7 - Mutations and DNA Repair

The function of DNA is to store genetic information. To store genetic information effectively, it is important that the sequence of DNA remains unchanged from generation to generation. However, rare mistakes occur during DNA replication; recall one mistake is made every 10 to 100 million nucleotides incorporated into daughter DNA strands in bacteria. Further, environmental agents, such as ionizing radiation, ultraviolet light, and a myriad of chemicals can damage DNA. DNA damage is often detrimental; DNA damage worsens protein activity and negatively affects phenotype. As a result, it may seem that it would be beneficial to fix all replication errors or DNA damage: however, changes in the DNA sequence can be advantageous in some cases. Alterations in the DNA sequence can create new gene variants (alleles) in a population. Evolution requires this diversity of alleles; natural selection chooses which allele combinations survive and reproduce in a particular environment. For this reason, a low level of mutation is required for evolutionary change.

In this section, we will define mutation, discuss various ways mutations can be classified, learn the mechanisms used by a bacterial cell to repair damaged DNA, and discuss situations in which defective DNA repair leads to disease in humans.

A. Mutations

What is a mutation?

A **mutation** is a change in the genetic material that is inherited by daughter cells at the conclusion of cell division. In most cases, an incorrect nucleotide is replaced with the correct one before the cell divides; therefore, based on the definition above, this type of change would not be considered a mutation. Only nucleotide changes that remain after cell division are mutations.

When environmental agents (ionizing radiation, ultraviolet light) damage the DNA double-helix, **induced mutations** occur. In other cases, DNA polymerases inadvertently incorporate an incorrect nucleotide into the daughter DNA strand during DNA replication. This DNA replication error is called a **spontaneous mutation**.

Key Questions

- What is the definition of mutation?
- What is the difference between an induced and a spontaneous mutation?

Germline and somatic mutations

What determines if a mutation is inherited by offspring? **Gametes cells** arise from specialized cells called **germline cells**. In contrast, the non-gamete cells within the body (muscle cells, neurons, epithelial cells, etc.) are **somatic cells**. Therefore, when a mutation arises in a germline cell (**germline mutation**), the mutation is transmitted via the gametes to

the individual's offspring. In contrast, a mutation generated in a somatic cell (**somatic mutation**) is not transmitted to the individual's offspring.

The timing of the mutation can also influence its transmission to offspring. If a mutation arises in a cell during early embryonic development, then chances are good that the mutation will be found in both the germline and somatic cells. Since this mutation is found in germline cells, the mutation will then be transmitted to the next generation. However, when a mutation arises late in development or in the adult organism, the mutation may only be found in one type of somatic cell and will be absent from the germline cells. As a result, this somatic mutation will not be transmitted to the offspring.

The timing of the mutation can also affect the severity of the phenotype. If a somatic mutation arises early in the development of a particular organ, it is likely that the majority of cells that make up the organ will contain the mutation. When an organ is made up of mostly mutant cells, the phenotype of the organ is impacted, often in a negative way. Alternatively, if only a subset of cells in the tissue contains the mutation, the phenotypic effect can be comparatively mild.

Key Questions

- What is the difference between a germline and a somatic mutation?
- How likely is a germline mutation transmitted from parents to their offspring? How likely is a somatic mutation transmitted from parents to their offspring?
- If a mutation arises early in development, is it more or less likely to be transmitted to the germline?

Point Mutations

Sometimes a mutation alters a single nucleotide in the DNA molecule. When only one nucleotide has been altered, the mutation is classified as a **point mutation**. When a purine nitrogenous base is changed to another purine (e.g., adenine to guanine) or a pyrimidine nitrogenous base is exchanged for another pyrimidine (e.g., cytosine to thymine), the point mutation is a **transition** mutation. Alternatively, when a purine is exchanged for a pyrimidine (or vice versa), the point mutation is a **transversion** mutation.

Silent, Missense, and Nonsense Mutations

When the mutation occurs in the intergenic or the repetitive DNA sequences within chromosomes, the mutation often has no or mild phenotypic consequences. Even though these mutations outside of the structural genes do not alter the cell's phenotype, these mutations are useful as they define the **DNA fingerprint** of an individual. Each person's DNA fingerprint contains a unique collection of mutations within the intergenic and repetitive DNA sequences (see Part 1).

We will focus our attention on mutations that occur within structural genes. Remember that structural genes are transcribed into messenger RNA (mRNA) molecules. The mRNA is then translated when the ribosome reads three nucleotide-long **codons** to generate the amino acid sequence of the encoded protein (see Part 11). When one nucleotide has been exchanged for another nucleotide, but there is no change in the amino acid sequence of the encoded protein, we refer to the mutation as a **silent** mutation (see **Figure 7.1**). Silent mutations occur because the genetic code is **degenerate**, that is, more than one codon specifies the same amino acid.

WILD-TYPE   5′   AUG . UUC . GUG . CAC . UUA . AUC . UAG   3′

              MET . PHE . VAL . HIS . LEU . ILE . STOP

SILENT   5′   AUG . UU**U** . GUG . CAC . UUA . AUC . UAG   3′

MET . **PHE** . VAL . HIS . LEU . ILE . STOP

MISSENSE   5'   AUG . UUC . GUG . **A**AC . UUA . AUC . UAG   3'

MET . PHE . VAL . **ASN** . LEU . ILE . STOP

NONSENSE   5'   AUG . UUC . GUG . CAC . U**A**A . AUC . UAG   3'

MET . PHE . VAL . HIS . **STOP**

FIGURE 7.1 Base Substitutions Can Affect Gene Structure and Function. The sequence of the wild-type RNA is indicated, with the amino acid sequence of the translated protein shown below it. In each mutation, the effect on the amino acid sequence of the encoded protein is indicated in bold. Silent mutations do not change the encoded amino acid. Missense mutations change a single amino acid in the encoded protein. Nonsense mutations change a codon into a stop codon.

When a point mutation substitutes a single amino acid, the mutation is a **missense** mutation. Missense mutations can sometimes cause disease. For example, **sickle cell anemia** is an autosomal recessive disease caused by a missense mutation in the structural gene that makes the beta globin protein subunits within hemoglobin. In sickle cell anemia,

the missense mutation in the beta globin gene replaces the amino acid glutamic acid in codon six with the amino acid valine.

Sometimes a point mutation results in the formation of a premature stop codon (**nonsense mutation**). Nonsense mutations are usually more severe in phenotype than missense mutations because nonsense mutations cause the encoded protein to be shorter in length than normal. These shorter proteins are often nonfunctional.

Frameshift Mutations

In some cases, nucleotides are inserted into the sequence of a structural gene (i.e., an **insertion** mutation) or nucleotides are **deleted** from the sequence of the structural gene. Because the ribosome reads the encoded mRNA one codon at a time during translation (see Part 11), some insertions or deletions will change the **reading frame** of the mRNA molecule. When the reading frame changes, the entire amino acid sequence of the encoded protein changes from the point of the insertion/deletion (**indel**) site onward. A mutation that changes the reading frame is called a **frameshift mutation** (see **Figure 7.2**).

WILD-TYPE   5'   AUG . UUC . GUG . CAC . UUA . AUC . UAG   3'

MET . PHE . VAL . HIS . LEU . ILE . STOP

FRAMESHIFT   5'   AUG . UUC . G**C**U . GCA . CUU . AAU . CUA . G   3'

MET . PHE . **ALA** . **ALA** . **LEU** . **ASN** . **LEU** ….

FIGURE 7.2 Frameshift Mutations Can Affect Gene Structure and Function. By inserting a single cytosine base in the third codon (underlined), a reading frame shift occurs that changes the amino acid sequence of the protein and eliminates the stop codon.

B. DNA Repair Systems

Most alterations to the DNA sequence are recognized by cellular **DNA repair systems** immediately. Some of these DNA repair systems recognize changes in the width of the DNA double-helix. For example, the DNA double-helix always has a purine (adenine or guanine) paired with a pyrimidine (cytosine or thymine). Purine-pyrimidine base pairing results in a DNA molecule that is 2 nanometers (nm) wide. Repair enzymes run along the length of double stranded DNA as it is replicating, ensuring that the replicated DNA is 2 nm wide. If two purines are paired, a bulge in the DNA double-helix occurs, and if there are two pyrimidines paired, the double-helix narrows. Variations in the width of the DNA double-helix signal that damage has occurred, and DNA repair is required.

There are multiple DNA repair systems because there are different types of mutations to correct. Even though each DNA repair system has different components and recognizes a different aberration in the DNA, there are some features in common among the DNA repair systems. First, each pathway contains an enzyme that recognizes the DNA damage. Next, the DNA damage and often a few extra nucleotides are removed from one of the two DNA strands. Finally, the excised nucleotides are replaced by a DNA polymerase and the final phosphodiester bond in the damaged DNA strand is formed by DNA ligase.

There are six DNA repair systems:

1. **Proofreading**
2. **Mismatch repair**
3. **Base excision repair (BER)**
4. **Nucleotide excision repair (NER)**
5. **Homology directed repair (HDR)**
6. **Non-homologous end joining (NHEJ)**

You will only need to know about the first four DNA repair systems in our current BIO375 class. The homology directed repair (HDR) and non-homologous end joining (NHEJ) content will be added to BIO375 in the future.

Proofreading

The first-line DNA repair system involves the DNA polymerases we discussed in Part 6. When a DNA polymerase synthesizes the daughter DNA strand, the DNA polymerase occasionally inserts an incorrect nucleotide. Recall that DNA polymerases possess 3' to 5' exonuclease activity; the DNA polymerase can reverse directions, remove the incorrect nucleotide in the 3' to 5' direction, and then move 5' to 3' again, replacing the incorrect nucleotide with the correct one.

Mismatch Repair

If an incorrect base pair is formed during DNA replication and the mistake is not removed by proofreading, the **mismatch repair** system is activated to fix the mistake. Mismatch repair recognizes which of the two nucleotides in the base pair is correct and then removes the incorrect nucleotide. For example, suppose that cytosine in the daughter DNA strand has been paired accidently with adenine in the template DNA strand during DNA replication. When mismatch repair recognizes this mistake, it now faces a dilemma: which of the two bases is incorrect? If mismatch repair

randomly chooses the nucleotide to correct, then 50% of the time, it will remove the adenine in the template DNA strand, instead of the cytosine in the daughter DNA strand.

In mismatch repair, the cell must first distinguish the template and daughter DNA strands. In the bacterium *E. coli*, the template DNA strand is methylated by **DNA adenine methyltransferase (Dam)**, while newly synthesized daughter DNA strands are unmethylated. The proteins in the mismatch repair system recognize the methylated parental strand, and then mismatch repair removes the mismatched nucleotide within the unmethylated daughter DNA strand.

In addition to Dam, there are seven other proteins that make up the mismatch repair system in the bacterium *E. coli*: MutS, MutL, MutH, MutU, ExoI, DNA polymerase, and DNA ligase. These seven proteins function as follows:

1. The **MutS** protein slides along the DNA double helix and finds the mismatch.
2. The **MutH** protein binds to a nearby DNA sequence containing a methylated adenine in the template DNA strand. In essence, MutH is the protein that distinguishes the template from the daughter DNA strand.
3. The **MutL** protein binds to both MutS and MutH forming a loop in the DNA.
4. MutH is an endonuclease that makes a single-stranded DNA break in the backbone of the unmethylated (daughter) DNA strand. This break occurs between the G and A bases in the 5'-GATC-3' sequence in the daughter DNA strand.
5. The **MutU** protein binds to the MutH cut site in the daughter DNA strand. MutU has DNA helicase activity and functions to separate the daughter DNA strand from the parental DNA strand at the MutH cut site.
6. The **ExoI** protein is an exonuclease that degrades the damaged daughter DNA strand in the 5' to 3' direction starting at the cut site produced by MutH. Degradation of the damaged daughter DNA strand continues until ExoI removes the mismatched nucleotide.
7. The gap in the daughter DNA strand is filled in by a **DNA polymerase** (i.e., either the DNA polymerase holoenzyme or DNA polymerase I).
8. The final covalent bond in the newly synthesized daughter DNA strand is formed by **DNA ligase**.

Mismatch Repair
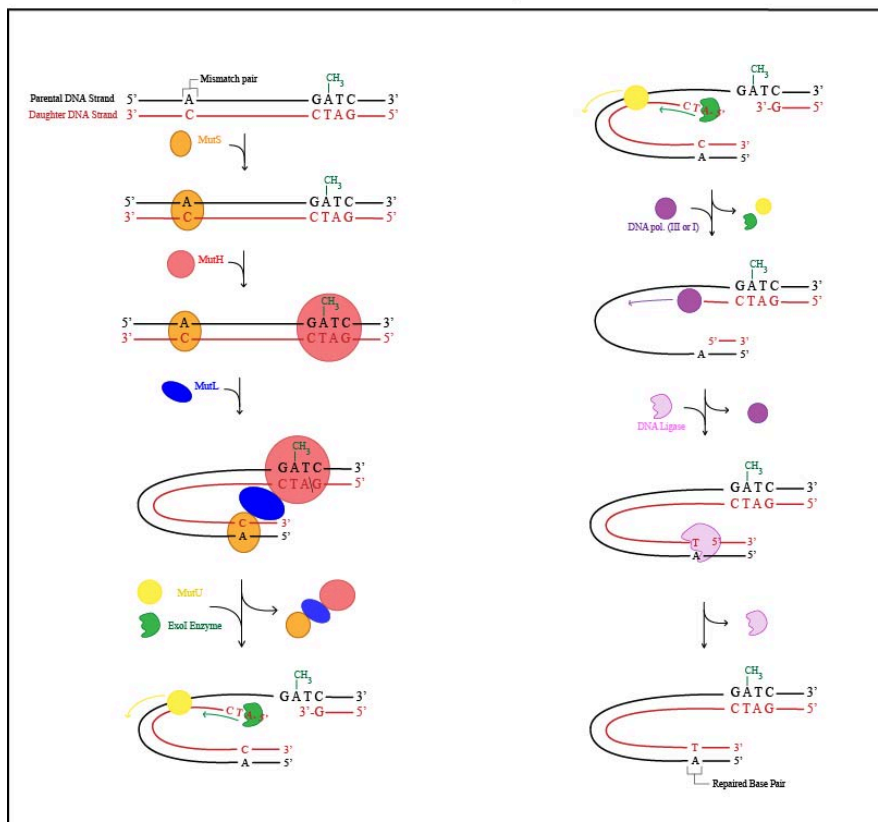
Figure 7.3 - Mismatch repair.

Base Excision Repair (BER)

Base excision repair (BER) corrects abnormal nitrogenous bases that are sometimes formed in DNA. For example, occasionally cytosine in DNA can be spontaneously converted into uracil. Note that the conversion of cytosine to uracil is a transition mutation that does not distort the width of the DNA double helix. BER in the bacterium *E. coli* repairs this transition mutation as follows:

1. The enzyme **uracil DNA glycosylase (UNG)** recognizes uracil in the DNA. UNG cleaves the covalent bond that links uracil to deoxyribose (i.e., breaks the covalent bond attached to the 1' carbon of deoxyribose).   The uracil base is released from the rest of the nucleotide, creating an **abasic** site in the DNA. Note that the nucleotide at the abasic site still contains deoxyribose and the phosphate group; the nucleotide is just missing the uracil nitrogenous base.
2. The abasic site is detected by , **apurinic/apyrimidinic endonuclease 1 (APE1)**, which cleaves the phosphodiester bond at the 5' end of the abasic site. This nick in the DNA backbone generates the free 3'-OH group required by DNA polymerase I.
3. **DNA polymerase I** uses its 5' to 3' exonuclease activity to remove the nucleotide (i.e., the deoxyribose sugar and phosphate group) at the abasic site. DNA polymerase I then uses its DNA synthesis activity (5' to 3' polymerase activity) to replace the removed nucleotide with the correct nucleotide.
4. **DNA ligase** forms the final phosphodiester bond in the DNA strand.

Note that there are similar pathways to repair other unconventional nitrogenous bases that are sometimes found in DNA (e.g. to remove the nitrogenous base hypoxanthine (H) that forms spontaneously from adenine). Instead of UNG, another gycosylase releases hypoxanthine; the other enzymes in BER work the same as described above.
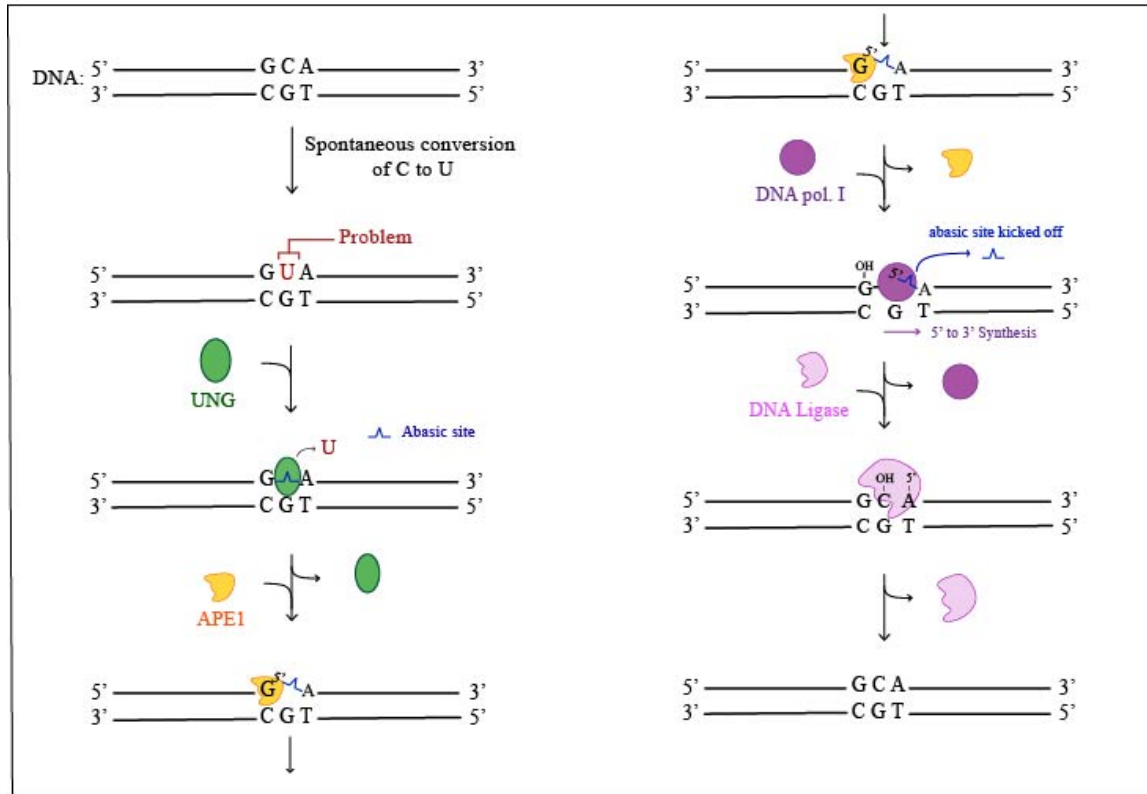
Base Excision Repair

Figure 7.4 - Base excisior repair (BER).

> ## Key Questions
>
> - What is an abasic site?
> - Explain the function of UNG, APE1, DNA polymerase I, and DNA ligase in base excision repair.

Nucleotide Excision Repair (NER)

Ultraviolet (UV) light exposure can lead to the formation of a **pyrimidine dimer** mutation in DNA that distorts the structure of the DNA double-helix. A pyrimidine dimer occurs when UV light causes additional covalent bonds to form between two adjacent pyrimidines in the same DNA strand. When this type of mutation occurs, the hydrogen bonds between the two adjacent pyrimidines and the bases in the other DNA strand are broken. During DNA replication or transcription, when the DNA or RNA polymerase encounters a pyrimidine dimer in the template DNA strand, the polymerases either stop replication or transcription altogether, or incorrect nitrogenous bases are placed in the synthesized strand opposite the pyrimidine dimer.

In *E. coli*, pyrimidine dimers can be repaired with the **nucleotide excision repair (NER)** system. NER involves six proteins: UvrA, UvrB, UvrC, UvrD, DNA polymerase, and DNA ligase. These six proteins remove a segment of DNA including the pyrimidine dimer and then replace the removed nucleotides via DNA replication. NER occurs as follows:

1. A complex consisting of two **UvrA** proteins and one **UvrB** protein scans the double stranded DNA in search of a pyrimidine dimer.
2. Once a pyrimidine dimer is identified, the UvrA/UvrB complex pauses over the dimer. The UvrA proteins are released, while **UvrC** attaches to UvrB at the dimer site.
3. UvrC is an endonuclease that cuts the damaged DNA strand on each side of the pyrimidine dimer.
4. **UvrD**, which is a DNA helicase, separates the two DNA strands, releasing the short segment of damaged DNA, including the pyrimidine dimer itself. UvrB, UvrC, and UvrD are released.
5. A **DNA polymerase** (i.e., either the DNA polymerase III holoenzyme or DNA polymerase I) fills in the gap, using the other DNA strand as a template.
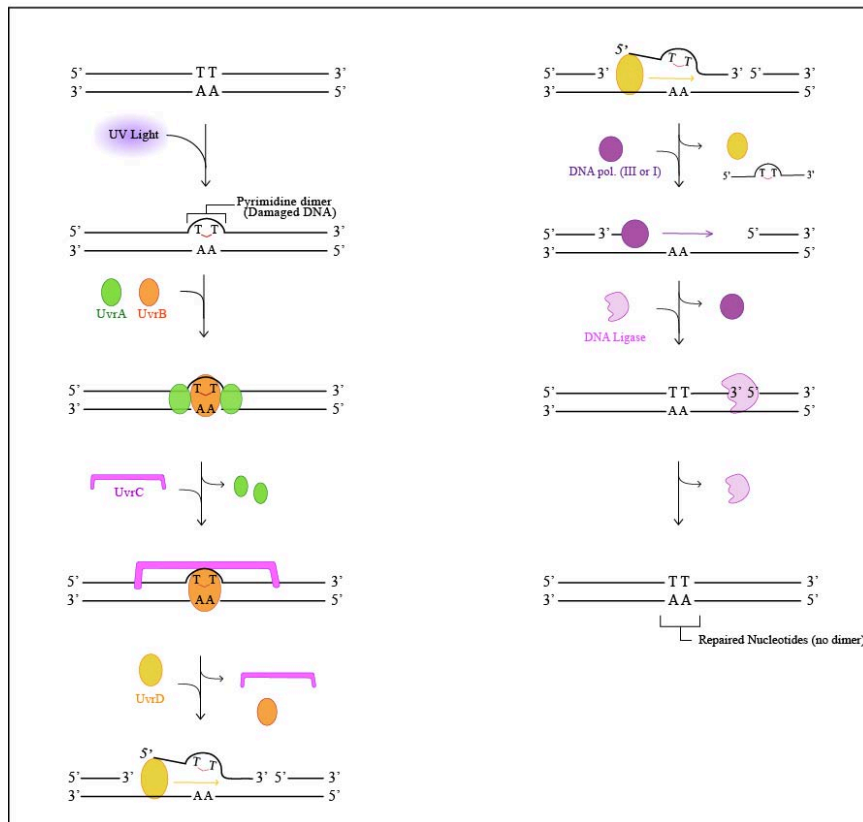6. **DNA ligase** forms the final covalent bond in the newly synthesized DNA.

Figure 7.5 - Nucleotide excision repair (NER).

Key Questions

- Which nitrogenous bases can potentially form pyrimidine dimers?
- What are the functions of the UvrA, UvrB, UvrC, UvrD, DNA polymerase, and DNA ligase proteins in NER?
- Which Uvr protein is the helicase that unwinds and releases the damaged DNA?

Homology Directed Repair (HDR) (future content)

Nonhomologous End Joining (NHEJ) (future content)

Human Genetic Diseases Associated with Faulty DNA Repair

The DNA repair mechanisms in the bacterium *E. coli* described above have counterparts in human cells. Although the proteins are not identical, much of the repair process is similar. Mutations in the genes that give rise to mismatch repair enzymes have been noted in human cells. When both copies of the repair gene are inactivated, cancer can occur. Specifically, a type of inherited cancer called **hereditary nonpolyposis colorectal cancer** occurs because of mutations in the genes that produce human mismatch repair proteins. Mutations in the base excision repair system in humans have also been implicated in **colon cancer**.

Epithelial cells in the skin are constantly exposed to UV light from the sun. As a result, the nucleotide excision repair system plays an important role in repairing pyrimidine dimers that form in the DNA of epithelial cells. **Xeroderma pigmentosum (XP)**, an autosomal recessive disease, occurs when one of the seven human genes involved in nucleotide excision repair are inactivated by mutation. Xeroderma pigmentosum patients are sensitive to sunlight (due to the inability to repair DNA damage caused by UV light) and are predisposed to forming skin cancer.

Insert Xeroderma pigmentosum image here.

Key Questions

- What inheritance pattern is most often associated with diseases caused by defects in the DNA repair systems? Why do you think this is so?

C. Trinucleotide Repeat Expansions and Disease

In humans, there are DNA sequences in which a three nucleotides is repeated consecutively along a DNA strand (**trinucleotide repeat**). In most cases, a parent transmits these repeats to their offspring without any change in the repeat number. However, in a few human genetic diseases, this repeat can expand over the course of generations due to slippage of the DNA polymerases during replication. Once the expansion exceeds a threshold size, the function of the encoded protein is altered, causing disease.

Huntington's disease (HD) is an example of a neurodegenerative disease in humans caused by the expansion of a 5'-CAG-3' repeat within the structural gene *HTT*. Although the exact function of the encoded HTT protein is unknown, it is believed that *HTT* plays an important role in the function of neurons. During protein synthesis, 5'-CAG-3' encodes the amino acid glutamine. In unaffected people, there can be between 10 and 35 repeats of the 5'-CAG-3' codon in the *HTT* gene. These repeats lead to a string of 10-35 glutamine amino acids within the encoded protein. As long as the number of 5'-CAG-3' repeats remains below 35, the encoded HTT protein functions normally. However, in some families, the number of repeats can extend beyond 35, ranging from 36 to 120 repeats. The longer string of encoded glutamine amino acids causes the HTT protein to degrade into small, toxic fragments. As these toxic protein fragments accumulate, neurons die prematurely. Over time, brain activity is altered, leading to symptoms of Huntington's disease, such as uncontrolled body movements, emotional problems, and decreased ability to learn and to make decisions. The number of 5'-CAG-3' repeats in the *HTT* gene correlates with the severity of the disease; a patient with more 5'-CAG-3' codon repeats typically has more severe symptoms than a person with fewer repeats.

Insert a Huntington's disease figure here.

Review Questions

Fill in the blank:

1. A bulge in the DNA double strand occurs when two _____ form a base pair.
2. Exposure to radiation and chemicals is responsible for causing _____ mutations.
3. A single nucleotide substitution in a structural gene that exchanges an amino acid for a stop codon is called a _____ mutation.
4. A deletion of four nucleotides from the coding region of a gene would result in a _____ mutation.
5. In each DNA repair system, _____ forms the final covalent bond in the damaged DNA strand after the correct nucleotide has been added.
6. If a mutation occurs in the DNA such that guanine is paired with uracil, then the most likely DNA repair system to recognize and correct this mutation would be _____ .
7. In mismatch repair, _____ recognizes and binds to the methylated adenine on the parental DNA strand.
8. _____ is the DNA helicase used in nucleotide excision repair.
9. _____ is an example enzyme that has 3′ − 5′ exonuclease activity.
10. Xeroderma pigmentosum is caused because of a defect in the _____ repair system.
11. CAG encodes the amino acid _____ and when the number of CAG repeats exceeds _____, neurological symptoms appear consistent with Huntington's disease.