA replication and mitosis) in all descendants of that cell. If the mutation is present in a gene that is not used in that cell type, the mutation may have no effect. On the other hand, the mutation may lead to

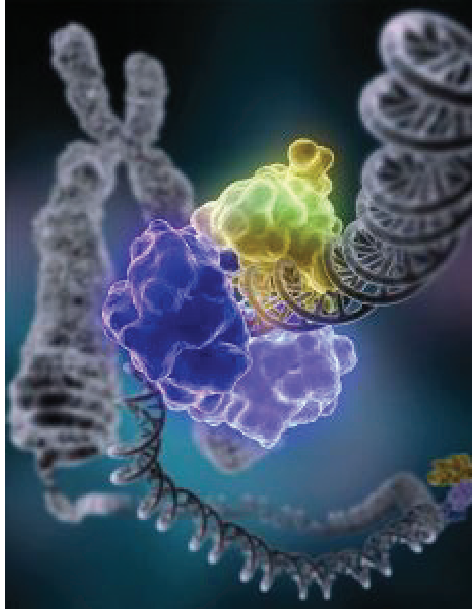

Figure 8.20: DNA repair. Shown is a model of DNA ligase repairing chromosomal damage. **DNA ligase** is an enzyme that joins broken nucleotides together by catalyzing the formation of a bond between the phosphate group and deoxyribose sugar of adjacent nucleotides in the DNA backbone. (22)

a serious medical condition such as cancer. Mutations and cancer will be discussed in the next lesson.

Beneficial Mutations

A very small percentage of all mutations actually have a positive effect. These mutations lead to new versions of proteins that help an organism and its future generations better adapt to changes in their environment. The genetic diversity that results from mutations is essential for evolution to occur. Without genetic diversity, each individual of a species would be the same, and no one particular individual would have an advantage over another. Adaptation and evolution would not be possible. **Beneficial mutations** lead to the survival of the individual best fit to the current environment, which results in evolution. This will be discussed in the evolution chapter.

Mutations and Cancer

During the discussion of the cell cycle, cancer was described as developing due to unregulated cell division. That is, **cancer** is a disease characterized by a population of cells that grow and divide without respect to normal limits. These cancerous cells invade and destroy adjacent

tissues, and they may spread throughout the body.

Nearly all cancers are caused by mutations in the DNA of the abnormal cells. These mutations may be due to the effects of **carcinogens**, cancer causing agents such as tobacco smoke, radiation, chemicals, or infectious agents. These carcinogens may act as an environmental "trigger," stimulating the onset of cancer in certain individuals and not others. Do all people who smoke get cancer? No. Complex interactions between carcinogens and an individual's genome may explain why only some people develop cancer after exposure to an environmental trigger and others do not. Do all cancers need an environmental trigger to develop? No. Cancer causing mutations may also result from errors incorporated into the DNA during replication, or they may be inherited. Inherited mutations are present in all cells of the organism.

Oncogenes and Tumor Suppressor Genes

Mutations found in the DNA of cancer cells typically affect two general classes of genes: oncogenes and tumor suppressor genes. In "normal," non-cancerous cells, the products of **proto-oncogenes** promote cell growth and mitosis prior to cell division; thus, proto-oncogenes encode proteins needed for normal cellular functions. Mutations in proto-oncogenes can modify their expression and the function of the gene product, increasing the amount of activity of the product protein. When this happens, they become oncogenes; thus, the cells have a higher chance of dividing excessively and uncontrollably. Cancer-promoting **oncogenes** are often activated in cancer cells, giving those cells abnormal properties. The products of these genes result in uncontrolled cell growth and division, protection against programmed cell death, loss of respect for normal tissue boundaries, and the ability to become established in diverse tissue environments. Proto-oncogenes cannot be removed from the genome, as they are critical for growth, repair and homeostasis. It is only when they become mutated that the signals for growth become excessive.

In "normal" cells, the products of **tumor suppressor genes** temporarily discourage cell growth and division to allow cells to finish routine functions, especially DNA repair. Tumor suppressors are generally transcription factors, activated by cellular stress or DNA damage. The function of such genes is to stop the cell cycle in order to carry out DNA repair, preventing mutations from being passed on to daughter cells. However, if the tumor suppressor genes are inactivated, DNA repair cannot occur. Tumor suppressor genes can be inactivated by a mutation that either affects the gene directly or that affects the pathway that activates the gene. The consequence of the lack of DNA repair is that DNA damage accumulates, is not repaired, and inevitably leads to cancer.

Several Mutations to Cause Cancer

Typically, a series of several mutations in these genes that activate oncogenes and inactivate tumor suppressor genes is required to transform a normal cell into a cancer cell (**Figure** 8.21). Cells have developed a number of control mechanisms to overcome mutations in proto-oncogenes. Therefore, a cell needs multiple mutations to transform into a cancerous cell. A mutation in one proto-oncogene would not cause cancer, as the effects of the mutation would be masked by the normal control of mitosis and the actions of tumor suppressor genes. Similarly, a mutation in one tumor suppressor gene would not cause cancer either, due to the presence of many ”backup” genes that duplicate its functions. It is only when enough proto-oncogenes have mutated into oncogenes and enough tumor suppressor genes have been deactivated that the cancerous transformation can begin. Signals for cell growth overwhelm the signals for growth regulation, and the cell quickly spirals out of control. Often, because many of these genes regulate the processes that prevent most damage to the genes themselves, DNA damage accumulates as one ages.

Usually, oncogenes are dominant alleles, as they contain gain-of-function mutations. Meanwhile, mutated tumor suppressors are generally recessive alleles, as they contain loss-of-function mutations. A proto-oncogene needs only a mutation in one copy of the gene to generate an oncogene; a tumor suppressor gene needs a mutation in both copies of the gene to render both products defective. There are instances when, however, one mutated allele of a tumor suppressor gene can render the other copy non-functional. These instances result in what is known as a **dominant negative effect**.

Lesson Summary

- Mutations may be due to environmental factors (mutagens) or may occur spontaneously.
- Typical mutagens include chemicals, such as those inhaled by smoking, and radiation, like X-rays, ultraviolet light, and nuclear radiation.
- Germline mutations can be passed on to descendants; somatic mutations cannot be transmitted to the next generation.
- Chromosomal alterations are large changes in the chromosome structure. There are 5 types of chromosomal alterations: deletions, duplications, insertions, inversions, and translocations.
- Point mutations occur at a single site within the DNA; examples of these include silent mutations, missense mutations, and nonsense mutations.
- A deletion or insertion in the DNA can alter the reading frame.
- Loss-of-function and gain-of-function mutations may result in altered function of the gene product or protein.
- Beneficial mutations may accumulate in a population, resulting, over time, in evolution.
- Harmful mutations can result in errors in protein sequence, creating partially or com-

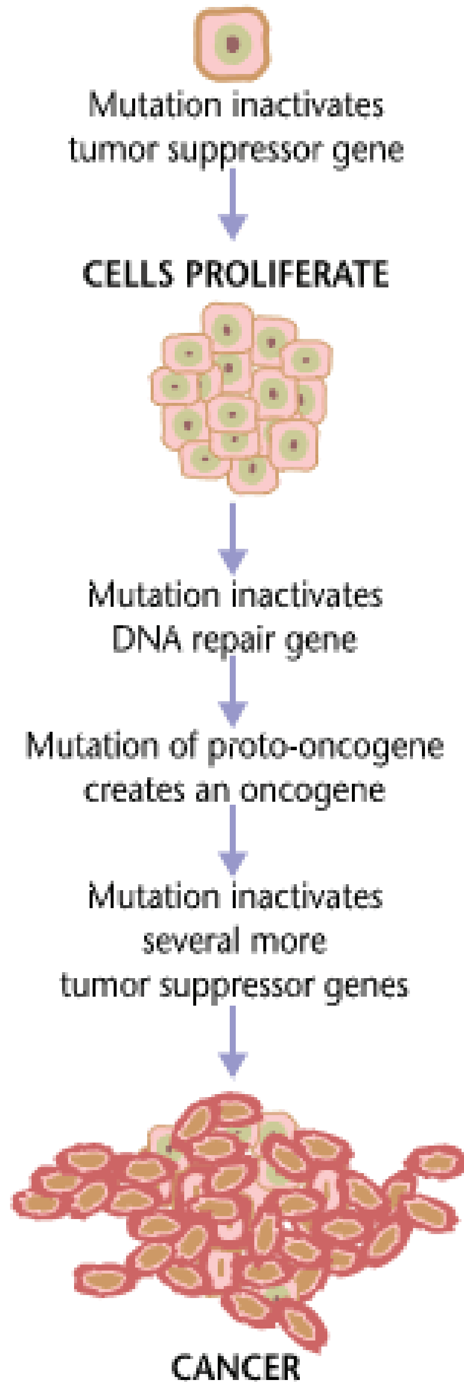

Figure 8.21: Cancers are caused by a series of mutations. Each mutation alters the behavior of the cell. In this example, the first mutation inactivates a tumor suppressor gene, the second mutation inactivates a DNA repair gene, the third mutation creates an oncogene, and a fourth mutation inactivates several more tumor suppressor genes, resulting in cancer. It should be noted that it does not necessarily require four or more mutations to lead to cancer. (7)

pletely non-functional proteins.
- Nearly all cancers are caused by mutations in the DNA of the abnormal cells.
- In non-cancerous cells, proto-oncogenes promote cell growth and mitosis prior to cell division; thus, proto-oncogenes encode proteins needed for normal cellular functions.
- In non-cancerous cells, tumor suppressor genes temporarily discourage cell growth and division to allow cells to finish routine functions, especially DNA repair.
- Mutations in proto-oncogenes and tumor suppressor genes may lead to cancer.
- Usually mutations in multiple genes are necessary to develop cancer.

Review Questions

1. Define mutation.
2. What are some common causes of mutations?
3. List some common types of mutations.
4. Describe some common chromosomal alterations.
5. Discuss potential consequences of point mutations, deletions and insertions.
6. List and describe three common types of point mutations.
7. What are effect-on-function mutations?
8. What is a germline mutation? A somatic mutation?
9. Explain why some mutations are harmful and some beneficial.

Further Reading / Supplemental Links

- Campbell, N.A. and Reece, J.B. *Biology*, Seventh Edition, Benjamin Cummings, San Francisco, CA, 2005.
- Biggs, A., Hagins, W.C., Kapicka, C., Lundgren, L., Rillero, P., Tallman, K.G., and Zike, D., *Biology: The Dynamics of Life*, California Edition, Glencoe Science, Columbus, Oh, 2005.
- Nowicki S., *Biology*, McDougal Littell, Evanston, IL, 2008.
- Kimball's Biology Pages:
- http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/
- How Cancer Starts:
- http://www.cancerhelp.org.uk/help/default.asp?page=97
- Mutations and evolution:
- http://www.nwcreation.net/geneticrecombination.html
- http://en.wikipedia.org

Vocabulary

allele An alternative form of a gene.

beneficial mutation A mutation that leads to a new version of a protein that helps an organism and its future generations better adapt to changes in their environment.

cancer A disease characterized by a population of cells that grow and divide without respect to normal limits.

carcinogen Cancer causing agent, such as tobacco smoke, radiation, chemicals, or infectious agents.

chromosomal alterations Large changes in chromosome structure.

deamination A mutation due to the spontaneous deamination of 5-methycytosine.

deletion Removal of a large chromosomal region, leading to loss of the genes within that region; also the removal of one or more nucleotides from DNA.

depurination A mutation due to the loss of a purine base (A or G).

DNA ligase An enzyme that joins broken nucleotides together by catalyzing the formation of a bond between the phosphate group and deoxyribose sugar of adjacent nucleotides in the DNA backbone.

dominant negative mutation Mutation that results in an altered gene product that acts in a dominant manner to the wild-type gene product.

duplication Leads to multiple copies of a chromosomal region, increasing the number of the genes located within that region; also know as amplification.

frameshift mutation Mutations which alter the mRNA reading frame.

gain-of-function mutation Mutation that results in the gene product or protein having a new and abnormal function.

gene pool The complete set of unique alleles in a species or population.

genetic disorder A condition caused by mutations in one or more genes.

germline mutation Mutation in the DNA within a gamete; can be passed on to descendents.

insertion Chromosomal alteration involving the addition of material from one chromosome to a nonhomologous chromosome; also a mutation which adds one or more nucleotides into the DNA.

inversion Chromosomal alteration reversing the orientation of a chromosomal segment.

loss-of-function mutation Mutation that results in a gene product or protein having less or no function.

missense mutations Point mutation which codes for a different amino acid.

mutagen An environmental factor which causes a mutation; includes certain chemicals and radiation.

mutation A change in the DNA or RNA sequence.

nonsense mutation Point mutation which codes for a premature stop codon.

oncogene Cancer promoting gene; the products of these genes result in uncontrolled cell growth and division, protection against programmed cell death, loss of respect for normal tissue boundaries, and the ability to become established in diverse tissue environments.

point mutations Exchange one nucleotide for another; known as base substitution mutations.

proto-oncogenes Genes whose products promote cell growth and mitosis prior to cell division.

silent mutations Point mutation which codes for the same amino acid.

somatic mutation A mutation in a body cell, not in a gamete; cannot be transmitted to the next generation.

splice site mutation Mutation in the coding region of a gene that alters splicing of the mRNA.

tautomerism A mutation due to the changing of a base by the repositioning of a hydrogen atom.

transition A purine to purine, or a pyrimidine to pyrimidine change.

translocation Chromosomal alterations involving the interchange of genetic material between nonhomologous chromosomes.

transversion A purine is replaced by a pyrimidine, or a pyrimidine is replaced by a purine.

tumor suppressor gene Gene whose product temporarily discourage cell growth and division to allow cells to finish routine functions, especially DNA repair.

Points to Consider

- Now that we have discussed DNA, protein synthesis and mutations, can you think of a mechanism that allows different cell types to have different proteins?
- What about during development? Why does a developing embryo need different proteins at different times of development?
- We have discussed oncogenes and tumor suppressor genes. Can you think of a specific cellular mechanism in which defects in these genes lead to cancer?

8.4 Lesson 8.4: Regulation of Gene Expression

Lesson Objectives

- Describe general mechanisms of gene expression.
- Differentiate between a cis-regulatory element and a trans-acting factor.
- Define a transcription factor.
- Define an operon.
- Describe how the lac operon regulates transcription.
- Describe the role of the TATA box.
- Express the importance of gene regulation during development.
- Describe the role of homeobox genes and gap genes.
- Discuss gene regulation in terms of the development of cancer.

Introduction

Each of your cells has about 22,000 genes. In fact, all of your cells have the same genes. So do all of your cells make the same proteins? Do all 22,000 genes get turned into proteins in every cell? Of course not. If they did, then all your cells would do the same thing. You have cells with different functions because you have cells with different proteins. And your cells

have different proteins because they "use" different genes. The regulation of gene expression, or gene regulation, includes the mechanism to turn genes "on" and transcribe the gene into RNA. Any aspect of a gene's expression may be regulated, from the onset of transcription to the post-translational modification of a protein. It is this regulation that determines when and how much of a protein to make, giving a cell its specific structure and function.

Mechanisms of Regulation

Any step of gene expression may be modulated, from the DNA-RNA transcription step to post-translational modification of a protein. Following is a list of stages where gene expression is regulated:

- Chemical and structural modification of DNA or chromatin
- Transcription
- Translation
- Post-transcriptional modification
- RNA transport
- mRNA degradation
- Post-translational modifications

In this lesson, we will focus on regulation at the level of transcription. We have previously discussed that during transcription RNA polymerase reads the DNA template to make a complementary strand of RNA. The genes to which RNA polymerase binds is a highly regulated process. When RNA polymerase binds to a gene, it binds to the **promoter**, a segment of DNA that allows a gene to be transcribed. The promoter helps RNA polymerase find the start of a gene.

Gene regulation at the level of transcription controls when transcription occurs as well as how much RNA is created. This regulation is controlled by **cis-regulatory elements** and **trans-acting factors**. A cis-regulatory element is a region of DNA which regulates the expression of a gene or multiple genes located on that same strand of DNA. These cis-regulatory elements are often the binding sites of one or more trans-acting factors, usually a regulatory protein which interacts with RNA polymerase. A cis-regulatory element may be located in a gene's promoter region, in an intron, or in the 3' region.

A **regulatory protein**, or a **transcription factor**, is a protein involved in regulating gene expression. It is usually bound to a cis-regulatory element. Regulatory proteins often must be bound to a cis-regulatory element to switch a gene on (activator), or to turn a gene off (repressor).

Transcription of a gene by RNA polymerase can be regulated by at least five mechanisms:

- **Specificity factors** (proteins) alter the specificity of RNA polymerase for a promoter

or set of promoters, making it more or less likely to bind to the promoter and begin transcription.

- **Repressors** (proteins) bind to non-coding sequences on the DNA that are close to or overlap the promoter region, impeding RNA polymerase's progress along the strand.
- **Basal factors**, transcription factors that help position RNA polymerase at the start of a gene.
- Enhancers are sites on the DNA strand that are bound by activators in order to loop the DNA, bringing a specific promoter to the initiation complex. An **initiation complex** is composed of RNA polymerase and trans-acting factors.
- **Activators** (proteins) that enhance the interaction between RNA polymerase and a particular promoter.

As the organism grows more sophisticated, gene regulation becomes more complex, though prokaryotic organisms possess some highly regulated systems. Some human genes are controlled by many activators and repressors working together. Obviously, a mutation in a cis-regulatory region, such as the promoter, can greatly affect the proper expression of a gene. It may keep the gene permanently off, such that no protein can be made, or it can keep the gene permanently on, such that the corresponding protein is constantly made. Both of these can have detremental effects on the cell.

Prokaryotic Gene Regulation

In prokaryotes, a combination of activators and repressors determines whether a gene is transcribed. As you know, prokaryotic organisms are fairly simple organisms with much less DNA. Prokaryotic genes are arranged in **operons**, a region of DNA with a promoter, an operator (defined below), and one or more genes that encode proteins needed to perform a certain task. To maintain homeostasis (and survive), the organism must quickly adapt changing environmental conditions. The regulation of transcription plays a key role in this process.

For a bacteria, many aspects of gene regulation are due to the presence or absence of certain nutrients. In prokaryotes, repressors bind to regions called **operators** that are generally located immediately downstream from the promoter. Activators bind to the upstream portion of the promoter.

The Lac Operon

The **lac operon** (**Figure** 8.22) is an operon required for the transport and metabolism of lactose in *E. coli*. The lac operon is regulated by the availability of lactose. The lac operon consists of a promoter, an operator, three adjacent structural genes which code for enzymes and a terminator. The three genes are: lacZ, lacY, and lacA. All three genes are controlled by the same regulatory elements.

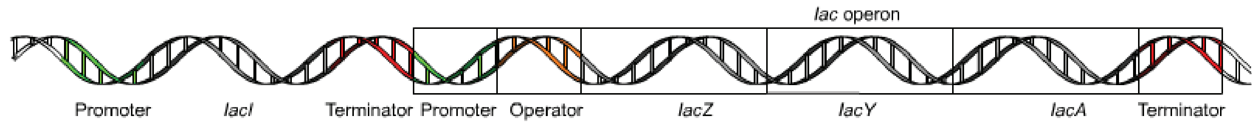

Figure 8.22: The lac operon. The lac operon contains genes for three enzymes, lac, lacY, and lac A, as well as the promoter, operator, and terminatory regulatory regions. (10)

In bacteria, the lac repressor protein blocks the synthesis of enzymes that digest lactose when there is no lactose present (**Figure** 8.23). When lactose is present, it binds to the repressor, causing it to detach from the DNA strand.

Specific control of the lac operon depends on the availability of lactose. The enzymes needed to metabolize lactose are not produced when lactose is not present. When lactose is available, and therefore needs to be metabolized, the operon is turned on, RNA polymerase binds to the promoter, and the three genes are transcribed into a single mRNA molecule. However, if lactose is not present (and therefore does not need to be metabolized), the operon is turned off by the lac repressor protein (**Figure** 8.23).

The lacI gene, which encodes the lac repressor, lies near the lac operon and is always expressed (constitutive). Therefore, the lac repressor protein is always present in the bacteria. In the absence of lactose, the lac repressor protein will bind to the operator, just past the promoter in the lac operon. The repressor blocks the binding of RNA polymerase to the promoter, keeping the operon turned off (**Figure** 8.23).

When lactose is available, a lactose metabolite called allolactose binds to the repressor. This interaction causes a conformational change in the repressor shape and the repressor falls off the operator, allowing RNA polymerase to bind to the promoter and initiate transcription (**Figure** 8.23).

Eukaryotic Gene Regulation

As previously discussed, all your cells have the same DNA (and therefore the same genes), yet they have different proteins because they express different genes. In eukaryotic cells, the start of transcription is one of the most complex aspects of gene regulation. Transcriptional regulation involves the formation of an initiation complex involving interactions between a number of transcription factors, *cis*-regulatory elements, and **enhancers**, distant regions of DNA that can loop back to interact with a gene's promoter. These regulatory elements occur in unique combinations within a given cell type, resulting in only necessary genes being transcribed in certain cells. Transcription factors bind to a DNA strand, allowing RNA polymerase to bind and start transcription.

Each gene has unique cis-regulatory sequences, only allowing specific transcription factors to bind. However, there are common regulatory sequences found in most genes. The **TATA**

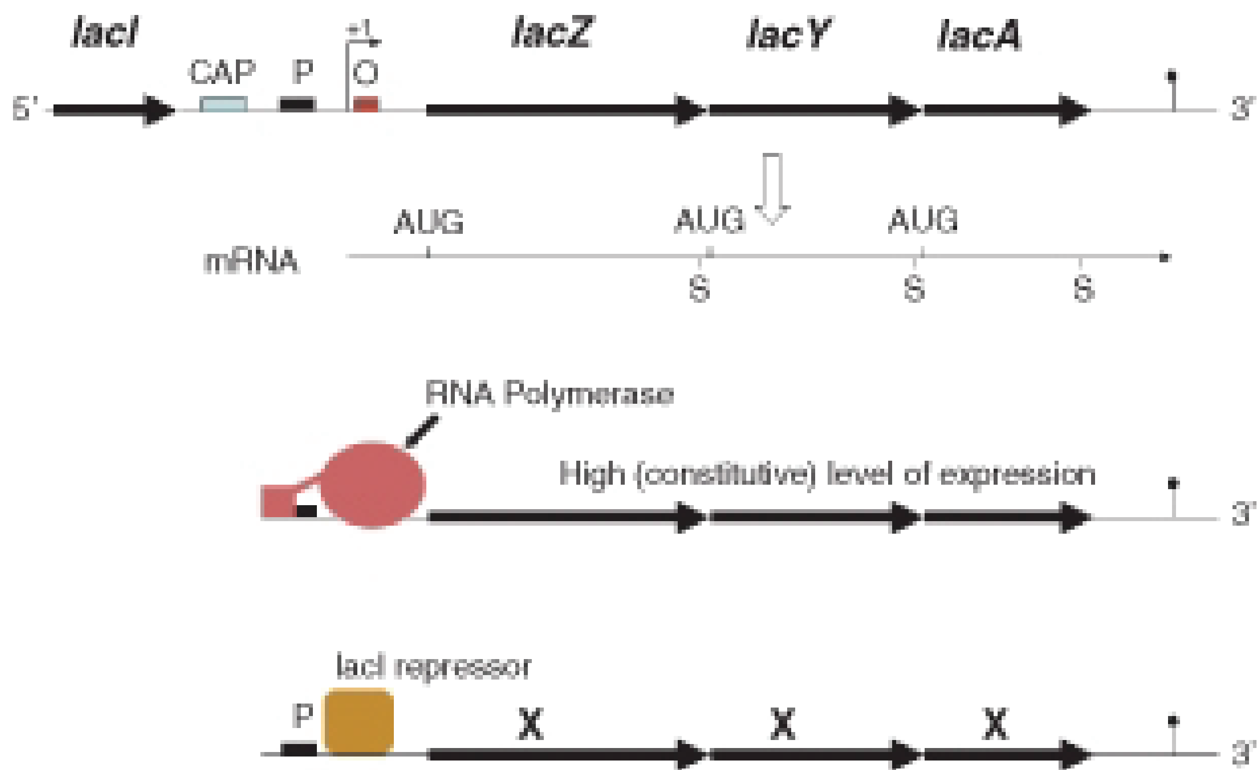

Figure 8.23: Regulation of the lac operon. When lactose is present, **RNA polymerase** (red) binds to the promoter (P) and the three genes are expressed, producing a single mRNA for the three genes. When lactose is unavailable, the lac repressor (yellow) binds to the operator (O) and inhibits the binding of RNA polymerase to the promoter. The three genes are not expressed. For an animation of the Lac Operon, see http://vcell.ndsu.edu/animations/lacOperon/first.htm. (15)

box is a *cis*-regulatory element found in the promoter of most eukaryotic genes. It has the DNA sequence 5'-TATAAA-3' or a slight variant, and has been highly conserved throughout evolution. When the appropriate cellular signals are present, RNA polymerase binds to the TATA box, completing the initiation complex. A number of transcription factors first bind to the TATA box while other transcription factors bind to the previously attached factors, forming a multi-protein complex. It is only when all the appropriate factors are bound that RNA polymerase will recognize the complex and bind to the DNA, initiating transcription.

One of the more complex eukaryotic gene regulation processes is during development. What genes must be turned on during development so that tissues and organs form from simple cells?

Regulation of Gene Expression During Development

What makes the heart form during development? What makes the skin form? What makes a structure become an arm instead of a leg? These processes occur during development because of a highly specific pattern of gene expression. This intensely regulated pattern of gene expression turns genes on in the right cell at the right time, such that the resulting proteins can perform their necessary functions to ensure proper development. Transcription factors play an extremely important role during development. Many of these proteins can be considered master regulatory proteins, in the sense that they either activate or deactivate the transcription of other genes and, in turn, these secondary gene products can regulate the expression of still other genes in a regulatory cascade. Homeobox and gap genes are important transcription factors during development.

Homeobox Genes

Homeobox genes contain a highly conserved DNA sequence known as a homeobox and are involved in the regulation of genes important to development. A homeobox is about 180 base pairs long; it encodes a 60 amino acid domain within the protein (known as the homeodomain), which can bind DNA. Proteins with a homeodomain are therefore transcription factors. These factors typically switch on series of other genes, for instance, the genes needed to encode the proteins to make a leg.

A particular subgroup of homeobox genes are the **Hox genes**. Protein products of Hox genes function in patterning the body, providing the placement of certain body parts during development. In other words, Hox genes determine where limbs and other body segments will grow in a developing fetus or larva. Mutations in any one of these genes can lead to the growth of extra, typically non-functional body parts in invertebrates. The Antennapedia mutation in *Drosophila* results in a leg growing from the head in place of an antenna. A mutation in a vertebrate Hox genes usually results in miscarriage.

Gap Genes

A **gap gene** controls the shape of a developing zygote early in its development. The products of these genes produce gaps in a rather uniform arrangement of cells (**Figure** 8.24). One example of this is the Kruppel gene, which regulates the activity of a number of other genes. Gap genes encode transcription factors, and the Kruppel gene is a zinc-finger protein. A **zinc finger** is a DNA binding region within the protein. A zinc finger consists of two antiparallel sheets and an helix with a zinc ion, which is important for the stability of this region. Gap genes control the expression of other genes within specific regions of cells in the developing organism. This allows specific genes to be expressed in certain cells at the appropriate stage of development.

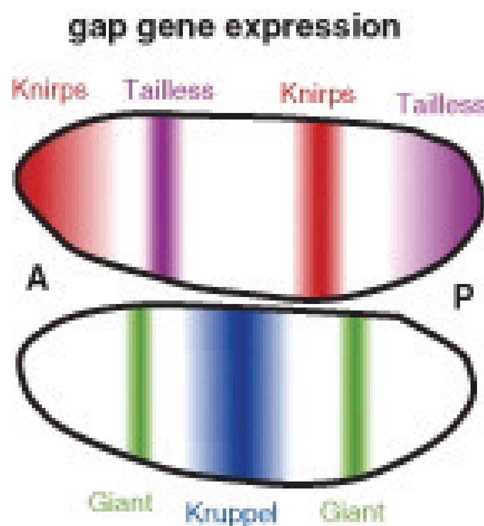

Figure 8.24: Gap gene expression. Shown is the expression pattern of four gap genes, Kruppel, Giant, Knirps, and Tailless, in a developing *Drosophilia* embryo. Note how the expression of these genes creates an unique pattern resulting in gaps in what was a rather uniform arrangement of cells. (25)

Regulation of Gene Expression in Cancer

As discussed in the last lesson, carcinogenesis depends on both the activation of oncogenes and deactivation of tumor suppressor genes. At least two separate mutations are necessary to develop cancer. For example, a mutation in a proto-oncogene would not necessarily lead to cancer, as normally functioning tumor suppressor genes would counteract the effects of the oncogene. It is the second mutation in the tumor suppressor gene that could lead to uncontrolled cell growth and possibly cancer. Both oncogenes and tumor suppressor genes play an important role in gene regulation and cell proliferation (**Figure** 8.25).

Oncogenes

The products of proto-oncogenes are required for normal growth, repair and homeostasis. However, when these genes are mutated, they turn into oncogenes and play a role in the development of cancer. Proto-oncogenes may be growth factors, transcription factors, or other proteins involved in regulation. A very common oncogene, ras, is normally a regulatory GTPase that switches a signal transduction chain on and off. Ras and Ras-related proteins are products of oncogenes found in 20% to 30% of human tumors.

Ras is a **G protein**, a regulatory GTP hydrolase that cycles between an activated and inactivated form. When a growth factor binds to its receptor on the outside of the cell, a signal is relayed to RAS. As a G protein, Ras is activated when GTP is bound to it. The active Ras then passes the signal to a series of protein kinases, regulatory proteins that eventually activate transcription factors to alter gene expression and produce proteins that stimulate the cell cycle (**Figure** 8.25). Many of the genes and proteins involved in signal transduction pathways are interconnected to ras. Any mutation that makes ras more active or otherwise interrupts the normal signal transduction pathways (**Figure** 8.25) may result in excessive cell division and cancer.

Tumor Suppressor Genes

An example of a tumor suppressor gene is p53, which encodes a 53,000 dalton protein, The p53 gene is activated by DNA damage. DNA may be damaged by ultraviolet light, and any damaged DNA may be harmful to the cell. Mutations causing problems with any of the components of **Figure** 8.25 may lead to the development of cancer. So that damaged DNA is not replicated, the cell cycle must be temporarily stopped so that the DNA can be repaired. The p53 tumor suppressor gene encodes a transcription factor that regulates the synthesis of cell cycle inhibiting proteins (**Figure** 8.25). p53 often activates a gene named p21, whose protein product temporarily stops the cell cycle. If the DNA can not be repaired, p53 activates other genes that lead to cell death, or apoptosis. This prevents the cell from passing on damaged DNA. If the p53 tumor suppressor gene is defective, as by mutation, DNA damage in the cell may accumulate and the cell may survive to replicate the damaged DNA. The damaged DNA would then be passed to other cells through many cell divisions, and cancer could develop.

Lesson Summary

- A cis-regulatory element is a region of DNA which regulates the expression of a gene or multiple genes located on that same strand of DNA.
- The cis-regulatory elements are often binding sites for one or more trans-acting factors, usually a regulatory protein which interacts with RNA polymerase.
- Repressor proteins bind to non-coding sequences on the DNA that are close to or

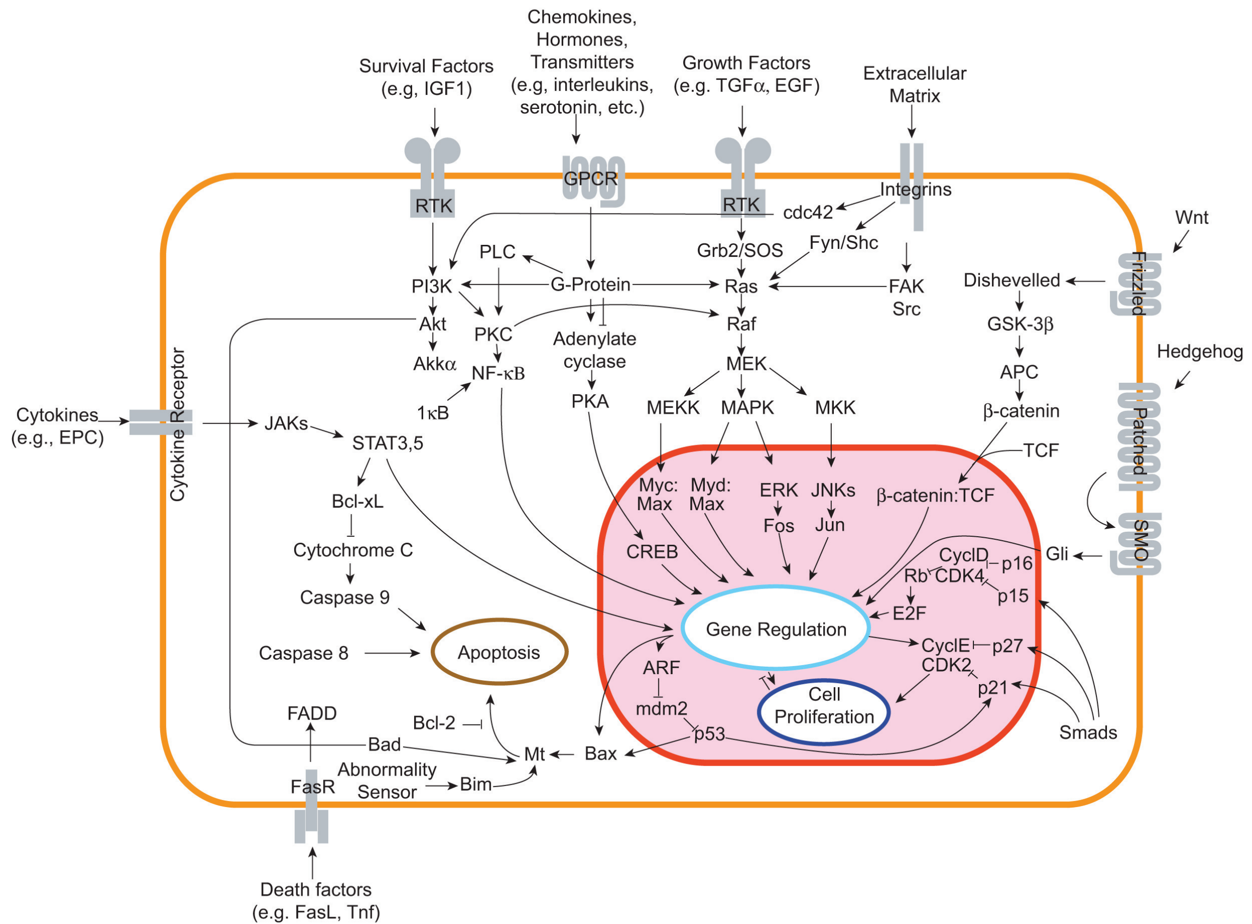

Figure 8.25: Signal transduction pathways. Ras (upper middle section) activates a number of pathways but an especially important one seems to be the mitogen-activated protein kinases (MAPK). MAPK transmit signals downstream to other protein kinases and gene regulatory proteins. Note that many of these pathways are initiated when a signal binds to its receptor outside the cell. Most pathways end with altered gene regulation and cell proliferation. The p53 tumor suppressor protein is shown at the lower section of the figure stimulating p21. The complexity of the pathways demonstrate the significant role these play in the cell. (18)

overlap the promoter region, impeding RNA polymerase's progress along the strand.
- Enhancers are sites on the DNA strand that are bound by activators.
- Prokaryotic genes are arranged in operons, regions of DNA with a promoter, an operator, and one or more genes that encode proteins needed to perform a certain task.
- The regulation of the lac operon is a key example of prokaryotic gene regulation. When lactose is present, RNA polymerase binds to the promoter and the operon is turned on; when lactose is unavailable, the lac repressor binds to the operator and the operon is turned off.
- Homeobox genes are involved in the regulation of genes important to development. They encode transcription factors.
- Gap genes control the shape of a developing zygote early in its development.
- At least two separate mutations are necessary to develop cancer. These mutations may occur in proto-oncogenes and/or tumor suppressor genes.

Review Questions

1. What is meant by gene expression, and why is this an important cellular mechanism?
2. How do cis-regulatory elements and a trans-acting factors work together?
3. Define a transcription factor.
4. Define an operon. Give an example.
5. Describe how the lac operon regulates transcription.
6. Describe the role of the TATA box.
7. Why is gene regulation an important aspect of development?
8. What is a homeobox gene? A gap gene? Why are these genes important?
9. Why does altered gene regulation have a potential role in the development of cancer?

Further Reading / Supplemental Links

- Campbell, N.A. and Reece, J.B. *Biology*, Seventh Edition, Benjamin Cummings, San Francisco, CA, 2005.
- Biggs, A., Hagins, W.C., Kapicka, C., Lundgren, L., Rillero, P., Tallman, K.G., and Zike, D., *Biology: The Dynamics of Life*, California Edition, Glencoe Science, Columbus, Oh, 2005.
- Nowicki S., *Biology*, McDougal Littell, Evanston, IL, 2008.
- http://en.wikipedia.org

Vocabulary

activator Protein that enhances the interaction between RNA polymerase and a particular promoter.

basal factor Transcription factor that helps position RNA polymerase at the start of a gene.

cis-regulatory element A region of DNA which regulates the expression of a gene or multiple genes located on that same strand of DNA.

enhancer Site on the DNA strand that can be bound by activator(s) in order to loop the DNA, bringing a specific promoter to the initiation complex.

G protein A regulatory GTP hydrolase that cycles between an activated and inactivated form; activated when GTP is bound to it.

gap gene Gene that controls the shape of a developing zygote early in its development; encodes transcription factors.

homeobox A 180 base pair long highly conserved segment of DNA; encodes a 60 amino acid domain within the protein (known as the homeodomain), which can bind DNA.

homeobox genes Genes that contain a highly conserved DNA sequence known as a homeobox and are involved in the regulation of genes important to development; encodes transcription factors.

hox genes Genes that function in patterning the body, providing the placement of certain body parts during development.

initiation complex Complex needed to start transcription in eukaryotes; composed of RNA polymerase and trans-acting factors.

lac operon An operon required for the transport and metabolism of lactose in *E. coli*.

operator A region of prokaryotic DNA where a repressors binds.

operon A region of prokaryotic DNA with a promoter, an operator, and one or more genes that encode proteins needed to perform a certain task.

promoter A segment of DNA that allows a gene to be transcribed; helps RNA polymerase find the start of a gene.

repressor Protein that binds to non-coding sequences on the DNA that are close to or overlap the promoter region, impeding RNA polymerase's progress along the strand.

RNA polymerase The enzyme that transcribes the DNA to make RNA.

specificity factor Protein that alters the specificity of RNA polymerase for a promoter or set of promoters, making it more or less likely to bind to the promoter and begin transcription.

TATA box A *cis*-regulatory element found in the promoter of most eukaryotic genes; when the appropriate cellular signals are present, RNA polymerase binds to the TATA box, completing the initiation complex.

transcription factor A protein involved in regulating gene expression; usually bound to a cis-regulatory element; also known as a regulatory protein.

zinc finger A DNA binding region within certain proteins encoded by a gap gene; consists of two antiparallel sheets and an helix with a zinc ion, which is important for the stability of this region.

Points to Consider

- The next chapter is *Human Genetics*. Discuss why an understanding of human genetics is an important medical issue for our society.
- We have extensively discussed mutations and cancer. There are many other phenotypes due to mutations in the human genome. Why is understanding mutations in humans important?
- What do you think the Human Genome Project is? What could some implications of The Human Genome Project be?

Image Sources

(1) CK-12 Foundation. . CC-BY-SA.

(2) http://en.wikipedia.org/wiki/Image:Griffith_experiment.svg. GNU-FDL.

(3) http://en.wikipedia.org/wiki/Image:Pre-mRNA_to_mRNA.png. GNU-FDL.

(4) CK-12 Foundation.
http://commons.wikimedia.org/wiki/Image:Transcription.png. Public Domain.

(5) http://en.wikipedia.org/wiki/Image:Schema_ARNt_448_658.png. GNU-GPL.

(6) *Summary of translation.*. Public Domain.

(7) NIH. http://en.wikipedia.org/wiki/Image:Cancer_requires_multiple_mutations_from_NIH.png. Public Domain.

(8) http://www.genome.gov//Pages/Hyperion/DIR/VIP/Glossary/Illustration/mutation.cfm. Public Domain.

(9) http://en.wikipedia.org/wiki/Image:Tevenphage.png. CC-BY-SA 2.5.

(10) http://en.wikipedia.org/wiki/Image:Lac_operon1.png. Public Domain.

(11) http://en.wikipedia.org/wiki/Image:NA-comparedto-DNA_thymineAndUracilCorrected.png. CC-BY-SA.

(12) CK-12 Foundation. http://en.wikipedia.org/wiki/Image:HersheyChaseEx.png. CC-BY-SA 2.5.

(13) CK-12 Foundation. *http://en.wikipedia.org/wiki/Image:Guanine_chemical_structure.png http://en.wikipedia.org/wiki/Image:Cytosine_chemical_structure.png http://en.wikipedia.org/wiki/Image:Thymine.png*. GNU-FDL.

(14) NIH. *http://en.wikipedia.org/wiki/Image:JamesDWatson.jpg*. CC-BY 2.5,Public Domain.

(15) http://en.wikipedia.org/wiki/Image:Lac_operon.png. GFDL.

(16) http://en.wikipedia.org/wiki/Image:3d_tRNA.png. CC-BY-SA 3.0.

(17) http://commons.wikimedia.org/wiki/File:DNA_replication_en.svg. Public Domain.

(18) http://en.wikipedia.org/wiki/Image:Signal_transduction_v1.png. GNU-FDL.

(19) *The four stages of protein folding.*. Public Domain.

(20) http://en.wikipedia.org/wiki/Image:DNA_Overview.png. GNU-FDL.

(21) http://en.wikipedia.org/wiki/Image:RNA-codons.png. Public Domain.

(22) http://en.wikipedia.org/wiki/Image:DNA_Repair.jpg. Public Domain.

(23) http://en.wikipedia.org/wiki/Genetic_code. CC-BY-SA.

(24) CK-12 Foundation. http://en.wikipedia.org/wiki/Image:DNA_chemical_structure.svg. GNU-FDL.

(25) http://commons.wikimedia.org/wiki/Image:Gap_ene_expression.png. Public Domain.