INTRODUCTION

This content review covers nucleic acids, DNA replication, and analysis of Gene Expression. The Med-Pathway.com web site has challenging MCAT style questions and passages that address this topic.

The biology and chemistry of nucleic acids is fundamental to science and medicine. The AAMC content outline extensively covers nucleic acids, ranging from chemical and physical principles to applications in molecular biology. Therefore, nucleic acids will be seen in both the biological and physical sciences sections. Therefore, this section covers material in biochemistry, organic chemistry, molecular biology as well as physical principles. The following will be reviewed here:

- DNA bases and relationship to uracil
- □ Nucleotide base modifications
- □ Purine and pyrimidine synthesis and degradation
- **DNA Replication and Repair**
- **DNA** tautomers
- **DNA** sequencing
- □ Nucleic acid hybridization
- □ Analyzing gene expression

Uracil and DNA

There are multiple mechanisms for introducing uracil into DNA. Errors in DNA polymerase function result in the misincorporation of uracil into DNA. Despite the fact that replicative DNA polymerases are notoriously faithful, the error rate has been estimated to be 10⁸. This means that for every 100 million bases replicated, one mistake is made. However, other mutations can arise. For example, spontaneous hydrolysis of cytosine creates uracil as shown below. Note how the imino form of cytosine is attacked by water, generating a tetrahedral intermediate that collapses to release ammonia.



To maintain genomic integrity any uracil that is incorporated into DNA is recognized and removed by DNA repair mechanisms. Uracil DNA glycosylase, an enzyme that cleaves glycosidic bonds, recognizes misincorporated uracil and creates an apyrimidinic site (AP site) as shown. This transduces DNA repair signals that result in the replacement of uracil with a complementary base. The abasic site is fixed via DNA polymerase. The mechanism of hydrolysis of the glycosidic bond is shown in the image.



Some enzymes use this deamination chemistry to catalyze the conversion of cytosine into uracil. In B cells of lymphoid origin, activation-induced deaminase (AID) uses DNA substrates to catalyze the conversion of cytosine into uracil. By creating a CG to a UG mismatch, AID induces mutations in the genomes of B cells. This occurs at the immunoglobulin (Ig) locus and results in the generation of antibody diversity through a process known as isotype switching.

Modification of bases. Many RNA and DNA bases are modified. RNA molecules, especially tRNA molecules, have numerous base modifications, including pseudouridine (Ψ) and methylation of cytosine and adenine. 5-methylcytosine (5mC) is the most common DNA modification and is often detected near promoter elements at the 5' end of genes. The presence of 5mC correlates with inhibition of gene expression, indicating that DNA modification influences the expression of genes without changes in the actual DNA sequence. This phenomenon formulates the core principle of epigenetics, a topic covered in the biological section of the MCAT.

Nucleosides are converted into nucleotides through the action of various kinases. These enzymes hydrolyze ATP and transfer phosphate moieties to substrates such as adenosine. The conversion of adenosine to AMP, ADP, and ATP is shown. Cells use nucleotides and their derivatives in numerous processes. This includes the well-established function in DNA and RNA, roles in carbohydrate activation (UDP-Glucose), signaling (cAMP and cGMP), enzyme co-factors (NAD⁺ and FADH2), and energy currency (GTP, ATP).



At the dinucleotide level, the ribonucleotide reductase (RNR) enzyme changes the 2' OH group into a hydrogen atom (H), the form compatible for DNA synthesis after conversion into a triphosphate. Deoxynucleotide diphosphates (dNDPs) are converted into deoxynucleotide triphosphates (dNTPs) through the action of more kinases.

Nucleotide synthesis and degradation

Purine and pyrimidine metabolic pathways are both depicted below. The cellular pools of these nucleotides are formed from both de novo synthesis and scavenging pathways. For synthesis, both purines and pyrimidines require 5-phosphoribosyl 1-pyrophosphate (PRPP), an activated precursor that can be seen from its high energy pyrophosphate bonds. Glutamine provides a nitrogen atom that is added to PRPP to generate 5-phosphoribosylamine (PRA). Understand that in addition to its role in protein structure, glutamine is a primary source of nitrogen (i.e. NH₃). For this, cells express glutaminase, an enzyme catalyzing the release of NH₃ from the glutamine, converting this substrate to glutamate.

For purine synthesis, PRPP is converted into 5-phosphoribosylamine (PRA), a molecule possessing the nitrogen atom that will eventually serve as the heterocyclic atom in the glycosidic bond of the nucleotide. Converting PRA into purine nucleotides requires the further addition of nitrogen and carbon to generate inosine, the branch point precursor in purine synthesis. Additional atoms added to the purine ring are derived from amino acids aspartate and glutamine as well as tetrahydrofolate (THF₄), a co-enzyme that donates one carbon moieties to substrate molecules.



De novo pyrimidine synthesis also requires PRPP. However, pyrimidine synthesis begins with carbamoyl phosphate, an activated precursor derived from glutamate, CO_2 , and ATP. This high energy compound is converted into the intermediate carbamoyl aspartate (Green arrows in figure) via aspartate transcarbamoylase (ATCase) an enzyme that catalyzes the first committed step in pyrimidine synthesis.

Carbamoyl aspartate is a precursor to orotic acid (orotate), itself a precursor to UMP. Both cytosine and thymine are derived from uracil. We focus on thymine as the pathway of converting uracil into thymine is an important drug target. As shown, thymidylate synthase (TS) uses tetrahydrofolate (TH₄) to add a methyl group to uracil. Inhibitors of TS (5 fluorouracil, or 5-FU) have been shown to be effective in cancer treatment through dampening DNA synthesis and subsequent cell division (See below). Addition, the dihydrofolate reeducates (DHFR) enzyme is a target of methotrexate (MTX), an often used therapeutic (see below).



Purine catabolism

The degradation pathways for purines (GMP and AMP) shown below reveal that purine catabolism converges at the production of xanthine, a precursor to uric acid. Uric acid is the end product of purine catabolism in humans but other species can further process it into allantoin.

Several distinct genetic diseases affecting purine metabolism have been described. This includes adenosine deaminase (ADA) deficiency, a rare condition successfully treated through gene therapy. Lack of ADA is causal for severe combined immunedeficiency disease (SCID) and is also the genetic defect underlying "The Boy in The Bubble".



Excessive levels of urate crystallize in joints and tendons, causing a painful, inflammatory condition known as gout. Uric acid is a diprotic acid with two pK_a values of approximately 5.0 and 10.0. Under normal conditions, uric acid is not very soluble. Therefore, any condition that increases serum uric acid levels can cause the metabolite to precipitate. This often occurs in the big toe and other joints and is known as hyperuricemia, or gout. One treatment for gout is to take oral inhibitors of xanthine oxidase (i.e. allopurinol). Pyrimidine degradation



Uracil is degraded in a sequence of steps first carried out by the enzyme dihydropyrimidine dehydrogenase (DPD). Up to 5% of the general population has some form of genetic deficiency in the DPD gene and this is an important determinant in 5-FU therapy. DPD generates 5,6 dihydrouracil, the first product in uracil catabolism. 5-FU is also degraded through this pathway as shown. We have already seen how this pyrimidine analog and cancer chemotherapeutic agent inhibits thymidylate synthase and prevents DNA synthesis. However, in high concentrations, the drug can be quite toxic. Thus, it is important to have an active DPD enzyme to properly eliminate the drug. However, up to 5% of the general population has some form of genetic deficiency in the DPD gene, making it important to know the status of DPD in patients before prescribing 5-FU as a treatment strategy.

DNA replication

DNA replication occurs at distinct cis-acting elements of DNA. Such origins of replication are used to nucleate the duplication of the genome. Two replication forks are generated at each origin and DNA replication occurs in a semiconservative manner as elegantly shown in the Meselson and Stahl experiment. We assume that you are familiar with this experiment from high school biology.

DNA polymerases replicate in the 5' to 3' direction (elaborated on below). Simple prokaryotic cells such as E. coli use a single origin of replication, but in eukaryotes, multiple origin sites are used. In yeast, the origins are rich in AT residues, but in higher eukaryotes origins appear more complex and likely include chromatin as well as DNA topological features. At the origin, localized regions of single stranded DNA serve as templates for RNA primers as shown with the red arrows in the figure. DNA polymerases cannot initiate DNA synthesis from free DNA templates; they require primers. The primers are

synthesized by primase. The primer-template junction recruits a number of factors (blue bubbles) including helicases and DNA polymerase.



Leading and Lagging strand synthesis

Multiple factors are recruited to the origin of DNA replication. A simplified schematic of this complex process is shown below. Helicases use the energy of ATP to unwind the DNA strands. Unwinding DNA creates a topological problem as the DNA is supercoiled. As helicases unwind the DNA, the DNA ahead of the fork becomes overwound. Topisomerase enzymes reduce the supercoiling that, if left unattended, would form "knots" and impede DNA replication.

Although DNA replication is semi-conservative, both DNA strands are not replicated at the same rate. This is because DNA polymerases only synthesize DNA in the 5' to 3' direction. The leading strand is continuously synthesized in the 5' to 3' direction as the helicase provides single stranded DNA substrate. Although not shown in the image, single stranded DNA is usually bound to various proteins to protect it from hydrolysis.

The directionality of DNA polymerase necessitates that the lagging strand is synthesized in a discontinuous fashion. As the helicase provides single stranded template, new RNA primers are laid down for DNA polymerase. Thus, the lagging strand is synthesized in "Okazai fragments", a phenomenon that leaves small gaps between each replicative event. After the removal of the RNA primer, DNA polymerases fills the gaps using ligase enzymes, ultimately sealing the phosphodiester bonds to generate nascent DNA.



Telomeres and the end replication problem

The ends of linear chromosomes are known as telomeres and are composed of hundreds of repeating units. In humans, the sequence is TTAGGG. Telomeres lie in heterochromatic DNA, often term "silenced" DNA and shorten as the cell ages. Shortened telomeres induce senescence.

Telomeres pose a special problem for the replication apparatus as removal of RNA primers at the ends exposes a single stranded, 3' overhang as shown below. Given that the DNA polymerases synthesize in the 5' to 3' direction, telomeres possess a single stranded component that cannot be replicated by conventional DNA polymerases. Eukaryotes have solved this end replication problem by employing the telomerase, an enzyme that uses RNA as an integral component in catalysis. By providing an RNA molecule complementary to the repetitive telomeric core sequence, the single stranded DNA can be "primed"

for replication as there is now a free 3' OH group for the enzyme to extend the nucleotide chain.



TELOMERES AND THE END REPLICATION PROBLEM

DNA polymerases use deoxynucleotide triphosphate (dNTP) substrates to replicate DNA. Replicative DNA polymerases use the familiar Watson-Crick base pairing rules (see below; A::T and G:::C) to generate complementary strands. DNA polymerases use RNA primers to initiate DNA replication.



Recall that the RNA primers bind to complementary sequences in the template DNA providing the foundational substrate for polymerases to replicate in a 5' to 3' direction.



To incorporate a new base into a growing DNA chain, a replicative polymerase will bind to both template and a complementary dNTP. The addition of a new phosphoester bond occurs through a nucleophilic addition-elimination reaction that uses the free nucleophilic 3'OH group of the chain nucleotide to attack the electrophilic α phosphate in the incoming nucleotide (see image). PPi is released and its hydrolysis completes the reaction. Multiple features of the double helix provide stabilizing features: hydrogen bonding between complementary bases, the phosphoester backbone, and base stacking interactions.



DNA repair

Although DNA polymerases are highly accurate enzymes, they occasionally introduce errors. In many cases, these errors can be corrected by the polymerase itself. However, as the DNA is constantly being assaulted by numerous agents, additional DNA repair mechanisms have evolved to fix the genome. Such DNA repair mechanisms include oxidative damage, UVinduced dimerization (i.e. thymine-thymine dimers), DNA mismatch repair, and DNA double strand break (DSB) repair. We will examine DSB repair in more detail.

There are two main methods for fixing DSBs: non homologous DNA end joining (NHEJ) and homologous recombination (HR). NHEJ ligates the broken DNA ends together irrespective of any lost information. Thus, NHEJ is error prone. In contrast, HR uses a sister chromatid to provide for a high fidelity repair event.

DSBs are created through various methods including ionizing radiation. An elaborate mechanism exists to repair DSBs. The breaks are sensed due to topological changes introduced into the DNA molecule as a consequence of a genotoxic insult. At the site of the break, exonucleases trim the DNA duplex to leave 3' overhangs. This creates a single stranded DNA molecule coated with a protein called Rad51. This generates a nucleoprotein filament endowed with the amazing property of being able to find a homologous sequence. Such a sequence can provide any missing genetic information that may have arisen from the DNA damage.

As shown below, the free 3' OH group in the nucleoprotein filament forms a "D loop" structure with a homologous template that would be present in S and G_2 phases of the cell cycle. Using the sister template, a new strand of DNA is synthesized through branch migration. After resolution, this generates a recombinant DNA product that has repaired the DNA break.



DNA repair and cancer

In addition to exogenous sources such as ionizing radiation, DSBs are endogenously generated during DNA replication. This is shown below and has important implication for cancer development. As DNA replication occurs in an oxidative environment, numerous single stranded gaps ("nicks") are always being created and fixed. One important factor for repairing such nicks is an enzyme known as poly ADP ribose polymerase (PARP) (see figure). In most cases, PARP can fix the single stranded break prior to the arrival of the new replication fork. However, in those cases where the replication fork beats PARP to the nicked DNA, a DSB is formed as shown. Under normal conditions, the DSB is repaired via HR as discussed above. However, in people that have genetic deficiencies in the HR apparatus (such as biallelic inactivation of the breast cancer susceptibility gene BRCA1), the break is not properly fixed. As a result dysfunctional HR events occur. This results in translocations, events that generate aberrant chromosomes including those with two centromeres (dicentrics). In many cases, such aberrant chromosomes result in apoptosis, but a small percentage can endow the cell with the increased proliferative capacity seen in cancer.



Understanding the relationship between PARP and HR has created a major pharmaceutical push to develop PARP inhibitors to treat BRCA1-driven breast cancers. Think about how this might work. The Med-Pathway Passage Workbook contains a passage on this exciting topic in cancer biology. One thing that drops out of this is that PARP inhibitors specifically target those cells deficient in HR (i.e. BRCA1^{-/-} cells).

DNA is organized into nucleosomes

DNA synthesized in S phase is organized into nucleosomes. This is also commonly referred to as "the beads on the string model". Here, approximately 146 base pairs of DNA are wound around a histone octamer. There is also a linker region of DNA that is approximately 80 base pairs. The combination of DNA, protein, and even RNA is known as chromatin. As nucleosomes compact DNA, there must be regulatory mechanisms that allow for the DNA replication, repair, and recombination machinery to have access. Indeed, the Ntermini of histones ("tails") are postranslationally modified by a number of enzymes that deposit various groups including phosphate, acetyl, and methyl groups. Various "reader" proteins typically recognize such groups. Once recognized, genes are either expressed, repressed or even repaired in the case of DNA damage.



Tautomers and DNA structure

Each of the DNA bases exist in an equilibrium between two major "tautomeric" forms. The image shows (Panel A) this for thymine, a pyrimidine that exists in both the keto and enol tautomeric forms. Note the relationship between the keto and enol forms as isomers. Panel B shows the mechanism of conversion of a ketone into an enol through a base-catalyzed pathway. In this scheme, a base (i.e. enzyme side chain or solvent molecule), removes a proton, leaving an additional lone pair of electrons on the heterocyclic nitrogen atom. These electrons are in resonance with the pi electrons in the carbonyl bond where they can form an enolate intermediate (Panel B). The enolate readily picks up a solvent proton to generate the enol tautomer. Tautomers are not confined to the keto/enol equilibrium and also exist in DNA in the amino/imino forms.



Biological significance of DNA tautomerism

Under most circumstances the keto form is far more prevalent at equilibrium. However, in some cases the DNA polymerase can incorporate the rare tautomeric form into DNA, a condition that can cause mutations as shown in Panel C. In the enol form, thymine can hydrogen bond to guanine, creating a T:::G mutation.

Exploiting the 3'OH group

DNA polymerases have an absolute requirement for elongating DNA chains. Therefore, by chemically manipulating the 3'OH group, the capacity to replicate can be regulated. During the 1980s HIV epidemic, scientists used the compound azidothymidine (AZT) to negatively regulate the viral DNA polymerase, a specialized enzyme known as reverse transcriptase. The structure of AZT is shown below. AZT is an early member of a class of compounds known as chain terminators. These important molecules are currently employed in HIV therapy.



Med-pathway.com Nucleic Acids & DNA Replication Content Review The MCAT Masters

DNA sequencing by the Sanger dideoxy technique

In the 1970s, Fred Sanger developed dideoxynucleotide triphosphates (ddNTPs or ddG, ddA, and ddT) for determining the sequence of DNA in a sample. These nucleotides have 3' H groups substituted for the 3'OH group normally used by template-driven DNA polymerase enzymes. Hence the term "dideoxy". The 3'OH is essential for growing nucleic acid chains. When a DNA polymerase reaction encounters a dideoxynucleotide (ddNTP), the nucleic acid chain is terminated upon incorporation of the ddNTP due to the absence of the free 3' OH group.

The Sanger DNA sequencing technique is described and illustrated below. Purified template DNA is enzymatically radiolabeled at the 5' end with the addition of a phosphate group (represented by *). Template DNA is incubated in a cocktail containing multiple reaction components. This includes a suitable buffer containing Mg⁺² and each of the four dNTPs. After mixing in the enzyme (Step 1), the final reaction mixture is then aliquoted into four separate tubes (Step 2), each of which is spiked with a single ddNTP. In a single tube each reaction proceeds until the DNA polymerase adds a dideoxynucleotide into the nucleic acid chain. Because both normal deoxynucleotides and a specific dideoxy chain terminating nucleotide is present in a given reaction, a series or "nested set of fragments" is generated for each nucleotide. Each fragment ends with the incorporation of a chain terminating ddNTP. As each fragment varies in size, they can be separated by acrylamide/urea gel electrophoresis. In this procedure, small DNA fragments migrate faster than larger nucleic acid chains. After running the sequencing gel electrophoresis experiment, the following profile is obtained:



Nucleic acid hybridization

Nucleic acid hybridization is a fundamental aspect of biology and has been exploited in multiple, medically important techniques. This includes the polymerase chain reaction (PCR) and Southern blotting as well as fluorescent in situ hybridization (FISH). These techniques are outlined in the figure and discussed below.

As the Watson-Crick model of DNA shows that the two antiparallel strands of DNA are held together through hydrogen bonding, DNA strands are easily separated with heat. The complementary single strands of DNA (ssDNA) reassociate after slow cooling. Alternatively, the single stranded DNA molecules can be used to amplify DNA through the technique of PCR (Polymerase Chain Reaction). In this procedure, primer sequences complementary to portions of the ssDNA are added. The mixture is cooled to allow annealing. In the presence of DNA polymerase enzyme and nucleotide triphosphates, the template DNA is amplified. Afterwards, the amplified DNA Med-pathway.com Nucleic Acids & DNA Replication Content Review The MCAT Masters

is heated, more primer anneals, and more DNA is amplified. After repeated cycles, a small sample of DNA can be exponentially amplified to significant levels used in various diagnostic techniques.

Alternatively, the ss DNA can be probed for the existence or alteration of various target sequences using techniques like Southern blotting or FISH. By fixing the single stranded DNA sequences in situ (i.e. membrane or glass slide), radiolabeled or fluorescent probes (i.e. Southern blot or FISH, respectively) can be used to query the presence of DNA sequences. Such techniques have wide applications ranging from forensic science to cancer diagnostics.



Analyzing gene expression

Hybridization techniques are critical for analyzing gene expression. In a Northern blotting experiment, mRNA molecules are isolated and separated via gel electrophoresis in a manner analogous to Southern blotting. When single stranded, radiolabeled DNA probes are added, stable, complementary DNA-RNA hybrids can be detected. Therefore, the Southern blotting

technique is for detecting DNA-DNA interactions and Northern blotting is used to detect DNA-RNA interactions.

Quantitative or real time PCR (qPCR or RT-PCR) is widely used to determine gene expression levels. Unlike conventional PCR, RT-PCR is quantitative. The procedure is shown below. In Panel A, total mRNA is isolated from cell lines through the use of poly T affinity purification. As the vast majority of mRNA molecules are equipped with poly A tails, poly T affinity purification separates mRNA from other RNA species including ribosomal RNA, which is abundant. Poly A+ RNA is composed of mRNA molecules devoid of introns; only the sequences encoded by the exons are isolated. The isolated mRNA molecules are used as templates in reactions catalyzed by reverse transcriptase. Originally discovered in retroviruses, reverse transcriptase uses mRNA as a substrate to generate complementary DNA (cDNA) molecules. This reaction occurs through an RNA-DNA hybrid intermediate. The enzyme is also equipped with a RNA nuclease activity that degrades the RNA in the RNA-DNA hybrid, generating a single stranded DNA template that is used as a substrate for the DNA polymerase function of reverse transcriptase.

Panel B shows how the mRNA levels are quantitated. In one manifestation of



the procedure, a fluorescent probe is hybridized to the internal region of the target gene whose expression levels are being measured. During the polymerization step of PCR, the polymerase will destroy the probe hybridized to the target DNA thereby liberating the fluorophore from the quenching agent. The amount of released fluorophore is related proportionally to the amount of input DNA and is measured in the qPCR technique.