The Biochemistry Survival Guide

NeighborhoodGeeks

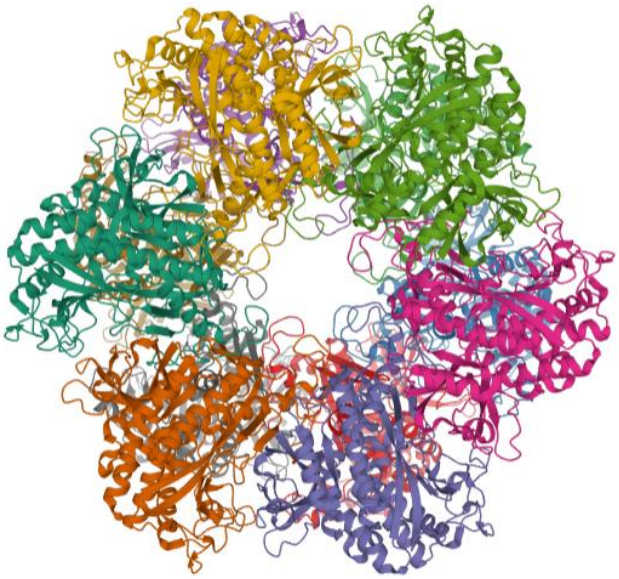


Figure 1. Glutamine Synthetase, PDBID: 2BVC[1]

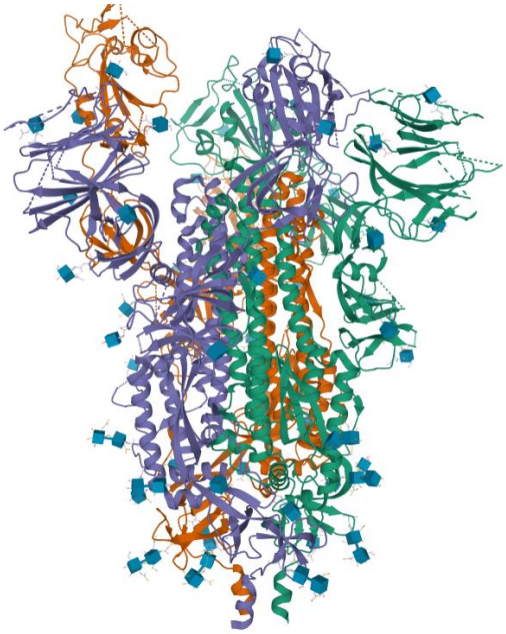


Figure 2. SARS-CoV-2 Spike Protein, PDBID: 6VYB[2]

Table of Contents

Focus 0: Foreword from the Author……6

Chapter 0-1: Strategies for Biochemistry……6

Chapter 0-2: Amino Acid Notecards……7-25

Chapter 0-3: Enzyme Classes……26-27

Focus 1: Protein Structure and Function + Biochem Basics

Chapter 1: Intermolecular Forces and Common Biological Functional Groups...28-37

Chapter 2: Arrangement of Atoms in Space (Stereochemistry)…….38-43

Chapter 3: Titration Curves and Protein PI……44-53

Chapter 4: Protein Structure Basics……...54-59

Focus 2: Enzyme Mechanisms, Inhibition, and Saturation

Chapter 5: Hemoglobin + Myoglobin: Structure, Function, and Saturation…61-64

Chapter 6: Enzyme Catalytic Strategies and Michaelis Menten Kinetics Basics…65-73

Chapter 7: Competitive Inhibition…….74-77

Chapter 8: Uncompetitive Inhibition…….78-81

Chapter 9: Mixed Inhibitors and Allosteric Inhibitors……...82-84

Focus 3: Membrane Structure and Function

Chapter 10: Cell Membrane Basics……86-92

Chapter 11: Types of Transporters…….93-96

Chapter 12: Free Energy of Transport……97-99

Focus 4: DNA Technology and Common Biochemical Techniques

Chapter 13: Polymerase Chain Reaction + Plasmids……101-106

Chapter 14: DNA Sequencing (Sanger and Next Generation) …….107-112

Chapter 15: Gel Electrophoresis and Column Chromatography……...113-118

Chapter 16: Western Blot……..119-120

Chapter 17: DNA Microarrays……..121

Chapter 18: Yeast-Two Hybrid……122-123

Chapter 19: CRISPR/Cas9……124-126

Focus 5: Carbohydrate Metabolism

Chapter 20: Carbohydrate Chemistry Basics……128-136

Chapter 21: Glycolysis……...137-150

Chapter 22: Krebs Cycle……....151-158

Chapter 23: Electron Transport Chain……...159-165

Chapter 24: Gluconeogenesis……166-171

Chapter 25: Pentose Phosphate Pathway……...172-177

Chapter 26: Glycogen Biosynthesis……...178-181

Chapter 27: Glycogen Breakdown……182-185

Focus 6: Lipid Metabolism

Chapter 28: Beta Oxidation……187-195

Chapter 29: Ketone Bodies……196-200

Chapter 30: Fatty Acid Biosynthesis……201-206

Chapter 31: Cholesterol Biosynthesis……207-212

Chapter 32: Cholesterol Transport……213-216

Chapter 33: Cholesterol Regulation……...217-219

Focus 7: Amino Acid/ Protein Metabolism

Chapter 34: Exogenous (externally supplied) Protein Breakdown……221

Chapter 35: Endogenous (internally made) Protein Breakdown……222-223

Chapter 36: Urea Cycle……..224-231

Chapter 37: Amino Acid Breakdown……232-235

Chapter 38: Amino Acid Biosynthesis……..236-241

Focus 8: Nucleic Acid Metabolism

Chapter 39: Nucleic Acid Biosynthesis (Purines) ……243-247

Chapter 40: Nucleic Acid Biosynthesis (Pyrimidines) …….248-250

Chapter 41: Nucleic Acid Breakdown and Salvage……...251-255

Focus 9: Cancer Biology

Chapter 42: Cancer Fundamentals……257-259

Chapter 43: Insulin-IRS1 Pathway……260-262

Chapter 44: HER-2 Pathway……..263-264

Chapter 45: Wnt Signaling Pathway……..265-267

Chapter 46: How Chemotherapeutics Work (Case Example: Tamoxifen)...268-270

Chapter 47: Future Directions for Cancer Research……271-273

Focus 10: SARS-CoV-2

Chapter 48: Virology Basics……275-279

Chapter 49: Immunology Basics……..280-282

Chapter 50: RNA-dependent RNA Polymerase inhibitors……..283-284

Chapter 51: Viral protease inhibitors……...285-287

Chapter 52: Cell-entry inhibitors……..288-289

Chapter 53: Vaccine Development and Stages of Clinical Trials……290-294

Focus 11: DNA Replication

Chapter 54: Initiation of DNA Replication in Prokaryotes……..296-297

Chapter 55: Elongation during DNA Replication in Prokaryotes……298-301

Chapter 56: Comparison between Replication in Prokaryotes and Eukaryotes…302

Focus 12: DNA Repair

Chapter 57: Mismatch Repair……304-306
Chapter 58: Base Excision Repair……..307-309
Chapter 59: Nucleotide Excision Repair……310-311
Chapter 60: Direct Repair (Prokaryotes only) ……...312-313
Chapter 61: Homologous Recombination……..314-316
Chapter 62: Trans Lesion DNA Repair……..317
Focus 13: Transcription and Translation
Chapter 63: Transcription in Prokaryotes……..319-322
Chapter 64: RNA Processing in Eukaryotes for mRNA, rRNA, and tRNA..323-325
Chapter 65: Genetic Code……..326-329
Chapter 66: Translation in Prokaryotes……..330-335
Focus 14: Gene Regulation and Epigenetics
Chapter 67: Histone Structure and Function……..336-338
Chapter 68: Lac Operon…….339-341
Chapter 69: Trp Operon…….342-343
References……344

Focus 0: Foreword from the Author

Chapter 0-1: Strategies for Biochemistry

Congrats on graduating from Organic Chemistry! In this course, you are going to discuss some of the practical applications of that knowledge, specifically in the context of biological systems, which I find to be super cool. This textbook will cover what is generally done in the generic two semester sequence of Biochemistry and will cover everything that premed students must know for the MCAT exam plus some extra cool tidbits on SARS-CoV-2, cancer, and molecular techniques. Here are my general pieces of advice and strategies for dealing with the Biochemistry course content:

1. Constantly engage in the material. Biochemistry is seen all around you so thinking about it and rationalizing things you see and hear should be relatively easy. Especially now with the SARS-CoV-2 pandemic taking the world by storm, it is OUR job as a collective to educate the public and explain to them using non-technical language why certain pieces of misinformation are incorrect and scientifically flawed. Use this pandemic as an opportunity for growth both in scientific communication and in understanding Biochemistry.
2. Make narratives. Storytelling has long captivated humankind; use the strength of storytelling to help you on your journey through Biochemistry. Especially regarding metabolic pathways. These are some of the more difficult parts of the course, but developing some kind of story surrounding it and going from start to finish for each pathway will make the material less daunting and more fun.
3. Don't forget your basics. Acids will still react with bases, intermolecular forces are still a thing, polarity does not go away. Never forget the fundamental tools that you have been cultivating since General Chemistry.

In the spirit of the SARS-CoV-2 pandemic, here are my *Fundamental Principles of Biochemistry*:

Minimize energy and charge

ATP powers unfavorable reactions

Structure informs function

Cellular processes avoid waste and maximize efficiency

Just as in the Organic Chemistry survival guide, I will periodically refer back to MASC to explain certain biochemical phenomena. Good luck and happy studying!

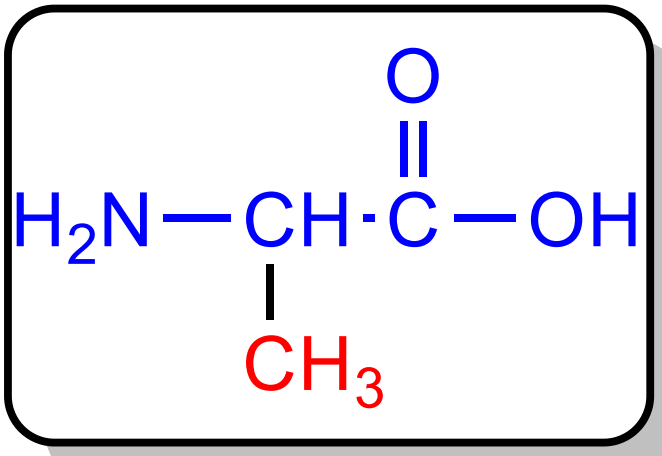
Chapter 0-2: Amino Acid Notecards

Amino acids are the building blocks of all the proteins in your body, as such, they are incredibly important to any aspiring biochemist and therefore I have consolidated all of the important information about each amino acid in these notecards. A couple of trends that we should take note of:

1. All amino acids with OH groups can get phosphorylated
2. All amino acids that are β-branched stabilize β-sheets
3. All basic amino acids are positively charged
4. All acidic amino acids are negatively charged

A good way of remembering all of the nonpolar amino acids by their one letter abbreviation is FAMILY VW + G. All other amino acids are either polar or ionic at neutral pH. All amino acids are comprised of an alpha amino group, a carboxylic acid group, and a side chain on the alpha carbon. These side chains define the properties of that amino acid.

Alanine, Ala, A



Properties:

Nonpolar

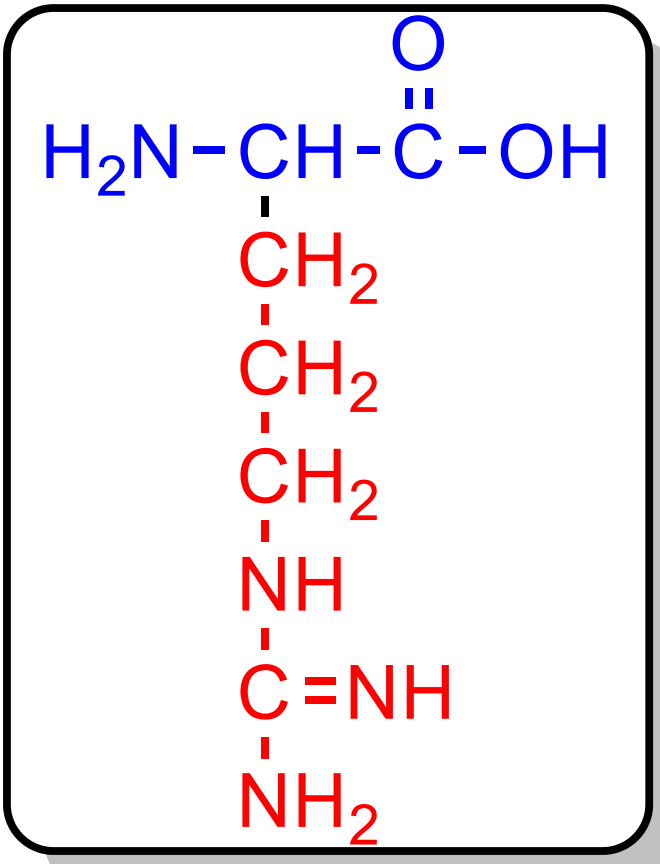
Aliphatic

Hydrophobic

Not Bulky

Stabilizes α helix

Arginine, Arg, R



Properties:

Positively charged

Basic

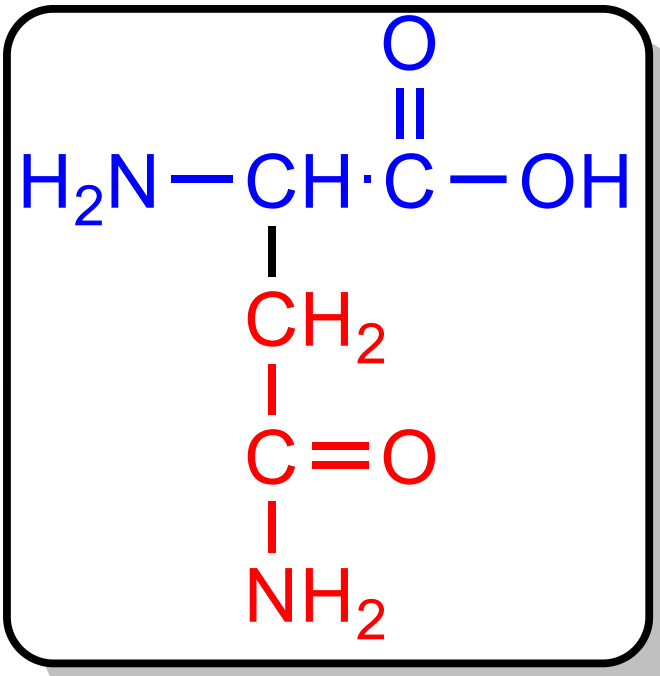
Aliphatic

Hydrophilic

Moderately bulky

Cap C-term of α helix

Asparagine, Asn, N



Properties:

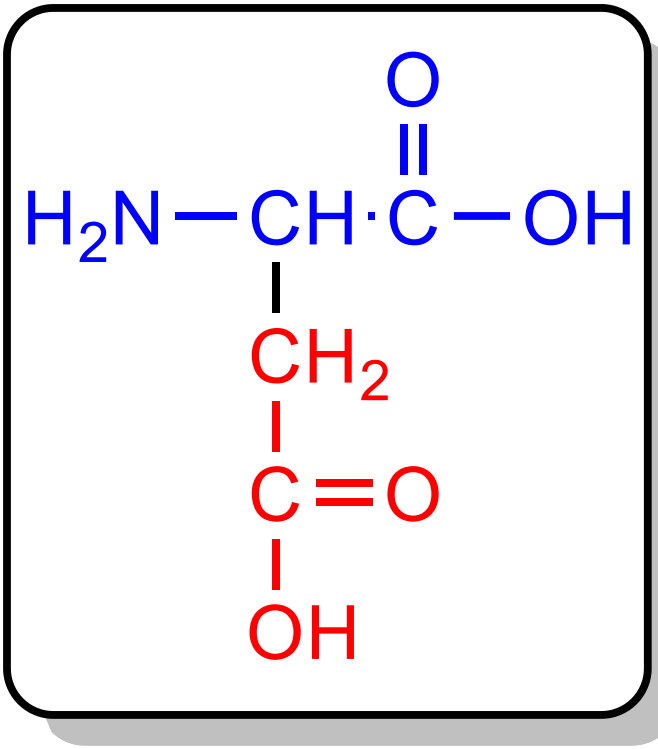
Polar

Aliphatic

Hydrophilic

Moderately bulky

Aspartic Acid, Asp, D



Properties:

Negatively charged

Acidic

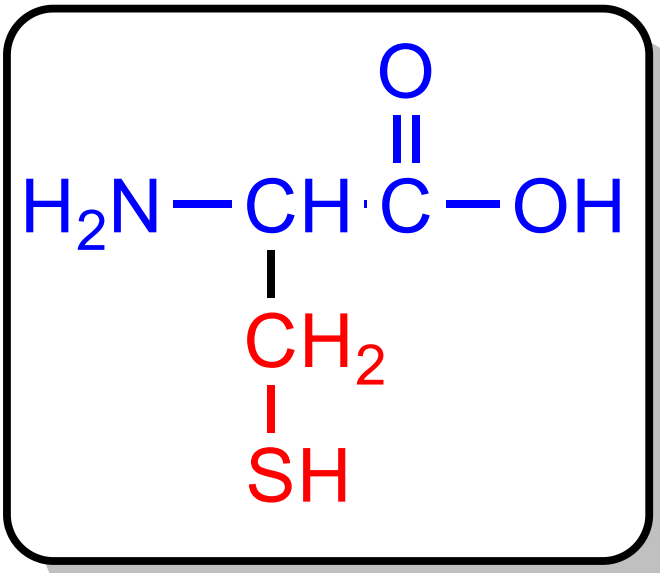
Aliphatic

Hydrophilic

Moderately bulky

Cap N-term of α helix

Cysteine, Cys, C



Properties:

Polar

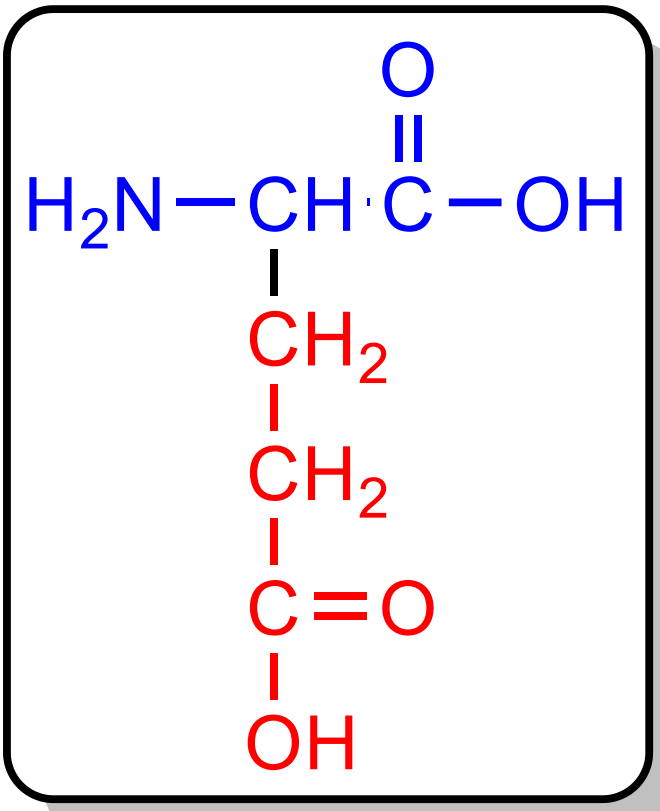
Aliphatic

Hydrophilic

Not bulky

Can form disulfide bonds

Glutamic Acid, Glu, E



Properties:

Negatively charged

Acidic

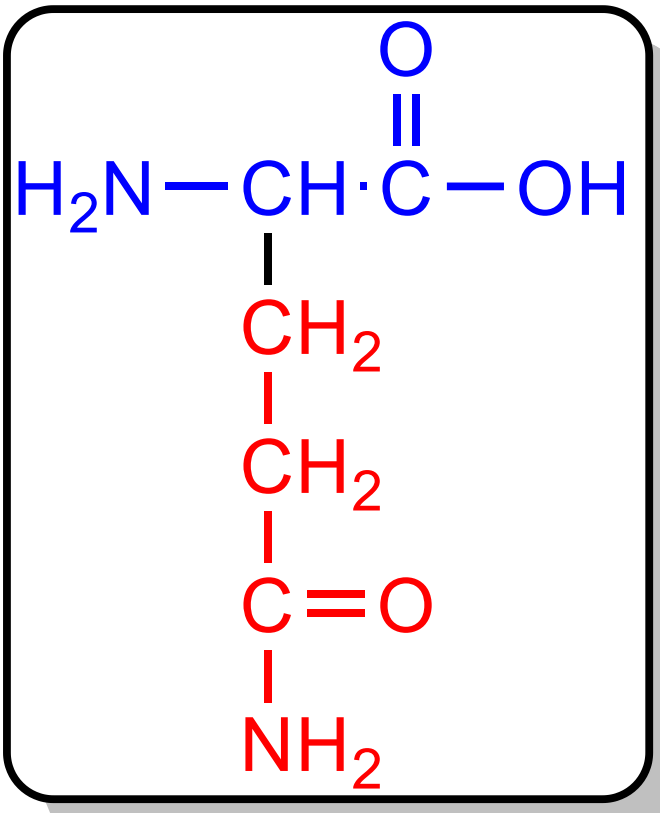
Aliphatic

Hydrophilic

Moderately bulky

Cap N-term of α helix

Glutamine, Gln, Q



Properties:

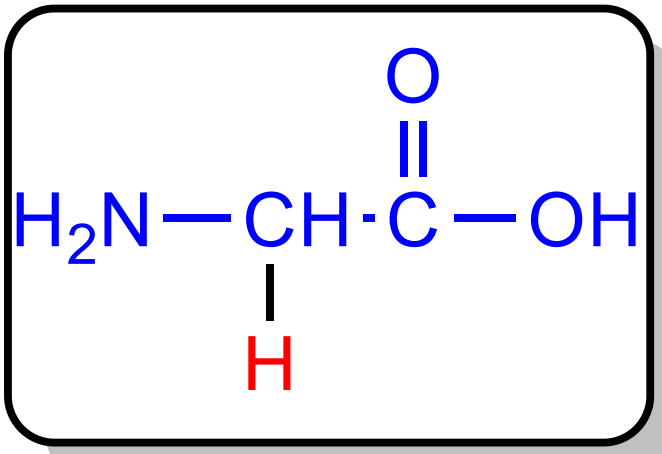
Polar

Aliphatic

Hydrophilic

Moderately bulky

Glycine, Gly, G



Properties:

Nonpolar

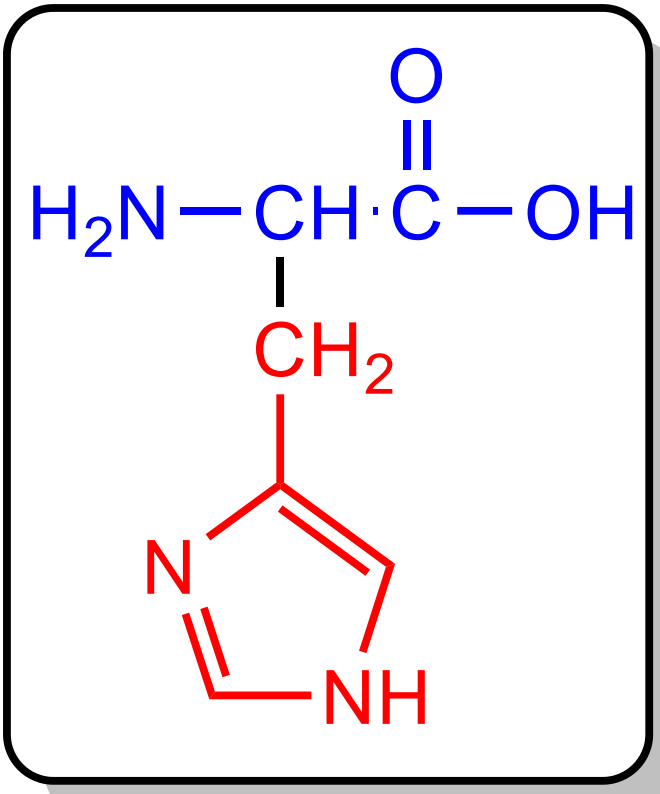
Aliphatic

Hydrophobic

Not Bulky

Heavily destabilizes α helix (very flexible)

Histidine, His, H



Properties:

Positively charged

Basic

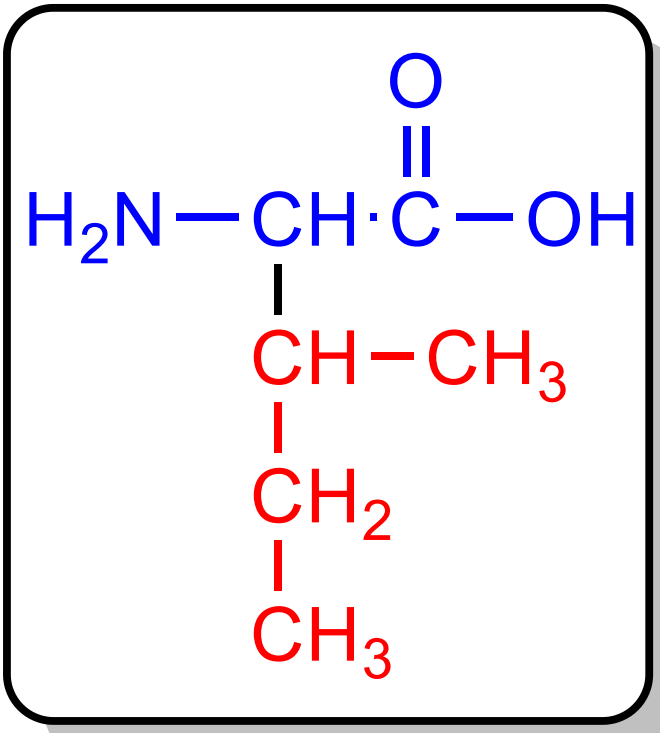
Aliphatic

Hydrophilic

Moderately bulky

Cap C-term of α helix

Isoleucine, Ile, I



Properties:

Nonpolar

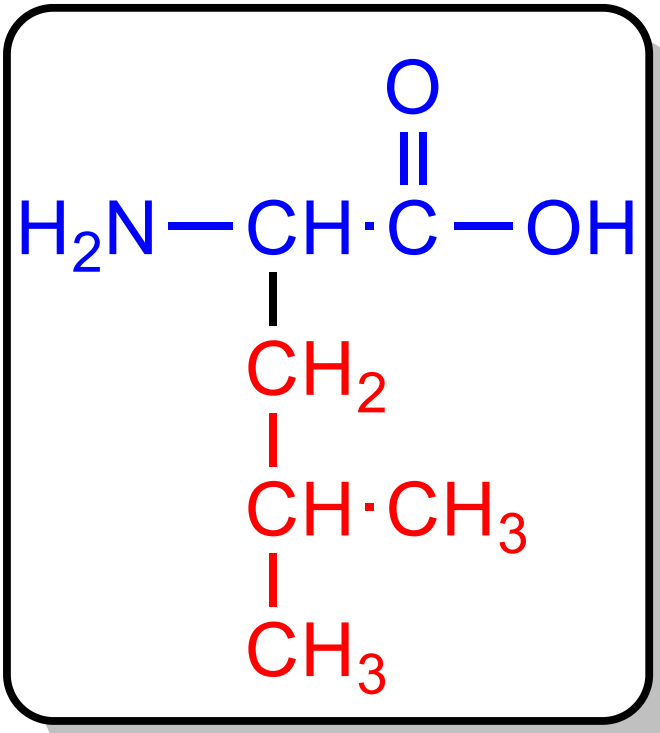
Aliphatic

Hydrophobic

Moderately Bulky

Stabilizes β Sheets

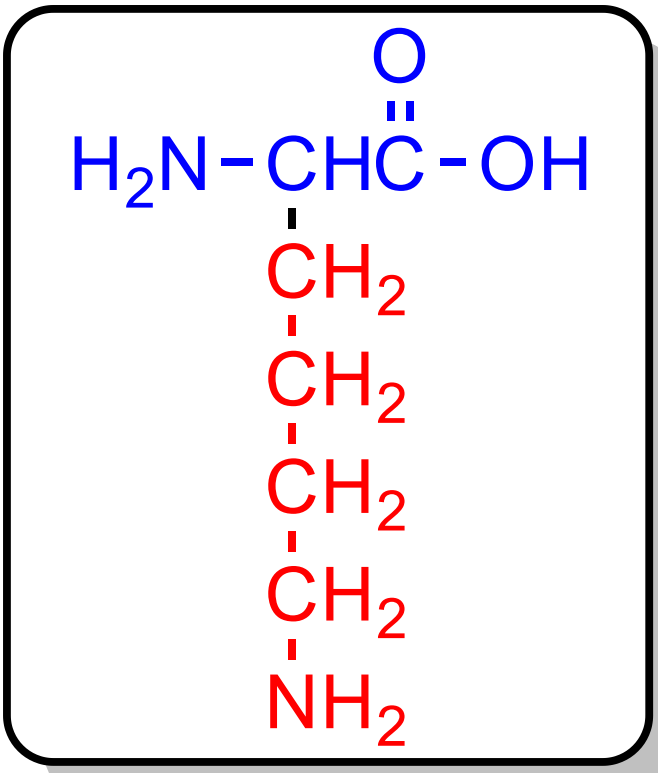
Leucine, Leu, L



Properties:

Nonpolar
Aliphatic
Hydrophobic
Moderately Bulky

Lysine, Lys, K



Properties:

Positively charged

Basic

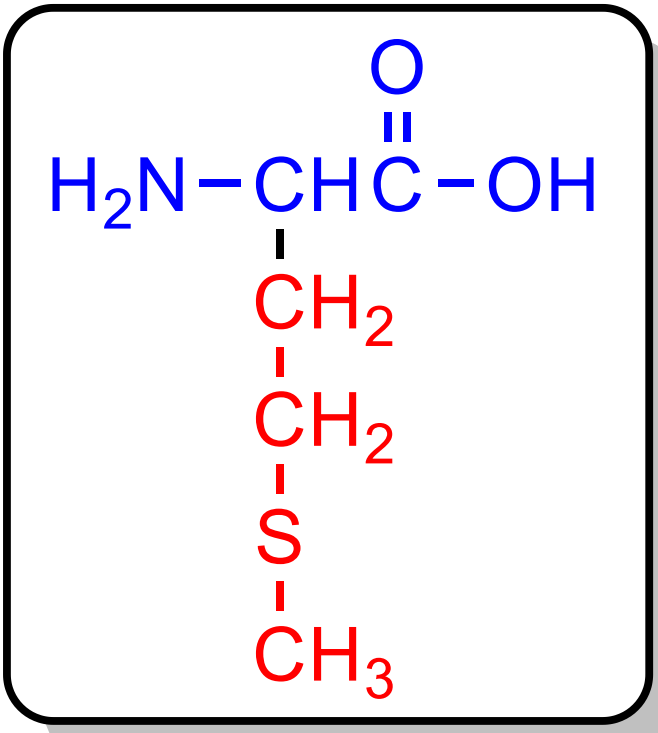
Aliphatic

Hydrophilic

Moderately bulky

Cap C-term of α helix

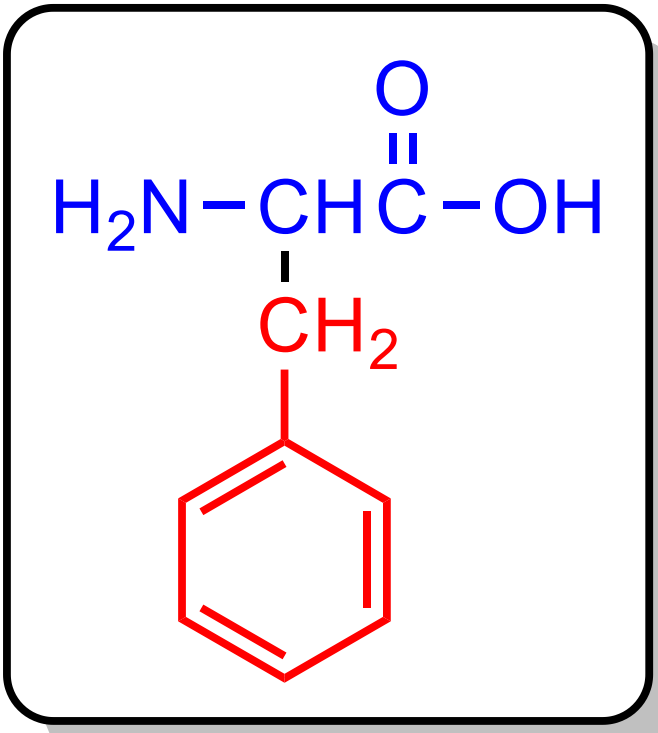
Methionine, Met, M



Properties:

Nonpolar
Aliphatic
Hydrophobic
Moderately Bulky

Phenylalanine, Phe, F



Properties:

Nonpolar

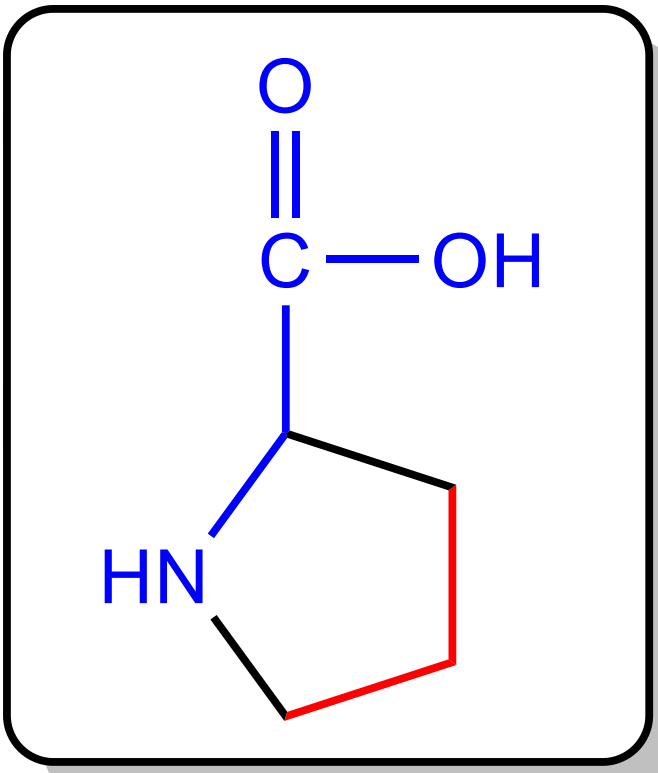
Aromatic

Hydrophobic

Very Bulky

Destabilize α helix

Proline, Pro, P



Properties:

Nonpolar

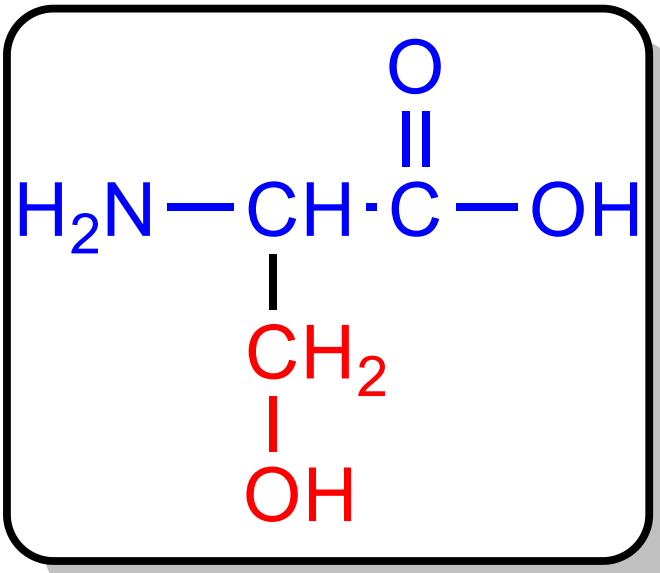
Aliphatic

Hydrophobic

Very rigid

Heavily Destabilizes α helix

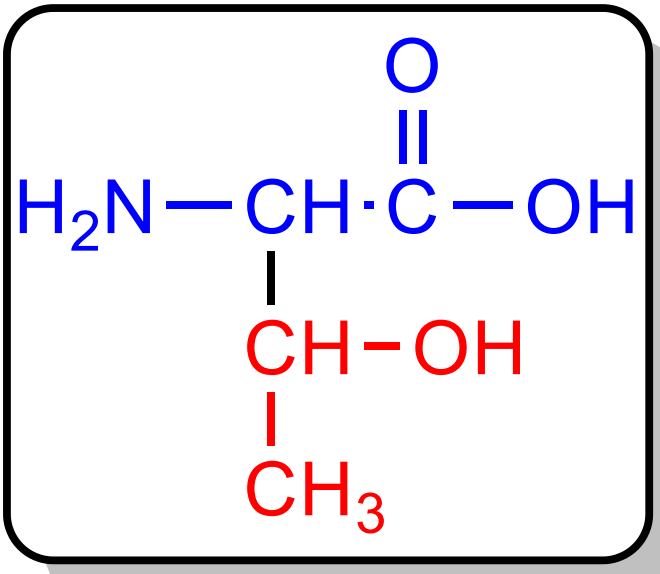
Serine, Ser, S



Properties:

Polar
Aliphatic
Hydrophilic
Not bulky
Can get phosphorylated

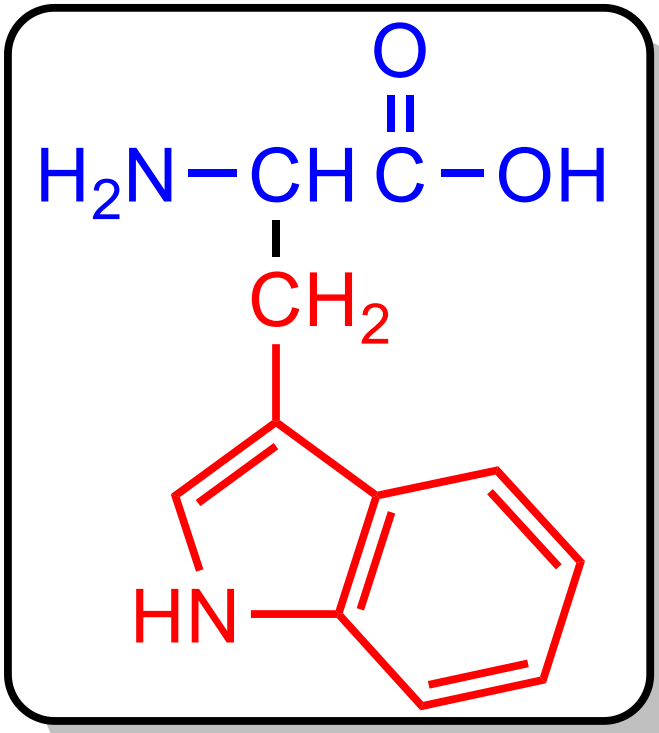
Threonine, Thr, T



Properties:

Polar
Aliphatic
Hydrophilic
Not bulky
Can get phosphorylated

Tryptophan, Trp, W



Properties:

Nonpolar

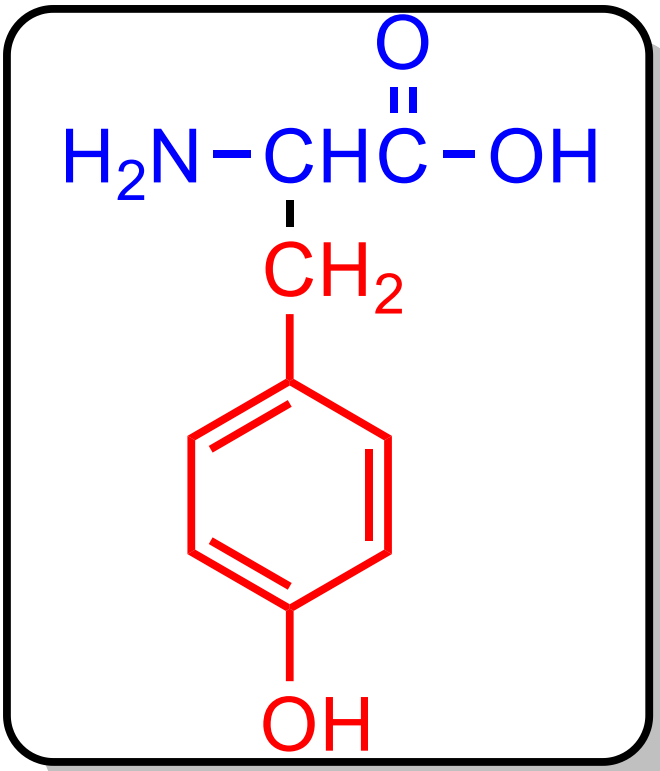
Aromatic

Hydrophobic

Very Bulky

Destabilize α helix

Tyrosine, Tyr, Y



Properties:

Nonpolar

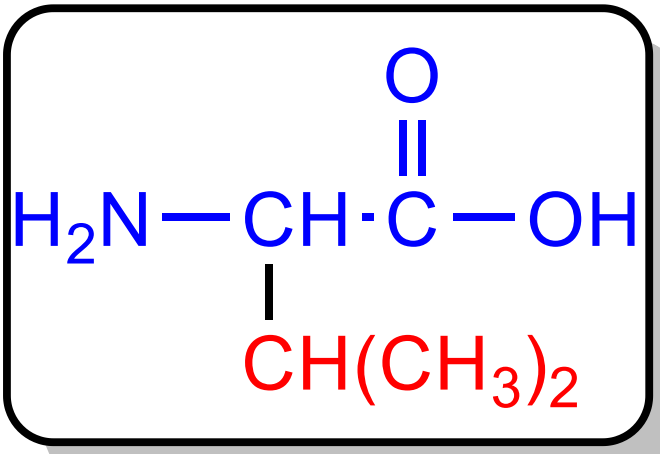
Aromatic

Hydrophobic

Very Bulky

Can get phosphorylated

Valine, Val, V



Properties:

Nonpolar
Aliphatic
Hydrophobic
Moderate Bulky

Chapter 0-3: Enzyme Classes

This chapter will focus on the different types of enzyme names and what those names mean about their function. This is effectively a cheat sheet that allows you to predict the chemistry of a given enzyme given just the name. This is simply meant as a reference. Typically, the enzyme name will consist of the substrate and then the type of reaction that it will do to it. For example, alcohol dehydrogenase will oxidize alcohols to aldehydes/ketones. It is simply an alcohol dehydrogenase because it can bind to a variety of small chain alcohols such as methanol and ethanol.

Kinase: phosphorylates an OH functional group on a protein

Phosphatase: removes a phosphate group on a protein

Synthase: makes a compound without using ATP

Synthetase: makes a compound with ATP or an equivalent

Mutase: moves a functional group to a different carbon

Isomerase: creates a structural isomer (i.e. glucose to fructose)

Sulfotransferase: adds a sulfate group to an OH group on a protein or small molecule

Sulfatase: removes a sulfate group

Transferase: moves a functional group from one molecule to another

Dehydrogenase: oxidizes a chemical bond using NAD^+ and FAD

Reductase: reduces a chemical bond using NADH and $FADH_2$

Oxidase: oxidizes a chemical bond using molecular oxygen (O_2)

Phosphorylase: break a chemical bond using phosphate anion

Hydratase: adds an OH group to a pi bond

Lyase: breaks a molecule into two, oxidizes one half and reduces the other half, no net change in oxidation state and therefore does not need a redox cofactor.

Ligase: joins two macromolecules together, basically glues them together

Polymerase: creates a nucleotide strand

Topoisomerase: removes or adds supercoils to DNA either using ATP or not using ATP for type II and type I respectively.

Endonuclease: removes nucleotides in the middle of a nucleotide strand

Exonuclease: removes nucleotides at the ends of a nucleotide strand

Acetylase: adds an acetyl group

Methylase: adds a methyl group

Helicase: unwinds DNA double helix

Primase: adds an RNA primer

Carboxylase: Adds a CO_2 group with the help of a biotin cofactor

Carboxykinase: Removes a CO_2 group and adds a phosphate using ATP or GTP

Thiolase: Uses free CoA-SH to cleave a carbonyl compound into two thioesters

Aminotransferase: Replaces a double-bonded oxygen with an amino group

Decarboxylase: Removes CO_2 from a molecule

Dehydratase: Removes H_2O from a molecule, typically creates a pi bond

Protease: Cleaves amide bonds in protein molecules

Reverse transcriptase: makes mRNA into cDNA

Integrase: incorporates (typically viral) DNA into the host's genome

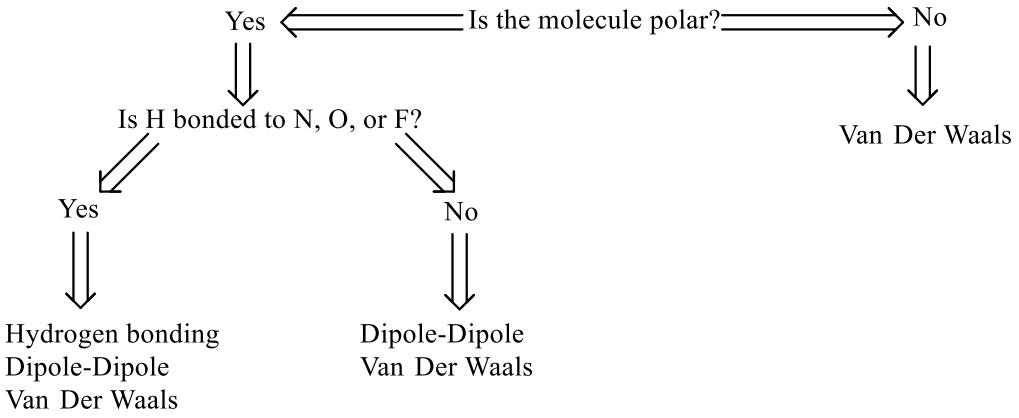
Focus 1: Protein Structure and Function + Biochem Basics

Chapter 1: Intermolecular Forces and Common Biological Functional Groups

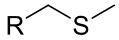
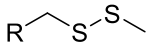
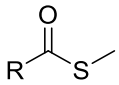
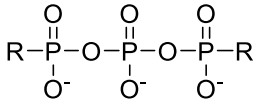
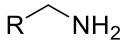
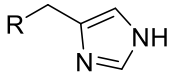
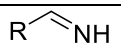
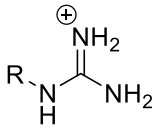
Intermolecular forces (IMF) are the crux of Biochemistry overall and are incredibly important for rationalizing protein structure and protein-drug interactions. Because of the importance these forces have in the realm of Biochemistry, we will do a quick review of what we know from General Chemistry and Organic Chemistry and then introduce some of the common biological functional groups as well as the IMFs they can participate in. Much of this is already covered in the General Chemistry or Organic Chemistry survival guide, so for a more thorough of this material, you can consult those guides. Broadly, there are four different types of IMFs that are relevant in Biochemistry. These forces are listed below in order of increasing strength (with the strongest on the bottom):

1. Van Der Waals interactions
2. Dipole-dipole interactions
3. Hydrogen bonding interactions
4. Ionic bonding interactions

I take it that most students who are in Biochemistry have a working understanding of how and which molecules are polar, therefore, to determine which IMFs any given molecule can participate in, you can use the following flowchart:



Now using this paradigm, we can introduce the common biological functional groups and determine which kinds of IMFs they could participate in. A list of the common biological functional groups is given below:

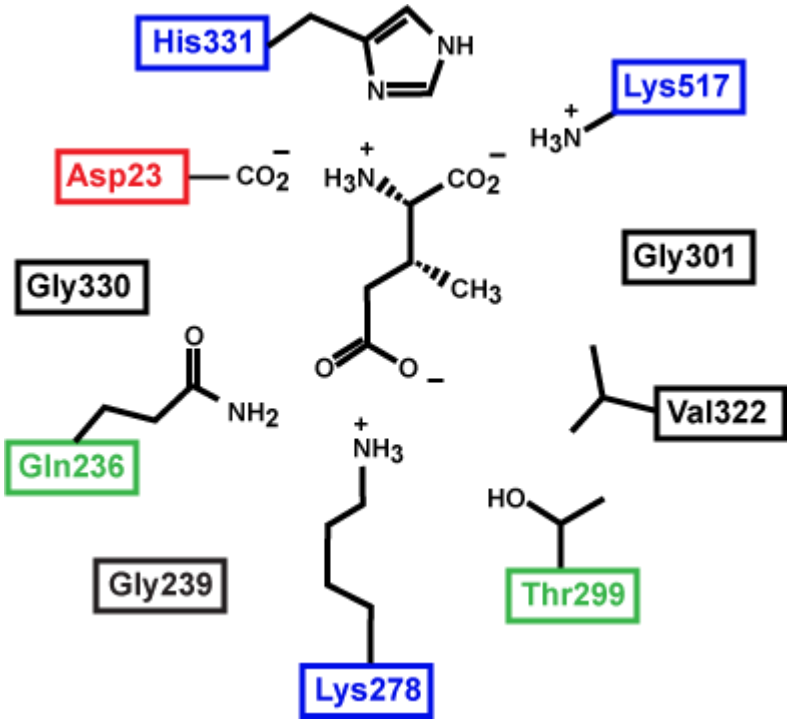
Functional Group	Example Structure	IMFs Possible	Chemical Reactions
Thiol		Dipole-dipole, Van Der Waals	Acid/base chemistry, redox chemistry to make disulfides
Sulfide		Dipole-dipole, Van Der Waals	Not much
Disulfide		Dipole-dipole, Van Der Waals	Redox chemistry to make thiols
Thioester		Dipole-dipole, Van Der Waals	HIGH ENERGY BONDS, can be used to power cellular machinery
Phosphoanhydride		Ionic, Dipole-dipole, Van Der Waals	HIGH ENERGY BONDS, can be used to power cellular machinery
Amine		Hydrogen bonding, dipole-dipole, Van Der Waals	Acid/base chemistry, nucleophilic
Imidazole		Hydrogen bonding, dipole-dipole, Van Der Waals	Acid/base chemistry, nucleophilic
Imine		Hydrogen bonding, dipole-dipole, Van Der Waals	Electrophilic
Amide		Hydrogen bonding, dipole-dipole, Van Der Waals	Peptide bond formation
Guanidinium		Ionic, Hydrogen bonding, dipole-dipole, Van Der Waals	Electrophilic
Carbonyl		Dipole-dipole, Van Der Waals	Electrophilic
Carboxylic acid		Hydrogen bonding, dipole-dipole, Van Der Waals	Acid/base chemistry, in basic form nucleophilic
Alcohol		Hydrogen bonding, dipole-dipole, Van Der Waals	phosphorylation

As you can see, many of these functional groups are capable of hydrogen bonding. This is no coincidence because hydrogen bonding is the strongest intermolecular force other than ionic

bonds and this is what can help drive protein folding. Maximizing hydrogen bonding and other intermolecular forces stabilizes the protein and therefore minimizes its energy (**M**ASC).

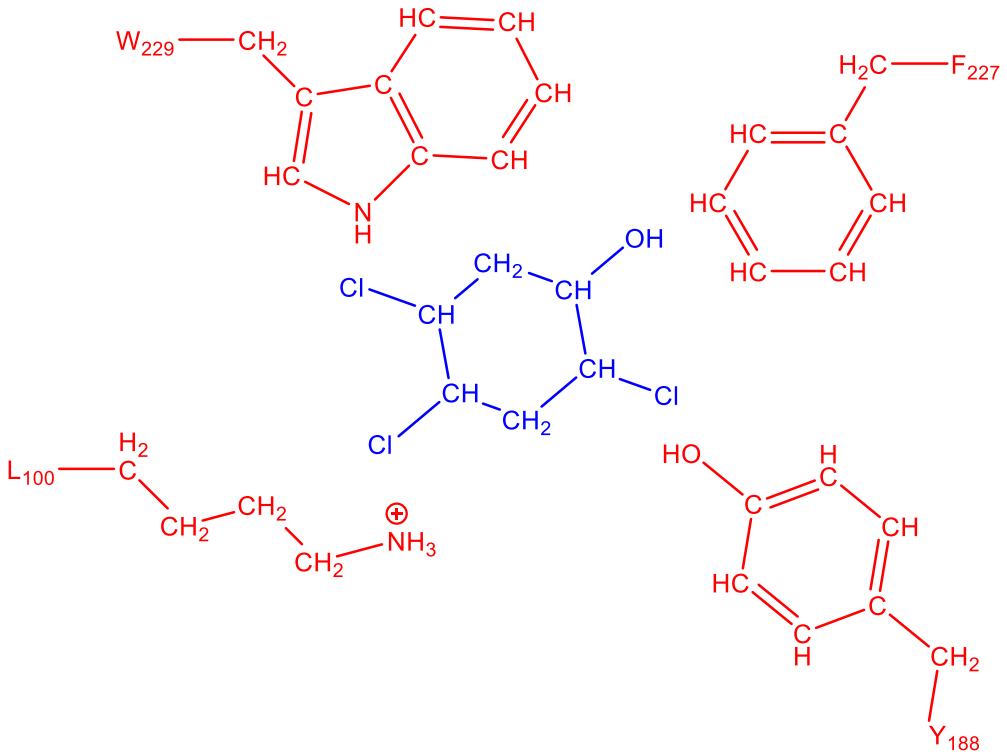
Here are some practice problems from the General Chemistry Survival Guide that are relevant to Biochemistry:

1. Identify all relevant intermolecular forces present in the following enzyme's active site:



2. It is your first day as a medical research scientist and you are attempting to design a drug to target HIV reverse transcriptase to prevent the spread of the HIV virus within the human body, the active site of the enzyme (the part of the enzyme with which the drug interacts) is shown below and your predecessor's molecule is shown in the active site as well:

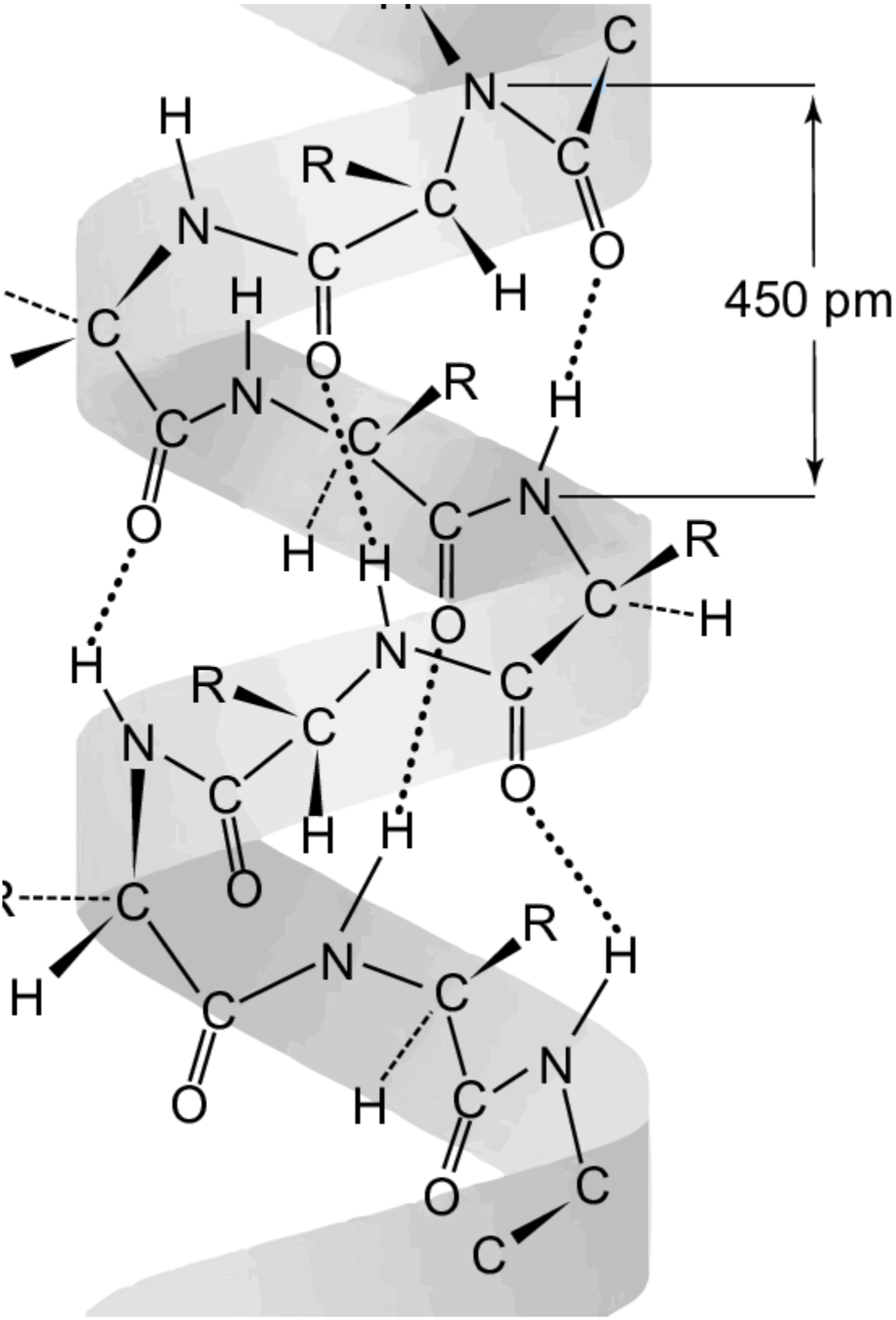
Red = Active site of enzyme
Blue = Drug compound



a. Identify all intermolecular forces that are keeping your predecessor's molecule in the active site.
b. Suggest ways to improve its binding affinity

3. Cellulose and glycogen are biomolecules that are both polymers of glucose, meaning that they are made of long chains of glucose molecules. Despite them both being polymers of glucose, cellulose is much tougher and stronger than glycogen. Explain this behavior in terms of intermolecular forces.

4. The picture below shows the secondary structure of a protein, identify the critical intermolecular forces that are keeping this secondary structure stable:

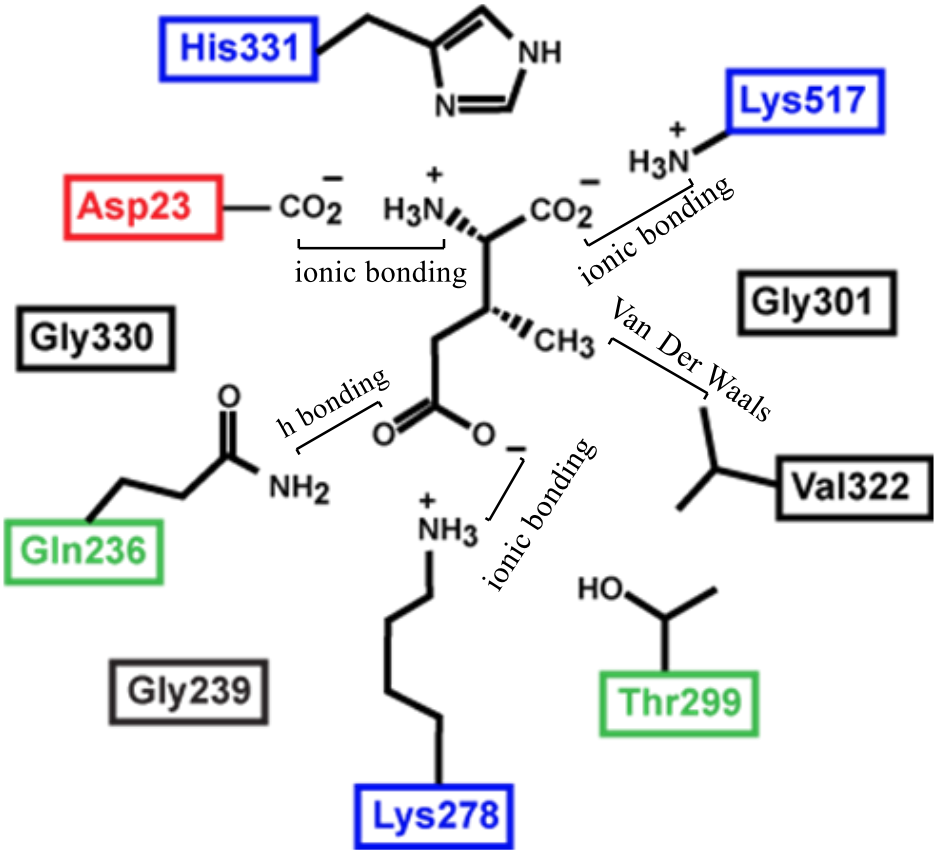


5. A drug shows great promise in treating the coronavirus. When it is administered into the blood, it is found that patients with severe acidosis, or acidic blood, respond poorly to this

medication while patients with severe alkalosis, or basic blood, respond substantially better than patients with neither condition. Suggest an explanation for this occurrence on the premise of intermolecular forces.

Answers:

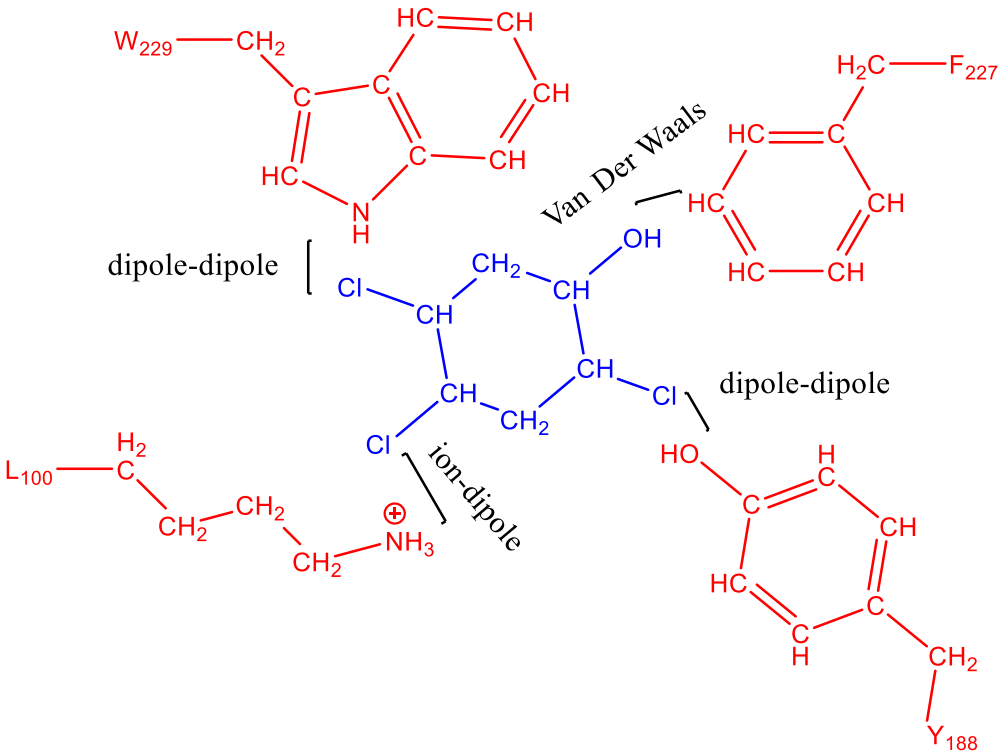
1.



2.

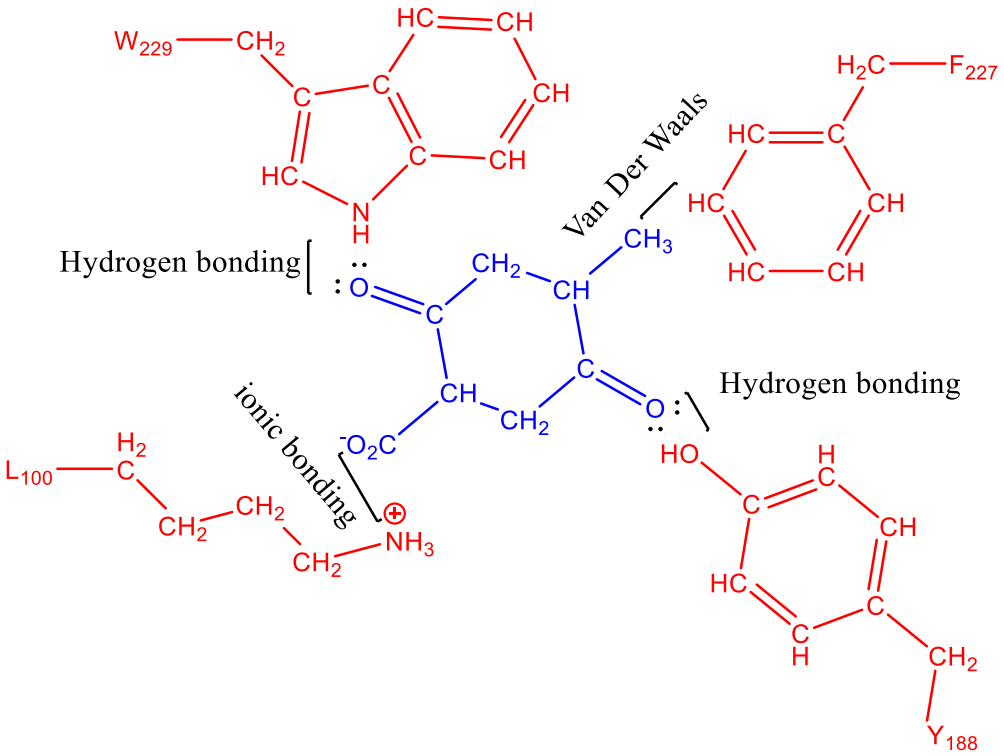
Red = Active site of enzyme
Blue = Drug compound

a.



b. A lot of possible answers here, the key is to change the blue molecule such that it has stronger intermolecular forces, one such response would be something like this:

Red = Active site of enzyme
Blue = Drug compound



The reason why changing the Cl's to O's makes the dipole-dipole interaction into hydrogen bonding is because O can act as a hydrogen bond acceptor while Cl cannot because it is not one of the special elements that can participate in hydrogen bonding (N,O, and F).

3. This was a tricky question admittedly, but it has to do with the orientation of the hydrogen bonds. You had to be able to figure that out because both molecules are polymers of glucose, so there is no difference between them except the orientation of the glucose. Because cellulose is stronger and tougher than glycogen despite being made of the same thing is because cellulose can orient its glucose chain so that it has 180 degree hydrogen bonds which maximizes their strength and makes the material stronger.

4. This alpha helical structure (the specific type of secondary structure shown in the picture I gave you) is held together by NH-O hydrogen bonds as shown by the dashed lines. These intermolecular forces are very strong and they can drive the formation of special structures like helixes that otherwise would not exist. We know it has to be hydrogen bonding because the NH portion of the protein acts as a hydrogen bond donor and the O acts as a hydrogen bond acceptor.

5. This is not a real situation, but if the situation was true, then it would stand to reason that the acidic blood was causing the loss of a negative charge that was critical for the drug to bind to it. Likely the drug interacted with the coronavirus with some type of ionic bonding or ion-dipole interactions that caused very strong binding and therefore the drug was very potent. The basic solution caused the negative charge of whatever the drug target was to become more pronounced and therefore the drug was able to bind to stronger, strengthening the interaction and making the drug work better.

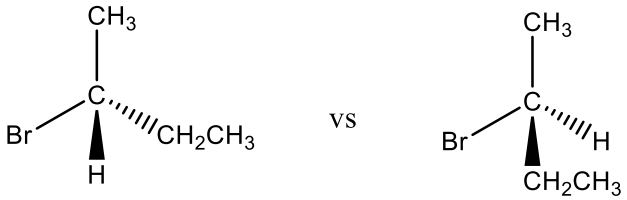
Chapter 2: Arrangement of Atoms in Space (Stereochemistry)

Enzymes and other proteins rely on their specific three-dimensional structure to do their function (MA**S**C), because of this, only specific molecules with the right stereochemistry can bind to them. Here we will discuss very briefly how to assign stereochemistry and list some examples where the wrong stereochemistry can have drastically different biological effects. For a more comprehensive review of this material, you can check out the Organic Chemistry survival guide. Most of the material in this chapter will be sampled from that guide with some tie-ins to Biochemistry where appropriate.

First let's go over when we know stereoisomerism exists. In this course there are two forms of isomerism that are important, R/S and cis/trans. For a compound to have R/S isomerism, there needs to be an asymmetric carbon, or a carbon with four different groups on it. For a compound to have cis/trans isomerism, each alkene carbon must have two different groups on it. The way that we assign R/S and cis/trans is very similar, but we will go through the R/S method first. To assign chirality to an asymmetric carbon, you need to do the following things:

1. Identify all four of the distinct groups
2. Assign priority to the groups on the premise of molecular weight of the atom attached directly. Keep doing this and going down the chain until you get to the point of first difference. You must evaluate both the atom in question and also the atoms to which it is bonded to.
3. If the group has a double or triple bond, pretend that carbon is bonded to two or three of that element.
4. Number all groups 1-4
5. Move your finger from 1 to 2 to 3 to 4.
6. If the hydrogen is on a dashed line, a clockwise motion of the finger would mean an R configuration, a counterclockwise motion would mean an S configuration. If the hydrogen is on the wedged line, flip the configuration assignment.

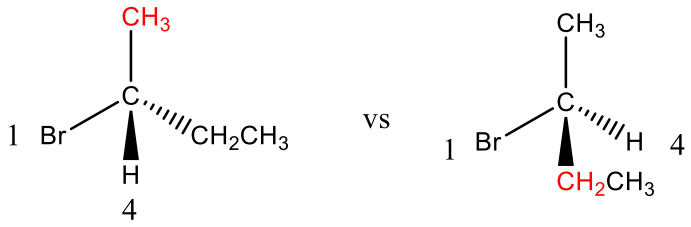
Here is an example with a worked out explanation:



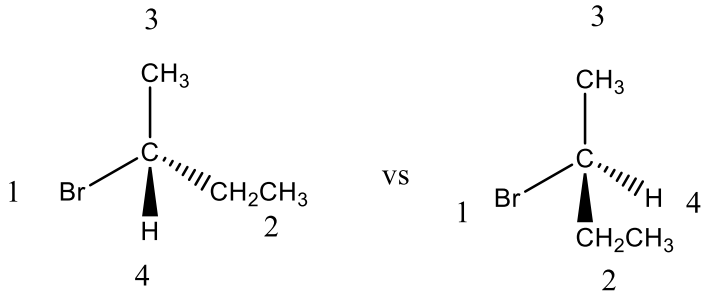
let's first try to assign priority to the groups attached to the asymmetric carbon:

first we look at the atoms directly attached to the asymmetric carbon and we evaluate their molecular weights, with the highest molecular weight getting highest priority. From this we know that Br is the highest priority group and H is the lowest priority group, but how do we

distinguish between CH_3 and CH_2CH_3? We have to go another atom down in the chain and see which group has higher priority groups attached. So far we have this:



We now need to evaluate the groups attached to the carbons indicated in red. The CH_3 group has 3 H's attached, while the CH_2CH_3 group has 2 H's and a C attached, because C > H the CH_2CH_3 group gets higher priority. Therefore we have this for the priority of the groups:

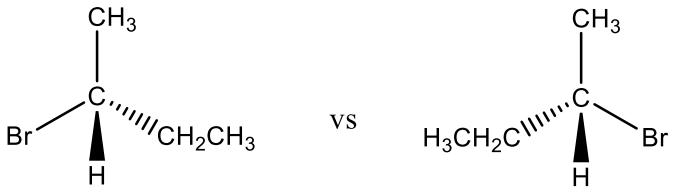


Now that we have the groups assigned proper priority, we simply need to move our finger from 1 to 2 to 3 to 4, if the direction of that rotation is clockwise and the lowest priority group is on a dashed line, the configuration is R and the configuration is S if the rotation is counterclockwise. Those assignments are *only if the lowest priority group (usually a H) is on a dashed line*, if it is on a bold-faced line you need to flip the assignment, if it is on a regular line, rotate the molecule so it is on either a bold-faced or dashed line.

For the molecule on the left, going from 1-2-3-4 is a counterclockwise rotation, which ordinarily would mean an S configuration, BUT, the lowest priority group (H) is on a bold-faced line, therefore the configuration is the opposite. The chirality assignment for the molecule on the left is R.

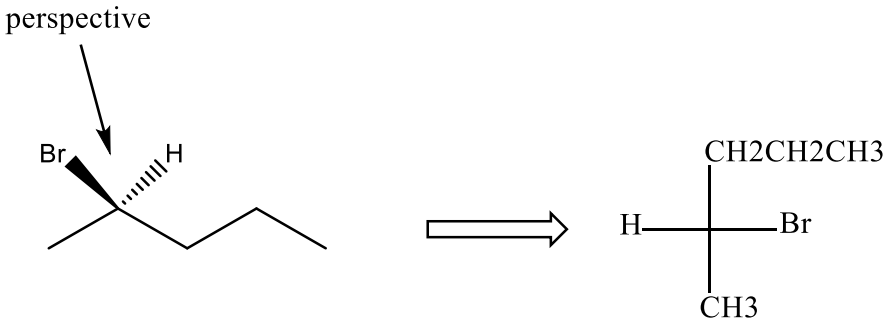
For the molecule on the right, going from 1-2-3-4 is a counterclockwise rotation as well, because the lowest priority group is on a dashed line, the configuration assignment is normal and the molecule is S. This relationship is very important, *switching the dash and bold-faced lines will cause inversion of stereochemistry*.

These two molecules (the one on the left and right) are referred to as *enantiomers*, and they are perfect mirror images of each other, meaning, it would also be valid to depict them like this:

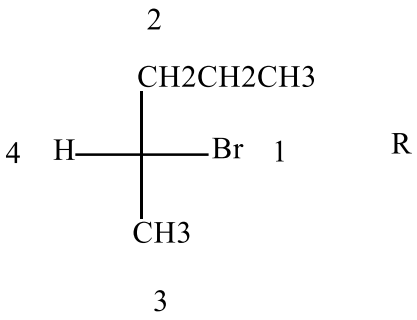


In Biochemistry, oftentimes we do not write it like this, but instead we write certain compounds like sugars in Fischer projections, which we will cover now.

Fischer projections are frequently encountered in carbohydrate chemistry and are very useful in determining the stereochemistry of specific chiral centers. Fundamentally, the Fischer projection requires the chemist to view the molecule from above or below the dash and wedge lines for each chiral carbon. For example,

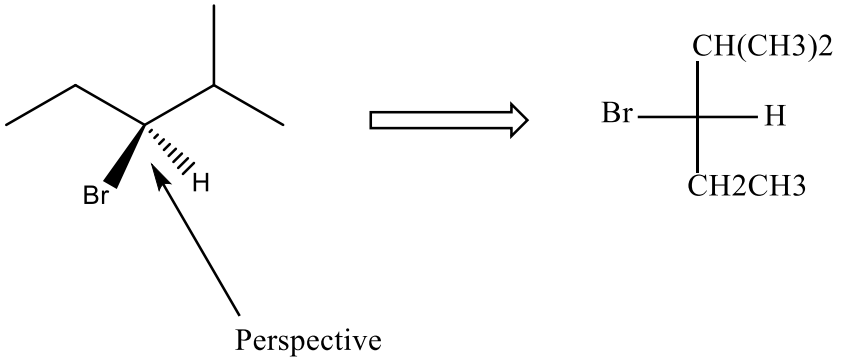


To do a proper Fischer projection, you effectively need to see in between the Br and H and view the molecule from above. If you do this, the H is to your left, the Br is to your right, the CH_3 is downward, and the rest of the chain is facing upward. To assign chirality to this carbon, you do the exact same procedure as normal (assign priorities and see if it turns clockwise or counterclockwise). If the H is on a horizontal line, you must flip whatever assignment you give it because if it is on a horizontal line, it is on a bold line from your perspective. Therefore for this example, the following would be your priority list and subsequent assignment:

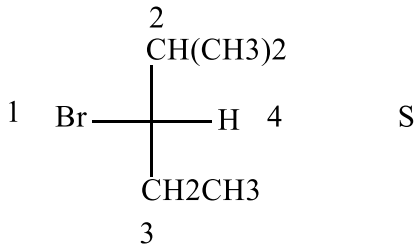


R

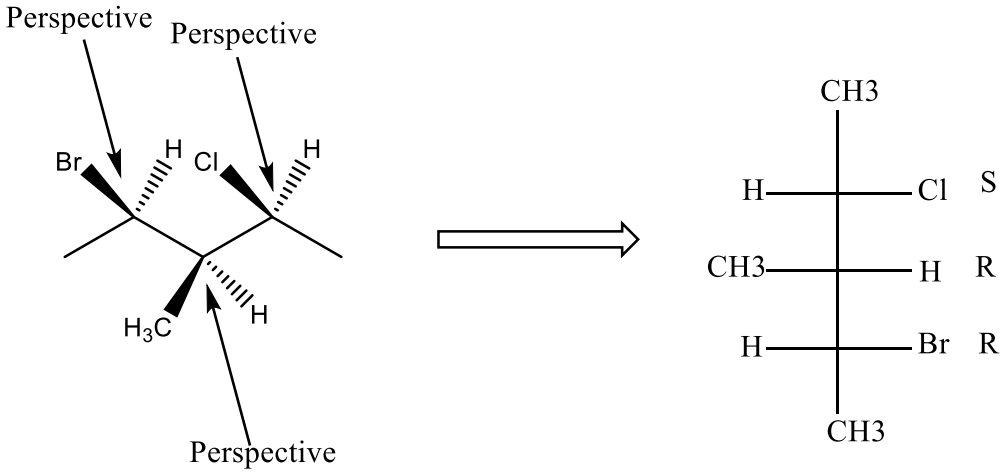
If you do it the traditional way, you will get the same result. This projection comes in handy particularly when you are faced with a situation where the lowest priority group is not on the bold or dashed line and is also very helpful when you are dealing with long chains of chiral carbons such as carbohydrates. You can also do this if the dash and wedge are facing downwards:



The configuration of that chiral center is S according to the Fischer and using the traditional way of solving chiral configurations. The Fischer way is shown below:

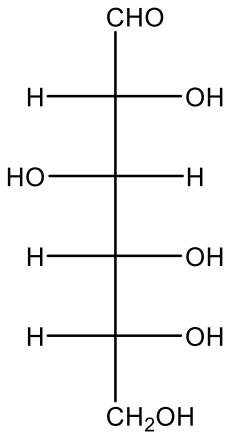


If there is a combination of several chiral centers in succession, you have to constantly adjust the perspective to accommodate the direction of the dash and bold lines. For example:

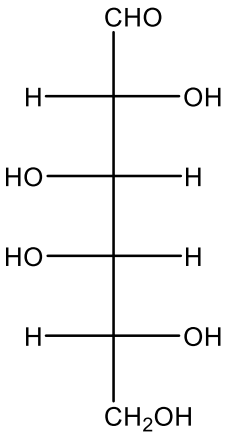


The chirality of the centers from top to bottom would be SRR according to the Fischer and traditional way. In the above Fischer projection, each of the intersection points represents a chiral carbon in the parent compound. In this way, we can make a single Fischer projection that gives the stereochemistry for the entire compound. Now we can assign the R/S chiralities to a carbohydrate molecules to make it more “Biochemistry-like”:

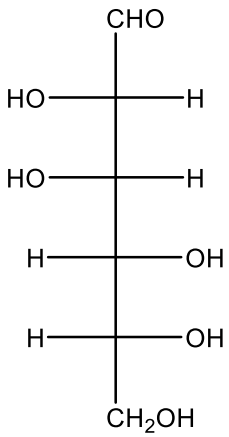
Glucose



Galactose



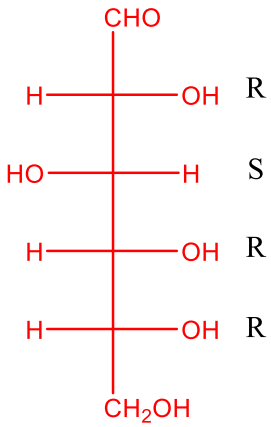
Mannose



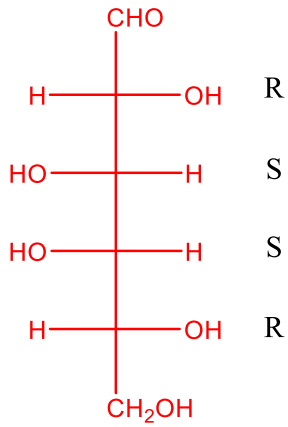
Assign the R/S chirality to each chiral center (asymmetric carbon) in the molecules above.

Answer:

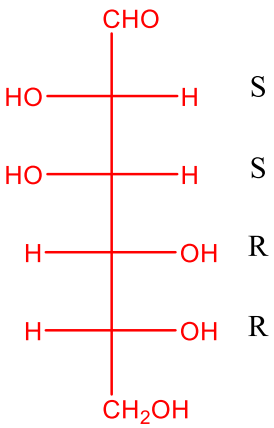
Glucose



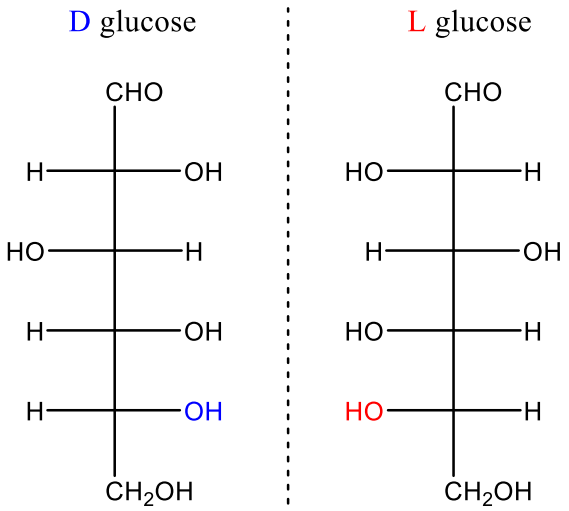
Galactose



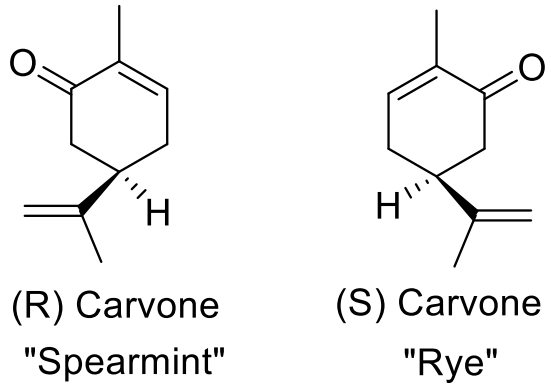
Mannose



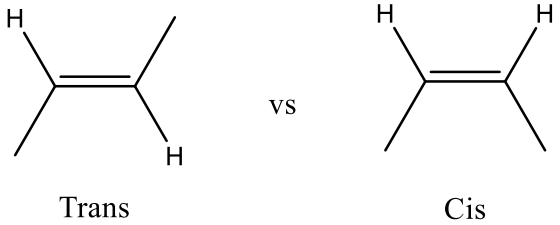
To make this more Biochemistry-like, we will now introduce the concept of D and L notation. If the aldehyde carbon, or more generally, the most oxidized carbon, is on the top and the OH on the bottom-most chiral carbon faces right, then the molecule is said to have the D configuration. Conversely if the OH on the bottom-most chiral carbon faces left, then the molecule is said to have the L configuration. Therefore, the molecules on the top are all the D configuration because all of them have the OH group on the bottom facing right. This is how all naturally occurring carbohydrates are, they are all the D configuration. The opposite is true for naturally occurring amino acids, they are all the L configuration. An example of the enantiomer of D glucose is L glucose, shown below:



R/S stereoisomerism is incredibly because stereoisomers have very different biological effects. One example of this are the stereoisomers of carvone, which show that the olfactory receptors, or smell receptors, are chiral. The R stereoisomer of carvone smells like spearmint while the S stereoisomer smells like rye, two very distinguishable odors.



Alkenes can also have stereochemistry in that they can be either cis or trans. The *cis* isomers are those that have the highest priority groups pointing in the **same** direction while the *trans* isomers are those that have them pointing in **opposite** directions like so:

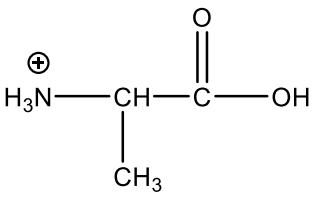


The prioritization of the groups is done in the exact same way as the R/S system.

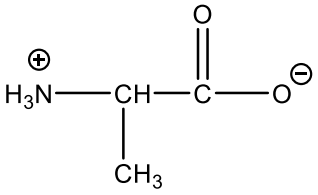
Chapter 3: Titration Curves and Protein PI

Amino acids are all at the very minimum diprotic, meaning they can give two protons. This is because of their amino and carboxylic acid groups. Because they have these two functionalities, amino acids can exist as either the fully protonated form, the zwitterionic form (neutral overall charge), or the fully deprotonated form shown below:

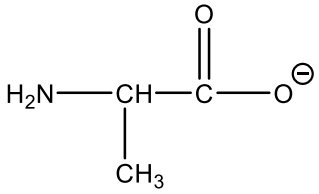
Fully protonated
pH 0



Zwitterion
pH 7



Fully deprotonated
pH 14

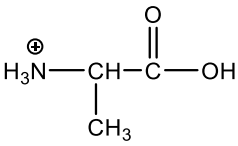


Because of this unique chemistry, amino acids are effectively diprotic (or triprotic for the acidic and basic amino acids) acids and will have a titration curve as such. Each functional group that has an acidic proton has a corresponding pKa and this value corresponds to the pH, which has the conjugate acid/base pair in equal concentrations. This can be seen mathematically using the Henderson Hasslebach equation:

$$pH = pKa + \log\left(\frac{CB}{CA}\right)$$

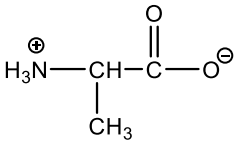
The amino acid above, alanine, has two acidic regions, the amino group and the carboxylic acid group. These two functional groups have pKas of 2.34 and 9.69, therefore, the following would be the case:

Fully protonated
pH 0



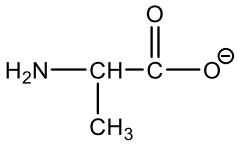
at pH 2.34:
zwitterion =
fully protonated

Zwitterion
pH 7

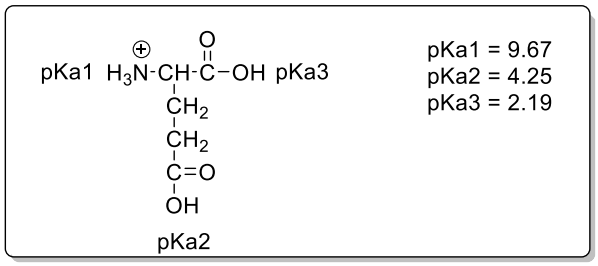


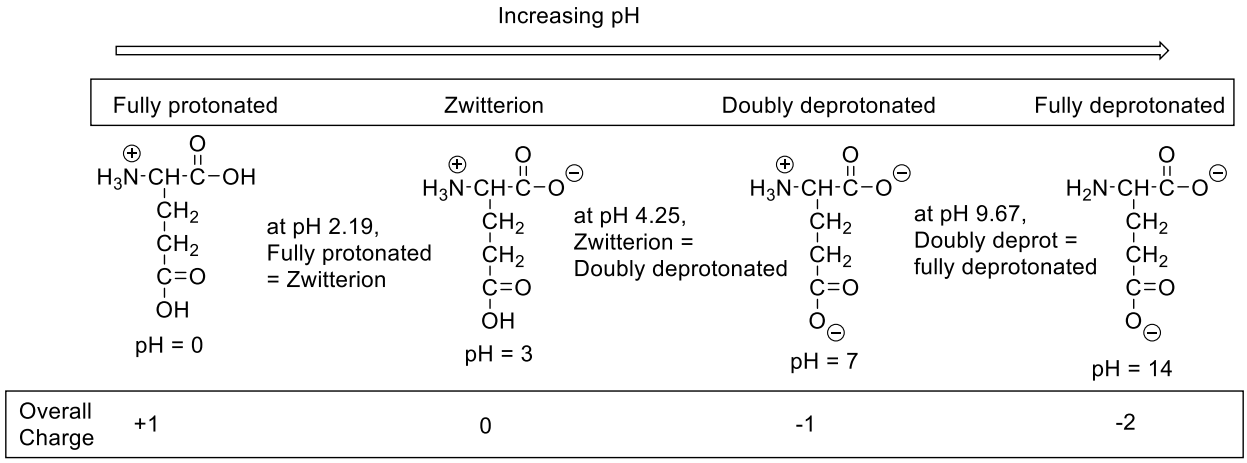
at pH 9.69:
zwitterion =
fully deprotonated

Fully deprotonated
pH 14



I chose the pH values of 0, 7, and 14 because they represent the two extremes (0 and 14) and because 7 is in between the two pKa values for alanine. Any value in between the two pKa values would work since the zwitterion is the predominant species in that entire range of pH values. This will be discussed in greater detail in the titration curve portion of the chapter. We can do the same thing with an acidic amino acid such as glutamic acid like so:





Just as before, I chose 3 and 7 because they were between the two flanking pKa values ($2.19 < 3 < 4.25$ and $4.25 < 7 < 9.67$). The trend regarding the overall charge is important to note, as pH increases, the overall charge becomes more negative by increments of one. It always follows positive to zero to negative. This allows us to define our pI. The pI is the pH at which the amino acid or peptide chain is electrically neutral. To get the pI, it is a simple averaging process:

$$pI = \frac{pKa_{low} + pKa_{high}}{2}$$

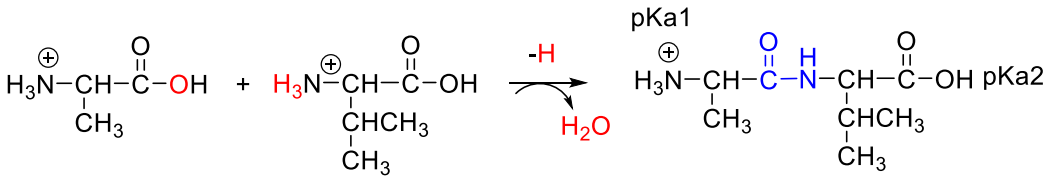
In this equation, pKa_{low} represents the pKa that is just below the pH where the amino acid or peptide is neutral while the pKa_{high} represents the pKa that is just above the pH where the amino acid or peptide is neutral. For glutamic acid, we can solve for the pI by taking the average of pKa_2 and pKa_3 because the amino acid is neutral at pH 3 like so:

$$pI = \frac{2.19 + 4.25}{2} = 3.22$$

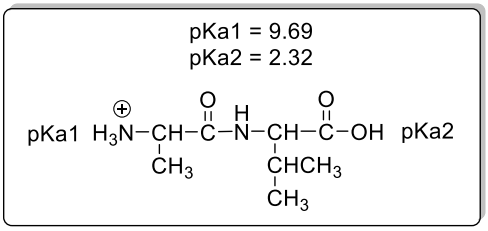
Try to see if you can calculate the pI for alanine given the data that I provided you above.

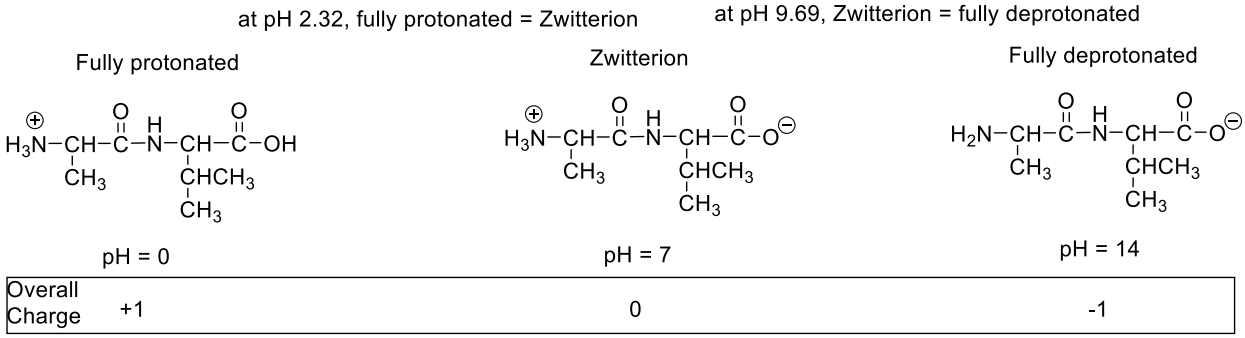
If you got 6.02 then you did it correctly. The pI is a unique property for each amino acid and each peptide. Because of this, the pI can be used to separate peptides from each other using isoelectric focusing. In this process, an electric field is applied across a pH gradient that has well defined pH ranges in each section. When the pH matches the pI, the peptide stops moving and in this way, the peptide can be separated from other peptides that have different pI values.

So far, we have only talked about pI in the context of amino acids, but how does this change when we form polypeptide chains? The answer is that it doesn't change all too much, we will start with a dipeptide to show you how it changes and then extend the argument to a peptide of any length. Consider the dipeptide comprised of alanine and valine below:



When the peptide bond forms between alanine and valine, the carboxylic acid group of alanine becomes an amide and the amino group of valine becomes part of the amide (blue) as well. Because of this, the pKa of these groups are now irrelevant because amides are not acidic and the only thing that we have to worry about are the pKa of the N terminal residue (alanine in this example) and the C terminal residue (valine in this example). Because there are no acidic or basic residues in this peptide, we do not have to worry about side chains. We will get to that later in the next example. For right now, let's just focus on calculating the pI of the valine-alanine dipeptide by doing the same method as before:

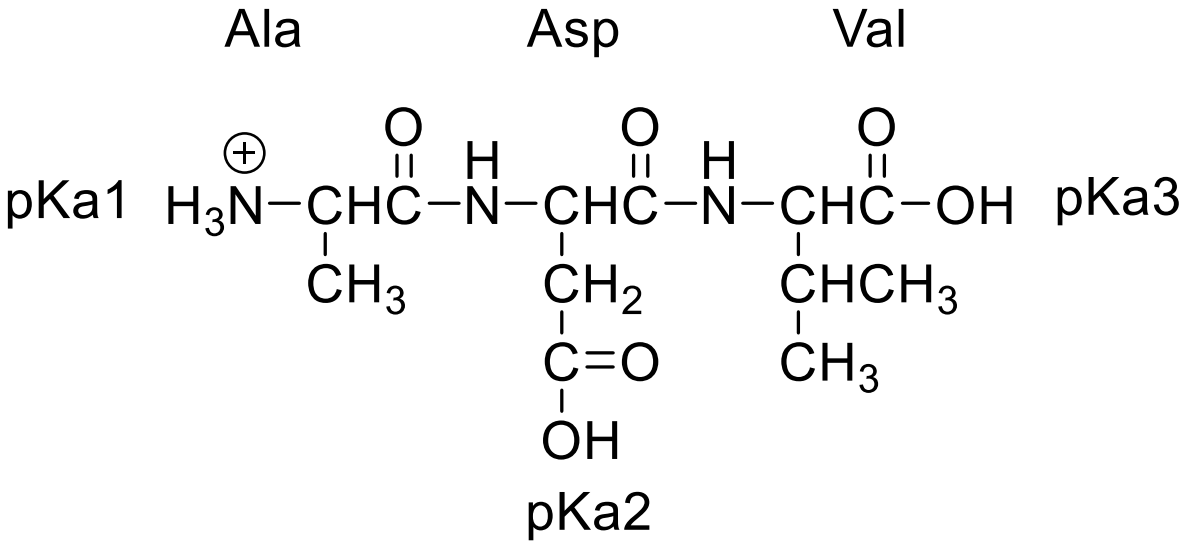




Because the peptide is neutral at pH 7, we must look for the pKas that are closest above and below 7, which in this case are 2.32 and 9.69. To calculate the pI, we simply just take the average of these two:

$$pI = \frac{2.32 + 9.69}{2} = 6.00$$

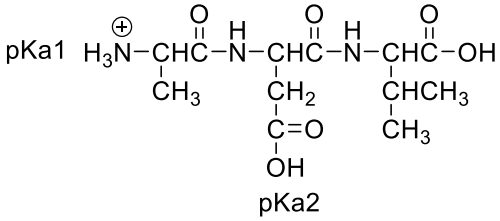
Now we can begin to discuss the pI of polypeptide chains that have acidic or basic amino acids in them. We can consider the tripeptide comprised of alanine, aspartic acid, and valine shown below:



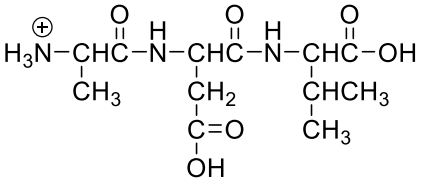
Before we tackle this problem, if aspartic acid is an acidic residue, what do you think its effect on the pI will be? Will it increase or decrease it? Well the pI is the pH that makes the peptide neutral and since acids become negative at lower pH values, logically the pI should decrease with acidic peptide incorporation. Let's see if this logical intuition works out (it does).

Ala Asp Val

pKa1 = 9.69
pKa2 = 3.65
pKa3 = 2.32



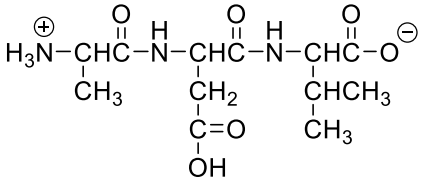
Fully protonated



pH = 0

Overall charge = +1

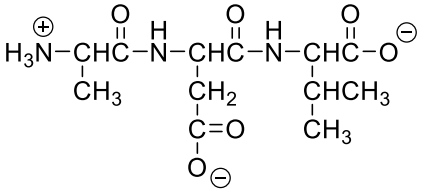
Zwitterion



pH = 3

Overall charge = 0

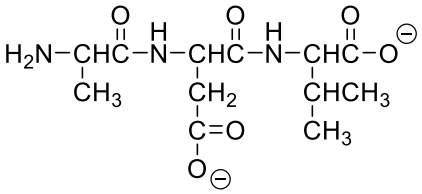
Doubly deprotonated



pH = 7

Overall charge = -1

Fully deprotonated



pH = 14

Overall charge = -2

Increasing pH

Now the peptide is neutral at pH = 3, this defines our neutral zone and therefore we must find the closest pKas that are above and below this pH. This turns out to be 2.32 and 3.65. Now the pI can be easily calculated just as before:

$$pI = \frac{2.32 + 3.65}{2} = 2.99$$

Now we can compare the two pIs, before the aspartic acid was put into the peptide, the pI was 6.00 and with the aspartic acid the peptide has a pI of 2.99. We effectively halved the pI just by including one acidic residue in the peptide. Effectively there are only three things that we need to consider when we do these problems:

1. The pKa of the amino group of the N terminal residue
2. The pKa of the carboxylic acid group of the C terminal residue
3. The pKa of all acidic or basic sidechains

Practice Questions:

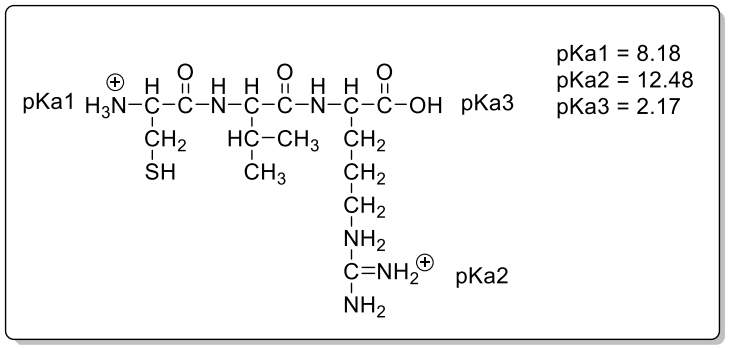
For all of these questions, use the following pKa values[3]

Amino acid	pKa of Carboxylic acid	pKa of amino	pKa of side chain
Glycine	2.34	9.60	--
Alanine	2.34	9.69	--
Valine	2.32	9.62	--
Leucine	2.36	9.60	--
Isoleucine	2.36	9.60	--
Methionine	2.28	9.21	--
Proline	1.99	10.60	--
Phenylalanine	1.83	9.13	--
Tryptophan	2.83	9.39	--
Asparagine	2.02	8.80	--
Glutamine	2.17	9.13	--
Serine	2.21	9.15	--
Threonine	2.09	9.10	--
Tyrosine	2.20	9.11	--
Cysteine	1.96	8.18	--
Aspartic Acid	1.88	9.60	3.65
Glutamic Acid	2.19	9.67	4.25
Lysine	2.18	8.95	10.53
Arginine	2.17	9.04	12.48
Histidine	1.82	9.17	6.00

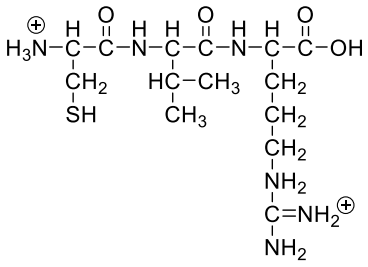
1. Find the pI for the tripeptide made of Cysteine-Valine-Arginine
2. Find the pI for the dipeptide made of Glutamic acid-arginine

Answers:

1.



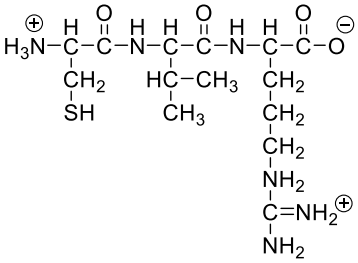
Fully protonated



pH = 0

Overall charge = +2

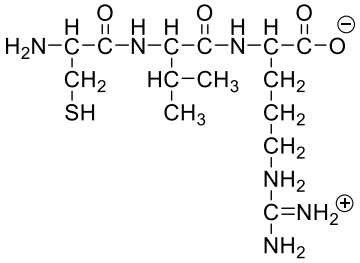
Monodeprotonated



pH = 5

Overall charge = +1

Zwitterion



pH = 9

Overall charge = 0

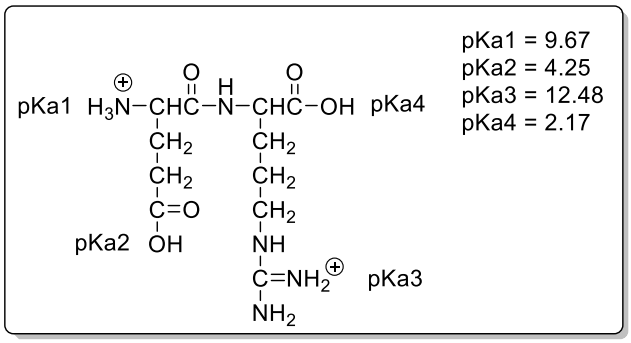
Increasing pH

Neutral zone

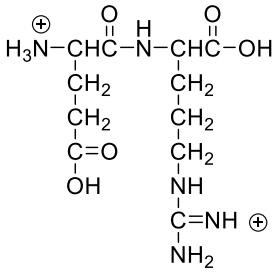
Closest pKa above pH 9 is 12.48
Closest pKa below pH 9 is 8.18

$$pI = \frac{8.18 + 12.48}{2} = 10.33$$

2.



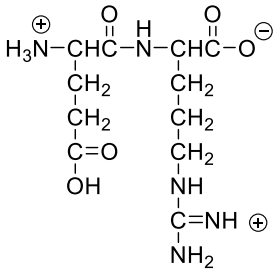
Fully protonated



pH = 0

Overall charge = +2

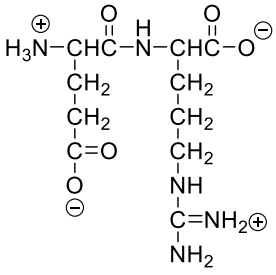
Monodeprotonated



pH = 3

Overall charge = +1

Zwitterion



pH = 6

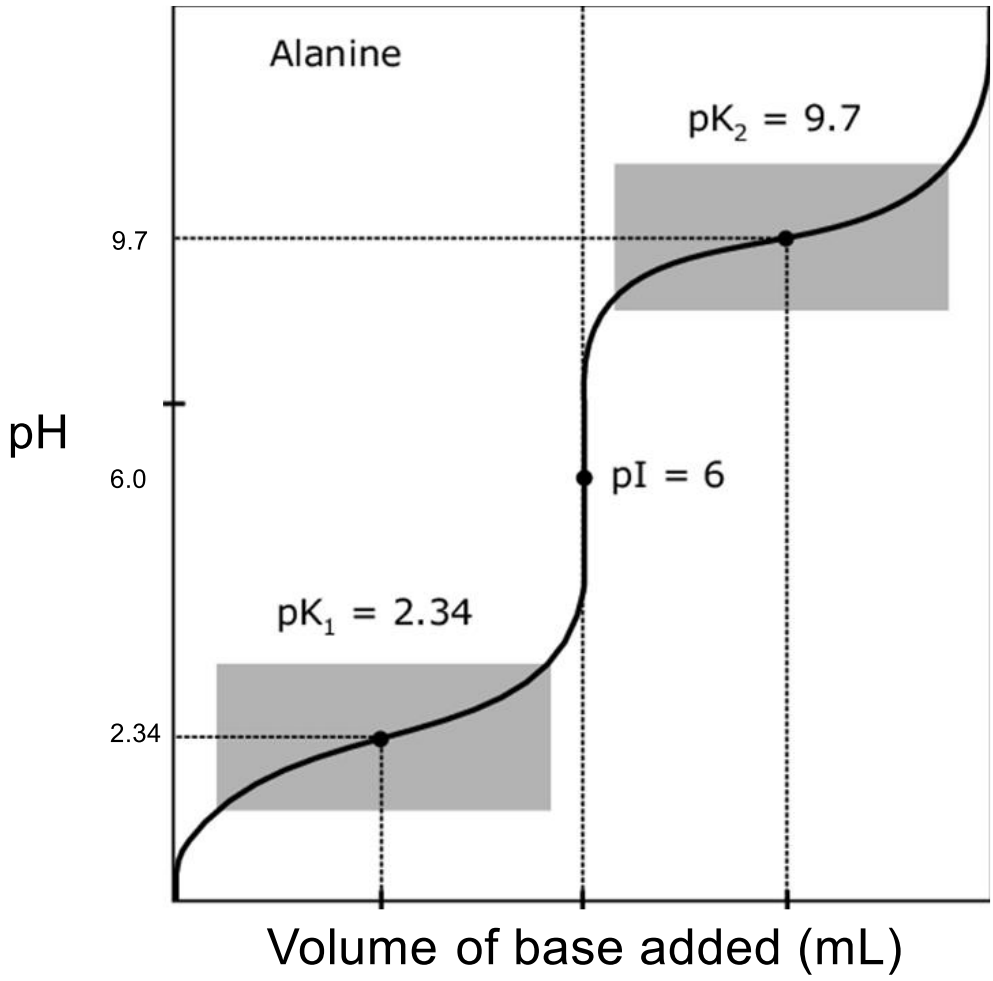
Overall charge = 0

Neutral Zone

Closest pKa above 6 is 9.67
Closest pKa below 6 is 4.25

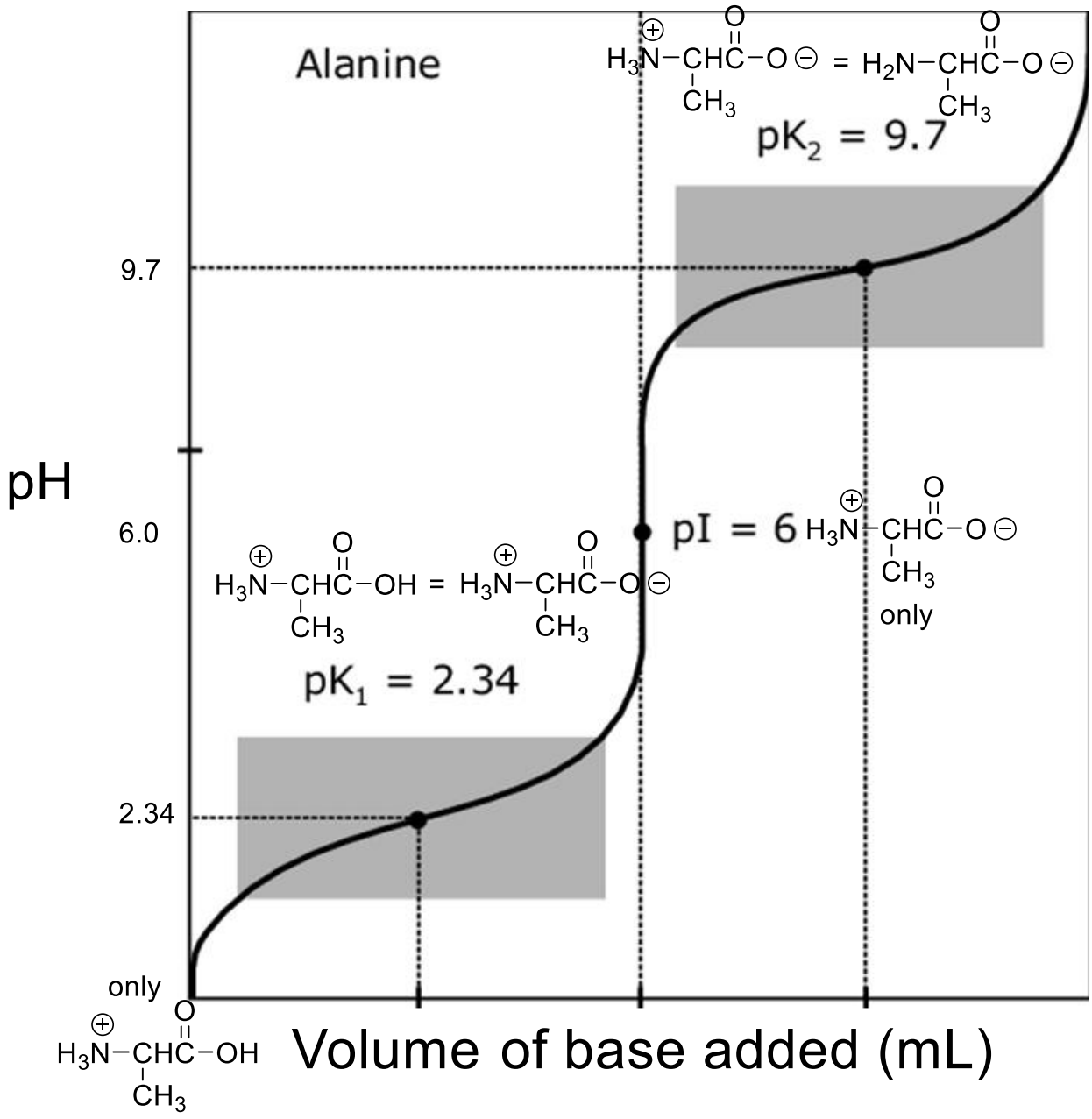
$$pI = \frac{9.67 + 4.25}{2} = 6.96$$

Now we can bring this knowledge to bear on something that is every Analytical Chemistry students' nightmare, titration curves. Titration curves plot pH against the volume of base added to the solution and they show the acidic behavior of the analyte. The fundamental principle behind titration curves is that bases will react with the most acidic functional groups first (lowest pKa) until all of that functional group is deprotonated then it will go to the next most acidic group etc. until everything is deprotonated. These titration curves can be used to determine the pKa because the pH will change the least (aka the slope will be the flattest) as the pH approaches the pKa. An example of the titration curve for alanine is shown below:[3]



We can start interpreting this titration curve by starting at pH = 0. At this pH, all of the alanine is in the fully protonated state, however, as the pH increases, more and more of the carboxylic acid group gets deprotonated until the half of this group is deprotonated. Once half of the carboxylic acid group is deprotonated, the pH = pK_1 shown in the diagram. After this point, the deprotonated Zwitterionic form predominates because more than half of the carboxylic acid group in alanine is deprotonated. Eventually all of the carboxylic acid group is deprotonated and

the only thing in the solution is the Zwitterionic form, that is when the pH = pI. Eventually, the base will start taking the proton off of the NH_3^+ group of alanine until half of the NH_3^+ becomes NH_2, that corresponds the pH = pK_2 in the diagram. After pH = pK_2, the fully deprotonated form predominates like so:



The same logic applies for amino acids that have acidic or basic side chains and the same analysis can be done for those amino acids except those amino acids will have more “flat” regions because they have additional pKas.

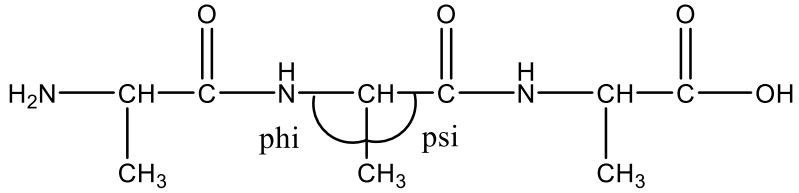
Chapter 4: Protein Structure Basics

In this chapter, we will discuss the basic terminology and types of local structures that can occur in proteins. Broadly, there are four levels of protein structure; these are the primary, secondary, tertiary, and quaternary structures.

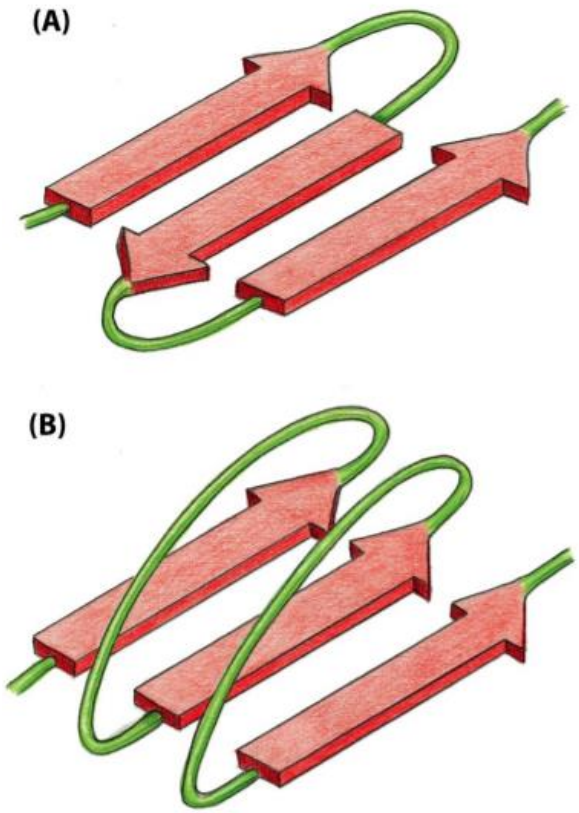
The primary structure is simply the amino acid sequence that comprises the peptide. The secondary structure is referring to how the hydrogen bonds of the peptide bonds allow the peptide to fold; there are two different types of secondary structures: alpha helices and beta sheets. These different structural motifs will be covered later in this chapter in great detail and we will determine how one can determine protein secondary structure on the premise of its amino acid sequence alone. The tertiary structure is how the side chains interact via IMFs to give the protein its overall shape. The quaternary structure is how different monomer units or subunits combine to yield the overall protein complex. An example of quaternary structure is hemoglobin, which is a dimer of dimers (ααββ). Overall protein shape is determined by several IMFs and covalent bonds:

1. Hydrogen bonds
2. Salt bridges (Ionic bonds between amino acid residues)
3. Dipole-dipole interactions
4. Van Der Waals forces in the interior of the protein
5. Disulfide bonds between two cysteine residues. These are the strongest driving force if available because this is a covalent bond.

The secondary structure is where we will focus our attention since this is the only structural component of peptides that we can predict using simply the amino acid sequence. If you recall from Organic Chemistry, single bonds have free rotation, therefore the alpha carbon can rotate about the amide nitrogen and carbonyl carbon within a polypeptide. The angle between the alpha carbon and the amide nitrogen is referred to as the phi angle and the angle between the alpha carbon and the carbonyl carbon is the psi angle.



Specific secondary structures will have specific phi/psi angles and are restricted to these angles to allow their formation. The first secondary structure we will discuss beta sheets and then we will discuss alpha helices because alpha helices are more complicated. There are two different kinds of beta sheets, parallel (B), and antiparallel (A) as shown below:[4]



The antiparallel type of beta sheet maximizes hydrogen bonding because it allows the hydrogen bonds to be collinear, which maximizes their strength as shown below:

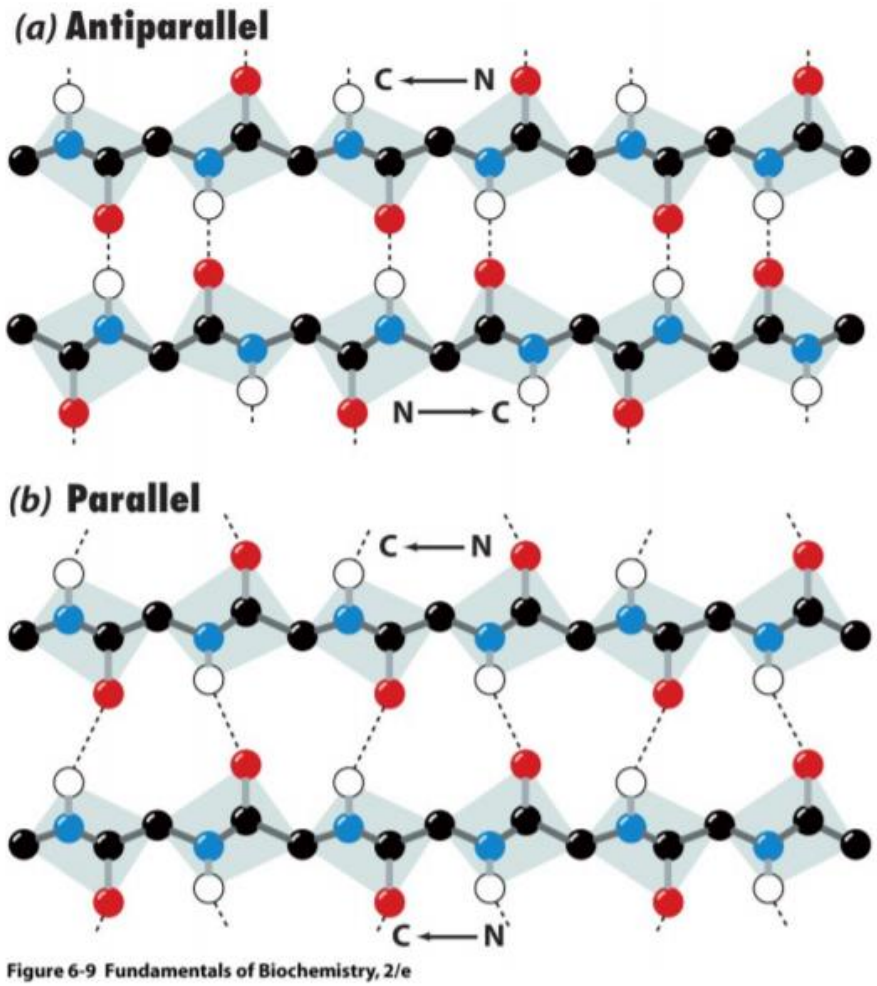
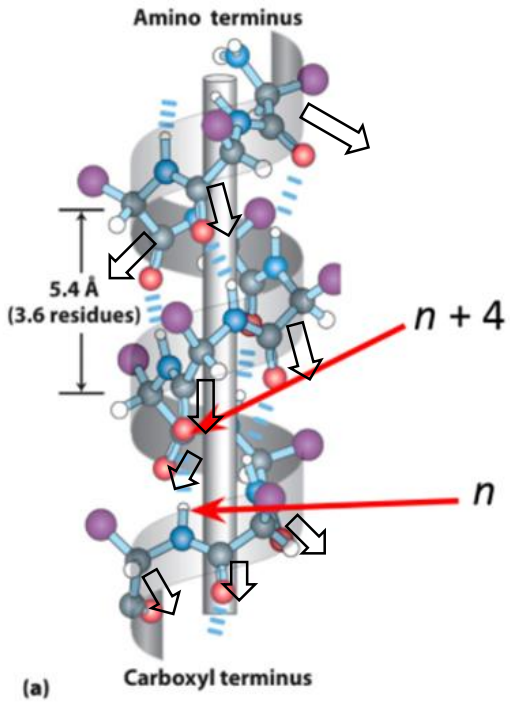


Figure 6-9 Fundamentals of Biochemistry, 2/e

Beta sheets are stabilized by beta-branched amino acids as shown in the amino acid notecards. Some materials are comprised entirely of beta sheets, such as silk fibroin. Silk fibroin is comprised of repeating glycine and alanine amino acids and are antiparallel which makes it very strong. The side chains of the amino acids in the beta sheet alternate facing up and down in beta sheets, because of this, amphipathic beta sheets can be made by alternating between polar and nonpolar amino acids. By doing this, one side of the beta sheet will be polar and interact with polar molecules/water and the other side will be nonpolar and interact with nonpolar molecules/membranes.

Alpha helices are incredibly complex by comparison and have many factors to perpetuate their formation over beta sheets. Alpha helices make a complete turn every 4 residues and are typically right-handed. Hydrogen bonding and favorable intermolecular forces are required for amino acids that are four residues apart because of this geometry. Amphipathic alpha helices can be made by making one side of the helix polar and the other nonpolar. This can be done by following the pattern PPNNPPNNPP… where P stands for polar and N stands for nonpolar. Because of this very specific geometry that alpha helices must take on, the extremely flexible amino acid glycine and the extremely rigid amino acid proline are commonly referred to as “helix breakers” and cannot be incorporated into stable alpha helices. In other words, we have to

look for amino acids that are in the “Goldilocks” region, not too flexible that they will not maintain the proper phi/psi angles and not too rigid that they cannot maintain the alpha helix phi/psi angle. Side chains of the amino acids that are present in the alpha helix face outward and the peptide bonds are aligned roughly parallel. Due to this parallel arrangement, alpha helices are unlike beta sheets in that they have a dipole moment. Shown below is an example of an alpha helix with the dipoles caused by the peptide bonds shown using arrows:[5]



Though the dipole moment roughly cancels out in the horizontal direction, all arrows shown above point down and therefore the N terminus of any alpha helix is partially positive and the C terminus of any alpha helix is partially negative. Due to this intrinsic dipole moment of the macromolecule, the two termini oftentimes must be “capped” to increase stability and minimize the dipole of the alpha helix (**M**ASC). To do this, alpha helices often contain acidic amino acids in the N terminal region and basic amino acids in the C terminal region. In between alpha helices and beta sheets, there are often beta turns. These beta turns are comprised of prolines and glycines and are often four amino acids long.

In summary, there are four things that must be present for an alpha helix to be stable:

1. Favorable interactions between any amino acid and the amino acid four residues ahead
2. Nonbulky amino acids four residues apart
3. N terminus with acidic residues and C terminus with basic residues
4. No prolines and glycines to break the helix

Because both alpha helices and beta sheets require specific and well-defined phi/psi angles, several analytical techniques have been developed to assess the phi/psi angles of peptides to elucidate the secondary structural motifs they possess. One such technique is called Circular

Dichroism spectroscopy, commonly referred to as “CD spec”. This technique measures the change in molar absorptivity of a peptide exposed to polarized light. These signals are assessed over wavelengths from 190 to 250 nm and this information can be used to see secondary structure characteristics of the peptide in question. Another technique is making a Ramachandran plot, which shows the distribution of phi/psi angles in the protein of interest with certain regions of the plot indicating specific local secondary structures. Overall protein structure is commonly assessed by X-ray crystallography and NMR. X-ray crystallography is a great method for getting a static picture of the protein, however, the protein needs to form a highly ordered crystal which is a difficult thing to accomplish and some proteins are not amenable to high crystallinity. NMR is a great method for capturing the protein in action, in other words, determining the dynamic motion of proteins in solution. However, protein NMR is incredibly complex even for small proteins and therefore this method is limited when used for larger proteins such as pyruvate dehydrogenase. More recently (i.e. in 2017) a newer method has been developed and used for determining the structure of biomolecules. This technique is called cryo-EM, or cryogenic electron microscopy. This technique is beneficial because it circumvents the need to crystalize the protein to get a high resolution structure of the protein and thus could be a superior means of macromolecular structure determination. In this technique, the analyte of interest, typically a protein in this context, is cooled to extremely cold temperatures in amorphous ice and electrons are shot at it via the transmission electron microscope to get an image of the protein at near-atomic scale.

Practice questions:

1. What type of secondary structure would the following amino acid sequence produce: Ala-Val-Asp-Glu-Leu-Ile-His-Lys
2. What type of secondary structure would the following amino acid sequence produce: Ile-Thr-Ile-Ile-Leu-Ser-Gly
3. What type of secondary structure would the following amino acid sequence produce: Pro-Gly-Pro-Gly
4. What type of secondary structure would the following amino acid sequence produce: Ala-Gly-Ala-Gly-Ala-Gly
5. What type of secondary structure would the following amino acid sequence produce: Phe-Phe-Phe-Phe-Phe
6. What type of secondary structure would the following amino acid sequence produce: Asp-Asp-Ser-Thr-Thr-Cys-His-His

Answers:

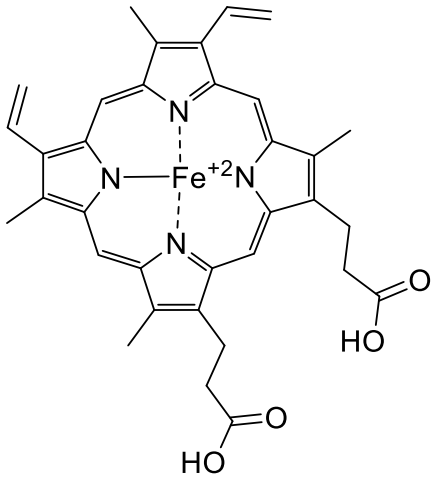
1. Amphipathic alpha helix because it is capped, has favorable interactions for residues four things apart, has no helix breakers, and follows the pattern NNPPNNPP
2. Beta sheet because it has many beta-branched amino acids, contains glycine which is a helix breaker, and is not capped.
3. Beta turn because it has alternating proline and glycine
4. Beta sheet, this is silk fibroin’s repeating formula
5. Beta sheet because it has many bulky amino acids four residues apart and is not capped

6. Alpha helix because it is capped, has favorable interactions for residues four things apart, and has no helix breakers. It is not amphipathic because it does not follow the pattern seen in number 1.

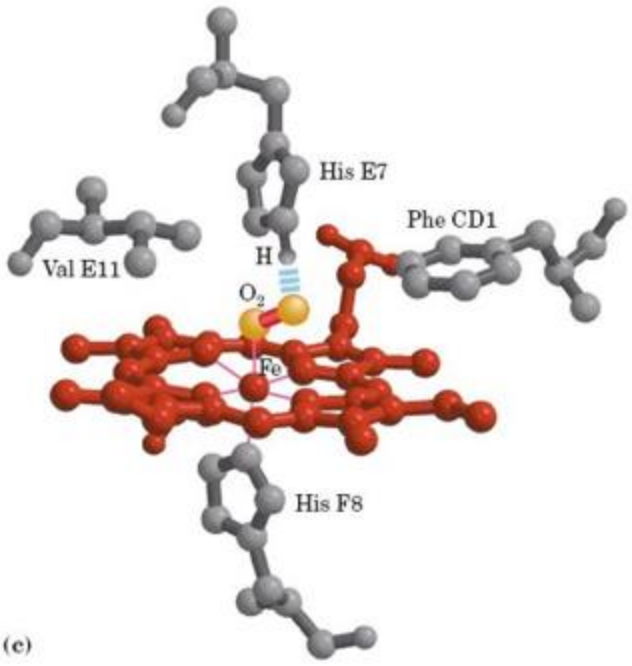
Focus 2: Enzyme Mechanisms, Inhibition, and Saturation

Chapter 5: Hemoglobin + Myoglobin: Structure, Function, and Saturation

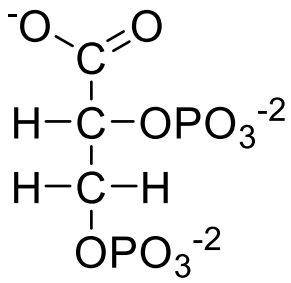
Proteins serve a wide variety of functions within the body, one of them is the transport of oxygen. In humans, the oxygen carrier molecule of the body is hemoglobin. Hemoglobin is a large macromolecular protein complex that has four subunits in total, two alpha and two beta subunits. These subunits resemble myoglobin, which is the oxygen carrier/storage unit in muscle cells. Both of these proteins contain a heme cofactor. Cofactors are molecules that are necessary for protein function that are not covalently bound to the protein. This heme cofactor is shown below and involves an iron center that can get oxidized in the presence of oxygen from Fe^{2+} to Fe^{3+}



The heme cofactor acts as the oxygen binding domain of the protein and is absolutely required for oxygen transport. The heme cofactor in the hemoglobin protein is ligated with His F8 to make the iron have an approximately octahedral coordination upon oxygen binding. Other molecules are prevented from binding the heme cofactor by sterically bulky amino acids such as Phe CD1 and Val E11. These amino acids along with His E7 are responsible for hemoglobin's ability to bind oxygen and prevents it from binding toxins such as CO as effectively as heme by itself. Carbon monoxide, CO, is poisonous because it displaces oxygen in hemoglobin and therefore oxygen cannot be delivered to your brain and muscle, ultimately leading to death.[5]



Hemoglobin's oxygen affinity is modulated by the presence of oxygen and 2,3-bisphosphoglycerate. When there is an abundance of oxygen present in the blood near hemoglobin, hemoglobin goes from the T to the R state. The R state is the conformation of hemoglobin that has a high affinity for oxygen. Think R for "ready". This T→R transition is done by breaking ionic bonds between the interface of the alpha1-beta2 subunit interface. When hemoglobin enters an area in the body where oxygen concentration is low, the T state gets more stabilized and the oxygen bound to heme gets released and delivered to those tissues. In this way, hemoglobin can "sense" which tissues are in the most need of oxygen and can deliver it accordingly. The body can also synthesize 2,3-bisphosphoglycerate, which will stabilize the T state by promoting additional ionic bonds between the negatively charged groups of 2,3-bisphosphoglycerate and the positively charged oxygen-binding cavity in each hemoglobin subunit. This is frequently seen when people are adjusting to higher altitudes, because the oxygen availability is much lower as one goes higher in altitude, the ability of hemoglobin to release oxygen becomes much more important. In other words, whatever oxygen that hemoglobin binds needs to get released to the tissues and the T state becomes more important. The structure of 2,3-bisphosophoglycerate is shown below:



These salt bridges formed in the oxygen-binding domain of hemoglobin are more pronounced with decreasing pH. Lower pH allows proximal His residues near the heme cofactor get protonated and this facilitates additional ionic bonds forming and stabilizing the T state of hemoglobin. This is especially important for delivering oxygen to muscle tissues because these tissues actively undergo lactic ACID fermentation during exercise and therefore the blood near muscles will be more acidic during energy-intensive activities and oxygen will be delivered to these tissues more easily by hemoglobin. This is called the Bohr effect and this is important for improving the efficiency of oxygen delivery to actively respiring tissues and for delivering oxygen from the lungs overall.

The fraction of hemoglobin that is saturated with oxygen can be calculated mathematically and is denoted as θ. The mathematical relationship is shown below:

$$\theta = \frac{p_{O2}}{p_{50} + p_{O2}}$$

In this equation, p_{O2} is the partial pressure of oxygen and p_{50} is the partial pressure of oxygen that saturates 50% of hemoglobin molecules. The lower the p_{50} the stronger the oxygen binding affinity that oxygen carrier molecule in question has. The percent oxygen delivered to the tissues would be the difference in θ in the lungs versus in the tissues. These only change because of the partial pressure of oxygen in the lungs and tissues.

Practice Questions:

1. Calculate the % saturation for oxygen carrier X that has a p_{50} of 60 mmHg if the p_{O2} is 75 mmHg.

Questions 2-4 please refer to the saturation curve data provided

P_{O2} (mmHg)	θ
1.0	0.09
2.0	0.17
5.0	0.33
10.0	0.50
25.0	0.71
164	0.94
250	0.96

2. What is the p_{50} for this oxygen carrier?
3. What is the θ when the p_{O2} is 1250 mmHg?
4. What percent oxygen is delivered to muscle tissue using this oxygen carrier if the p_{O2} in the lungs is 25.0 mmHg and the p_{O2} in the muscle is 2.0 mmHg?
5. Which oxygen carrier has a higher affinity for oxygen based on the data provided below:

P_{O2} (mmHg)	θ Protein A	θ Protein B	θ Protein C
1.0	0.09	0.33	0.01
2.0	0.17	0.50	0.01

5.0	0.33	0.71	0.03
10.0	0.50	0.83	0.06
25.0	0.71	0.93	0.13
164	0.94	0.99	0.50
250	0.96	0.99	0.60

Answers:

1. Simply plug and chug into the equation:

$$\theta = \frac{75mmHg}{60mmHg + 75mmHg} = 0.55$$

2. The p_{50} is the partial pressure of oxygen when $\theta = 0.5$, therefore the p_{50} for this oxygen carrier is 10.0 mmHg.
3. Simply plug and chug now that you know the p_{50} like so:

$$\theta = \frac{p_{O2}}{p_{50} + p_{O2}} = \frac{1250mmHg}{10mmHg + 1250mmHg} = 0.992$$

4. Calculate θ when p_{O2} is 25 mmHg and when it is 2.0 mmHg then subtract the two like so:

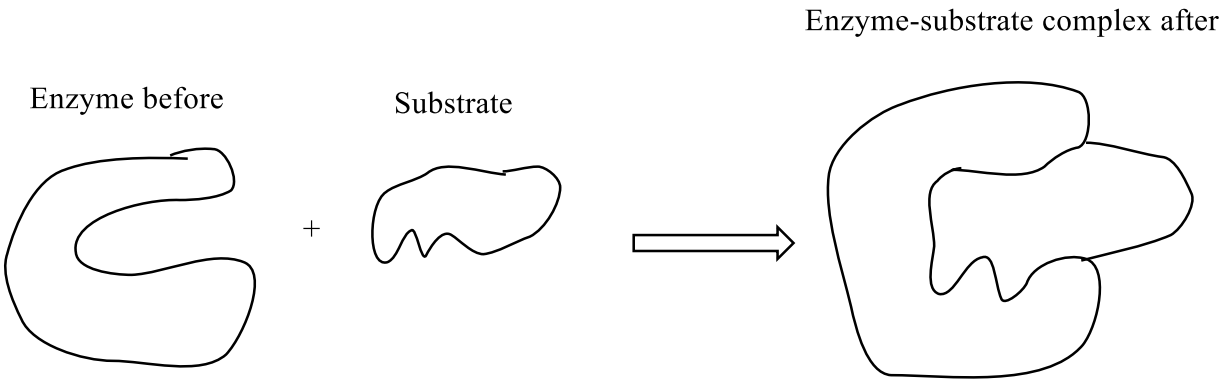
$$0.71 - 0.17 = 0.54$$

   So 54% of oxygen is delivered to the tissues.

5. Look for the p_{50} for each protein, because protein B has the lowest p_{50} (2.0 mmHg) it has the strongest affinity for oxygen.

Chapter 6: Enzyme Catalytic Strategies and Michaelis Menten Kinetics Basics

Enzymes are a class of proteins that act as catalysts and speed up biological reactions. Without enzymes, life as we know it would cease to exist, but how is it that enzymes can catalyze so many types of reactions but only be made of 20 amino acids? What strategies do enzymes employ and what thermodynamically speaking allows for these biological reactions to be spontaneous? These questions will be (partially) answered in this chapter. First, we need to discuss how enzymes bind their substrate, and for that, we will turn to the induced-fit model. Originally, it was thought that each enzyme has a particular substrate and that these two molecules fit together like a key into a lock. We now know that this is not true and that one enzyme can in fact, catalyze many reactions with many different substrates. To account for this observation, the new working theory of enzyme dynamics is that it is an induced-fit model. This means that the enzyme originally has a shape that will not accommodate perfectly any substrate, but upon substrate binding, the enzyme will change its shape to match it better and in this way maximize its binding affinity (MA**S**C).



Now that the substrate is bound to the enzyme, there are many different catalytic strategies that enzymes employ:

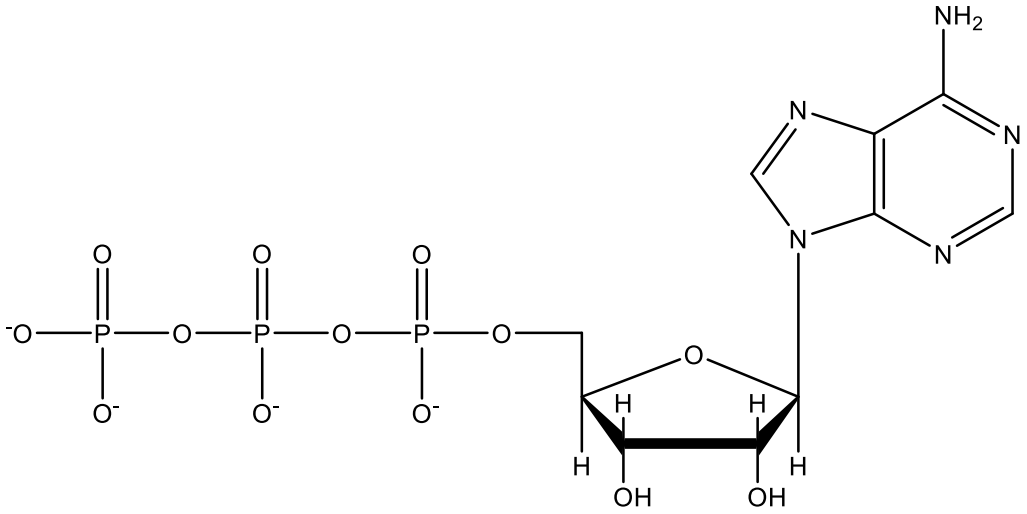
1. Acid/Base Catalysis
2. Covalent Catalysis
3. Metal Ion Catalysis

Acid/Base catalysis often uses acidic or basic residues in the active site to protonate or deprotonate certain functional groups to make the reaction more favorable. Think back to Organic Chemistry when we wanted to hydrolyze an ester or an amide, that required an acid catalyst. In this context, instead of using HCl, we are using Asp and Glu residues to protonate the carbonyl oxygen to facilitate the reaction. The most common residues in this catalytic strategy are Asp, Glu, His, Lys, and Arg, although sometimes serine and tyrosine can be used.

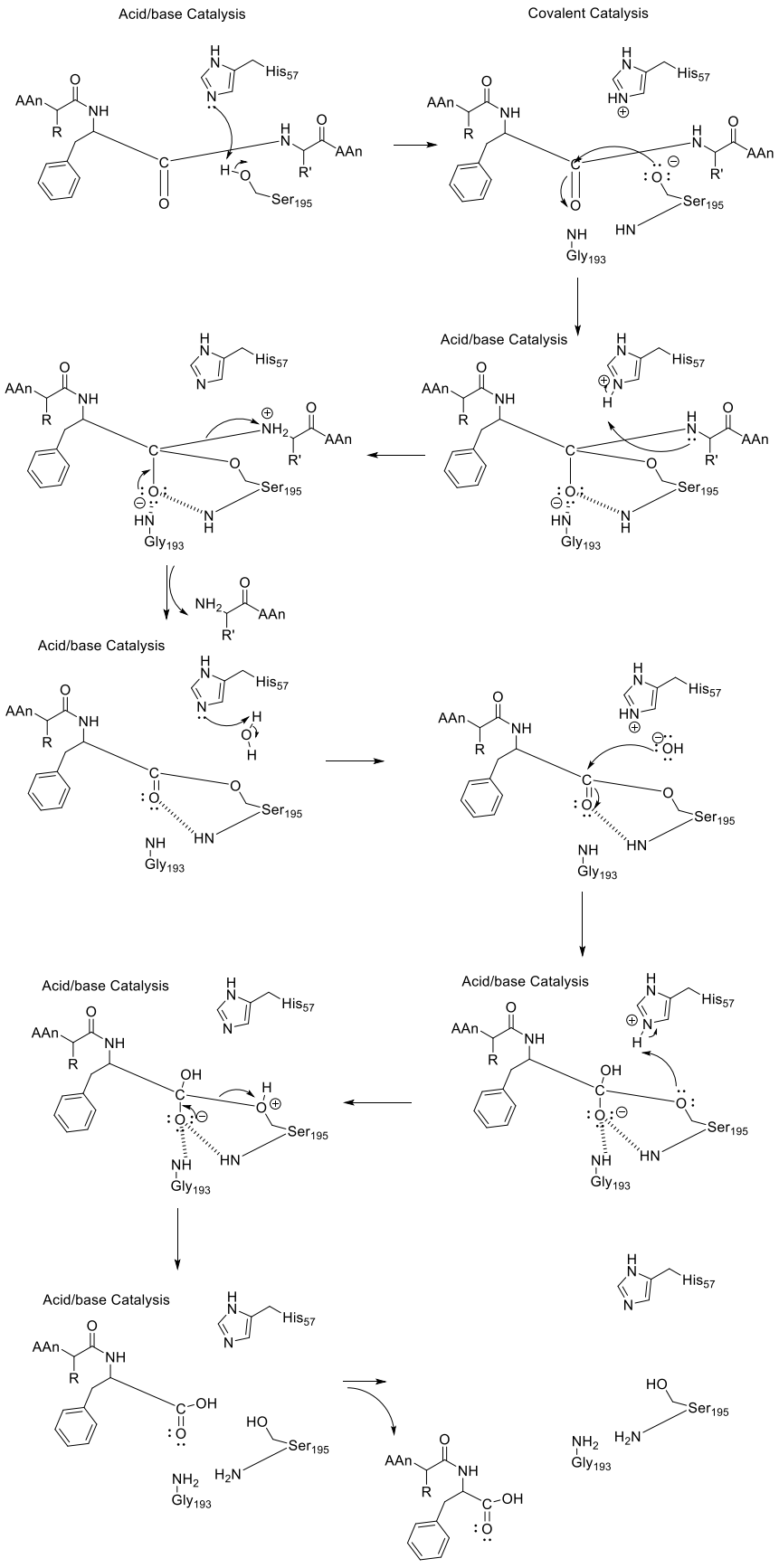
Covalent catalysis typically involves attaching the reactant to the enzyme to make it a more reactive carbonyl compound that can get displaced by a nucleophile. These reactions make use of a nucleophilic acyl substitution reaction mechanisms using serine, cysteine, or any amine/carboxylate groups in the enzyme's active site. Because enzymes are catalysts, they must

be recycled and therefore after the reaction concludes, those residues must return back to their initial state (protonated/deprotonated). This is frequently paired with acid/base catalysis.

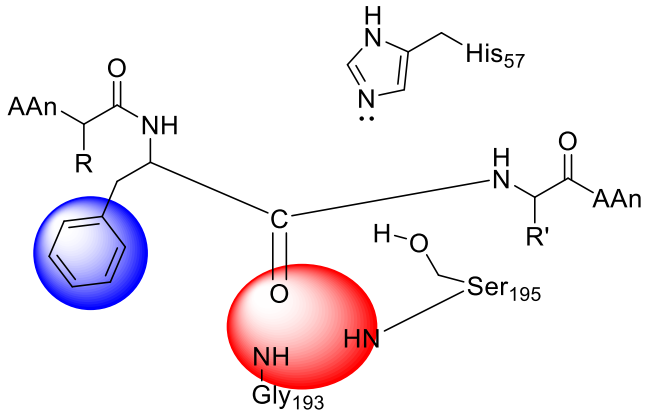
Metal ion catalysis uses metal ions such as Mg^{2+}, Zn^{2+}, Cu^{+}, etc. to stabilize negative charge, facilitate the conformational change in the enzyme that would allow reactive groups to be near the reactant, and activate nearby nucleophilic residues. They can also be redox-active centers and facilitate oxidation and reduction reactions such as the Fe^{2+} in the CYP450 enzymes. Recall that negative charges attract positive charges and therefore positively charged metal ions can be used to stabilize anions (**M**ASC). This is frequently seen in ATP-hydrolyzing enzymes, which typically have an obligate Mg^{2+} cofactor, the structure of ATP is shown below:



To see these catalytic strategies in action, we will look at the enzyme mechanism for chymotrypsin. This enzyme is responsible for breaking peptide bonds next to aromatic amino acids in a polypeptide chain (protein) when protein is taken in from the diet of the organism. The mechanism for chymotrypsin is shown below with the portions that correspond to acid/base catalysis and covalent catalysis indicated:


Acid/base Catalysis
Covalent Catalysis
Acid/base Catalysis
Acid/base Catalysis
Acid/base Catalysis
Acid/base Catalysis
Acid/base Catalysis
His57
Ser195
Gly193
AAn
R
R'

The way that this enzyme recognizes the aromatic amino acids in the peptide chain is through its hydrophobic pocket, it stabilizes the anionic oxygen using the oxyanion hole that has ion-dipole interactions with the N-H bonds of Ser^{195} and Gly^{193} (**M**A**S**C). The overall structure of the enzyme's active site is shown below:



The blue region is the hydrophobic pocket and the red region is the oxyanion hole that allows chymotrypsin to recognize the aromatic amino acids and facilitate peptide bond cleavage by pairing acid/base catalysis with covalent catalysis.

Enzymes bind their substrate tightly, but not too tightly. If enzymes were to bind their substrates too tightly then they would not let go of them and the products would not get released. There has been a lot of work done to characterize enzyme kinetics, or the rate of enzymatic reactions. In this text, we will only address Michaelis Menten kinetics, which is the basic model of enzyme kinetics that many people use even to this day. When studying enzyme kinetics, one can either measure the rate of appearance of product or disappearance of reactant, in either case, a rate is determined initially and this rate is referred to as V. But enzymes are not infinitely fast, in other words, there is a maximum rate that they can achieve that is limited by the chemistry and the kinetics of moving the enzyme to fit the substrate (induced-fit), this is referred to as V_{max}. The final property that determines the kinetics of the enzyme is the affinity of that enzyme for that particular substrate, this is measured as the K_m, which is the concentration of substrate that is required to reach half of the maximum rate. These terms come together into the Michaelis Menten equation, which predicts the rate of an enzyme given the concentration of substrate (S), maximum rate (V_{max}), and affinity for substrate (K_m):

$$V = \frac{(V_{max})(S)}{K_m + S}$$

Does this equation remind you of anything? If you said the oxygen saturation equation we learned in the previous chapter, you would be correct, look at the striking similarities between the two equations:

$$\theta = \frac{p_{O2}}{p_{50} + p_{O2}} \; vs \; V = \frac{(V_{max})(S)}{K_m + S}$$

This should come as no surprise, both of the subjects of these equations are saturable proteins that serve a particular function. The only difference is that one refers to the amount of saturated protein and the other refers to the rate at which the proteins are getting saturated. Effectively p_{O2} and S are analogous and K_m and p_{50} are analogous. Therefore, enzymes that have a lower K_m will have a stronger affinity for their substrate and the same kinds of questions can be asked of you between the two systems. In addition to K_m, V_{max} and S, there are other kinetic parameters for enzymes, these are E_{tot} and k_{cat}. E_{tot} is simply the total concentration of enzyme in the reaction solution and k_{cat} is the turnover number, the maximum number of reactions that one enzyme can do per second in excess substrate concentration, the product of these two quantities will therefore give the maximum velocity as follow:

$$V_{max} = E_{tot}(k_{cat})$$

The quotient of K_m and k_{cat} is often times used to measure catalytic efficiency as follows:

$$catalytic\ efficiency = \frac{k_{cat}}{K_m}$$

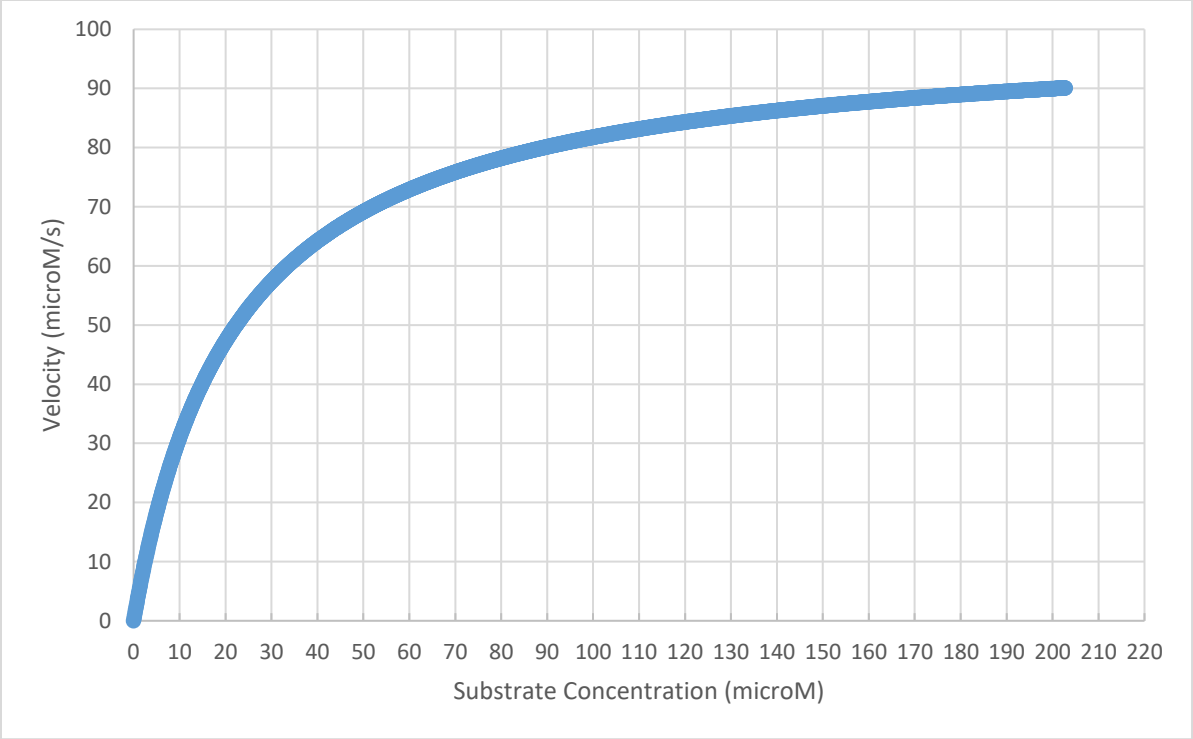
For this number to be accurate, the concentration of enzyme needs to be known very precisely because k_{cat} is dependent on the total enzyme concentration. Because of this very strong dependency, oftentimes k_{cat} and therefore catalytic efficiency is difficult to reproduce exactly.

Practice Questions:

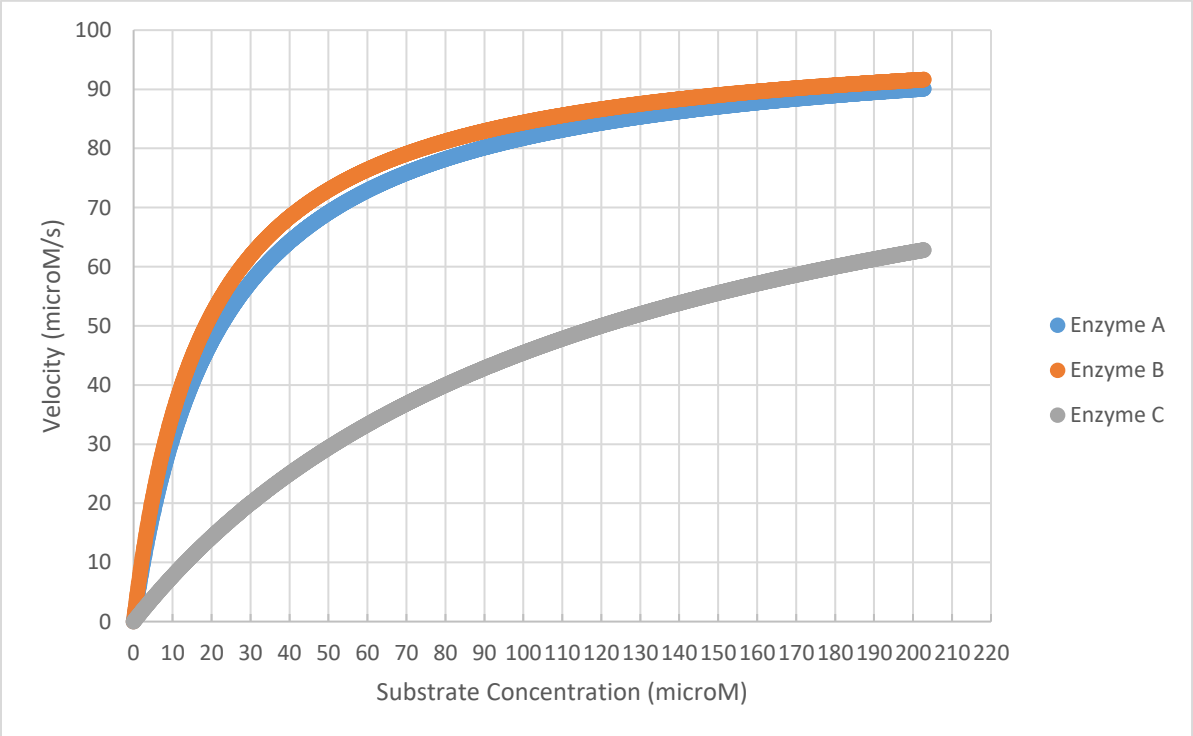
1. What is the K_m for an enzyme that has a maximum rate of 10 mmol/sec and a velocity of 5 mmol/sec at a concentration of 10 nM substrate? Show the easy way and the hard way of calculating this.
2. Which enzyme has the strongest affinity for their substrate according to the kinetic data shown below:

Substrate (mM)	Enzyme A Velocity (mM/s)	Enzyme B Velocity (mM/s)	Enzyme C Velocity (mM/s)
0.1	0.46	1.8	6.230
0.5	2.4	6.7	10.38
1	4.5	10.	11.33
5	17	17	12.22
10	25	18	12.34
25	36	19	12.41
100	50	20.	12.46

3. Based on the Michaelis Menten plot shown below, calculate the K_m, V_{max} and k_{cat} assuming that the concentration of enzyme is 10 microM.



4. Which enzyme has the highest affinity for its substrate according to the Michaelis Menten plot below (assume same maximum velocity):



5. Which enzyme above has the highest catalytic efficiency (assume same enzyme concentration)? Show your work.

6. Suppose you collect a sample of cells from a river and you want to evaluate the enzymatic activity of a particular from the crude cell lysates. Which kinetic parameter(s) can you not determine?

Answers:

1. If the maximum velocity is 10 mmol/s and the velocity recorded is 5 mmol/s then that means we are half the maximum velocity, therefore the substrate concentration and the K_m are one in the same. K_m = 10 nM. The harder way is doing the algebra to show this

$$V = \frac{V_{max}(S)}{K_m + S}$$

We need to solve for K_m here:

$$V(K_m + S) = V_{max}(S)$$

$$V(K_m) + V(S) = V_{max}(S)$$

$$V(K_M) = V_{max}(S) - V(S)$$

$$K_m = \frac{V_{max}(S) - V(S)}{V}$$

$$K_m = \frac{10mmols^{-1}(10nM) - 5mmols^{-1}(10nM)}{5mmols^{-1}} = \frac{100mmols^{-1}nM - 50mmols^{-1}nM}{5mmols^{-1}}$$

$$K_m = \frac{50mmols^{-1}nM}{5mmols^{-1}} = 10nM$$

2. The enzyme that has the strongest affinity for the substrate is the one that has the lowest K_m and therefore we must compare the K_m for all the proteins. Protein A has a K_m of 10 mM, protein B has a K_m of 1 mM, and C has a K_m of 0.1 mM. Therefore enzyme C has the highest affinity.
3. The plot shown appears to level off at 100 microM/s therefore that is the V_{max}. Since 100 is the V_{max}, 50 microM/s would be one half that. The concentration of substrate that would give 50 microM/s as the velocity is 10 microM, therefore that is the K_m of the enzyme. To calculate the k_{cat}, we simply need to divide the max velocity by the total enzyme concentration like so:

$$k_{cat} = \frac{50microMs^{-1}}{10microM} = 5\ s^{-1}$$

So the k_{cat} is 5 s^{-1}.

4. Since we can clearly see that the blue and orange plateau at 100 microM/s, that is the maximum velocity for all of the enzymes of interest. Therefore, we need to determine the concentration of substrate that gets each enzyme to 50 microM/s to get the K_m for each enzyme. This corresponds to approximately 10 microM for enzyme A, 5 microM for

enzyme B, and 50 microM for enzyme C. Therefore enzyme B has the highest affinity for its substrate.

5. Since catalytic efficiency is dependent on k_{cat} and K_m,

$$catalytic\ efficiency = \frac{k_{cat}}{K_m}$$

we need to calculate or get a gist for the numerical values for each of those. K_m we just calculated in the previous question, so we know those. K_{cat} we have not calculated yet, but we can calculate it using the formula:

$$V_{max} = k_{cat}(E_{tot})$$
$$k_{cat} = \frac{V_{max}}{E_{tot}}$$

But in this question and the question before, we were told to assume that they had the same total enzyme concentration and the maximum velocity, therefore all of these enzymes have the exact same k_{cat}. Therefore, the catalytic efficiency is only driven by K_m, in which case enzyme B has the highest catalytic efficiency.

6. If you are working with crude lysates, you cannot know with high accuracy the concentration of enzyme that you are working with, therefore k_{cat} and catalytic efficiency cannot be calculated. Everything else can be calculated (K_m and V_{max}).

As you probably saw in the previous practice problems, sometimes it is hard to see exactly what the maximum velocity is based on the Michaelis Menten plots. Likewise, it is also hard to determine the K_m. It is effectively just a guessing game and educated estimations. To combat this, Lineweaver Burke plots were created. These plots allow for the accurate determination of K_m and V_{max} using the x-intercept and y-intercept respectively. The Lineweaver Burke equation is shown below:

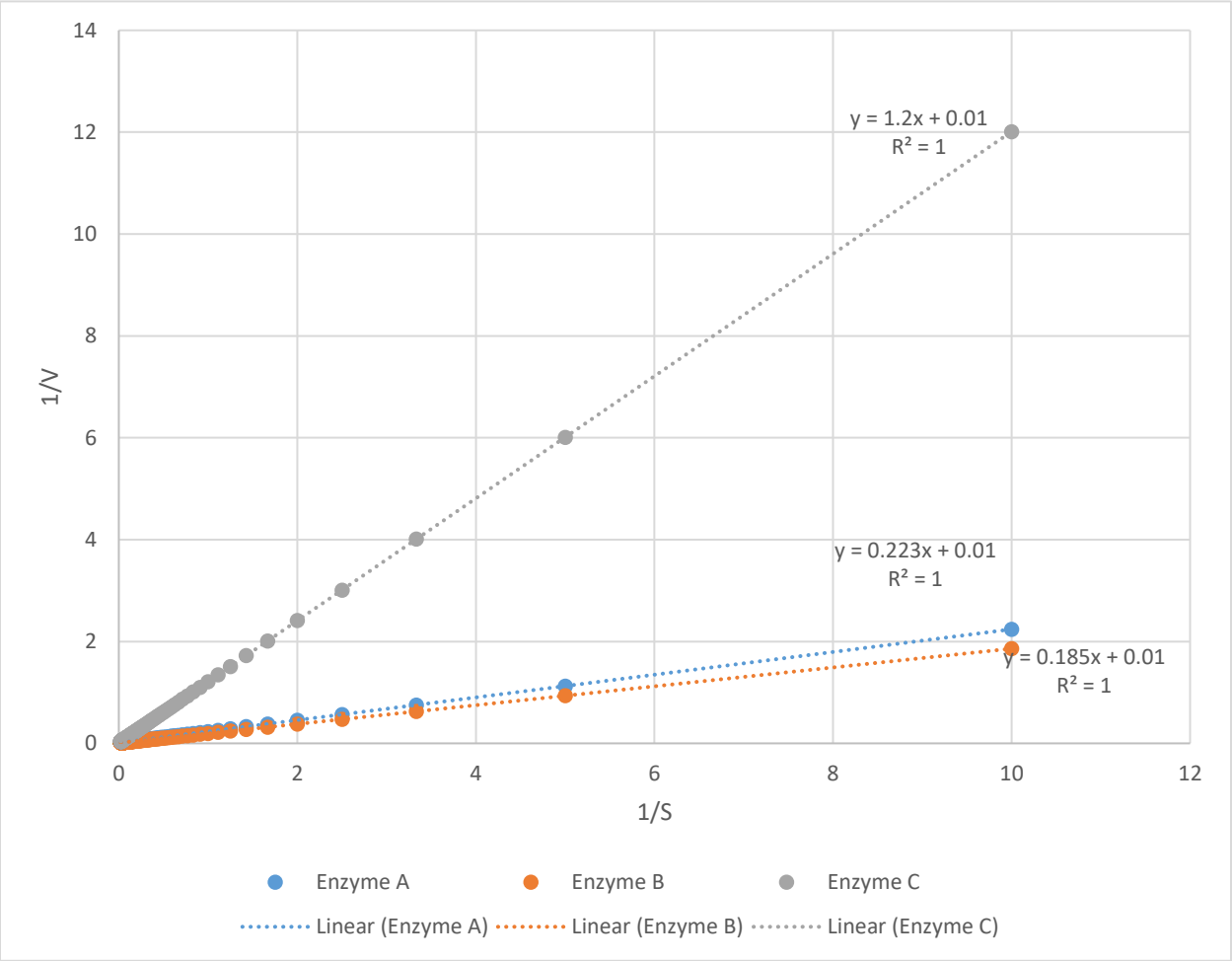
$$\frac{1}{V} = \frac{K_m}{V_{max}}\left(\frac{1}{S}\right) + \frac{1}{V_{max}}$$

This gives a linear plot with 1/V on the y and 1/S on the x. For Lineweaver Burke plots:

$$K_m = -\frac{1}{x_{intercept}}$$

$$V_{max} = \frac{1}{y_{intercept}}$$

Examples of Lineweaver Burke plots are shown below for the enzymes that we discussed in Practice Question 4.



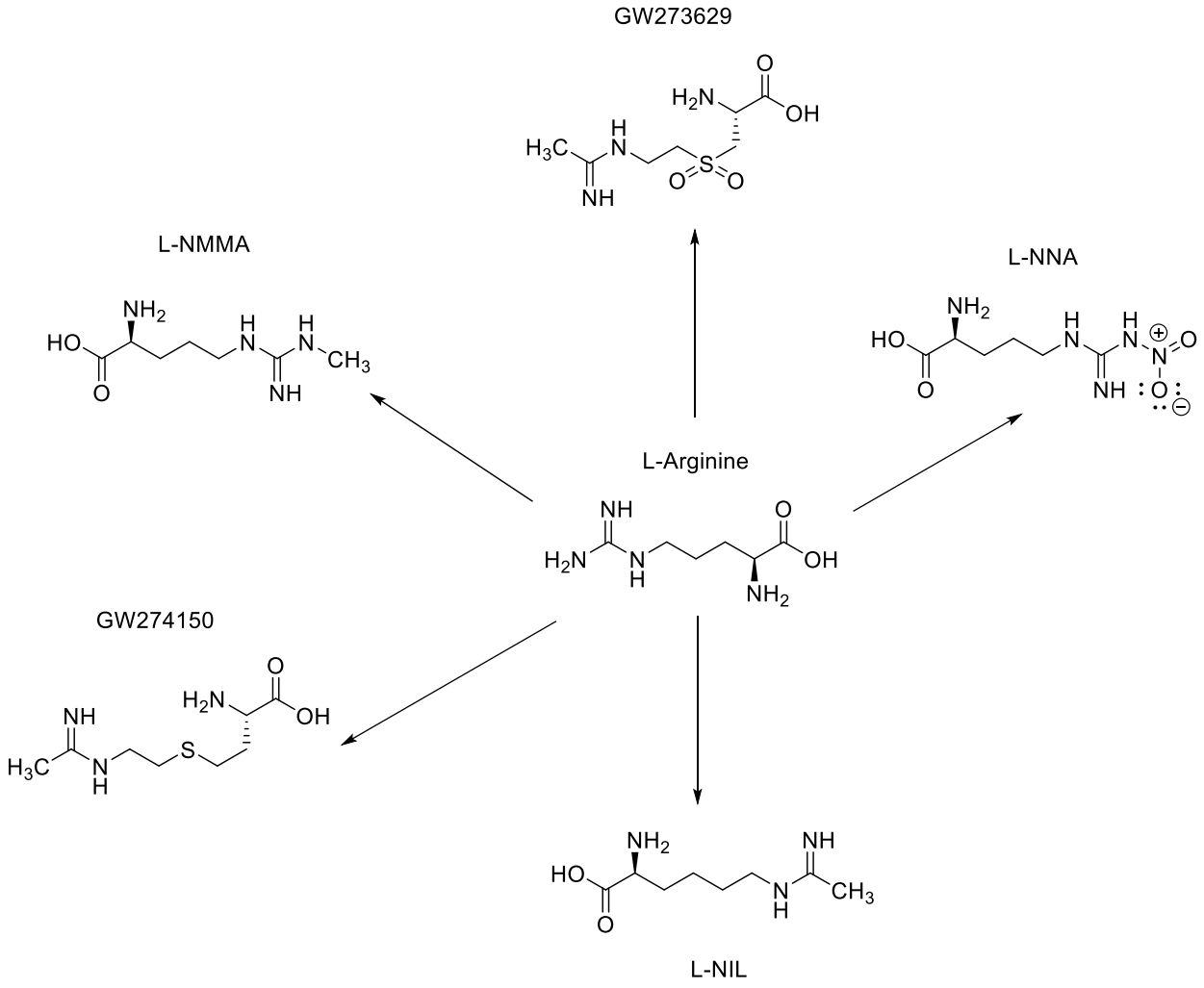
As you can see, they all have the same y intercept, which is why I said to assume they all have the same maximum velocity (because they do). This was incredibly difficult to see for enzyme C because it has a much higher K_m, which is seen by the much higher slope that enzyme C has compared to the other two trend lines and the much smaller x-intercept. The benefit of the Lineweaver Burke plot is that it allows for a quantitative rather than semi-quantitative analysis of the enzyme in question, it takes the guesswork out of determining the K_m and V_{max}. However, the Lineweaver Burke plots are not without their issues, they suffer from the fact that by taking the reciprocal, errors in experimental protocol are increased substantially. In other words, experimental error is amplified in reciprocal space and so the K_m and V_{max} values are oftentimes not as accurate as you would hope. This can sometimes be combatted by fitting the regular Michaelis Menten data to the equation using regression analysis and extracting the necessary information from the fit. But since you don't have access to Excel during the exam, most professors will ask you to do Lineweaver Burke analysis that they provide because without computers this is the best way to quantify these parameters.

Chapter 7: Competitive Inhibition

Now that we covered what ordinarily happens, we can start introducing things to mess certain aspects of the model up. The simplest way to do this is to discuss inhibitors. Inhibitors come in four different flavors:

1. Competitive inhibitors
2. Uncompetitive inhibitors
3. Mixed inhibitors
4. Allosteric inhibitors

Competitive inhibitors will be covered in this chapter with the other types of inhibitors discussed in future chapters. Competitive inhibitors, as the name suggests are molecules that will compete with the natural substrate for enzyme binding. Typically these molecules resemble the substrate that they are competing against and they are designed to have additional IMFs with the active site of the enzyme. Examples of competitive inhibitors for the Nitric oxide synthase enzymes, which takes L-Arginine as its substrate are shown below:



These competitive inhibitors bind to the same active site as L-Arginine, but they form stronger interactions with the enzyme that prevents the enzyme from performing chemistry on the natural substrate. These inhibitors can be overcome by increasing the amount of substrate because the more substrate that is available to the enzyme, the more likely by chance for it to be closer to the substrate than the inhibitor. Because competitive inhibitors can be overcome by increasing substrate concentration, the V_{max} parameter for the enzyme in question DOES NOT CHANGE. However, because the competitive inhibitor forms stronger attractive interactions with the enzyme, perceptually the enzyme does not like the substrate as much and therefore K_m will increase. The extent that K_m increases is called α and the apparent K_m, otherwise known as K_{mapp} can be calculated using the following equation:

$$K_{m_{app}} = \alpha K_m$$

Therefore for competitive inhibitors:

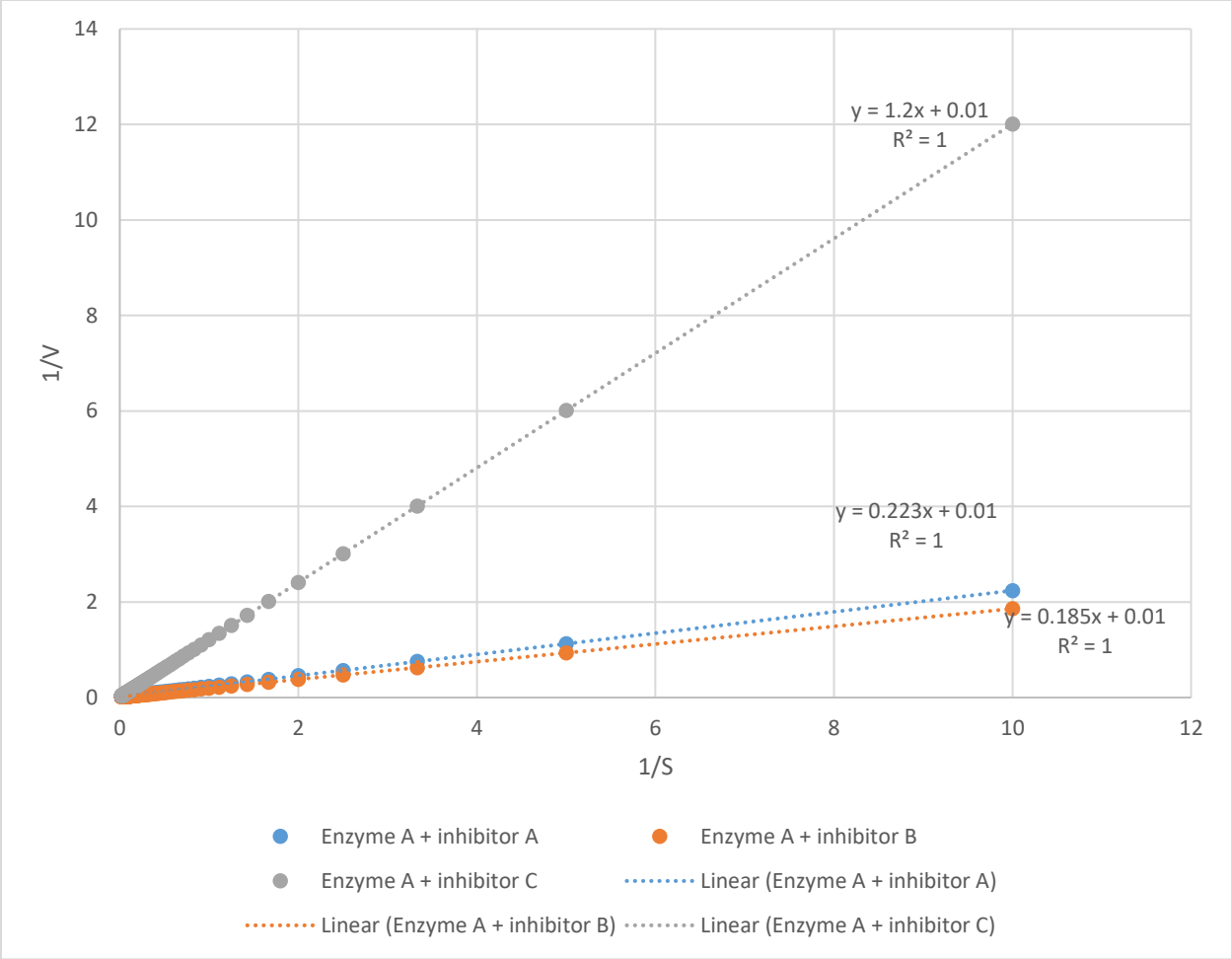
1. V_{max} does not change
2. K_m increases by α

Practice Questions:

1. Determine which inhibitor is the most potent based on the following data:

Substrate (mM)	Velocity Enzyme A (mM/s)	Velocity Enzyme A + Inhibitor A (mM/s)	Velocity Enzyme A + Inhibitor B (mM/s)
0.1	9.09	1.96	0.249
0.5	33.3	9.10	1.23
1	50	16.7	2.44
5	83.3	50	11.1
40	97.6	88.9	50
160	99.4	97.0	80
295	100	100	100

2. Determine which inhibitor is the most potent based on the following data:



Answers:

1. We have to determine what the alpha value is and/or what the K_{mapp} is for each specific scenario. Because inhibitor B has the highest alpha value and highest K_{mapp} it is the strongest inhibitor.
2. We have to determine the K_{mapp} for each scenario by calculating the x-intercept, when we do this, we find that inhibitor C causes the strongest inhibition.

Chapter 8: Uncompetitive Inhibition

Before we discussed competitive inhibitors and how they effect V_{max} and K_m, now we will discuss uncompetitive inhibitors. Uncompetitive inhibitors are those inhibitors that will bind to the enzyme-substrate complex, which is when the enzyme first binds the substrate but no chemistry is happening yet. When these inhibitors bind the enzyme-substrate complex, the inhibitor causes a conformational change in the enzyme that prevents substrate release. Because these inhibitors do not compete for binding the enzyme active site, they CANNOT be overcome with increasing substrate concentration like competitive inhibitors can. This makes them unique and more difficult to deal with, therefore, many drugs developed against a specific molecular target are generally not uncompetitive inhibitors. Indeed, uncompetitive inhibitors could be poisons if they covalently attach to the enzyme-substrate complex because then they cannot be removed and the enzyme is permanently deactivated. Because of their very strong effects on the enzyme's apparent ability to catalyze its reaction, uncompetitive inhibitors will decrease both the K_m and V_{max}. The decrease in V_{max} makes sense because the inhibition cannot be overcome by increasing substrate concentration, therefore the V_{max} must decrease. The decrease in K_m is a byproduct of the fact that it is easier to saturate the enzymes because once the substrate is bound, it is stuck there due to the uncompetitive inhibitor. Recall that K_m is the concentration of substrate that will give half the maximum velocity and is a proxy for binding affinity. If the enzyme binds the substrate with high affinity, it will not let go as easily, which is exactly what the uncompetitive inhibitor does. The extent that the uncompetitive inhibitor effects K_m and V_{max} is referred to as α' and the new values for these kinetic parameters are given by the following equations:

$$K_{m_{app}} = \frac{K_m}{\alpha'}$$

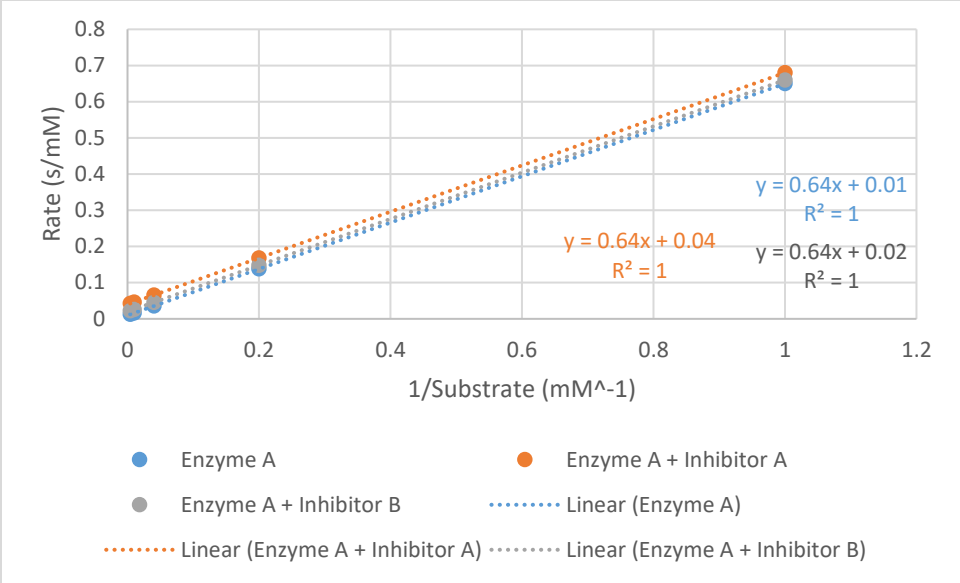
$$V_{max_{app}} = \frac{V_{max}}{\alpha'}$$

Practice questions:

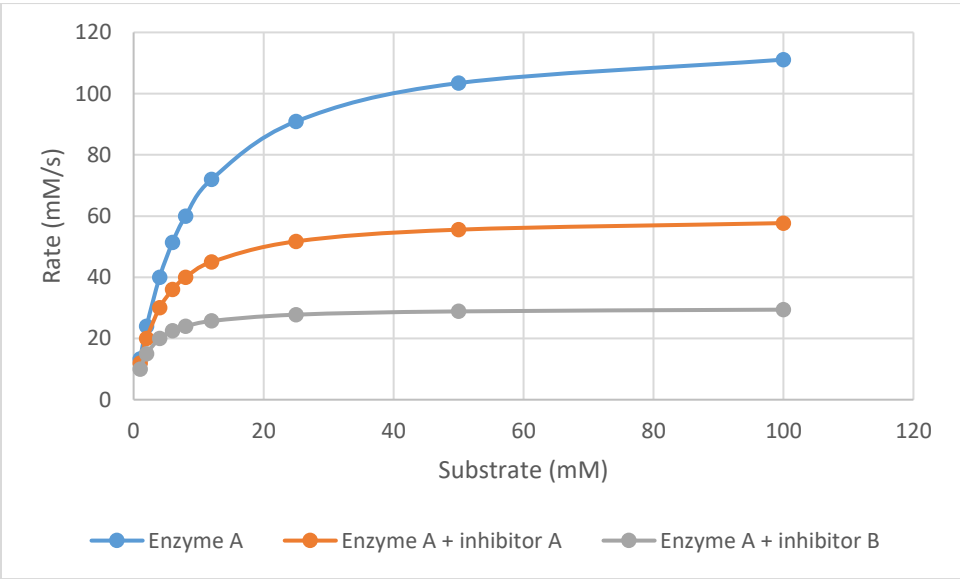
1. Determine which inhibitor is the most potent:

Substrate (mM)	Velocity of Enzyme A (mM/s)	Velocity of Enzyme A + inhibitor A (mM/s)	Velocity of Enzyme A + inhibitor B (mM/s)
1	40	33	25
2	67	50	33
4	100	67	40
25	172	93	48
225	197	99	49.8
572	200	100	50

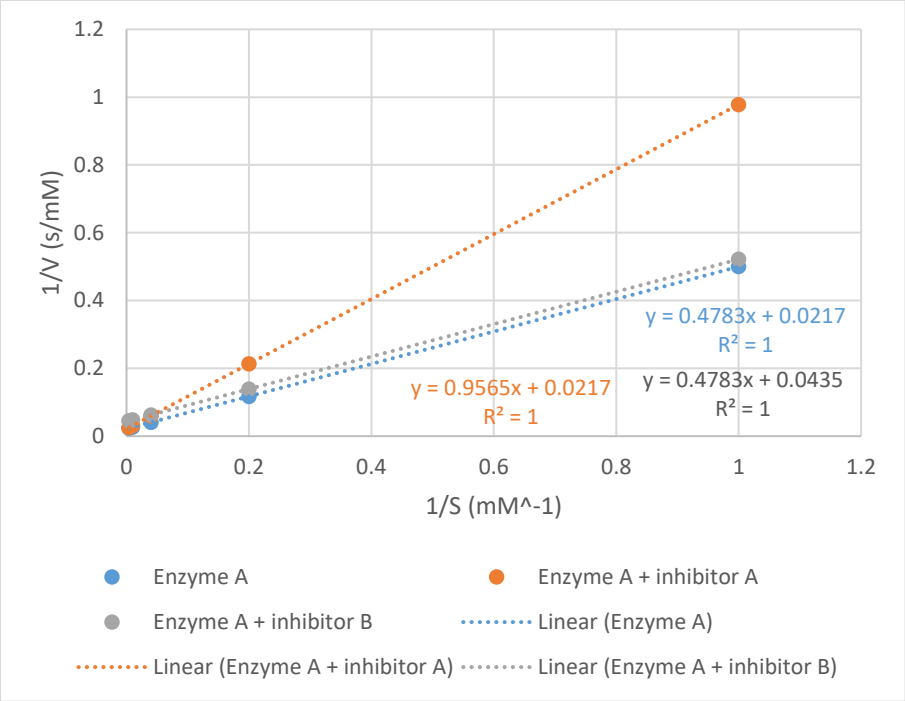
2. Based on the Lineweaver Burke Plots shown below, which inhibitor is the most potent:



3. Based on the Michaelis-Menten Plots shown below, which inhibitor is the most potent:



4. Based on the Linewaver Burke Plots shown below, which inhibitor is an uncompetitive inhibitor? What kind of inhibitor is the other?



Answers:

1. We have to determine the K_m and the V_{max} for each situation. The enzyme by itself normally has a K_m and V_{max} of 4mM and 200 mM/s respectively. The enzyme with inhibitor A has a K_m and V_{max} of 2mM and 100 mM/s respectively. The enzyme with inhibitor B has a K_m and V_{max} of 1mM and 50 mM/s respectively. Therefore, we can conclude that inhibitor B has a higher α' and therefore would be more potent.
2. Just as before, we have to determine the K_m and V_{max} for the original enzyme and then compare it to the K_{mapp} and V_{maxapp} for each inhibitor. The original enzyme has a K_m of 64 mM and a V_{max} of 100 mM/s by using the x-intercept and the y-intercept respectively in the Lineweaver Burke plot. Inhibitor A brings down these values to 16 mM and 25 mM/s and inhibitor B brings down these values to 32 mM and 50 mM/s. Therefore, inhibitor A is the more potent of the two inhibitors since it decreases K_m and V_{max} the most.
3. This one can be solved just visually, since inhibitor B causes the maximum velocity to decrease the most (inhibitor B curve plateaus at a smaller value), that inhibitor is the more potent of the two.
4. For an inhibitor to be uncompetitive, it must decrease BOTH K_m and V_{max}, therefore, we must determine the K_m and V_{max} of the original enzyme and the enzyme in the presence of the two inhibitors. Visually, the line of best fit shows that the y-intercept does not change between the original enzyme and the inhibitor A curve, therefore inhibitor A cannot be an uncompetitive inhibitor. The inhibitor B curve shows that the y-intercept does change therefore this could be an uncompetitive inhibitor, but we can determine the K_{mapp} to confirm this. The enzyme originally has a K_m of 22 mM and a

V_{max} of 46 mM/s while the inhibitor B curve shows a K_m of 11 mM and a V_{max} of 23 mM/s. This is consistent with an α' value of 2. Inhibitor B is therefore an uncompetitive inhibitor. The inhibitor A curve shows a K_m of 44 mM and a V_{max} of 46 mM/s, therefore it is a competitive inhibitor. Competitive inhibitors *always* look like parallel lines in the Lineweaver Burke plot.

Chapter 9: Mixed Inhibitors and Allosteric Inhibitors

The two large categories of inhibitors were described in the previous two chapters, this chapter will introduce the hybrid of the two categories (mixed inhibitors) and an entirely separate category of inhibitors (allosteric).

Mixed inhibitors are inhibitors that can work as both uncompetitive and competitive inhibitors and as such they will have both α and α' in their K_m and V_{max} equations:

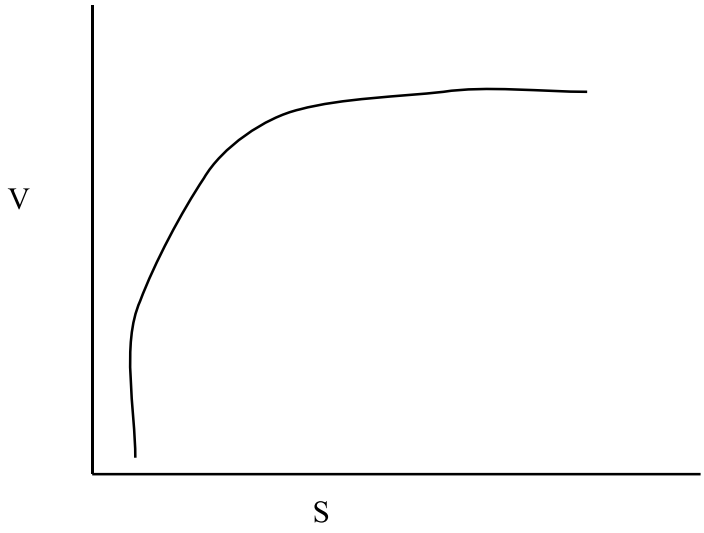
$$K_{m_{app}} = \frac{\alpha K_m}{\alpha'}$$

$$V_{\max_{app}} = \frac{V_{max}}{\alpha'}$$

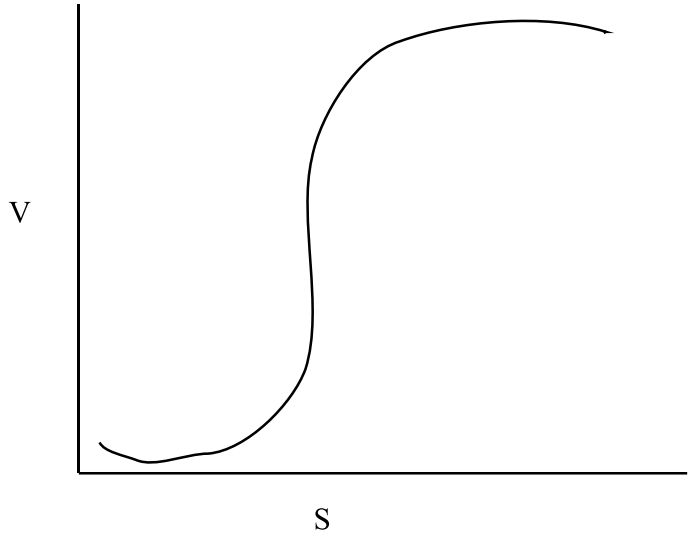
There is a special category of mixed inhibitor, which is where it can act equally as a competitive inhibitor and an uncompetitive inhibitor. These inhibitors are referred to as noncompetitive inhibitors and they are very rare in practice. Unlike competitive inhibitors, mixed inhibitors do not bind the active site directly, but instead bind an alternative site, known as the allosteric site on the enzyme of interest, however, their effect on the enzyme's activity is the same.

Allosteric inhibitors are inhibitors that solely bind the allosteric site of the enzyme, which is a site that is entirely separate from the active site and acts solely as a regulatory element. Proteins that are allosterically regulated typically display non Michaelis Menten kinetics and instead display a sort of sigmoidal kinetic curve like so:

Simple Michaelis Menten Kinetics



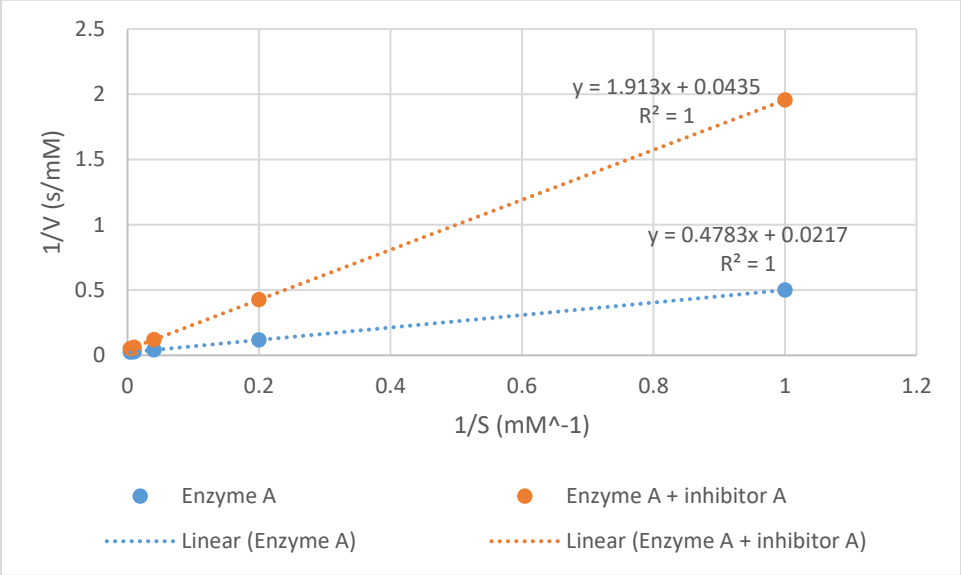
Sigmoidal Kinetics (Allosteric Enzymes)



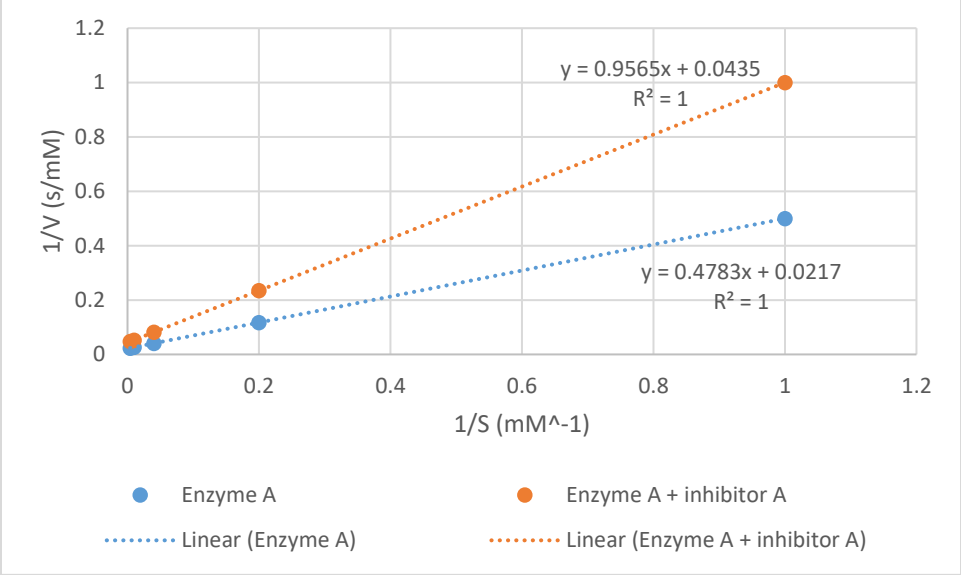
Notice how the sigmoidal curve looks like an S, S for sigmoidal. This is a telltale sign that the enzyme in question is an allosterically regulated enzyme or protein. Allosterically enzymes behave this way because they generally have two different states, the R and T states and these two states have distinct conformations (structures) that have either higher or lower binding affinity for the substrate. This overall behavior is referred to as cooperativity and is a behavior that hemoglobin shares. Many enzymes in biological systems are allosterically regulated and we will discuss these specifically once we delve into biological processes/metabolism. One of the most common mechanisms of allosteric inhibition is feedback inhibition, where the product acts as an allosteric inhibitor of the enzyme that produced it. In this way, the living organism can prevent over production of a particular product (MAS**C**).

Practice Questions:

1. Determine α and α' based on the following kinetic data:



2. Determine the type of enzyme inhibition displayed by inhibitor A based on the kinetic data shown below:



Answers:

1. First we have to determine the K_m and V_{max} for the enzyme by itself by calculating the x-intercept and y-intercept in the Lineweaver Burke plot. Doing this gives that the K_m for the enzyme is 22 mM and the V_{max} is 46 mM/s. Doing the same analysis for the kinetics data in the presence of the inhibitor shows that the apparent K_m is now 44 and the V_{max} is 23. This implies that the α' is 2 because the V_{max} decreased by a factor of 2. Now knowing the α' = 2, the α value can be solved using the mixed inhibitor equation:

$$K_{m_{app}} = \frac{\alpha K_m}{\alpha'}$$
$$K_{m_{app}} * \alpha' = \alpha K_m$$
$$\alpha = \frac{K_{m_{app}} \alpha'}{K_m} = \frac{44(2)}{22} = 4$$

2. First we need to know the V_{max} and K_m for the enzyme by itself and then we have to calculate it for the enzyme in the presence of the inhibitor. Doing this we find that the K_m and V_{max} for the enzyme by itself is 22 mM and 46 mM/s respectively and only the V_{max} changes to 23. Because only the V_{max} changes by a factor of 2, the inhibition must be noncompetitive.

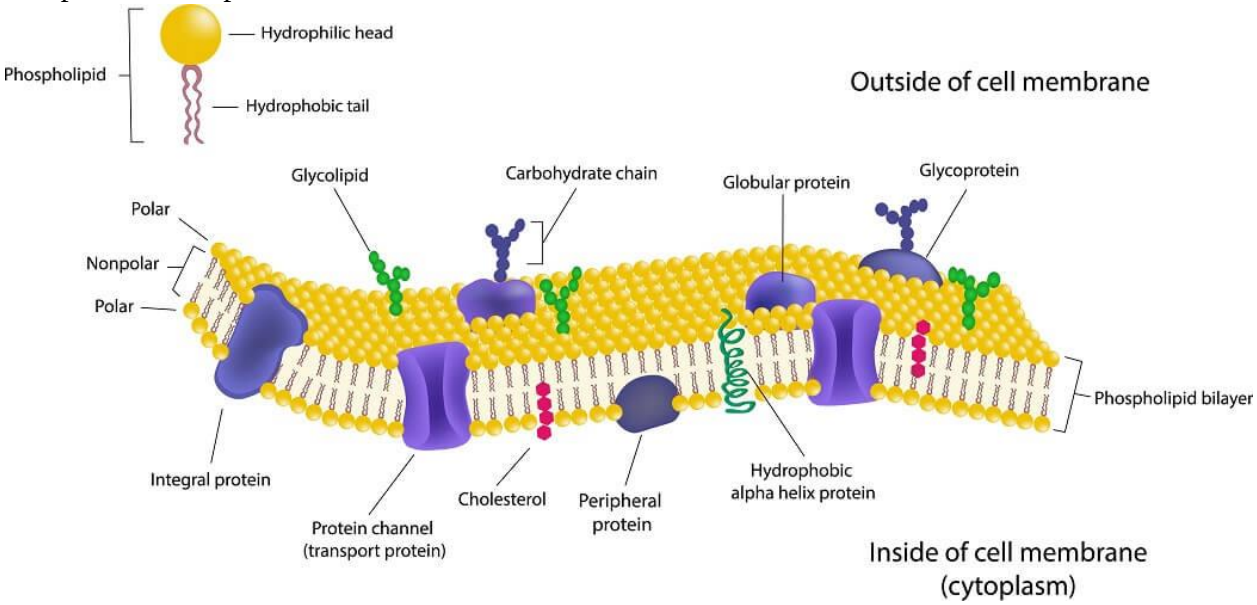
Summary of the different inhibitor types:

Parameter	Competitive	Uncompetitive	Mixed	Allosteric
K_m	$K_{m_{app}} = \alpha K_m$	$K_{m_{app}} = \frac{K_m}{\alpha'}$	$K_{m_{app}} = \frac{\alpha K_m}{\alpha'}$	If in T state higher K_m, if in R state lower K_m
V_{max}	$V_{max_{app}} = V_{max}$	$V_{max_{app}} = \frac{V_{max}}{\alpha'}$	$V_{max_{app}} = \frac{V_{max}}{\alpha'}$	If in T state lower V_{max}, if in R state higher V_{max}
Sigmoidal curve?	No	No	No	Yes

Focus 3: Membrane Structure and Function

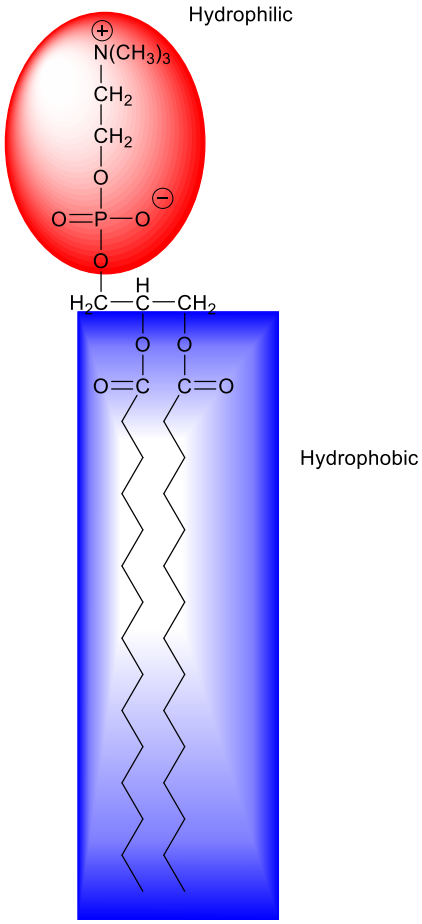
Chapter 10: Cell Membrane Basics

Broadly speaking the cell membrane is comprised of phospholipids, proteins, glycoproteins, glycolipids etc. that allow the cell to select what comes in and out of it and can be used to identify certain types of cells (glycoproteins and glycolipids are super important in this regard). A representative picture of a cell membrane is shown below[6]

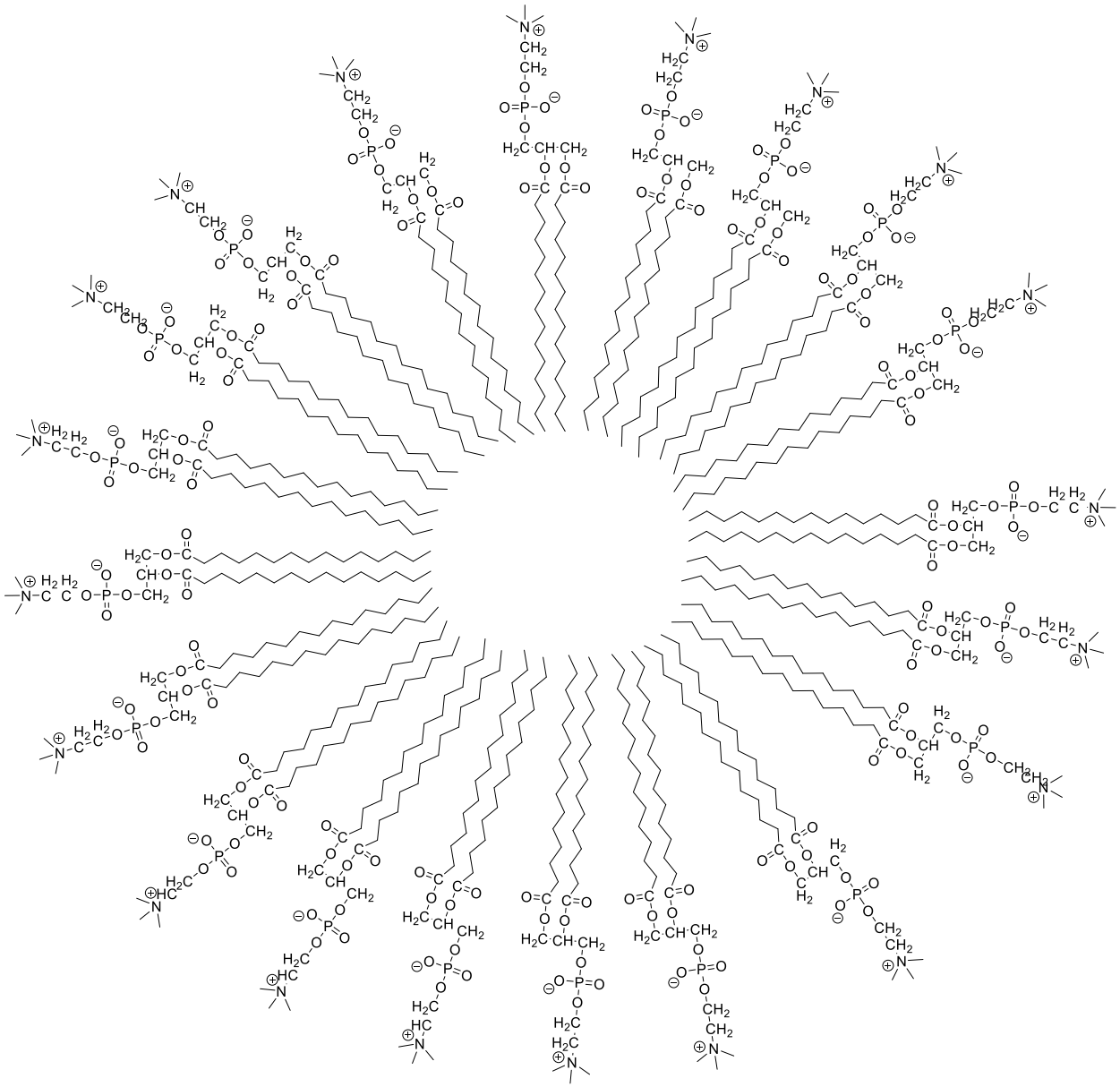


Because of the very important role cell membranes have governing biology; we will take an in-depth look at how membranes form in the first place, how the composition of that membrane determines its function and properties, and the means by which cells use proteins to transport important biomolecules to and from the cytoplasm to the extracellular space.

First, we need to look at what composes most mammalian cell membranes, phospholipids, and how these molecules are able to self-assemble into these super useful membranes. Phospholipids, like the name suggests, are comprised of a phosphate group covalently attached to lipids. All phospholipids are comprised of a phosphate head group, glycerol backbone, and attached carbon chains. The most common of these phospholipids in bacteria and mammalian cells is phosphatidylcholine shown below:



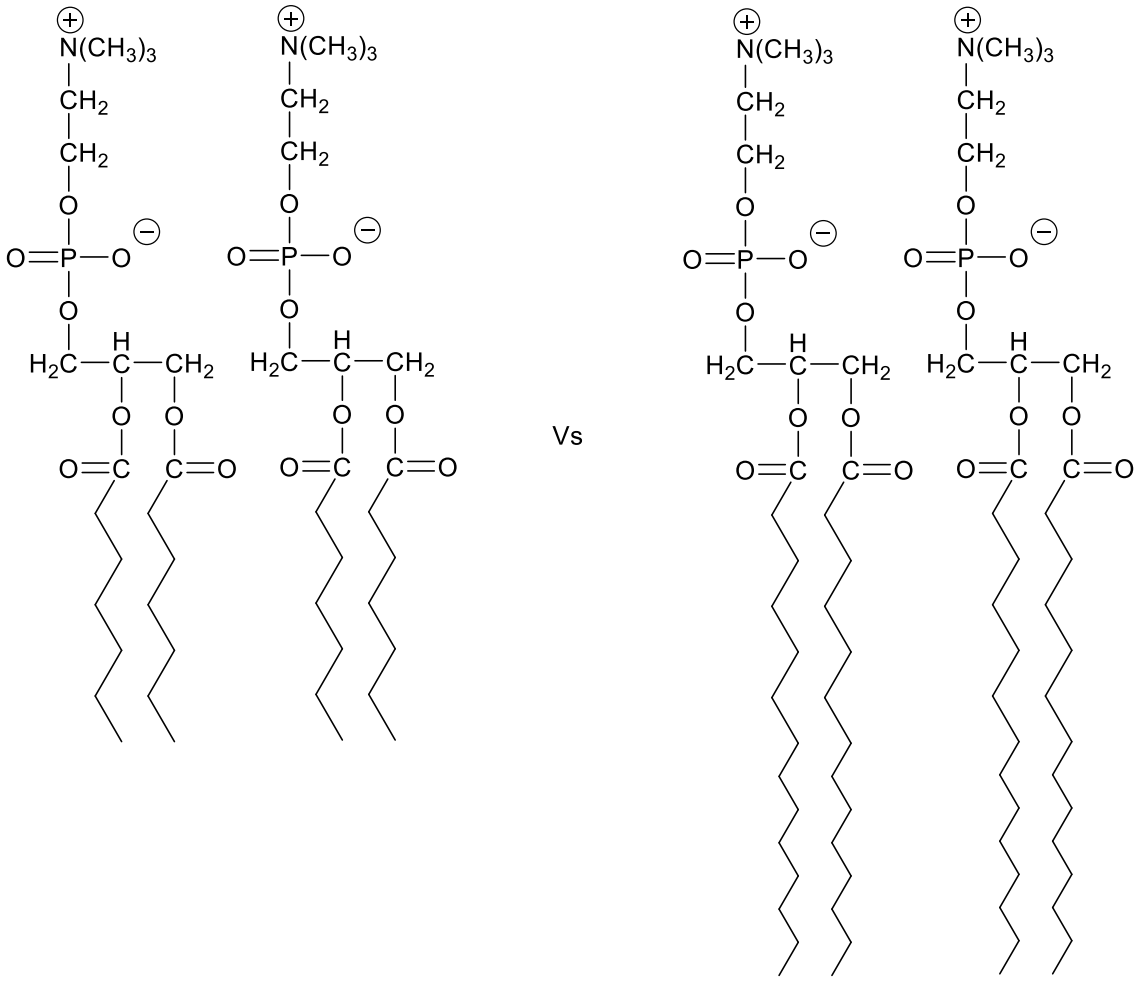
This biomolecule, as shown in the figure above, has two regions: a hydrophilic (water-loving) head and a hydrophobic (water-hating) tail. These two regions allow the phospholipid to form membranes spontaneously because the hydrophilic head group will point towards the water while the hydrophobic tails will face inward towards each other like so:



What is shown above is a micelle, a cell membrane would be comprised of an additional layer of phospholipids, with tails facing towards the tails shown above and heads facing towards the center of the circle. This is referred to as the phospholipid bilayer and the driving force for this spontaneous self-assembly is the polarity and intermolecular forces that attract each group together. The ionic head groups are attracted to the water because of strong hydrogen bonding interactions and ion-dipole interactions while the nonpolar hydrocarbon tails are attracted to each other via Van Der Waals forces. Therefore, because polar likes polar and nonpolar likes nonpolar, the self-assembly into this bilayer represents an energy minimum (**M**ASC) and will spontaneously form.

Many different types of phospholipids can compose cell membranes. These can vary by length of the carbon chain in their hydrophobic tails, by the number of pi bonds they have, and by

incorporating cholesterol. First, we will consider the effect carbon chain length has on membrane flexibility and we will look at this through the lens of intermolecular forces. Which type of phospholipid would have the strongest forces of attraction between the tail regions of the two shown below:

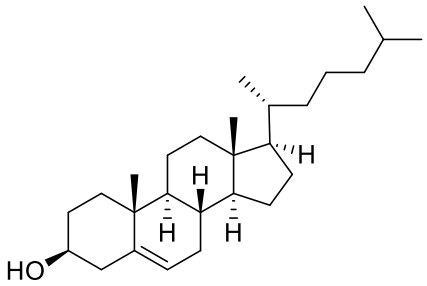


To answer this question, we have to recall what Van Der Waals forces are and how they relate to molecular weight. Van Der Waals forces are momentary dipoles that are formed by electrons occupying one side of the atom more than the other. These momentary dipoles attract the opposite charge just like any other dipole would and the more of these dipoles you have the stronger the force of attraction. Therefore, because the higher the molecular weight makes it more likely to have more momentary dipoles, higher molecular weight molecules have stronger Van Der Waals forces, which means that of the two phospholipids shown above, the one on the right has stronger attractive forces than the one on the left. This may seem trivial, but it is very important for organisms that live in very hot environments. *In very hot environments, the membrane is constantly moving very quickly and therefore it can easily melt if the hydrophobic tails are not tightly bound together*. Conversely, if an organism is in a very cold environment, the membrane is not moving much at all and the risk is not that it will melt, but instead that it will be too rigid to function properly (move proteins where they need to go, incorporate external

nutrients, etc.). In this scenario, the shorter fatty acid chain is preferable because these phospholipids are more flexible and can more efficiently prevent the membrane from becoming too rigid. It is effectively a Goldilocks situation, the membrane needs to be not too flexible that it will collapse but not too rigid that it will not move at all. But biology is smart and has evolved over billions of years to optimize the phospholipids that it uses in its membranes to accommodate the type of environment the organism is exposed to (MAS**C**).

The next parameter that can tweak membrane flexibility is the number of pi bonds in the fatty acid chain, these pi bonds are always ***cis*** in biological systems and because of this, they interfere with lipid packing. Due to them interfering with lipid packing, the attractive forces in the hydrophobic tails are weaker and this will increase flexibility.

Finally, the organism can incorporate cholesterol into their membranes. The structure of cholesterol is shown below:



It is also a very nonpolar molecule because of the large amount of hydrocarbons present in its structure. Because it is nonpolar it can be easily incorporated into the hydrophobic tail region of the phospholipid bilayer. Cholesterol is effectively a double agent, in both hot and cold temperatures, cholesterol will stabilize the membrane. Meaning, that regardless of the temperature, cholesterol will make the membrane happier. More flexible in cold environments and more rigid in hot environments.

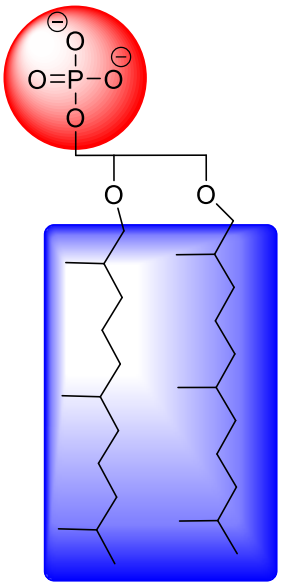
Practice Questions:

1. Suppose you are a bacteria living in a very cold environment, rank the following lipid compositions from best to worst:
    - Very long, saturated phospholipids
    - Very long, unsaturated phospholipids
    - Short length, saturated phospholipids
    - Short length, unsaturated phospholipids
2. Rank the lipid compositions from best to worst if the bacteria was in a very hot environment

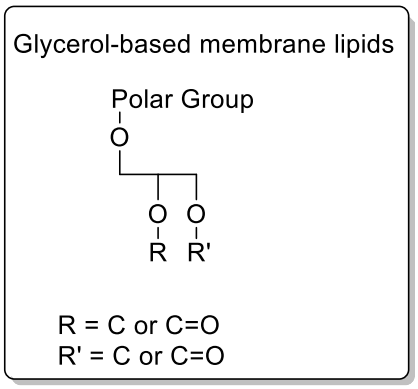
Answers:

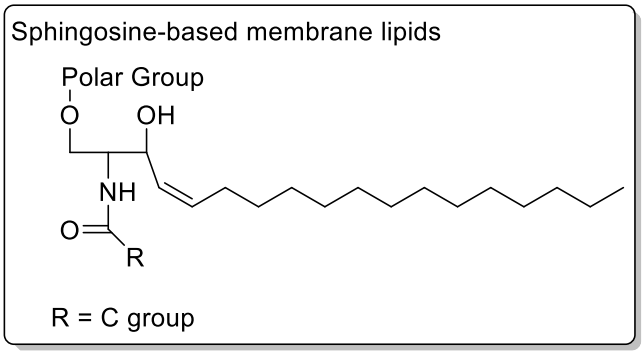
1. If we are in a very cold environment, the risk is that the membrane will be too rigid, therefore we should optimize the membrane's flexibility by having shorter fatty acids and having them be unsaturated (have pi bonds). Therefore the order from best to worst should be short length, unsaturated; short length, saturated; very long, unsaturated; very long, saturated.
2. In a very hot environment, the risk is that the membrane will be too flexible, therefore we should optimize the membrane's rigidity. This is just the opposite of the last problem, so the new order would be: very long, saturated; very long, unsaturated; short, saturated; short, unsaturated.

Other types of organisms use different types of lipid-based molecules to incorporate into their membranes. Archaea for instance use a glycerol-lipid that has *ether* linkages instead of the typical *ester* linkages found in bacteria and eukaryotes. An example of this lipid molecule is shown below:



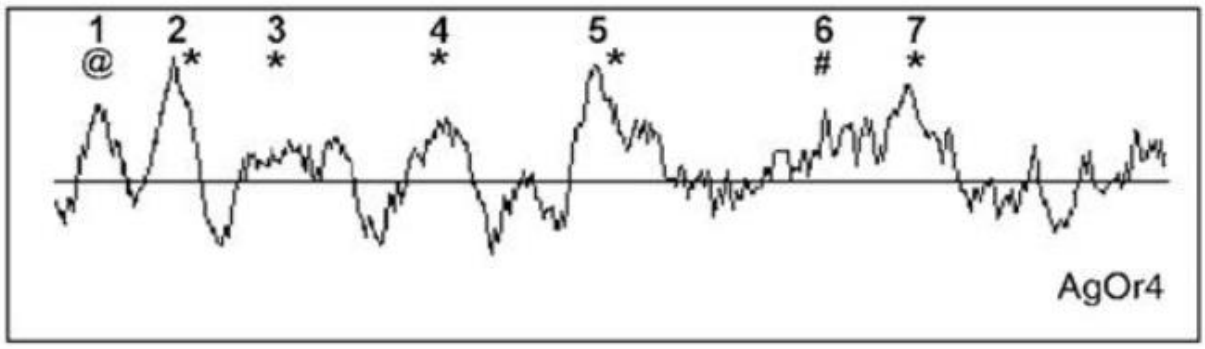
These ether linkages are important because they are more chemically resistant than the corresponding esters found in mammalian cell membranes. Recall from the Organic Chemistry survival guide that ethers are "exclusionary" and will only react with very strong acids. This accounts for the fact that archaea are frequently found in extreme environments and are capable of surviving in high temperature and pressure conditions (MAS**C**). Most membrane lipids have glycerol as a backbone, however, there are other membrane lipids that instead use a sphingosine backbone, a side-by-side comparison of the two backbones is provided below:





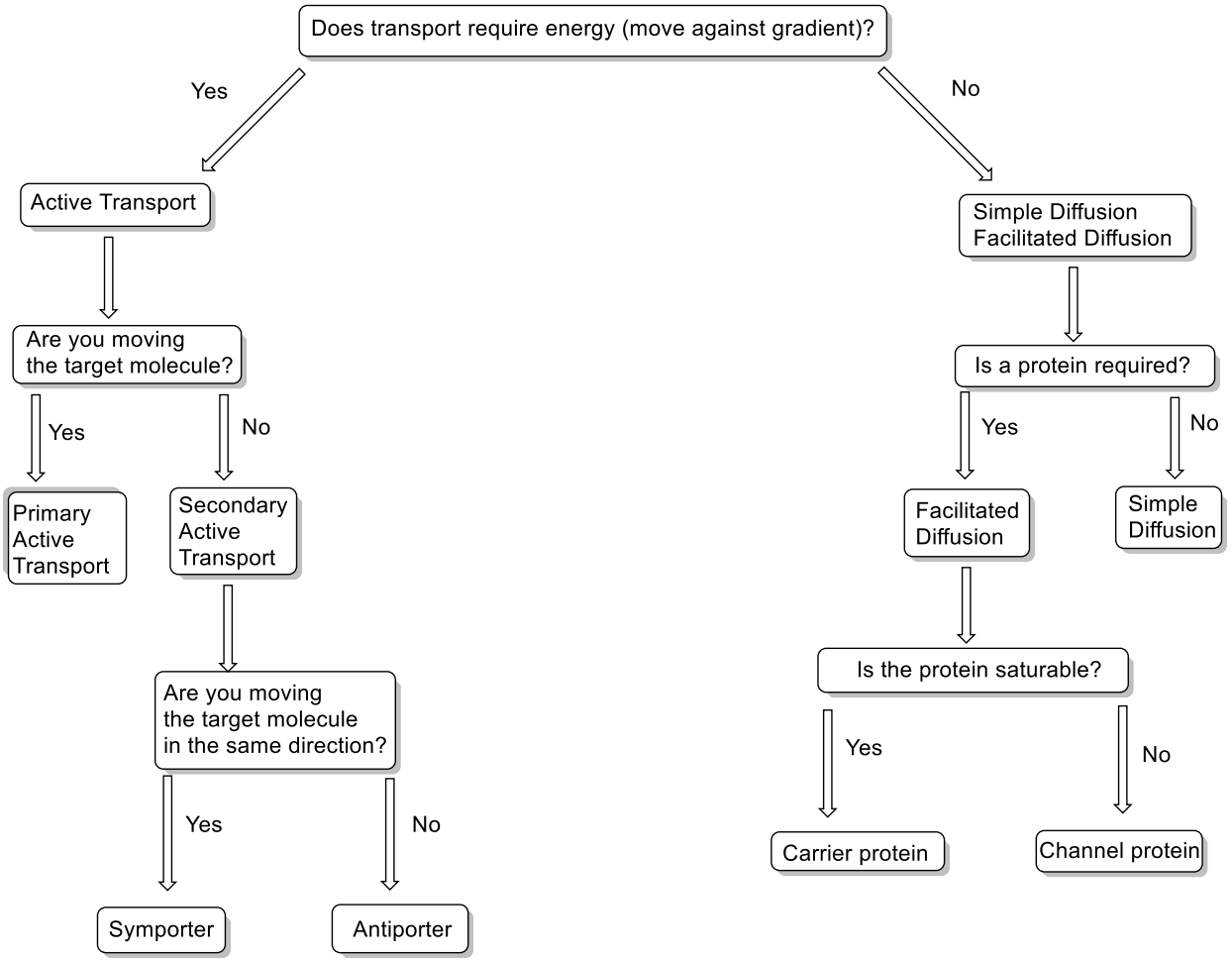
The major difference is that the sphingosine is much more nonpolar; it has a much longer alkyl chain of 18 carbons total compared to 3 carbons of glycerol. Sphinosine also has an amide linkage as opposed to the ether or ester linkages for the glycerol-based membrane lipids, this makes them more chemically resistant than esters, but less chemically resistant than ether lipids. These sphingolipids are commonly found in the myelin sheath of neurons and glycosphingolipids (sphingolipids with carbohydrates as their polar group) are used to determine blood type. All structural lipids are degraded by the lysosome using phospholipases that are used to cleave specific bonds in the membrane lipid.

Cell membranes have a whole host of proteins within them, these proteins can be transmembrane proteins, that go from the cytoplasm to the extracellular space or peripheral proteins that are anchored to one side of the membrane. The number of times that transmembrane proteins cross the membrane can be determined using a hydropathy plot, which takes a running average of the hydrophobicity of the amino acid residues and plots it against residue number. The number of times that the hydropathy plot crosses the x-axis is the number of times that it crosses the cell membrane. An example of a hydropathy plot is shown below for a GPCR protein which are known to be 7-pass transmembrane proteins:[7]

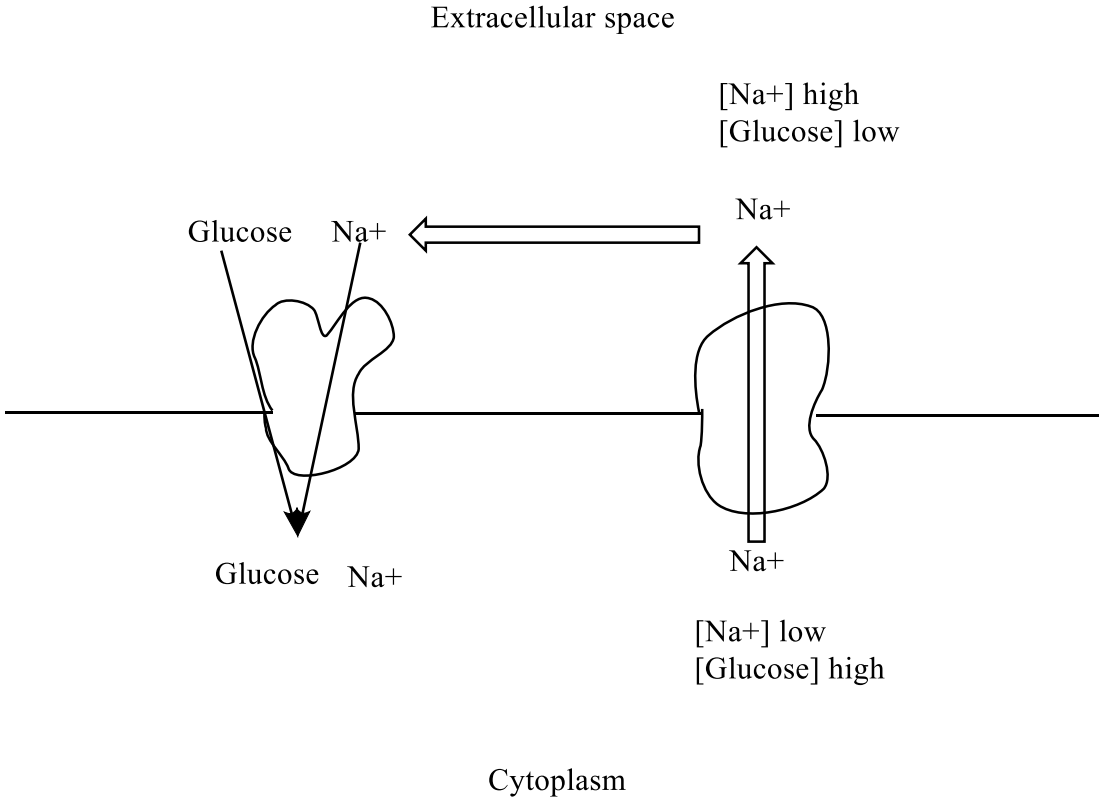


Chapter 11: Types of Transporters

Because the cell membrane is a phospholipid bilayer, molecules that want to cross it must be sufficiently small and nonpolar OR have a dedicated protein transporter. There are a number of different transport proteins and different method of transport. Broadly, there are three different transportation strategies that molecules can use: simple diffusion, facilitated diffusion, and active transport. Simple diffusion is just when the molecule crosses the cell membrane without using a protein, this is done only by nonpolar molecules and gases like O_2, these molecules always flow with their concentration gradient (from high to low) and therefore, there is no energy used. Facilitated diffusion is when the molecule crosses the cell membrane using a protein and with its concentration gradient (again from high to low), just as before, there is no energy used in this process either. Facilitated diffusion can occur through two different types of proteins, carrier proteins and channel proteins. Carrier proteins are proteins that will change their shape like your esophagus to essentially "swallow" the molecule of interest and move it into the cell. Because this process requires time, carrier proteins are saturable just like enzymes. Channel proteins on the other hand are simply proteins that are always open and allow the target molecule to move in and out of the cell with ease, these are NOT saturable. Because they are not saturable and could transport a very high volume of molecules very quickly, channel proteins are often gated either mechanically or by voltage. This is used in nerve cells to regulate action potentials by controlling the transport of Na^+ and K^+ using voltage-gated ion channels. Active transport on the other hand, involve protein pumps that will use ATP to power the movement of molecules *against their concentration gradient*. Because this moves molecules against their gradient, this process requires the energy stored in the high energy phosphoanhydride bonds of ATP (M**A**SC). This process of pumping molecules against their gradient can be done on the desired molecule itself, in which case it is primary active transport, OR the gradient that is established by the ATPase pump can be used to drive the facilitated diffusion of the target molecule, this is *secondary active transport*. The most commonly thought of primary active transport protein is the Na^+/K^+ ATPase pump, which will pump Na^+ out of the cell and move K^+ into the cell (think sweaty banana) in a $3Na^+:2K^+$ ratio. The most commonly thought of secondary active transport protein is the Na^+/glucose symporter protein, which uses the Na^+ gradient established by the Na^+/K^+ ATPase to move glucose into the cell. Secondary active symport transporters move the two molecules in the same direction, while antiport transporters move the two molecules in opposite directions. Let's put this all together in a flow chart:



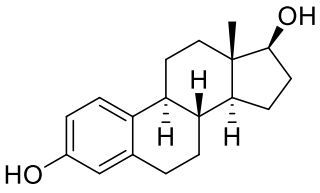
I know secondary active transport is a little confusion at first, so here is a diagram using the Na^+/Glucose symporter as our model case:



Because the two darker arrows are pointing in the same direction, that is to say, towards the cytoplasm, this secondary active transport process is considered a symporter process.

Practice Questions:

1. Identify the molecules listed below would be able to move across the membrane using passive diffusion, explain why on the premise of polarity:
   Water, oxygen gas, nitrogen gas, CO_2, glucose, ammonia, Cl^-, estradiol (structure shown below).



2. Protein X uses ATP to power the movement of Ca^{2+} to allow neurotransmission to occur, what type of transport is this?
3. Protein Y does not use ATP itself, but uses a concentration gradient established by an ATPase pump moving ion Z to power the movement of fructose against its concentration gradient and moves fructose in the opposite direction of Z. What type of transport is this?

Answers:

1. All nonpolar and gaseous molecules would be able to move into the cell via simple diffusion, therefore O_2, N_2, CO_2, and estradiol would be able to move via passive diffusion.
2. Primary active transport because it uses ATP and moves the target molecule directly.
3. Secondary active transport because it uses ATP but does not move the target molecule directly with the ATPase pump.

Chapter 12: Free Energy of Transport

As previously discussed, some transport processes require energy and others do not and this had to do with the preexisting concentration gradient. However, how do we mathematically show that this is the case? Moreover, what other factors contribute to the energy required to move certain molecules from the cell to the extracellular space? This can be expressed mathematically in two parts: the concentration component and the electrical component:

$$\Delta G = RTln\left(\frac{C_{to}}{C_{from}}\right)$$

$$\Delta G = zF\Delta\Psi$$

The first equation represents the concentration aspect of the gradient and the second equation represents the electrical aspect of the gradient, combined together gives the overall free energy of transport:

$$\Delta G = RTln\left(\frac{C_{to}}{C_{from}}\right) + zF\Delta\Psi$$

For a process to be spontaneous, the free energy change (ΔG) must be < 0 aka negative. Now we will analyze each aspect of the total equation in turn and rationalize why what we know to be true works out mathematically. This is a super useful skill especially for those of you who will eventually take p-chem.

We know that moving molecules from high concentration to low concentration is thermodynamically favorable and therefore spontaneous (ΔG < 0), therefore, if C_{from} is greater than C_{to} the overall expression should be negative.

$$\Delta G = RTln\left(\frac{C_{to}}{C_{from}}\right)$$

In this equation, R is the gas constant expresses as J/molK, which is 8.314 J/molK. T is the temperature in KELVIN. Just as a refresher, you can convert from Celsius to Kelvin using the following equation:

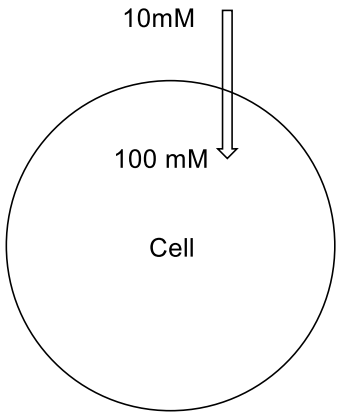
$$K = 273.15 + C$$

Let's suppose that for a particular cell, the glucose concentration on the outside of the cell is 10mM and the glucose concentration on the inside of the cell is 100 mM. If we wanted to find the free energy change of moving the glucose molecule from the outside of the cell into the inside of the cell at 37 degrees Celsius, we can do that using the equation above:

$$\Delta G = \left(8.314\frac{J}{molK}\right)(273.15 + 37)K * \ln\left(\frac{100mM}{10mM}\right) = 5937\frac{J}{mol} = 5.937\frac{kJ}{mol}$$

This checks out, because we know that we need to couple glucose transport with Na^+ transport for it to occur from last chapter. We already knew that transporting glucose this way by itself would not be possible, in other words, that it was nonspontaneous aka ΔG > 0.

I find that with these transport problems, drawing a picture goes a LONG way to understand what is going on, so here is a diagram for the situation we just did:



Obviously using this picture, it is clear to see that moving glucose in this manner is not favorable, it is moving *against* its concentration gradient.

Now we can evaluate the electrical component of the equation, this is the part that personally caused me the most confusion when learning this, and I think it had a lot to do with me trying to memorize specific circumstances rather than just logically thinking the equation through. We will not do route memorization here, I think that is horrible and that is not what guides of this kind are for, they are to help understand the material. All you need to know for understanding this part of the equation is the thing that everyone knows since like middle school, that opposite charges attract and like charges repel. Therefore, we can intuit that if like charges repel then transporting a negative charge to an area that is more negative would be unfavorable, and it is. Let's prove that using the equation:

$$\Delta G = zF\Delta\Psi$$

This equation looks really scary, but it isn't too bad. Z is the charge of the ion, for Cl^{-1} it is -1, for Ca^{2+} it is +2. F is Faradays constant, which is 96,500 J/Vmol and $\Delta\Psi$ is the potential difference between where the ion is going versus where it came from. This potential difference is effectively just the difference in charge between the two locations, if the difference is negative than the place the ion is going is more negative, if it is positive then the ion is going to a more positive place. This NEED to be expressed as V because we need the units to cancel. Now we can apply this to a situation involving Cl^- transport. Suppose we want to transport Cl^- ion from the extracellular space to the cytoplasm of the cell if the transmembrane potential is -67 mV (inside of cell is more negative). Assume that the concentration of Cl^- is roughly the same on either side of the cell membrane (this is not true in actual cells, but we just care about the electrical component of the gradient for right now so we will just ignore it).

$$\Delta G = (-1)\left(96500\frac{J}{Vmol}\right)(-67mV)\left(\frac{1V}{1000mV}\right) = 6465.5\frac{J}{mol} = 6.4655\frac{kJ}{mol}$$

This makes sense because we are moving the Cl^- to a more negative location and these negative charges will repel the Cl^- because as we know like charges repel, thus making the process thermodynamically unfavorable and the free energy change greater than 0. Now that we understand how these two parts of the equation play a role in determining the overall free energy change, we can get to combining the two of them with some practice questions.

Practice Questions:

1. Calculate the free energy change for transporting Ca^{2+} from the sarcoplastic reticulum which has a concentration of Ca^{2+} of 12 μM to the extracellular space which has a concentration of 12,000 μM. The intermembrane potential is +67 mV and the temperature is 37 Celsius.

Answer:

1. $\Delta G = RTln\left(\frac{C_{to}}{C_{from}}\right) + zF\Delta\Psi$ plugging in the known variables gives us

$$\Delta G = \left(8.314\frac{J}{molK}\right)(273.15 + 37)K \ln\left(\frac{12{,}000\mu M}{12\mu M}\right) + (+2)(96500\frac{J}{Vmol})(0.067V)$$

$$\Delta G = 30743\frac{J}{mol} = 30.743\frac{kJ}{mol}$$

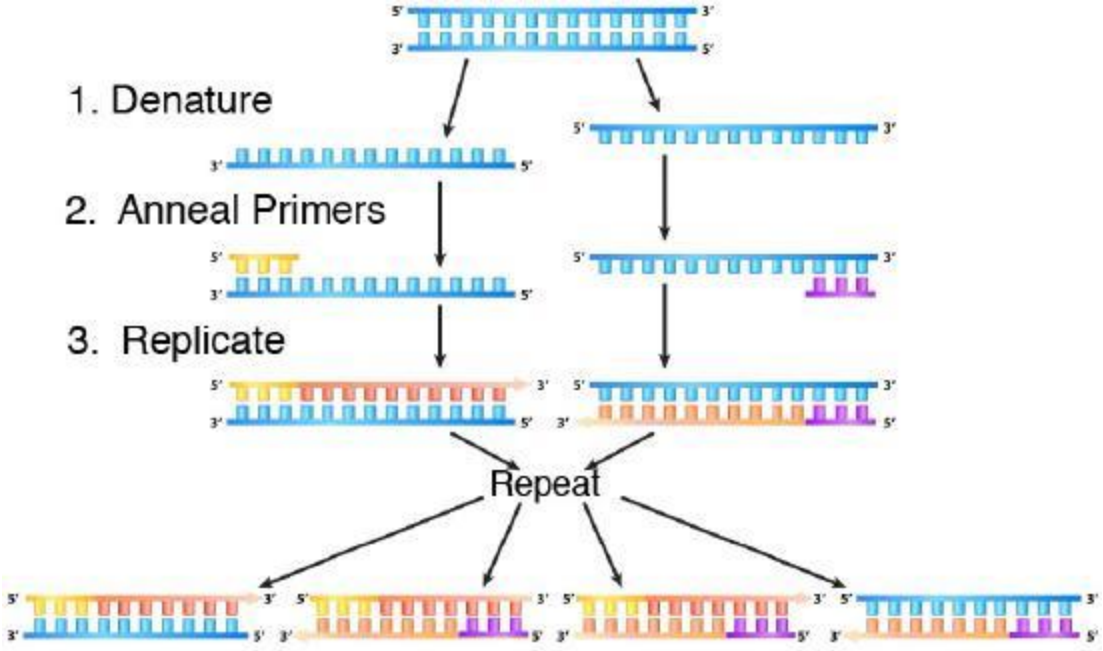
Which makes sense because the Ca^{2+} is going against not only its concentration gradient, but ALSO its electrical gradient.

Focus 4: DNA Technology and Common Biochemical Techniques

Chapter 13: Polymerase Chain Reaction + Plasmids

Some of the most important DNA technologies that are polymerase chain reaction (PCR) and plasmid transfection/transduction. First, we will discuss PCR and then we will discuss plasmid transfection and how it can be used to genetically engineer organisms to express certain genes on command.

PCR is a biochemical technique that is used to amplify a target sequence of nucleic acids, usually DNA. This technique is commonly used to diagnose SARS-CoV-2 infections, the virus that causes the Covid-19 pandemic, and therefore it is important to understand how PCR works on a molecular level. At its core, PCR simply emulates how DNA replication typically occurs. It uses similar enzymes, requires the same molecules to work, and functions more or less the same way. There are three stages of PCR, these are the denaturation, annealing, and elongation steps pictured below:[8]



In the denaturation step, the DNA target molecule is heated to break the hydrogen bonds between the DNA bases to give two separate strands.

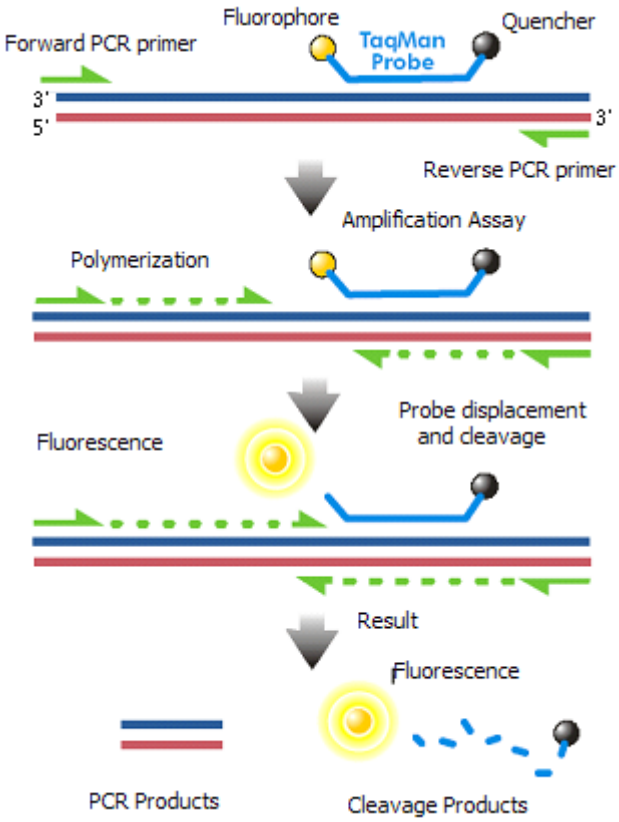
Because the DNA is antiparallel, meaning that each strand has opposite directionality (one is 3'→5' and the other is 5'→3'), we need to have two primers, one that will pair with the 3'→5' segment and another that will bind the other. These are depicted as the yellow and purple primers shown in the figure above and are frequently referred to as the forward and reverse primers. These primers will bind the DNA at the target sites because of base complementarity. Recall from introductory biology that A pairs with T and G pairs with C. The step where the primers anneal to the target DNA strand is known as annealing and this typically occurs at a much lower temperature than the denaturation step because we do not want the primer to dissociate (leave) from the DNA strand it is paired with.

Then in the elongation step, a DNA polymerase is used to add nucleotides until the end of the strand is reached and the process is repeated a certain number of times or cycles. This DNA polymerase is typically a thermostable DNA polymerase because these reactions occur at elevated temperatures. The most commonly used DNA polymerase for this step is the Taq polymerase, though other polymerases have been discovered and developed that are more accurate and faster than Taq so by the time this text is available on the website, Taq will likely no longer be the polymerase of choice. For any PCR reaction to occur, these things must be present:

1. Thermostable DNA polymerase
2. dNTPs aka your A, T, G, and C bases
3. DNA target sequence (typically manufactured in a lab)
4. Forward and reverse primers
5. Buffer solution and Mg^{2+} **

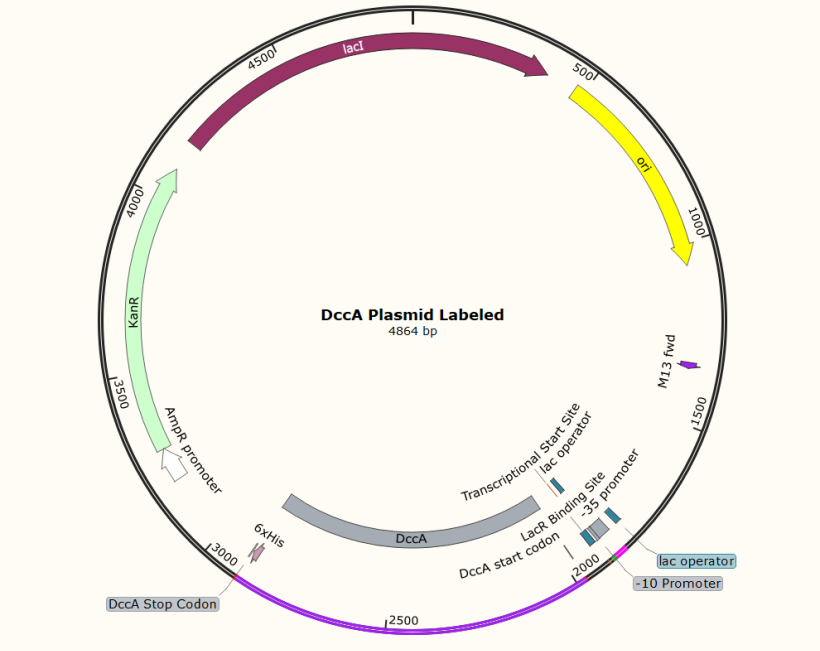
**: Mg^{2+} is required for the DNA polymerase mechanism, which we will cover later in this textbook. Please see Chapter 49 for more details regarding the DNA polymerase mechanism.

This is very useful for amplifying DNA sequences, but what about RNA sequences? To amplify RNA sequences, we have to use RT-PCR. The RT stands for reverse transcriptase. This enzyme is effectively a copying machine that will make a complementary DNA strand (cDNA) from an RNA template. This is especially useful for diagnosing Covid patients because SARS-CoV-2 is an RNA virus and the PCR test looks for the viral RNA. Once the cDNA is produced by reverse transcriptase, the PCR reaction mentioned above is done using forward and reverse primers specific to the cDNA molecule of interest. For quantification, the PCR reaction is done using fluorescent probes. These fluorescent probes consist of an oligonucleotide sequence that has a fluorescent molecule covalently attached and a fluorescence quencher. This prevents fluorescence from occurring when the fluorescent probe is in tact. Once the Taq polymerase extends and hits the fluorescent probe region, the fluorescent molecule gets released and is separated from the quencher molecule, see diagram below:[9]



This allows detection using fluorimeters. Quantification is possible because the more PCR amplifications that occur, the more fluorescence that is measured. In other words, the amount of fluorescence is proportional to the number of cDNA molecules that were originally in the reaction. This is what allows the PCR test for SARS-CoV-2 to be so accurate, all of these probes and primers are SPECIFIC to the SARS-CoV-2 RNA sequence (or more accurately, the cDNA copy of the RNA sequence) and therefore we can get highly accurate results by running the PCR for a standardized number of cycles. This whole process of quantifying RNA this way is referred to as qRT-PCR or quantitative reverse transcriptase polymerase chain reaction and this is how Covid tests work.

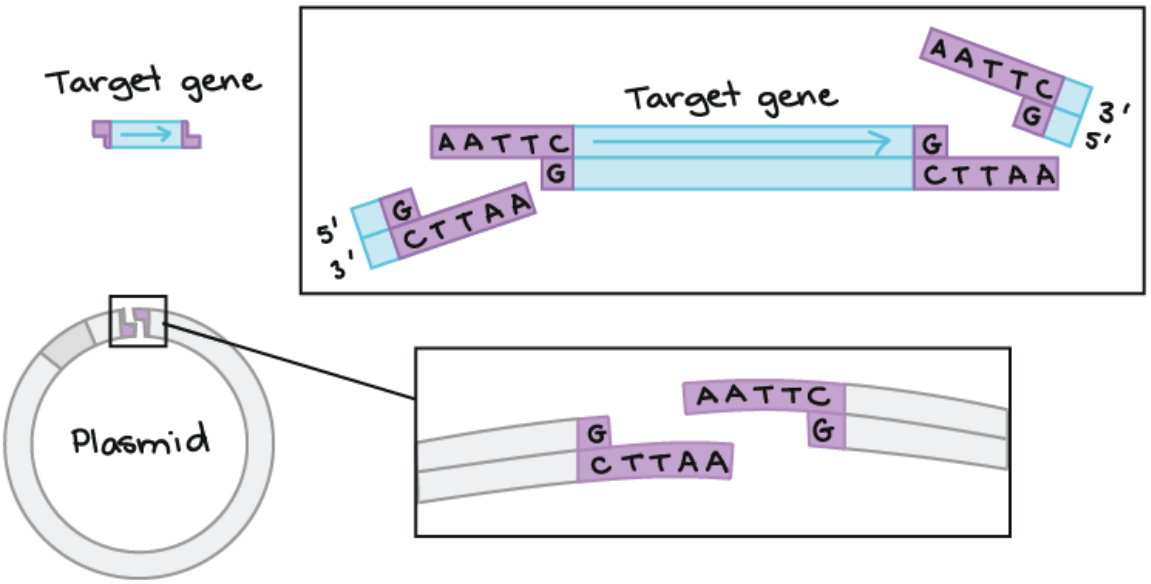
But PCR is useful not only for clinical diagnosis of coronavirus or amplifying desired target sequences of DNA, but it is also useful for site-directed mutagenesis that allows us to genetically engineer organisms to produce a mutated (chimeric) protein of interest. Before we discuss this incredibly useful way of using PCR, we should discuss the basics of plasmids and how they can be used for genetic engineering. Plasmids are pieces of circular DNA that are frequently found in bacteria, an example of a plasmid is shown below:



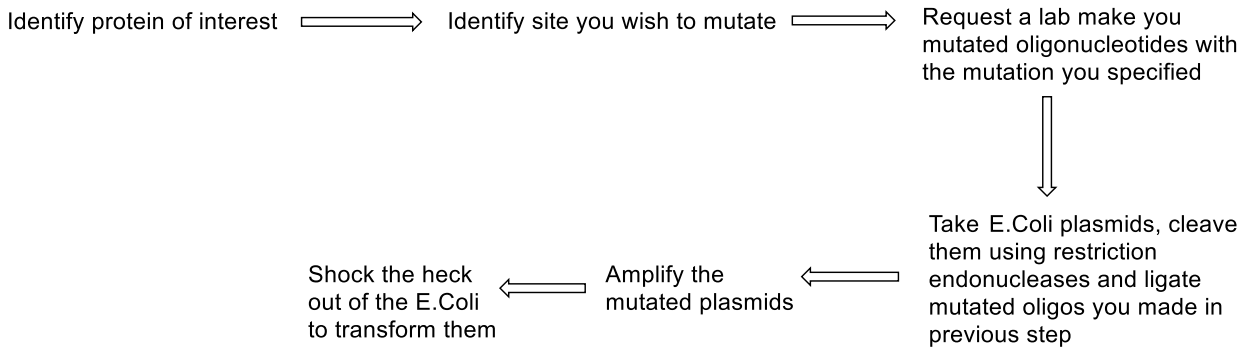
Here the target protein is DccA, which is a type of haloalkane dehalogenase. These enzymes are important for bioremediation efforts, if you are curious about the specifics, just email me. This was part of a research project that I did in undergrad. Expression plasmids typically have the following regions:

1. Antibacterial resistance gene, in this case we had a kanamycin resistance gene (KanR). This allows you to know which bacteria have the plasmid because they will not die when exposed to kanamycin.
2. Origin of replication so the plasmid can replicate (Ori)
3. Target protein with some sort of tag at the end, in this case we had a His tag and the target protein was DccA. His tags will be discussed in further detail in Chapter 15, suffice it to say that the His tag allows for purification of the DccA protein.
4. Inducible operator, in this case we used the lac operator and this is more or less standard. For more details about the lac operon, see Chapter 66. This allows the experimenter to control expression of the target gene.

Custom plasmids can be manufactured using restriction endonucleases that will cleave off specific DNA sequences leaving “sticky ends.” These sticky ends are exposed DNA bases that can be used to allow new DNA to be inserted, see the figure below.[10]



For the production of the plasmid shown above, the lac genes were taken out and replaced with the DccA gene. This DccA gene had mutations at a particular target site of the experimenter's choosing, I chose S157 to mutate so I had to mutate codon 157 to give all possible amino acids in a process called *iterative saturation mutagenesis*. These new recombinant plasmids can be amplified using the standard PCR technique we just discussed and can be transformed into bacteria using electroporation. Transformation is the process by which expression plasmids are inserted into bacteria and electroporation is simply exposing the bacteria to high voltage to make their cell membranes more permeable. A diagram showing the whole process is shown below:



Practice Questions:

1. Explain what would happen to the PCR yield if the forward and reverse primers annealed to themselves.
2. Explain what would happen to the PCR yield if the forward and reverse primer melting temperature is below the annealing temperature.
3. Explain what would happen to the fluorescence measurements in qRT-PCR if the fluorescent probes dimerized.
4. Explain what would happen to the fluorescence measurements in qRT-PCR if the fluorescent portion of the probe does not release.

5. What would happen the number of bacterial colonies exposed to kanamycin if electroporation was unsuccessful in the experiment I mentioned above?

Answers:

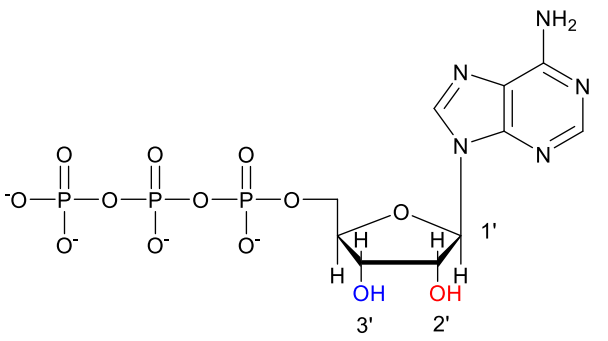
1. If the primers annealed to themselves then they could not bind the target DNA sequence and therefore the extension portion of PCR would not be able to occur. The end result would be PCR yields will be lower.
2. If the forward and reverse primers melt below the annealing temperature then they cannot anneal the target DNA sequence and therefore extension cannot occur. The end result is PCR yields are lower.
3. If the fluorescence probe dimerizes then the Taq polymerase cannot hit into the probe and therefore the fluorescence of the sample should decrease.
4. If the fluorescent portion of the probe does not get released then it will stay close to the quencher which would prevent it from fluorescing. Therefore, the fluorescence of the sample should decrease.
5. If the electroportation was not successful, then the bacteria would not have the plasmid. If they do not have the plasmid then the bacteria would lack the kanamycin resistance gene and therefore the colonies would decrease/die off.

Chapter 14: DNA Sequencing (Sanger and Next Generation)

Now that we know how to detect and amplify DNA/RNA, we can now discuss how we know the sequence of DNA. First, we will discuss the older method of Sanger sequencing and then we will discuss the newer method (NGS).

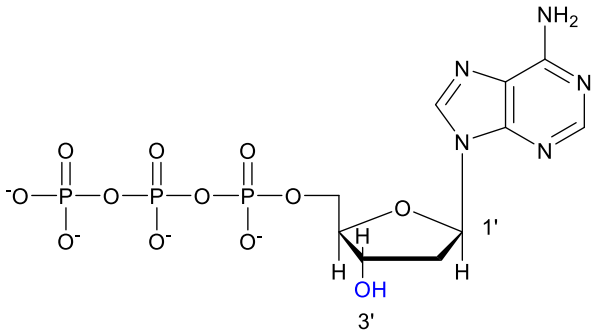
Sanger sequencing is the first method of determining a DNA sequence that was widely used and the premise of Sanger sequencing is that DNA polymerization can ONLY occur if there is a 3'OH on the nucleotide being added by the polymerase. Let's go over what that means specifically using the following structures:

ATP



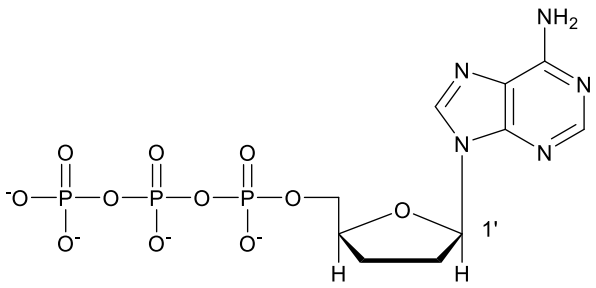
in RNA

dATP



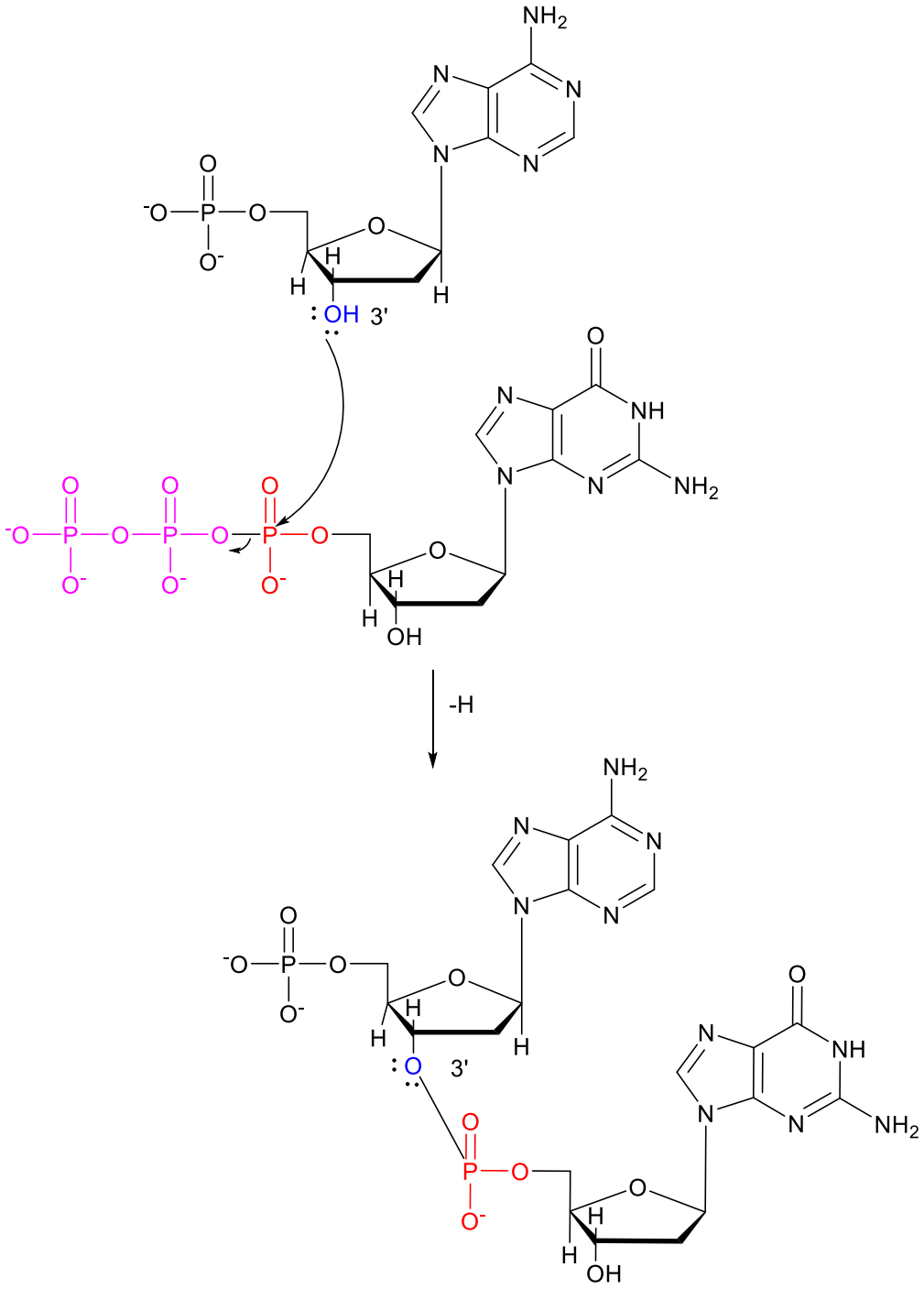
in DNA

ddATP



Cannot be used for polymerization

DNA polymerization (and RNA polymerization) need to have the 3’ OH group because that is what attacks the next 5’ phosphate on the next nucleotide in the mechanism shown below:



Because the ddATP does not have that 3’OH group, it cannot extend the chain and therefore the DNA strand terminates at that position. In Sanger sequencing, the dideoxynucleotides used are fluorescent and therefore, the DNA strands can be run on a gel to visualize them and determine their length. Because the DNA polymerization process gives a complementary strand, whatever dideoxynucleotide (ddNTP) stops the chain, the complementary nucleotide had to be in the

original strand. For example, if ddATP stops the DNA strand at the third base, then the third base of the target DNA strand is a T because T pairs with A. Varying what ddNTP gets put in the polymerization mixture will give an idea as to which DNA base is at that position. To start off, four reaction mixtures are created, one for each ddNTP that can be added along with a primer, DNA polymerase, and dNTPs that can be used for ordinary polymerization. Then the reaction products of these mixtures are ran on a DNA agarose gel to determine their length and the DNA sequence can be determined. Shorter DNA strands will appear at the bottom while longer DNA strands will appear at the top because larger DNA strands will experience more resistance in the thick agarose gel when moving. This is best explained with an example. Suppose the DNA sequence we wish to sequence is the following:

3'-AGCGTACCAGTAC-5'

If that is the DNA sequence we wish to solve for, then we can add a primer starting like so:

3'-AGCGTACCAGTAC-5'

5'-TCG-3'

From this primer we can extend. If we were in the ddCTP reaction mixture, then wherever there is a G base in the target strand, we could stall the base after. The important thing is that the ddNTPs can be incorporated into the growing strand of DNA, but they cannot extend and they can interfere with this process at any part of the elongation process. Likewise if we were in the ddGTP reaction mixture, wherever there is a C base in the target strand, we would stall after etc. etc. Let's consider the ddCTP reaction mixture and let's look at the different nucleotide strands we could get:

5'-TCGC-3'

5'-TCGCATGGTC-3'

5'- TCGCATGGTCATG-3'

Therefore, in the ddCTP lane, we would have three distinct bands that correspond to the oligonucleotide sequences shown above. Because they all have different lengths, they will all occur at different locations in the gel. Now we can consider the ddATP reaction mixture:

5'-TCGCA-3'

5'-TCGCATGGTCA-3'

5'- TCGCATGGTCATG-3'

Just as before, we would get three distinct bands and two of them would be unique to the ddATP lane.

Now we can bring this all together for all four lanes like so, the highlight represents fluorescence:

Base Number After Primer	ddATP	ddGTP	ddCTP	ddTTP
1	Continue	Continue	Stop after	Continue
2	Stop after	Continue	Continue	Continue
3	Continue	Continue	Continue	Stop after
4	Continue	Stop after	Continue	Continue
5	Continue	Stop after	Continue	Continue
6	Continue	Continue	Continue	Stop after
7	Continue	Continue	Stop after	Continue
8	Stop after	Continue	Continue	Continue
9	Continue	Continue	Continue	Stop after
10	Continue	Stop after	Continue	Continue

Therefore, from these data, we can determine that the DNA strand that we polymerized was:

5’-TCGCATGGTCATG-3’

And therefore the complementary strand (target DNA) would be:

3’-AGCGTACCAGTAC-5’

Which it is! This same data in the table above could be shown visually on a gel like so:

ddATP	ddGTP	ddCTP	ddTTP
	———		
			———
———			
		———	
			———
	———		
	———		
			———
———			
		———	

Essentially the ddNTPs serve as fluorescent markers that show you where the nucleotide gets added in the overall DNA strand and the DNA polymerase will add these nucleotides because it doesn't know any better. Once it adds these nucleotides, they are effectively going to jam the polymerase and prevent it from extending further, leaving the fluorescent marker at the end, which experimenters can visualize using fluorescence measurements and running it on an agarose gel.

Next generation sequencing functions practically the same, however, each ddNTP has a different chromophore attached to it so that they give a unique color when fluorescing and the reaction does not get separated by an agarose gel, but instead by a capillary electrophoresis instrument that can automate the sequencing process and recognize the different chromophores in the different DNA sequences.

Practice Questions:

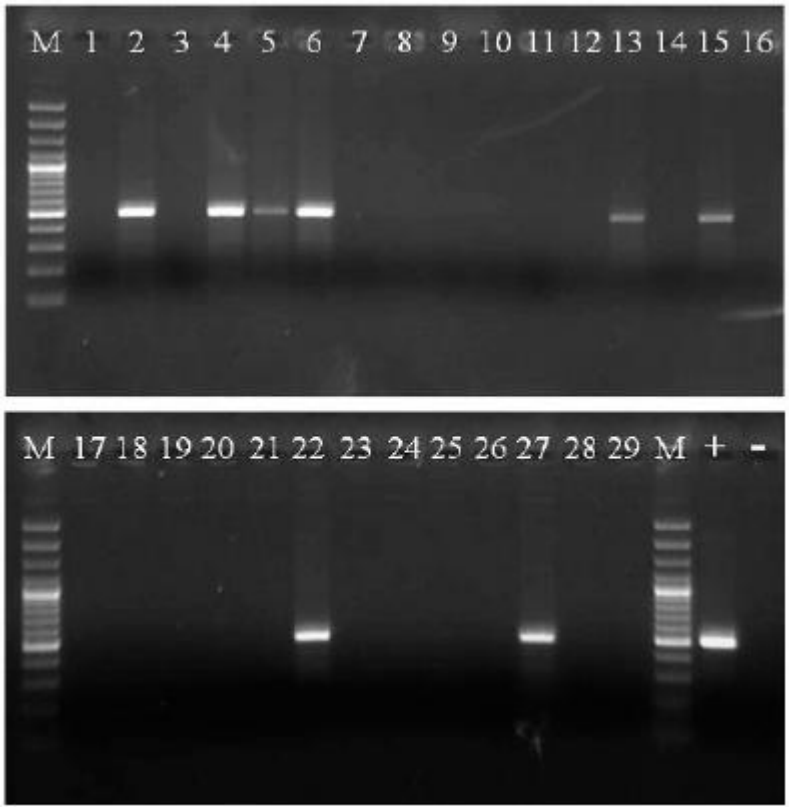
1. What would happen to the Sanger Sequencing experiment, if the primers were not added?
2. What would happen to the ddATP lane if ddATP was not added?
3. What would happen to the ddATP lane if dATP was not added?

Answers:

1. If the primers aren't added then you can't extend the chain and therefore all lanes would have no signal.
2. If ddATP was not added to the ddATP lane, then the chain would not terminate in that lane and you would just get one very long DNA sequence that is complementary to the target sequence.
3. If dATP was not added to the ddATP lane, then the chain would terminate after the very first T base in the target DNA sequence (also known as the first A base incorporated by the DNA polymerase) because the only thing that COULD get incorporated into the chain at that point is the ddATP, which as we know always terminates the chain.

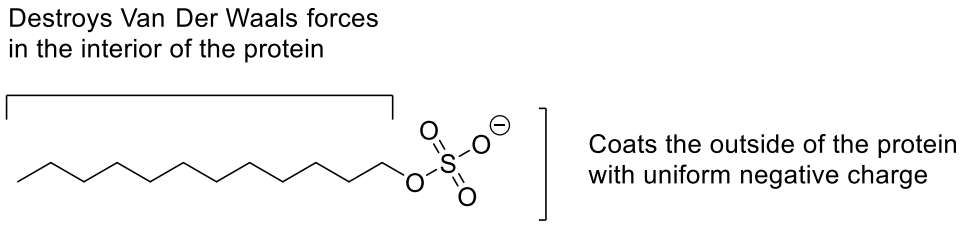
Chapter 15: Gel Electrophoresis and Column Chromatography

One of the most important ways to visualize and determine the size of proteins and DNA is through gel electrophoresis. This technique involves running the protein or DNA across a thick, viscous gel using a voltage source. Both DNA and proteins need to be stained before analysis so that visualization is possible. For a DNA gel electrophoresis, typically the DNA is stained with ethidium bromide. This fluorescent dye is flat and can bind to the DNA double helix, allowing for fluorescence measurements of the DNA in the gel. For protein gel electrophoresis, typically the protein is stained with Coomassie blue dye. In both types of gel electrophoresis, there are several lanes that are loaded via micropipette and the first lane is typically a standardized ladder that allows the experimenter to approximate the size of the proteins or DNA present in each lane. An example DNA gel is shown below where the M is the ladder and 1-16 are the lanes containing the DNA sample of interest:[11]



The fundamental principle behind both types of gel electrophoresis is the same. Negatively charged objects move from negative to positive. Because DNA is always negatively charged due to its phosphate backbone, it will always move from negative to positive since opposite charges attract. The gel that it moves through is very viscous and therefore it constantly slows down the DNA as it travels through it (think of you walking through quicksand as an example). Therefore, shorter pieces of DNA will move faster and father; they encounter less resistance to movement when flowing through the gel. Protein gel electrophoresis uses the same concept, however,

because not all proteins have the same shape and because all proteins are not negatively charged, the proteins need to be denatured and coated with a negative charge. To add the negative charge and help facilitate denaturing the protein (making the proteins linear), SDS is added to the protein sample. SDS is a detergent molecule that will break up the hydrophobic interactions of the protein of interest and cover the exterior of the protein with its negatively charged sulfate groups. The structure of SDS is shown below:



The protein will also be exposed to beta-mercaptoethanol, which is a mild reducing agent and will irreversibly remove disulfide bonds that can help stabilize the protein structure. Along with these two reagents, the protein is also heated to break all of the hydrogen bonding, dipole-dipole, and ionic bonding interactions that could help give the protein its native conformation. Now that all the proteins are linearized (aka denatured) and coated with a uniform negative charge, they can be separated using gel electrophoresis. Just as before, the shorter proteins will travel farther in the gel compared to the longer proteins and in this way, proteins can be separated by size. The only difference between protein gel electrophoresis and DNA gel electrophoresis is the type of gel and type of dye. In protein gel electrophoresis, a polyacrylamide gel and Coomassie blue dye is used while in DNA gel electrophoresis, an agarose gel and ethidium bromide dye is used. Provided below is a table comparing the two methods:

----------	Protein Gel Electrophoresis	DNA Gel Electrophoresis
Dye	Coomassie blue	Ethidium bromide
Additional reagents used	SDS and betamercaptoethanol	None
Gel material	Polyacrylamide	Agarose
Separates molecules on basis of	Size	Size
Molecules move from blank to blank	Negative to positive	Negative to positive

Practice Questions

1. Suppose you have a protein, X, that has several disulfide bonds. How would the protein gel look if you forgot to add betamercaptoethanol compared to if you had done it properly?
2. Suppose you connected the terminals improperly (negative on bottom and positive on top) how would the results look like on a DNA gel? Protein gel? How could you tell definitively that there was something wrong without checking how you connected the terminals?

3. Some proteins are dimeric in their native conformation. How would these proteins look like using an SDS-PAGE gel? (SDS-PAGE is the type of protein gel electrophoresis that we have discussed so far). Does an SDS-PAGE gel give an accurate indication about these proteins size? Why or why not?
4. There is a separate type of protein gel electrophoresis called blue native polyacrylamide gel electrophoresis (BN-PAGE) in which SDS is not added and the protein is studied in its native conformation. How would this gel look in comparison to an SDS-PAGE gel for proteins that are multimeric (have several subunits)? Why then do you think this technique is useful?

Answers:

1. If the protein of interest, in this case protein X, has a lot of disulfide bonds, then forgetting to add betamercaptoethanol would cause those disulfide bonds to be retained. Because they are retained, the shape of the protein is not linear and therefore is likely more compact/similar to its native shape. Because the protein is more compact, it will encounter less resistance as it moves through the polyacrylamide gel matrix and therefore it should appear further down the gel than it ordinarily would if this was done properly.
2. If the terminals are not connected properly and the positive is on the top while negative is on the bottom, then the protein and/or DNA would not separate at all. Instead, they would congregate at the top because they would be much more attracted to the positive terminal than the negative terminal (they are both negatively charged and opposites attract). This would appear to suggest that all the proteins or DNA in your sample are extremely large. Without looking at how the terminals are connected, you can tell you messed it up because the ladder would also be effected by this and therefore the ladder (which you know has many proteins or DNA segments of *different* lengths/masses) would show one band strictly at the top with no differentiation. Or the proteins/DNA would leave the gel entirely out of the top of the instrument.
3. If proteins are dimeric in their native conformation, SDS-PAGE would separate the two subunits because SDS-PAGE is a denaturing technique. Therefore, we would have two separate bands on the gel if the two subunits are of different length/mass or one band if the two subunits are of the same length/mass. In either case, the gel does not give an accurate indication of the mass of the overall protein because the two subunits are separated.
4. The BN-PAGE gel would have the multimeric protein appear much higher up on the gel because it is heavier compared to the SDS-PAGE technique. This helps show the true weight of the protein and can help determine if a protein is made of subunits of the same mass because in that case it would appear higher up on the gel compared to the band of the subunit in the SDS-PAGE.

Colum chromatography is used to purify a protein sample from a crude cell lysate. In all forms of column chromatography, a stationary phase and a mobile phase is used. The stationary phase is typically the physical surface upon which proteins can bind. The mobile phase is the solvent that is being moved through the column. Typically, in these procedures, the eluent is collected as it comes out of the column and different fractions are stored in some kind of glass container. Proteins can be separated by the following three characteristics:

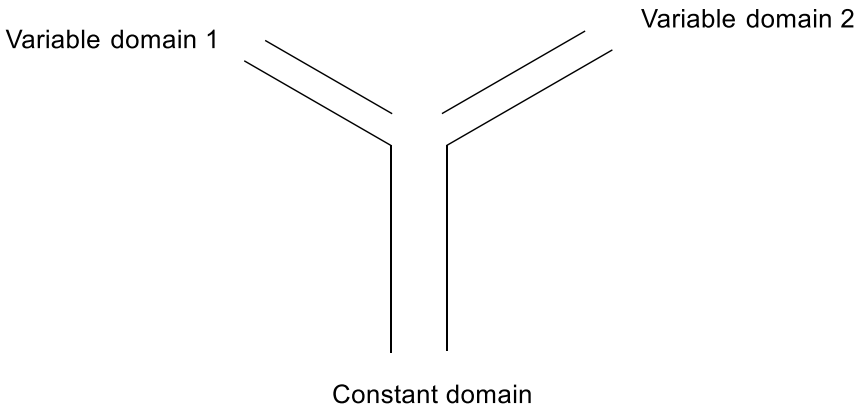
1. Charge
2. Size
3. Binding affinity to substrate

To separate proteins by charge, typically ion-exchange chromatography (IEC) is used. In this technique, proteins are exposed to a column that contains a stationary phase (where the proteins could get trapped) that has either negatively charged or positively charged ions on its surface. To trap a protein that has a net *negative* charge, one would use *anion exchange chromatography* and if you wanted to trap a protein that has a *positive* charge, one would use *cation exchange chromatography*. Protein charge comes out due to the charge of the acidic and basic residues on its surface, therefore, to remove the protein of interest from the column, the pH simply needs to be higher or lower to remove these charges (negative means lower, positive means higher).

To separate proteins by size, typically size-exclusion chromatography (SEC) is used. SEC uses a porous polymer matrix (stationary phase) with tiny holes in it that causes smaller proteins to get stuck for longer in the stationary phase. Therefore, larger proteins will elute first while smaller proteins will elute last because they get stuck in the nooks and crevices of the polymer matrix. All proteins will eventually elute over time, so many fractions must be collected from this type of chromatography.

The most effective means of purifying proteins is by their affinity to a particular molecule or substrate. This is done through affinity chromatography and this technique uses a stationary phase coated with a particular molecule or antibody that will bind the protein of interest. The most common method of affinity chromatography is immunoaffinity chromatography where the stationary phase is covered with antibodies that are specific for their protein target. Antibodies are Y-shaped proteins that are effectively targeted molecular missiles that will recognize and bind to a particular target with very high affinity. A representative structure of an antibody is shown below:

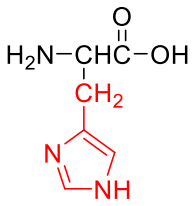
Variable domains are those domains responsible for substrate binding



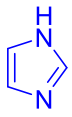
Constant domains are the regions of the antibody
that are the same for all antibodies produced by a given organism

Antibodies work by recognizing a particular part of the target called an epitope. Antibodies that only recognize one epitope are called monoclonal antibodies, while antibodies that recognize several different epitopes are called polyclonal antibodies. Regardless of the antibody type, these proteins are extremely specific to their target and will effectively only bind the thing they are designed to recognize. Therefore, immunoaffinity chromatography is incredibly powerful because only the protein of interest will bind the stationary phase and all other proteins will be washed away. To elute from immunoaffinity chromatography, the pH can be changed to alter the shape of the protein of interest and cause it to dissociate from the antibody. The second most commonly used form of affinity chromatography is Ni-NTA columns. These columns are used when the protein of interest has a His-tag. These proteins are commonly genetically engineered variants that the experimenter wants to purify. Because His-tags are not frequently found in other proteins naturally, Ni-NTA columns are incredibly powerful ways of separating the genetically engineered protein from the rest of the proteins. This technique works because histidine is very similar to imidazole, which are both known to bind Ni^{2+} with very high affinity. The structure of histidine and imidazole are shown below:

Histidine



Imidazole



His tags are made of many repeating His residues, each of which will like the Ni^{2+} in the column and will stick to it. To elute the His-tag protein, imidazole can be used to elute the protein

because imidazole will compete for binding the column and in doing so will displace the His residues on the target protein, causing it to leave the column.

Practice Questions:

1. Develop a procedure using at least two techniques mentioned in this chapter to separate out A from the following peptide sequences:

A: EEVATYD

B: VVSSCWYYMPPGQW

C: RHYRWMK

D: HHHHHWYILLP

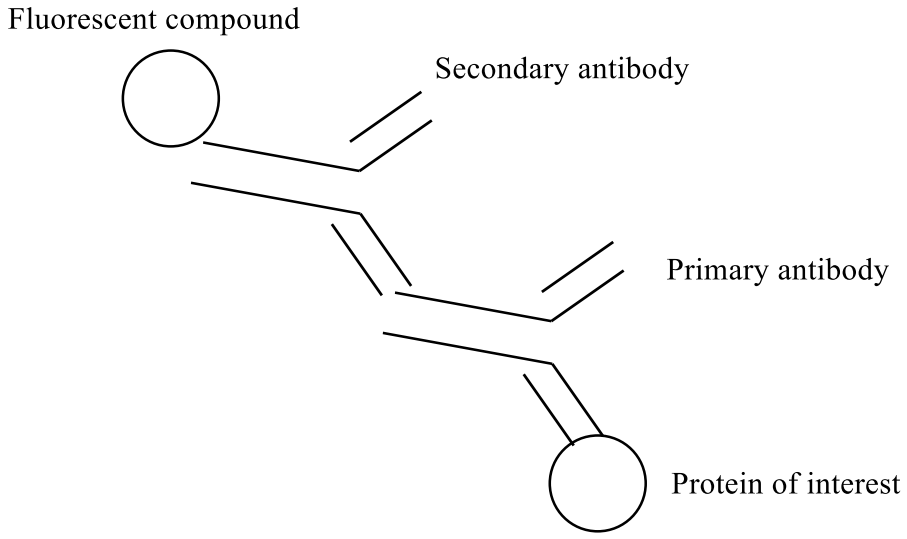
2. How would we separate out D?
3. How would we separate out C?

Answers:

1. A has many negatively charged amino acids, therefore anion exchange chromatography could be used to selectively separate A from the other proteins.
2. D has a polyHis tail and therefore we could use Ni-NTA chromatography to separate D from the other proteins.
3. C has many positively charged amino acids, therefore cation exchange chromatography could be used to separate C.

Chapter 16: Western Blot

We have covered now how to separate proteins on a gel, but how do we know if the protein we want is on the gel? To answer this question, we can do a western blot. A western blot uses the precision of antibodies with the separating power of SDS-PAGE to identify proteins very specifically. To do a western blot, first an SDS-PAGE gel is performed, then the proteins on the gel are transferred onto a membrane using an electric field. This membrane is then covered with a blocking agent (usually milk or something similar) to prevent nonspecific antibodies from binding. This blocked membrane is stained with an antibody that will specifically bind the protein target. This antibody that binds the protein target is referred to as the primary antibody. Then the membrane is stained with a secondary antibody, which will recognize and bind the primary antibody and allow for visualization typically through fluorescence. The secondary antibody is effectively the reporter antibody. A general schematic of a protein-primary antibody-secondary antibody complex is shown below:



These complexes can be analyzed for fluorescence using a fluorimeter and visualized to see if the protein of interest is there. Because antibodies are targeted molecular missiles, this technique allows for very accurate detection of proteins.

Practice Questions:

1. Describe how you can detect protein X starting from a cell lysate. Protein X has a pI of 4, a molecular weight of 450 kDa (very large), and binds ATP. Explain step-by-step how you can separate protein X from other proteins in the cell lysate and detect it.
2. What is a potential complication of using denatured proteins in a western blot analysis?
3. What would happen if the secondary antibody is not added?

Answers:

1. First, we can separate protein X using anion exchange chromatography. If we use a pH 7 buffer during anion exchange chromatography, the protein would be negatively charged because the pH is above the pI. This would allow anion exchange chromatography to separate out protein X. The protein can be eluted by changing the pH of the mobile phase (buffer) from 7 to 2, this would make the protein positively charged and therefore would quickly leave the column (it is no longer electrostatically attracted to the stationary phase). Second, we can separate protein X using size-exclusion chromatography, because it is a very large protein, it will elute very quickly in SEC and therefore will occur in the first few fractions. Third, we can separate protein X using affinity-chromatography. To do this, we would use a stationary phase coated with ATP, this would cause protein X to bind it with high affinity. We can then elute protein X either by introducing more ATP to the buffer solution, changing the pH of the solution, or introducing another ATP-binding protein to compete for stationary phase binding. Once we have a highly purified sample of protein X by using the methods described above, we can run a protein SDS-PAGE gel and do a western blot to identify the protein of interest. Technically we do not need to purify it using those techniques but it was good review. We could alternatively just run the SDS-PAGE gel directly on the cell lysate and then do a western blot to detect protein X using a primary antibody that is specific to protein X and a fluorescent secondary antibody.
2. Antibodies are specific for proteins because of the epitope they bind. If the epitope they bind is not present in the denatured conformation of the protein then they will not bind the target protein and then you are out of luck.
3. If the secondary antibody is not added then you would falsely think that the protein is not present in the sample because there would be no fluorescence.

Chapter 17: DNA Microarrays

Now that we can detect and purify proteins, we can shift our attention to overall gene expression using DNA microarrays. DNA microarrays are used to measure gene expression of several hundred genes and/or to determine differential gene expression between organisms or of the same organism in different stages of life or disease. These DNA microarrays work by putting tiny spots of highly specific synthetic DNA that can base pair with fluorescently labeled cDNA or RNA from a particular organism. These synthetic DNA probes are referred to as reporters and will not fluoresce until the target sequence base pairs with it and in this way gene expression can be determined by fluorescence detection. Because each reporter is highly specific for its RNA target, this technique is both accurate and an incredibly high throughput method of assessing gene expression en masse. An added benefit of fluorescence detection is that the degree of fluorescence also gives quantitative insight into how much that gene is expressed and therefore specific disease states such as cancer can be differentially examined using this method. This is, in fact, how many oncogenes/tumor suppressor genes were identified early on in cancer research.

Practice Questions:

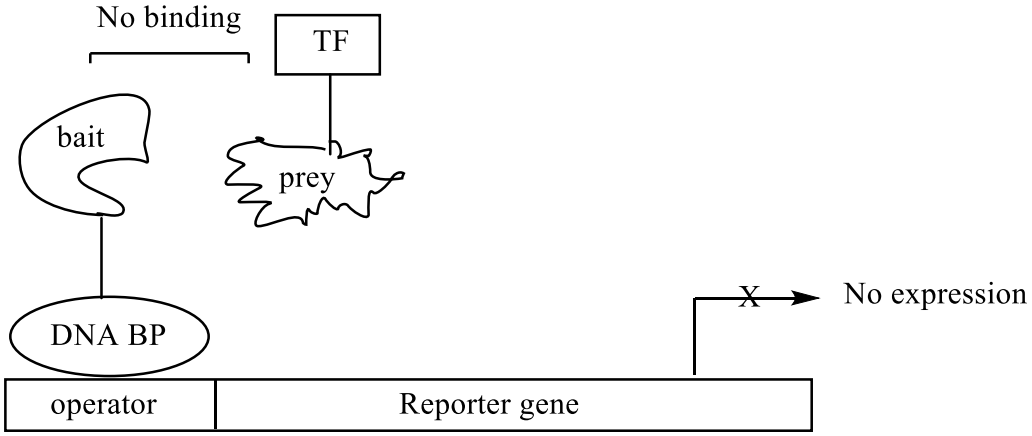
1. Suppose you are at your first day of work in Professor X's cancer research laboratory trying to identify potential oncogenes in Hepatocellular Carcinoma. You isolate mRNA from normal cells and tumor cells then reverse transcribe them to cDNA using reverse transcriptase and you label the normal cell cDNA with a green fluorescent dye and the tumor cell cDNA with a red fluorescent dye. How can you use DNA microarrays to determine differential gene expression?
2. Suppose you do the same thing as the question above, but instead of using two different colored dyes, you use the same green fluorescent dye for the normal cells and tumor cells. How would you have to change the procedure to account for this difference? (Hint: how many DNA microarrays would you have to use here?)

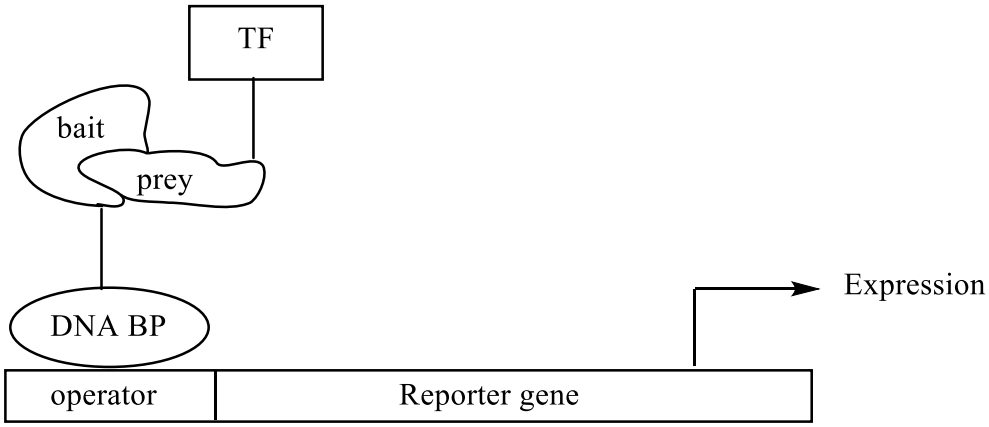
Answer:

1. If normal = green and cancer = red, then the spots that have stronger red fluorescence are the genes that are overexpressed in cancer aka oncogenes.
2. If there is no colorimetric difference between normal and cancer cells, you CANNOT use one microarray. Instead, two microarrays need to be used, one for normal and one for cancer cell cDNA. Then the two microarrays can be measured for fluorescence independently and compared, the spots that have higher fluorescence in the cancer plate relative to the normal plate are the gene products that are overexpressed in cancer aka oncogenes.

Chapter 18: Yeast-Two Hybrid

Sometimes, biologists/biochemists want to determine if two proteins interact with or bind each other. To do this, the Yeast-Two hybrid system can be used. This system has three components, the bait, prey, and reporter gene. The bait is one of the proteins that you want to evaluate and it is covalently attached to a DNA-binding protein that will prevent expression of a reporter gene. This protein can be made with recombinant DNA technology that we discussed previously. The prey is the other protein of interest and it is bound to the transcription factor that will activate expression of the reporter gene. This can also be made via recombinant DNA technology. If the two proteins interact, then the reporter gene will be transcribed and translated into its protein product, which we can detect. Under NO OTHER CIRCUMSTANCES will the reporter gene be expressed, therefore, expression of the protein product MUST indicate that the two proteins (bait and prey) bind each other. A general schematic of the yeast-two hybrid system is shown below:





Practice Questions:

1. Suppose you are a researcher in Dr. Y's laboratory and you are attempting to determine if proteins A-D interact with each other. To do this, you employ the yeast-two hybrid method and link this system to a gene that will express green-fluorescent protein (GFP) that will make the cells glow green when viewed under a fluorescence microscope. You obtain the following results:

Protein Combo	Fluorescence? (Y/N)
A + B	N
A + C	Y
A + D	N
B + C	Y
B + D	Y
C + D	N

   Which proteins interact with each other? Why or why not?

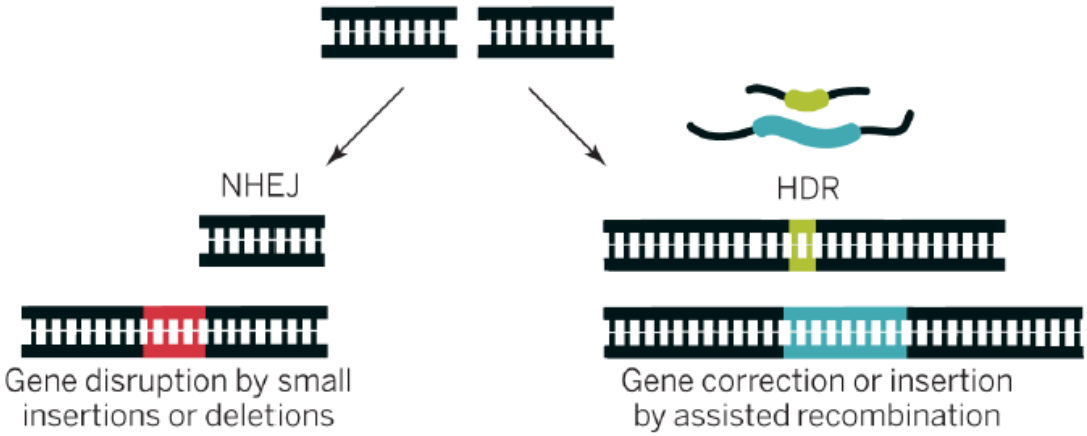
Answers:

1. Only those combos that result in fluorescence have protein pairs that bind each other. Therefore A binds C, B binds C, and B binds D. All other protein combos do not result in fluorescence therefore they do not bind.

Chapter 19: CRISPR/Cas9

This chapter is one of the MOST important chapters in this text because of the implications this technology has for human health. CRISPR/Cas9 is a method of gene editing that is easy to employ and has relatively high fidelity (accuracy). Because of its ease of use and wide variety of applications ranging from agriculture to cancer to sickle cell anemia to malaria, CRISPR/Cas9 has largely supplanted other methods of gene editing such as zinc-finger nucleases (ZFN) and TALENs. Here we will describe the basic principles of CRISPR/Cas9, how it works, and why its two creators won the Nobel Prize in Chemistry for it in 2020.

Before we discuss the exact mechanism of the CRISPR/Cas9 system, we have to discuss what happens when the DNA gets both its strands cut (double stranded break). When this happens, the cell goes into panic mode and there are two pathways that it can undergo: the error-prone nonhomologous end-joining pathway (NHEJ) and the accurate homology-directed repair pathway (HDR) shown below:[12]



The pathway that it undergoes has to do with if there are any homologous (similar) DNA near the double stranded break. If there is not then it undergoes NHEJ and if there is then it goes through HDR. When CRISPR/Cas9 is used, new genetic information can be incorporated at specific sites using the HDR pathway if you also insert DNA that can base pair with the DNA involved with the double stranded break.

The CRISPR/Cas9 system is comprised of three different components: gRNA, Cas9, and a target gene containing a PAM sequence. The guide RNA or gRNA is a strand of RNA that can base pair with the target gene and direct the Cas9 scissor protein to the target site. The target gene has to have a PAM sequence, which is a short 3 nucleotide sequence that is recognized by Cas9, the molecular scissor. This PAM sequence varies depending on the Cas9 that is used, however, the most commonly used Cas9 is from *Streptococcus pyogenes* and recognizes the PAM sequence NGG, which occurs every 8 nucleotide bases. N stands for aNy nucleotide (A,G,T, or C). Once the gRNA binds the target gene and the Cas9 recognizes the PAM sequence, Cas9 will cut the DNA at that exact spot, removing the target gene. Once the target gene is removed by the Cas9

scissor, either NHEJ or HDR pathways ensue and the gene is knocked out (NHEJ) OR replaced with a new gene (HDR). The HDR pathway can be promoted by using Cas9 nickases, which are mutated versions of the Cas9 protein that has one of its two catalytic sites deactivated. Two Cas9 nickases used together can create jagged or "sticky" ends at the cut site and thereby promote HDR.

CRISPR/Cas9 is an incredibly powerful tool because it can be used in prokaryotes, eukaryotes, mice, mosquitoes, and humans. One of the most promising applications of CRISPR/Cas9 is its ability to introduce functional genes and therefore correct genetic illnesses such as sickle cell anemia. Sickle cell anemia is a genetic disease that is the result of a single amino acid substitution in hemoglobin. This mutation causes the amino acid at position 6 to change from glutamic acid to valine. This causes hemoglobin to misfold and not be an efficient oxygen carrier. This misfolded hemoglobin results in red blood cells appearing sickle-like instead of perfectly round, this causes tremendous pain and prevents efficient oxygen delivery to the person's tissues. Because of the well-defined nature of this genetic disorder, scientists know exactly where to target in the genome to CURE this disease and can use CRISPR/Cas9 to replace the GTG codon (valine) for a GAG codon (glutamic acid) and effectively convert abnormal cells into healthy cells. At the time of this writing, there is an ongoing clinical trial using CRISPR/Cas9 to help sickle cell anemia patients and a related disease β-thalassaemia. CRISPR/Cas9 can also be used to determine what genes are the drivers of cancer growth. By removing certain genes in mice and observing the development of tumors, experimenters can determine the effects those genes have on cancer growth. And finally, CRISPR/Cas9 can be used to make mosquitoes, the vector that gives humans the *Plasmodium* parasite that causes malaria, either partially immune to *Plasmodium* infection or infertile causing population decline of the vector and therefore reduced disease spread. Because of the wide variety of applications, CRISPR/Cas9 has become a staple within the biological scientific community.

Practice Questions:

1. Suppose you are a researcher in professor Z's laboratory and you are attempting to screen for driver genes (genes that when abnormally expressed cause cancer growth), when you knockout genes A-D in mice exposed to CCl_4, a carcinogen, you observe the following results:

Gene knocked out	Tumor growth increased or decreased
A	Increased
B	Decreased
C	Increased
D	Increased

   Which genes are tumor-suppressor genes (genes that prevent tumor growth), which genes are oncogenes (genes that promote tumor growth)?

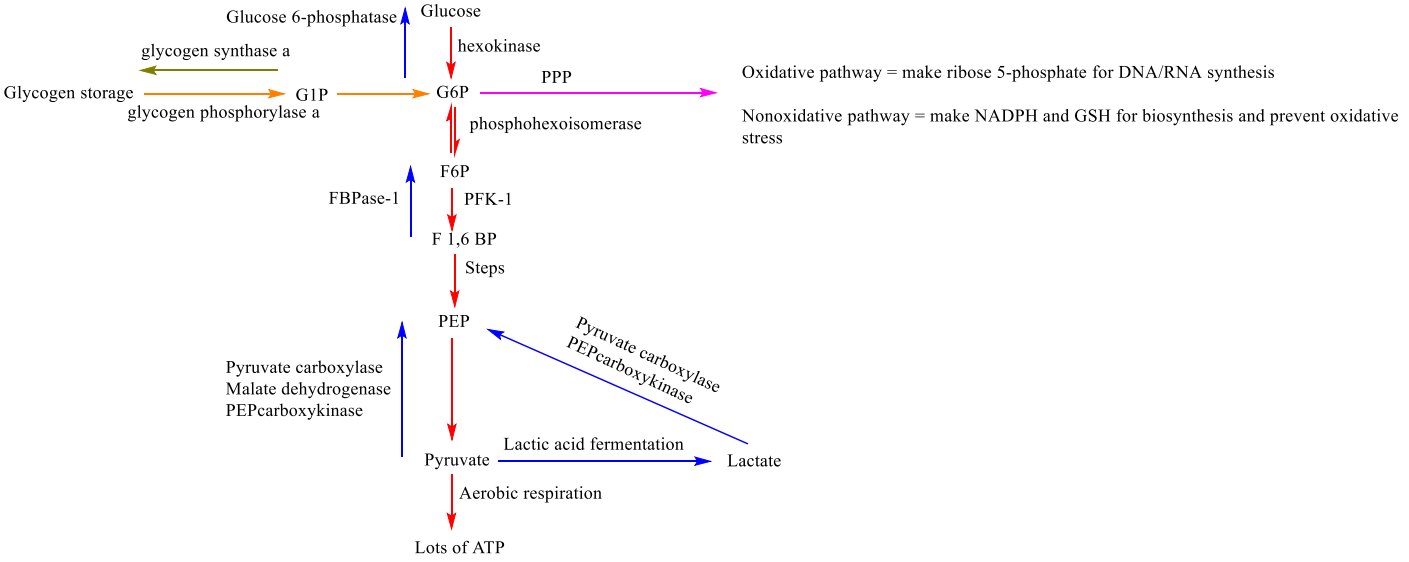
2. What are some safety concerns with CRISPR/Cas9?

Answers:

1. Knocking out (removing) genes that prevent tumor growth will cause the cancer to grow faster, therefore A, C, and D are all tumor-suppressor genes while B is an oncogene.
2. CRISPR/Cas9 may have off-target effects if the gRNA is not specific enough to the target site. Because of this gRNA design is critical and must be optimized to minimize off-target effects before use in humans.

Focus 5: Carbohydrate Metabolism

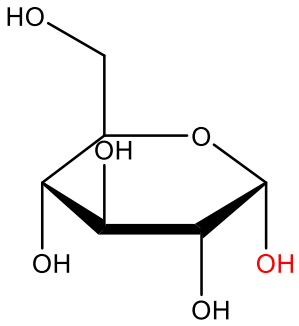
Red = Glycolysis and Aerobic Respiration
Blue = Gluconeogenesis and Anaerobic Respiration
Pink = Pentose Phosphate Pathway
Orange = Glycogen Breakdown
Green-Brown = Glycogen synthesis



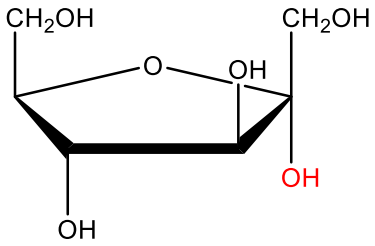
Chapter 20: Carbohydrate Chemistry Basics

Now we get into what I call “real biochemistry”, we have covered the basics so now we can start looking at metabolic pathways and understand the chemical logic and enzymatic mechanisms that are used to make life as we know it happen. Our number one preferred energy source is carbohydrates. In this chapter, we will discuss the basics of carbohydrate chemistry as well as some vocabulary regarding their structure. Carbohydrates are effectively sugar molecules, think glucose, fructose, sucrose, and lactose for some examples of these. Sugar molecules are also called saccharides, monosaccharides are those sugars that only involve one sugar ring, while disaccharides and polysaccharides are sugar molecules that are composed of two or more sugar rings covalently connected together via glycosidic bonds. Glucose and fructose are examples of monosaccharides while sucrose and lactose are examples of disaccharides. The structure of these carbohydrates are shown below:

α D-Glucose

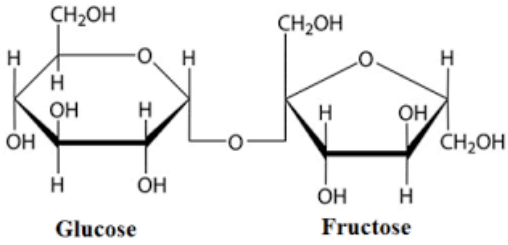


α D-Fructose

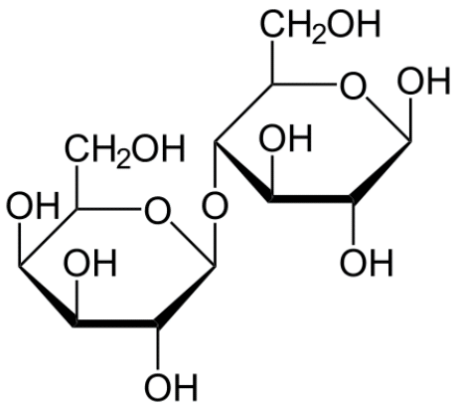


These are the alpha forms because the OH groups on the anomeric carbons are pointing down (labeled red)

Sucrose

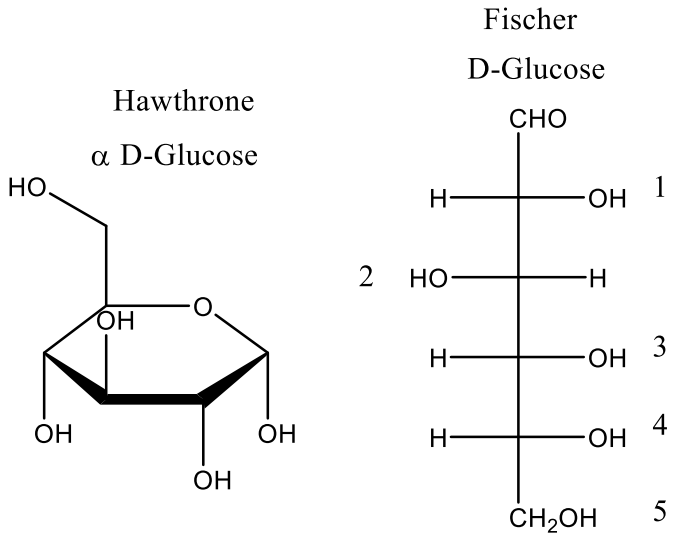


Lactose

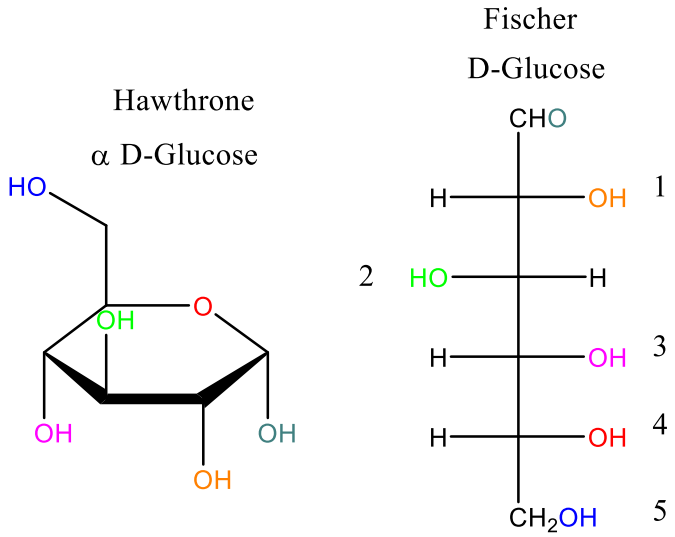


Sugars, as we mentioned before when we discussed stereochemistry exist in both cyclic and linear forms. Knowing how to convert the linear into the cyclic and the cyclic into the linear is therefore very important to understanding their chemistry. Like we said before, the cyclic forms occur when a hydroxyl group attacks either an aldehyde or ketone in the linear form of the carbohydrate. When the carbonyl is a ketone, the sugar is referred to as a ketose and when the carbonyl is an aldehyde, the sugar is referred to as an aldose. The carbons that get attacked by the OH groups are always referred to as the anomeric carbons. The OH group that is attached to the anomeric carbons can be oriented either up or down in the cyclic form. If the OH group is

oriented up then it is considered a β carbohydrate and if it is facing down then it is considered an α carbohydrate. Now that we have the basic vocabulary down for monosaccharides, we can discuss how to convert between the Fischer Projection (Linear) to the Hawthorne projection (cyclic). To do this, we will look at an example and see what conclusions we can draw. Consider α D-glucose linear and cyclic forms shown below:



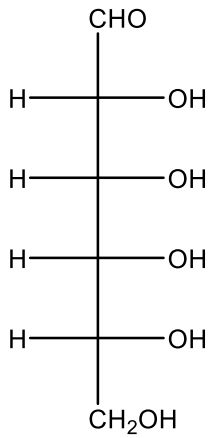
As we said before, we have to attack the carbonyl with one of the OH groups to make the cyclic form. Which OH must be used to attack the aldehyde group on the top for us to make a 6-membered ring as shown on the Hawthorne projection? Because we want a 6-membered ring to form, we need to count from the OH group to the carbonyl to see how large the ring would be. For us to count 6 total atoms, OH group number 4 has to be used. Because carbonyl groups are planar, the OH group can attack from either the top or the bottom face, causing the formation of the α and β forms of glucose respectively. I will label all OH groups in different colors in both forms so that you can see where exactly they all are:

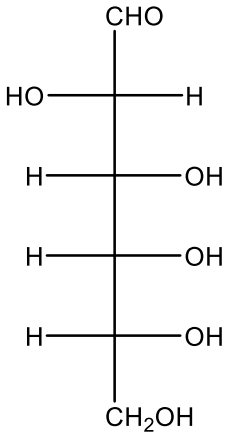


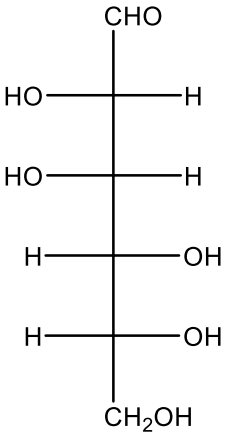
What do we notice about the OH groups that are pointing right in the Fischer versus the OH groups that are facing down in the Hawthorne? That's right, they are the same ones! Therefore, in general, the OH groups that point right in the Fischer point down in the Hawthorne. Likewise, the OH groups that point to the left in the Fischer point up in the Hawthorne. Now we can effectively convert between Fischer and Hawthorne projections with ease.

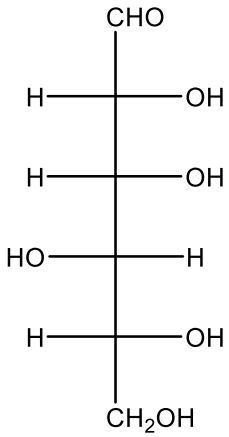
Practice Questions:

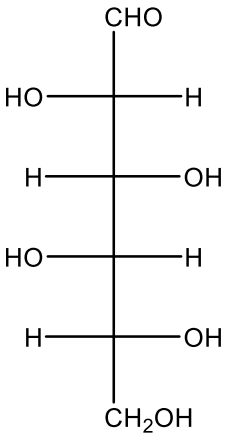
1. Draw the Hawthorne projection for the following carbohydrates (assume alpha 6-membered ring forms):

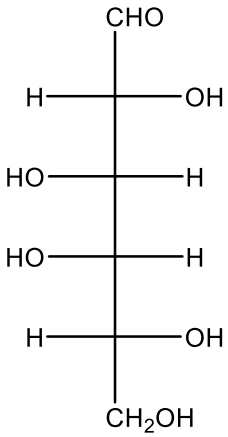




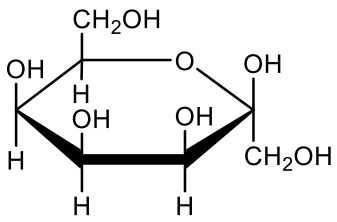
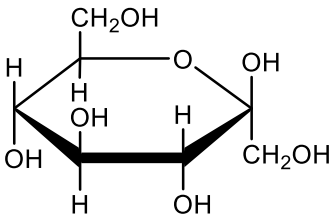
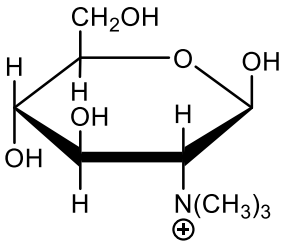
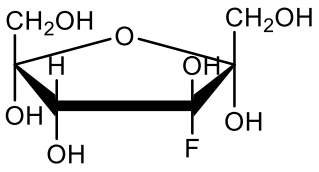
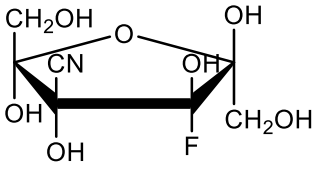
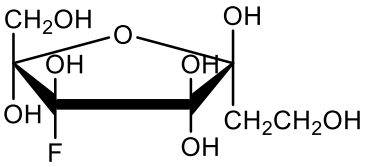






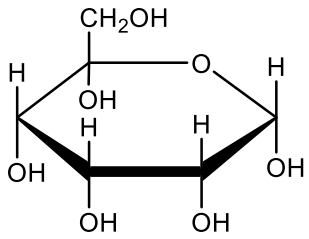
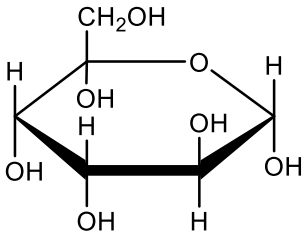
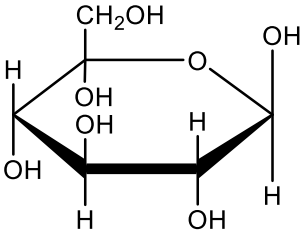
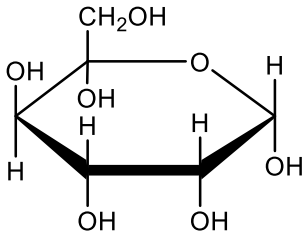
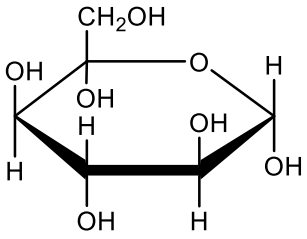
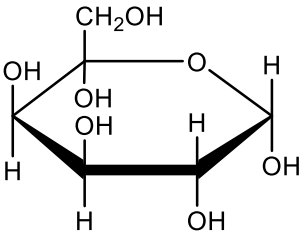


2. Draw the Fischer projections for the following carbohydrates:

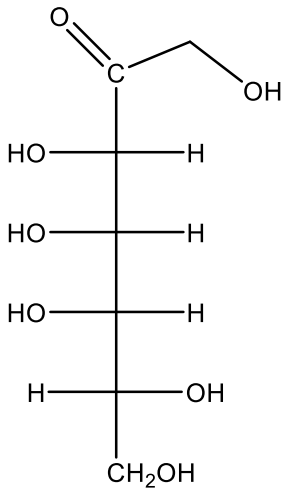







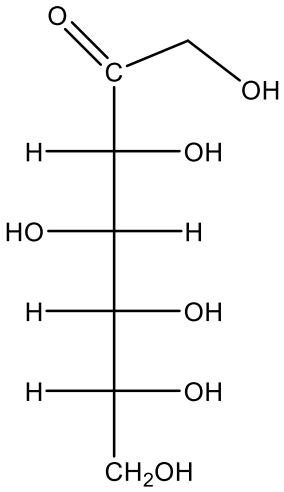
Answers:

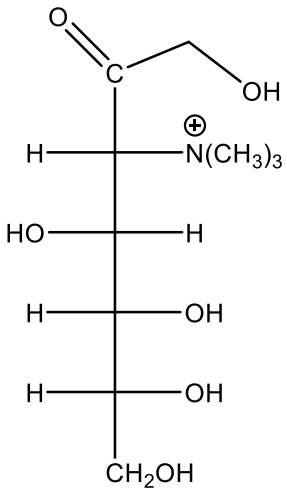
1.

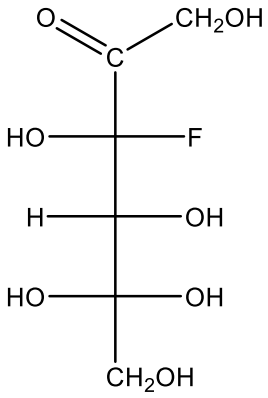







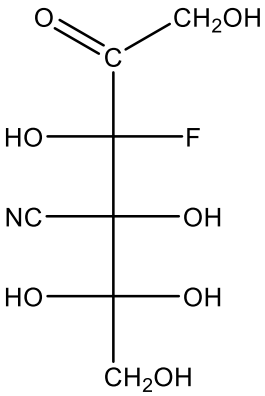
2.

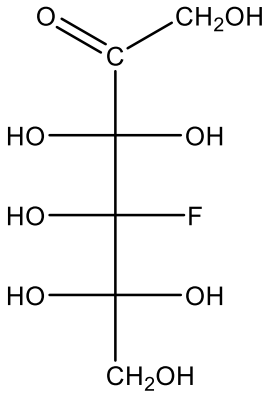












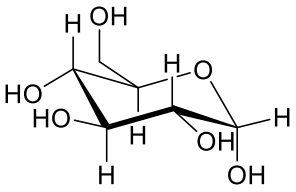
Monosaccharides can exist in the 6-membered ring form (pyranose) or 5-membered ring form (furanose). These two forms refer to the simple organic molecules pyridine and furan shown below:

Pyridine

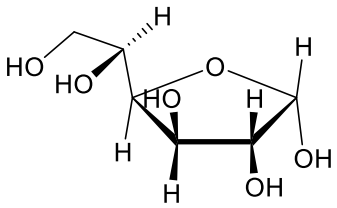
Furan

Shown below are the pyranose and furanose forms of D-glucose:

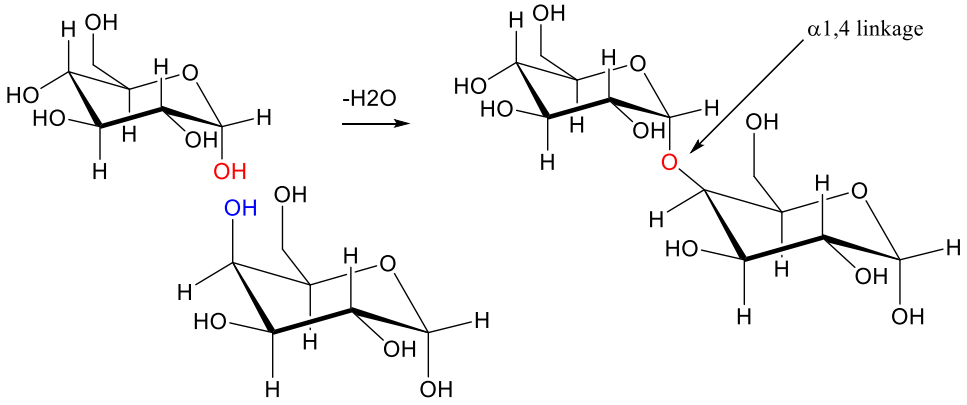
Pyranose

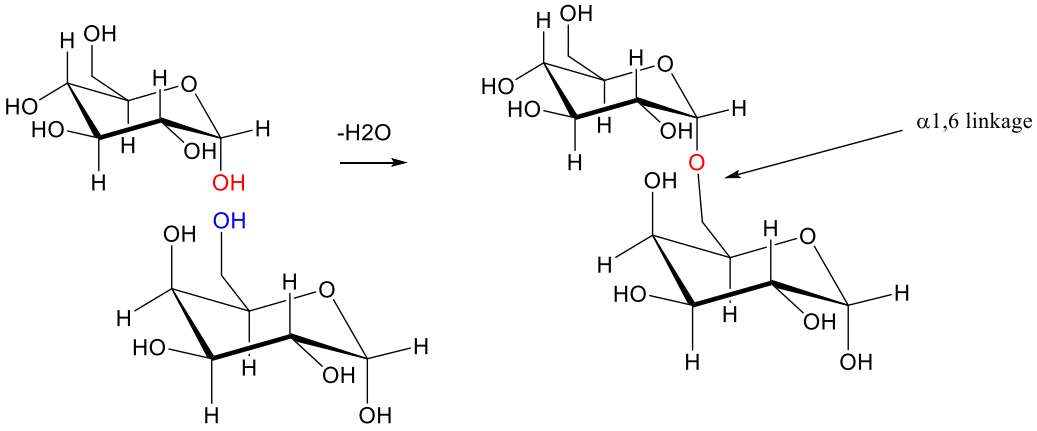


Furanose

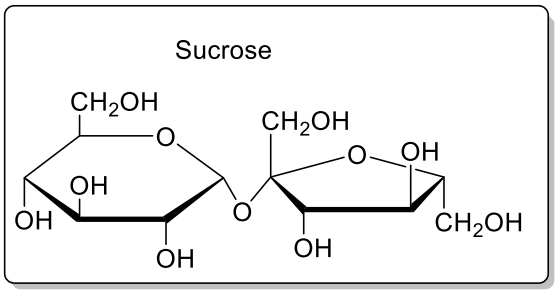


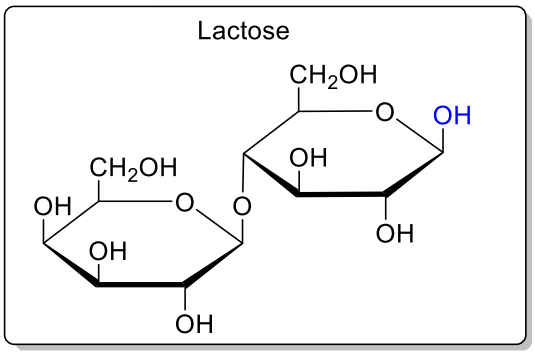
Now that we covered the basics of monosaccharides, we can discuss some of the terminology used to describe disaccharides and polysaccharides as a whole. When monosaccharides are bonded together covalently, they form glycosidic linkages. The way that these two monosaccharides bond to each other determines the type of glycosidic linkage that they form. If the glycosidic linkage forms between the alpha anomeric carbon of one monosaccharide and the OH group on the four position (counting clockwise from the anomeric carbon) of another alpha monosaccharide, that glycosidic linkage is an α1,4 linkage. Conversely, if the alpha anomeric OH of one monosaccharide and the OH group on the six position of another alpha monosaccharide form a glycosidic bond, that is termed an α1,6 linkage. Examples of these are shown below:





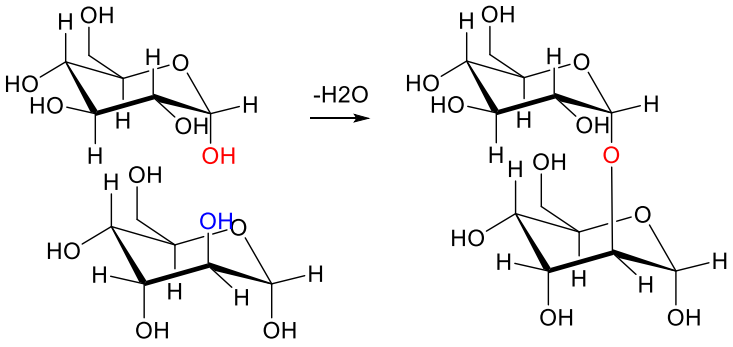
Polysaccharides can be described as either reducing or nonreducing sugars. If the polysaccharide is reducing, then its anomeric carbon can be oxidized to a carbonyl and therefore there MUST be an OH group on that carbon. If the anomeric carbons of all the sugars in the compound are involved with glycosidic bonds then that sugar does not have any free anomeric OH groups and therefore that sugar is said to be nonreducing. Sucrose is an example of a nonreducing sugar while lactose is an example of a reducing sugar. The important OH group in lactose that makes it a reducing sugar is highlighted in blue in the diagram below:

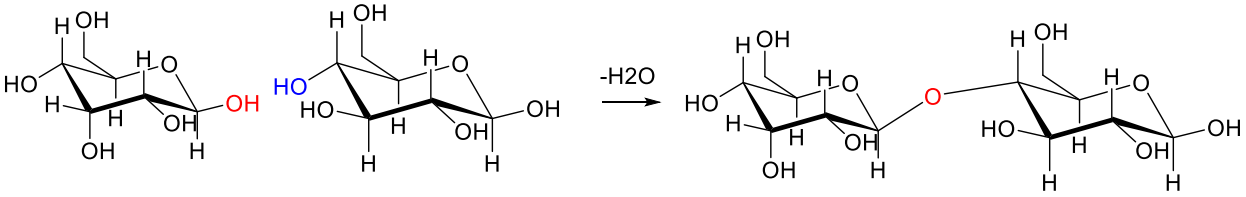


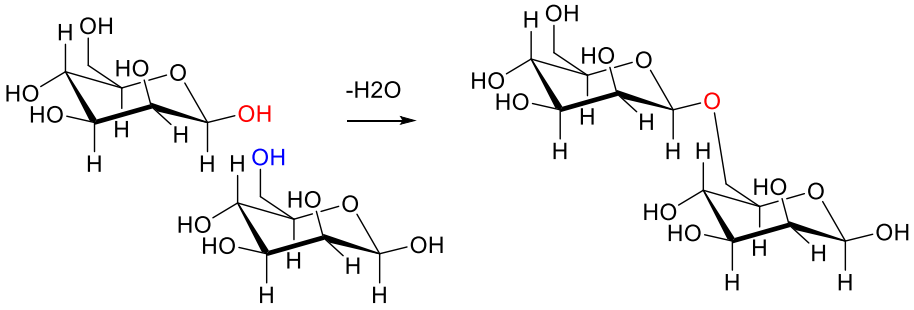


Practice Questions:

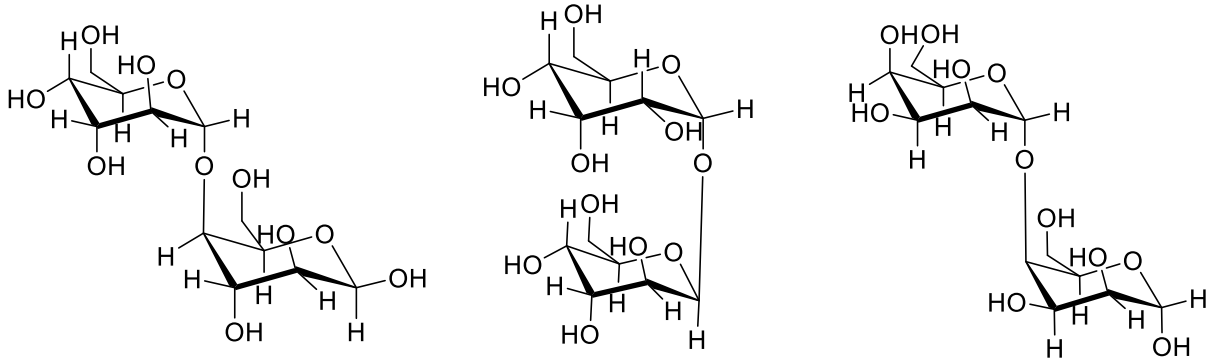
1. Identify the types of glycosidic linkages in the following polysaccharides:







2. Identify the reducing sugars in the following list of polysaccharides:

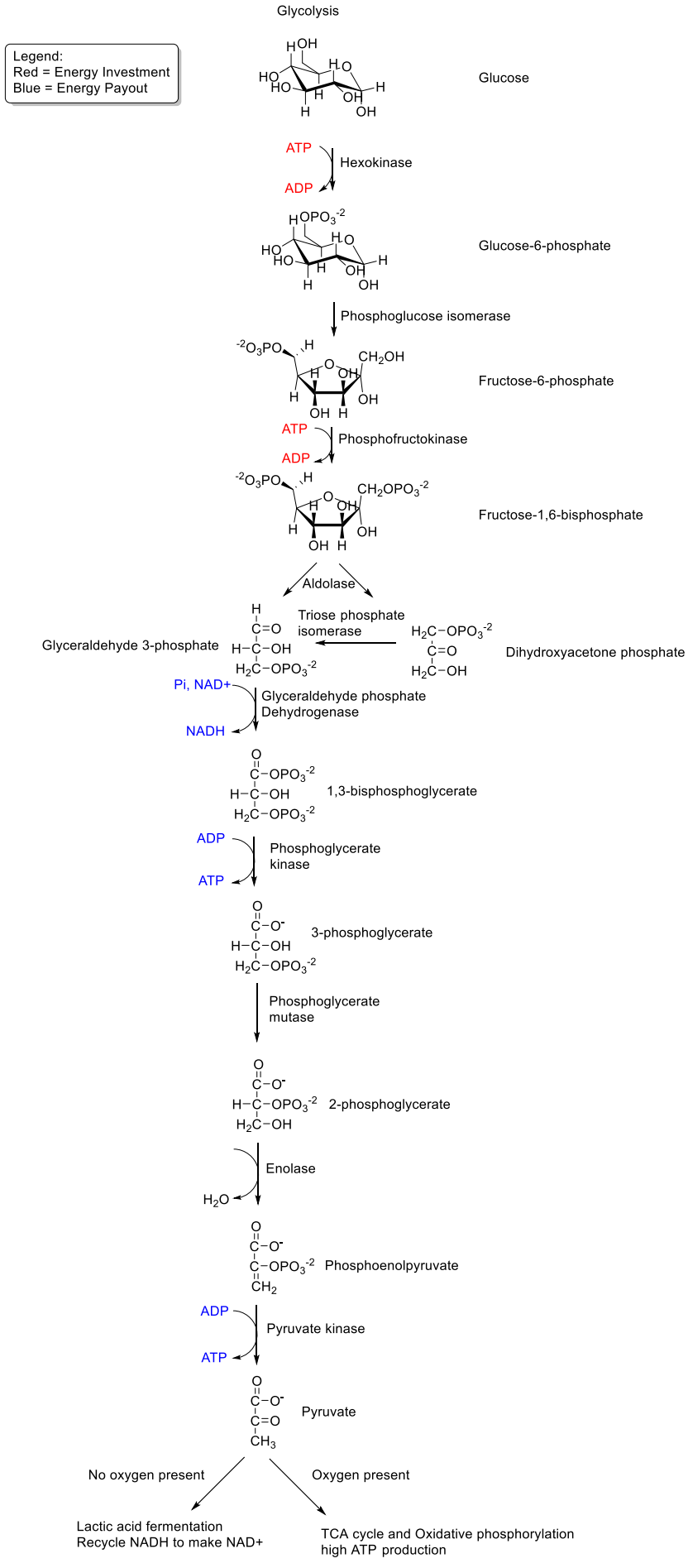


Answers:

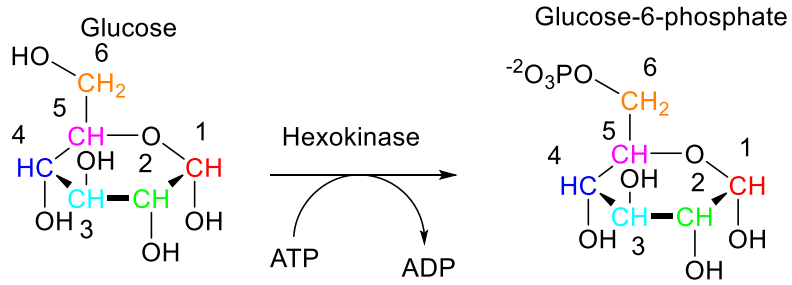
1. Alpha 1,2; beta 1,5; beta 1,6
2. The first and third ones are reducing sugars because they have a free anomeric hydroxyl group

Chapter 21: Glycolysis

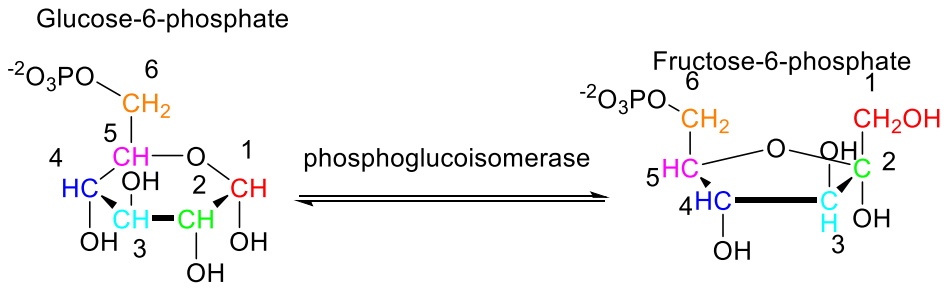
Glycolysis is an extremely old pathway, every organism on the planet does glycolysis and this suggests that we all derived from some common ancestor in which this pathway first evolved. Glycolysis literally means “cutting sugar” and that is effectively what we are doing. Glycolysis is the process where glucose, the main carbohydrate energy source of the body, is broken down to pyruvate and the body gains a net total of 2 ATP and 2 NADH. An overview of glycolysis is shown below and we will go step-by-step and explain all of the reactions that occur as well as the major regulatory enzymes present in this pathway:


Glycolysis
Legend:
Red = Energy Investment
Blue = Energy Payout
Glucose
ATP
Hexokinase
ADP
Glucose-6-phosphate
Phosphoglucose isomerase
Fructose-6-phosphate
ATP
Phosphofructokinase
ADP
Fructose-1,6-bisphosphate
Aldolase
Triose phosphate isomerase
Glyceraldehyde 3-phosphate
Dihydroxyacetone phosphate
Pi, NAD+
Glyceraldehyde phosphate Dehydrogenase
NADH
1,3-bisphosphoglycerate
ADP
Phosphoglycerate kinase
ATP
3-phosphoglycerate
Phosphoglycerate mutase
2-phosphoglycerate
Enolase
H2O
Phosphoenolpyruvate
ADP
Pyruvate kinase
ATP
Pyruvate
No oxygen present
Oxygen present
Lactic acid fermentation
Recycle NADH to make NAD+
TCA cycle and Oxidative phosphorylation
high ATP production

The very first step of glycolysis is the conversion of glucose to glucose-6-phosphate. However, this reaction is energetically unfavorable and therefore we need to use ATP to power this reaction to go forward (MA**S**C). Because we are hydrolyzing ATP in this reaction, hexokinase must have a Mg^{2+} cofactor. This phosphate group that we add to glucose is important because it will activate the carbon that receives it and allows it to ultimately power ATP synthesis later on in the pathway. In this way, we spend ATP to make ATP, in a sort of ATP stock market investment. Because this enzyme uses ATP and ATP is effectively the energy currency of the cell, this step is a major regulatory step and as such hexokinase is allosterically regulated by many things: its product, glucose-6-phosphate inhibits the enzyme while glucose activates it. This should make sense because this reaction uses energy, so if we already have a lot of glucose-6-phosphate then why should we produce more (MAS**C**)? Likewise, if we have a lot of glucose then we should be using it to make energy (MAS**C**). This reaction not only chemically activates glucose, but it also traps it in the cell. Glucose-6-phosphate is negatively charged and therefore cannot cross the cell membrane to go back into the extracellular space. The hexokinase reaction is shown below:

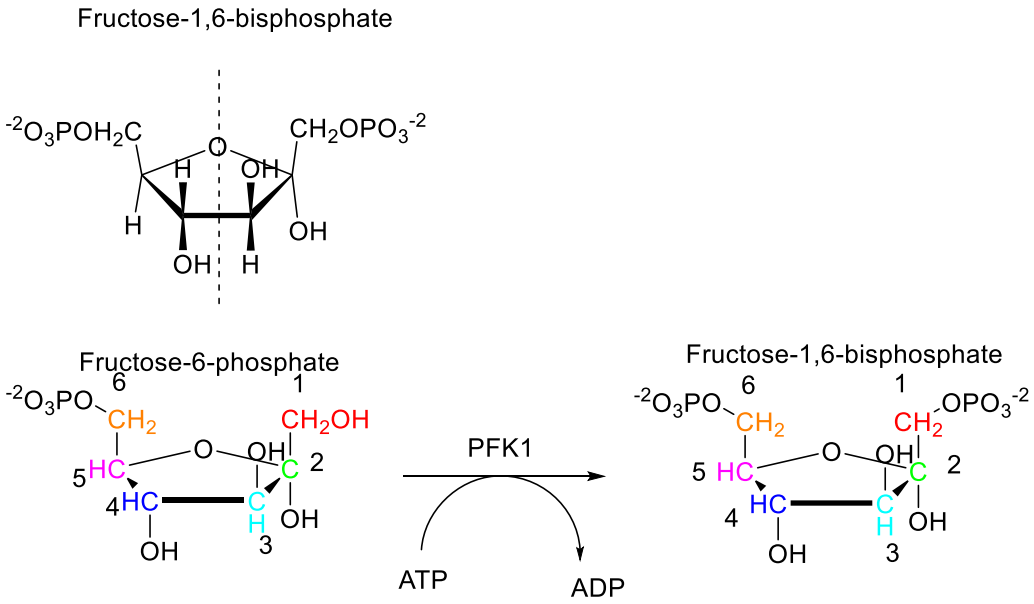


The next step in glycolysis is isomerizing glucose-6-phosphate to fructose-6-phosphate by phosphoglucoisomerase. This reaction is simply isomerization and therefore it is completely reversible. This reaction is important, however, because fructose is symmetrical down the center while glucose is not. This symmetry allows the pathway to minimize the amount of steps it needs to use as we will soon see. A general theme for all metabolic pathways is that symmetry is your friend because it allows you to minimize the chemistry and maximize the energy output. The phosphoglucoisomerase reaction is shown below:

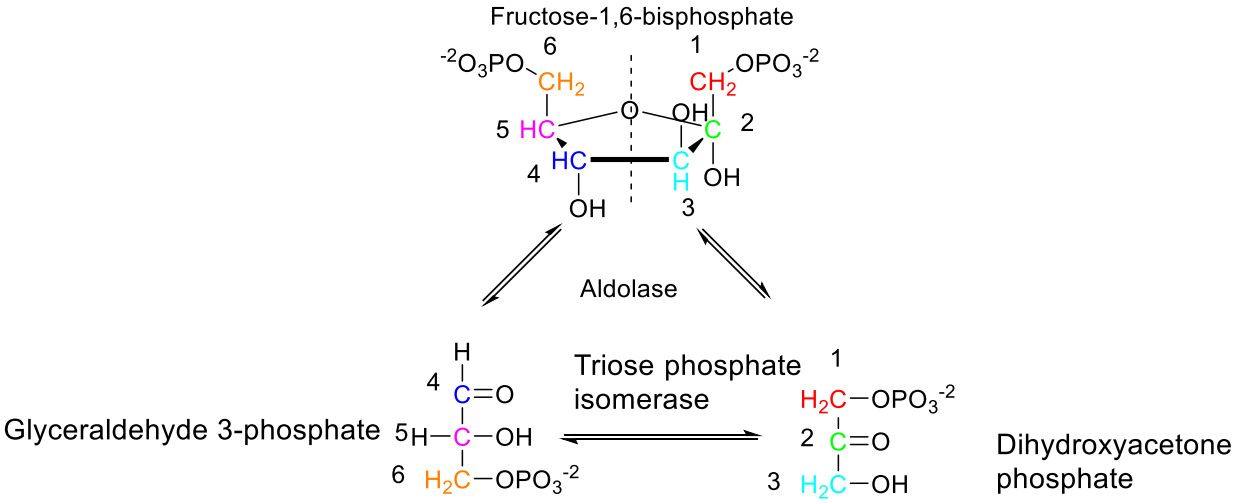


The third step in glycolysis is the second investment step, which is catalyzed by phosphofructokinase-1 (PFK1). This reaction converts fructose-6-phosphate to fructose 1,6-bisphosphate, which is a symmetrical molecule and therefore minimizes the chemistry after this

step. This reaction uses ATP hydrolysis just like hexokinase to power the reaction (M**A**SC) and therefore it also uses a Mg^{2+} cofactor. Just like hexokinase, this is also a highly regulated enzyme because it uses ATP and is irreversible. PFK1 is regulated allosterically by ATP (inhibits), PEP (inhibits), citrate (inhibits), AMP (activates), and fructose 2,6-bisphosphate (activates). These regulators make sense when one considers the overall purpose of glycolysis, to make ATP! Obviously if we have a lot of ATP then we don't need to go through glycolysis and therefore ATP should inhibit PFK1 (MAS**C**). Likewise, if we have a lot of AMP that indicates that the cell is energy-starved and therefore glycolysis is needed to replenish the ATP supply (MAS**C**). PEP and citrate are both downstream products, PEP is in glycolysis and citrate, as we will learn later on, is part of the Krebs cycle. Both of these molecules are responsible for creating energy and therefore if we have a lot of them, we do not need to use PFK1 (MAS**C**). Fructose 2,6-bisphosphate is a molecule that is made by PFK2 and it happens when fructose-6-phosphate levels are high. If fructose-6-phosphate levels are high that means we should be using it up; fructose-6-phosphate by itself is not useful for anything and therefore we should promote PFK1 activity by creating this fructose 2,6-bisphosphate molecule. The symmetry of fructose 1,6-bisphosphate and the reaction of PFK1 are shown below:

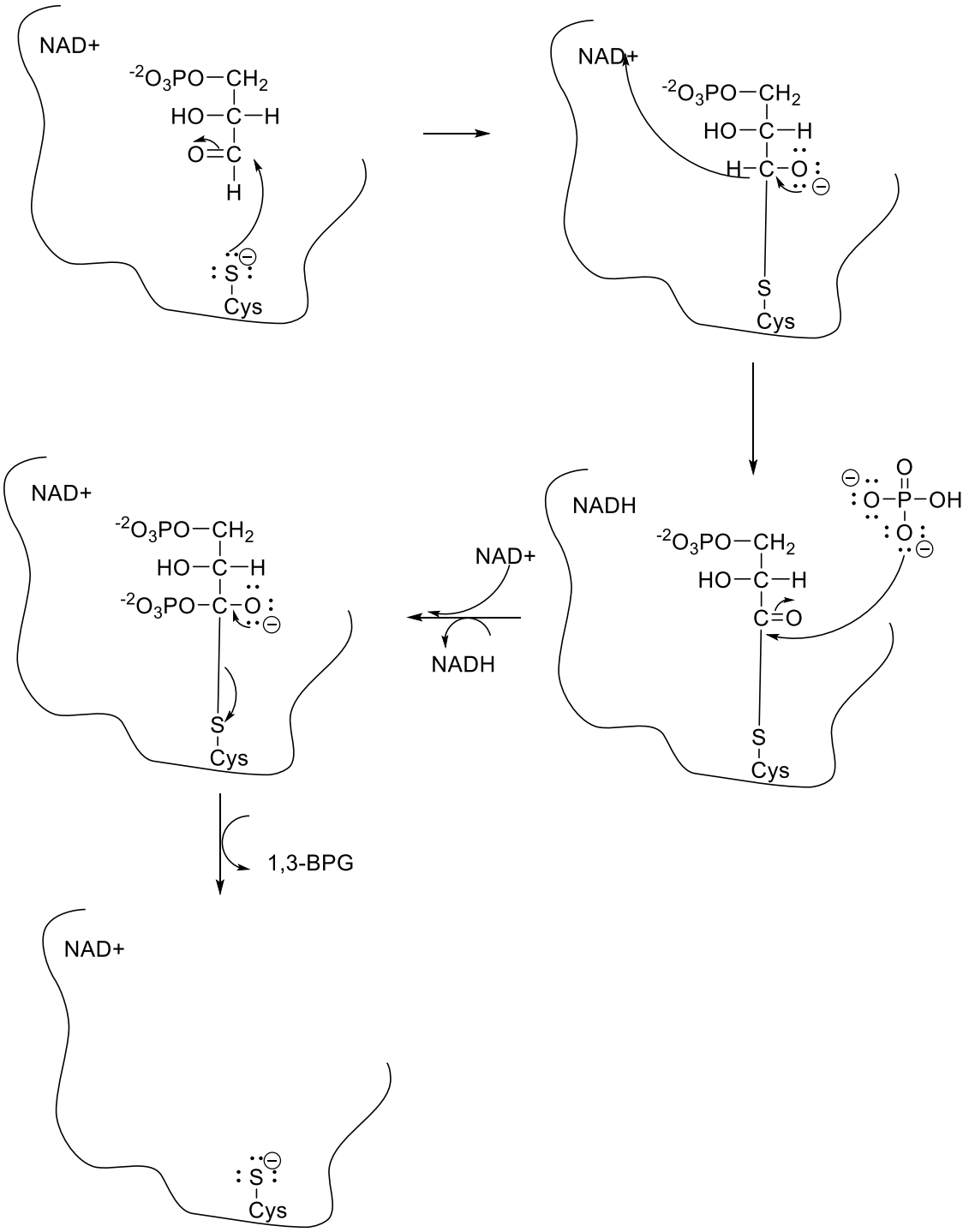


Because of this symmetry, the next step involving aldolase allows the cell to create two 3-carbon compounds that are structurally very similar. This reaction takes fructose-1,6-bisphosphate and creates glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP). The parts of fructose 1,6-bisphosphate that create both trioses are shown below in color to help distinguish them. This reaction is reversible. The overall reaction of aldolase and triose phosphate isomerase is shown below:



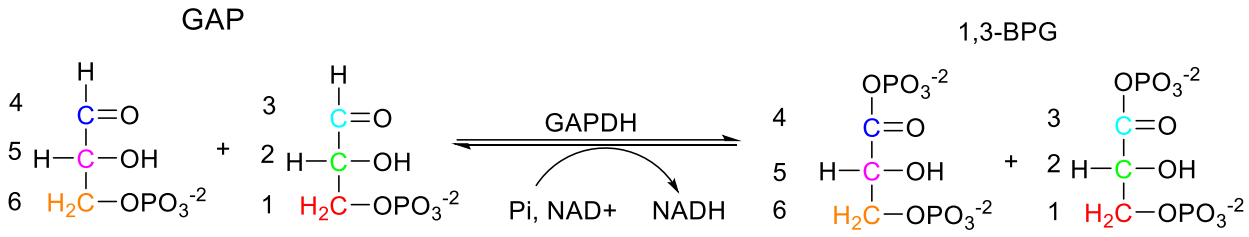
The dihydroxyacetone phosphate (DHAP) is converted to GAP because metabolic processes are more efficient the less steps and less branching you have. By converting DHAP to GAP, the cell is maximizing the efficiency of glycolysis because without this step, the cell would have to create an entirely separate pathway to metabolize DHAP, which not efficient (MAS**C**).

Once all of the carbons are converted to GAP, we can start creating energy from this process. The first energy producing step in glycolysis is the GAPDH step. This uses inorganic phosphate (PO_4^{3-} or P_i) and NAD^+ to oxidize glyceraldehyde 3-phosphate to 1,3-bisphosphoglycerate (1,3BPG). This reaction mechanism is important to know. It goes through a high energy thioester intermediate to drive the reaction forward. The mechanism is shown below:

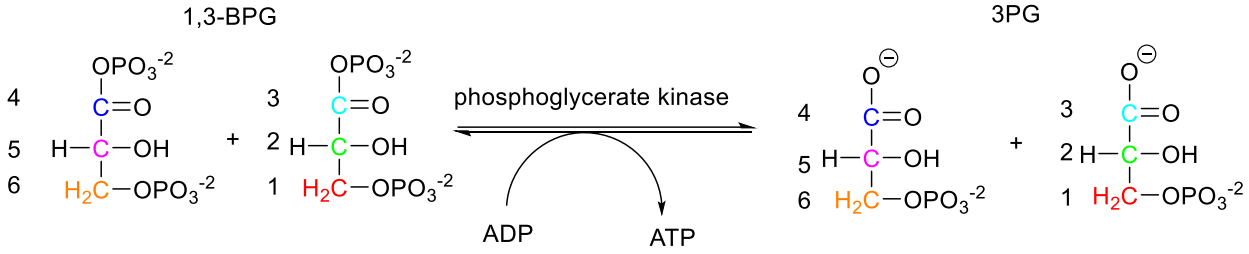


Just like in organic chemistry, we will go through this reaction mechanism step by step because it is an important mechanism to know. First, the cysteine uses its nucleophilic thiolate (S^-) to attack the carbonyl carbon of GAP. Recall that anything with a negative charge is nucleophilic and recall that anything with a polar pi bond is electrophilic so this is simply the nucleophile attacks electrophile pattern that you should be all too familiar with from orgo. Next, the tetrahedral intermediate collapses back to the carbonyl and the C-H electrons are being donated

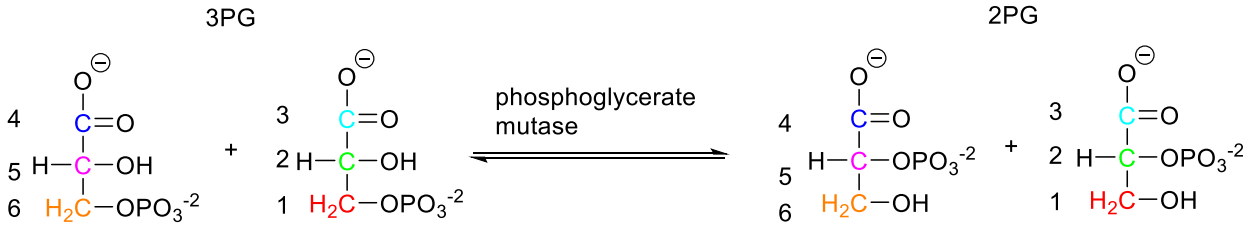
to the NAD^+. This is the oxidation step, recall that oxidation is the loss of electrons or the loss of hydrogen, which is exactly what happened here. The resulting carbonyl is a thioester and this thioester has a great leaving group (S). Recall that the weaker base is the better leaving group, because S^- is very stable, it will happily leave and get replaced by the phosphate when it nucleophilically attacks the thioester. The NADH that gets produced in this reaction will eventually be used in the electron transport chain to yield 2.5 ATP. The overall reaction of GAPDH is shown below:



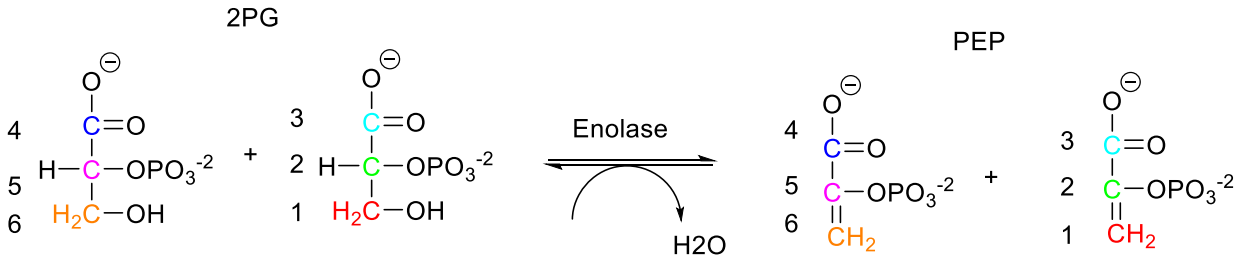
Once this 1,3-BPG is made, it can be used to phosphorylate ADP to make ATP and 3-phosphoglycerate (3PG). This reaction is catalyzed by phosphoglycerate kinase and is reversible due to low 1,3-BPG concentrations. Kinases are enzymes that will transfer phosphate groups from one molecule to another. The 1,3-BPG molecule is extremely high energy because the negative charges of the two phosphate groups repel each other and the carbonyl oxygen is also partially negative, leading to more electrostatic repulsion in the molecule. The overall reaction of phosphoglycerate kinase is shown below:



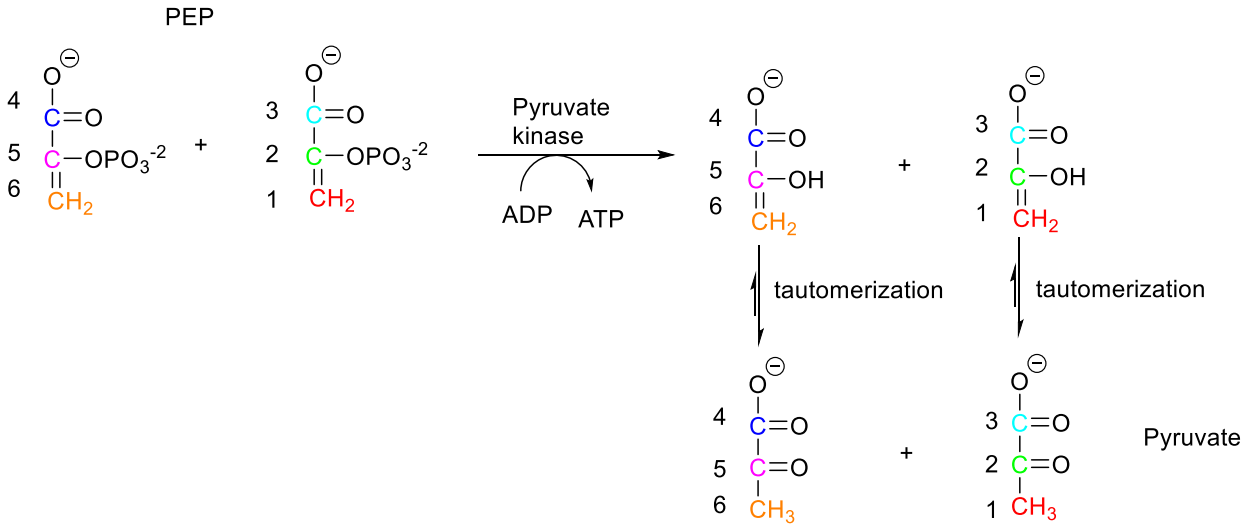
Once 3PG is made, it is converted to 2PG to allow for more electrostatic repulsion that will drive the formation of ATP in the last step of glycolysis. This reaction is catalyzed by phosphoglycerate mutase. Mutases always catalyze the apparent transfer of one functional group on one carbon to another carbon. In this case it is moving the phosphate group from the 3 carbon to the 2 carbon. This is a reversible reaction as well, as such it is not regulated. The overall reaction of phosphoglycerate mutase is shown below:



2PG is definitely high energy, but to make it more high energy, enolase is used to add a double bond between the 2 and 3 carbons, this makes PEP. This reaction is unfavorable and reversible, it is pulled forward only because PEP is kept at a low concentration. The double bond helps increase the energy of the molecule because when the phosphate is hydrolyzed off, it will result in an enol that will rapidly tautomerize to the ketone and as we know from Organic Chemistry, C=O bonds are incredibly stable and favorable. The overall reaction of enolase is shown below:



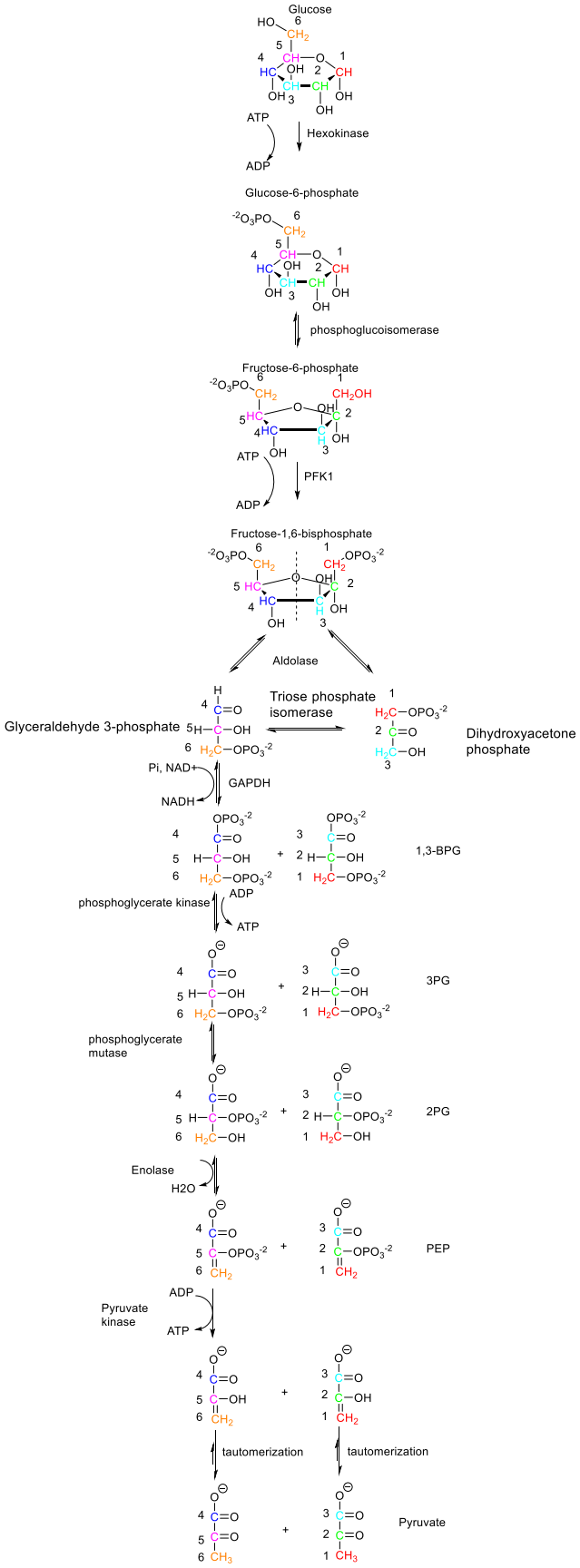
Finally, the very last step of glycolysis is the pyruvate kinase reaction. This converts PEP to pyruvate and makes ATP in the process. This is an irreversible reaction and is driven forward by the incredibly stable pyruvate molecule it makes and the minimization of electrostatic repulsion. As the last irreversible reaction, it is subject to a wide variety of regulation. Pyruvate kinase is activated by fructose 1,6-bisphosphate and is inhibited by ATP, acetyl-CoA, and fatty acids. This should make sense because again, pyruvate kinase's purpose is to make ATP, therefore if we already have a lot of ATP there is no reason to make more of it. Acetyl-CoA is the downstream product after pyruvate goes through the PDH complex (details coming soon) and therefore this is a form of feedback inhibition. A lot of fatty acids means that the cell already has a lot of energy since making those fats required a lot of energy in the first place. Conversely, if we have a lot of fructose 1,6-bisphosphate then that means the cell is actively going through glycolysis and therefore the overall pathway needs to move faster. Because pyruvate kinase is an irreversible step of glycolysis, it needs to be activated under these conditions (MAS**C**). The overall reaction of pyruvate kinase is shown below:



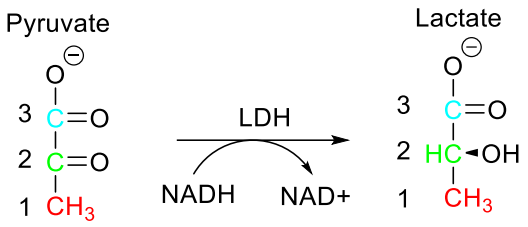
A summary of the regulation of glycolysis is shown in the table below:

Enzyme	Inhibited	Activated
Hexokinase	Glucose-6-phosphate	Glucose
PFK1	ATP, PEP, Citrate	AMP, Fructose 2,6-bisphosphate
Pyruvate kinase	ATP, Acetyl-CoA, fatty acids	Fructose 1,6-bisphosphate

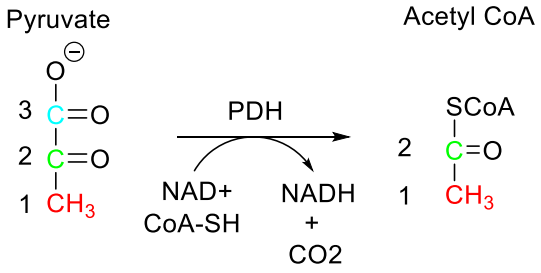
The carbons in glucose are labeled below at each step of glycolysis for your reference. This will help you if you are asked about radiolabeling:


Glucose
ATP
Hexokinase
ADP
Glucose-6-phosphate
phosphoglucoisomerase
Fructose-6-phosphate
ATP
PFK1
ADP
Fructose-1,6-bisphosphate
Aldolase
Triose phosphate isomerase
Glyceraldehyde 3-phosphate
Dihydroxyacetone phosphate
Pi, NAD+
GAPDH
NADH
1,3-BPG
phosphoglycerate kinase
ADP
ATP
3PG
phosphoglycerate mutase
2PG
Enolase
H2O
PEP
ADP
Pyruvate kinase
ATP
tautomerization
tautomerization
Pyruvate

Once pyruvate is produced, it has one of two possible fates. If oxygen is not present, then it will be converted to lactate using lactate dehydrogenase in the reaction shown below:



If oxygen is present, then it will go into the mitochondria and be made into acetyl-CoA using pyruvate dehydrogenase complex in the reaction shown below.



Pyruvate dehydrogenase complex is a HUGE enzyme complex that is comprised of three enzymes (E_1, E_2, and E_3) each of which catalyzes their own reaction, the details of which we will go into now.

The E_1 enzyme is responsible for decarboxylating pyruvate and attaching the acetyl group of pyruvate to TPP, which is a necessary prosthetic group (covalently bound cofactor). It is also responsible for transferring the acetyl group to E_2's lipoyllysine via a redox reaction and forming a thioester in the lipoyllysine prosthetic group. The E_2 reaction is responsible for producing acetyl-CoA by cleaving the thioester and the reduced lipoyllysine is produced. This reduced lipoyllysine gets oxidized by the FAD cofactor, this produces oxidized lipoyllysine and $FADH_2$. The $FADH_2$ is used to reduce NAD^+ to NADH in the E_3 reaction. The overall enzyme complex reaction is shown below (this is not my image, this is the intellectual property of Lehninger's Principles of Biochemistry, 7th edition and W.H. Freeman and Company):[5]

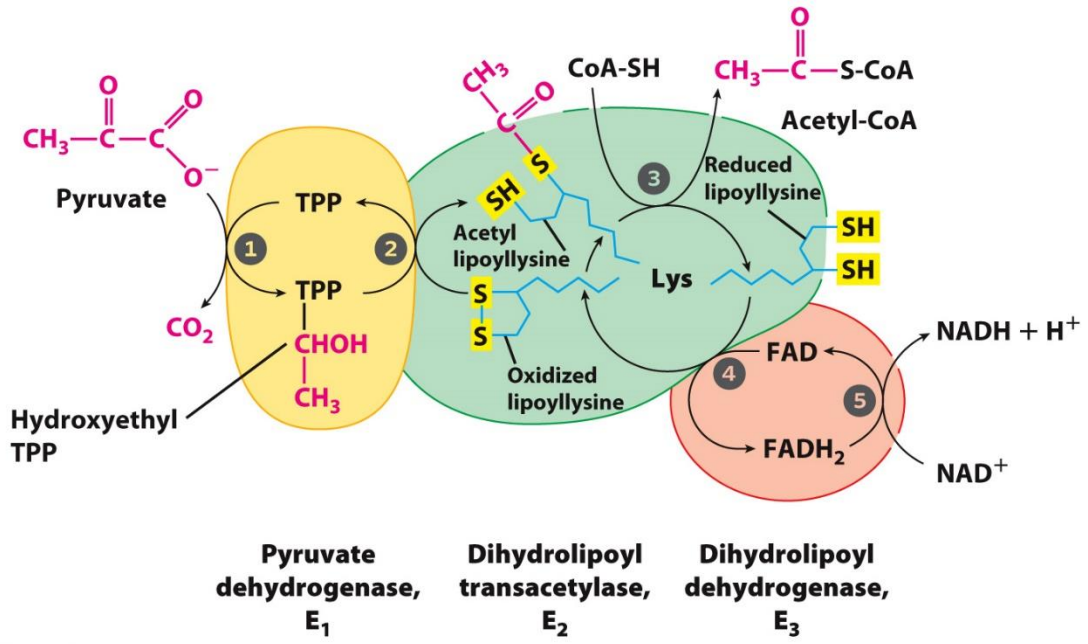


Figure 16-6
Lehninger Principles of Biochemistry, Seventh Edition
© 2017 W. H. Freeman and Company

The PDH complex is an incredibly important enzyme that determines the fate of pyruvate. As such, it is highly regulated by a wide variety of compounds that are summarized below:

Compound name	Inhibitor or activator
ATP	Inhibitor
Acetyl-CoA	Inhibitor
NADH	Inhibitor
Fatty acids	Inhibitor
AMP	Activator
CoA-SH	Activator
NAD^+	Activator
Ca^{2+}	Activator

This form of regulation is consistent with the function of the PDH complex. Ultimately we only want the PDH complex to work if we want to go through the Krebs cycle, however, the purpose of the Krebs cycle is to produce energy and, more specifically, NADH/ $FADH_2$. Because that is the overall goal, anything that indicates to the cell that we already have a lot of energy would indicate that we should NOT use PDH (MAS**C**). This explains why ATP, NADH, and Fatty acids are inhibitors. They are all indicative that the cell has an excess amount of energy and therefore there is no point to go through Krebs cycle. Acetyl-CoA is the product of the reaction, therefore, if we already have an excess of acetyl-CoA there is no reason to make more of it and so PDH will be inhibited (MAS**C**). Conversely, molecules that indicate that we have an energy deficit or that the cell NEEDS energy would activate this complex. That explains why AMP and NAD^+ are allosteric activators of the PDH complex. CoA-SH is a reactant of the PDH complex and therefore the more of it that is available to the cell the more the PDH complex will want to

work to use it because CoA-SH by itself is useless. The only odd ball is Ca^{2+}, however, this can also be rationalized because muscle contractions require Ca^{2+} release and therefore high levels of Ca^{2+} indicates that the cell needs energy and needs it fast. PDH is also regulated by phosphorylation. When ATP levels are high, PDH kinase will add a phosphate to PDH and deactivate it. This is an additional level of control that the cell has over the Krebs cycle.

Practice Questions:

1. Suppose you are an experimenter and you radiolabel C-3 of glucose, where does that radiolabel end up in GAP? What about pyruvate? What about acetyl-CoA? What if I label C-1? What about C-5?
2. Yeast, unlike humans, undergo alcoholic fermentation rather than lactic acid fermentation when there is no oxygen present. Explain why when fermenting grapes to make wine, no oxygen should be present? Explain why no alcohol is produced if the winemaker does not put in Na_2HPO_4 to the yeast mix.
3. Suppose there is an organism that only produces 2 ATP from glycolysis. Which enzyme is likely inactive? How do you know? How could you test this?
4. Insulin is a hormone that indicates to your cells that there is a lot of glucose in the BLOOD. How do you suspect this hormone would regulate glycolysis (would it increase or decrease it)?
5. Glucagon is a hormone that indicates to your cells that there is very little glucose in the blood. How do you suspect this hormone would regulate glycolysis (would it increase or decrease it)?
6. What are the three irreversible steps of glycolysis?
7. How many ATP and NADH would be produced if 2314 glucose molecules went through glycolysis? Please only give the net figures for this. How many ATP equivalents?
8. What catalytic strategy(ies) does GAPDH use?

Answers:

1. Follow the chart I gave. C-3 in GAP would be the carbonyl carbon (4 and 3 are linked). C-3 would be the CO_2^- group of pyruvate. Acetyl-CoA is made by decarboxylating the CO_2^- group, therefore it would end up in CO_2. If you labeled C-1, then that would be the carbon with the phosphate group in GAP, it would be the methyl group in pyruvate, and the methyl group in acetyl-CoA. C-5 would be the same as C-2 because they are linked. This would be the middle carbon in all of the compounds (GAP, pyruvate, and carbonyl of acetyl-CoA).
2. In order for alcoholic fermentation to take place, there must be no oxygen, therefore in order to make the yeast produce alcohol oxygen cannot be there. If there is no P_i source

in the medium, then the yeast cannot do the GAPDH reaction and therefore glycolysis would stop and so too would alcoholic fermentation (no pyruvate = no ethanol).

3. If the organism only makes 2 ATP in the payoff phase then that means one half of the payoff phase is not occurring (we have 2 GAPs ordinarily and these 2 GAPs give us 4 ATP, 2 each). If this is the case then that means that triose phosphate isomerase is not functioning. To test this, we can radio label carbon 3 of glucose. If triose phosphate isomerase is not working, then there should be no radiolabeled CO_2 being produced because carbons 1-3 of glucose would be stuck in DHAP and would not move forward to pyruvate and eventually as CO_2 once glycolysis is over.
4. Insulin says there is a lot of glucose the cell can take in, therefore it should increase the number of glucose transporters on its surface and up the rate of glycolysis so that the cell can take advantage of the large glucose supply in the blood (MAS**C**). This is actually true btw.
5. Glucagon would have the opposite effect, it would tell the cells that we have no glucose that the cells can take in from the blood and therefore glycolysis is not worth doing (no glucose = we can't do glycolysis). Remember, cellular processes are optimized to minimize waste and maximize efficiency (MAS**C**). If there is no glucose available to the cell, then why the heck would they make glycolysis go faster?
6. The three irreversible steps of glycolysis are hexokinase (glucose to glucose-6-phosphate), phosphofructokinase (fructose-6-phosphate to fructose-1,6-bisphosphate), and pyruvate kinase (PEP to pyruvate).
7. For each glucose molecule, we produce 2 net ATP and 2 net NADH. Therefore we simply multiply 2 by however many glucose molecules:

$$\frac{2ATP}{glucose} * \frac{2314glucose}{1} = 4628\ ATP$$

$$\frac{2NADH}{glucose} * \frac{2314glucose}{1} = 4628\ NADH$$

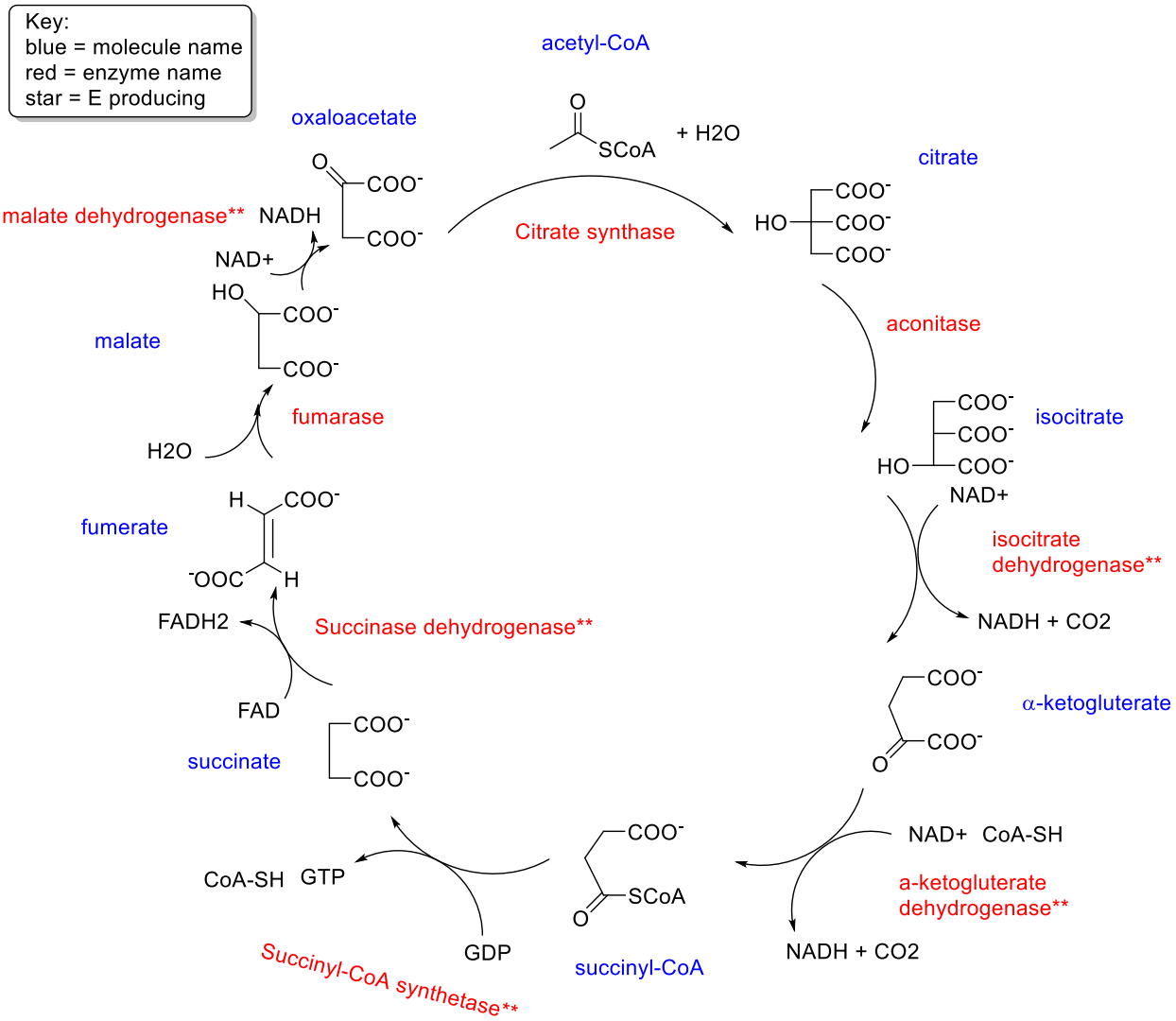
To get the ATP equivalents, we have to convert the NADH to ATP using the relationship that 1NADH = 2.5 ATPs like so:

$$ATP_{total} = 4628\ ATP + 4628NADH * \left(\frac{2.5ATP}{NADH}\right) = 16198\ ATP$$

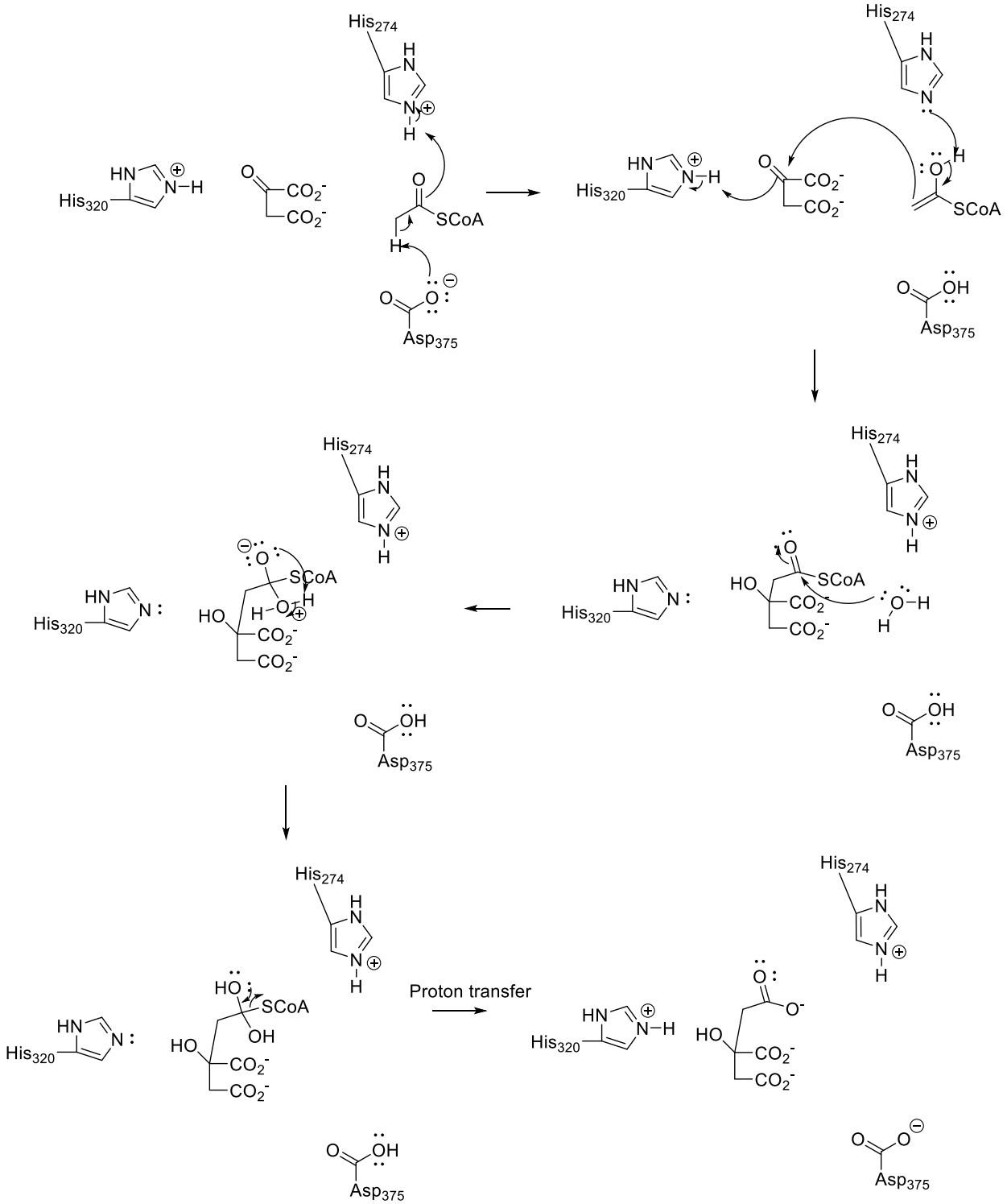
8. GAPDH uses covalent catalysis

Chapter 22: Krebs Cycle

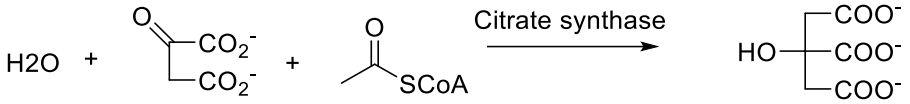
Once we have converted pyruvate to acetyl-CoA in the mitochondria, that acetyl-CoA gets transported to the mitochondrial matrix, which is the very center of the mitochondria. Once there, acetyl-CoA will go through the Krebs cycle. The most important thing to keep in mind with the Krebs cycle is the main goal: to make NADH and $FADH_2$. These two molecules will power the next step of carbohydrate breakdown, the electron transport chain. It is always important to keep the big picture in mind because without it, these pathways become a lot more difficult. An overview of the Krebs cycle is shown below. Just as before, we will go through all of the steps and discuss the major regulatory enzymes in this process.



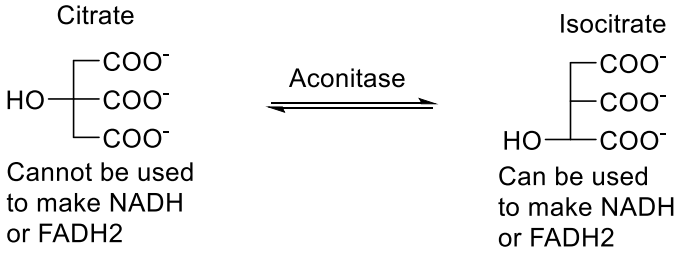
The very first reaction in the Krebs cycle is the creation of citrate from acetyl-CoA and oxaloacetate. This reaction is highly favorable and therefore is irreversible. The mechanism of this reaction is shown below:



Because this is the irreversible entry point into the Krebs cycle, this enzyme is highly regulated just like most other irreversible enzymes. Again, it is important to keep in mind the overall goal of the Krebs cycle when we discuss regulation because the overall goal can help rationalize why the regulations make sense. The Krebs cycle is used to make NADH and ultimately to make ATP through the electron transport chain. As such, molecules that indicate that energy is in surplus would decrease citrate synthase's activity and the opposite is true for molecules that indicate that energy is in demand. Citrate synthase is inhibited by citrate itself (feedback inhibition), ATP, NADH, and succinyl-CoA while it is activated by ADP. These can all be rationalized using the overall goal of Krebs cycle. ATP, NADH, and succinyl-CoA all indicate that the cell either has energy directly (ATP) or has the means to make a lot of energy in the future (NADH and sucinyl-CoA). ADP on the other hand indicates the cell lacks energy and therefore we need to go through Krebs cycle (MAS**C**). The overall citrate synthase reaction is shown below:

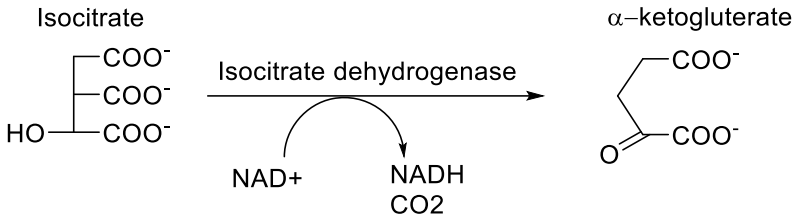


The second step of the Krebs cycle is the isomerization of citrate to isocitrate using aconitase. This reaction is unfavorable and is reversible, the reaction is driven forward by keeping isocitrate concentrations low. This isomerization step is important because the tertiary alcohol of citrate CANNOT be oxidized and therefore it is useless if we want to make NADH or $FADH_2$. Isocitrate has a secondary alcohol, which as we know from organic chemistry can get oxidized to a ketone in the presence of an oxidizing agent (NAD^+ or FAD). The overall reaction is shown below:

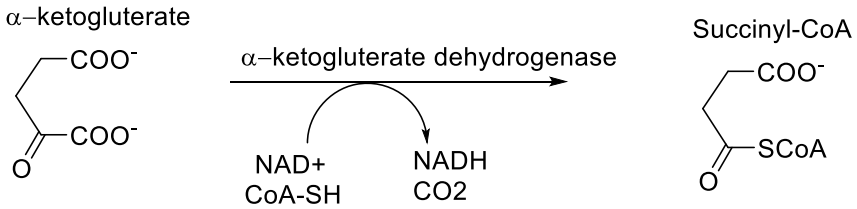


The third step is the production of α-ketogluterate from isocitrate via isocitrate dehydrogenase. Whenever dehydrogenase is the ending of the enzyme name, either NADH or $FADH_2$ is produced because the enzyme catalyzes a redox reaction. As a general rule, when a C-O bond is being oxidized to C=O, NADH is produced and if a C-C bond is oxidized to C=C, $FADH_2$ is produced with very few exceptions. In this reaction, the secondary alcohol of isocitrate is being oxidized to a ketone and therefore we are oxidizing a C-O bond and thus NADH is produced. This reaction is highly favorable and is irreversible. Like all irreversible reactions, this enzyme is highly regulated using a similar paradigm to citrate synthase. Isocitrate dehydrogenase is inhibited by ATP and activated by Ca^{2+} and ADP for much the same reasons that citrate synthase was regulated this way. This reaction releases CO_2 from the middle carbon of isocitrate. Releasing CO_2 is an important part of the Krebs cycle because it allows for further oxidation of the carbon chain to occur later on (thus producing NADH and $FADH_2$), which as we know, is the

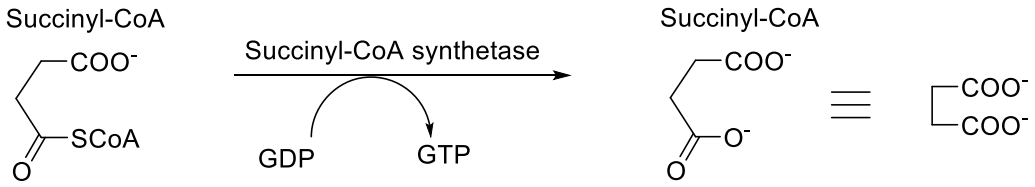
driving force for the Krebs cycle to occur in the first place. The overall reaction of isocitrate dehydrogenase is shown below:



The next step in the Krebs cycle is another oxidative decarboxylation reaction that is catalyzed by α-ketogluterate dehydrogenase. This enzyme converts α-ketogluterate to succinyl-CoA using CoA-SH and NAD^+. This reaction is also irreversible and highly favorable, therefore, it is highly regulated. Like many enzymes we have seen so far, it is inhibited by its products (succinyl-CoA and NADH) and activated by Ca^{2+}. Like citrate synthase, isocitrate dehydrogenase, and PDH, this enzyme is regulated this way to only work when the cell needs energy (MAS**C**). The overall reaction is shown below:

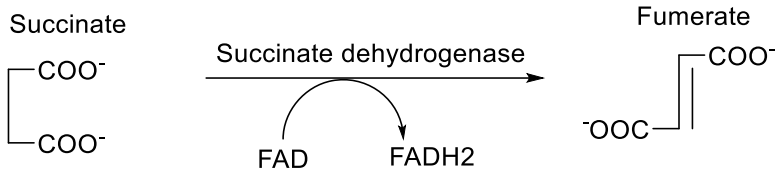


The thioester present in succinyl-CoA makes it so that it has the energy to produce an ATP equivalent (GTP) through the reaction catalyzed by succinyl-CoA synthetase. This reaction is slightly favorable, but overall is reversible, which is why even though it is named succinyl-CoA synthetase, it actually makes succinate not succinyl-CoA. The overall reaction of this enzyme is shown below:

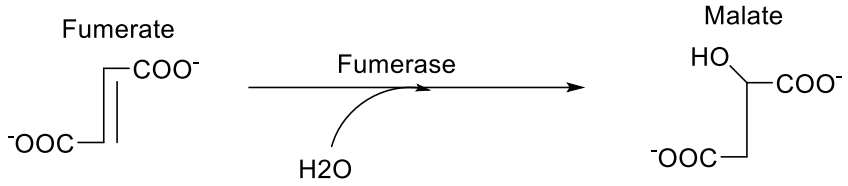


Succinate is perfectly symmetrical and therefore the two CH_2 carbons are equivalent and so are the two CO_2^- carbons.

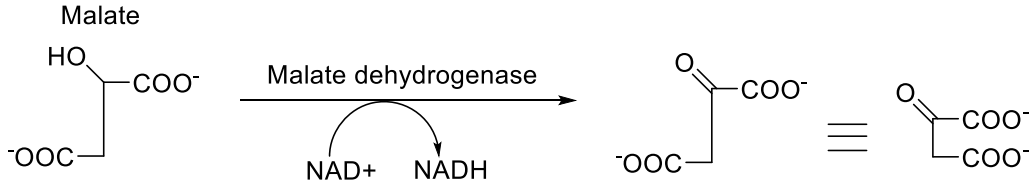
Succinate has not C-O bonds to oxidize, therefore, the only bonds that could get oxidized are the C-C bonds and that is exactly what happens. Succinate dehydrogenase catalyzes the oxidation of succinate to fumarate and produces $FADH_2$ in the process. This reaction is reversible and the $FADH_2$ is used directly in complex II of the electron transport chain (details to come soon). The overall reaction is shown below:



Now that the new C=C pi bond is made, that pi bond can get hydrated to make an alcohol, which can ultimately get oxidized later on to produce NADH. This hydration process is catalyzed by fumerase. This reaction is reversible and stereospecific for L-malate. The reaction is shown below:



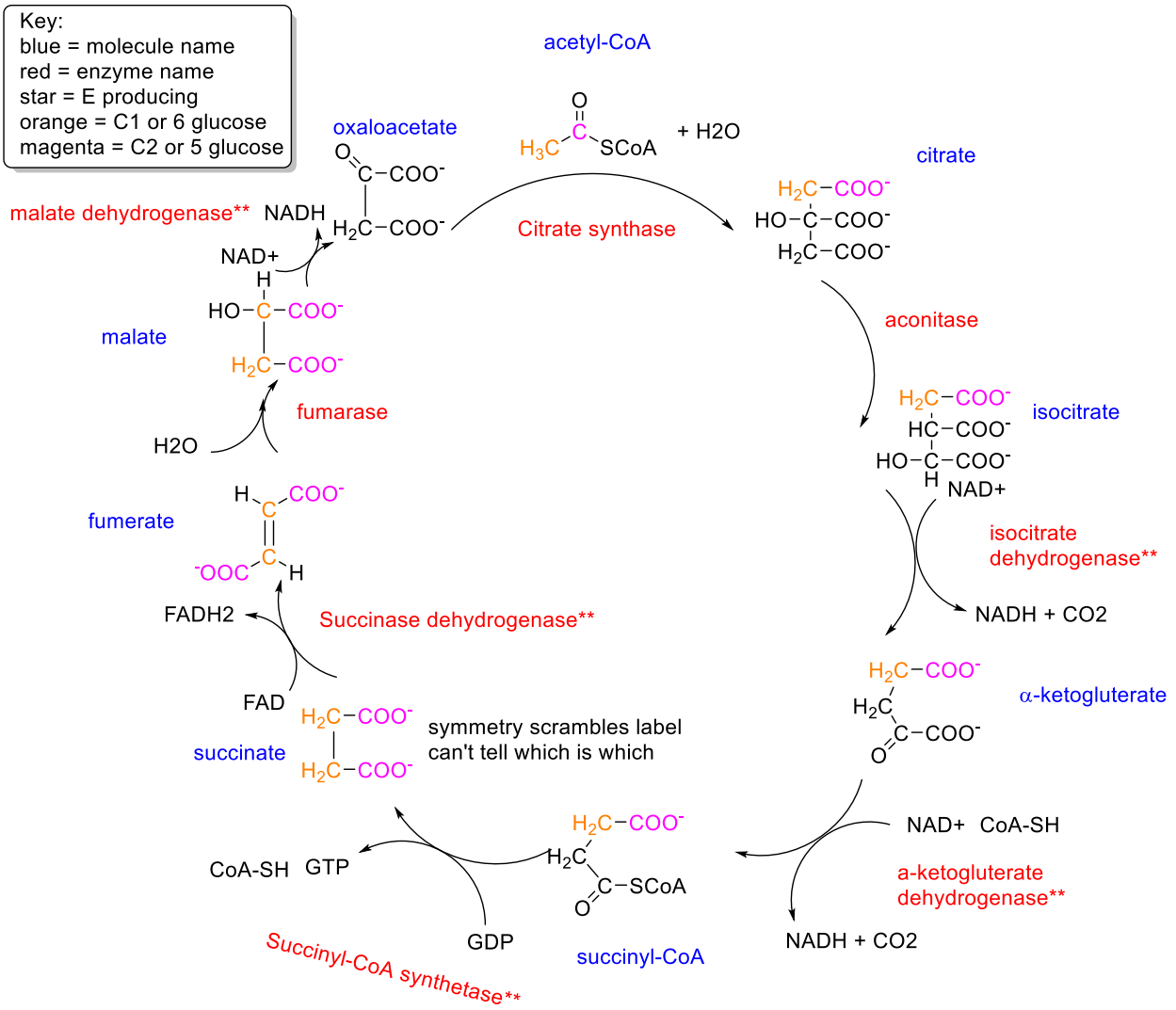
Finally, the very last step of the Krebs cycle is the oxidation of malate to oxaloacetate via malate dehydrogenase. Because this reaction is oxidizing a C-O bond, it produces NADH. However, unlike the other dehydrogenases in this cycle, this reaction is reversible and unfavorable because the OH group can form intramolecular hydrogen bonds with the CO_2^- group adjacent to it in malate, but cannot do that in oxaloacetate. This reaction is pulled forward by keeping oxaloacetate concentrations very low. The overall reaction is shown below:



A quick summary of the Krebs cycle:

1. Main purpose is to make NADH and $FADH_2$ (per acetyl-CoA, we produce 3 NADH and 1 $FADH_2$)
2. All the irreversible reactions are highly regulated (PDH, citrate synthase, isocitrate dehydrogenase, and α-ketogluterate dehydrogenase)
3. Groups are moved around/removed and C=C pi bonds are made to make room for more oxidation
4. Acetyl-CoA carbons are NOT removed.
5. Succinate dehydrogenase functions as a dual agent, works in Krebs cycle and the electron transport chain

Here is a radiotracing guide to help you:



Because succinate is symmetrical, there is no way of knowing which CH_2-CO_2^- pair was the original one that had the radiolabel, therefore everything after succinate the radiolabel loses meaning. The two CH_2 of succinate are equivalent (either one could be orange) and the two CO_2^- groups are equivalent (either one could be magenta). I kept going with it until malate, but I doubt that your professor will ask you about a radiolabel anywhere after succinate. To do so would make no sense because you could not predict accurately where the radiolabel would end up.

To help you with remembering the regulations of each enzyme, I have provided a table below that summarizes all of the activators and inhibitors of each of the major regulatory enzymes of the Krebs cycle.

Enzyme	Inhibitors	Activators

Citrate synthase	NADH, succinyl-CoA, ATP	ADP
Isocitrate dehydrogenase	ATP	Ca^{2+}, ADP
α-ketogluterate dehydrogenase	Succinyl-CoA, NADH	Ca^{2+}

These enzymes are important not only because they are the only irreversible reactions in the Krebs cycle, but also because they are important branch points for other cellular processes. Citrate is important for fatty acid synthesis and α-ketogluterate is important for amino acid synthesis and breakdown as we will see later on in this text. Because citrate and α-ketogluterate have important roles in other cellular functions, the enzymes that produce them and break them down are highly regulated.

Practice Questions:

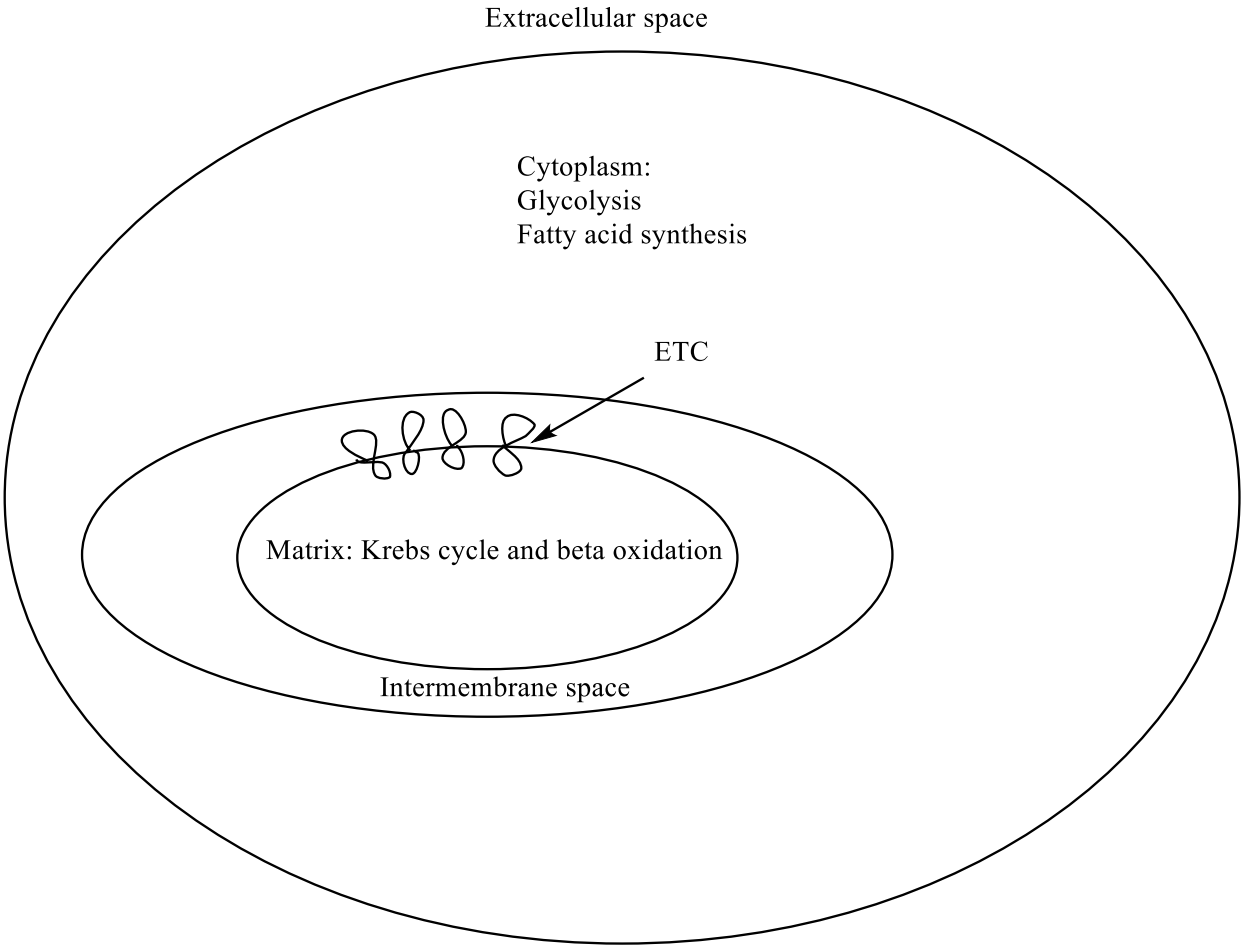
1. Suppose you radiolabeled carbon 2 of glucose, where does the radiolabel end up in citrate? α-ketogluterate?
2. Suppose you design an inhibitor of pyruvate dehydrogenase complex. What would be the effect of this inhibitor on the rate of lactic acid fermentation? Why?
3. Under what conditions of ATP and NADH would the Krebs cycle be most active? Most inactive? Explain.
4. Suppose you design an inhibitor of citrate synthase, what would the effect of this inhibitor have on the rate of fatty acid synthesis? Why?
5. How many acetyl-CoA molecules are required to make 36 NADH and 12 GTP from the Krebs cycle? How many glucose molecules would this be equivalent to?

Answers:

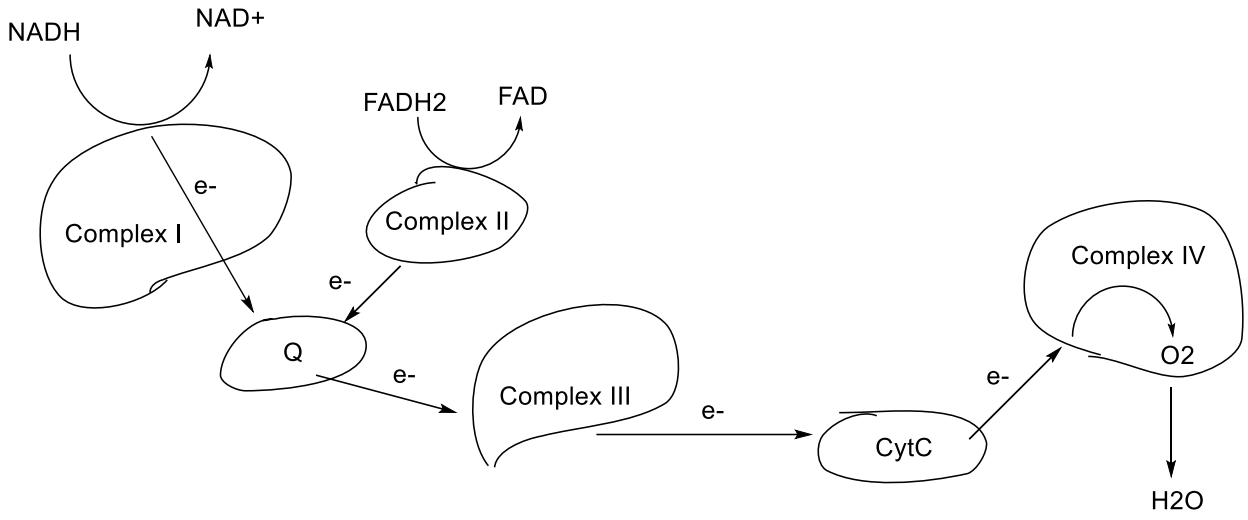
1. Use the radiolabeling chart, it will be the terminal carbonyl group carbon for both citrate and α-ketogluterate.
2. If PDH no longer works, then aerobic respiration (Krebs cycle and electron transport chain aka ETC are not able to occur), therefore the cell must resort to anaerobic respiration aka lactic acid fermentation. Thus, the rate of lactic acid fermentation would increase substantially to compensate for the ATP loss due to lack of aerobic respiration.
3. Low ATP and low NADH = most active Krebs cycle, high ATP and high NADH = least active Krebs cycle. This is because Krebs cycle is active only when the cell needs energy and since ATP and NADH are markers that tell us the energy needs of the cell, if they are low that means the Krebs cycle would have to work in over drive. On a molecular level, however, consider the fact that without ATP and NADH, there are fewer allosteric inhibitors for the enzymes citrate synthase, isocitrate dehydrogenase, α-ketogluterate dehydrogenase, and PDH complex, therefore flux through Krebs cycle should be faster.
4. If you inhibit citrate synthase then you would decrease citrate production. Without citrate production, fatty acid synthesis cannot occur (details soon to come).
5. Each acetyl-CoA molecule makes 3 NADH and 1 GTP, therefore we would need 12 acetyl-CoA to make 36 NADH and 12 GTP. For every glucose molecule, we get 2 acetyl-CoA and therefore we would need 6 glucose molecules to make 12 acetyl-CoAs.

Chapter 23: Electron Transport Chain

This is the main event, ladies and gents. The main ATP payoff phase of aerobic respiration and the reason why we need oxygen so badly. From now on, I will be referring to this as the ETC for short for the sake of brevity. The ETC has four main protein complexes and these protein complexes use NADH and $FADH_2$ to pump H^+ or protons from the matrix of the mitochondria to the intermembrane space. To give you a nice map of the cell and where all of these metabolic processes take place, I have made a little schematic for you shown below (not to scale):



Even though there are four complexes in the ETC, there are two entry points neither of which use all four. The two entry points are the result of the two different types of electron carriers that can be used to power the ETC (NADH and $FADH_2$). If the electrons are being supplied from NADH, then the electrons will flow from complex I to complex III to complex IV. If the electrons are supplied from $FADH_2$, then the electrons flow from complex II to complex III to complex IV. This should make sense because as you should recall, complex II is the same as succinate dehydrogenase, which makes $FADH_2$. The two pathways are shown schematically below:



We will cover these two pathways of electron flow individually. But just like the Krebs cycle, we have to contextualize this whole pathway with what the main goal is. The main goal of the ETC is to make more protons be in the intermembrane space than there are in the matrix. This creates an electrochemical gradient that the cell can use to power ATP synthesis using ATP synthase (recall the ΔG of transport equation).

First, we will discuss the NADH pathway, which involves electron movement from complex I to III to IV. When NADH approaches complex I, it donates its two electrons to FMN, which is a redox cofactor for the large protein complex. This FMN will pass electrons one at a time to Fe-S clusters that act as the electron highway throughout the complex. Ultimately, these electrons are passed to ubiquinone (Q), which acts as the exit from the electron highway. When the electrons move through the complex, the redox energy produced by these electron transfers is used to power the movement of FOUR protons from the matrix to the intermembrane space.

We can actually calculate the energy produced by an electron transfer event (redox reaction), by using the equation:

$$\Delta G = -nFE_o$$

Here n is the number of electrons transferred, F is Faraday's constant (96500 J/Vmol), and E is the cell potential. To calculate the cell potential, we have to look at the reduction potentials of the two chemicals in the reaction. The more positive the reduction potential, the more that compound wants to accept electrons (get reduced). Let's consider the reduction potentials for NADH and FMN to calculate the energy released upon electron transfer from NADH to FMN. To do this, we have to first write out the chemical reaction like so:

$NADH + FMN + H^+ \rightarrow NAD^+ + FMNH_2$

Now that we have the chemical reaction, we can evaluate its redox half reactions to get the overall potential (E) of this reaction. Here the NADH is getting oxidized because it is losing electrons (remember LEOGER) and therefore the FMN is getting reduced.

The oxidation half reaction is:

$NADH + H^+ \rightarrow NAD^+ + 2e^- + 2H^+$ $\quad E_{reduction} = -0.320V$

The reduction half reaction is:

$FMN + 2H^+ + 2e^- \rightarrow FMNH_2$ $\quad E_{reduction} = -0.300V$

Because the NADH is getting oxidized, we have to invert the reduction potential from -0.320V to +0.320V and we can keep the FMN reduction potential the same since it is in fact getting reduced. Now we can calculate the total E for the reaction like so (using this method we just add):

$$E_0 = +0.320V - 0.300V = 0.020V$$

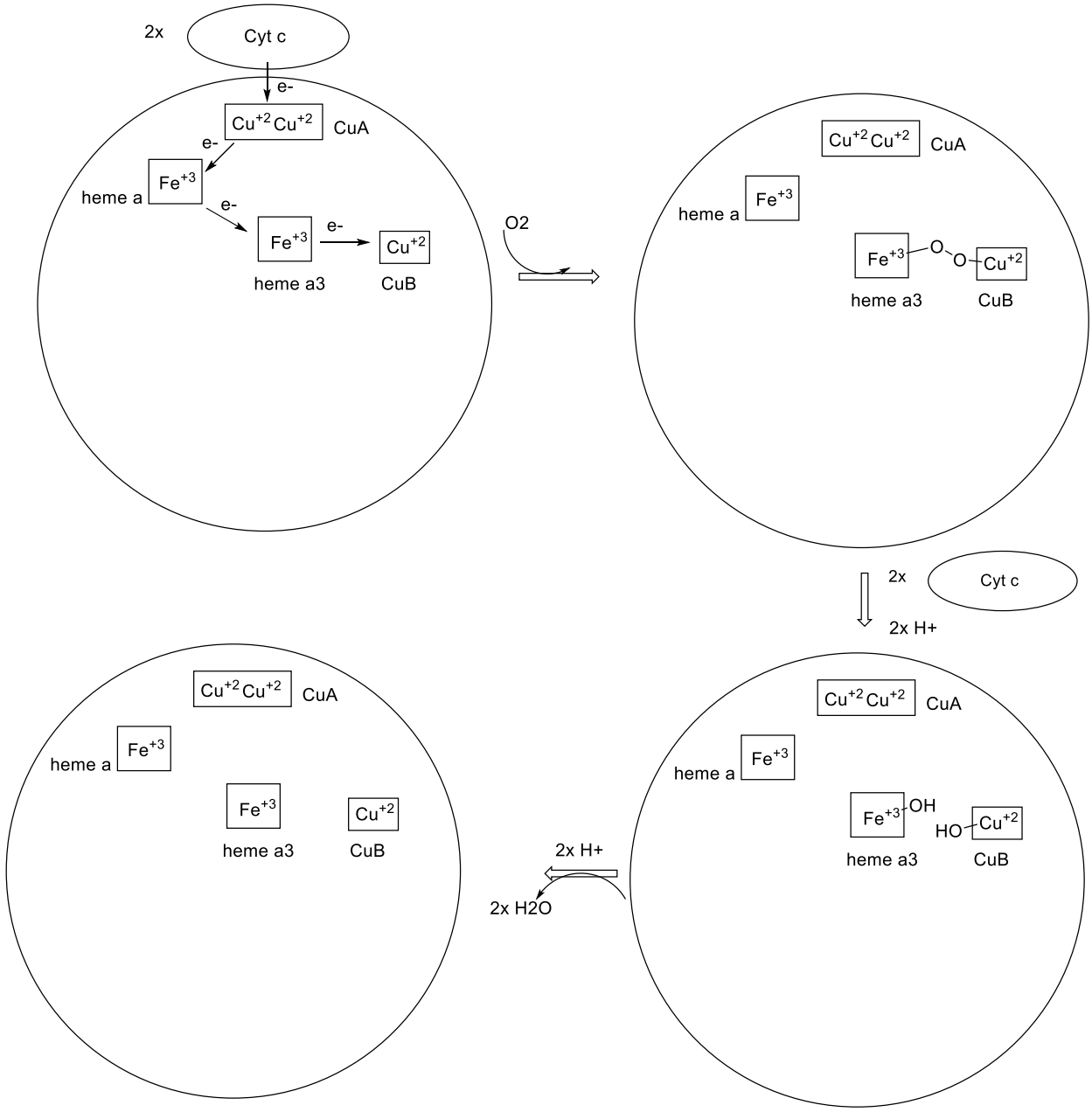
And therefore we can calculate the free energy change:

$$\Delta G = -(2)\left(96500\frac{J}{Vmol}\right)(0.020V)$$

$$\Delta G = -3860\frac{J}{mol}$$

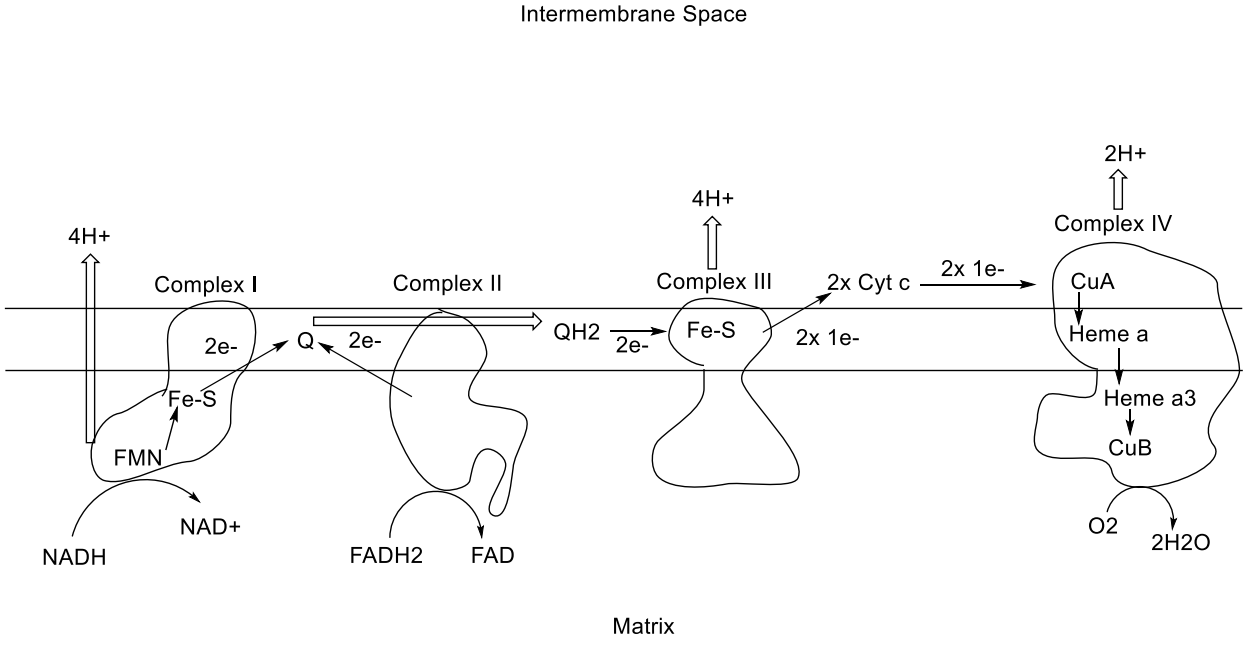
Once Q gets reduced to QH_2, it moves through the membrane and transports its two electrons to complex III. Once it is near complex III, the electrons from ubiquinone get transferred to Fe-S clusters that are present in complex III. These again act as electron highways and transports the electrons to cytochrome C (Cyt c). Cyt c is a heme protein that has a redox active Fe center that can be oxidized (Fe^{3+}) or reduced (Fe^{2+}). Cyt c is different from ubiquinone in two key ways, first, it moves in the intermembrane space NOT the membrane and second, it transfers only a single electron rather than two (this is because the reduced state is Fe^{2+} not Fe^+). Once the electrons are passed to Cyt c, the energy released by the electron transfer is used to power the movement of another FOUR protons from the matrix into the intermembrane space. Of those four protons, two of them come directly by electron transfer from QH_2 to the Fe-S clusters while the other two come from rereducing another molecule of Q to QH_2.

Once Cyt c gets reduced, a pair of reduced Cyt c will move over to complex IV, where it can transfer their two electrons to Cu_a, a pair of two copper ions present in complex IV. These two copper ions will transfer their electrons to heme a, which is a nearby heme cofactor. These electrons will then be transferred to heme a_3 and Cu_b leaving them both reduced. Now that they are reduced, molecular oxygen will bind both the Fe in the heme and the Cu in Cu_b and form a peroxide bridge. This peroxide bridge is unstable and so when another pair of reduced Cyt c's come by to drop off their electrons again, those electrons will be used to reduce the peroxide bond to give two O^-'s on both the Fe and the Cu. At the same time, two protons are transfer from the matrix to protonate these two O^-'s to make them OH groups. Two more protons are transfer from the matrix to make the two OH groups water molecules, taking with them electrons and making the Fe and Cu return to their oxidized state. The end result is 4 protons transferred, two to form the water molecules and two pumped into the intermembrane space. A schematic diagram of this process is shown below:



The second pathway is very similar to the first. The only difference is the first step, instead of going through complex I first, we go through complex II first. In complex II, the electrons from $FADH_2$ are transferred to Fe-S clusters, just like the other enzyme complexes and are shuttled to ubiquinone to form QH_2. Unlike complex I, however, this electron transfer DOES NOT pump ANY protons and that is why $FADH_2$ does not give as much ATP as NADH. Everything after complex II is the exact same as before, so I will not repeat it.

The overall picture of the electron transport chain is this:



Once this electron transport chain process is finished, an electrochemical gradient is established due to the H^+ pumped into the intermembrane space. This electrochemical gradient can be used to power ATP synthase, the enzyme that is responsible for making ATP. Now we will describe the actual mechanism of ATP synthesis using the ATP synthase enzyme.

ATP synthase is comprised of two functional units, the F_1 unit which faces the matrix and catalyzes the creation of ATP and the F_0 unit which inside the membrane and is responsible for transporting protons from the intermembrane space to the matrix to power the F_1 unit. The F_1 unit has three sets of dimers that comprise it and these dimers can be in three different conformations that serve different functions (MA**S**C). These conformations are the open, loose, and tight conformations. The open conformation allows ATP to leave, the loose conformation binds ADP and P_i, and the tight conformation squishes the ADP and P_i together to make ATP. When a proton is moved into the F_0 unit, that causes the central stalk of the ATP synthase enzyme to rotate 120 degrees. This rotation causes all three of the dimers to change conformation in the order L-T-O (i.e. if one dimer was in L it would go to T, if the other was in T it would go to O, and if the last was in O, it would loop back around to L). Therefore, the cost to produce 1 ATP just looking at the ATP synthase machinery is 3 protons. This is because if we start at O we would go to L with 1, go to T with another, and then return back to O with the third to restart the cycle. This turns out to be insufficient, however, because a fourth proton is required to move NADH from the cytosol to the mitochondria and to move ATP out of the mitochondria. Recall that the NADH produced in glycolysis is in the cytoplasm, while all of this is occurring in the mitochondria so we need to somehow move the NADH into the mitochondria. It turns out that there are two ways to do this: the malate-aspartate shuttle and the glycerol-3-phosphate shuttle, the details of which are not super important to know so I will leave that to you to search up on your own if you are curious.

Summary of the electron-transport chain:

1. Complex I, III, and IV pump protons into the intermembrane space.
2. The proton gradient made by the above complexes powers ATP synthesis
3. Complex IV consumes oxygen and makes it into water at the end of the ETC
4. ATP synthase uses its F_1 subunit to make ATP and uses its F_0 subunit to transport protons into the matrix and power this machinery.

Practice Questions:

1. How many protons are required to make 1 ATP molecule?
2. How many protons are pumped per NADH molecule? $FADH_2$? If there is a difference, explain why.
3. Suppose you made a drug that caused the innermembrane of the mitochondria to become porous. That is to say that protons can move freely from the matrix to the intermembrane space and vice versa. What would the effect of this compound have on O_2 consumption? What about ATP production?
4. Suppose you made a drug that inhibits ATP synthase and prevented it from making ATP and moving protons from the intermembrane space to the matrix. What would the effect of this compound have on O_2 consumption? What about ATP production? What about glycolysis and Krebs cycle? Explain.
5. Calculate the number of protons pumped for 1 glucose molecule.

Answers:

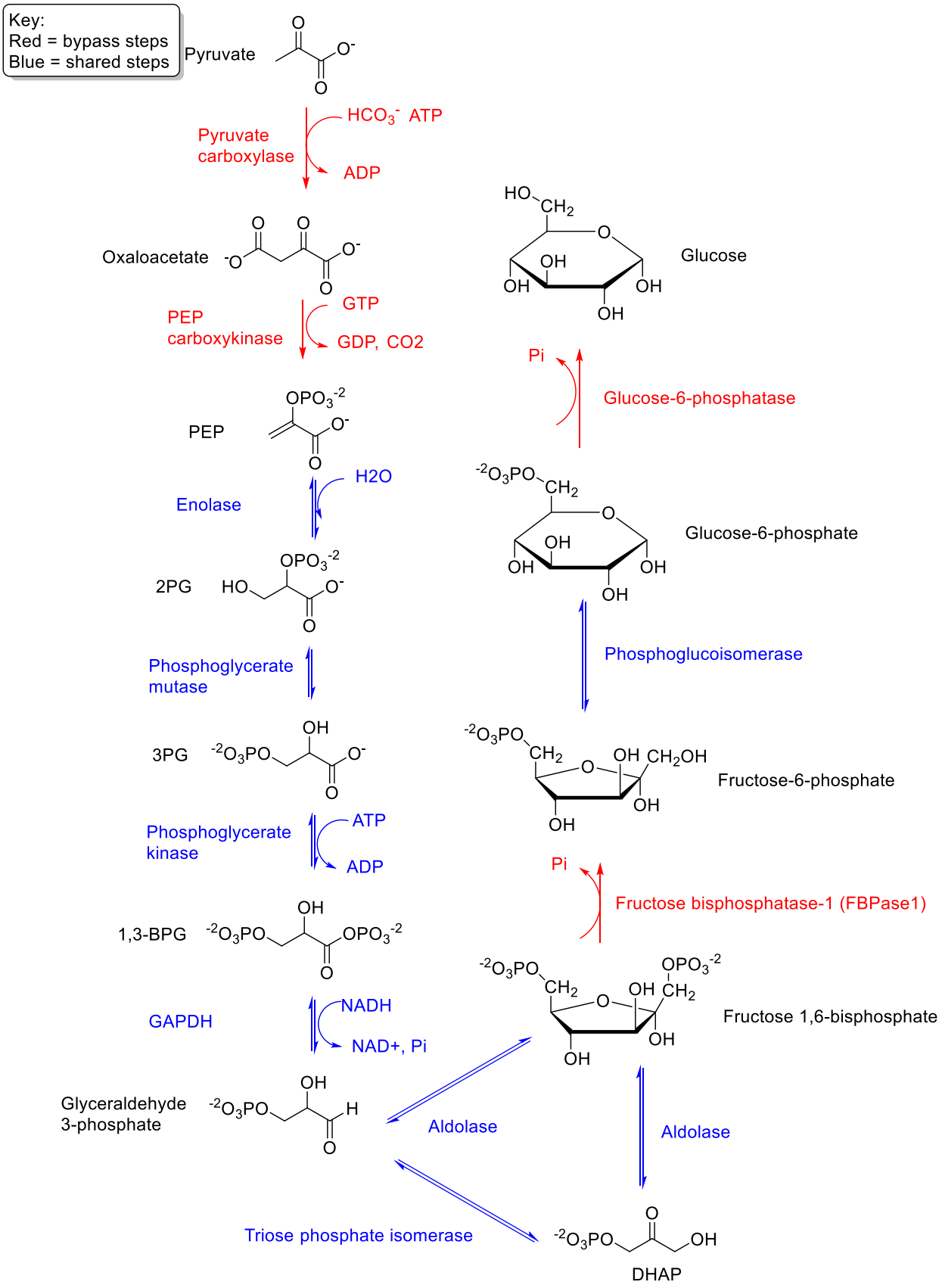
1. 4 protons are required to make 1 ATP molecule, 3 to move the F_1 subunit three times and 1 to move the NADH and ATP to the mitochondria and the cytoplasm respectively.
2. Each NADH molecule pumps 10 protons (4 from complex I, 4 from complex III, and 2 from complex IV) while $FADH_2$ only pumps 6 protons (4 from complex III and 2 from complex IV). The difference is because complex II does not pump any protons into the intermembrane space.
3. If the membrane is porous then that means the electrochemical gradient cannot be established. This means ATP synthesis would stop. However, because the ETC doesn't "realize" the membrane is porous, it will continue to use up NADH and $FADH_2$ to reduce O_2 to water and therefore O_2 consumption would increase like normal.
4. If the drug inhibits ATP synthase, then the protons that are pumped by complex I, III, and IV will remain in the intermembrane space. More and more protons would be pumped until eventually the energy supplied by the redox chemistry of the ETC is not sufficient to pump any more protons across the innermembrane. When this happens, all of the ETC stops working, meaning complex IV does not work anymore and oxygen can no longer get consumed by the cell. This makes oxygen consumption decrease and ATP production decrease. The end result is that NADH and $FADH_2$ accumulate in the cell since they have nowhere to go anymore (ETC doesn't work) and so NADH will allosterically inhibit citrate synthase, α-ketogluterate dehydrogenase, and PDH complex. At the same time,

however, ADP levels and AMP levels will rise in the cell and that will activate several glycolytic enzymes such as PFK1 and pyruvate kinase and Krebs cycle enzymes. So the effect on overall carbohydrate metabolism is a wash, you would likely have to see experimentally what would happen.

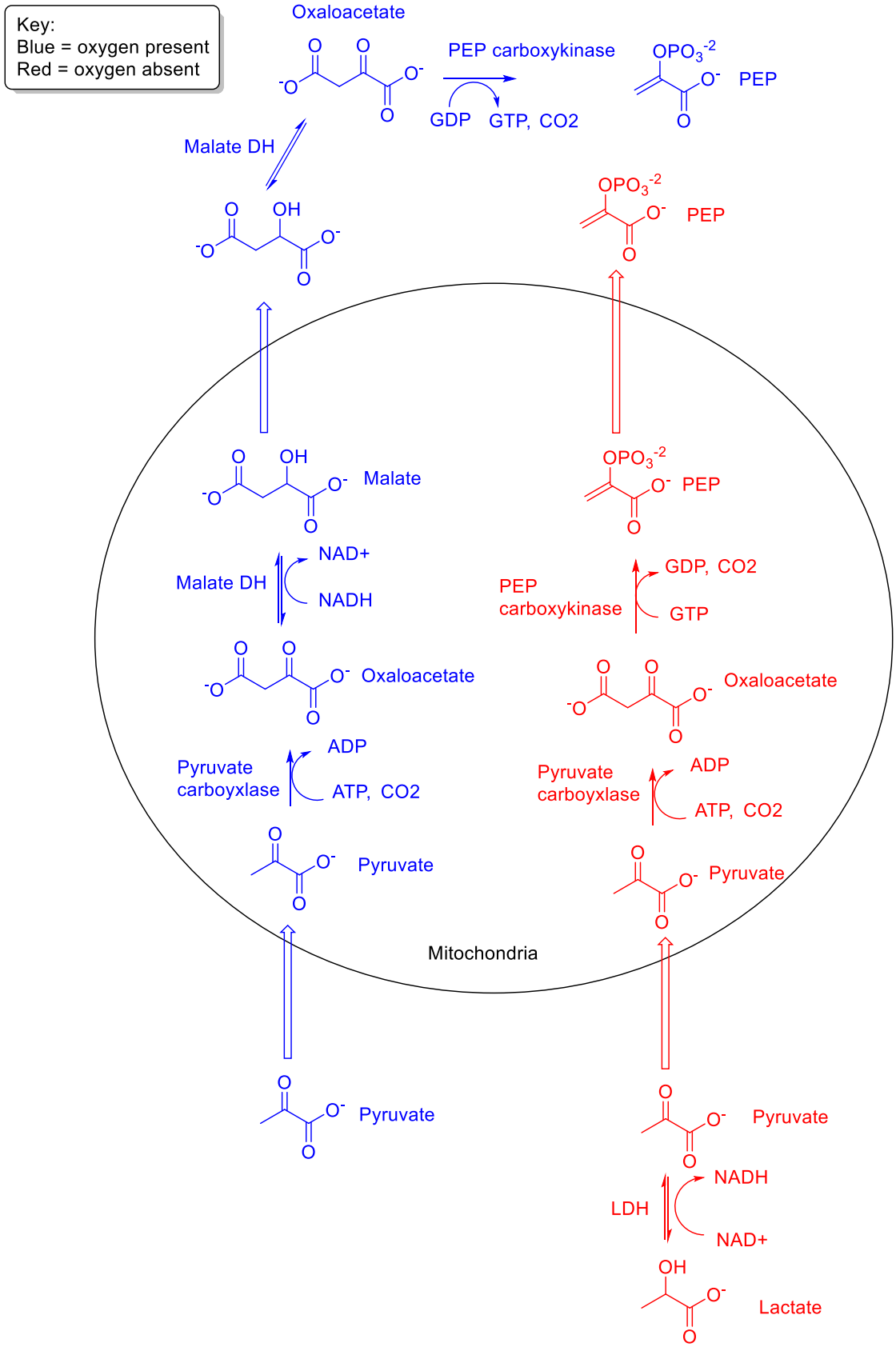
5. For every 1 glucose molecule, we create 2 NADH from glycolysis, 2 acetyl-CoAs which each give 3 NADH and 1 $FADH_2$. Therefore each glucose molecule creates 2 + 3(2) = 8 NADH and 2 $FADH_2$. Each NADH pumps 10 protons and each $FADH_2$ pumps 6 protons, therefore 8(10) + 6(2) = 80 + 12 = 92 protons pumped per glucose molecule.

Chapter 24: Gluconeogenesis

We have just covered how sugar is broken down, so now we will discuss how sugars are made. In this chapter, we will specifically discuss how glucose is made. The process in which glucose is made is referred to as gluconeogenesis and luckily for you, you know most of the steps already to this process. Gluconeogenesis uses the reversible steps of glycolysis to go backwards. This is evolutionarily favorable because it avoids having to make new enzymes that could take millions of years to perfect. In this way, the cell maximizes its resources and can do many processes with fewer enzymes (MAS**C**). There are three irreversible steps of glycolysis, however, and therefore we will discuss how gluconeogenesis gets around them. Like all metabolic processes, we have to consider the overall goal of gluconeogenesis before we delve into the specifics so that we can contextualize the pathway. Gluconeogenesis is done only when there is a lack of glucose in the body and the body has enough energy to make glucose. This happens if someone is on a low carb diet. The liver is the glucose storage unit of the body and therefore this organ specifically has high rates of gluconeogenesis. Glycolysis and gluconeogenesis are fundamentally opposing pathways and therefore, to avoid making glucose at the same time as you break it down, the major enzymes for both pathways are highly regulated and gluconeogenesis is primarily done in the liver while glycolysis is mainly done in the muscle and brain. An overall scheme of gluconeogenesis is shown below:


Key:
Red = bypass steps
Blue = shared steps
Pyruvate
Pyruvate carboxylase
HCO3- ATP
ADP
Oxaloacetate
PEP carboxykinase
GTP
GDP, CO2
PEP
Enolase
H2O
2PG
Phosphoglycerate mutase
3PG
Phosphoglycerate kinase
ATP
ADP
1,3-BPG
GAPDH
NADH
NAD+, Pi
Glyceraldehyde 3-phosphate
Aldolase
Triose phosphate isomerase
DHAP
Aldolase
Fructose 1,6-bisphosphate
Pi
Fructose bisphosphatase-1 (FBPase1)
Fructose-6-phosphate
Phosphoglucoisomerase
Glucose-6-phosphate
Pi
Glucose-6-phosphatase
Glucose

The starting point of gluconeogenesis is pyruvate. To start its journey to becoming glucose, the pyruvate needs to be first converted to oxaloacetate and ultimately PEP. We need to make it PEP because otherwise it cannot use the glycolysis machinery to make glucose (MAS**C**). However, pyruvate can come from different places depending upon if the cell is in the presence of oxygen or not. If the cell is in an oxygenated environment, then the pyruvate will enter the mitochondria and become oxaloacetate using pyruvate carboxylase. Pyruvate carboxylase, like all carboxylases uses biotin as a cofactor. Adding the CO_2^- group is not energetically favorable, however, so the cell needs to use ATP to power this reaction (M**A**SC). We need the oxaloacetate we just produced to be in the cytoplasm because that is where the glycolytic enzymes are located, therefore, the oxaloacetate will be converted to malate using malate dehydrogenase in the mitochondria. This malate can be moved into the cytoplasm and then be oxidized to oxaloacetate. This is a common method of moving carbons from the mitochondria to the cytoplasm and is referred to as the malate shuttle. Once the oxaloacetate is moved into the cytoplasm, it can be converted to PEP using PEP carboxykinase at the expense of GTP (which is energetically equivalent to ATP). If the cell does not have O_2 available, then instead of going through malate, the oxaloacetate will be converted directly to PEP in the mitochondria. A summary of this process is shown below:


Key:
Blue = oxygen present
Red = oxygen absent
Oxaloacetate
PEP carboxykinase
GDP
GTP, CO2
PEP
Malate DH
Malate
NAD+
NADH
Oxaloacetate
ADP
Pyruvate carboxylase
ATP, CO2
Pyruvate
Mitochondria
PEP carboxykinase
GDP, CO2
GTP
Pyruvate carboxylase
LDH
Lactate
Cytoplasm

It should make sense that the pyruvate becomes oxaloacetate and malate in the mitochondria if oxygen is available because Krebs cycle is only active if oxygen is present and both oxaloacetate and malate are Krebs cycle intermediates. That is why the pathway with oxygen absent does not use malate DH at all; the Krebs cycle doesn't work if there is no oxygen!

Once the PEP is made, the glycolytic enzymes for the reversible steps can be used all the way up to fructose 1,6-bisphosphate. That brings us to the second bypass step, which uses FBPase-1. FBPase-1 is a phosphatase, which are enzymes that remove phosphates from proteins or small molecules. This specific phosphatase is responsible for removing the phosphate from the 1 carbon of fructose 1,6-bisphosphate. This enzyme is highly regulated in conjunction with PFK-1. It is inhibited strongly by AMP, while PFK-1 is inhibited by ATP and citrate but activated by ADP and AMP. This makes sense because gluconeogenesis is expensive and consumes energy, therefore, if AMP levels are high that indicates the cell has no energy and therefore gluconeogenesis should not occur (MAS**C**). This step is also regulated by fructose 2,6-bisphosphate levels, if you recall from the glycolysis chapter and just like how there is a PFK-1 and FBPase-1, there is a PFK-2 and FBPase-2. They are perfectly analogous, PFK-1 and PFK-2 both promote glycolysis, PFK-1 because it is directly involved with the pathway and PFK-2 because it makes fructose 2,6-bisphosphate, which allosterically activates PFK-1. Likewise, FBPase-1 and FBPase-2 will inhibit glycolysis either by destroying fructose 1,6-bisphosphate (FBPase-1) or by destroying the activator of glycolysis (FBPase-2). As one would expect, glucagon and insulin levels control the expression of these proteins. Glucagon indicates the sugar is low, if sugar is low then that means the cell needs to make sugar; gluconeogenesis should increase. Insulin on the other hand indicates that sugar is high, if sugar is high then we don't need to make more of it; gluconeogenesis should decrease (MAS**C**). That is exactly what happens. Glucagon activates PKA, which will phosphorylate and activate FBPase-2 while insulin will activate a protein phosphatase to remove the phosphate group from FBPase-2.

Now that fructose 1,6-bisphosphate is converted to fructose-6-phosphate, it can be isomerized to glucose-6-phosphate using phosphoglucoisomerase just like in glycolysis (just in reverse again). We are sooooo close now to making that delectable glucose, we just need to remove that pesky phosphate group. How do we remove phosphate groups again? Oh yeah through a phosphatase of course! So we just use glucose-6-phosphatase to remove the phosphate group from glucose-6-phosphate and voila we have some nice glucose. This glucose is oftentimes used by cells in different tissues such as the brain or muscles, which is why gluconeogenesis is physiologically necessary.

Here are the key takeaways from gluconeogenesis:

1. It is done to replenish glucose supplies
2. It costs a total of 4 ATP, 2 GTP, and 2 NADH (need two pyruvate to make 1 glucose)
3. It is activated by glucagon but inhibited by insulin
4. Major regulatory step is the PFK step (FBPase-1 and 2)
5. Most of the steps are simply glycolysis but in reverse because many of the enzymes of glycolysis are reversible.

Practice Questions:

1. Type I diabetes is characterized by a person's inability to produce their own insulin. How do you suspect this affects the rate of gluconeogenesis in diabetics? Glycolysis? Explain.
2. How much ATP, GTP, and NADH would be used to create 10 molecules of glucose?
3. Suppose you added radiolabeled CO_2 to a sample of cells actively undergoing gluconeogenesis. Where would this radiolabel end up in glucose?

Answers:

1. If diabetics cannot make their own insulin, that means that the major circulating hormone is glucagon. The body effectively has convinced itself that there is a major shortage of sugar even if there is sugar present and therefore the body is always thinking sugar is not there. Because glucagon is the major circulating hormone, PKA (protein kinase A) will be activated, this protein kinase will phosphorylate FBPase-2 which will activate it. This will cause fructose 2,6-bisphosphate to be destroyed and therefore glycolysis will slow down and gluconeogenesis will speed up.
2. For each molecule of glucose, we need 4 ATP, 2 GTP, and 2 NADH. We want to make 10 glucose molecules from pyruvate and therefore we would need 40 ATP, 20 GTP, and 20 NADH.
3. The only place where CO_2 is incorporated is in the pyruvate carboxylase reaction, however, to make PEP, that CO_2^- group that was added in that reaction must be removed to add the phosphate group and therefore the radiolabel would NOT end up in glucose.

Chapter 25: Pentose Phosphate Pathway

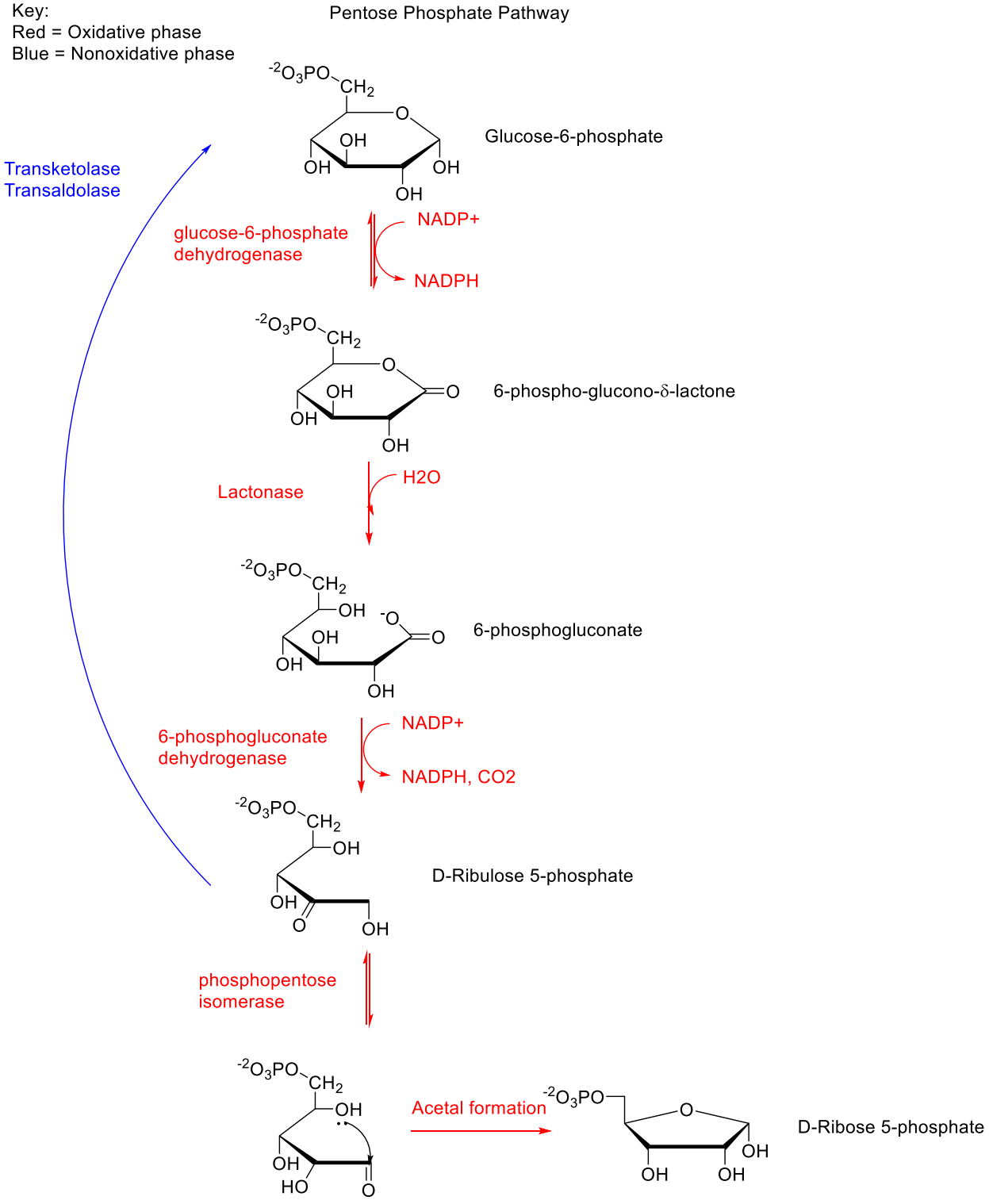
The pentose phosphate pathway (PPP) is an incredibly important pathway for dividing cells. As such, many scientists have thought to inhibit this pathway to target cancer cells (see the cancer focus for more details). Here we will discuss the PPP in great detail and discuss how it is important for nucleotide biosynthesis, fatty acid synthesis, and for controlling oxidative damage. Like all metabolic pathways to this point, we will first discuss the broad goals of the PPP (which we just outlined) before we go into the nitty gritty so to speak. There are two overall goals for the PPP:

1. To create NADPH to power fatty acid synthesis and recycle glutathione
2. To create ribose-5-phosphate to start nucleotide biosynthesis

An overall scheme of the PPP is shown below:

Key:
Red = Oxidative phase
Blue = Nonoxidative phase

Pentose Phosphate Pathway



The PPP occurs in two different phases depending on the needs of the cell. If the cell needs NADPH more than ribose-5-phosphate then the cell will primarily do the nonoxidative phase of PPP. If the cell needs ribose-5-phosphate more than NADPH then the cell will primarily do the oxidative phase (MAS**C**). First we will discuss the oxidative phase since this is the most straight forward part of the pathway then we will discuss the nonoxidative phase.

Glucose-6-phosphate is the start of the PPP, as such hexokinase functions not only as an entry point into glycolysis, but ALSO of PPP, therefore hexokinase is a major target for cancer therapy. Once the glucose-6-phosphate is made, it can be oxidized using glucose-6-phosphate dehydrogenase. Like all dehydrogenases, it makes a reduced cofactor, in this case NADPH. This is consistent with our analysis that C-O bonds being oxidized to C=O gives us NADH or NADPH. Once the oxidation is complete we have made a cyclic lactone based on glucose called 6-phospho-glucono-δ-lactone. Because this is the beginning of the PPP, glucose-6-phosphate dehydrogenase is highly regulated and is allosterically inhibited by NADPH so that cells that do not need NADPH do not do PPP when it is not necessary (MAS**C**).

Once the lactone is made, it can be hydrolyzed to give the linear carboxylic acid and the OH group. This is important because the carboxylic acid will be decarboxylated in the next step. This reaction is catalyzed by lactonase and results in the formation of 6-phosphogluconate.

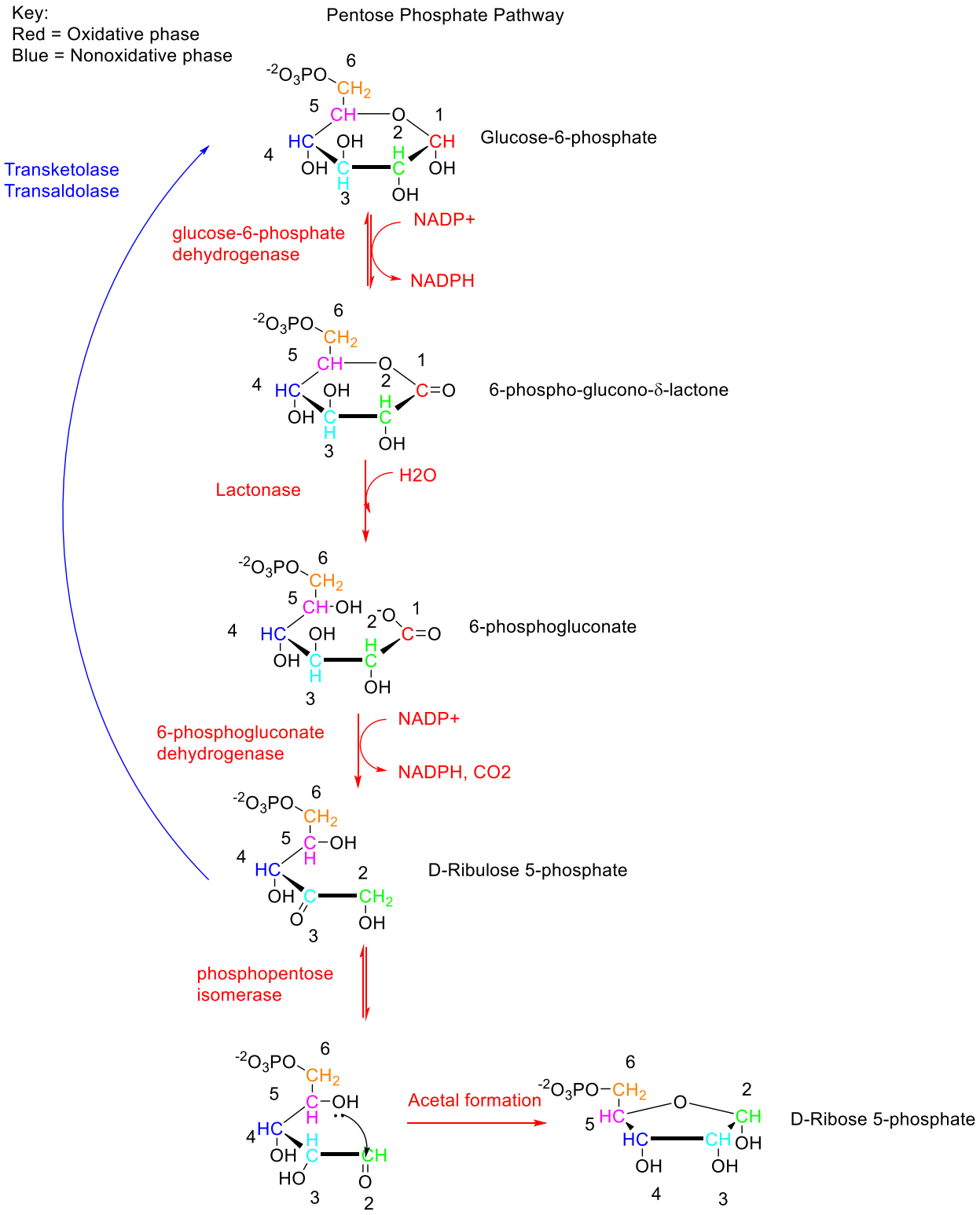
This linear carboxylic acid can be oxidized using 6-phosphogluconate dehydrogenase. We need to remove one carbon because we started with 6 carbons in glucose, but we will ultimately end with 5 carbons in ribose-5-phosphate. To do this, we need to oxidize the carbon that is two carbons away from the carboxylic acid group. Recall from organic chemistry that decarboxylation reactions only occur with 1,3-dicarbonyls. This reaction gives us D-ribulose 5-phosphate and an extra NADPH.

The problem with D-Ribulose 5-phosphate is that if the terminal OH group were to attack the carbonyl to give an acetal, the acetal would be an unstable four-membered ring. To correct for this, phosphopentose isomerase is used to move the carbonyl to the terminal carbon and allow for the acetal to form a much more stable five-membered ring and give the final produce D-Ribose 5-phosphate. This D-Ribose 5-phosphate can be used for DNA nucleotide biosynthesis and therefore it is important for dividing cells who need to replicate their DNA rapidly. The NADPH that is produced can be used to reduce GSSG, the oxidized form of glutathione to 2 GSH, the reduced form. The reduced form of glutathione is able to scavenge free radicals and hydrogen peroxide using glutathione peroxidase and prevent oxidative stress/DNA damage.

The nonoxidative phase of the PPP is a bit more difficult, so we will not cover it in much depth here. Suffice it to say that ribulose 5-phosphate is able to undergo several transketolase and transaldolase reactions to get back to glucose 6-phosphate. These enzymes are able to move carbon chains from different ketose and aldose sugars to give different chain lengths without touching the other functional groups in the molecule. In this way, ribulose 5-phosphate, a five

carbon ketose sugar can be converted in several steps to glucose 6-phosphate, a six carbon aldose sugar.

Provided below is a radiotracing guide to help you determine where radiolabels will end up in the PPP:


Key:
Red = Oxidative phase
Blue = Nonoxidative phase
Pentose Phosphate Pathway
Glucose-6-phosphate
glucose-6-phosphate dehydrogenase
NADP+
NADPH
6-phospho-glucono-δ-lactone
Lactonase
H2O
6-phosphogluconate
6-phosphogluconate dehydrogenase
NADP+
NADPH, CO2
D-Ribulose 5-phosphate
phosphopentose isomerase
Acetal formation
D-Ribose 5-phosphate
Transketolase
Transaldolase

Practice Questions:

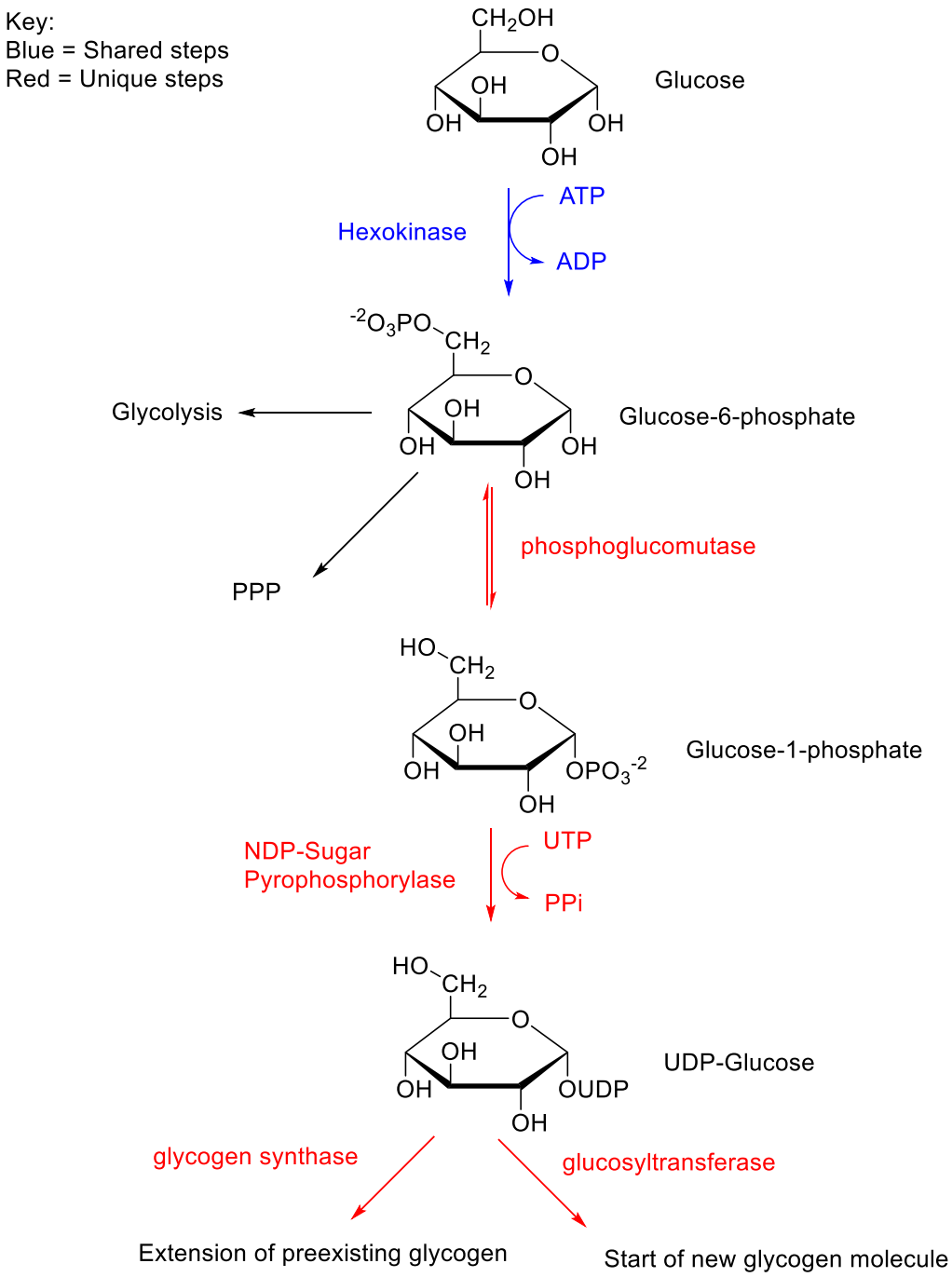
1. Cancer cells are characterized by uncontrolled cell growth. Given the tendency for these cancer cells to continuously divide, which phase of the PPP do you suspect would be the most important for them?
2. Under which circumstance would the PPP be most active? Actively dividing cells low in NADPH, actively dividing cells high in NADPH, nondividing cells low in NADPH, or nondividing cells high in NADPH.
3. If I radiolabeled carbon 1 of glucose, where does that radiolabel end up in D-Ribose 5-phosphate?

Answers:

1. If we need to constantly divide, then we need to constantly make new DNA. Without Ribose 5-phosphate, we cannot do that and therefore I suspect that the oxidative phase of the PPP would be the major part.
2. The PPP would be most active when the cell is actively dividing and has a low NADPH supply. This is because glucose 6-phosphate is allosterically inhibited by NADPH, therefore, if NADPH is low the enzyme will no longer be inhibited and therefore the pathway will go faster.
3. Look at the chart, carbon 1 leaves as CO_2. It does not appear anywhere in Ribose 5-phosphate.

Chapter 26: Glycogen Biosynthesis

Sometimes the cell doesn't need the glucose it has and instead wants to store it. The glucose storage unit on the cellular level is its glycogen supply. Glycogen is a long polymer of glucose that contains α1,4 and branched α1,6 linkages. It is a highly branched polymer and this allows it to maximize its surface area and allow for efficient glycogen breakdown (MA**SC**). In this chapter, we will discuss the steps that the cell must take to make glycogen from scratch and how the cell adds on glucose units to a preexisting glycogen polymer. Like all metabolic pathways, we need to start our discussion on when it would be logical to store glucose in the form of glycogen. Glycogen's only role is to store glucose, therefore, we should not be making glycogen if we have no glucose and we should not be making glycogen if we have no energy. Therefore, this pathway should be active under the opposite conditions, that is to say, high glucose and high energy (ATP and NADH). An overview of the glycogen synthesis pathway is shown below, just as with the other pathways we will discuss each step in detail and figure out the chemical logic behind why it works:



Luckily, this is one of the shorter pathways because it shares a step with glycolysis and the PPP, which is the hexokinase reaction. Once glucose-6-phosphate is made, it has to be isomerized using phosphoglucomutase. This will place the phosphate from the 6 carbon to the 1 carbon. This may seem trivial, but it is important because all of the linkages (α1,4 and the branched α1,6) in glycogen start from the 1 carbon and therefore we need to have a good leaving group on the 1 carbon to facilitate that chemistry (**M**A**S**C).

It turns out that this leaving group is not quite good enough to allow for the full range of glycogen synthesis chemistry and therefore the glucose needs to be labeled with UDP, a much better leaving group. This reaction is catalyzed by NDP-Sugar pyrophosphorylase and makes UDP-glucose. This UDP-glucose can be used either to make a new glycogen molecule via combining with glycogenin or it can be used to extend a pre-existing glycogen molecule using glycogen synthase. The decision here has to do with the needs of the cell. If we have no glycogen, then it needs to combine with glycogenin using glucosyltransferase to make a new glycogen molecule (MAS**C**). Conversely, if we have a glycogen molecule that isn't at its full length yet, we will use glycogen synthase to add the glucose in an α1,4 linkage. If we want to add a branchpoint, we would simply use the branching enzyme.

Glycogen synthesis has two major regulatory steps, that is the hexokinase step and the glycogen synthase step. This is because both of these are irreversible reactions and are commitment steps in the glycogen synthesis pathway.

Hexokinase we have already discussed the regulation of, it is stimulated by glucose levels and it is inhibited by glucose-6-phosphate levels. Hormonally, it is activated by insulin for two reasons. First, insulin directly activates hexokinase and second, insulin causes more glucose receptors (GLUT4 specifically) to be on the cell membrane so that it is easier to get glucose into the cell. This should make sense because insulin tells the body that there is a lot of glucose available in the blood and it would be a waste not to use it (MAS**C**), therefore processes that will break down this glucose and store what remains are logical to perpetuate under these circumstances.

Glycogen synthase is also regulated hormonally by insulin. Glycogen synthase is deactivated when it is phosphorylated, therefore insulin will inhibit GSK3, glycogen serine kinase 3, the protein kinase responsible for phosphorylating and shutting down glycogen synthase. Insulin also promotes PP1, protein phosphatase 1, which will remove the phosphate groups from glycogen synthase. Insulin, therefore, is a double whammy, it will inhibit the kinase and promote the phosphatase. Consistent with its role of storing excess glucose, glycogen synthase is indirectly activated by glucose and glucose-6-phosphate levels, both of which activate PP1 and it is inhibited by glucagon and epinephrine which both deactivate PP1. Glucagon, as we mentioned, tells the body that there is a lack of glucose, which is why we would want to prevent glycogen synthesis. Epinephrine on the other hand indicates that there is an urgent matter at hand, it is more commonly known as adrenaline. When adrenaline is coursing through your veins, that indicates to the body that we need to make energy and get the heck out of there as fast as possible. Under these circumstances, it would be illogical to waste energy making glycogen (MAS**C**) and therefore it makes sense that adrenaline would inhibit glycogen synthesis.

A summary of the regulators of this pathway are summarized below:

Molecule	Hexokinase	Glycogen synthase
Glucose	Activate	Activate
Glucose-6-phosphate	Inhibit	Activate
Insulin	Activate	Activate
Glucagon	Inhibit	Inhibit

Here are the key takeaways from glycogen synthesis:

1. It occurs when there is excess glucose in the cell and in the blood (insulin levels are high).
2. It costs 1 ATP and 1 UTP per glucose
3. It shares the hexokinase step with glycolysis and PPP
4. Glucose-6-phosphate must be isomerized to glucose-1-phosphate to allow for the α1,4 and branched α1,6 linkages to occur in glycogen.

Practice Questions:

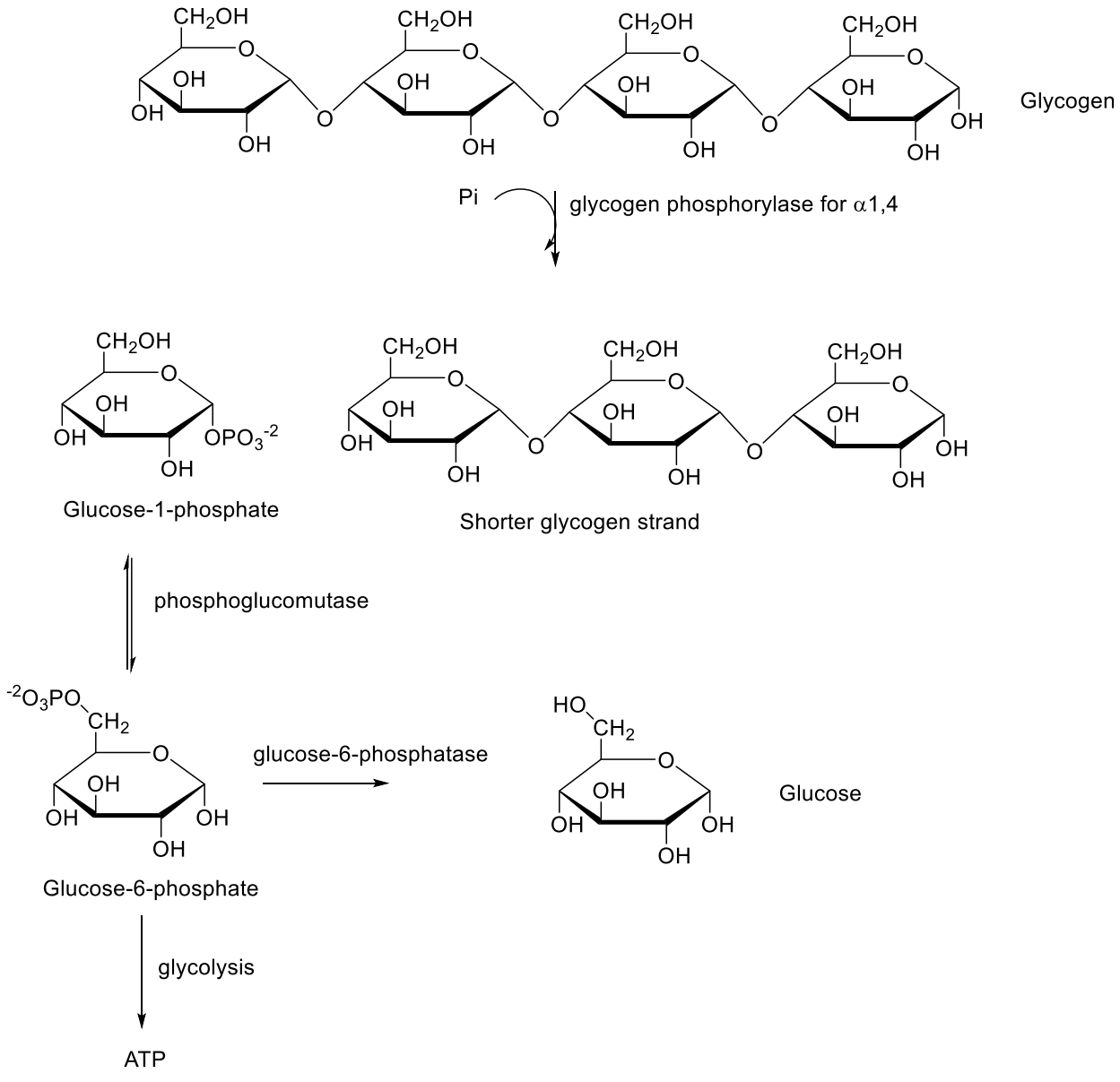
1. Type 1 diabetes is characterized by the inability to produce insulin. How would this disorder affect glycogen synthesis levels? Explain using logic and on a molecular level.
2. Suppose PP1 is mutated so that it is perpetually turned on, how would this mutation affect glycogen synthesis levels? What about GSK3?
3. Some cancer cells overexpress insulin receptors on their surface. How would this affect glycogen synthesis levels?

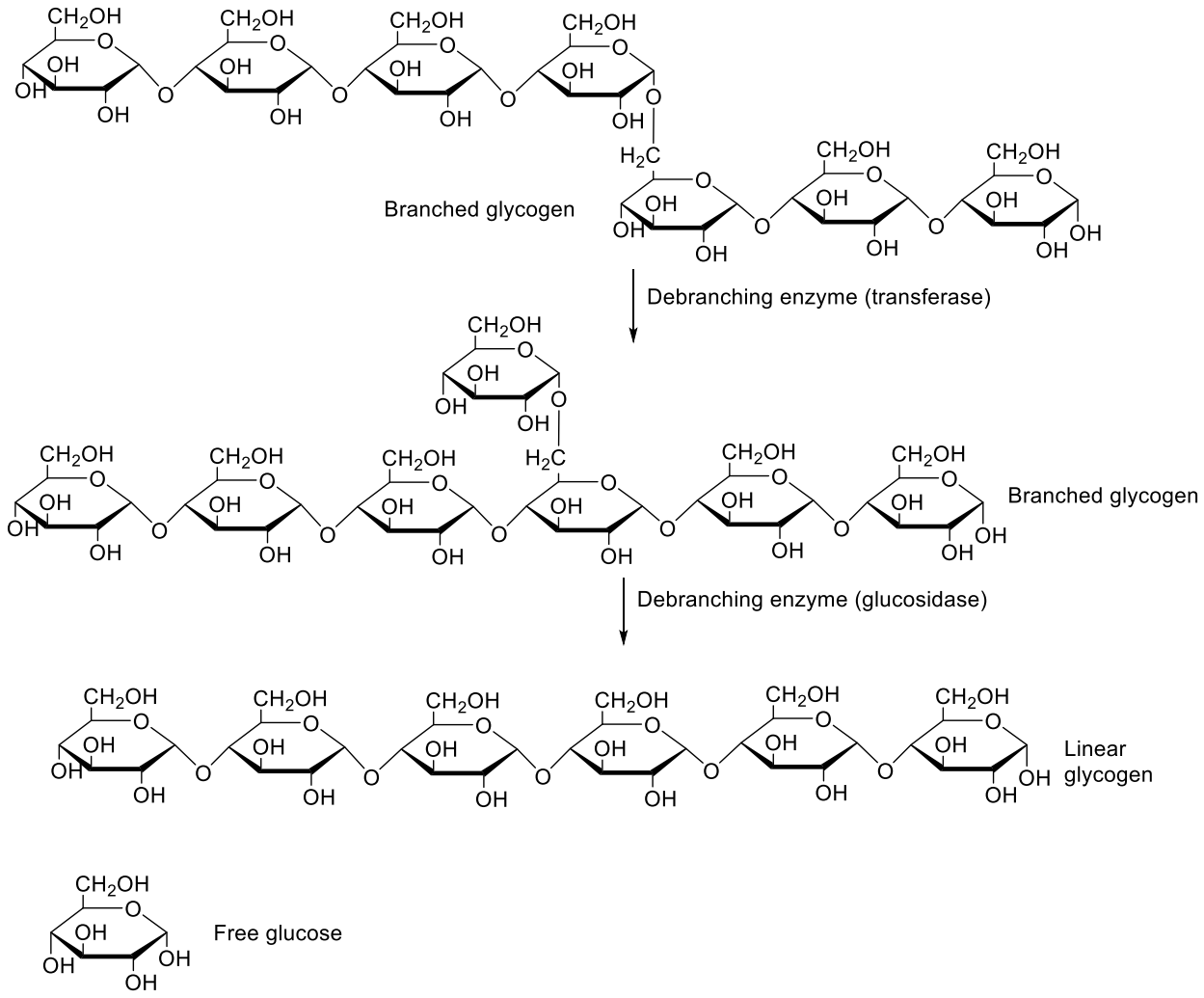
Answers:

1. If you cannot produce insulin then the body would perpetually think it is in a low sugar state and therefore glycogen synthesis would not occur to an appreciable extent. Mechanistically, without insulin, the major circulating hormone would be glucagon, which will deactivate PP1, which is the protein responsible for increasing the activity of glycogen synthase. Glucagon also inhibits hexokinase expression which would therefore decrease glycogen synthesis levels.
2. If PP1 is permanently turned on, then it would constantly remove the phosphates from glycogen synthase and therefore glycogen synthase would be permanently turned on as well. This would overall increase glycogen synthesis levels. GSK3 would have the opposite effect, namely, that if it were overactive it would constantly phosphorylate and deactivate glycogen synthase, thereby decreasing overall glycogen synthesis levels.
3. If there are more insulin receptors on the surface of the cell, then the cell would be much more sensitive to insulin levels. The end result would be that the cell would think it is in a high-sugar environment and therefore glycolysis and glycogen synthesis would be upregulated (increased). Recall that insulin will activate PP1 and deactivate GSK3, both of these effects cause glycogen synthesis to increase overall.

Chapter 27: Glycogen Breakdown

Now that we discussed how to make glycogen, we can discuss how to break it down. Like with all metabolic processes, we have to consider what the overall goal is and when we would want to do this. Glycogen is the glucose storage place in the cell, as such, it can be used to give the cell glucose when none is available in the blood or the cell is otherwise in need of energy. As such, we would want to break down glycogen when glucose is not present in the blood (glucagon levels high) and when the cell needs energy (low ATP and NADH). An overview of glycogen breakdown is shown below:





Glucose is released from glycogen using glycogen phosphorylase, which will cleave the α1,4 glycosidic bond and create one glucose-1-phosphate and a shortened glycogen molecule. This glucose-1-phosphate can be isomerized to glucose-6-phosphate, which can be used for glycolysis or PPP. Glycogen also has branches due to its α1,6 linkages. To remove those branches, the cell uses the debranching enzyme. This enzyme transfers over the last three glucose molecules in the branch to the nonreducing sugar end of the glycogen chain and then cleaves the α1,6 linkages to give a free glucose molecule, NOT glucose-1-phosphate as shown above.

To avoid breaking down glycogen at the same time as you are making it, glycogen phosphorylase, like glycogen synthase is hormonally regulated. Unlike glycogen synthase, however, glycogen phosphorylase is activated by phosphorylation. Glucagon and epinephrine(adrenaline) activate glycogen phosphorylase by activating its kinase, phosphorylase b kinase. Insulin on the other hand will activate protein phosphorylase 1 (PP1) and will remove the phosphate from glycogen phosphorylase and therefore deactivate it. In this way, the

breakdown of glycogen is controlled based off the availability of sugar so that when sugar is high, glycogen is made and when sugar is low glycogen is broken down (MAS**C**).

Here are the key takeaways from the glycogen breakdown pathway:

1. Glycogen phosphorylase breaks α1,4 linkages and releases glucose-1-phosphate.
2. The debranching enzyme removes three glucose residues from the branch, adds those glucoses to the main chain, then releases free glucose from glycogen by breaking α1,6 linkages.
3. Glycogen phosphorylase is the main regulatory enzyme in the glycogen breakdown pathway. It is activated by glucagon and adrenaline and deactivated by insulin.

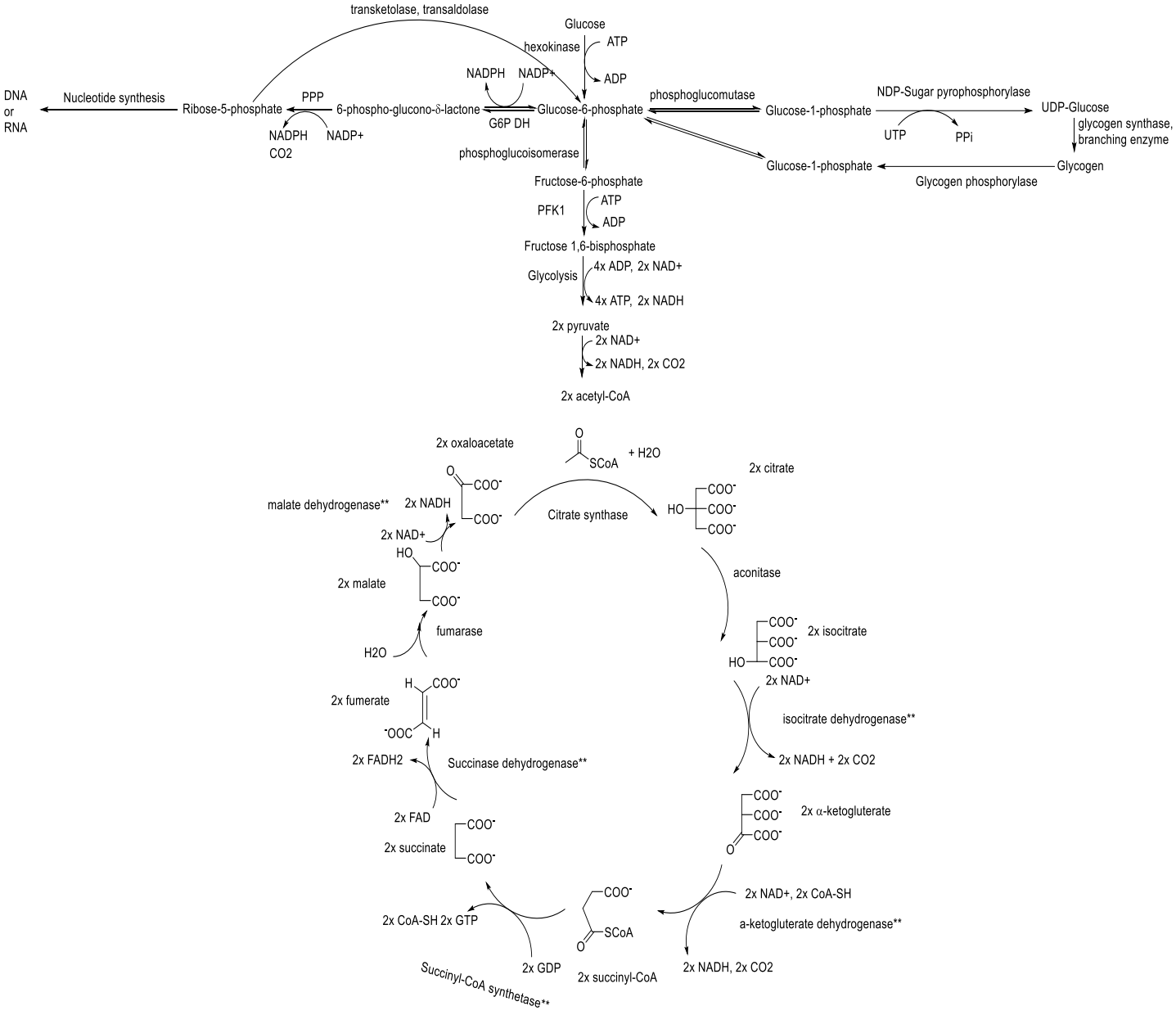
Practice Questions:

1. Name the ideal conditions under which glycogen breakdown would occur the fastest in terms of molecule concentration and energy needs of the cell.
2. Which of the following conditions would glycogen breakdown occur fastest: you are being chased by a murderer or you just had a large Italian feast and are in bed.

Answers:

1. Low glucose and low energy. Low glucose means that the cells need to replenish their sugar supplies and low energy means the cell needs to mobilize their glucose reserves so that they can undergo glycolysis and aerobic respiration to make ATP.
2. If you are being chased by a murderer, you (usually) would have a huge spike in adrenaline, therefore, the adrenaline would activate the phosphorylase b kinase and that would activate glycogen phosphorylase.

Macroscopic Overview of Carbohydrate Metabolism Overall:

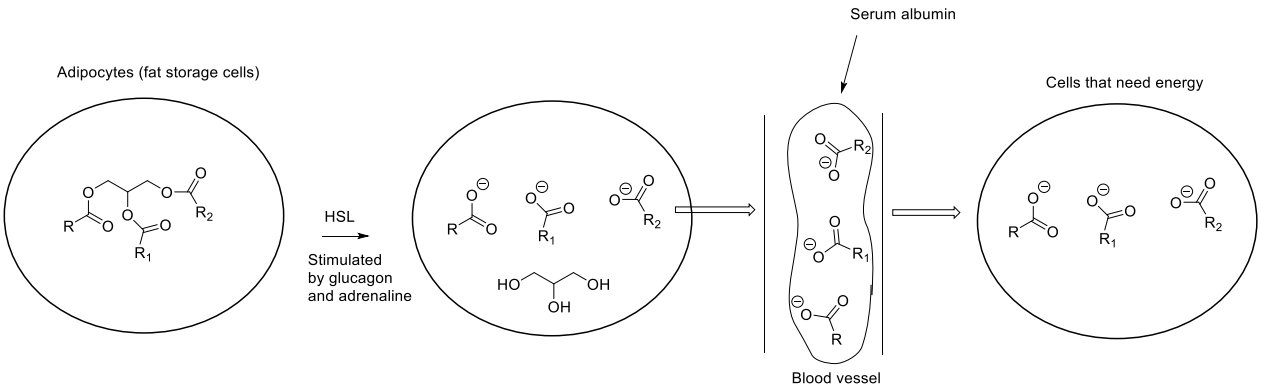


Focus 6: Lipid Metabolism

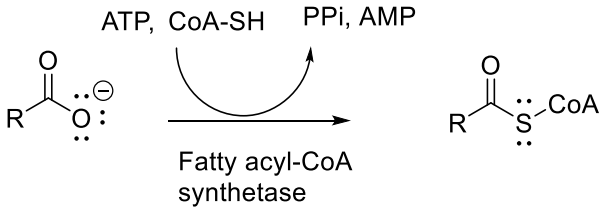
Chapter 28: Beta Oxidation

Now that we have discussed all of carbohydrate metabolism, we will turn our attention to fat metabolism. When there is no more glucose to use for energy, your body turns to fats to make energy. One of the premier pathways for using fat to make energy is called beta oxidation and this process occurs in the mitochondrial matrix.

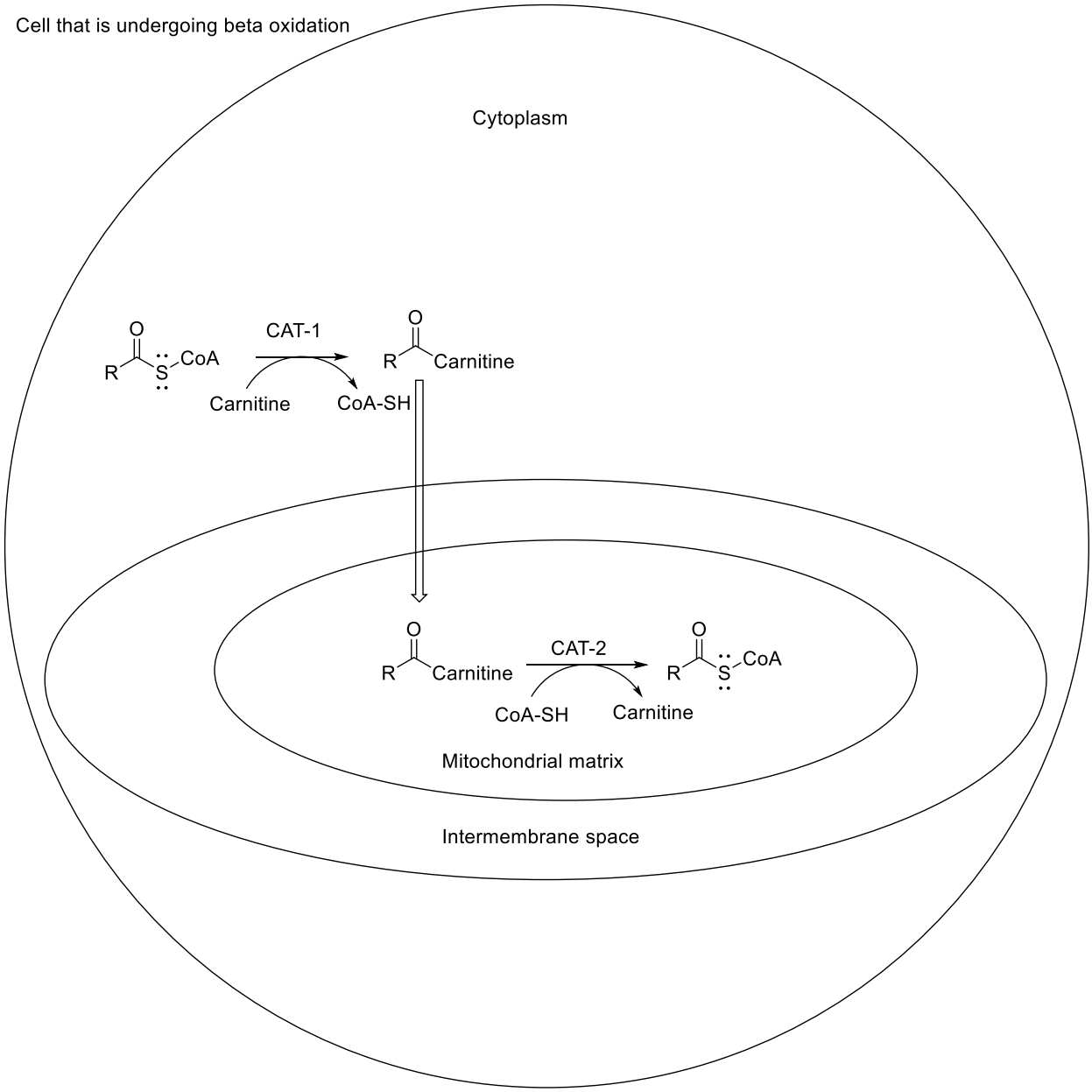
Fat in the form of triacylglycerides is stored in specialized cells called adipocytes. These adipocytes store fat for long-term energy storage, however, they lack the machinery necessary to metabolize these fats to make ATP. Effectively adipocytes is to liver cells as triacylglycerides are to glycogen. So, the fats that are in adipocytes need to be mobilized to cells that can use them to make energy. The way that fatty acids are transported to tissues is shown below:



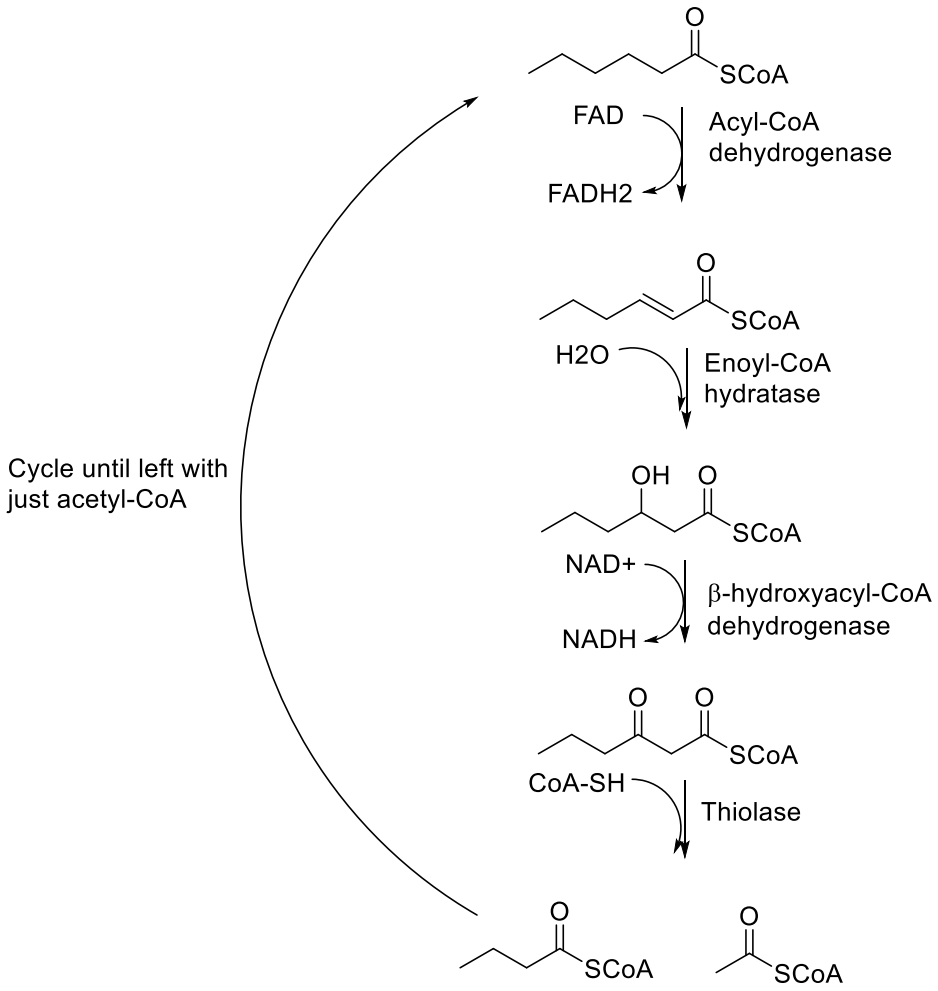
HSL, or hormone sensitive lipase, is the enzyme that is responsible for freeing up fatty acids from triacylglycerides. Once the fatty acids are removed from the glycerol backbone, they are mobilized into the blood, where they hitch a ride on serum albumin that shuttles them off to cells that need energy (liver cells for instance since these do a lot of beta oxidation). Once these fatty acids are in the cell, they need to be activated to fatty acyl-CoAs using ATP (M**A**SC). This reaction is catalyzed by fatty acyl-CoA synthetase in the reaction shown below:



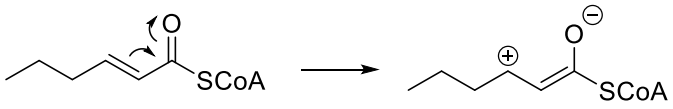
Once they are activated into fatty acyl-CoAs they can be shuttled to the mitochondria using the carnitine transport system, which is summarized below:



This is effectively a two-step process involving the two CAT enzymes (CAT-1 and CAT-2). First, the fatty acyl-CoA needs to be made into a fatty acyl-carnitine because only the carnitine version can be transported into the mitochondria (we will explain why once we discuss fatty acid synthesis). This reaction is catalyzed by CAT-1. Once the fatty acyl-carnitine is made it is moved into the mitochondrial matrix, where the second step unfurls. This involves making the fatty acyl-CoA back from the carnitine analogue using CAT-2. Beta oxidation has an ABSOLUTE requirement of using fatty acyl-CoAs, therefore the fatty acyl-carnitine must be converted back to the CoA analogue before beta oxidation will occur. Now, with the fatty acyl-CoA in the mitochondria, beta oxidation can proceed rapidly. The steps of beta oxidation are shown below:



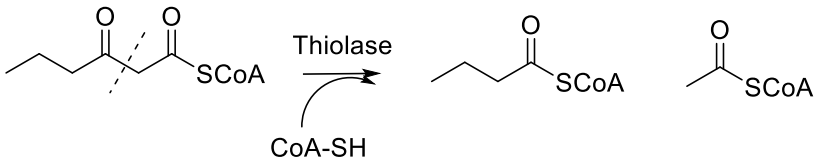
The first reaction oxidizes the C-C bond between the α and β carbons using acyl-CoA dehydrogenase. Because we are oxidizing a C-C bond, we produce an $FADH_2$ in this step. The trans pi bond we just formed is important because it creates an α-β unsaturated thioester. This allows the water to add to the β position in the next step because if you recall from Organic Chemistry, the β position is electrophilic and will be attacked by weak bases due to the resonance form shown below:



The addition of water across the pi bond we made is catalyzed by enoyl-CoA hydratase and this places an OH group on the beta position stereospecifically to create the L isomer.

Now that we have the OH group in the BETA position, what do you think we are going to do next in BETA OXIDATION? Well oxidize it of course! But if we are oxidizing that carbon, we would be oxidizing a C-O bond, and therefore we would expect that the β-hydroxyacyl-CoA dehydrogenase that catalyzes this oxidation to produce an NADH, which it does. See, I don't lie to you.

Finally, to remove the acetyl-CoA that we made, we use thiolase to cleave the bond between the two carbonyl carbons and we add a SCoA group to the carbonyl carbon that gets cut like so:



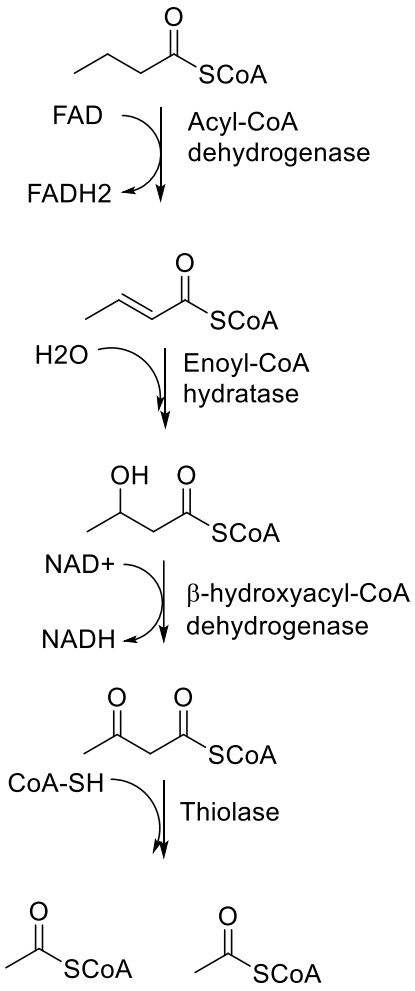
Therefore, for any even saturated fatty acid of carbon length, n, we make the following acetyl-CoA, NADH, and $FADH_2$:

$$acetyl - CoA = \frac{n}{2}$$

$$NADH = \frac{n}{2} - 1$$

$$FADH_2 = \frac{n}{2} - 1$$

The minus 1 fudge factor for NADH and $FADH_2$ comes in because the very last acetyl-CoA molecule does not get made through oxidation, but instead gets made via the thiolase reaction which does not produce either NADH or $FADH_2$. This can be shown visually, if we consider the beta oxidation of a four carbon fatty acid:



See? We produced two acetyl-CoAs, but only 1 NADH and 1 $FADH_2$. We can also calculate the ATP yield starting from the free fatty acid down to the acetyl-CoA molecules we make by using the relation that every acetyl-CoA gives 10 ATP equivalents (1 GTP, 1 $FADH_2$, and 3 NADH). This gives us the following formula:

$$ATP_{total} = ATP_{NADH} + ATP_{FADH_2} + ATP_{acetylCoA} - 1$$

$$ATP_{total} = 2.5\left(\frac{n}{2} - 1\right) + 1.5\left(\frac{n}{2} - 1\right) + 10\left(\frac{n}{2}\right) - 1$$

$$ATP_{total} = \frac{2.5n}{2} - 2.5 + \frac{1.5n}{2} - 1.5 + \frac{10n}{2} - 1$$

$$ATP_{total} = \frac{14n}{2} - 5$$

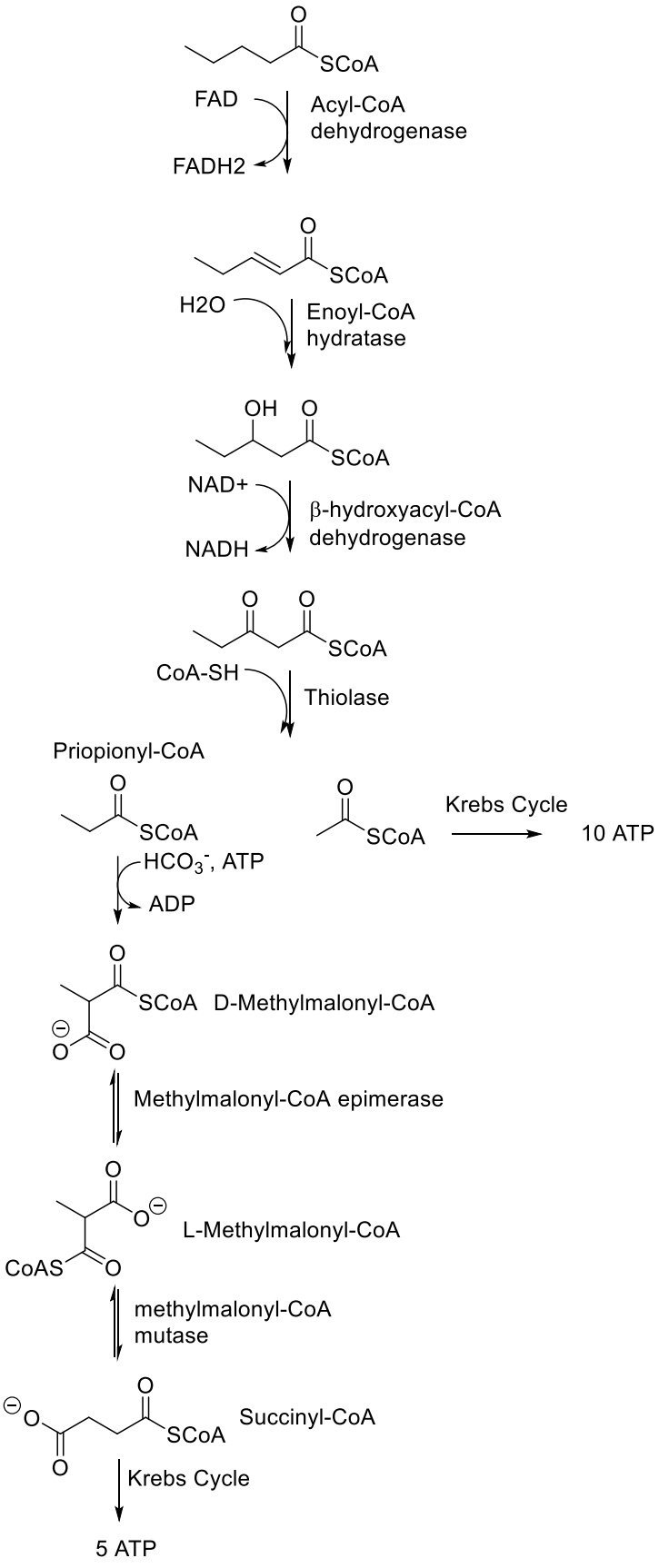
$$ATP_{total} = 7n - 5$$

The minus 1 fudge factor here is from the fact that we had to CoAylate the fatty acid using fatty acyl-CoA synthetase and that cost us 1 ATP. These formulas can be used for any even saturated fatty acid, but that begs the question, what happens if we have an unsaturated fatty acid and what

happens if we have an odd fatty acid? We will get to both of those questions shortly. We will start with the unsaturated question first since that is by far the easiest to explain.

If you have an unsaturated fatty acid, then you already have a pi bond in between two carbons in the fatty acid. However, nature usually makes cis double bonds rather than trans double bonds, which are required for the next step of beta oxidation, so a special enzyme has to make the pi bond trans. This enzyme is called enoyl-CoA isomerase. For the pi bonds that are not in the correct positions (are not between the alpha and beta carbons), enoyl-CoA reductase reduces those pi bonds to sigma bonds using NADPH. In either case, the pi bond that was between the alpha and beta carbons prevents acyl-CoA dehydrogenase from producing $FADH_2$ (the pi bond is already there), so we lose 1 $FADH_2$ overall.

If we have an odd-numbered fatty acid, everything proceeds as normal, except now instead of producing n/2 acetyl-CoAs, we produce one less because the very last carbon chain is three carbons long. This is propionyl-CoA and this is carboxylated using priopionyl-CoA carboxylase at the expense of ATP to make succinyl-CoA after isomerization, which can be used to feed into the Krebs cycle. This is not preferred because succinyl-CoA enters the Krebs cycle after two NADH are already produced so we lose 5 ATP from this (relative to acetyl-CoA). This process is shown below:


SCoA
FAD
FADH2
Acyl-CoA dehydrogenase
SCoA
H2O
Enoyl-CoA hydratase
OH
SCoA
NAD+
NADH
β-hydroxyacyl-CoA dehydrogenase
SCoA
CoA-SH
Thiolase
Priopionyl-CoA
SCoA
SCoA
Krebs Cycle
10 ATP
HCO3-, ATP
ADP
SCoA
D-Methylmalonyl-CoA
Methylmalonyl-CoA epimerase
L-Methylmalonyl-CoA
CoAS
methylmalonyl-CoA mutase
Succinyl-CoA
SCoA
Krebs Cycle
5 ATP

The key takeaways from beta oxidation are as follows:

1. Beta oxidation occurs primarily in the liver and in other tissues when glucose supplies are very low.
2. Fat is mobilized from adipocytes, which are the fat storage areas of the body, using HSL. Activated by glucagon.
3. Beta oxidation creates a lot of NADH, $FADH_2$, and acetyl-CoA to feed into the Krebs cycle and the ETC but very little ATP directly.

Practice Questions:

1. Explain mechanistically why diabetics have a higher rate of beta oxidation than nondiabetics.
2. What cofactor does propionyl-CoA carboxylase have? (hint: recall E_1 of the PDH complex).
3. How many NADH, $FADH_2$, and acetyl-CoA molecules would be made from a 22 carbon fatty acid? How many ATP equivalents is that? What if the fatty acid was 21 carbons long? How would that change it? What if the 22 carbon long fatty acid had one unsaturation between the alpha and beta carbons? Explain.
4. Suppose you isolate mitochondria from a mutated cell line that has an unknown mutation. You noticed that when the cells were cultured, they rapidly died off when glucose supplies were low, so you suspected that there may be something wrong with their beta oxidation pathway. When you supplied their mitochondria with fatty acyl-CoAs, no ATP was produced, however, fatty acyl-carnitine was detected in the mitochondrial matrix. What enzyme(s) could be mutated to cause this? How do you know?
5. Suppose you did the same experiment as before except with a different mutated cell culture. This time regardless of the type of fatty acid you provide to them (fatty acid or fatty acyl-CoA) no ATP is produced and no fatty acyl-carnitine is detected. Which enzyme(s) could be mutated to cause this? How do you know?
6. Suppose you did the exact same experiment as before on a third mutated cell culture. This time, when you added the fatty acyl-CoA, ATP is produced, but when you add the free fatty acid, no ATP is produced. What enzyme(s) could be mutated to cause this? How do you know?
7. Suppose you mutated enoyl-CoA hydratase to make the D hydroxyl acyl-CoA instead of the L hydroxyacyl-CoA. How do you suspect this would influence ATP yield, NADH yield, and $FADH_2$ yield from beta oxidation? Why?
8. What enzyme is similar to acyl-CoA dehydrogenase that we have discussed so far? (Hint: think Krebs cycle).
9. What enzyme is similar to enoyl-CoA hydratase that we have discussed so far? (Hint: think Krebs cycle).
10. What about β-hydroxyacyl-CoA dehydrogenase? Same hints as before.

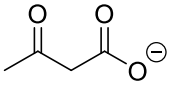
Answers:

1. In diabetics, the major circulating hormone is glucagon, therefore, glucagon is constantly signaling to adipocytes to release fatty acids into the blood stream, which will get shuttled to other tissues for beta oxidation.
2. Biotin. All carboxylases use biotin as a cofactor, an example of another carboxylase that uses biotin is E_1 of PDH.
3. 11 acetyl-CoA, 10 NADH, 10 $FADH_2$. These equate to 150 ATP for the 22 carbon long saturated fatty acid. If the carbon chain was instead 21 carbons long, then we would make 9 acetyl-CoA, 1 propionyl-CoA, 9 NADH, and 9 $FADH_2$. This equates to 9(10) + 1(5) + 9(2.5) + 9(1.5) ATP equivalents or 131 ATP. If there was one unsaturation, then we would lose 1 $FADH_2$ from the original total, so the total would now be 11 acetyl-CoA, 10 NADH, and 9 $FADH_2$. This would equate to 148.5 ATP. This is because with one unsaturation, we would lose out on the first step of beta oxidation that makes one $FADH_2$.
4. If fatty acyl-carnitine is detected in the mitochondrial matrix, then that means CAT-2 is mutated and no longer functions. This is because without CAT-2, the carnitine cannot be replaced with SCoA and therefore beta oxidation cannot occur. Without beta oxidation, no ATP can be produced, which is the exact scenario we see here.
5. If no fatty acyl-carnitine is detected, that means CAT-1 is mutated and no longer functions because without CAT-1, the fatty acid or fatty acyl-CoA cannot enter the mitochondrial matrix and therefore no beta oxidation = no ATP.
6. If when fatty acyl-CoA is added ATP is produced but not when fatty acids are added, that indicates that beta oxidation works when fatty acyl-CoA is present but not when the fatty acid is there. If this is the case that means CAT-1, CAT-2, and all of the beta oxidation enzymes are working, but the preparatory stage does not work. This means that the fatty acyl-CoA synthetase does not function anymore due to the mutation and therefore the fatty acids cannot be made into fatty acyl-CoAs which we know do work to produce ATP.
7. Enzymes are stereospecific, if the wrong stereoisomer is produced then the next enzyme down the line, in this case, β-hydroxyacyl-CoA dehydrogenase, will no longer recognize the substrate and therefore the chemistry will not occur. Therefore, regardless of the length of the fatty acyl-CoA supplied to the mithchondria, the NADH, $FADH_2$, and acetyl-CoA yield would be 0, 1, and 0 respectively. It would get stuck in the β-hydroxyacyl-CoA form and no further chemistry could happen.
8. Acyl-CoA dehydrogenase is very similar to succinate dehydrogenase because they both make trans alkenes and produce $FADH_2$.
9. Enoyl-CoA hydratase is very similar to fumerase because they both add OH groups to the alkenes that were there.
10. β-hydroxyacyl-CoA dehydrogenase is very similar to malate dehydrogenase because they both produce NADH and they both oxidize an OH group at the beta position.

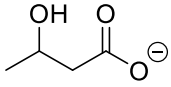
Chapter 29: Ketone Bodies

Some tissues cannot do beta oxidation. One example of cells that cannot do beta oxidation are neurons in the brain. For those cells, glucose is their main energy source, so if a person does not consume glucose and those cells are starved of it, they need some way of getting energy. Thus, the ketone body system has evolved (MAS**C**). I personally like to think of ketone bodies as mobilized acetyl-CoA equivalents because that is effectively what they are. There are three different ketone bodies that are produced and they are shown below:

Acetone



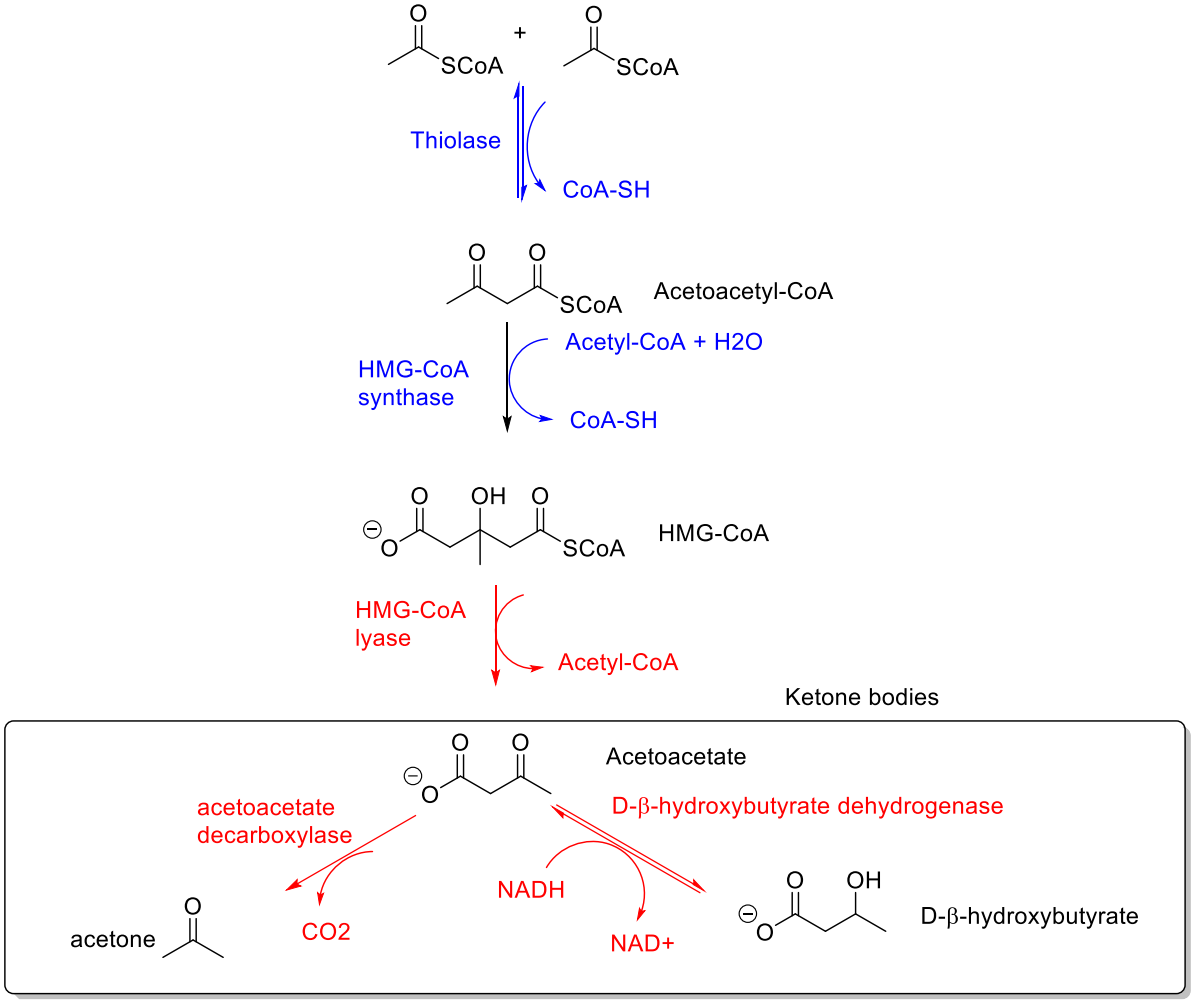
Acetoacetate



D-β-hydroxybutyrate

These ketone bodies are packaged versions of acetyl-CoA that can be transported to tissues that are unable to perform beta oxidation by themselves, but that have a working Krebs cycle that allows them to make ATP from acetyl-CoA. This process primarily occurs in the liver and an overview of the ketone body forming reactions are shown below:

Key:
Blue = Shared steps with cholesterol synthesis
Red = Unique steps



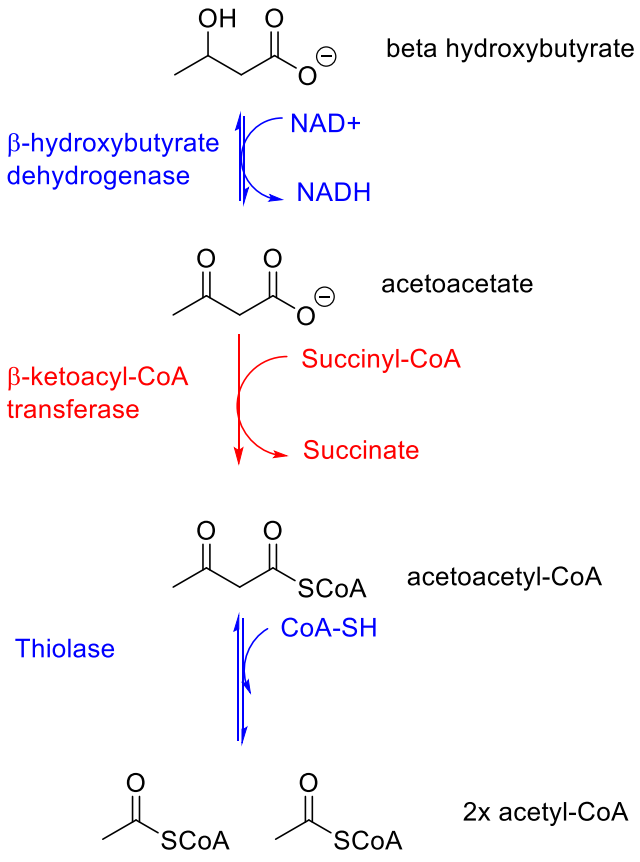
Acetone is not useful for energy production because it lost the CO_2^- group and therefore it is exhaled, but the other two ketone bodies can be moved into the blood stream and transported to the brain for energy production. Now we will go through the chemical logic and steps for each of the ketone body reactions. The very first reaction is a reaction that you have seen before. It uses thiolase to combine two acetyl-CoA molecules together to form acetoacetyl-CoA, this is effectively the reverse reaction of the last step in beta oxidation. This is done to remove one of the thioesters. Recall that thioesters are incredibly reactive carboxylic acid derivatives because thiols are great leaving groups, because of their high reactivity, thioesters would quickly react with proteins and water in the bloodstream and would not reach their destination. Therefore, it is important that this process remove thioesters and replace them with unreactive bonds like C-C bonds.

The next reaction produces HMG-CoA, which is a major intermediate for cholesterol synthesis (details to come). The HMG-CoA is made using the aptly named HMG-CoA synthase by using another molecule of acetyl-CoA and water to hydrolyze off one of the thioesters. This adds two carbons to the chain and makes a free OH group beta to either carbonyl. This is very useful because now when the HMG-CoA lyase is used, it produces the desired acetoacetate ketone body and releases an acetyl-CoA. This acetoacetate can be decarboxylated to give acetone, which will be exhaled, or reduced to D-β-hydroxybutyrate. Now the ketone bodies can be exocytosed (excreted) from the cell and released into the blood stream, where they will be transported to the brain. These ketone bodies are incredibly acidic and if enough of them are present in the blood, one can develop ketoacidosis, which can be life threatening.

Once the ketone bodies reach their destination, they can be converted back to two acetyl-CoA molecules and the tissues that received them can now make energy. This is done using the reactions described below, many of which are simply the same reactions we just learned but in reverse:

Key:
Blue = shared reactions from formation
Red = unique reactions

In the brain



The only unique reaction in the delivery reactions is the transferase reaction, which uses a Krebs cycle intermediate, succinyl-CoA, to transfer its CoA group to the acetate group of acetoacetate. This is necessary because the CoA group allows thiolase to break the product acetoacetyl-CoA back to its two acetyl-CoA component pieces. These two acetyl-CoA molecules can be used in the Krebs cycle to make ATP.

Key takeaways from the ketone body pathway:

1. It is only used when tissues such as the brain run out of glucose and need energy badly
2. It shares many of the same reactions as the cholesterol synthesis pathway (HMG-CoA being where they diverge)
3. Ketone bodies are mobilized acetyl-CoA delivery units
4. Ketone bodies are acidic and are released into the blood (can cause ketoacidosis)
5. Acetone is exhaled when ketone bodies are actively being produced

Practice Questions:

1. Suppose you are at your first day of rotations as a physician and a patient comes in presenting with fatigue and mild cognitive impairment. Blood tests show their pH is normal (i.e. within normal limits). Their blood sugar is also incredibly low. What pathway in their body is not functioning properly and how could you tell without testing their blood pH? (This is probably not accurate to real life, I am in no way a medical doctor. I am just a dude who likes biochemistry).
2. Which of the following situations would a person most likely produce the most ketone bodies: well fed and resting, well fed and running, starving and resting, or starving and running?
3. Over the years, many fad diets have come in and out of favor. One such fad diet is the keto diet, in which a person does not eat any carbohydrates. One possible side effect of the ketogenic diet is people’s breath stinks (yes even when they brush their teeth). Explain from a biochemical perspective why this is the case.

Answers:

1. The ketone body pathway is likely not working properly because their brain is not making energy (hence the cognitive impairment). This is corroborated with their incredibly low blood sugar levels. However, we know that this pathway is not working because the pH of the blood is normal, if this pathway was working as intended, the blood would likely be at least mildly acidic. If you did not test their blood pH you can probably try to detect acetone in their breath since usually people do not exhale acetone except for when they are producing ketone bodies.
2. If you are producing ketone bodies, you would not be well fed, instead you would be closer to starving and if you really need ketone bodies you would have to be doing

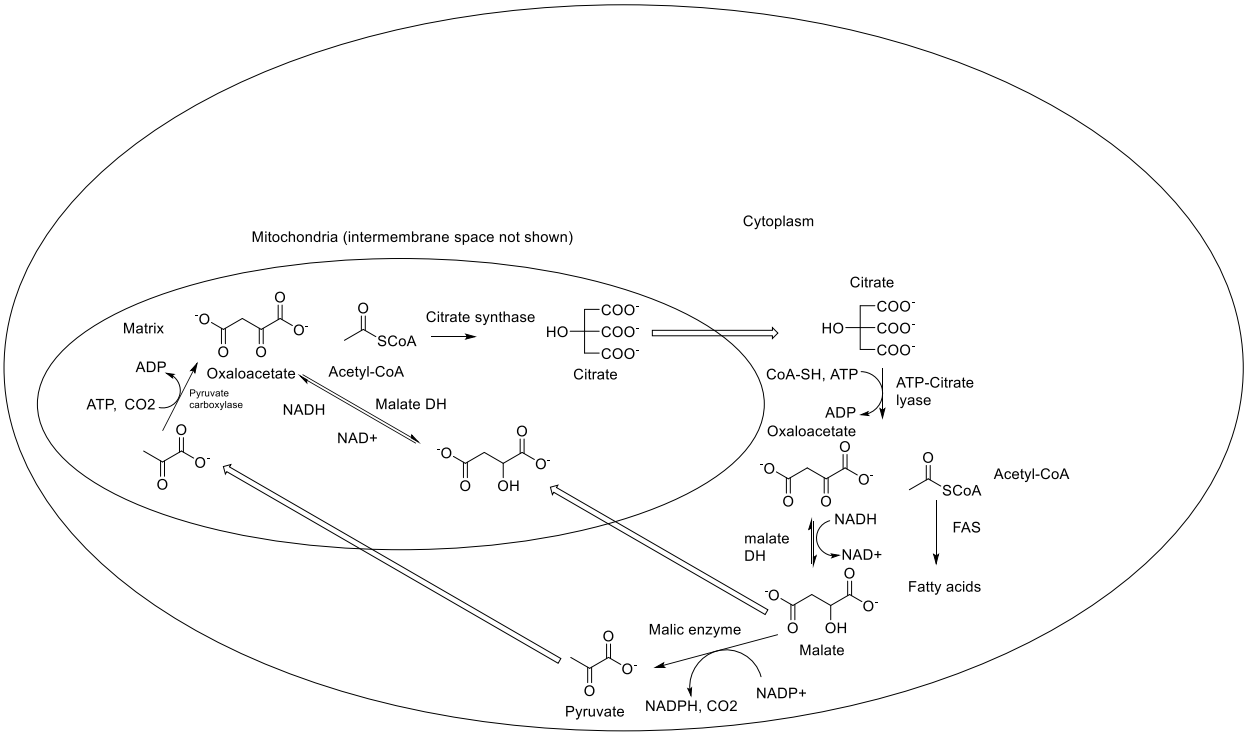
something that requires energy. Therefore, starving and running should give you the most amount of ketone bodies.

3. Acetone stinks, therefore, the ketogenic diet causes the body to have no glucose in the brain and therefore the liver needs to go into overdrive pumping out ketone bodies left and right. This translates to the breath of the person smelling of acetone (which is a ketone body), which is not pleasant. Moral of the story, fad diets can have some bad consequences.

Chapter 30: Fatty Acid Biosynthesis

Fats are incredibly important for cellular function. They serve as long-term energy storage, compose the cell membrane, and serve as water-repellant surfaces. As such, the ability of a cell to make its own fats is incredibly important. Here, we will discuss how cells make their own fatty acids. Luckily, this process is incredibly similar to the beta oxidation pathway and at the end of this chapter we will do a side-by-side comparison of the two processes. This process is primarily done if fatty acids are low, energy is high, or if the cell is actively dividing. Like all synthetic processes, this requires a lot of energy and therefore the cell will only do it if it has a surplus of ATP (MAS**C**).

In beta oxidation we went from fatty acids to acetyl-CoA, now with fatty acid synthesis, we are going from acetyl-CoA to fatty acids. Fatty acid synthesis occurs in the cytoplasm, NOT in the mitochondrial matrix, however. This poses a challenge because the cell's acetyl-CoA supplies are localized solely in the matrix of the mitochondria. This begs the question how do we move the acetyl-CoA from the mitochondria to the cytoplasm? It turns out that we can use the citrate shuttle. The citrate shuttle is described visually below and we will go through each step to explain how it works:



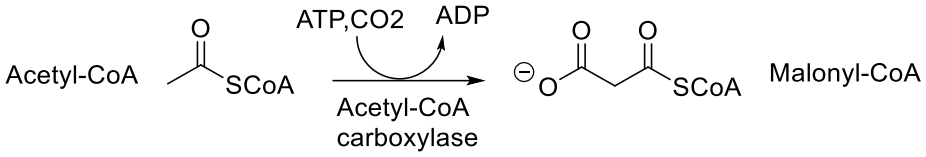
Now we will explain why we said before that citrate is important for fatty acid synthesis. When there is an excess amount of citrate, that indicates to the cell that it has a lot of energy and can therefore use that energy to make fatty acids. The first reaction of the citrate shuttle is making citrate using citrate synthase, no surprises there. Once this citrate is made, it can actually leave the mitochondria and enter the cytoplasm. Remember that many of the enzymes in the Krebs cycle are inhibited by citrate so if there is a lot of citrate, those enzymes do not work anymore

and so citrate will accumulate in the mitochondria. Citrate will only move to the cytoplasm if there is a surplus of it.

Once citrate is in the cytoplasm, it can be made back into acetyl-CoA and oxaloacetate using ATP-Citrate lyase in what is effectively the citrate synthase reaction but in reverse. Remember that making citrate was in irreversible step of the Krebs cycle, so we NEED ATP to power the reverse reaction (M**A**SC). This cytosolic acetyl-CoA can be used to make fatty acids via fatty acid synthase (FAS), which we will cover soon. The oxaloacetate, however, is stuck in the cytoplasm unless it is reduced back to malate using malate dehydrogenase. This is very similar to the malate shuttle seen in the first step of gluconeogenesis. Malate, as we saw in gluconeogenesis, can freely move into and out of the mitochondria and so it can directly move back into the mitochondria and get reoxidized to oxaloacetate using malate dehydrogenase to complete the cycle. If the cell needs NADPH to power fatty acid synthesis, however, malic enzyme can be used to make malate into pyruvate, which can also freely move into and out of the mitochondria. Once the pyruvate is in the mitochondria, it can be carboxylated using pyruvate carboxylase at the expense of ATP to make oxaloacetate again and complete the cycle once more (M**A**SC).

The net ATP loss is 1 ATP per acetyl-CoA not using the malic enzyme and is 2 ATP per acetyl-CoA using the malic enzyme. Because the malic enzyme pathway uses more ATP, it is only used if the cell needs NADPH. We will see later that NADPH provides the reducing power to make the fatty acid during fatty acid synthesis. The net NADH loss is 0 NADH per acetyl-CoA without malic enzyme and is 1 NADH per acetyl-CoA with it. Effectively we are swapping an NADH for an NADPH when we use the malic enzyme pathway at the expense of 1 ATP.

Now that we have the acetyl-CoA in the cytoplasm, we can get to making fatty acids. Most of the acetyl-CoA's that are used during fatty acid synthesis need to be activated by carboxylation using acetyl-CoA carboxylase in the reaction shown below:



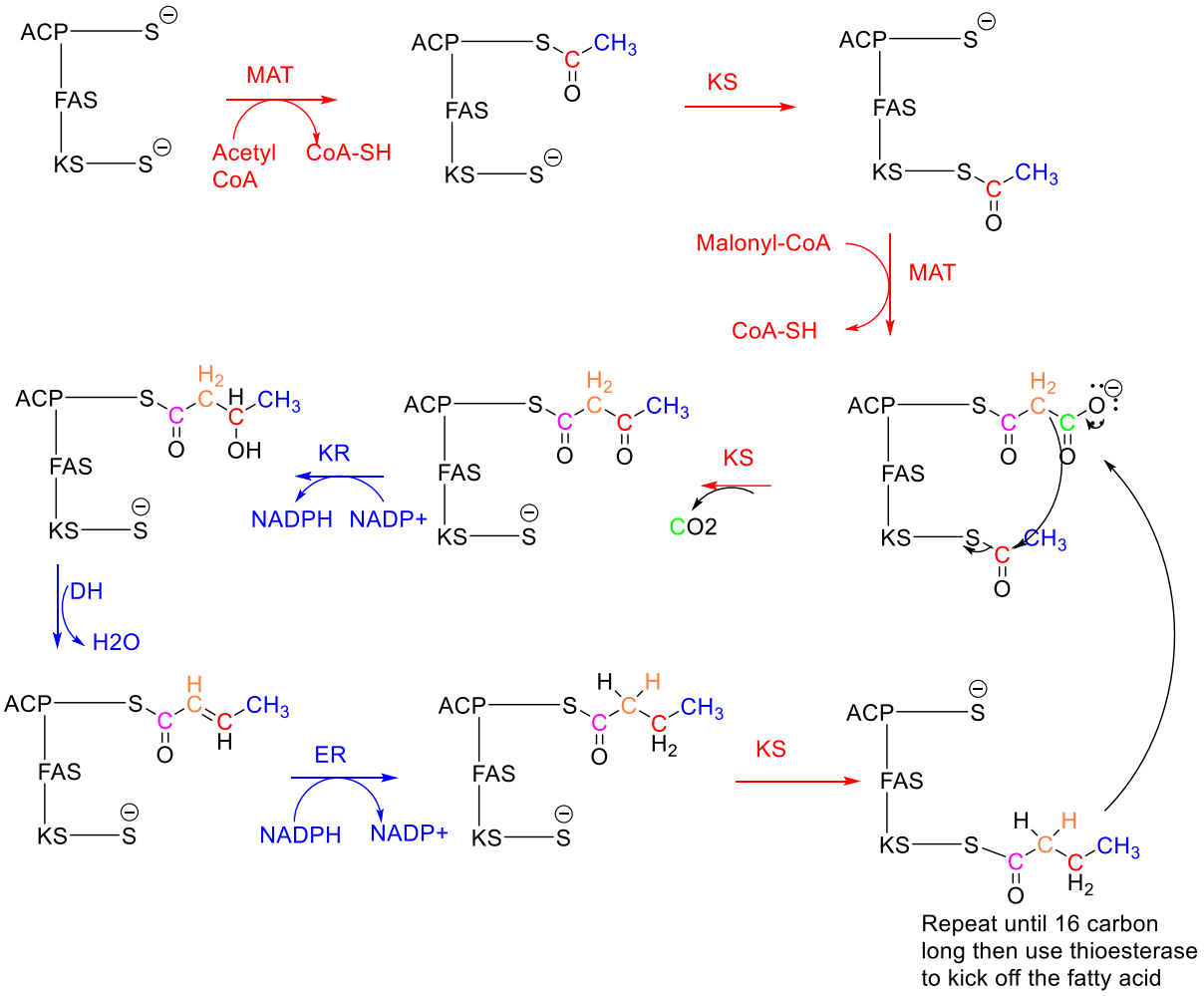
The one exception to this is the first acetyl-CoA, which DOES NOT need to be activated and can instead be added directly to the fatty acid synthase enzyme. Like all carboxylases, acetyl-CoA carboxylase uses biotin as a cofactor.

Because malonyl-CoA is only used for fatty acid synthesis and this reaction is irreversible, this is a commitment step and therefore acetyl-CoA carboxylase is highly regulated. Acetyl-CoA carboxylase is allosterically activated by citrate but inhibited by palmitoyl-CoA. This should makes sense because excess citrate in the cytoplasm indicates that fatty acid synthesis is necessary while palmitoyl-CoA is the product of fatty acid synthase and therefore if there is an excess amount of this then there is no reason to continue making it (MAS**C**). In addition to malonyl-CoA being used to make fats, it is also an inhibitor of CAT-1, preventing beta oxidation.

Acetyl-CoA carboxylase is also regulated hormonally in the way that you would expect; glucagon and adrenaline turn it off while insulin turns it on. Glucagon indicates that blood sugar is low and therefore beta oxidation needs to occur to make energy. To avoid a futile cycle (making fatty acids at the same time as you break them down), glucagon activates PKA and this will phosphorylate acetyl-CoA carboxylase and turn it off. Adrenaline works the same way as glucagon, this should make sense because if you are in a fight or flight situation, you need energy and because fatty acid synthesis uses up ATP, you would not want to do fatty acid synthesis (MAS**C**). Adrenaline will also activate PKA and in turn, acetyl-CoA carboxylase will be phosphorylated and deactivated. Conversely, insulin indicates to the cell that there is a lot of glucose and energy levels are therefore high, so insulin will facilitate fatty acid synthesis by activating protein phosphatase 2A, which will dephosphorylate acetyl-CoA carboxylase and turn it on (MAS**C**). In this way, the cell can control the production of fatty acids depending on the energy demands of the cells.

The fatty acid synthase (FAS) enzyme is an enormous protein complex that is comprised of several different subunits, each with their own catalytic functions. There are two main components of the fatty acid synthase enzyme; these are the acyl carrier protein (ACP) and keto synthase (KS) domains. While the fatty acid is being synthesized, it is connected to a phosphopantetheine linker that creates another thioester. Therefore, placing any acetyl-CoA or malonyl-CoA onto the fatty acid synthase does not require any ATP or energy because it is simply replacing one thioester for another. To add carbons to the fatty acid synthase, the malonyl/acetyl-CoA ACP transferase (MAT) subunit is used. The KS domain is used to extend the carbon chain of the fatty acid. The rest of the subunits of the fatty acid synthase enzyme do the same reactions seen in beta oxidation but in reverse. An overview of the fatty acid synthase chemistry is shown below (diagram based off Palmer's Biochemistry textbook):

Key:
Blue = Shared steps
Red = Unique steps



First, the acetyl-CoA needs to be added onto the ACP, this is done using MAT. But to make way for the incoming carbons, KS is used to move the acetyl group from the ACP to the KS domain. Then MAT adds the malonyl group to the ACP. This malonyl group has a carboxyl group at the end that can decarboxylate because the compound is a β-carboxylate thioester. When the carboxyl group leaves, that allows the sigma bond between the orange and green carbon to break, allowing those electrons to attack the nearby thioester in the KS domain. This extends the carbon chain length by 2 every time and minimizes energy by breaking the thioester bond in the KS domain and releasing CO_2 in the process (**M**ASC). After this, the steps are effectively just beta oxidation but in reverse. The only difference here is that NADPH is always used regardless of if you are reducing a C=O bond or C=C bond. The β-hydroxythioester is also in the D configuration rather than the L configuration that is seen in beta oxidation. Once the carbon chain is completely reduced, it is moved back to the KS domain and the process is repeated until

a certain carbon chain length is reached, then thiolase is used to hydrolyze the thioester and release the fatty acid.

Now we can generalize to get the total NADPH and ATP used for any fatty acid of length n.

$$ATP = ATP_{citrate} + ATP_{malonylCoA}$$

$$ATP = \frac{n}{2} + \frac{n}{2} - 1$$

$$ATP = n - 1\ lost$$

$$NADPH = NADPH_{KR} + NADPH_{ER}$$

$$NADPH = 2(\frac{n}{2} - 1)$$

$$NADPH = n - 2\ lost$$

For the ATP used, we have to consider that for every acetyl-CoA we use 1 ATP to transport it from the mitochondria to the cytoplasm and we have to consider that for every acetyl-CoA except the first one that we use 1 ATP to make the malonyl-CoA. Each acetyl-CoA gives us two carbons, therefore, to translate that to the carbon count, n, we simply divide n by two to get us the number of acetyl-CoAs. Therefore $ATP_{citrate}$ is n/2 because each acetyl-CoA takes up 1 ATP. Likewise, $ATP_{malonyl\text{-}CoA}$ is n/2 – 1 because we exclude the first acetyl-CoA to this count (hence the -1). That gets us n-1 overall after combining the fractions.

For the NADPH used, we use 2 NADPH after the first acetyl-CoA is added. Therefore, it is two times (the number of acetyl-CoA – 1). That gets us n-2 overall after combining the fractions.

Here is a side-by-side comparison between fatty acid synthesis and beta oxidation:

Characteristic	Fatty acid synthesis	Beta oxidation
Location	Cytoplasm	Mitochondrial matrix
Configuration of beta hydroxy intermediate	D	L
ATP produced/consumed	Consumed	Produced
Redox cofactors produced/consumed	Consumed (NADPH)	Produced (NADH, $FADH_2$)
Activated by hormone *blank*	Insulin	Glucagon/adrenaline
Transport system	Citrate shuttle	Carnitine shuttle
Carbon chain activated by	Carboxylation	CoAylation
Number of enzymes	1	4

Here are the key takeaways from the fatty acid synthesis pathway:

1. It is done to make fatty acids, the main component of cell membranes
2. Insulin increases fatty acid synthesis while glucagon and adrenaline decrease it

3. Fatty acid synthesis uses the same kind of chemical logic as beta oxidation, but it catalyzes the reverse reactions.
4. Fatty acid synthesis is incredibly expensive
5. For fatty acid synthesis to occur, acetyl-CoA needs to be transported to the cytoplasm via the citrate shuttle

Practice Questions:

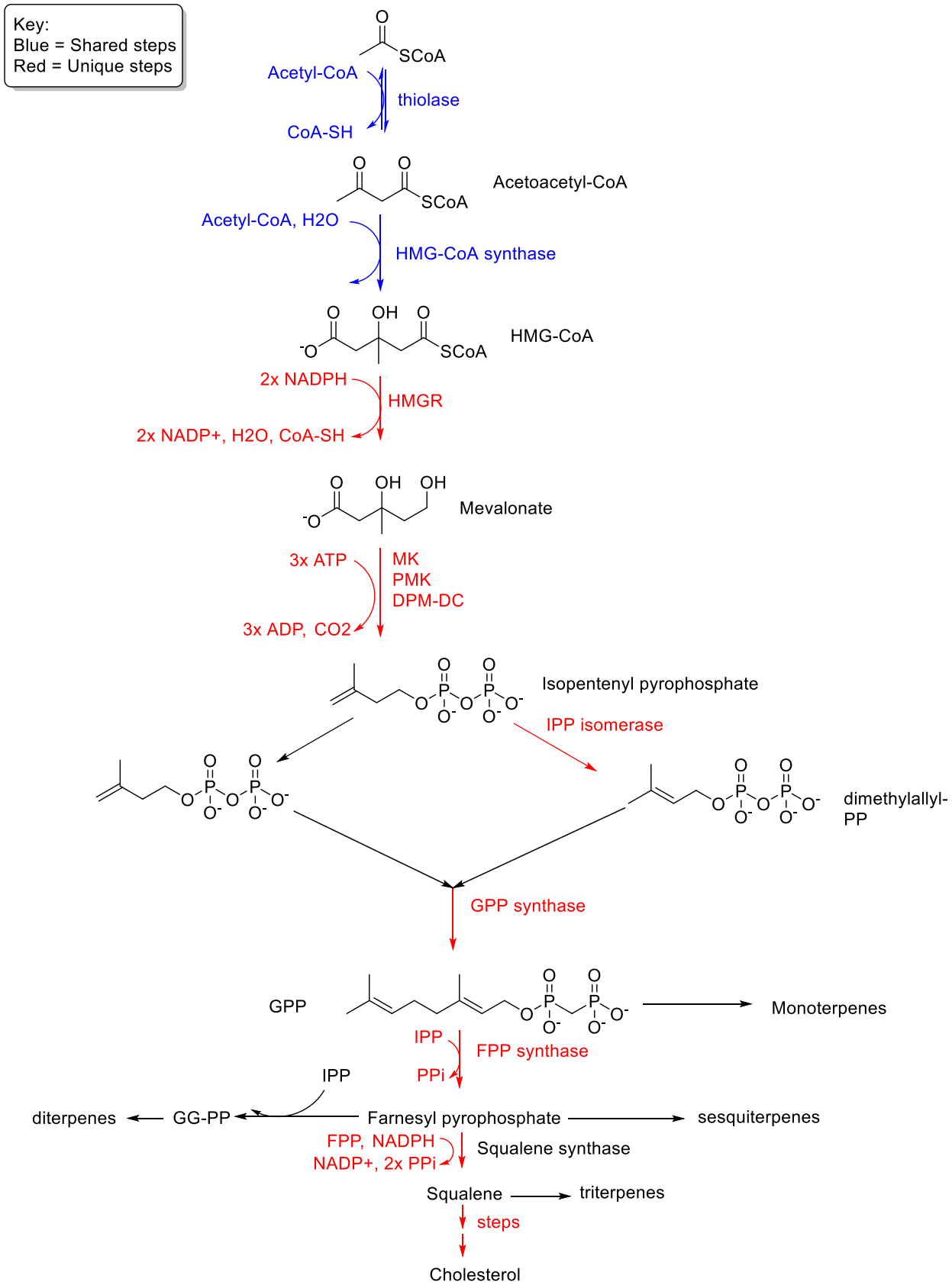
1. Through what pathways is NADPH produced?
2. Cancer is characterized by rapid cellular growth. Therefore, one would expect that fatty acid synthesis should be _____(increased/decreased) in cancer cells. Because of this, cancer drugs could be targeted against _____(enzyme name) to _____ (activate/deactivate) fatty acid synthesis.
3. During what situation would fatty acid synthesis most likely occur at the fastest rate: well fed and exercising, starving and exercising, well fed and resting, or starving and resting?
4. When blood sugar levels are high _____ (hormone) is released. _____ (hormone) binds to its receptor and activates _____ (protein), which _____ (phosphorylates/dephosphorylates) acetyl-CoA carboxylase. In doing so, it ____ (increase/decrease) fatty acid synthesis.
5. How many ATP are consumed in making a 14 carbon fatty acid from acetyl-CoA found in the mitochondria? NADPH?
6. Explain how beta oxidation and fatty acid synthesis have evolved to avoid a futile cycle.

Answers:

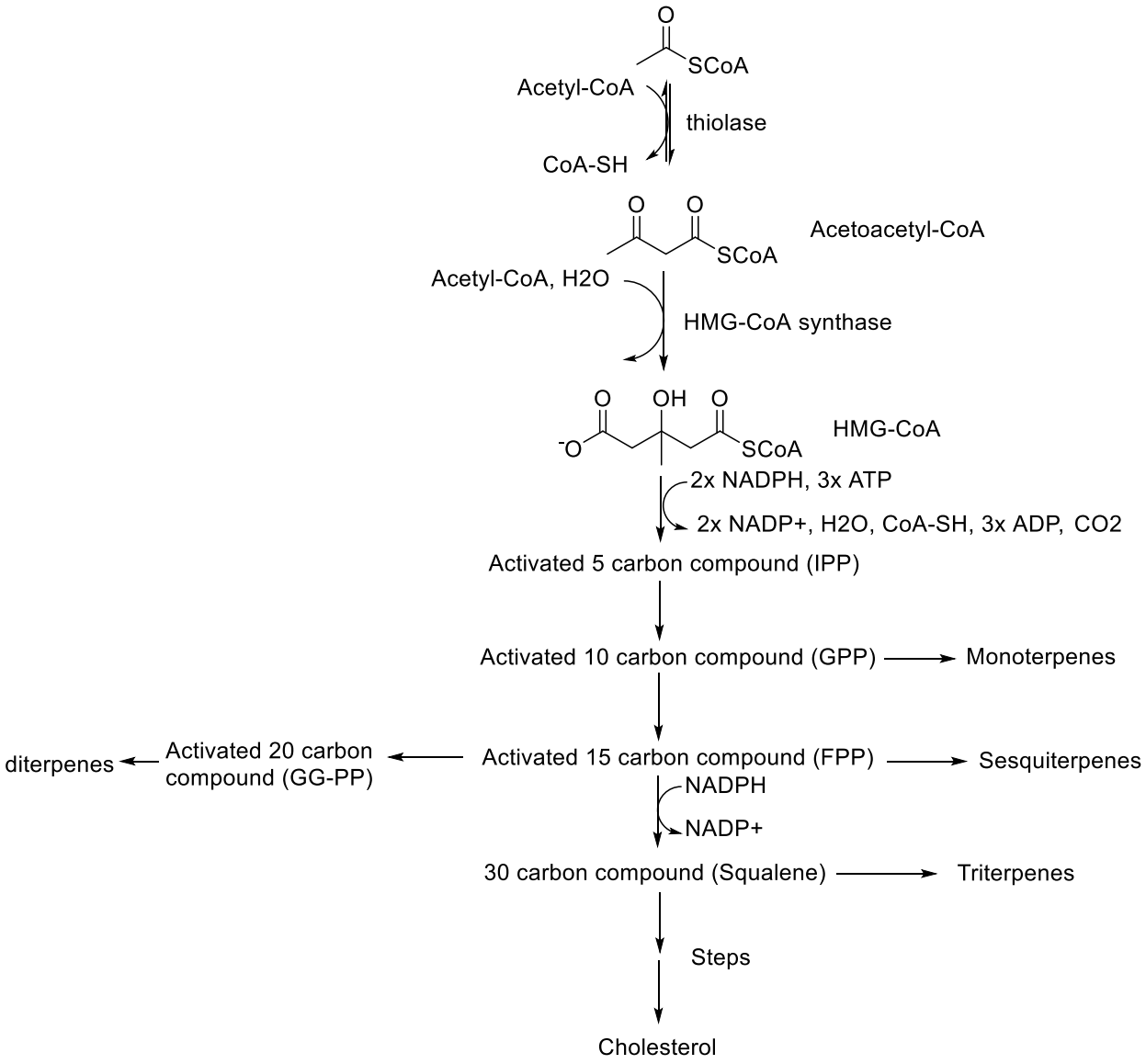
1. The PPP and malic enzyme produce NADPH to power fatty acid synthesis.
2. Increase, acetyl-CoA carboxylase, deactivate.
3. Well fed and resting. Well fed means that your blood sugar levels are high, promoting insulin to circulate in the blood. The insulin then binds its receptor and activated a protein phosphatase to remove the phosphate from acetyl-CoA carboxylase, thereby activating it and stimulating fatty acid synthesis. Resting indicates a low energy demand and therefore glucagon and epinephrine/adrenaline would not get released to an appreciable extent.
4. Insulin, Insulin, protein phosphatase 2A, dephosphorylates, increases.
5. 13 ATP lost, 12 NADPH lost.
6. Beta oxidation and fatty acid synthesis occur in entirely different places (mitochondria versus cytoplasm) and the product of acetyl-CoA carboxylase is malonyl-CoA, which blocks the transfer of fatty acyl-CoAs from the cytoplasm to the mitochondria via CAT-1.

Chapter 31: Cholesterol Biosynthesis

Cholesterol is incredibly important to maintain the structural stability of cell membranes and is a precursor to a wide variety of steroid hormones and bile acids. As such, it is important that cells can make their own cholesterol. Here, we will describe the process by which cells make cholesterol along with many other natural products (terpenes). Some of the reactions you have already learned because of the Ketone Bodies chapter. Like with the other metabolic pathways, we start with the overall goal of the process. Here, the overall goal is to make cholesterol and other nonpolar natural products. This pathway is subject to a tremendous amount of study because high cholesterol as we will discuss later is one of the leading causes of heart disease in the United States and therefore inhibition of these enzymes could help fix this issue. An overview of the cholesterol synthesis pathway is shown below:


Key:
Blue = Shared steps
Red = Unique steps
Acetyl-CoA
thiolase
CoA-SH
Acetoacetyl-CoA
Acetyl-CoA, H2O
HMG-CoA synthase
HMG-CoA
2x NADPH
HMGR
2x NADP+, H2O, CoA-SH
Mevalonate
3x ATP
MK
PMK
DPM-DC
3x ADP, CO2
Isopentenyl pyrophosphate
IPP isomerase
dimethylallyl-PP
GPP synthase
GPP
Monoterpenes
IPP
FPP synthase
PPi
IPP
diterpenes
GG-PP
Farnesyl pyrophosphate
sesquiterpenes
FPP, NADPH
NADP+, 2x PPi
Squalene synthase
Squalene
triterpenes
steps
Cholesterol

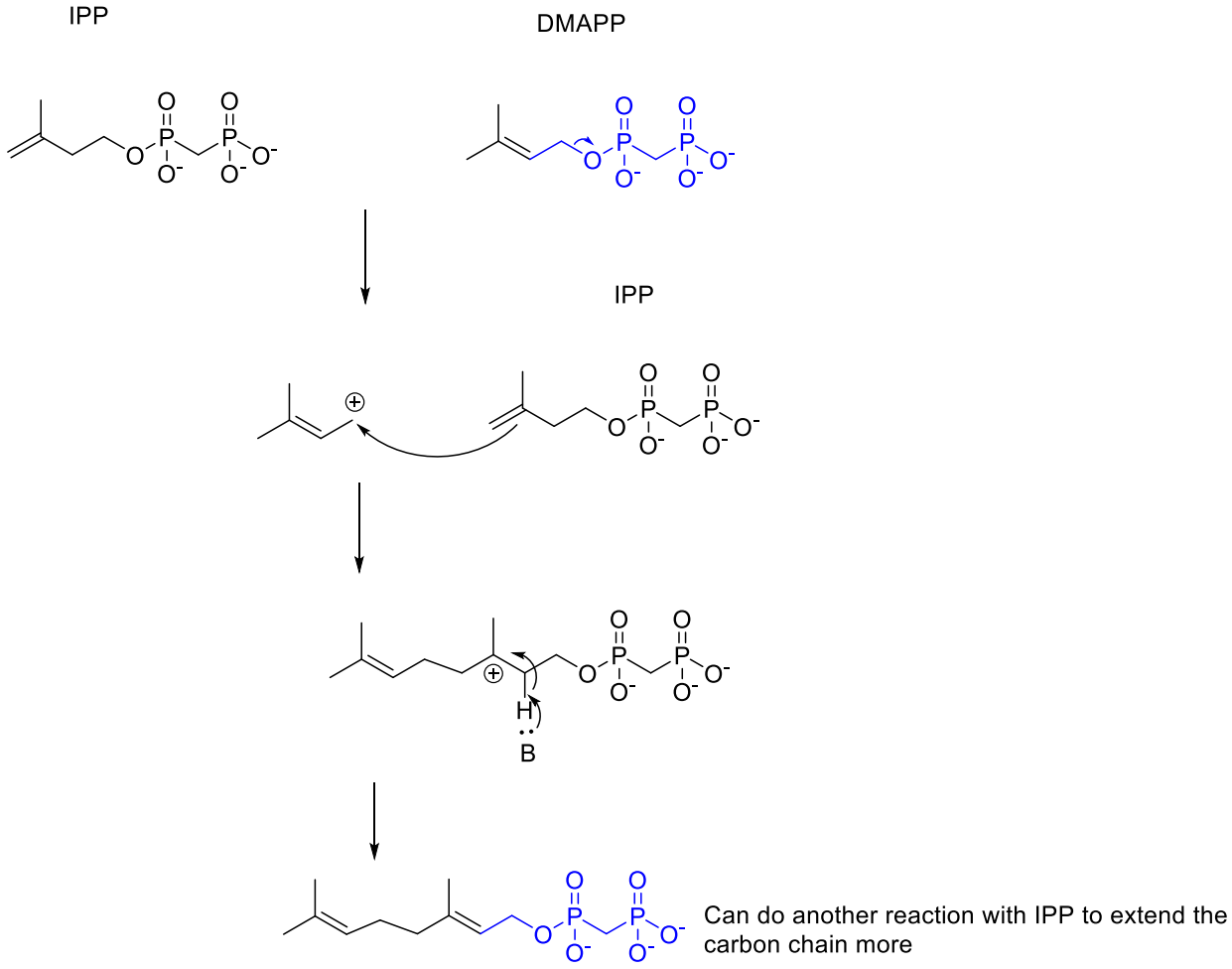
Since this is a lot to remember, this is a macroscopic view of cholesterol synthesis:



The first two steps of cholesterol synthesis are the same as the ketone body synthesis pathway so I will not go over those reactions here. Because the HMGR step is the first step that diverges from the ketone body pathway and it is an irreversible step, HMGR is a highly regulated enzyme. The details of its highly sophisticated regulation will be covered in another chapter; they are too long to be described here. This reaction produces mevalonate, which has two OH groups. These two OH groups are important because they are able to be phosphorylated and ultimately, the CO_2^- group is decarboxylated to give isopentenyl pyrophosphate or IPP. The pyrophosphate group is important because it is a good leaving group, recall from organic chemistry that weak bases are good leaving groups and the pyrophosphate is a very weak base. It is this leaving group that allows for the C-C bonds to form. As we see in the macroscopic view, we start with a 5

carbon compound and we end with a 30 carbon compound. The only way this can occur is if we form C-C bonds, which the pyrophosphate facilitates.

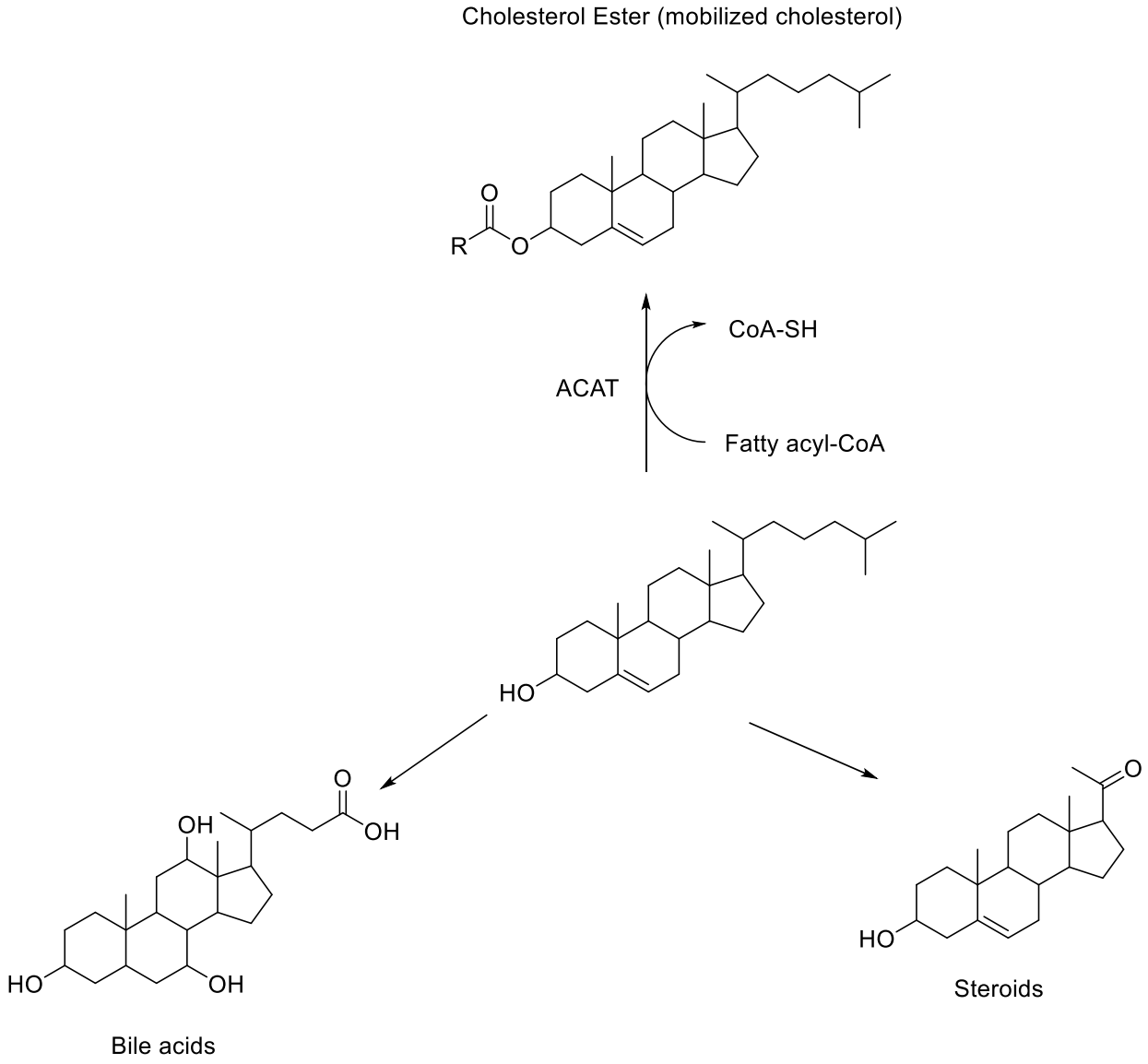
With that in mind, the isomerization of IPP to dimethylallyl pyrophosphate is important because of the way in which the C-C bonds are formed to give geranyl pyrophosphate or GPP. The mechanism of the C-C bond formation is shown below:



The isomerization allows the pyrophosphate to leave and gives rise to a much more stable primary allylic carbocation, with a tertiary allylic resonance form compared to the IPP, which would give rise to a primary carbocation, which is too unstable to exist (**M**ASC). Recall from organic chemistry that allylic carbocations are orders of magnitude more stable than regular carbocations and that tertiary allylic carbocations are among the most stable possible.

This geranyl pyrophosphate can undergo the same kind of reaction because if the pyrophosphate were to leave, it would again give rise to the primary allylic carbocation and the same reaction can occur if another isopentenyl pyrophosphate is nearby. This is exactly what farnesyl pyrophosphate (FPP) synthase catalyzes. This process repeats over and over again until we make squalene from two FPPs. This reaction mechanism is similar except NADPH acts as a reducing agent by removing one of the PP leaving groups and replacing it with an H.

After squalene is made, there are several reactions that take place to finally give cholesterol. These reactions are not necessary to know, so I will not include them here. I leave that to you to look up if you are so inclined. From cholesterol, a wide variety of physiologically important compounds can be made summarized below:



Here are the key takeaways from the cholesterol synthesis pathway:

1. Pyrophosphorylation of alcohol groups activates the carbon chain and allows for C-C bond formation
2. Cholesterol synthesis is a highly regulated process
3. Cholesterol synthesis intermediates gives rise to a wide variety of terpene natural products
4. Cholesterol itself is a precursor to steroid hormones and bile acids

5. Cholesterol can be mobilized by esterification of its OH group
6. Carbon count of the cholesterol intermediates increases by 5 until reach 15, then jumps to 30 to make squalene.

Practice Questions:

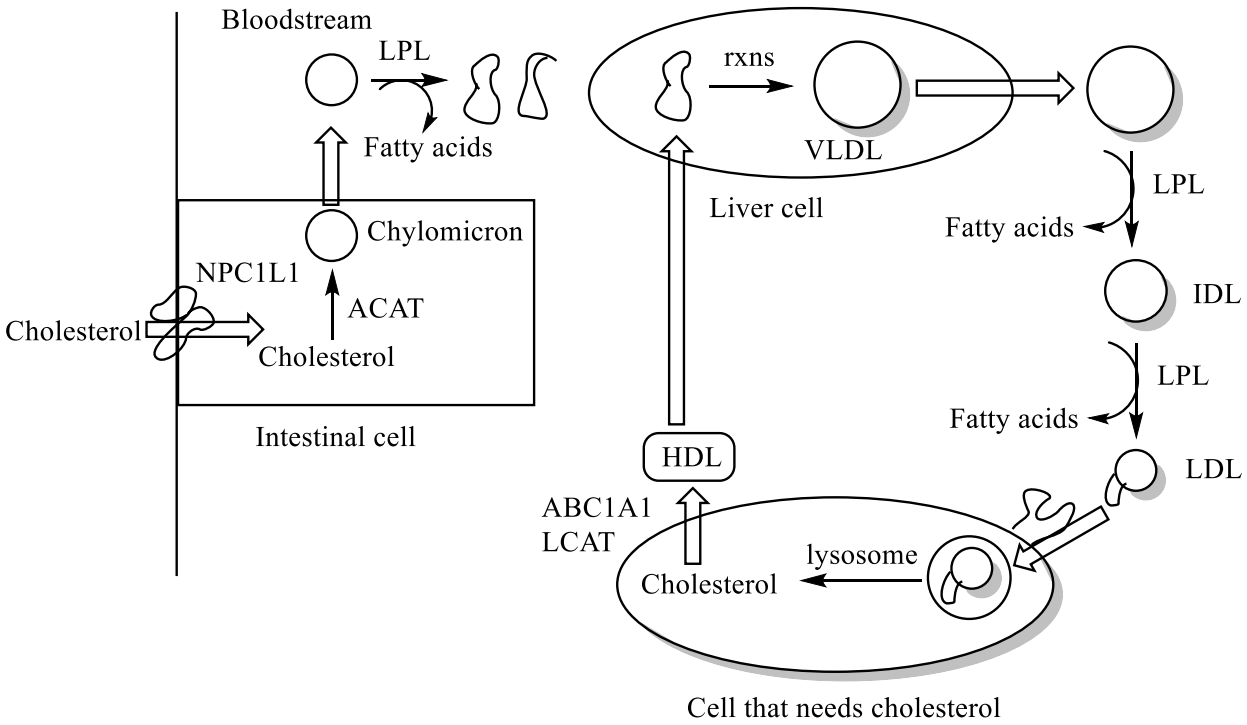
1. One possible post-translational modification that can be done to proteins is farnesylation. Much like phosphorylation, farnesylation greatly influences protein structure and function. Where do you think proteins that are farnesylated end up in the cell (location-wise)?
2. Terpenes are a large class of natural products, yet despite this, their synthesis is carried out through the same enzymes. Explain why this makes sense given the chemical logic of bond formation. (Hint: look at the geranyl pyrophosphate synthesis mechanism).
3. What hormone do you think would stimulate cholesterol synthesis? What hormone(s) do you think would inhibit cholesterol synthesis? Explain.
4. A certain kind of bacteria is typically found in warmer temperatures. When that bacteria is put in a colder environment, what do you suspect would happen to cholesterol biosynthesis? Would it increase or decrease? Explain.
5. What major enzyme do you suspect the major cholesterol medications target? Why?

Answers:

1. Farnesylation would add a very large nonpolar group to the protein, therefore, I would suspect that farnesylated proteins would be destined for membrane incorporation of some kind (could be the cell membrane or membrane of an organelle).
2. The C-C bonds that form during terpene biosynthesis occur because of carbocations. Carbocations are notoriously annoying because of their tendency to rearrange and have resonance forms. This explains why there are many possible terpene products from one enzyme.
3. Insulin would likely stimulate cholesterol synthesis while glucagon and epinephrine (adrenaline) would likely inhibit cholesterol synthesis. This is because of the different messages these hormones are sending. Insulin tells the cells in the body that there is plenty of glucose aka energy. Because energy is plentiful and cholesterol synthesis (or any synthetic pathway for that matter) uses energy, it is natural that insulin should be a sort of growth factor that tells the cell it is appropriate to make fats and other biomolecules. Glucagon and adrenaline on the other hand communicate the opposite message. They indicate the there is high energy demand, that the body is desperate for energy and therefore it should not waste its already limited ATP supply on making cholesterol it doesn't necessarily need.
4. If the bacteria is used to warmer temperatures but is put in a cold environment, the cell membrane is likely too rigid for the cold temperature and therefore it should increase its cholesterol production so that its membrane's stability can be maintained.
5. HMGR is the major enzyme that is targeted by cholesterol medication (statins) because it is the commitment step of cholesterol synthesis.

Chapter 32: Cholesterol Transport

Cholesterol is not only made in cells, but it is also ingested. The cholesterol that is ingested is also referred to as dietary cholesterol. When cholesterol is ingested, it is taken into the body by the cells in the intestine using the NPC1L1 receptor. There ACAT is used to esterify the OH group of cholesterol to make it a cholesterol ester, which is packaged into a chylomicron. A chylomicron is a type of lipoprotein or lipid droplet that is transported from the intestinal cells to the blood stream, where it is broken down using LPL (lipoprotein lipase) to release fatty acids. LPL breaks the chylomicron down into chylomicron fragments containing dietary cholesterol, which migrate to the liver, where the cholesterol is repackaged into VLDL. When the VLDL is released from the liver, it gets broken down progressively to IDL and LDL upon LPL attack. VLDL, IDL, and LDL are all different kinds of lipoproteins that have well-defined composition and characteristic proteins on their surface called apolipoproteins. VLDL and IDL are used to distribute fatty acids to target tissues. LPL is effectively the unpacker enzyme that allows these lipoproteins to deliver their fatty acids to a given target tissue. LDL on the other hand is used to deliver cholesterol to tissues and rather than getting broken down by LPL, it binds the LDL receptor using its ApoB-100 protein. This causes the LDL to taken into the cell using a vesicle (endocytosis). This vesicle fuses with a lysosome, the digestive organelle of the cell, that extracts all of the nutrients out of the LDL droplet. The remaining cholesterol stays in the cell and can be picked up using HDL. An overview of the cholesterol transport process is shown below:

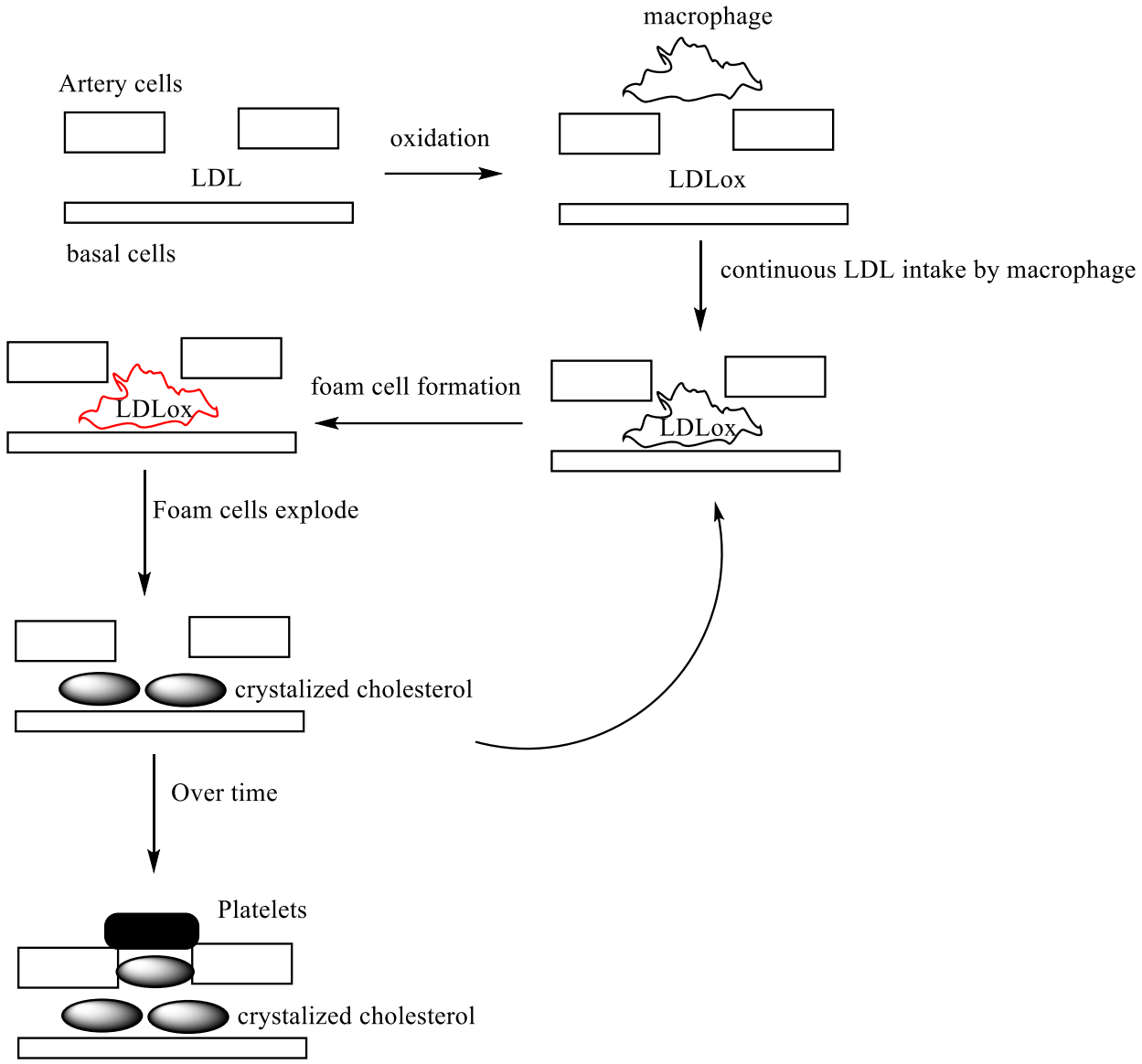


In both cases of cholesterol being put into a lipoprotein, it needs to be esterified. In the intestinal cells, this is done using ACAT. In other cells that are releasing excess cholesterol, this is done

using LCAT. Either way, the cholesterol needs to be esterified to make it sufficiently nonpolar to move into the center of the lipid droplet. Recall that lipids are extremely nonpolar while cholesterol has the hydroxyl group at the end that enhances its polarity. The HDL lipoprotein can be used to extract excess cholesterol and bring it back to the liver by using its ApoA-1 protein on its surface. This protein interacts with ABC1A1 and activates LCAT, which allows the excess cholesterol in nonliver cells to be packaged into HDL.

It is often said that there is “bad” and “good” cholesterol. The difference between these two types of cholesterol has to do with the type of lipoprotein they are a part of. The LDL cholesterol is dangerous if it gets caught in your arteries because it can lead to atherosclerosis. The mechanism of this will be described later. The HDL cholesterol is very beneficial because it takes excess LDL cholesterol back to the liver, preventing it from getting stuck in the arteries and occluding blood vessels. Therefore, we want LDL levels to be Low and HDL levels to be High. In short, LDL brings cholesterol from liver to tissues and can get stuck along the way while HDL brings cholesterol from tissues to the liver.

Atherosclerosis occurs when LDL gets stuck in blood vessels and gets oxidized via radical lipid peroxidation. This oxidized LDL causes nearby cells to release signals called cytokines into the blood that make macrophages attempt to take in the LDL. These macrophages, however, do not regulate the amount of LDL that is taken in because the oxidized LDL activates an unregulated receptor on their surface. This causes the macrophages that are there to become engorged in cholesterol and ultimately to form foam cells that explode and deposit cholesterol crystals. This causes even more macrophages to come in a positive feedback loop until eventually platelets start to come and occlude the blood vessel more and more over the course of time. This is called plaque and the more plaque that occurs in blood vessels the harder it is for blood to flow, until eventually the vessel is completely clogged and a heart attack or stroke occurs. Because the oxidation of LDL occurs through a radical mechanism, there are many people who propose that antioxidants can help prevent atherosclerosis because these antioxidants will sequester the radicals and prevent them from oxidizing the LDL that causes these plaques to form in the first place. Another very important thing to understand with this is that LDL IS ONLY BAD IF IT IS IN THE BLOOD! If LDL is in the cell, it does no harm at all, it is only when it is stuck in an artery and gets oxidized that it is bad. Therefore, anything that can promote LDL uptake by cells, prevent LDL oxidation, or prevent overall cholesterol uptake could theoretically be used to treat heart disease. An overview of the LDL atherosclerosis mechanism is provided below:



Practice Questions:

1. What are some proteins that can be inhibited to treat atherosclerosis? What about proteins that could be enhanced to treat atherosclerosis? Explain how each one works.
2. Explain why cholesterols need to be esterified to be packaged into the lipoproteins.
3. Compare and contrast LDL and HDL on the premise of composition, function, etc.

Answers:

1. Inhibition of NPC1L1 and ACAT could possibly treat atherosclerosis. NPC1L1 could treat atherosclerosis because LDL cholesterol is only harmful if it is in the blood, therefore inhbiting this transporter would prevent cholesterol incorporation into chylomicrons and therefore it cannot enter the blood. Instead cholesterol would be passed

in the urine, where it cannot cause plaque formation. ACAT is the enzyme responsible for the formation of cholesterol esters, without this enzyme, cholesterol cannot be incorporated into chylomicrons and would instead reside in the intestinal cells. Just as before, LDL is only harmful if it enters the blood stream and therefore if it is stuck in the intestinal cells, it cannot cause plaque formation. Activation of LDL receptor, LCAT, and ABC1A1 could also help treat atherosclerosis. The LDL receptor makes it so that the LDL does not stay in the blood for long and instead gets taken in by cells. As previously mentioned, this would help treat atherosclerosis because LDL can only cause damage if in the blood. LCAT and ABC1A1 are both responsible for HDL cholesterol transport and this would help the "good" cholesterol do its job.

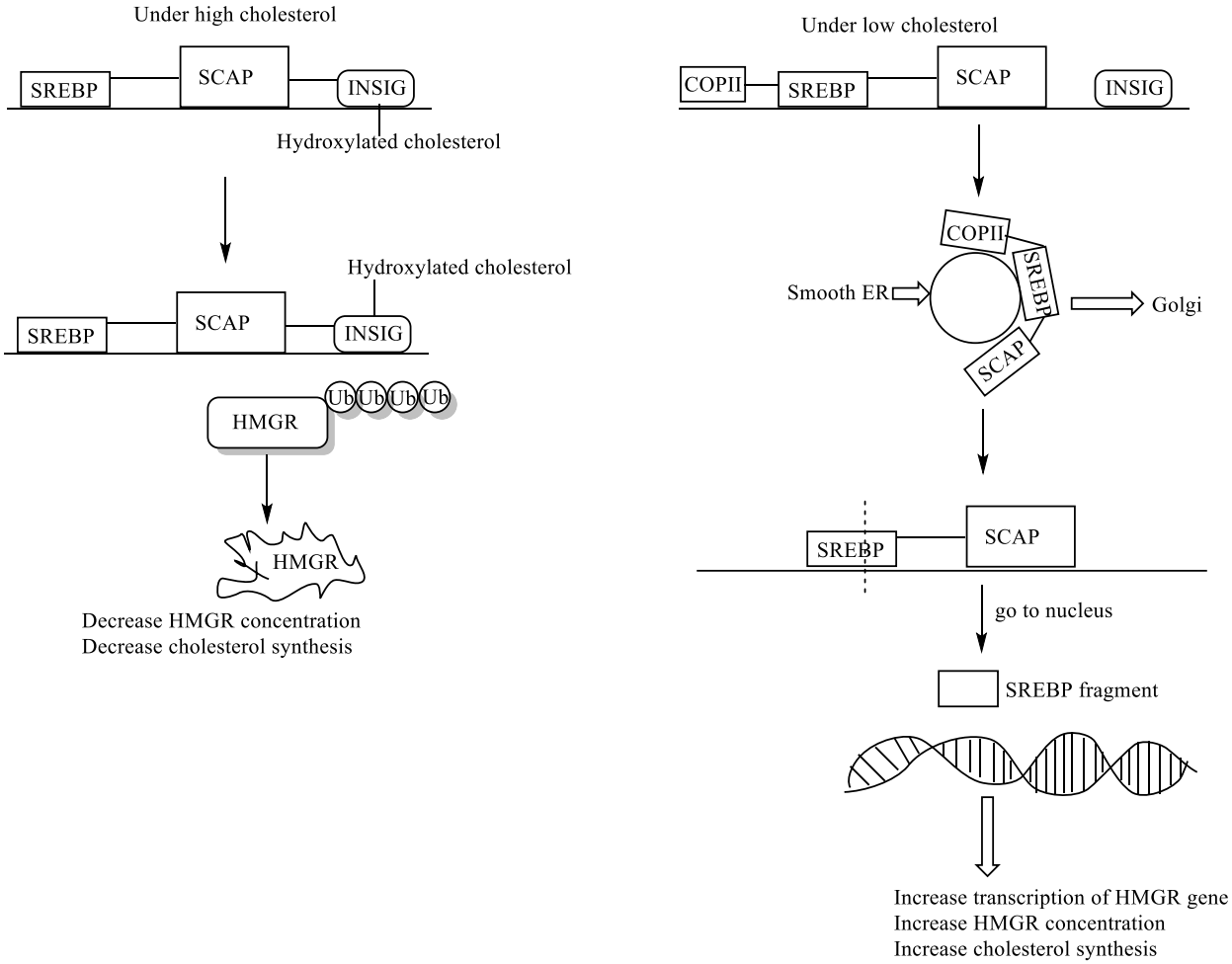
2. Cholesterol is not nonpolar enough to go into chylomicrons and lipoproteins directly. Instead it needs to have its polar OH group esterified to an ester moiety to make it sufficiently nonpolar to be incorporated into the interior of the lipid droplets.
3. LDL is mostly cholesterol while HDL is mostly protein. LDL causes atherosclerotic plaques while HDL prevents those plaques from forming. LDL is responsible for the delivery of cholesterol to tissues from the liver, while HDL is responsible for taking excess cholesterol from the tissues to the liver. LDL has the ApoB-100 apolipoprotein while HDL has the ApoA-1 protein on its surface.

Chapter 33: Cholesterol Regulation

HMG-CoA reductase (HMGR) is a highly regulated enzyme that is responsible for the commitment step in the cholesterol synthesis pathway. There are several ways that this enzyme is regulated and we will discuss the regulatory mechanisms in-depth in this chapter. For any enzyme, there are three ways that they can be regulated: post-translational modification, transcriptional regulation, and proteolytic degradation. Post-translational modification allows for short-term regulation of the enzyme of interest, while the other two means of regulation are for longer-term regulation and they control the concentration of the enzyme in the cell. We will cover all three of these methods of regulating HMGR activity.

HMGR is deactivated by phosphorylation. Because making cholesterol requires a lot of ATP and NADPH, the cell will only make cholesterol when it has a lot of ATP. This, like many other processes, means that glucagon and adrenaline will turn off HMGR by activating PKA that will phosphorylate HMGR and thereby deactivate it. Conversely, insulin will activate HMGR phosphatase and therefore increase HMGR activity. Another incredibly important protein kinase is responsible for regulating HMGR activity. This protein kinase is the AMP-dependent protein kinase (AMPK). This kinase, as the name suggests, is activated by high AMP levels. This functions similarly to glucagon and will deactivate energy-consuming processes and activate energy-producing ones (MAS**C**). Because cholesterol synthesis requires a tremendous amount of energy, AMPK will phosphorylate HMGR and deactivate it. Therefore, in short, AMPK, glucagon, and adrenaline all decrease HMGR activity while insulin increases HMGR activity.

HMGR is also transcriptionally regulated in the smooth endoplasmic reticulum. In the smooth ER, there are three main proteins that are embedded in the membrane: SREBP, SCAP, and INSIG. SREBP and SCAP form a stable complex that can bud off of the smooth ER when cholesterol levels are low while INSIG acts as an anchor for the SREBP-SCAP complex and can destroy HMGR. Under conditions of low cholesterol, INSIG does not bind the SREBP-SCAP complex and this allows SREBP-SCAP to use COPII move from the smooth ER to the Golgi apparatus. At the Golgi, the SREBP-SCAP complex is broken down and a small piece of SREBP is released. This small piece enters the nucleus and promotes expression of the HMGR gene, increasing HMGR concentration and thereby increasing cholesterol synthesis overall. Under conditions of high cholesterol, enzymes are able to make hydroxylated cholesterol that will bind to INSIG and activate it. This causes INSIG to anchor SREBP-SCAP to the smooth ER and causes HMGR to be ubiquitylated and destroyed. An overview of its transcriptional regulation is shown below:



Practice Questions:

1. What blood sugar levels and cholesterol levels would cause the highest cholesterol synthesis rate? Explain.
2. A mutation results in INSIG being unable to bind SREBP-SCAP. Explain how this mutation would affect HMGR concentration.
3. A mutation results in INSIG always binding SREBP-SCAP. Explain how this mutation would affect HMGR concentration.

Answers:

1. High blood sugar and low cholesterol levels would stimulate cholesterol synthesis the most because under these conditions insulin would be high (HMGR would be dephosphorylated and activated) and INSIG would be unable to bind SREBP-SCAP.
2. If INSIG cannot bind SREBP-SCAP, then regardless of cholesterol levels, SREBP-SCAP will bud off of the smooth ER and continuously stimulate expression of the HMGR gene. HMGR concentration should increase.

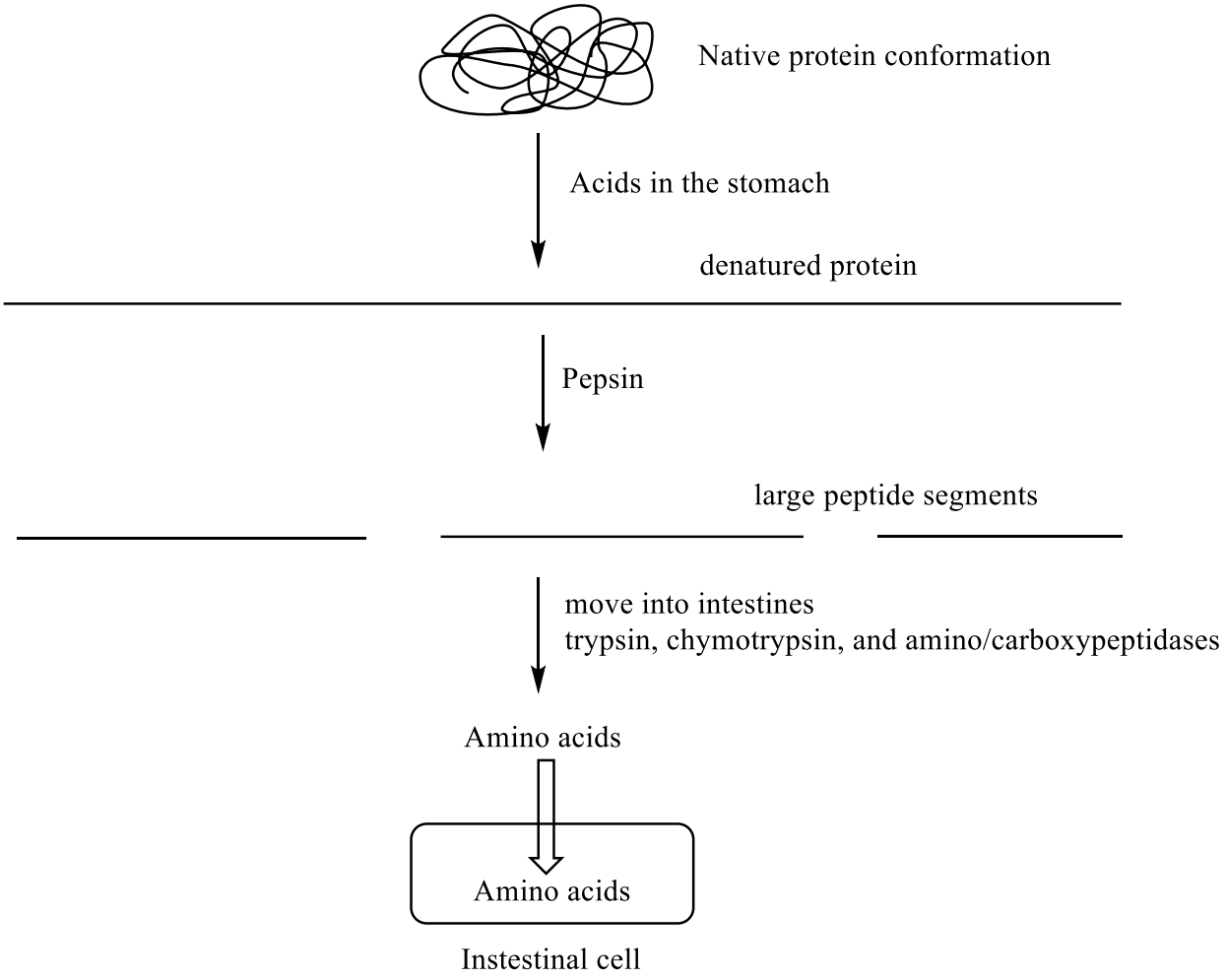
3. If INSIG always bind SREBP-SCAP, then regardless of cholesterol levels, the SREBP-SCAP complex cannot leave the ER or stimulate expression of the HMGR gene. HMGR concentration should decrease.

Focus 7: Amino Acid/ Protein Metabolism

Chapter 34: Exogenous (externally supplied) Protein Breakdown

When you eat an organism (be it plant or animal), you are also eating the proteins that compose it. When those proteins enter your stomach, they get partially degraded and eventually throughout your digestive tract they get broken down to the amino acids that comprise them. Here, we will discuss how dietary protein is broken down by the body into amino acids.

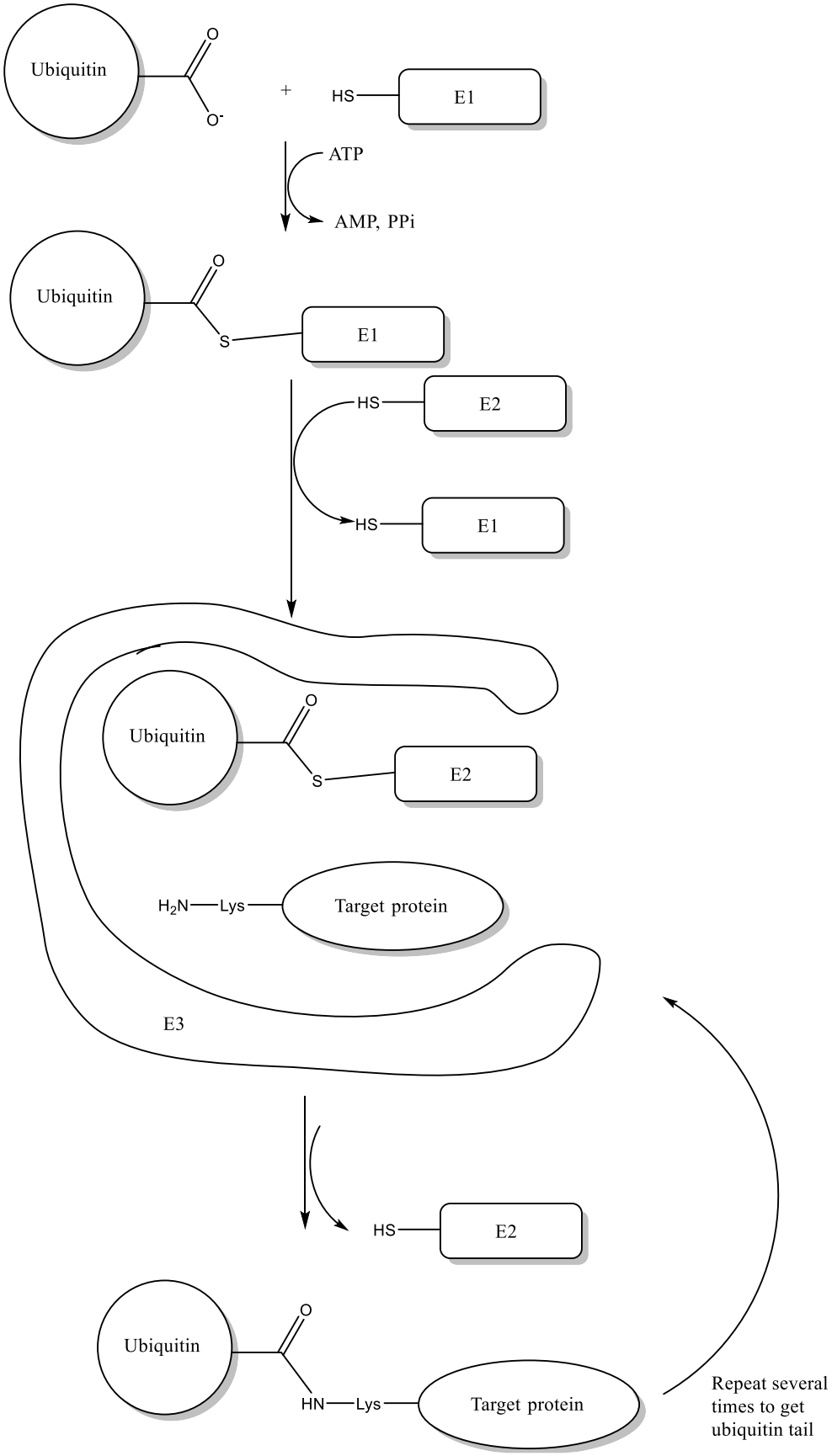
Digestion of proteins starts in the stomach, where the acidic environment protonates the residues on the surface of the protein. This causes the protein to denature and linearize, then pepsin can be used to chop it up into large pieces. Once these large peptide fragments are made, they move down the gastrointestinal tract into the intestines, where they are further broken down by trypsin, chymotrypsin, and amino/carboxypeptidases A and B. These enzymes break down the peptide fragments into their amino acid components and then these amino acids are transported into the intestinal cells via special protein transporters. An overview of this process is shown below:



Chapter 35: Endogenous (internally made) Protein Breakdown

When proteins are made in the cell and they need to be destroyed, the cells use an ubiquitination system to mark the protein for destruction. Ubiquitination is paired with proteosomal degradation. The proteasome is effectively the recycling center of the cell. It breaks down ubiquitinated proteins so that their amino acids can be used to make other proteins. The proteins that are made by the cell that are ubiquitinated are typically proteins that are either misfolded or not needed at the moment (MAS**C**). This form of protein breakdown allows the cells to effectively manage and recycle their amino acid supplies. Here we will discuss how the ubiquitination system functions to breakdown proteins.

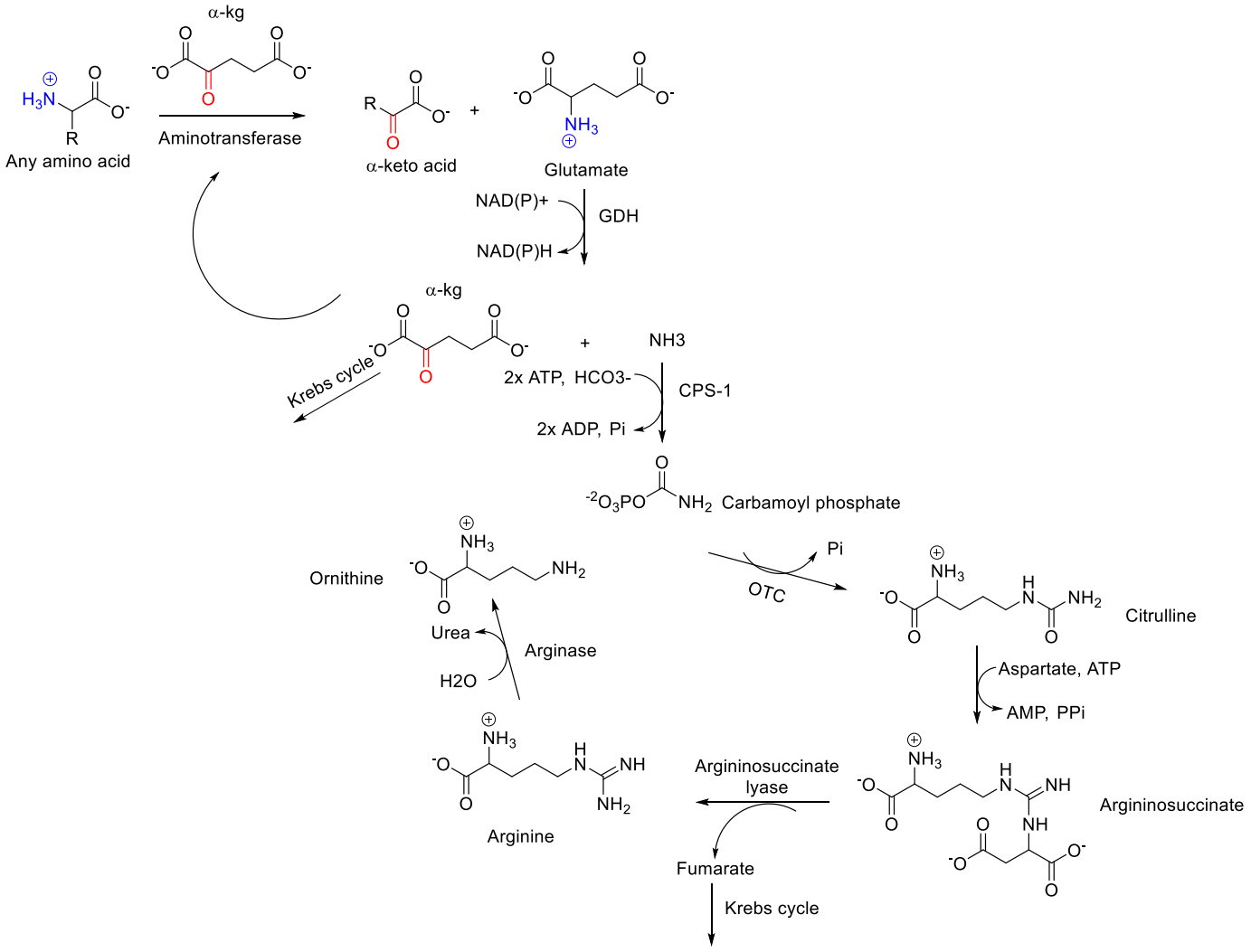
There are three reactions that take place to ubiquitinate a protein. These three reactions are catalyzed by E_1, E_2, and E_3. The E_1 reaction uses ATP to power the formation of a thioester bond between a carboxylic acid group present on ubiquitin and a thiol present on the E_1 enzyme (M**A**SC). Remember that thioesters are very reactive, therefore ATP must be used to make the unstable thioester. The E_2 reaction is simply a transthioesterification reaction and exchanges a thioester with E_1 for a thioester with E_2. Finally, the E_3 reaction connects the ubiquitin to a lysine residue on the target protein we want to destroy. This is repeated several times to add several ubiquitins and then the ubiquitinated protein is recognized by the proteasome and is destroyed. An overview of this process is shown below:


Ubiquitin
O
O-
+
HS
E1
ATP
AMP, PPi
Ubiquitin
O
S
E1
HS
E2
HS
E1
Ubiquitin
O
S
E2
H2N
Lys
Target protein
E3
HS
E2
Ubiquitin
O
HN
Lys
Target protein
Repeat several
times to get
ubiquitin tail

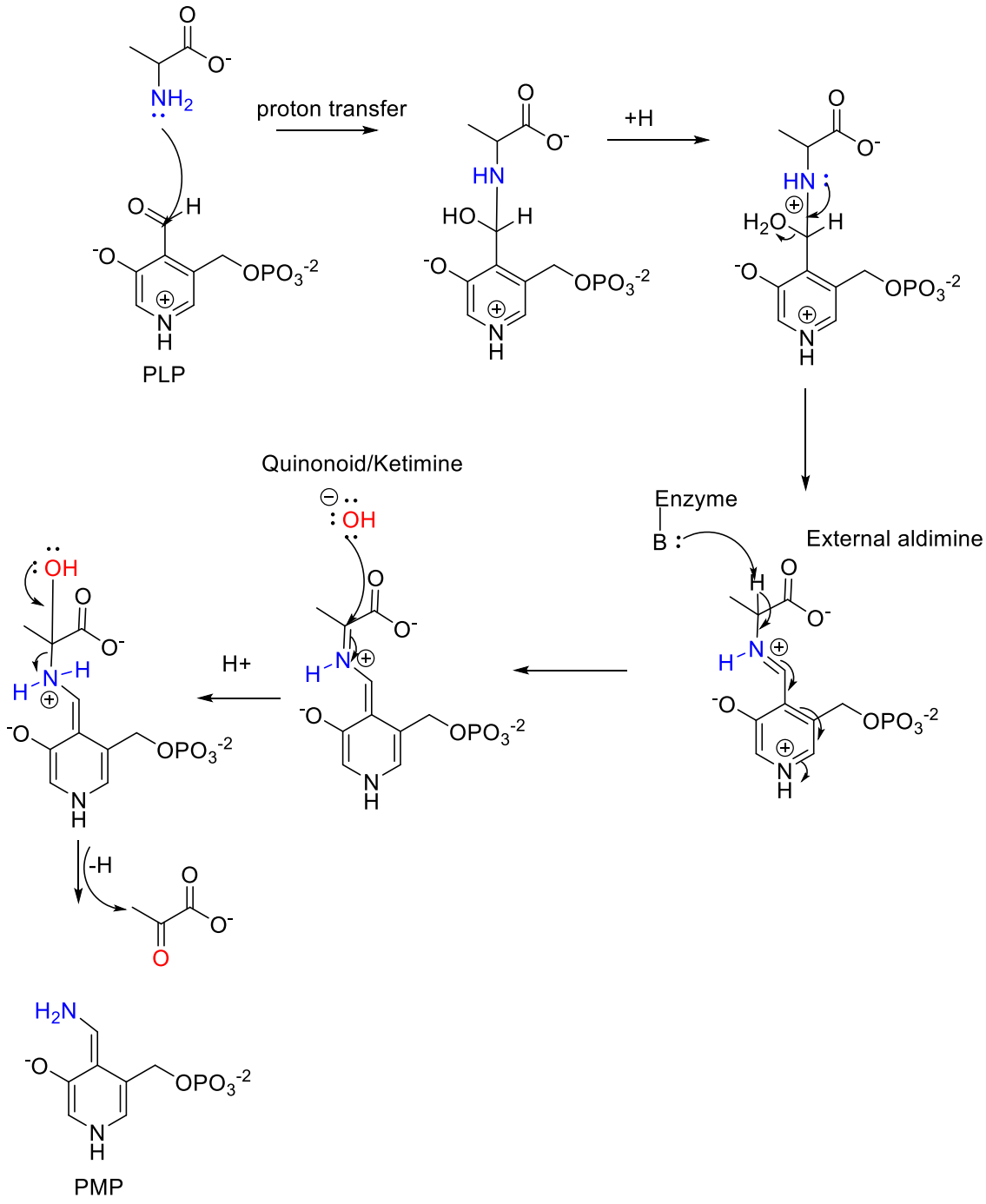
Chapter 36: Urea Cycle

Now that we have discussed how large proteins are broken down from the diet and the ones the cells themselves make, we can discuss how amino acids are broken down. Amino acids are broken down in three stages. First, all the amino acids funnel their NH_3 to α-ketogluterate to make glutamate using an aminotransferase. Second, the glutamate releases its NH_3 using glutamate dehydrogenase to recycle the α-ketogluterate supplies. Third, the NH_3 released from glutamate is used to make urea in the mitochondria and cytoplasm, this is known as the urea cycle. This urea is then excreted in the urine. Like all metabolic processes, we want to first discuss when this is most likely to be used. If there is a lot of excess amino acids or the cells have exhausted all carbohydrates and lipids for energy, this pathway will be most active (MAS**C**). This should make sense because if there is an excess amount of amino acids, then we can break them down for energy. However, proteins are not useful for energy production, so they are only broken down in cases of severe starvation since proteins are incredibly useful catalytic tools.

An overview of amino acid breakdown is shown below:

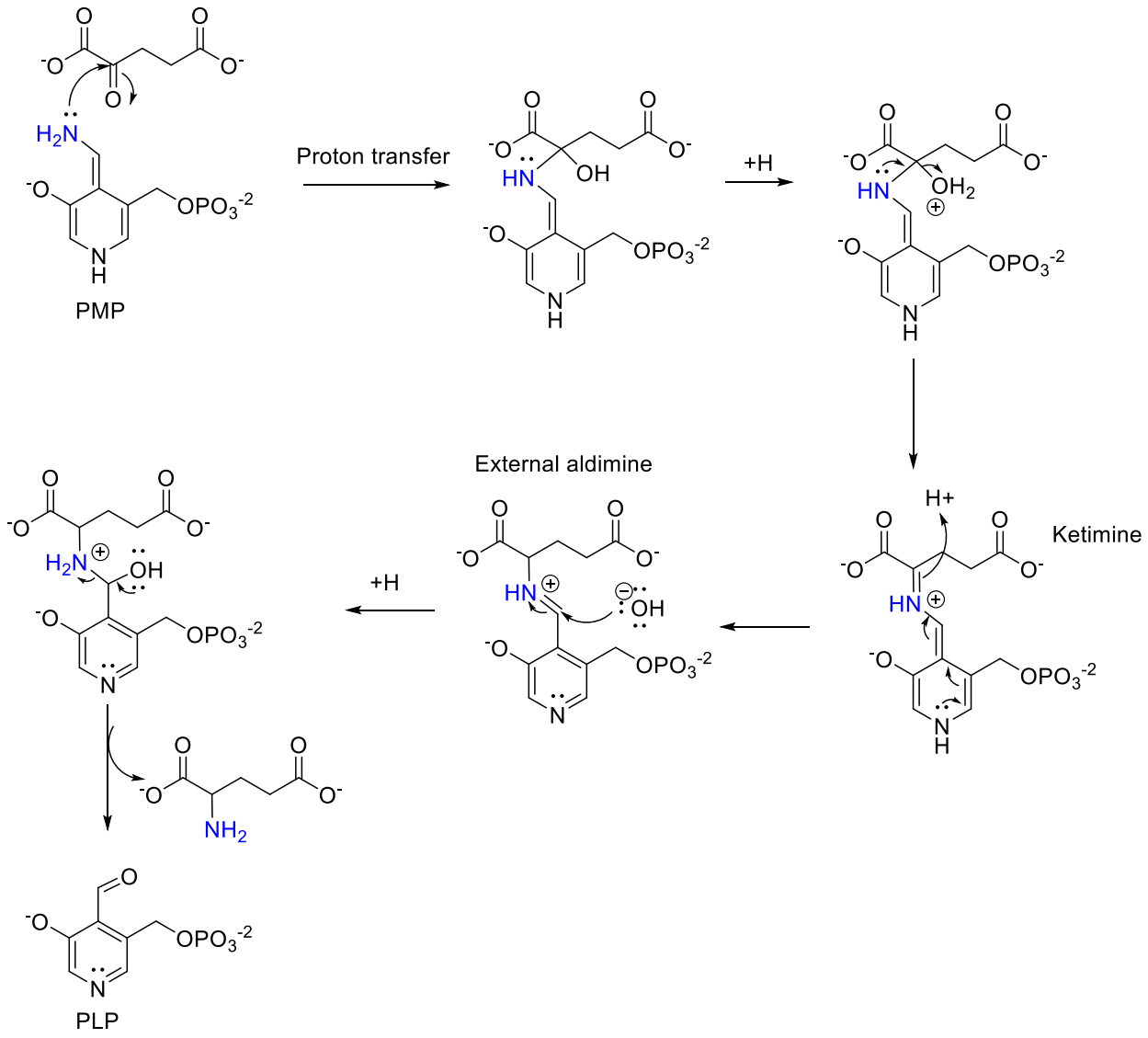


Central to this entire pathway is the ability to transfer amino groups (NH_3) from one carbon chain to another. This reaction is always catalyzed by aminotransferases, which use a PLP cofactor. The mechanism for these aminotransferases are important and therefore, we will discuss it in detail now. For this example, we will discuss the case where alanine is used to donate its NH_3 group to α-ketogluterate. This is simply an example case, ANY amino acid and alpha-keto acid can be used in this reaction and the mechanism would be the same. This enzyme family uses a ping-pong bi-bi mechanism, meaning that the reaction effectively occurs in two steps. First, the PLP turns the amino acid into an alpha-keto acid. This makes PLP have an amino group and therefore it is now called PMP. Second, this PMP is used to donate the amino group to a separate alpha-keto acid to make it an amino acid. The first step of donating the amino group from the amino acid to from PMP is shown below:



The first step in this mechanism is the amino group of the amino acid acts as a nucleophile and attacks the electrophilic aldehyde carbon of the PLP cofactor. Since the amine on the amino acid is a primary amine, it creates an imine product like we would expect based off our knowledge of organic chemistry. If you would like an organic chemistry explanation for why this reaction proceeds the way it does, please look at the organic chemistry survival guide posted on the website. This imine that forms is called the external aldimine because it is an imine that formed by using an amino group outside of the enzyme. It is an aldimine because the aldehyde was

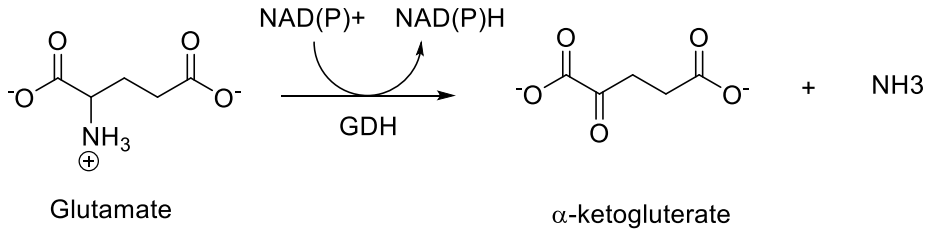
getting attacked by the amine group. To resolve the positive charge in the aromatic ring on the bottom of the PLP cofactor, a basic residue in the aminotransferase enzyme is used to deprotonate the acidic alpha carbon of what was the amino acid. This causes an electron cascade that places the electrons from the pi bond near the N in the aromatic ring on the N itself, resolving its positive charge (**M**ASC). This forms the quinonoid intermediate, also known as the ketimine (not ketamine, don't do drugs kids). This ketimine is also very electrophilic because the imine nitrogen is positively charged and is more electronegative than the imine carbon. This allows OH^- to attack it, resolving the positive charge on the imine nitrogen and negative charge on the oxygen. This creates a tetrahedral intermediate (**M**ASC). This tetrahedral intermediate will collapse once the nitrogen gets protonated because as we know from organic chemistry, positively charged groups want to leave very badly. The other driving force is the formation of the C=O bond, which like in orgo, is a very stable bond (**M**ASC). This releases the alpha-keto acid, in this case, that acid is pyruvate and creates the PMP cofactor that can be used to donate the NH_3 group to an alpha-keto acid acceptor. This reaction mechanism is shown below:



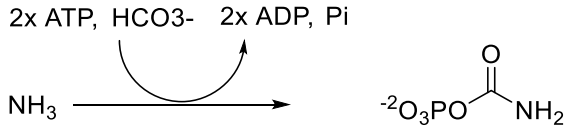
This mechanism is extremely similar to the first step. In fact, it is just the first step but in reverse. Here, the PMP is acting as the amino donor and the alpha-keto acid is acting as the amino acceptor, rather than vice versa for the first step. All of the chemical logic is the exact same as well. What is important to note here is that in both cases, the oxygen is supplied by a free OH^- anion rather than the oxygen that was there initially in the alpha-keto acid or aldehyde. On the other hand, the amino group is directly transferred. In other words, the amino group that was on the amino acid is the same amino group that is given to the alpha-keto acid.

While this reaction can occur for a wide variety of alpha-keto acids and amino acids. The main alpha-keto acid for this process is α-ketogluterate. This is because the process needs to have a main funneling point to minimize the number of enzymes it needs to use to free up the NH_3 from amino acids (MAS**C**). By funneling all of the amino groups from the amino acids to glutamate, the cell only needs to evolve a highly efficient glutamate dehydrogenase to free the NH_3, rather than evolve 21 different dehydrogenases for each amino acid. α-ketogluterate is the molecule of choice here because its concentration is highly regulated by energy demands; high energy demand translates to high levels of α-ketogluterate. In this way, the cell only breaks down amino acids to an appreciable extent when energy demands are extremely high and the cell has nothing else to fall back on for energy (no sugar or fat).

The glutamate dehydrogenase enzyme (GDH) is responsible for "oxidative deamination" of glutamate and catalyzes the reaction shown below:



GDH is a unique enzyme in that it can use the oxidizing power of both NAD+ and NADP+ to oxidize glutamate to α-ketogluterate. Once this ammonia is released it quickly has to get sequestered as carbamoyl phosphate. This reaction is catalyzed by carbamoyl phosphate synthase-1 or CPS-1 and uses 2 ATP and 1 HCO_3^- to create carbamoyl phosphate, a high energy compound (M**A**SC) in the reaction shown below:

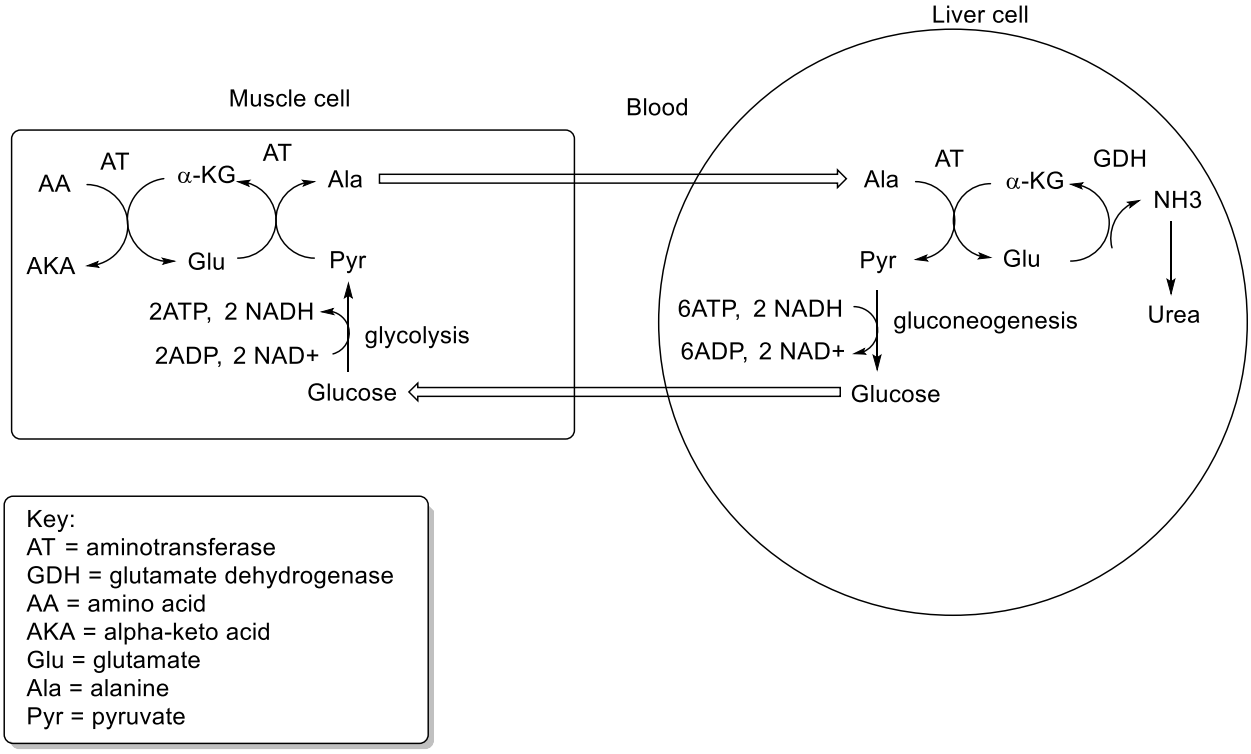


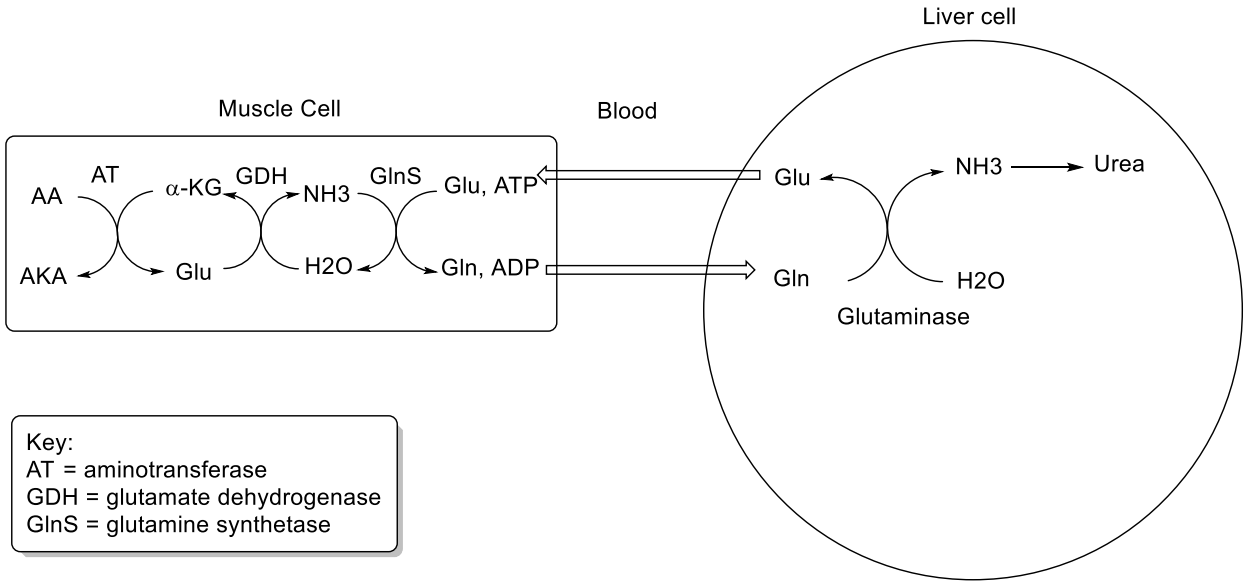
Now that the carbonyl carbon is activated via phosphorylation, the carbamoyl phosphate can be combined with ornithine to make citrulline using ornithine transcarbamylase (OTC). This releases the phosphate group and adds the citrulline group via the NH_2 group. But now we are at a dead end, there is no activated carbons in this compound that we can use to add to anything, so we need to use ATP to catalyze the condensation of citrulline with aspartate to create

argininosuccinate using argininosuccinate synthase (M**A**SC). This helps create that guanidine group on the right hand side of the molecule so that the cell can make arginine in the next step using argininosuccinate lyase. This accomplishes two things: first, it allows for the creation of additional Krebs cycle intermediates (fumarate) and second, it creates arginine, which is recognized by arginase and can be used to make urea and ornithine to complete the cycle. Most of these reactions occur in the cytoplasm except for the CPS-1 and OTC reactions.

This cycle is regulated at CPS-1 because that is the entry point into the urea cycle. The CPS-1 enzyme is most active when someone has a high protein diet or when they are starving and breaking down proteins to make energy. When someone is on a high protein diet or when they are starving, many amino acids are present in the cell, some of these get funneled into glutamate like we just discussed. Because there is now an increased amount of glutamate, there is a higher likelihood that it will react with other reactive compounds such as acetyl-CoA. When these two molecules react using N-Acetylglutamate synthase, it creates N-acetylglutamate, which will activate CPS-1. The N-acetylglutamate synthase enzyme itself is activated by arginine, which makes sense because arginine is the last step of the urea cycle. A lot of arginine present signals that the urea cycle stops are arginine, likely because there isn't enough carbamoyl phosphate being fed into the urea cycle. In this way, the cell can detect and respond to a dysfunctional urea cycle and correct itself (MAS**C**).

Sometimes, the nitrogen in the amino acids needs to be transferred to the liver, since the liver is the tissue that is mainly responsible for doing the urea cycle. The nitrogen can be shuttled to the liver through two different pathways, the glucose-alanine cycle and the glutamine cycle. These cycles are summarized below:





As you can see in the diagram, both pathways use the same kind of machinery. They both use aminotransferases to make a lot of glutamate. The only difference is how that glutamate is used. In the glucose-alanine shuttle, it is used to make alanine from pyruvate while in the glutamine shuttle it is used to make NH_3, which ultimately is used to make glutamine. The glucose-alanine shuttle is energetically inefficient; it uses up 4 ATP for every molecule of urea made (2 made in the muscle and 6 used in the liver). This is due to the large energetic cost of gluconeogenesis in the liver. Because of this inefficiency, the glutamine shuttle is the primary mode of transport of amino acids into the liver for use in the urea cycle (MAS**C**). The glutamine cycle only uses 1 ATP, so it is four times as energetically efficient as the glucose-alanine alternative. This explains why glutamine is the primary amino acid that is found in the blood; the glutamine shuttle is by far the most used means of transport.

Here are the key takeaways from the urea cycle:

1. Only two reactions occur in the mitochondria, with the rest occurring in the cytoplasm (CPS-1 and OTC occur in mitochondria).
2. CPS-1 is the major regulatory enzyme of the pathway and is allosterically activated by N-acetylglutamate, a glutamate analogue that is present during starvation or high protein diet.
3. Amino acids are moved to the liver for use in the urea cycle via glucose-alanine shuttle or glutamine shuttle.
4. All amino acids donate their NH_3 group to α-ketogluterate to make a whole lot of glutamate. This is effectively the funneling point for the pathway. This is done using an aminotransferase, which is an enzyme that uses the PLP cofactor.

Practice Questions:

1. How much ATP is produced by removing the NH_3 from alanine in the urea cycle? From the urea cycle and aerobic respiration? If there is a difference, explain.
2. How much ATP is produced/lost in the muscle tissue for transporting 16 NH_3 using the glucose-alanine shuttle? Glutamine shuttle?
3. Suppose there was a mutation in CPS-1 that caused it to no longer function, what would happen to the pH of the cell and the surrounding blood? Explain.

Answers:

1. Alanine makes pyruvate when using an aminotransferase. If we just look at the ATP produced from the urea cycle, then we would lose 3 ATP (2 from CPS-1 and 1 from argininosuccinate synthase). If we consider that we make a fumarate in this process, the fumarate would give us 1 NADH, which is equivalent to 2.5 ATP. Therefore, just looking at the urea cycle, we would net lose 0.5 ATP. However, if we oxidize the pyruvate using Krebs cycle and the ETC, then we would gain 1 ATP and 1 NADH from the PDH complex and each acetyl-CoA gives 10 ATP. So, if we consider the 3 ATP lost from the urea cycle, the 1 ATP gained from PDH, 1 NADH gained from PDH, 1 NADH gain from fumarate, and 10 ATP gained from acetyl-CoA, we can get that we gained net total of $ATP_{total} = 10 + 1 + 2(2.5) - 3 = 13\ ATP$

   If we convert NADH to ATP.

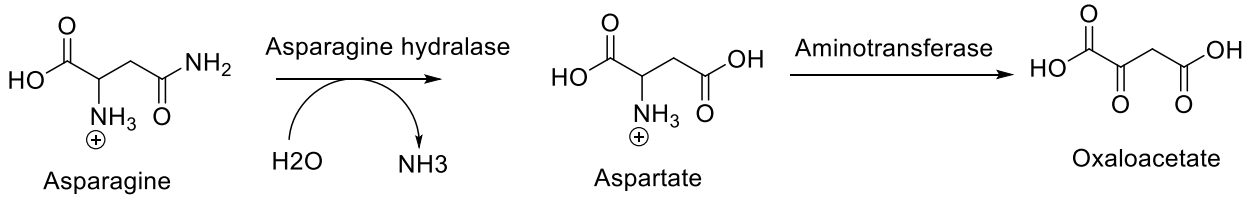
2. In the muscle tissue, the glucose-alanine shuttle has us use glycolysis to make pyruvate. Because we are transporting 16 NH_3, we only need 8 glucose molecules (each glucose molecule gives 2 pyruvate). No other reaction in the muscle tissue produces or consumes ATP, so really it is just a net gain of ATP and NADH from glycolysis in the muscle tissue from the 8 glucose molecules we need. Each round of glycolysis gives 2 ATP and 2 NADH, therefore we get 16 ATP and 16 NADH. Converting the NADH to ATP yields 56 ATP gained in the muscle tissue from the glucose-alanine shuttle. The glutamine shuttle on the other hand uses up 1 ATP per NH_3 transported to the liver because of the glutamine synthetase reaction. Therefore, using the glutamine shuttle, we would LOSE 16 ATP.
3. If CPS-1 was no longer working, then ammonia would begin to accumulate in the cell. The end result is that the pH of the cell and likely the extracellular space would increase because ammonia is basic. If left untreated, this could result in severe alkalosis and possibly death.

Chapter 37: Amino Acid Breakdown

We briefly discussed how amino acids are broken down in the previous chapter, or at least how they can donate their NH_3 group for the urea cycle. However, amino acids can be broken down to a wide variety of compounds that can be used for energy. Here, we will discuss some of those important amino acids that can be broken down into compounds that can be used for energy. Broadly, there are two different kinds of amino acids: ketogenic and glucogenic. Ketogenic amino acids are those amino acids that can be broken down to acetyl-CoA and therefore be made into ketone bodies. Glucogenic amino acids, as the name would suggest, are the amino acids that can be converted to glucose ultimately. An important fact to note is that amino acids can be both ketogenic and glucogenic. These degradation pathways that we will discuss are solely focused on making energy from amino acids, therefore this will only really occur if someone is on a high protein diet or if they are starving, just like the urea cycle. Here, we will only discuss the breakdown of asparagine, leucine, phenylalanine, and serine. Serine and asparagine are the two glucogenic amino acids while leucine is a ketogenic amino acid and phenylalanine is both ketogenic and glucogenic. The specific steps for these processes are not super important to know, but the end products are important. A comprehensive list of products for each of these amino acids is provided below:

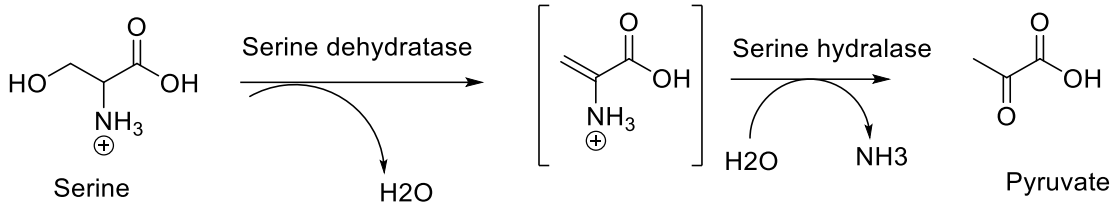
Amino acid	Ketogenic/Glucogenic	Product of degradation
Asparagine	Glucogenic	Oxaloacetate
Leucine	Ketogenic	Acetyl-CoA
Phenylalanine	Ketogenic and glucogenic	Acetoacetate, fumarate
Serine	Glucogenic	Pyruvate or 3-phospho-glycerate

We will start with asparagine since the breakdown of asparagine is relatively simple. This process is outlined below:

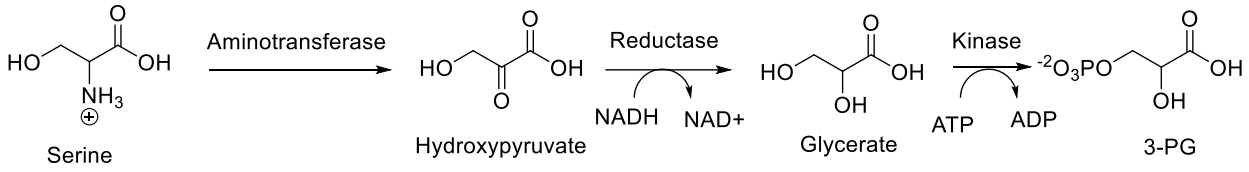


First, the asparagine is hydrolyzed to give the carboxylic acid from the amide. This gives aspartate as the product. Then, the NH_3 group is donated and oxaloacetate is formed using an aminotransferase.

Serine is the next simplest one, however, serine has two degradation pathways. One that produces pyruvate and another that produces 3-phosphoglycerate. We will discuss the pyruvate pathway first and then discuss the 3-phosphoglycerate pathway.

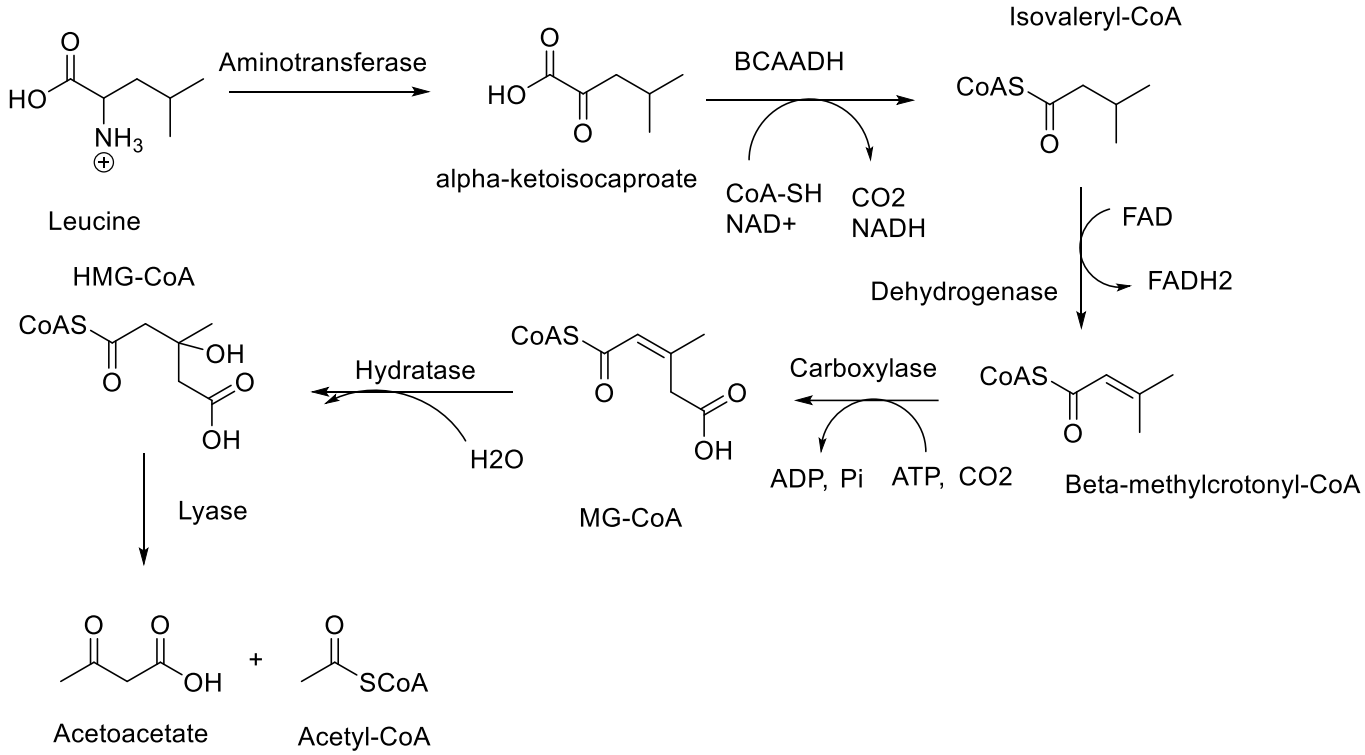


First, the serine is dehydrated using serine dehydratase. This yields the unstable intermediate aminoacrylate, which will spontaneously get hydrolyzed to give pyruvate by effectively replacing the NH_3 group with OH and the tautomerizing to give the ketone.



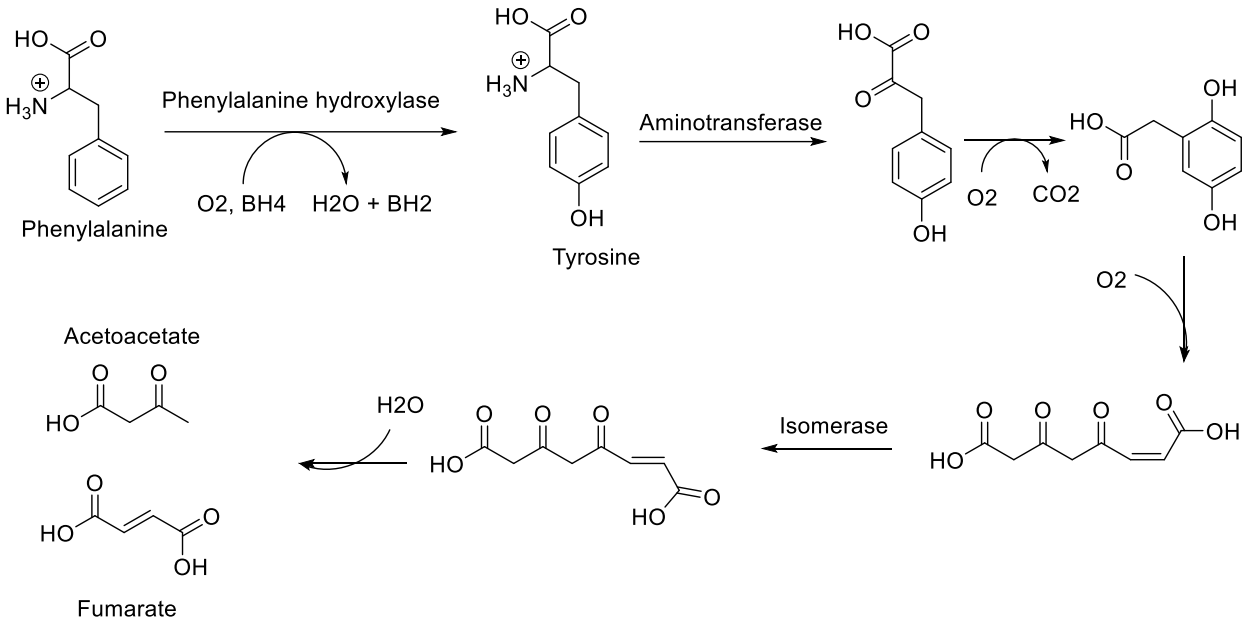
In the 3-PG pathway, the serine donates its amino group using an aminotransferase to make hydroxypyruvate. This gets the molecule closer to glycerate, which we can then phosphorylate to make 3-PG. To make it glycerate, however, the carbonyl carbon needs to get reduced, so a reductase is used to reduce the ketone to an alcohol. This glycerate is then phosphorylated using a kinase to yield 3-PG, this reaction is powered by ATP hydrolysis (M**A**SC). This 3-PG is an intermediate in glycolysis and can go on to yield ATP and pyruvate.

The leucine degradation pathway is a bit more complicated because it is a ketogenic amino acid. As such, it wants to get closer to an intermediate of ketone body synthesis. In the case of leucine, it wants to get closer to HMG-CoA so that it can be lyased and release acetoacetate and acetyl-CoA. The leucine degradation pathway is shown below:



Each step of the leucine degradation pathway gets the compound to resemble HMG-CoA more and more. First, the NH_3 group needs to get removed because there is no nitrogen in HMG-CoA, to do this, an aminotransferase is used to make alpha-ketoisocaproate. Then, one of the carbonyl groups needs to get removed and one of them needs to get replaced with a thioester. This is accomplished using branched chain alpha-keto acid dehydrogenase (BCAADH) using free CoA-SH and NAD^+ to create isovaleryl-CoA. This isovaleryl-CoA is dehydrogenated and a new carbon-carbon pi bond is made. This new pi bond is instrumental in allowing the OH to be added to form HMG-CoA in the end and for allowing the carboxylation reaction to occur farther away from the thioester carbon. Ultimately, HMG-CoA is made and is subsequently lyased to make the two ketone bodies it desires.

Finally, we will discuss the degradation of phenylalanine. There are many reactions in this pathway, but the only one that is super important is the phenylalanine hydroxylase reaction. This reaction allows the degradation of phenylalanine and tyrosine to be linked and this minimizes the number of enzymes the cell has to make to degrade the two of them (MAS**C**). An overview of the phenylalanine degradation pathway is shown below. Unlike other amino acid degradation pathways, we will not go in-depth on this method of degradation.

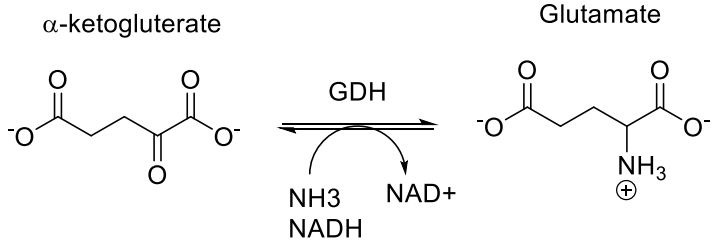


Many genetic diseases are linked to this pathway. If phenylalanine hydroxylase no longer functions, that leads to phenylketonuria, which causes impaired neurological development.

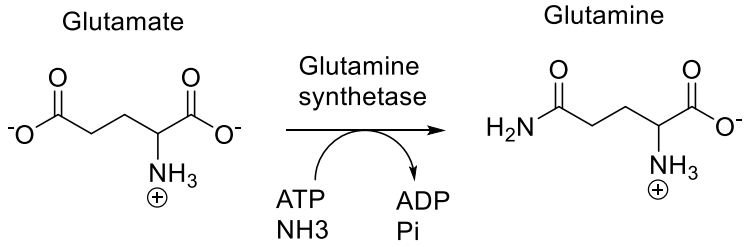
Chapter 38: Amino Acid Biosynthesis

Making amino acids is incredibly important for cells to create large proteins. Here, we will discuss amino acid biosynthesis of several amino acids and also discuss how amino acids serve as precursors to neurotransmitters. These pathways mostly get their nitrogen from nitrogen fixation, which is the process by which N_2 is reduced to NH_3 by nitrogen fixing bacteria. This NH_3 can get incorporated into important biomolecules such as amino acids that are then used to make proteins for the organism. The organisms that benefit from nitrogen fixing bacteria can themselves be eaten and in this way "fixed nitrogen" is available to all forms of life. In this chapter, we will focus our attention on the creation of glutamate, glutamine, proline, arginine, serine, glycine, and cysteine. A lot of these reactions are things that we saw before in the urea cycle and amino acid breakdown, but in reverse. In this way, the cell optimizes its efficiency by minimizing the number of enzymes it uses (MAS**C**).

First, we will discuss glutamate and glutamine synthesis since just like in the urea cycle, this is how most nitrogen gets funneled to make amino acids. Glutamate, as we previously discussed can be dehydrogenated using NAD^+ to make α-ketogluterate using glutamate dehydrogenase. What we did not discuss is that this reaction is actually reversible and therefore you can also go from α-ketogluterate to glutamate using the same enzyme in the reaction shown below:



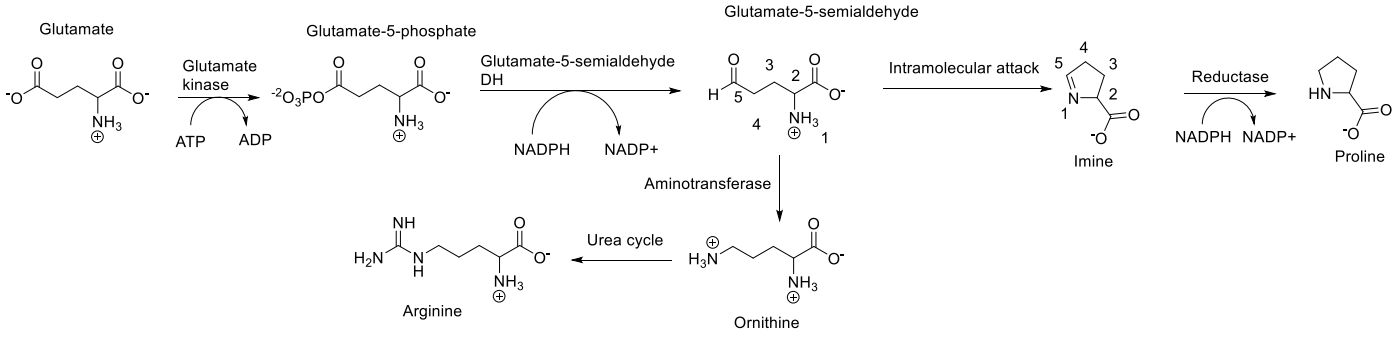
Once this glutamate is made, we can create other amino acids such as glutamine. Glutamine is made from glutamate using glutamine synthetase in the reaction shown below:



ATP is used to phosphorylate the carboxyl group on the left hand side to activate the carbonyl so that the ammonia can be used to displace the phosphate and create the amide. Recall from organic chemistry that carboxylates are incredibly unreactive and therefore creating the amide requires ATP since we are creating a more reactive compound (M**A**SC). Since this reaction uses ATP, the glutamine synthetase enzyme is highly regulated. Large glutamine levels allows the cell to create many amino acids and other urea cycle intermediates, as such, these molecules will allosterically inhibit glutamine synthetase. Some of these molecules are listed below:

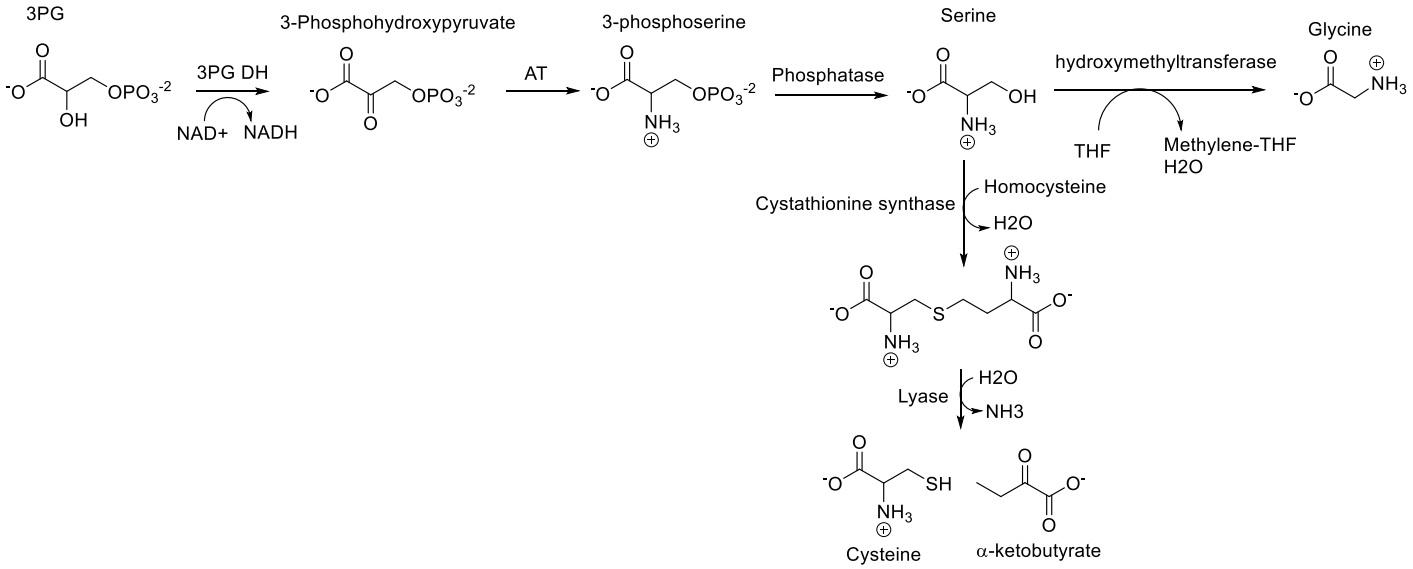
Molecule	Inhibitor/Activator?	Why?
AMP	Inhibitor	Glutamine synthetase uses energy, if energy low we don't want to use it as much.
Histidine	Inhibitor	Downstream product, if we have a lot already, no need to make glutamine.
Tryptophan	Inhibitor	Same as histidine
Carbamoyl phosphate	Inhibitor	Indicates the cell is actively trying to get rid of NH_3 not use it.
Glycine	Inhibitor	Tells the cell that we have excess amino acids already, no need to go through amino acid synthesis
Alanine	Inhibitor	Same as glycine

Glutamate can be used not only to make glutamine, but also proline and arginine. The synthesis of arginine uses the reactions of the urea cycle to bring ornithine to arginine and proline uses an intramolecular attack to make the ring. An overview of their synthesis is shown below:



The first step in both proline and arginine synthesis is the phosphorylation of glutamate by glutamate kinase. This step requires ATP to be used because we are activating the carbonyl carbon and thus making it more reactive (M**A**SC). This activated carbonyl carbon can get reduced using the hydride from NADPH in the reaction catalyzed by the glutamate-5-semialdehyde DH reaction. This creates an aldehyde on the left hand side of the compound, which is a reactive electrophile. This aldehyde group can be attacked by the amino group giving an imine (remember your orgo), which can get reduced to the amine using a reductase (remember reductive amination from orgo). This gives proline directly. Arginine synthesis is a bit different, rather than having the amino group attack the aldehyde, the aldehyde group is exchanged for another amino group to give ornithine. Ornithine, as we just discussed, is an intermediate in the urea cycle, the product of which is arginine. In this way, the cell uses the urea cycle for two main purposes: first, to get rid of excess ammonia as urea and second, to create arginine for use in polypeptides (MAS**C**).

Now we will discuss a more difficult chain of reactions, though some of these reactions you have seen before but in reverse. Here, we will discuss the synthesis of serine, glycine, and cysteine. Just like how glutamate, glutamine, proline, and arginine use α-ketogluterate as a starting material, serine, glycine, and cysteine use 3-phosphoglycerate as theirs. An overview of the synthesis of these amino acids are shown below:

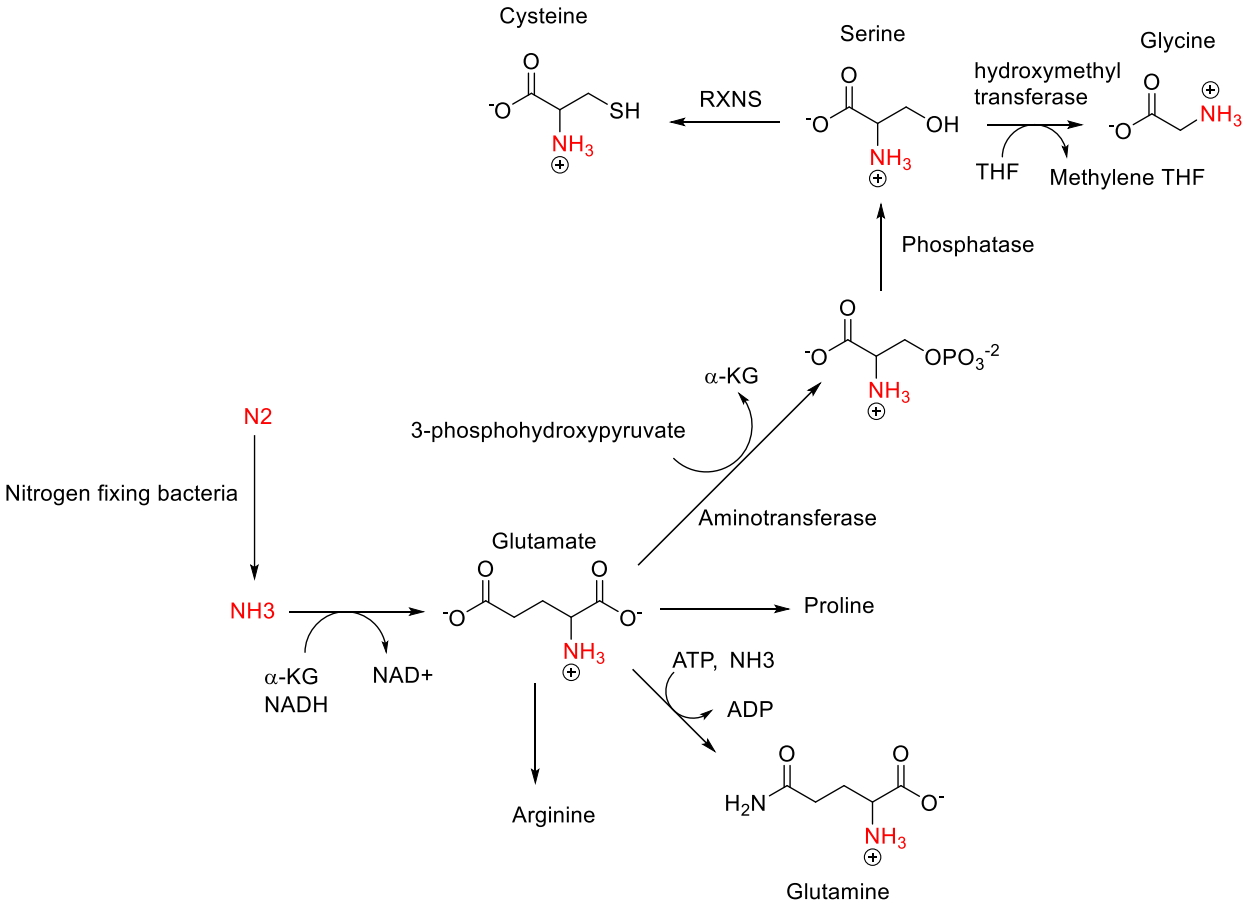


The very first step is to oxidize the alcohol group of 3PG using 3PG DH. This is important because we need to add the amino group to the carbon chain to make the amino acids we want. Once the alcohol group is oxidized to the ketone, an amino transferase can replace the ketone with an amine group to give 3-phosphoserine. This 3-phosphoserine is almost where we want it, all we need to do now is remove the phosphate using a phosphatase and we can get serine directly.

Serine is a versatile precursor to glycine and cysteine. Glycine can be made directly from serine using a hydroxymethyltransferase. This enzyme uses a tetrahydrofolate cofactor (THF), not to be confused with tetrahydrofuran that we see as a common organic solvent (damn you biochemists!). THF is an incredibly important cofactor for "one carbon metabolism", which is effectively the way cells can add or remove a single carbon from one compound to another. Methylene THF is incredibly important, as we will see soon enough, for thymidine biosynthesis while formyl THF is incredibly important for all purine biosynthesis (A and G bases). Disruption of one carbon metabolism was commonly used to develop chemotherapeutic agents at the dawn of the war on cancer (details to come in the Cancer focus). Cysteine is a bit more difficult to make from serine. First, the serine needs to combine with homocysteine to make cystathionine. This cystathionine has a sulfide bond that can be cleaved using a lyase to make the thiol and remove the ammonia of the longer segment of the sulfide to make it a ketone. This makes cysteine and α-ketobutyrate and in this way, the cell can make cysteine from serine.

PLP is not only used for aminotransferases, but it is also used in decarboxylases because of its ability to act as an electron acceptor (recall from orgo that decarboxylation reactions create lone pairs). These PLP-dependent decarboxylases are frequently used to make neurotransmitters from amino acids. Glutamate, for example, can be decarboxylated to give the neurotransmitter GABA. Histidine can be decarboxylated to give the neurotransmitter histamine, etc. etc.

Here is an overview of the nitrogen flow in living organisms:



Here are the key takeaways from amino acid biosynthesis:

1. Glutamate is the funneling point for ammonia and can be used to make arginine, proline, and glutamine.
2. Glutamine synthetase is highly regulated by energy demands (AMP-dependent) and amino acid levels (histidine, tryptophan, glycine, and alanine).
3. Nitrogen fixing bacteria take useless N_2 and make it into useful NH_3

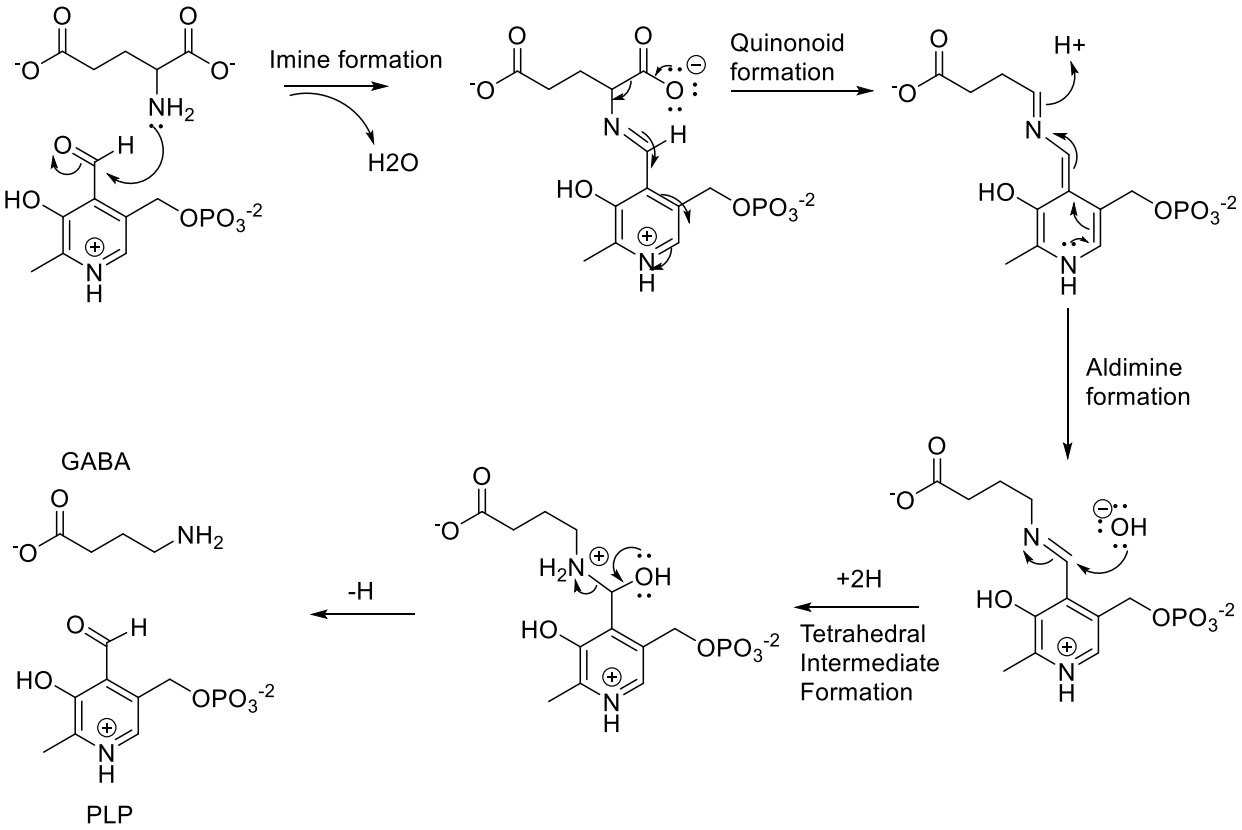
Practice Questions:

1. What is the energetic cost (in terms of ATP and redox cofactors) of synthesizing one arginine from glutamate? Proline?

2. Provide a reasonable mechanism to show how PLP can be used to facilitate decarboxylation. (Hint: it is very similar to how PLP is usually used in aminotransferases)
3. Fill in the blanks: Atmospheric nitrogen is reduced to _____ (chemical) using _____ (type of bacteria). This _____ (chemical) is incorporated into biomolecules such as _____ (chemical) using _____ (enzyme). This compound is important for several pathways, such as _____ (chemical) synthesis and the _____ (pathway). If there is excess ammonia in the cell, _____ (pathway) is promoted. This pathway starts by releasing the ammonia using _____ (enzyme). This free ammonia first has to get attached to an activated carbon using _____ (enzyme) at the expense of _____ (high energy molecule) and _____ (gas) before it can enter the _____ (pathway). Ultimately, in _____ (pathway), the ammonia is secreted as _____ (molecule). If ammonia is at normal levels and the cell needs to make proteins, _____ (pathway) is promoted. Glutamate is a versatile precursor to many chemicals such as _____ (chemicals).

Answers:

1. ATP cost of synthesizing one arginine from glutamate: 1 ATP from glutamate kinase, 1 ATP from argininosuccinate synthetase. NADPH cost: 1 NADPH from glutamate-5-semialdehyde DH. NADH gain: 1 NADH if we use fumarate in the Krebs cycle. Overall energetic cost: 2 ATP and 1 NADPH lost but 1 NADH gained. If we convert the redox cofactors to ATP equivalents, we get the following: 2.5 ATP – 2 ATP – 0 ATP = 0.5 ATP gained. An important thing to know is that NADPH CANNOT be used in the ETC, so it does not give any ATP. Proline ATP cost: 1 ATP from glutamate kinase. Proline NADPH cost: 1 NADPH from glutamate-5-semialdehyde DH and 1 NADPH from the reductase. Overall energetic cost: 1 ATP, 2 NADPH.
2. We will use glutamate as an example here to show how we can make GABA

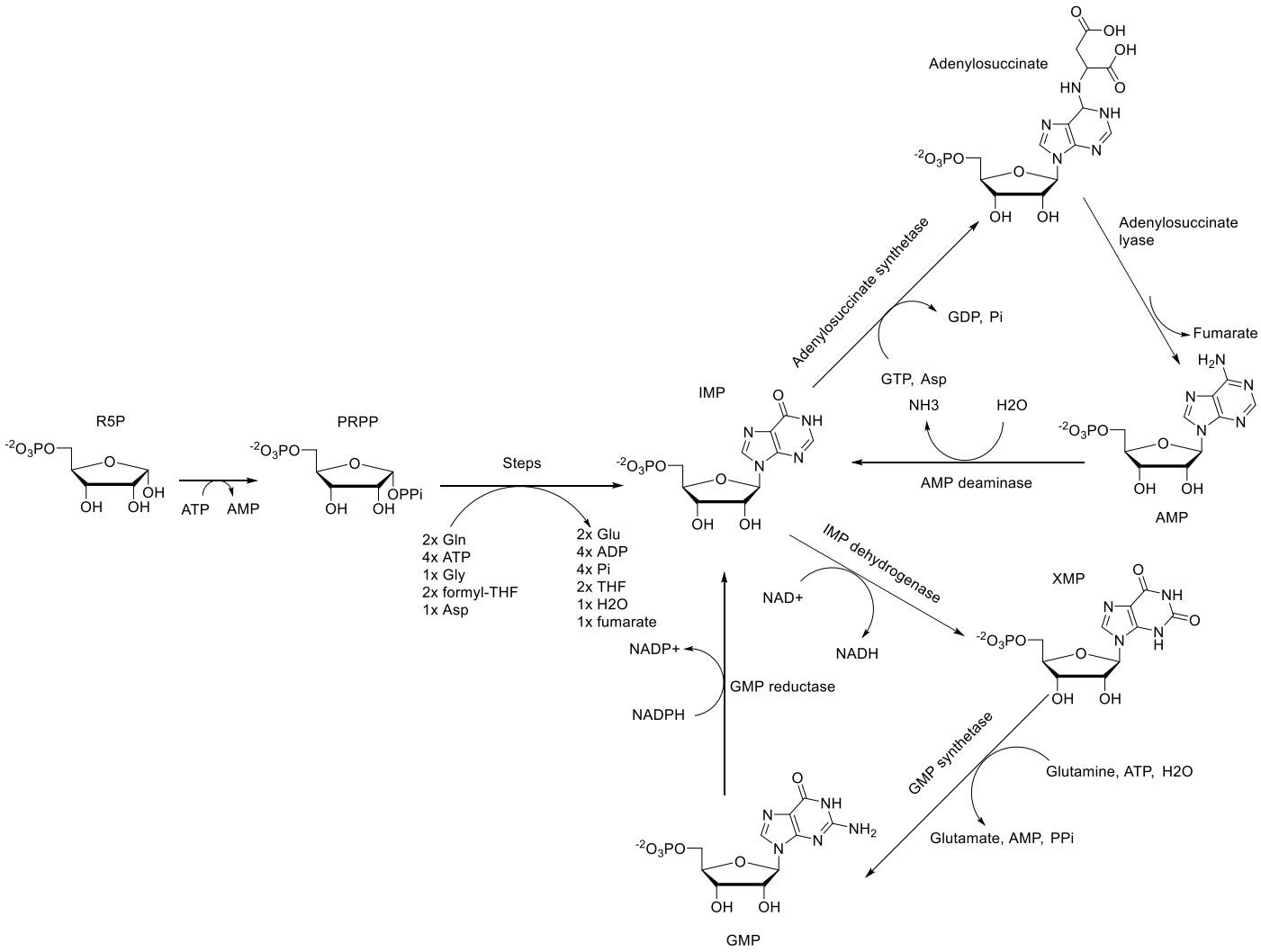


3. Atmospheric nitrogen is reduced to NH_3 or ammonia (chemical) using Nitrogen fixing bacteria (type of bacteria). This ammonia (chemical) is incorporated into biomolecules such as glutamate (chemical) using glutamate dehydrogenase (enzyme). This compound is important for several pathways, such as amino acid (chemical) synthesis and the urea cycle (pathway). If there is excess ammonia in the cell, the urea cycle (pathway) is promoted. This pathway starts by releasing the ammonia using glutamate dehydrogenase (enzyme). This free ammonia first has to get attached to an activated carbon using CPS-1 (enzyme) at the expense of ATP (high energy molecule) and CO_2 (gas) before it can enter the urea cycle (pathway). Ultimately, in the urea cycle (pathway), the ammonia is secreted as urea (molecule). If ammonia is at normal levels and the cell needs to make proteins, amino acid synthesis (pathway) is promoted. Glutamate is a versatile precursor to many chemicals such as glutamine, proline, arginine, and technically serine, glycine, and cysteine (chemicals).

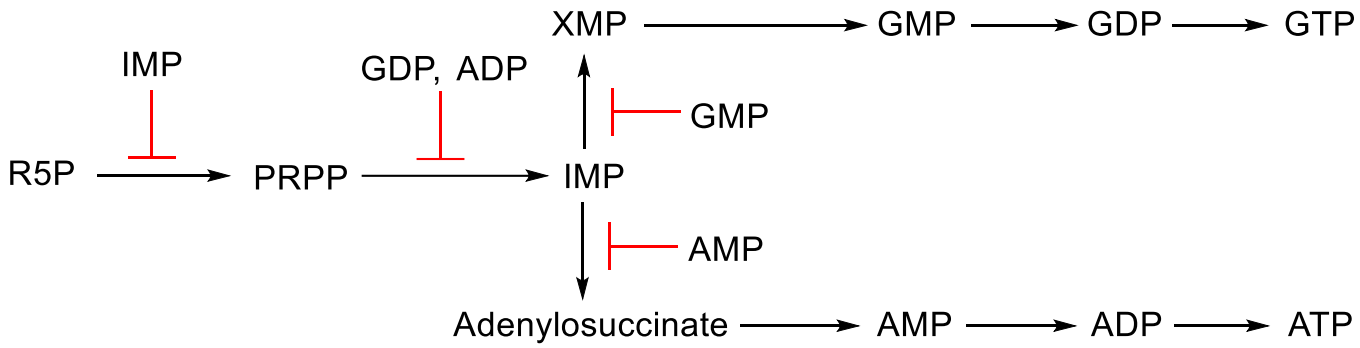
Focus 8: Nucleic Acid Metabolism

Chapter 39: Nucleic Acid Biosynthesis (Purines)

Now that we have discussed how amino acids, fats, and sugars are broken down and synthesized, it is time that we start discussing nucleic acids. But first, some definitions. Nucleic acids are subdivided into two categories, purines and pyrimidines. Purines are the A and G nucleotides and are characterized by a fused bicyclic nitrogenous base while pyrimidines are the T and C nucleotides that have a single ring nitrogenous base. In DNA, A pairs with T and G pairs with C and in RNA, T is replaced with U. These interactions are accomplished through very strong hydrogen bonding interactions that pull the two bases together. The G and C base pairs have 3 major hydrogen bonding interactions making them stronger than the A and T base pairs since those only have 2. In this chapter, we will discuss how the A and G nucleotides are made from scratch aka *de novo*. This process, as you can imagine, is incredibly costly and therefore only cells that have a lot of energy (or in the case of cancer, *think*, they have a lot of energy) will undergo this process. This pathway is also most active when a cell is actively dividing because nucleotides are necessary for replicating DNA to make daughter cells. An overview of purine biosynthesis is shown below:

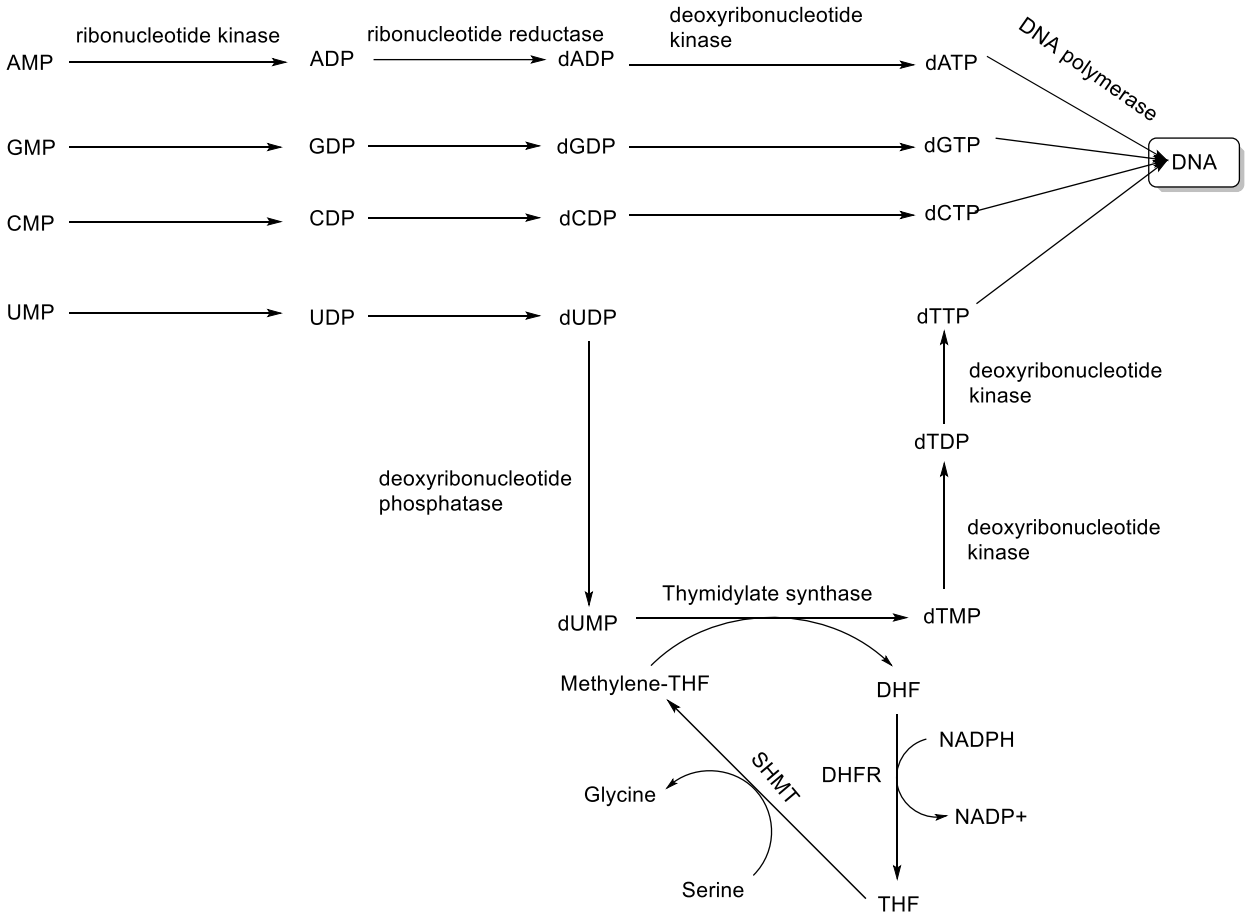


The specific steps to make the IMP are not necessary to know (there are too many), however, it is important to know that formyl-THF is required to make purines and that IMP is the major intermediate nucleotide that can be interconverted to either GMP or AMP. A lot of ATP is used because many carbon-carbon bonds are formed and therefore a lot of energy is required to make IMP. It is also important to know that the R5P or ribose-5-phosphate is the end product of the pentose phosphate pathway (blast from the past). Another important aspect of the interconversion of IMP to either GMP or AMP is that the opposing nucleotide triphosphate is used to catalyze their formation. For example, to make AMP, GTP is used to power the adenylosuccinate synthetase reaction and to make GMP, ATP is used to power the GMP synthetase reaction. In this way, ATP and GTP levels keep the other nucleotide in balance and the cell can maintain roughly equal concentrations of both (MAS**C**). This entire pathway is highly regulated so that nucleotides are not made when they are not necessary and that the levels of A and G nucleotides balance each other out. This regulatory mechanism is shown below, the red lines indicate inhibition:

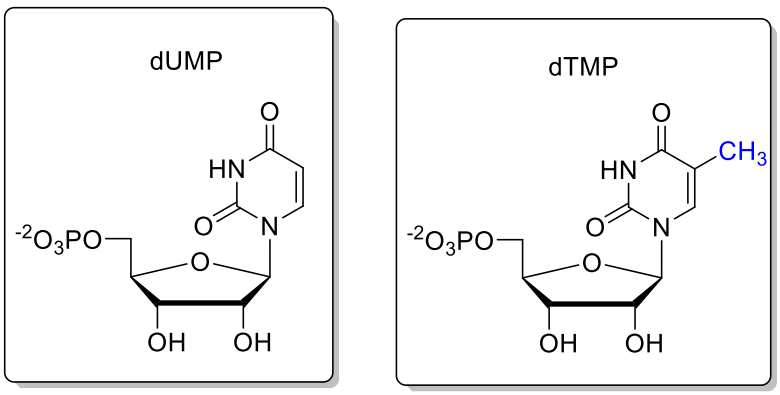


The common theme throughout all of this regulation is that A and G nucleotides and precursors thereof inhibit steps upstream in their production pathway. If IMP levels are high, for example, then the cell should not waste its ATP making PRPP and should instead do other processes (MAS**C**).

Once the GMP and the AMP are made, they can be phosphorylated to GDP and ADP using ribonucleotide kinases. If the cell wants to make nucleotides for DNA replication, however, the 2'OH group must be removed to make dNTPs. To do this, ribonucleotide reductases are used to make the deoxy analogues of GMP and AMP. The same can be done for CMP and UMP. However, UDP needs to be further processed to dUMP so that it can get methylated by thymidylate synthase to make dTMP. This dTMP can ultimately get kinased up to the triphosphate form (dTTP) and then get incorporated into DNA. A summary of this process is shown below:

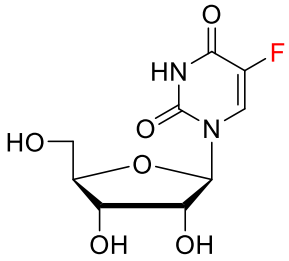


Here, the dUMP is converted to dTMP using thymidylate synthase. This enzyme uses a methylene-THF cofactor that allows for the enzyme to methylate the uridine base to make it a thymidine base. The uridine and thymidine bases are shown below in a side-by-side comparison:



The product of the thymidylate synthase reaction is dTMP and dihydrofolate (DHF). The methylene-THF needs to be recycled so that the enzyme can continue to function. The DHF is made back into the methylene-THF in a two-step process: first, the DHF is reduced back to THF using dihydrofolate reductase (DHFR), then the THF is made into methylene THF using the serine hydroxymethyl transferase enzyme that we just discussed in the previous chapter.

Because thymidine biosynthesis follows a different path than the rest of the nucleotides that are used to make DNA, this recycling pathway has been the target of many anticancer drugs. The most prominent of which is 5-fluorouracil shown below:



5-fluorouracil is a prodrug, meaning that it is inactive by itself, but becomes active once the body performs some chemistry on it. In the case of 5-fluorouracil, it becomes activated once it is phosphorylated and depending on the degree of phosphorylation, the drug will have different effects.

If it is 5-F-dUMP, it will permanently deactivate thymidylate synthase because in the reaction mechanism, the enzyme covalently bonds to the nucleotide and then the hydrogen that is supposed to be where the fluorine is gets removed. C-F bonds are incredibly strong and do not break as easily as C-H bonds and therefore the drug stays permanently attached to the enzyme, deactivating it in the process.

If it is 5-F-dUTP, it will trick the DNA polymerase (the DNA making machine) and become incorporated in the DNA double helix in place of T. This may not sound bad, however, 5-F-dUTP has nonstandard hydrogen bonding interactions with G bases and therefore it causes mutations to occur during replication (C goes to T). The more mutations that occur in the cell the more likely one of those mutations proves fatal and therefore the cancer cell dies because those are the cells that are dividing most rapidly.

Here are the key takeaways from purine biosynthesis:

1. It is very energetically expensive (uses up a lot of the cell's resources)
2. It starts with R5P being pyrophosphorylated (PP group is added) to PRPP.
3. IMP is the critical intermediate that can be made into AMP or GMP depending on the needs of the cell.
4. ATP powers the production of GTP while GTP powers the production of ATP so that the cell keeps a roughly equal concentration of both.
5. Nucleotides are made into bases that are appropriate for DNA incorporation using ribonucleotide reductases, thymidylate synthase, and deoxyribonucleotide kinases
6. Thymidylate synthase's methylene-THF cofactor needs to be recycled using DHFR and SHMT.

Practice Questions:

1. 5-fluorouracil was originally used as an anticancer drug, however, one of its major side effects was its nasty cytotoxicity (kills healthy cells as well as cancer cells) and therefore

many patients receiving 5-fluorouracil suffer from some bad side effects. Explain 5-fluorouracil's toxicity on the premise of its mode of action and structure.

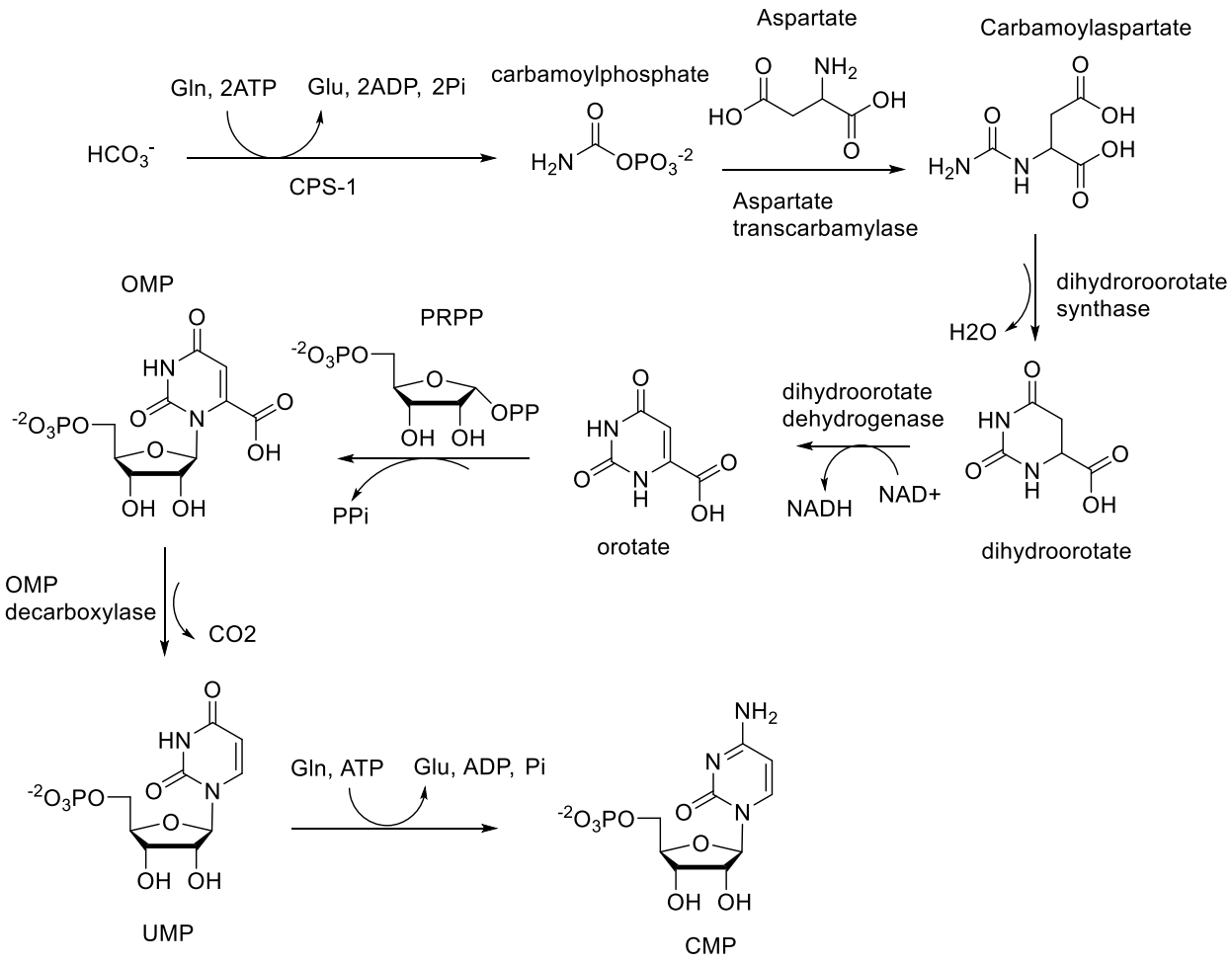
2. Suppose you developed a drug that would inhibit DHFR, how would this effect nucleotide biosynthesis and the proliferation of cells? Explain.
3. Many of these reactions of purine biosynthesis require NADPH, where does this NADPH come from? What cellular processes produce the NADPH that is used in nucleotide biosynthesis?

Answers:

1. 5-fluorouracil is cytotoxic because it simply prevents the production of T nucleotides so that DNA replication is halted by permanently deactivating thymidylate synthase. This is in no way cancer specific, the only reason why cancer cells are disproportionately affected by this drug is because those cells are the ones that are dividing constantly and therefore make the most use out of their thymidylate synthase. Because of its nonspecific mode of action, it will prevent the replication of not only cancer cells, but also healthy cells. Therefore, it will cause bad side effects. (we will discuss how scientists are trying to work around this issue later on).
2. If you inhibited DHFR, then you would disrupt thymidine synthesis and therefore DNA replication would grind to a halt. Because of this, cell division would be stalled and purine biosynthesis would also be halted. Without THF, methylene-THF cannot be made and therefore thymidylate synthase no longer works.
3. The NADPH that is used in purine biosynthesis comes from the PPP and malic enzymes in the cell.

Chapter 40: Nucleic Acid Biosynthesis (Pyrimidines)

So we just discussed how A and G nucleotides are made for DNA synthesis and for energy molecules such as ATP and GTP, here we will discuss how C and U (or T) are made in the body. Luckily this process is a bit less complicated than the purine synthesis part and as such this chapter will be much shorter than the previous. The same rules and regulations apply for pyrimidine synthesis as before, this is only done if the cell has a lot of energy and needs to divide because making these nucleotides costs a lot of energy. At the very end of this chapter, we will do a side-by-side comparison between the two synthesis pathways and highlight the ways in which they are both similar and different from one another. An overview of the pyrimidine synthesis pathway is shown below:



Net cost for UMP: 3ATP, 1Gln, 1Asp, 1NAD+, 1PRPP
Net cost for CMP: 4ATP, 2Gln, 1Asp, 1NAD+, 1PRPP

Extra ATP for the UMP and the CMP is taking into account the 1 ATP cost to make PRPP from R5P

The very first reaction uses an enzyme that we had already discussed, CPS-1. Just like in the urea cycle chapter, this enzyme is responsible for the production of carbamoylphosphate. This is

powered using ATP because we are making an unreactive HCO_3^- into a very reactive carbamoyl phosphate (M**A**SC). This carbamoylphosphate has an activated carbonyl carbon, which can combine readily with aspartate to give carbamoylaspartate. This is important because aspartate has the carboxylic acid groups in the right orientation to allow the nitrogen of the urea group to perform an intramolecular attack to form dihydroorotate. This does not require ATP for two reasons: first, the ring size that is formed is favorable since six-membered rings are stable and second, carboxylic acids are more reactive than ureas. Then to make the C-C pi bond, dihydroorotate dehydrogenase is used. This is uncharacteristic, typically if we are oxidizing a carbon-carbon bond, FAD is used as the redox cofactor. This is a unique scenario; however, because the carbon-carbon pi bond that forms is highly stabilized through resonance and so NAD^+ can be used here. This makes orotate that can easily attack the pyrophosphorylated carbon of PRPP with its amide nitrogen to make OMP. This OMP can be decarboxylated to produce UMP and then further reacted with glutamine and ATP to produce CMP.

Here is a side-by-side comparison of the two nucleotide biosynthesis pathways:

Characteristic	Purine Synthesis	Pyrimidine Synthesis
Needs PRPP	True	True
Needs formyl-THF	True	False
Needs glycine	True	False
Needs alanine	True	False
Needs aspartate	True	True
Needs bicarbonate	False	True
Forms the ring first then adds the ribose	False	True
Requires the same number of ATP equivalents (NTPs) regardless of the nucleotide made	True	False
Is more energetically expensive	False	True
Name of critical intermediate	IMP	OMP
Can go back to intermediate	True	False

Here are the key takeaways from the pyrimidine synthesis pathway:

1. UMP costs 1 less ATP to make than CMP
2. Pyrimidine and purine synthesis require PRPP and aspartate to work
3. In pyrimidine synthesis, the preformed ring is added directly to PRPP while in purine synthesis the ring is made while attached to PRPP.
4. OMP is the critical intermediate for pyrimidine biosynthesis and its downstream products, UMP and CMP are made irreversibly and linearly.

Practice Questions:

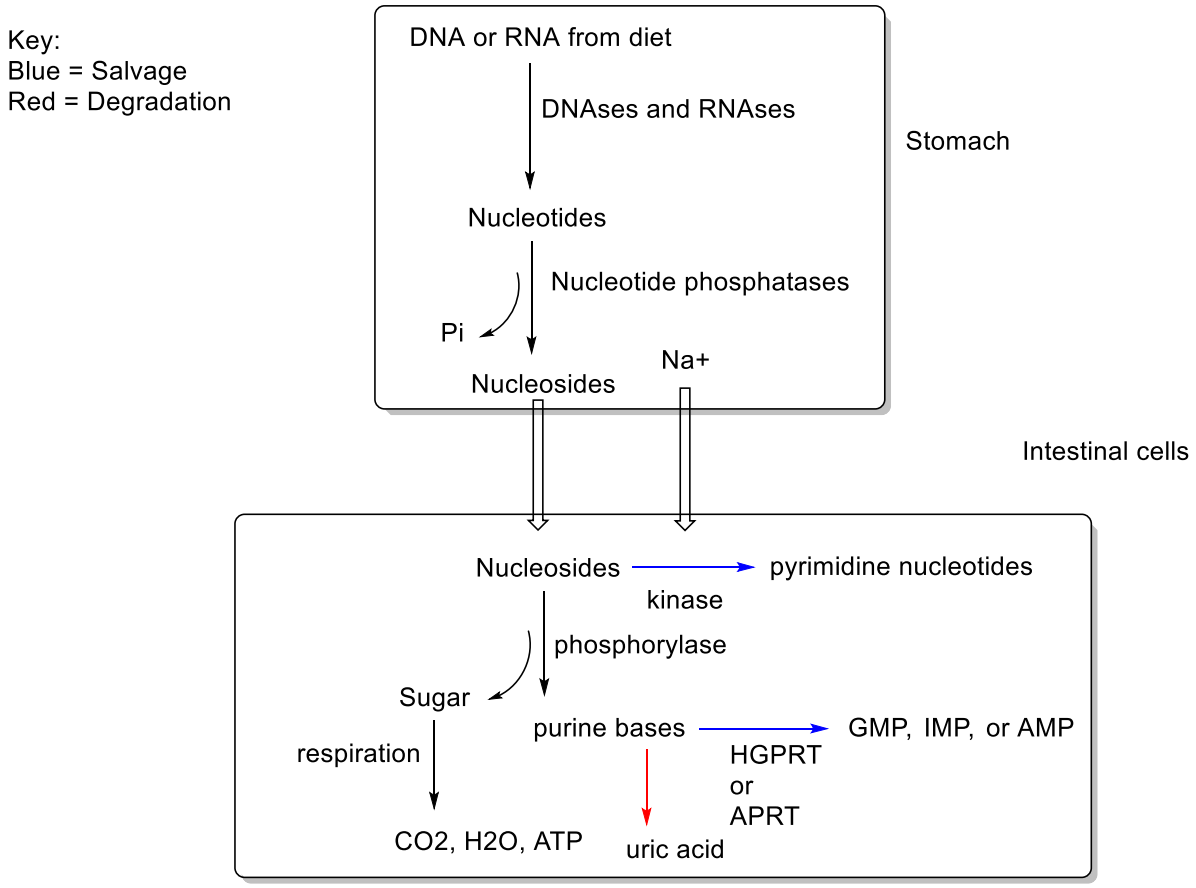
1. Suppose enzyme X makes formyl-THF and that you make a compound that inhibits that enzyme so that it no longer functions. Explain the impacts this would have on purine and pyrimidine synthesis and explain how you could determine the type of inhibition the compound displays (competitive versus noncompetitive versus uncompetitive).
2. Oncogenes are genes that code for a protein that, when overactive, give rise to rapid cell growth. Tumor suppressor genes are genes that code for a protein that, when underactive, give rise to rapid cell growth. Would the genes that code for the enzymes in purine or pyrimidine synthesis be considered oncogenes or tumor suppressor genes? Explain.

Answers:

1. Purine synthesis would be halted because it requires formyl-THF to work while pyrimidine synthesis would be unaffected because it does not require formyl-THF. To determine the type of inhibition, the substrate of the enzyme would be varied and the initial rates would be measured in the presence and absence of the inhibitor, then a Lineweaver Burke analysis can be performed to determine the K_m and V_{max} of the enzyme with and without the inhibitor. If the K_m increases and the V_{max} stays the same, the inhibitor is competitive. If the K_m and V_{max} decrease by the same factor, then the inhibitor is uncompetitive. If the K_m stays the same but the V_{max} decreases, then the inhibitor is noncompetitive.
2. The genes that code for the enzymes in these pathways would be considered oncogenes because if they were overactive, more and more nucleotides would be made that can be used to make new DNA during replication. This would allow for more rapid cellular division aka tumorigenesis.

Chapter 41: Nucleic Acid Breakdown and Salvage

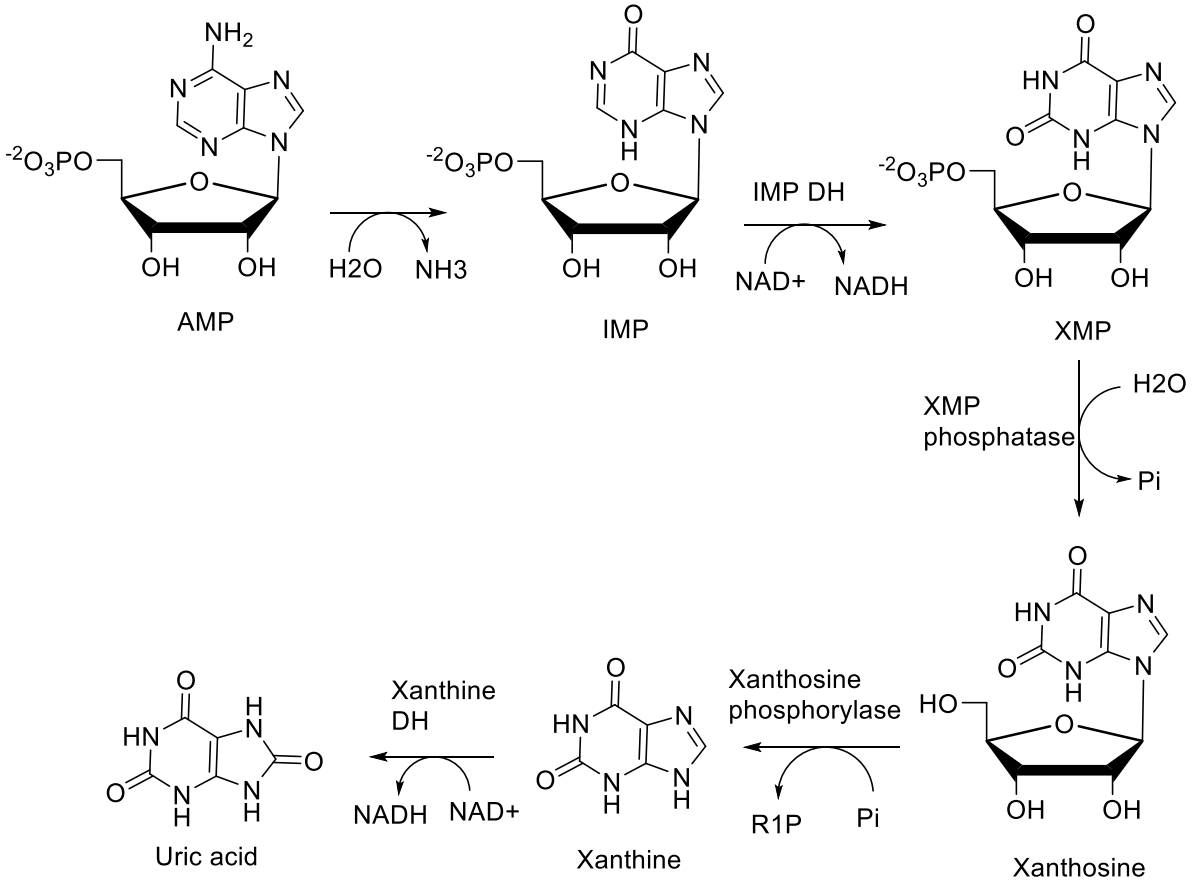
Now that we determined how nucleotides are made from scratch, we can discuss how organisms can salvage the nucleotides they ingest in their diet and how excess purine nucleic acids are broken down into uric acid. Many parasites lack a *de novo* nucleotide synthesis pathway and therefore rely exclusively on the salvage pathway. Because of this, the salvage pathway is an attractive target for antiparasite drug development. Luckily, the salvage pathway is relatively straightforward. The breakdown pathway uses many of the same reactions that we have already discussed and therefore is also more straightforward than you would think. This is a byproduct of the cell's natural tendency to minimize the number of enzymes it has to perfect over its evolutionary history (MAS**C**). An overview of the salvage pathway is shown below:



In the stomach and in the mouth, there are DNAses and RNAses that can break down DNA and RNA into nucleotides. These nucleotides will get their phosphate group removed to give nucleosides by using the phosphatases in the stomach. Ultimately, these nucleosides are moved into the intestine by a Na^+/nucleoside symporter, then the path diverges for pyrimidine and purine nucleotides. Pyrimidine nucleotide salvage is very simple, the nucleoside is simply phosphorylated using a protein kinase to get back the nucleotide. Purine salvage is a bit more difficult because it goes through two steps rather than one. Purine nucleosides in the intestinal cells gets its ribose or deoxyribose removed using a phosphorylase, this leaves only the base left.

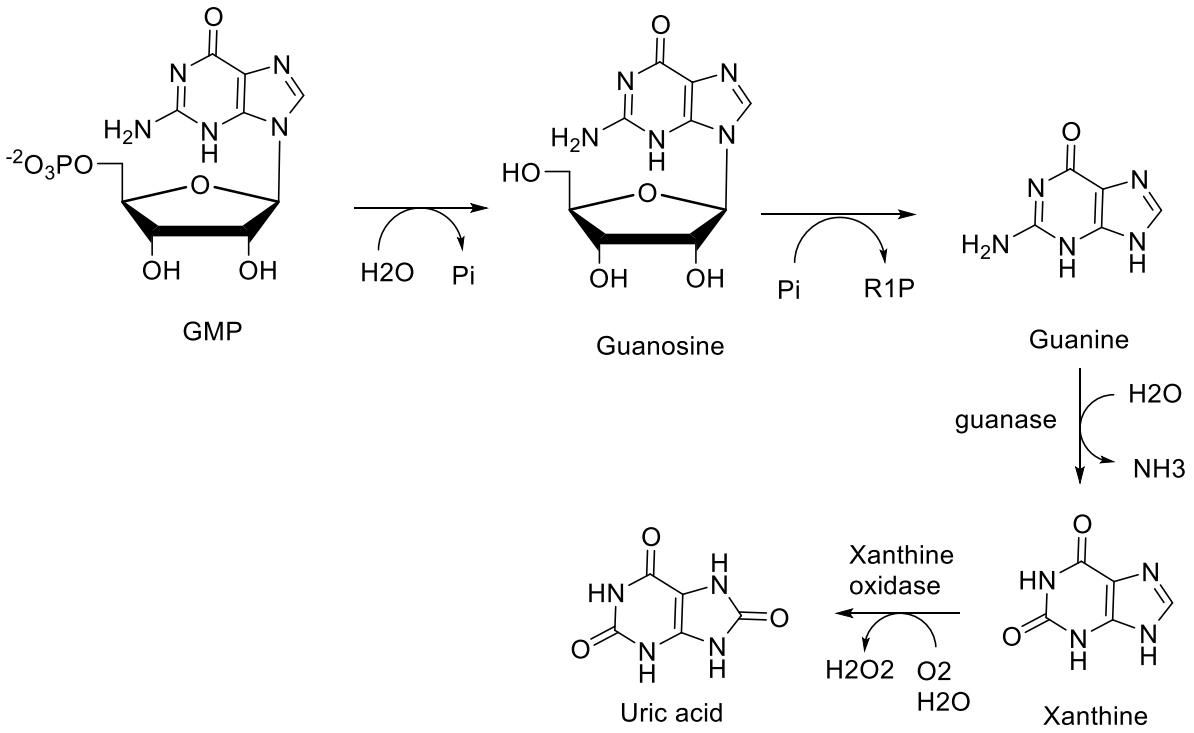
This purine base can either be degraded in the intestinal cell into uric acid or it can be made into GMP, IMP, or AMP using HGPRT and APRT.

Now we can discuss how purine bases are degraded into uric acid. There are many different ways that this can occur, but we will only focus on two of them (one for AMP degradation and one for GMP degradation) and then discuss the connection between uric acid and gout. The first degradation pathway is shown below:



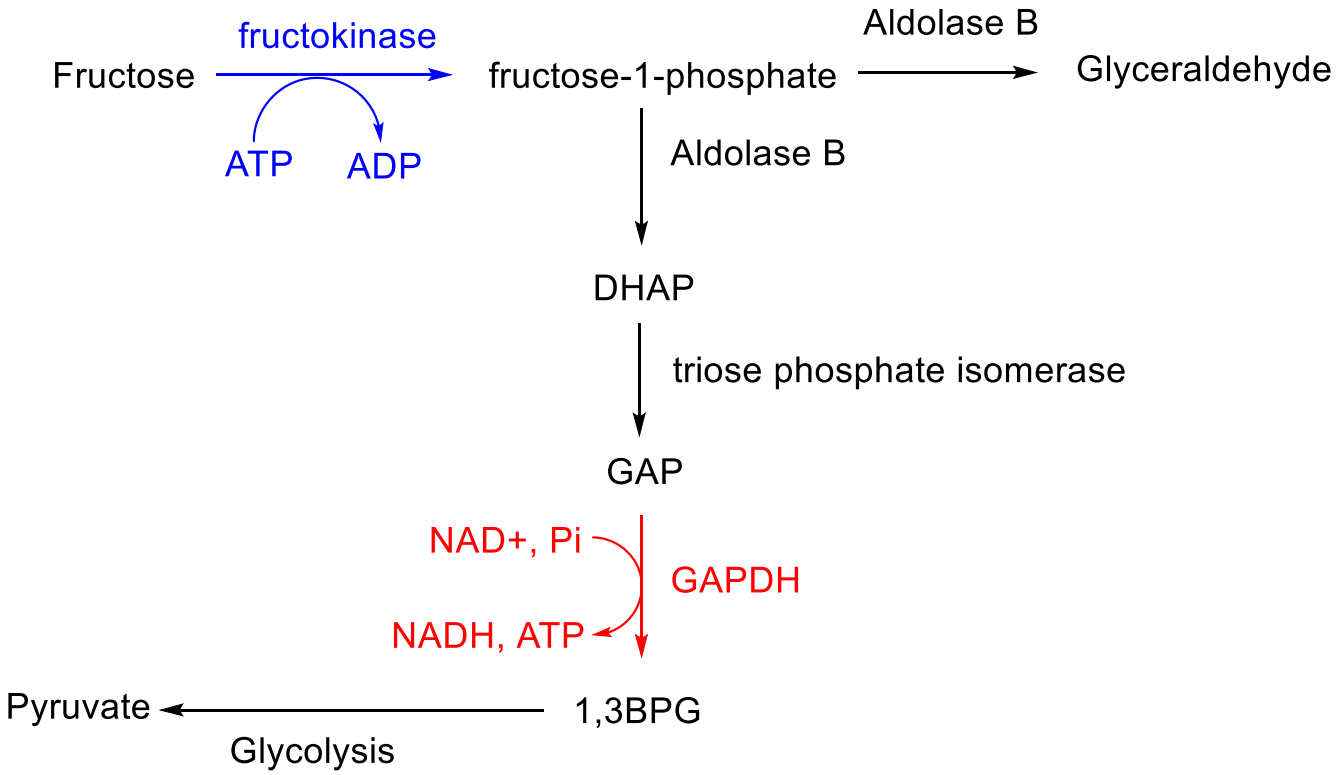
In both degradation pathways, the critical intermediate is xanthine because xanthine can be directly oxidized to give uric acid. Therefore, we need to first make the base resemble xanthine and then remove the ribose group. That is exactly the chemical logic for this degradation pathway. First, the amino group is replaced with a carbonyl since xanthine lacks that amino group, this makes IMP. Then the IMP is oxidized further to XMP and adds an additional carbonyl group adjacent to the amide nitrogen. Now that the base is xanthine, all that needs to happen is that phosphate group needs to get removed using a phosphatase and then the ribose needs to get removed using a phosphorylase. This isolates the xanthine base from XMP and allows it to be oxidized using xanthine dehydrogenase to uric acid.

The second degradation pathway is shown below:

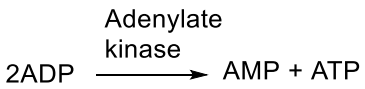


This pathway is much simpler than the previous one because the guanine base can be directly made into the xanthine base using guanase, so all that needs to happen is remove the phosphate and ribose group. The only difference between this pathway and the previous one is the very last step, instead of using xanthine dehydrogenase, xanthine oxidase is used, which makes H_2O_2 rather than NADH. Therefore, the energy gain from degrading GMP is less than degrading AMP.

Sometimes uric acid builds up in the blood and forms crystals. These crystals are incredibly sharp and can cause a tremendous amount of pain. This condition is known as gout and it most prominently appears in the big toe. Many things can cause gout, genetics plays a big role and so does diet. Ultimately, gout occurs when there is too much uric acid and therefore anything that is going to promote purine nucleotide degradation is going to increase the risk of gout occurring. One of the most prominent ways gout can form is a high fructose diet. While this may seem to not have any relation to nucleotide metabolism, we will discuss how ATP and AMP levels can determine how much uric acid is formed in the body in the following diagram, which shows how fructose is metabolized:



In the pathway shown above, the fructokinase reaction is incredibly fast (one could say it is "fast as fuck boi") and therefore the cell rapidly loses ATP and ADP accumulates. However, ATP is necessary to power the fructokinase reaction. Some of the ATP is made from the GADH reaction, however, that reaction is slow and therefore an additional reaction done by adenylate kinase is done to make more ATP to power the fructokinase machinery. The reaction is shown below:



This reaction effectively takes 2 ADP and makes 1 AMP and 1 ATP to make use of the excess ADP generated from the fructokinase reaction. This same AMP can be degraded into uric acid using the pathway we just discussed. The key takeaway is that when there is excess ADP, adenylate kinase will make AMP and ATP. This AMP makes uric acid, which accumulates into crystals and gout develops. This all has to do with the kinetics of the enzyme reactions shown in blue and red above. The blue reaction is several orders of magnitude faster than the red one, therefore ADP accumulates, that sets off adenylate kinase, that makes a lot of AMP and that AMP gets degraded to uric acid then BAM gout.

Here are the key takeaways for this entire chapter:

1. AMP and GMP are degraded to uric acid.
2. Xanthine is the critical intermediate of purine degradation

3. Too much uric acid leads to gout
4. Pyrimidines are salvaged as nucleosides and are kinased up to the nucleotides
5. Purines are salvaged as bases and HGPRT/APRT is used to bring them up to nucleotides

Practice Questions:

1. Gout is partially genetic. What kind of mutation in xanthine oxidase would result in gout being more likely? Increase or decrease in function? Explain.
2. What is the net ATP gain of degrading 21 AMP molecules to uric acid?
3. Would a drug that activates xanthine oxidase or inhibits xanthine oxidase treat gout? Explain.

Answers:

1. Gout is caused by increased uric acid levels. Xanthine oxidase is the enzyme that makes uric acid from xanthine, therefore, people who are genetically predisposed to developing gout have mutations in xanthine oxidase that cause an increase in function.
2. 2NADH molecules per AMP degraded to uric acid, there are 21 AMP molecules getting degraded, therefore there are 42 NADH molecules made. These NADH molecules can each be used to make 2.5 ATP, therefore there is a net gain of 105 ATP.
3. Inhibiting xanthine oxidase would treat gout because xanthine oxidase increases uric acid levels and that is the “problem molecule”. Therefore, inhibiting xanthine oxidase will decrease uric acid production thereby treating gout.

Focus 9: Cancer Biology

Chapter 42: Cancer Fundamentals

Cancer is not just one disease, but thousands. The complexity of cancer has made developing cures extremely difficult and has pushed humankind to understand the biological world in an unprecedented depth. Here, we will describe the basic principles of cancer biology and explain how cancer growth is incredibly remarkable.

Cancer is described by rapid cellular growth. Cancer cells are cells that have gone "rouge" and effectively have giant orgies in the body to rapidly divide and make copies of themselves. I like to think of them as narcissistic rouge cells. These cells are genetically different from other cells in the body because they are able to bypass the normal checkpoints and safeguards that prevent such rapid cell growth. The genes that drive cancer growth are aptly named "driver genes". These driver genes come in two different flavors: oncogenes and tumor suppressor genes. In an ordinary cell, you can think of oncogenes and tumor suppressor genes as the accelerator and the breaks of a car respectively. In an ordinary cell, the tumor suppressor genes prevent cell growth and the oncogenes stimulate cell growth. In a cancer cell, however, the accelerator is constantly on and the breaks are broken. In other words, oncogenes are overactive and tumor suppressor genes are inactive. The end result is that cancer continues to grow and spread regardless of the messages the external environment is telling the cell.

The way that cancer cells get their energy and the metabolic pathways that they invest the most time in is drastically different from normal cells. This is great because it allows drugs to be tailored to the cancer cell's specific metabolic preferences. The first difference between healthy cells and cancer cells was determined by Dr. Otto Heinrich Warburg, who discovered that cancer cells prefer glycolysis and anaerobic respiration (lactate fermentation) over all other means of obtaining energy. This observation is termed the "Warburg effect" and it has been used to develop diagnostic tools for healthcare professionals such as using [18F]-fluorodeoxyglucose positron emission tomography (PET) to see where cancer is located. Cancer cells also use different versions of some glycolytic enzymes to speed up their glycolytic rates. These different versions can be targeted because they are cancer-specific and therefore drugs targeting those enzymes will NOT affect normal cells. One example of an enzyme that is different in cancer cells compared to normal cells is pyruvate kinase. In cancer cells, pyruvate kinase is the M2 isoform while in normal cells, pyruvate kinase is the M1 isoform. This M2 isoform of pyruvate kinase is strongly activated by fructose 1,6-bisphosphate while the M1 isoform is weakly affected by fructose 1,6-bisphosphate levels. This allows the cancer cell to more rapidly undergo glycolysis and feed pyruvate into lactate fermentation to start the cycle all over again. Cancer cells also metabolize glutamine more than normal cells specifically by glutaminolysis, which produces glutamate. Glutamate is incredibly useful because, as we previously discussed, glutamate can be used to make α-ketogluterate to help replenish Krebs cycle intermediates and it can be used to make a wide variety of amino acids such as proline and arginine. This is consistent with the only goal of cancer: to grow. The more amino acids it has, the more proteins it can make and the faster it can divide.

Here are the key takeaways of the fundamentals of cancer biology:

1. Cancer has only one goal: grow
2. Oncogenes are the accelerators while tumor suppressor genes are the breaks to cell growth
3. Cancer loves to eat glucose and make lactate even if oxygen is there (Warburg effect)
4. Cancer mutates to make replication easier and faster, bypasses the normal safeguards

Practice Questions:

1. Some scientists have proposed that cancer patients go on a keto diet. A keto diet has high fat but low carb content. Explain the scientific reasoning behind this proposal.
2. What cellular processes do you suspect would be elevated in cancer cells? What processes do you think would be less active? What hormone accomplishes this? Explain.
3. Given your answer to the question above, which glycolytic enzyme do you suspect could be the best one to target? Explain.
4. Why does [18F]-fluorodeoxyglucose positron emission tomography (PET) allow doctors to see where cancer is? What is a downside of this technique, in other words, what region of the body would this technique not be ideal for? Explain.
5. Many cancer patients who are not receiving treatment notice that they lose muscle mass. Explain why cancer can lead to losing muscle mass.
6. When intracellular lactate levels are too high, lacatate is excreted out of the cell and into the surrounding environment. This would make the blood around the tumor acidic or basic?

Answers:

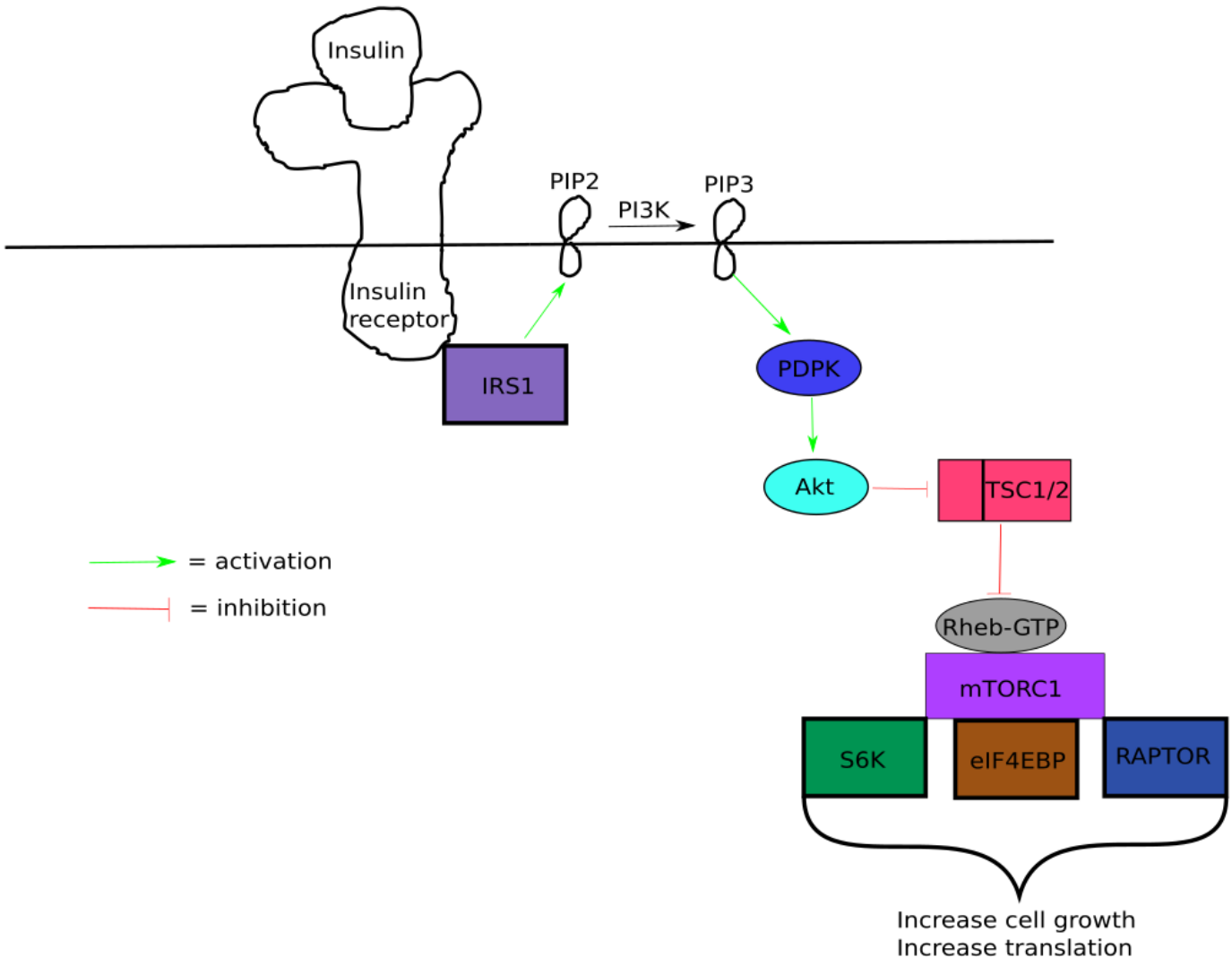
1. Cancer loves glucose and their mutations are aligned to maximize glycolytic output. They effectively put all their eggs in one basket, so cancer cells are not generally amenable to breaking down fats. The end result is that cancer cells become energy starved in the low carb environment and die off, while the healthy cells use the fat for energy like they normally would.
2. Glycolysis, PPP, *de novo* nucleotide synthesis, fatty acid synthesis, and amino acid synthesis should all be more active in cancer than in healthy tissue. Fatty acid breakdown, gluconeogenesis, Krebs cycle, and ETC should be slower in cancer than in healthy cells. Insulin is the major hormone that will facilitate this.
3. I would argue that hexokinase would be the most attractive target because hexokinase make glucose-6-phosphate that can be used for the PPP or the rest of glycolysis. Therefore, inhibiting it will prevent PPP and glycolysis from occurring. This means the cell lacks NADPH to make fats and cannot make ribose-5-phosphate to initiate nucleotide biosynthesis.
4. Radiolabeled fluorinated glucose can be used to see where cancer is located because cancer uses this glucose for energy more than other tissues (remember glycolysis is stimulated a lot in cancer cells). Therefore, the radiolabel will appear mostly in cells that

are actively undergoing glycolsysis the most, these tissues are generally cancerous. The issue with this is that the brain also does a lot of glycolysis and therefore the brain will light up like a Christmas tree irrespective of if there is cancer there.

5. Cancer likes to feed on glutamine, therefore, when all the glutamine runs out in the tumor cells, they can send out signals to the muscle to stimulate the glutamine shuttle. This constantly removes glutamine from muscles and therefore the muscle will have less protein and the person will lose muscle mass. This is referred to as cachexia and it is common for very aggressive cancers.
6. Lactate is acidic, therefore, if it is released from the cancer cells, the blood in the surrounding area will be more acidic.

Chapter 43: Insulin-IRS1 Pathway

Now that we have the fundamentals of cancer down, we can start to discuss the specific signaling pathways that give rise to cancer growth aka tumorigenesis. In all of these signaling pathways, we will specifically address how mTORC1 is stimulated. mTORC1 is the master regulator of cellular growth and its activity is controlled by its associated Rheb small G protein. Small G proteins are proteins that bind either GTP or GDP. When they are bound to GTP, they are activated and when they bind GDP, they are inactivated. Cellular signals and other cellular processes can alter the state that they are in. Here, Rheb activates mTORC1 and therefore when it is bound to GTP it will stimulate cell growth. mTORC1 has many downstream effector proteins that will increase translation and increase cellular growth, these are RAPTOR, S6K, and eIF4EBP. A general motif throughout all of these signaling pathways is that once the receptor binds its ligand, a phosphorylation cascade occurs where a bunch of proteins phosphorylate other proteins ultimately leading to some biological effect. An overview of this signaling pathway is shown below:



This pathway is not only true for insulin, but also insulin-like growth factors. It should make sense that insulin would increase cell growth because of what insulin signals to the body. Insulin says that there is an excess amount of glucose that is readily available to the body. In other words, the body has an excess of energy, with this excess energy, the body's cells can make more and more proteins and thus it is reasonable that they would grow and divide.

The first step in this pathway is insulin binding its receptor (duh). This specific receptor is a receptor tyrosine kinase (RTK). These types of receptors will phosphorylate themselves at tyrosine residues and this causes the recruitment of IRS1 aka insulin receptor substrate 1. IRS1 signals to the cell that there is a lot of glucose available in the blood and carries this message over to PI3K. PI3K is the PIP3 kinase enzyme that is responsible for phosphorylating the PIP2 phospholipid to make PIP3. PIP3 is an activator of PIP3 dependent protein kinase (PDPK), this kinase will activate Akt also known as protein kinase B. Akt is an incredibly important protein kinase in the cell, and its job is to promote cell growth and energy production. To do this, Akt will inhibit TSC1/2, a protein complex whose job it is to turn off mTORC1. With TSC1/2 deactivated by Akt, mTORC1 remains on and can stimulate cell growth and protein synthesis using its effector proteins RAPTOR, S6K, and eIF4EBP.

Practice Questions:

1. Identify all oncogenes and tumor suppressor genes in the Insulin-IRS1 pathway.
2. What would happen if the insulin receptor was mutated such that it is perpetually phosphorylated? Explain the effects this would have on the premise of the signaling pathway.
3. PTEN is a protein phosphatase that is responsible for converting PIP3 back into PIP2. What would this protein be considered, an oncogene or tumor suppressor gene? Explain.
4. PRAS40 is a protein that will bind to an inhibit RAPTOR. Would this protein be considered an oncogene or tumor suppressor gene? Explain.
5. Suppose you are working at a biotechnology firm and you are tasked with determining tumor suppressor genes and oncogenes in the insulin-IRS1 pathway to identify drug targets for a new protein drug you are helping manufacture. Explain how you could identify tumor suppressor genes and oncogenes using CRISPR/Cas9 technology with guide RNAs targeting IRS1, PI3K, Akt, and TSC1/2. In other words, what would you observe upon knocking those genes out in a tumor sample?

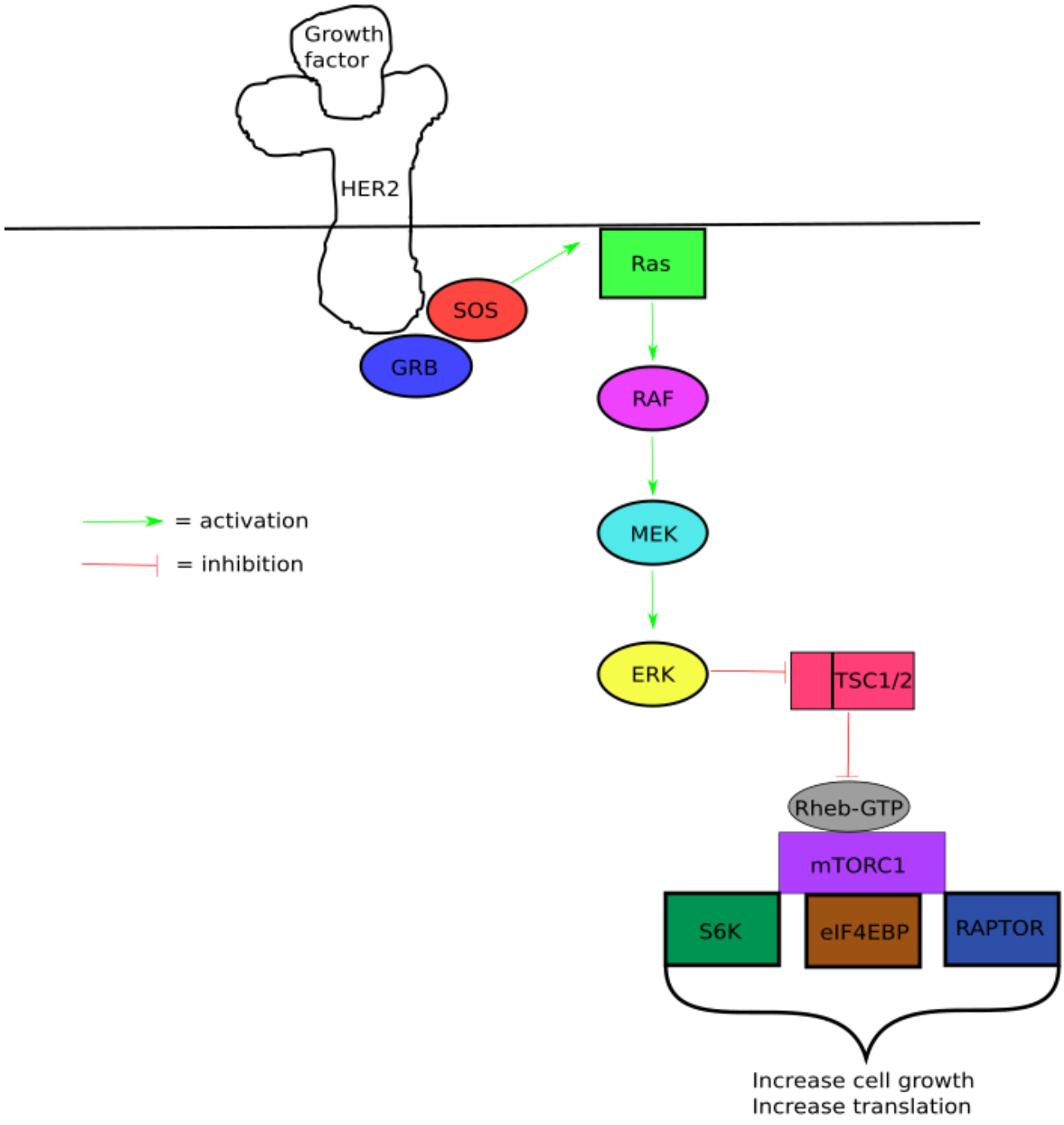
Answers:

1. Oncogenes = Insulin receptor, IRS1, PI3K, PDPK, Akt, Rheb-GTP, mTORC1, RAPTOR, S6K, and eIF4EBP. Tumor suppressor genes = TSC1/2.
2. If the insulin receptor was permanently phosphorylated then IRS1 would always be bound to it, meaning PI3K would always be active, leading to more Akt activity and ultimately more cell growth via mTORC1. This would likely be an oncogenic mutation that would promote tumorigenesis.

3. PTEN would decrease PIP3 levels, thereby decreasing PDPK activity and decreasing mTORC1 activity in turn. Overall PTEN would have to be a tumor suppressor gene because it decreases cell growth when active.
4. PRAS40 inhibits RAPTOR, which is a protein responsible for cell growth, therefore PRAS40 would have to be a tumor suppressor gene.
5. If you designed gRNAs that target IRS1, PI3K, Akt, and TSC1/2 and used a lentivirus to transfect tumor cells, you would see that the tumor cells that had IRS1, PI3K, and Akt knocked out would have less growth while the tumor cells that had the TSC1/2 gene knocked out would have more growth. This is because without the oncogenes to stimulate mTORC1, there will be less overall growth while if TSC1/2 is absent and cannot deactivate mTORC1, it will continue to stimulate cell growth leading to tumor growth.

Chapter 44: HER-2 Pathway

HER2 is a growth factor receptor that is involved with stimulating cell growth and division. It also activates mTORC1 as its downstream target and therefore functions similarly to the insulin-IRS1 pathway. HER2 is grossly overexpressed in 25% of breast cancer patients and therefore it is important to discuss the known signaling cascade associated with it. An overview of the signaling pathway is shown below:



The very first thing that happens in this pathway is that the growth factor binds the HER2 receptor and this causes the recruitment of two accessory proteins, GRB and SOS. These two proteins help to stimulate Ras, a small G protein that is anchored to the cell membrane. Ras phosphorylates RAF and that leads RAF to phosphorylate MEK and so on to ERK. ERK is very similar to Akt in that its job is to promote cell division and growth. To do this, ERK will inhibit TSC1/2 in the same kind of way that Akt did and that will prevent TSC1/2 from deactivating mTORC1 overall leading to rapid cell growth.

Here are the key takeaways from the HER2 signaling pathway:

1. HER2 promotes cell growth by activating mTORC1
2. ERK is the terminal protein that inhibits TSC1/2 and activating mTORC1
3. Ras is a small G protein that is anchored at the cell membrane

Practice Questions:

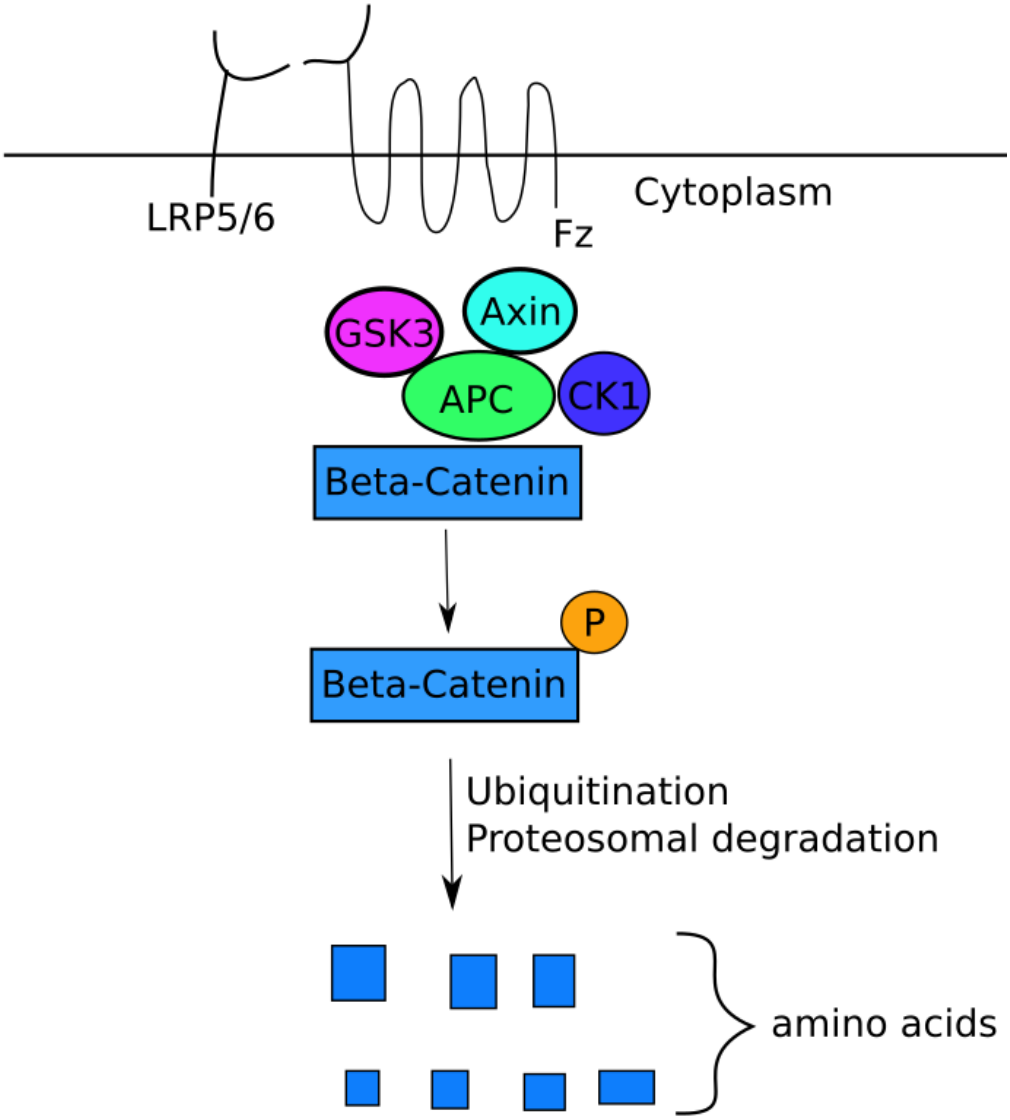
1. For each protein in the HER2 pathway, find an analogous protein in the insulin-IRS1 pathway (yes you can group proteins together).
2. Herceptin was the first drug that was largely successful in treating HER2 positive breast cancer. This drug works by binding the HER2 receptor but not activating its phosphorylation activity. Explain how this works to decrease cancer growth.

Answers:

1. GRB and SOS are similar to IRS1; Ras, REF, and MEK are similar to PI3K and PDPK; and ERK is similar to Akt
2. If Herceptin binds the HER2 receptor but does not trigger its autophosphorylation then GRB and SOS will not bind it and therefore Ras will not be active and ultimately mTORC1 will not be active either.

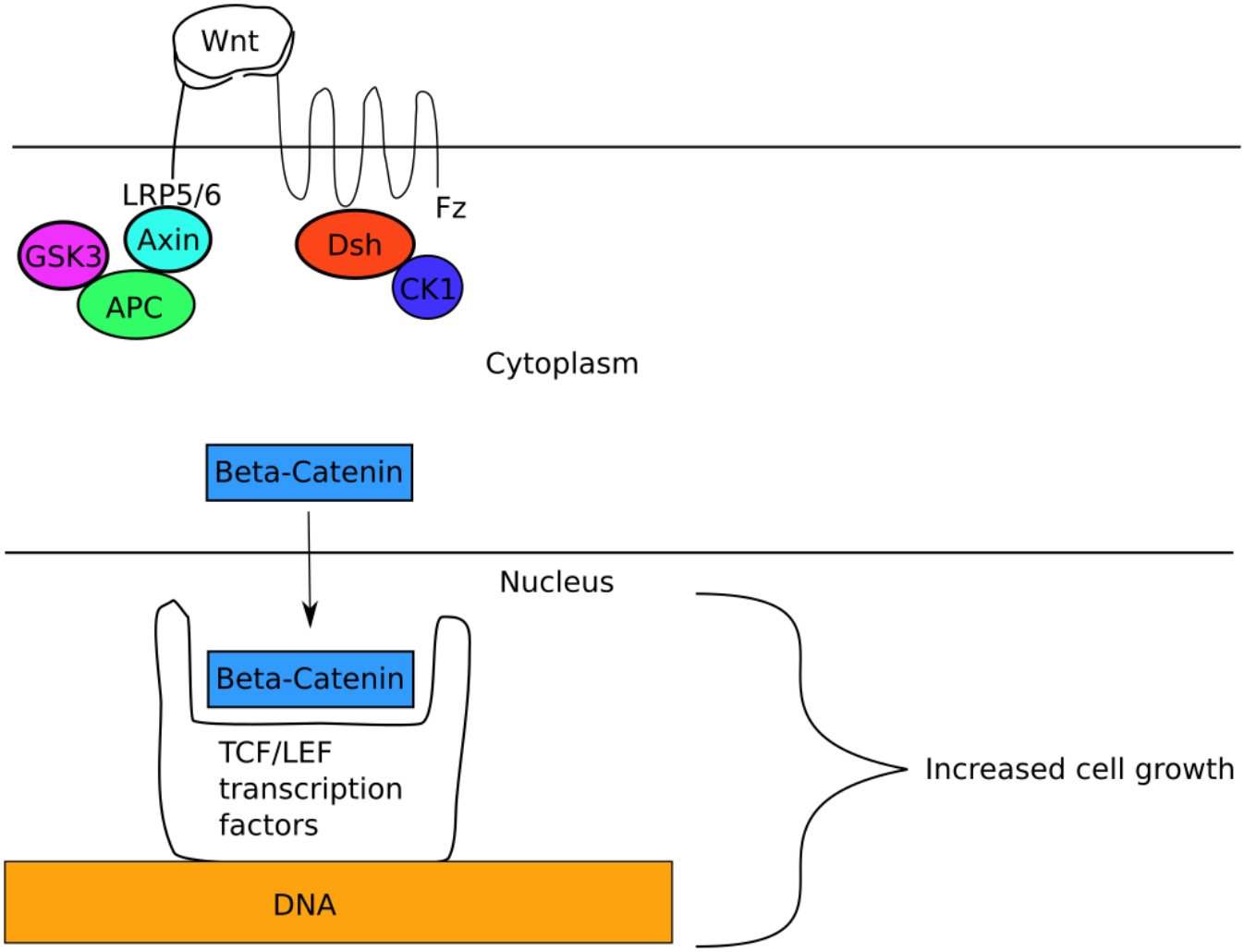
Chapter 45: Wnt Signaling Pathway

The Wnt signaling pathway is critical for regulating embryonic development and stem cell-likeness. Aberrant Wnt signaling has been associated with tumorigenesis specifically in colon cancer and we will describe how Wnt activates proteins and genes that are critical for promoting cancer. Unlike the other two cancer-inducing pathways, this pathway does not target mTORC1. Instead, the pathway targets β-catenin, which will move from the cytoplasm to the nucleus of the cell and activate transcription of genes that are important for cell growth and differentiation. Another difference between this pathway and the previous two is that this pathway does not use an RTK receptor. Before we go through what happens in the active pathway, we first have to discuss what ordinarily happens in the cell in the absence of a Wnt signal. Without Wnt binding its receptor on the cell membrane, β-catenin is phosphorylated and destroyed by the destruction complex shown below:



In the absence of Wnt signals, the APC forms a destruction complex that will phosphorylate β-catenin and this will cause its ubiquitination and ultimate destruction. Without any β-catenin to go to the nucleus and cause transcription of growth genes, cell growth is partially halted.

An overview of the active Wnt signaling pathway is shown below:



When a Wnt signal binds its receptors LRP5/6 and frizzled (Fz), it causes disheveled (Dsh) to bind. Dsh is important because it disassembles the APC destruction complex by recruiting CK1. Without CK1, the disassembled APC complex will bind to LRP5/6 using Axin, this effectively keeps it from destroying β-catenin. Now that β-catenin is free to move around the cell, it will move to the nucleus and promote transcription of growth genes by binding to TCF/LEF transcription factors. The net effect of this is increased cell growth and differentiation.

Here are the key takeaways from the Wnt/ β-catenin pathway:

1. β-catenin is destroyed by phosphorylation and ubiquitination in the absence of Wnt signals.
2. Wnt ligand binds frizzled and LRP5/6 causing the disassembly of the APC destruction complex, allowing β-catenin to move freely.
3. Free β-catenin will move to the nucleus where it will promote the transcription of growth genes.

Practice Questions:

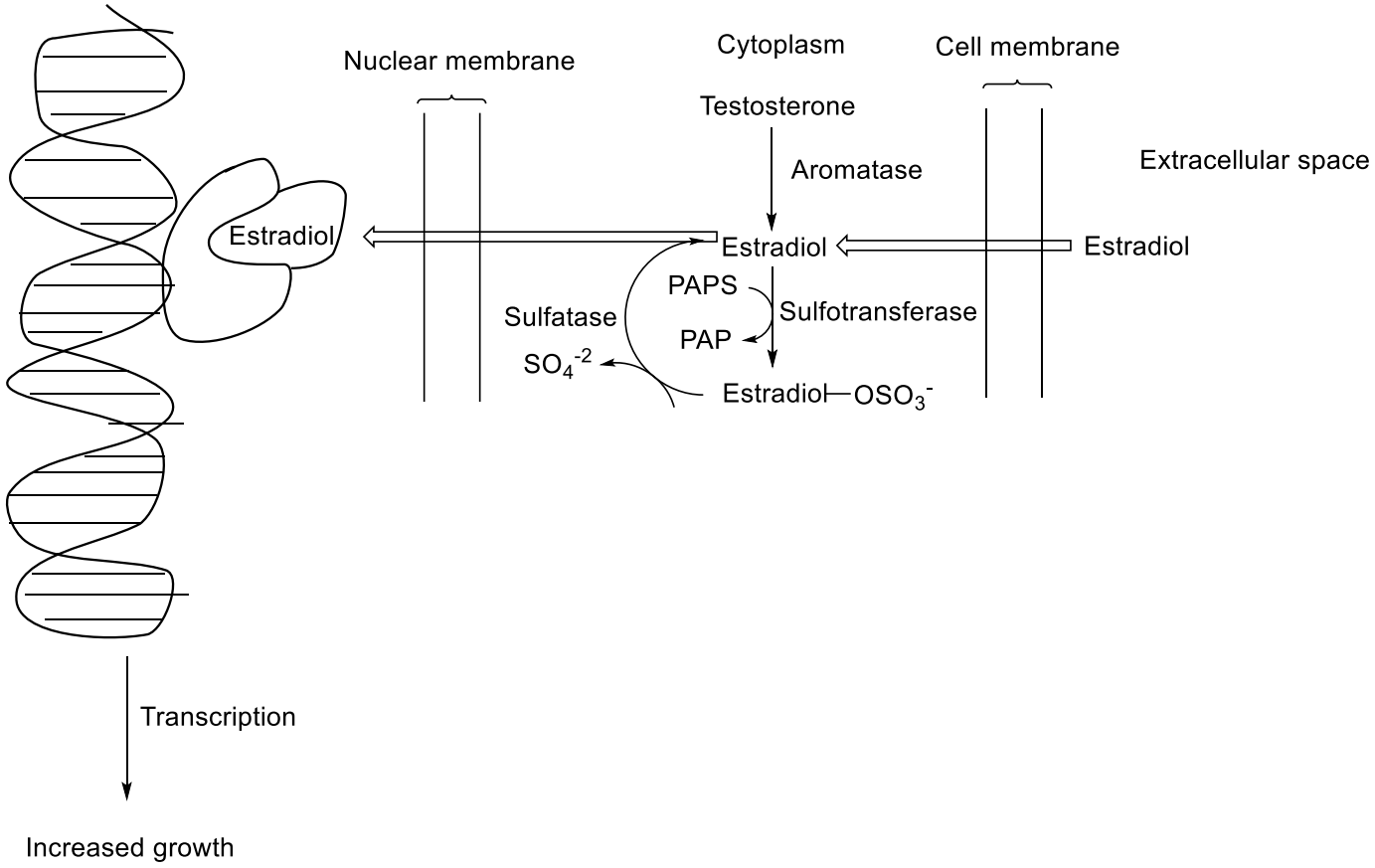
1. Identify all the tumor suppressor genes and oncogenes in the Wnt/ β-catenin pathway.

Answers:

1. Tumor suppressor genes: APC destruction complex genes, oncogenes: Fz, LRP5/6, and β-catenin

Chapter 46: How Chemotherapeutics Work (Case Example: Tamoxifen)

Broadly, anticancer drugs work by blocking a cancer-inducing signal or by killing cells that go through a metabolic process more than others. For example, 5-fluorouracil works by disrupting thymidine synthesis and therefore cells that are constantly dividing will be more affected by the drug. Since cancer cells divide more rapidly than normal cells, these cells are disproportionately killed. Another example is tamoxifen, which targets the estradiol (E_2) receptor. Estradiol is a critical steroid hormone that will bind to the E_2 receptor in the nucleus and activate transcription of many genes involved with cell growth. This is very common in breast cancer and therefore it was an attractive therapeutic target to treat breast cancer patients. An overview of the estradiol pathway is shown below:

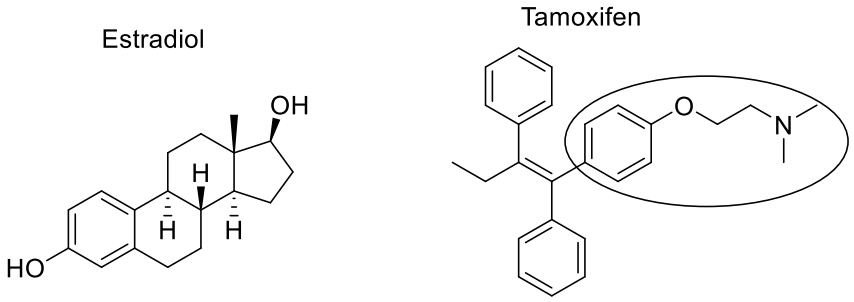


Estradiol can be introduced to the cell in two different ways. First, aromatase can be used to make estradiol from testosterone and second, estradiol can be received from outside the cell. Recall that testosterone can be made from cholesterol through a series of reactions because it is a steroid hormone and cholesterol is a precursor to steroid hormone synthesis. Once testosterone is made in the cell, an enzyme called aromatase can be used to make estradiol. Alternatively, estradiol can be received from outside the cell, in which case it will cross the membrane directly because it is nonpolar and enter the cytoplasm.

In either case, the estradiol can be moved into the nucleus, where it will bind its receptor and cause recruitment of a coactivator that will stimulate the transcription of growth genes. The cell

can control the effects estradiol has on the cell through sulfation. Much like how proteins can be activated or deactivated through phosphorylation, small molecules such as estradiol can be controlled through sulfation. The enzyme class that catalyzes the transfer of sulfate groups are called sulfotransferases and these enzymes will add a sulfate group to one of the OH groups on estradiol. This sulfate group prevents the estradiol from binding its receptor in the nucleus because now it is too bulky and because it is negatively charged. All sulfotransferases use a cofactor called PAPS, which is the sulfate donor molecule in the enzyme. The sulfated estradiol can be made back into the active form by a sulfatase, which will remove the sulfate group and allow the estradiol to move into the nucleus and do its job.

Tamoxifen is a E_2 receptor antagonist, which will bind the receptor, but prevent the coactivator from binding and therefore prevents transcription of the growth genes that are controlled by estradiol. Like antagonists in TV shows, tamoxifen screws everything up, it prevents transcription of the growth genes and therefore will prevent cancer growth. The structure of tamoxifen and estradiol are shown below with the pharmacophore of tamoxifen (active portion of the drug) circled:



The bulky nature of the benzene rings and alkene region of tamoxifen allows it to block the coactivator while the circled portion allows tamoxifen to bind the E_2 receptor because of that portion's more polar nature and ability to undergo strong dipole-dipole interactions. Tamoxifen has been widely used as a first-line treatment for breast cancer because of this unique ability.

Practice Questions:

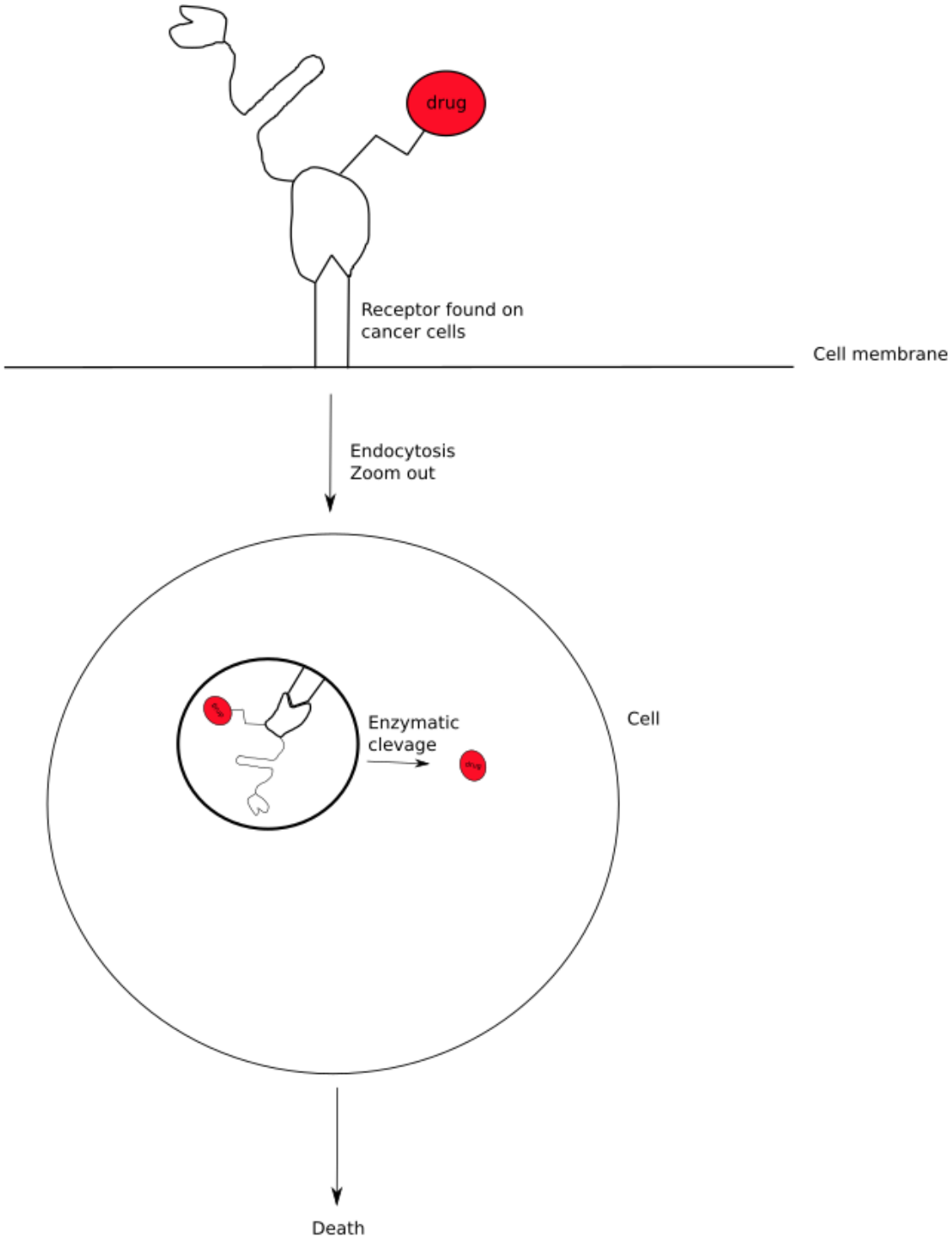
1. Explain how the sulfation and phosphorylation systems are similar but different. Make parallels between the different enzymes involved where appropriate.
2. Name at least two enzymes in the estradiol pathway that can be inhibited that would decrease estradiol binding its receptor in the nucleus. Explain why inhibiting those enzymes would possible decrease cancer growth.
3. Tamoxifen is a receptor antagonist in the breast, but a receptor agonist in the uterus. Receptor agonists are molecules that will bind and activate receptors while antagonists will bind but not activate the receptor. Explain this discrepancy of tamoxifen binding in different tissues and hypothesize a possible repercussion of this aberrant binding.

Answers:

1. Sulfation and phosphorylation are similar because they act as control mechanisms to mediate expression of either small molecules or proteins where appropriate (MAS**C**). They both work by introducing a bulky and negatively charged group to the molecular structure, which will tune its polarity, electrostatics, and intermolecular forces. Kinases are analogous to sulfotransferases, however, sulfotransferases need PAPS while kinases need ATP. Phosphatases are analogous to sulfatases.
2. Aromatase and sulfatase can be inhibited to prevent estradiol form binding its receptor in the nucleus. If aromatase is inhibited then testosterone that the cell makes cannot be converted to estradiol, therefore intracellular estradiol levels will decrease. If sulfatase was inhibited, then the sulfated estradiol cannot be made back into normal estradiol. Recall that sulfated estradiol cannot enter the nucleus and activate its receptor and therefore this would also decrease transcription of growth genes and thus decrease cancer growth.
3. Tamoxifen can be an agonist in the uterus and an antagonist in the breast because the structure of the estradiol receptor is different in the uterus than in the breast. This structural difference doesn't allow the tamoxifen to block the coactivator in the uterus and therefore when it binds, it acts the same as estradiol would. Because tamoxifen is a receptor agonist in the uterus, it will likely cause cancer in that area because it will over stimulate cell growth.

Chapter 47: Future Directions for Cancer Research

One of the biggest issues in cancer treatment is the nonselective nature of the anticancer drugs. One way to combat this is using antibody drug conjugates (ADCs). ADCs combine the molecular targeting power of the antibody with the destructive force of the drug that is attached to it. In an ADC, there are three parts: the antibody, the drug molecule, and the linker that connects the two. As we previously discussed, antibodies are effectively targeted molecular missiles, therefore we can control where the drug is delivered based off the antibody to which it is bound. Because of this very targeted nature, ADCs could allow the use of more powerful anticancer drugs because they will target just the cancer cells and leave the other cells alone. An important aspect of ADC design is the design of the linker region because the linker region is what gets cleaved to release the drug compound. If the linker region is weak and releases the drug prematurely, it could cause massive damage to the body. If the linker region is too strong, it won't release the drug at all and then the ADC will have no effect. Currently, there are many research groups and pharmaceutical companies that are focusing their efforts on ADCs and specifically the design of the linker region so that future ADC drugs can be used with high selectivity for cancer cells over normal cells and with high potency. At the time of this writing, there are 9 FDA approved ADCs, all of which are used to treat different kinds of cancers such as HER2 positive breast cancer (Kadcyla) and acute myelogenous leukemia (Mylotarg). A general theme throughout all of the ADCs that have been approved is that they target a receptor on the cancer cell that is substantially overexpressed or unique to the cancer cell itself, thus sparing healthy cells and destroying cancer cells. An overview of how ADCs work is shown below:



Practice Questions:

1. What are some potential downsides to using ADCs?
2. What are some potential challenges that you may come into when designing and making an ADC?

Answers:

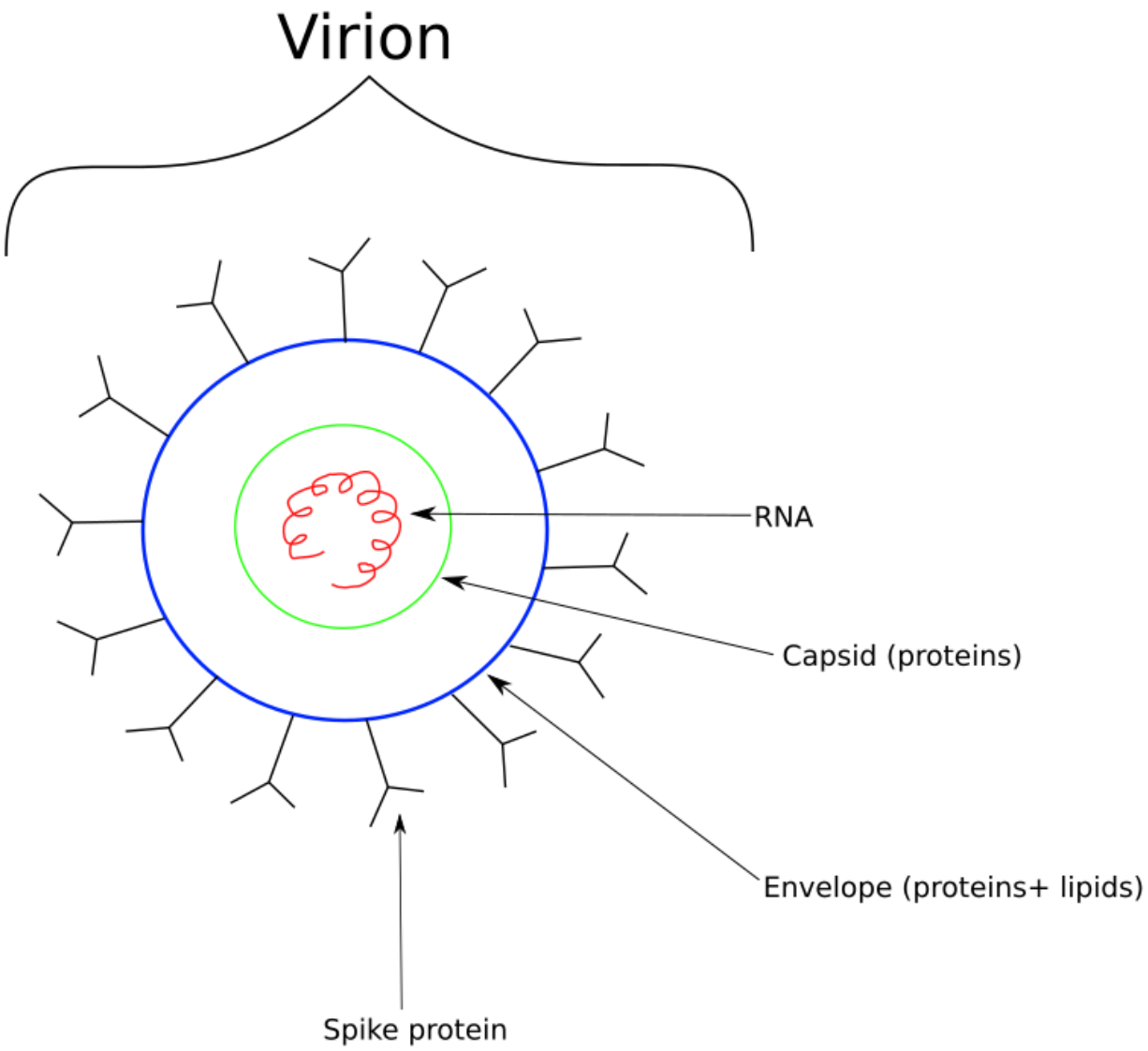
1. ADCs are likely super expensive because antibodies are expensive. ADCs may have an unstable linker that can release the drug prematurely, causing cytoxicity to normal cells. ADCs may interact with receptors at nontarget sites if the receptor is not specific enough to cancer cells.
2. Designing the ADC so that the linker is a specific spot is a challenging task. Remember that protein folding is important for antibody structure, so the linker needs to be added on to the antibody after it is expressed by the cell. If the linker is not in a specific spot, the ADC may be less effective because the linker bonds may not be as accessible to enzymes for cleavage or may not allow the drug molecule to interact with its molecular target.

Focus 10: SARS-CoV-2

Chapter 48: Virology Basics

At the time of this writing, the SARS-CoV-2 virus has claimed the lives of approximately 1.8 million people and has infected 82 million worldwide according to Johns Hopkins University. Because of the rising concerns and drastic public health measures taken to curb the spread of the virus in both the United States and globally, it is imperative that we discuss some basic virology so that we can further understand these pathogenic agents. We will start out by discussing viruses in general and, where appropriate, discuss the specifics of the SARS-CoV-2 virus. For the sake of brevity, I will refer to the SARS-CoV-2 virus as Covid19 for the remainder of this focus since that is the colloquial name for the virus and because it is easier to type.

Before we delve into the molecular mechanisms of infection, we have to start with a definition of what a virus is. A virus, in its most basic form, is a pathogen that is **required** to infect a host cell to survive and reproduce. Viruses *are NOT living* because they cannot perform their own metabolism and they cannot reproduce without first infecting a host. This makes them different from bacteria, which can replicate by themselves and survive without infecting people or other cells. For viruses to replicate, they need some form of information that they can pass down from one generation to the next. This information is referred to as the **viral genome** and this genome can be made of either DNA or RNA. Covid19 is an RNA-based virus while the herpesvirus is a DNA-based virus. Viruses are not just free-floating pieces of DNA or RNA, however, they encapsulate this genome in what is known as a **capsid** and sometimes an **envelope**. On the surface of a virus particle, also known as a virion, there are some molecular anchors (**spike proteins and sheath/tail fibers**) that allow the virus to attach to host cells so it can infect them. The Covid19 virus has spike proteins while viruses that infect bacteria use sheath and tail fibers. The structure of the coronavirus is shown below to give you some insight into what each of these pieces mean:



As indicated by the diagram above, the capsid is made of proteins while the envelope is made of proteins and lipids. The envelope in particular is made of a lipid bilayer, the same type that is seen in most cell membranes. This makes many enveloped viruses susceptible to alcohol; the alcohol would mess with the intermolecular forces that drive bilayer formation and therefore the virion would be destroyed and thus the virus would be "dead". Which brings us to the next definition we need to discuss. A virus is said to be "dead" if it cannot infect living cells anymore. This often means that the virion is either physically or chemically attacked by something that punctures the envelope or other critical part of the virion that prevents it from infecting more cells.

Here are the key takeaways from virus structure:

1. Active virus particles are referred to as virions
2. Each virion is comprised of a genome (DNA or RNA), a capsid, and an optional envelope

3. Capsids protect the DNA and are composed of proteins, envelopes protect the whole virion and have spike proteins or sheath/tail fibers attached that aid in cell entry.
4. Envelopes are made of phospholipids and can be destroyed by alcohol.
5. Viruses are considered “dead” when they cannot infect more cells.

Now that we understand virus structure, we can begin to talk about the “life cycle” of a virus. In other words, how viruses infect our cells and how they intelligently make more and more copies of themselves. For this discussion, think of a virus as a machine. It is a machine with only one purpose: replicate. The general strategy that viruses use here is to make the cell it infects create more viruses. This is because each virion only comes with instructions to make more viruses and nothing else; the virus NEEDS to use the host cell’s own machinery to replicate. This makes treating viral infections incredibly difficult because any drug we could use to halt viral growth would also be damaging our own cell’s enzymes. Thus, antivirals typically have VERY nasty side effects. Here is the general “life cycle” of a virus explained in 6 bullet points:

1. Virus attaches to the cell using a receptor
2. Virus enters the cell via endocytosis or fuses with the membrane
3. Virus’s genome gets released into the cytoplasm
4. Its genome gets replicated and the proteins its genome codes for get produced using the cell’s own machinery
5. The parts that make each new virus get glued together and activated using viral proteases
6. Host cell “explodes” or lyses because of all the new viruses it made

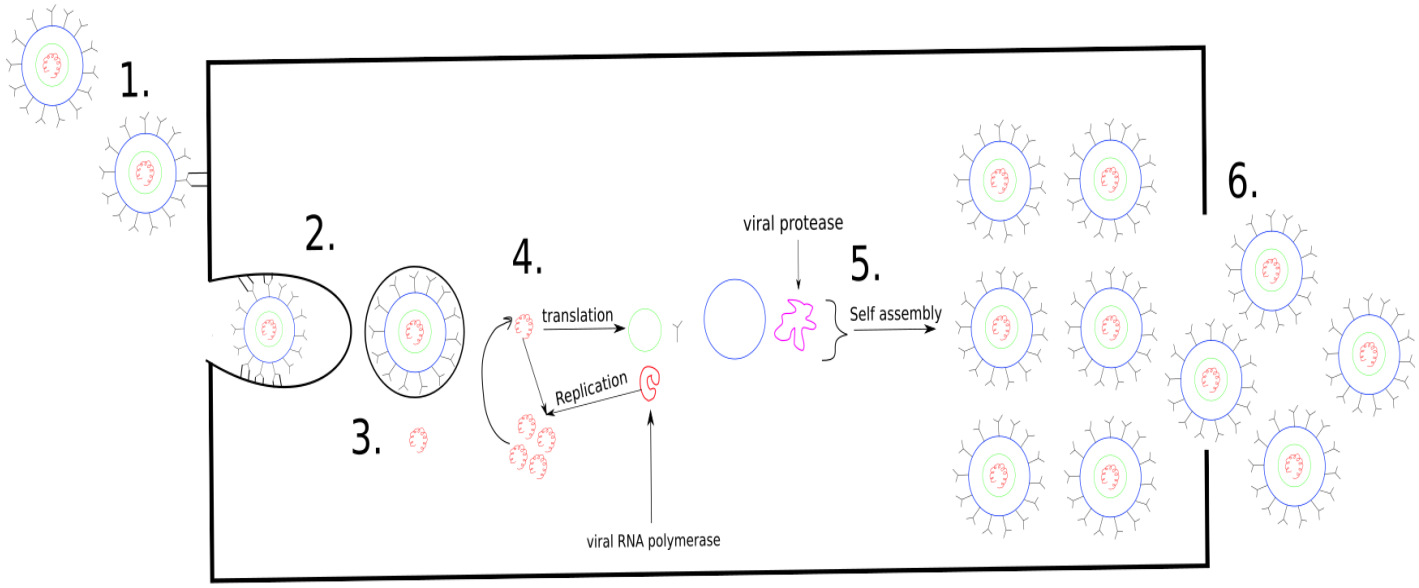
For DNA and RNA viruses, these steps are the exact same with one exception, DNA viruses need to have their DNA transcribed into mRNA so it can be translated into viral proteins. Remember the central dogma of cell biology: DNA to RNA to proteins. Typically, the viral genome is released into the cytoplasm because of the acidic pH of the endosomal vesicle. For *DNA viruses, this release occurs in the nucleus* so that **transcription** to mRNA can occur and for *RNA viruses, this release occurs in the cytoplasm* so that **translation** to proteins can occur directly after. There are a couple of incredibly important viral proteins that many viral genomes encode:

1. Viral RNA polymerase (copying machine)
2. Viral proteases (glue)
3. Spike proteins (anchor)
4. Capsid proteins (structure)

The viral RNA polymerase will copy the viral RNA so that the host cell will be worked into overdrive, making more and more viruses at an exponential rate. These proteins are generally the first to get released because they are integral for rapidly making more virions. The viral proteases are proteins that will cut and activate the structural proteins that make up the virion particle, think of these as the building proteins that glue the pieces of the virion together. They’re like Bob the Builder ® gone horribly wrong. The Covid19 virus uses the 3ClPro protease to activate its viral proteins. The spike proteins or the sheath/tail fibers are the anchors that allow the virion

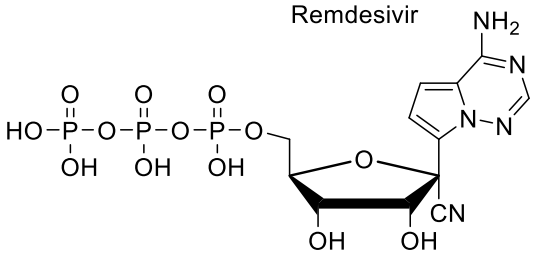
to recognize host cells. The Covid19 virus uses its spike protein to recognize the ACE2 receptor, for example. The capsid proteins are what protect the viral genome, therefore, these are an absolute must so that the virus can continue to infect host cells. The other proteins that are made by the viral genome are other structural proteins and these are generally made last in the cycle, just at the end of step 4.

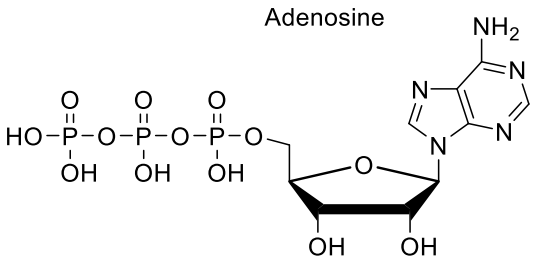
A general overview of the viral "life cycle" is shown below for Covid19:



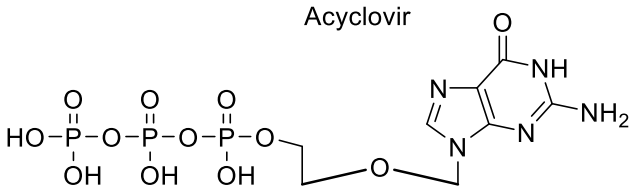
Practice Questions:

1. Explain from start to finish how Covid19 infects your cells.
2. Some drugs, such as Remdesivir, are designed to treat Covid19 target the virus's RNA polymerase. Explain how this drug interferes with the RNA polymerase on the premise of its structure shown below:





   Explain how interfering with the RNA polymerase helps treat Covid19 infection.
3. Viruses are generally less sophisticated than eukaryotic cells, therefore, they can be "tricked" into accepting molecules that cannot be used to extend the nucleotide chain. One of these molecules, acyclovir is shown below. Explain how this molecule prevents nucleotide chain extension on the premise of its structure.



4. Some vaccines are made using “inactivated viruses”, these viruses are virions that lack the genes that are necessary to replicate, but still have their spike protein or sheath/tail fibers intact. Explain how these “inactivated viruses” are made using DNA technology.

Answers:

1. Covid19 virions bind to the ACE2 receptor on the surface of your cells. This causes the cell to endocytose the virion, where the pH change in the endosome causes the virus’s RNA to shed into the cytoplasm. There, the viral RNA gets translated directly into proteins such as the virus’s RNA polymerase, capsid, proteases, spike protein, and structural proteins. The virus’s RNA polymerase is used to replicate the RNA of the virus over and over again and these new RNA molecules also get translated. This repeats itself until there are many virion parts in the cytoplasm. Then the viral proteases go to work assembling the virions, which eventually will cause the cell to lyse, releasing tons of new virus particles, which can go on to infect more host cells.
2. Remdesivir is an adenosine nucleotide analogue and therefore it can fool the viral RNA polymerase into accepting it to build the virus’s RNA. The problem with Remdesivir is that its structure is sufficiently different from the A base that it can halt the RNA polymerase shortly after its incorporation. Without a functional viral RNA polymerase, the virus cannot make more of its RNA and its RNA will be quickly destroyed by enzymes in the cytoplasm. No more viral RNA = no more instructions to make new viruses = no more new viruses created = infection over.
3. Acyclovir lacks the 3’OH group that is necessary for nucleotide extension during polymerization and therefore when it gets put into the growing nucleotide chain, it cannot be used to add any additional nucleotides and the RNA molecule is truncated at that position. This partially synthesized RNA will not code for all of the proteins necessary for viral replication and therefore the viral infection will slowly go away.
4. These “inactivated viruses” are made by infecting cells with artificially synthesized mRNA that codes for all of the virus EXCEPT some of its replication genes. Then the rest of the process occurs as normal during the viral life cycle and many viruses are made that lack the replication functionality and therefore are “inactivated”. This would illicit the same kind of immune response as a normal virus, but would lack the replication ability and therefore could not cause the viral disease. Remember, the replication happens using the cell’s own enzymes so all that matters to the body is what the virus looks like from a surface level, not what is inside it.

Chapter 49: Immunology Basics

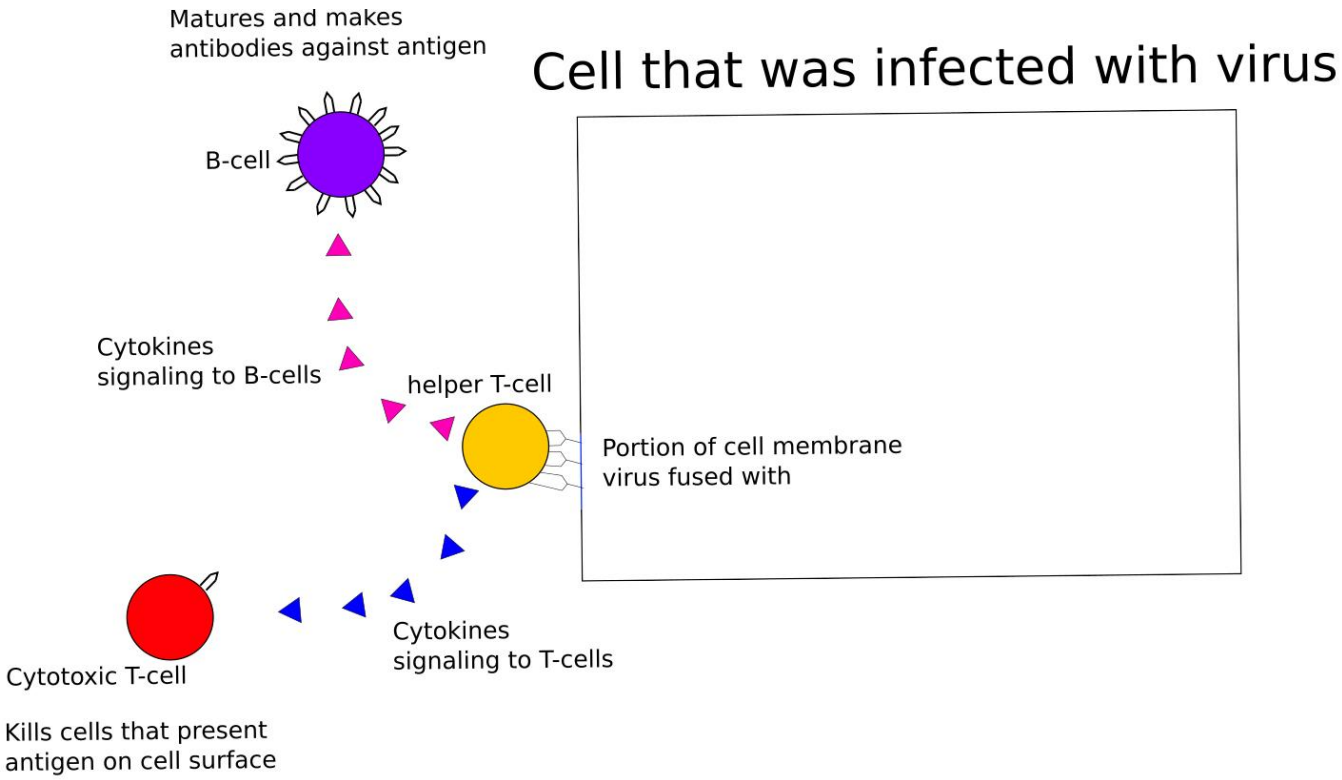
Now that we discussed how the virus spreads throughout the body, we can talk about how your body takes the fight to the virus. When you get a viral infection, it becomes an all-out war between your body's defense systems and the virus itself. Initially, the virus is at the upper hand because it recognizes your body, but your body does not recognize the virus. Broadly, there are two different ways the body can launch an attack on the virus or prevent infection in the first place: general responses and virus-specific responses. These are also referred to as "innate immune responses" and "adaptive immune responses" respectively.

General responses are the attacks the body throws at the virus that are not specific to that pathogen, these are effectively the first line of defense whenever you are exposed to a pathogen of any sort. These attacks and defense mechanisms are called the "innate immune responses" because they are innate in any person regardless of prior infection history. In other words, they come as a package deal with you being a healthy human. If you are exposed to any pathogen, they first need to get through the skin or find some other opening into your body. This is because the cells in your skin are impermeable to viral infection. If the virus gets passed the skin and it enters through the nose or some other mucous membrane, the body will combat this by producing a lot of "snot" and mucous. This mucous is used to lubricate where the virus is so that it can be transported to the throat and towards the stomach, where the virus can be killed using the very low pH of the stomach and antiviral compounds produced by the "good" bacteria in our gut. If the virus still survives and infects one of your cells, the cell will produce chemical signals that will attract natural killer cells that will cause it to destroy the infected cell. These natural killer cells are effectively homicidal maniacs that will destroy the body's own cells if they produce a signal saying that they are infected. The body can also produce interferons, specifically IFN-alpha and beta. These are used to trigger apoptosis of infected cells and can deactivate protein synthesis.

Virus-specific responses are attacks that, as the name suggests, are specific to the virus in question. These generally come in three forms: T-cells, B-cells, and antibodies. These responses are virus-specific because they recognize a particular molecule on the virus, known as an antigen. T-cells come in two different flavors, helper T-cells and cytotoxic T-cells. Both forms of T-cells will recognize the pathogen of interest by binding its antigen, but only the cytotoxic variant will destroy it. The helper T-cell is simply there to say "hey guys, I found something." Once the helper T-cell sends the message in the form of a cytokine, B-cells and cytotoxic T-cells will be brought to the area and are then used to destroy the pathogen. Once the body destroys the pathogen, or rather, the infected cell that presented the antigen, B-cells will produce antibodies that will recognize that specific antigen and bind to them recruiting more immune cells called macrophages to attack at that location. Over time, B-cells will mature to plasma cells, or memory B-cells, that will constantly produce highly specific antibodies that will bind to the antigen the next time it comes back into the body. After the infection is over, memory T-cells are present immediately after. These memory T-cells are simply the mature T-cells that survived after the majority of the other T-cells died off. Only about 10% of T-cells will survive to be memory T-cells that will persist after the infection is over. These memory T-cells will either be in the blood

or the lymph nodes. Those that circulate in the blood are effector T-cells while those that are in the lymph nodes are central T-cells, which can be activated to effector T-cells if the virus shows up again.

An overview of this process is shown below:



In short, the helper T-cell will recognize cells that have been infected with the virus and release signals to cytotoxic T-cells and B-cells. Cytotoxic T-cells will directly attack cells that have antigens on their surface while B-cells will act as antibody production factories and will pump out antibodies that will recognize the specific antigen from the virus. These antibodies will bind to the antigen on new virions and mark them for destruction by macrophages.

Practice Questions:

1. One of the most interesting thing about viruses is that they can mutate. These mutations can occur anywhere on the genome, however, not all mutations are created equal. What mutation would likely result in the largest effect on the adaptive immune response? Explain.
2. Autoimmune disease is when the body attacks its own tissue. Explain how this can occur using the principles of the adaptive immune response. In other words, explain step-by-step how autoimmune diseases manifest themselves.

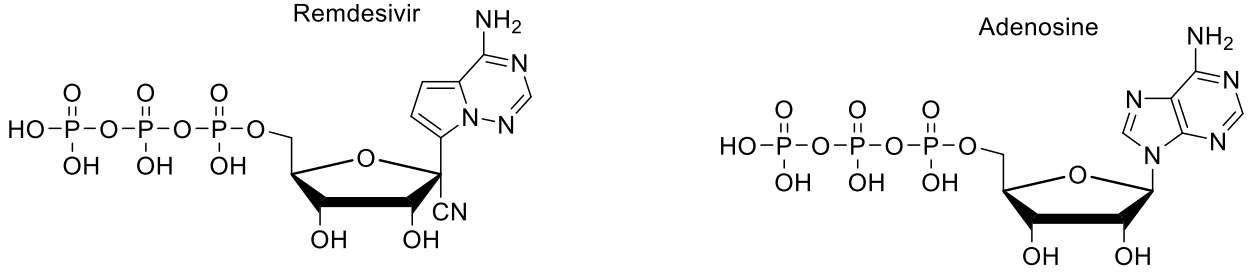
Answers:

1. Mutations in the spike protein or sheath/tail fibers would be the most likely to disrupt the adaptive immune response because the whole premise of this system is that the antigen (usually the spike protein or sheath/tail fibers) is recognized by antibodies and the helper T-cell/ cytotoxic T-cell. If the spike protein or sheath/tail fibers is altered, then the helper T-cell cannot recognize the virus anymore and therefore the adaptive immune response would be compromised.
2. Autoimmune diseases would occur if the antigen that the helper T-cell, cytotoxic T-cell, and B-cell's antibodies bind to is a protein that is displayed by normal cells. This would cause the exact same immune response as a viral antigen, but instead of killing infected cells, the body would kill healthy ones.

Chapter 50: RNA-dependent RNA Polymerase inhibitors

The Covid19 copying machine is the RNA polymerase that it uses to copy the viral genome that codes for more viruses. Therefore, jamming this copying machine could be key to stopping the production of more viruses and thus treating the disease.

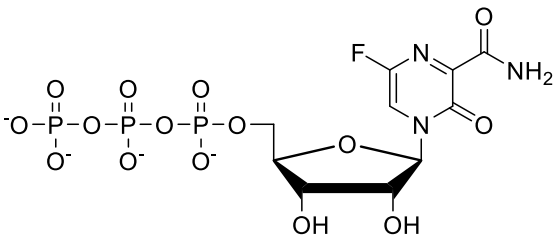
Some previously established drugs have been looked at to treat Covid19 that work by inhibiting its RNA polymerase. One of them you may have heard a lot about because it was the first drug to be FDA approved to treat the coronavirus, Remdesivir. For those of you who don't know, Remdesivir is not a new drug, in fact it was used as a general broad spectrum antiviral compound for viruses such as Ebola, SARS, and MERS. In all of these viruses, Remdesivir works the same way, it is an adenosine mimic that will prematurely stop the RNA polymerase from copying the viral genome. This produces a truncated RNA molecule that can still be translated into proteins, but these proteins are missing parts of their structure and as we know, structure informs function (MA**S**C). Thus by intercepting the viral life cycle during the replication phase, Remdesivir prevents the production of more virus particles and can therefore treat the disease. The Remdesivir molecule acts as a competitive inhibitor of the viral RNA polymerase and competes with adenosine for binding the RNA polymerase. The structure of Remdesivir and adenosine is shown below:



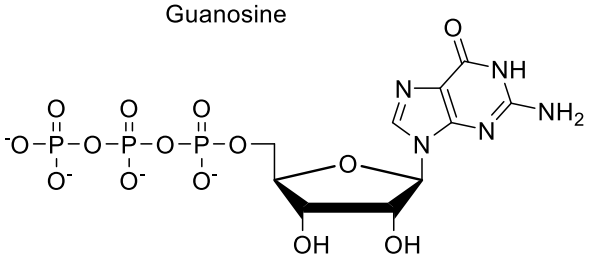
As you can see, the two molecules are incredibly similar to one another, they both are phosphorylated, both have a ribose ring, and both have a very similar base structure. The only real difference is the CN group on carbon 1 of Remdesivir that is notably absent in Adenosine. Remdesivir has proven to be an effective treatment for moderate Covid19 disease, the final phase 3 clinical trial analysis of over 1000 patients revealed that Remdesivir decreases median recovery time by 5 days, decreases short-term mortality by 5.2% and long-term mortality by 3.8%.[14]

Other drugs have also been looked at to treat Covid19, such as favipiravir, which in its active form acts as a guanosine mimic. The structures of the two of them are shown below:

Favipiravir



Guanosine



This molecule functions the same as Remdesivir, except instead of competing with adenosine, it is competing with guanosine to bind the RNA polymerase. Favipiravir, like Remdesivir, was previously used to treat other viruses. In the case of Favipiravir, it is an approved treatment for influenza in Japan. At the time of this writing, a growing number of countries are including Favipiravir as a treatment for Covid19 such as India. Currently, there are numerous clinical studies underway to determine the efficacy of Favipiravir to treat Covid19, some of which are in phase 3.

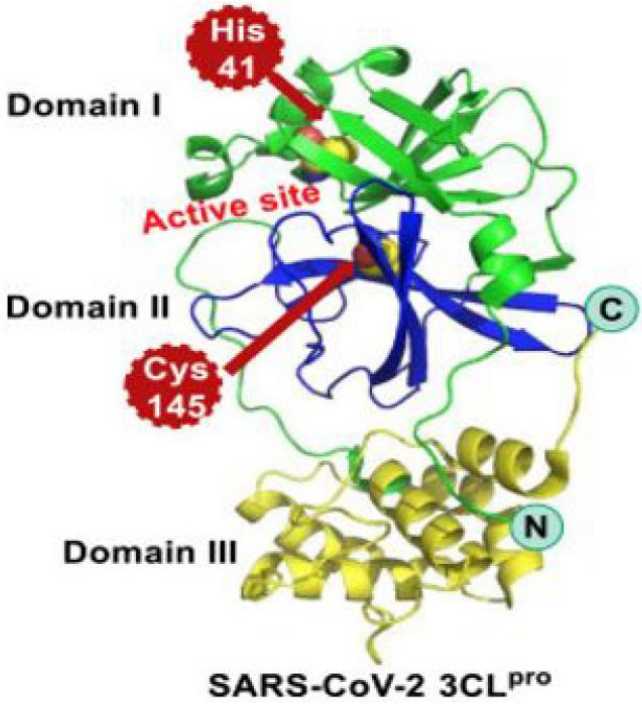
Practice Questions:

None, this is just FYI

Chapter 51: Viral protease inhibitors

Covid19 has two main proteases, the 3ClPro viral protease and the papain-like protease (PLpro). These proteases are important for viral assembly, remember that the viral proteases cut up the viral peptides into their active forms so they can come together to form active viruses. They are therefore attractive therapeutic targets. If we can inhibit these proteases then we can prevent them from making new virus particles and therefore the viral load would decrease and the disease will be easier to fight off. The main protease that researchers have targeted at the time of this writing is the 3ClPro protease.

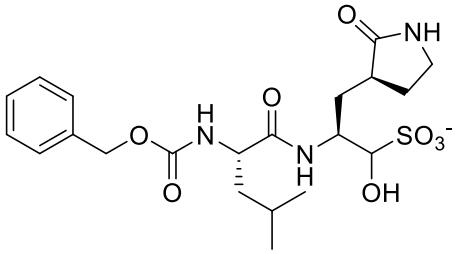
Before we discuss how the drug compounds targeting the 3ClPro enzyme works, we need to dissect the structure of the protease to understand how artificial compounds can disrupt its function. The 3ClPro is a cysteine protease, meaning it cleaves peptides using a thiolate (S^-) from its cysteine residue. Specifically, its catalytic dyad is His41 and Cys145. The entire protein structure is shown below:[13]



These two residues are absolutely required for the protease to work. Much of the protease inhibitors that are being looked at currently target 3ClPro because of its remarkable similarity to an analogous protein in the original SARS virus.

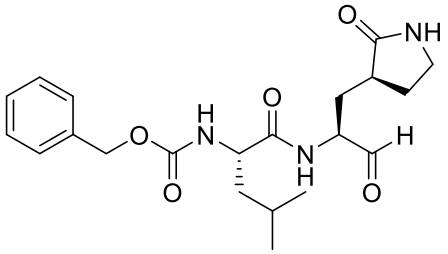
One of the most promising drug compounds to target the viral protease is GC376. This molecule is a peptide prodrug that is a derivative of phenylalanine-leucine-proline, the structure of the molecule is shown below:

GC376
"inactive"



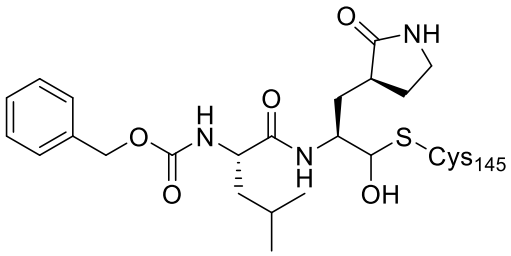
This molecule is called a prodrug because it first needs to get metabolized by the body before it can attack the 3ClPro protease. The active form is shown below:

GC373
"active"



This drug works because of the highly electrophilic aldehyde moiety that it introduces. Recall that aldehydes are some of the most reactive carbonyl compounds, coming close to anhydrides and acyl chlorides in terms of reactivity. Because of this, the aldehyde group is "primed" to get attacked by the thiolate present in the active site of the enzyme and preventing that same thiolate cleaving the peptide bonds of the viral proteins. This covalent bond formed between the enzyme and the drug is shown below:

Enzyme-drug complex



This drug can be classified as a "suicide substrate" because when it enters the active site of the enzyme, it cannot leave because of the strong covalent bond between the drug and the enzyme. This drug has not yet been tested in humans, but the *in vitro* data (the data collected from isolated cells) is promising. The active form of the drug was able to inhibit 50% of the viral proteases activity at just 0.40 μM concentration. This is also frequently referred to as the IC_{50}. The lower the IC_{50}, the more powerful the drug molecule is.

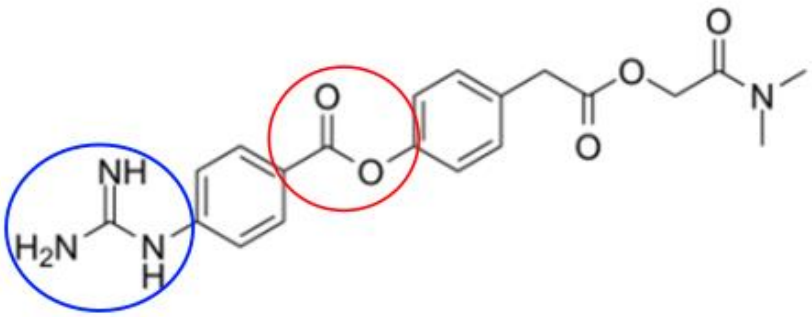
Practice Questions:

None, this is just FYI

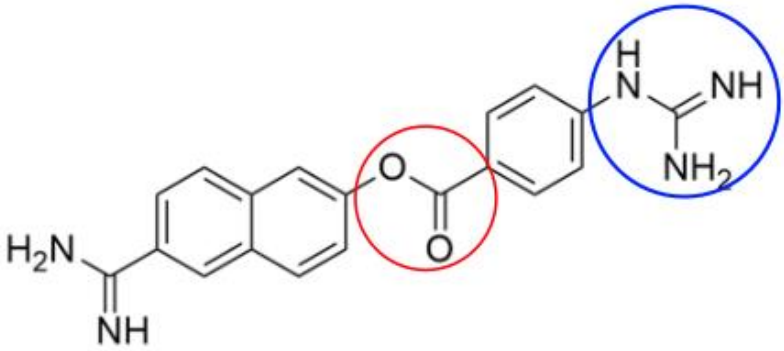
Chapter 52: Cell-entry inhibitors

Covid19 binds to human cells using the ACE2 receptor, however, in order for the virus to fuse with our cells and inject its genome, the spike protein needs to be activated by the TMPRSS2 serine protease on the host cell's membrane. This serine protease is highly expressed in the lung, where Covid19 causes the most havoc, thus inhibiting the activity of this serine protease could prevent Covid19 fusing with our cells and making more copies of itself. The principle behind this treatment strategy is that Covid19 is not harmful in the blood, it is only harmful if it infects our cells, which it can only do if TMPRSS2 is functioning.

Luckily, TMPRSS2 has been implicated in other diseases and therefore there are already many drugs that are designed to inhibit this protease. Since this protein is a serine protease, it uses the alkoxide O^- of its serine residue to attack and break peptide bonds in the active site. Therefore, if we can design a drug that has a more active carbonyl group that looks like a peptide, we can fool the protease and form a covalent intermediate just like we did with GC373 and 3ClPro in the last chapter. That is where Camostat and Nafamostat Mesylate come in. These two compounds are shown below:



Camostat Mesylate



Nafamostat Mesylate

These two drugs have two critical domains, the Guanidinium group (blue) and the ester group (red) connecting the two aromatic regions. The guanidium group aids in hydrogen bonding with the active site and promotes tighter binding while the ester group is what will get attacked by the serine residue in the active site, Ser441. This will create a covalent adduct just like we saw with GC373 and will permanently deactivate the serine protease. These two drugs are currently in phase III clinical trials done across the globe to see their effects both in isolation and in combination with other drugs thought to combat Covid19 such as Favipiravir. Preliminary results are promising and the *in vitro* results are strong. Nafamostat for example had an EC_{50} of 10 nM in lung cells. The EC_{50} is the concentration of a drug necessary to reduce the intended biological effect by 50%. The other great thing about both of these drugs is that they have been approved for treating pancreatitis in the past, so there is a robust safety history for both of these drugs being used in the clinic. If you are interested more in cell entry inhibitors, I wrote an entire report on it for my Biochemistry class, so just email if you want more information.

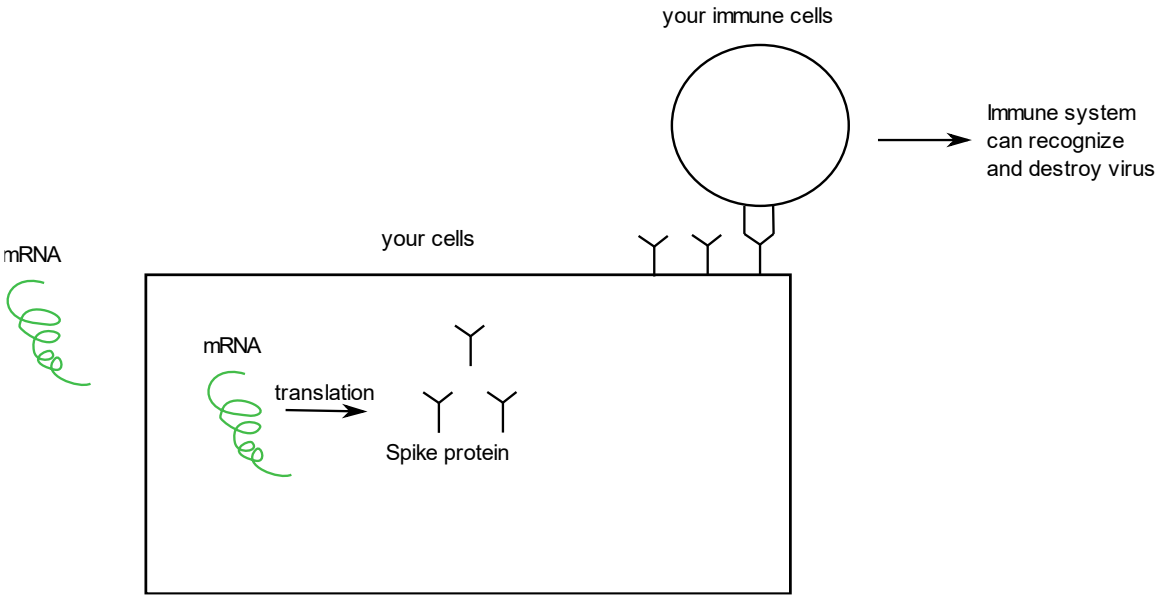
Chapter 53: Vaccine Development and Stages of Clinical Trials

There has been a lot of misinformation and confusion surrounding this virus since the beginning. It has become politicized and many nonscientists are coming up with outlandish statements trying to discredit the vaccines for Covid19. Therefore, unlike most chapters, we are going to first understand the basic science behind each vaccine that is currently either approved or pending approval by the FDA and explain why all of these conspiracy theories are absolutely outlandish and infuriate me to no end. I will also address specific concerns that many folks, such as my parents, have about the vaccine. My parents have no degree in science, however, they are highly educated people. My mother, for example, has two masters degrees and my father has a bachelors degree in Accounting. As Biochemists in training, it is our duty to explain to the public that these concerns they have are not founded in fact, but in misguided faith in people who have no science background whatsoever.

There are currently four major vaccines that are being developed: Pfizer, Moderna, Astrazenica/Sputnik, and Johnson & Johnson. These four vaccines can be further subdivided into two camps: mRNA vaccines (Pfizer and Moderna) and viral vector vaccines (Astrazenica and Johnson & Johnson). We will discuss the basic science behind these two vaccine camps and then we will go through specific bullet points that address common concerns many folks have about the vaccine. My hope is that this chapter is readable for people who are not science-savvy, so you can always give this chapter for your parents or other people in your life to read so that they are educated about the virus and the vaccine as a whole.

mRNA Vaccine Efforts (Pfizer + Moderna):

mRNA at its core are instructions that your cell uses to make proteins. Every protein that your cell makes is coded by an mRNA molecule. The m in mRNA stands for messenger because this molecule is the messenger that allows your cells to translate RNA to proteins. This goes back to the central law of cell biology: DNA to RNA to proteins. This is a strict order that does not get broken. The Pfizer and Moderna vaccines work by supplying your cells the instructions (mRNA) to make the spike protein of the coronavirus. Like we mentioned before, this spike protein is what the virus uses to get into your cells. Once the mRNA is in our cells, it starts getting translated into the coronavirus spike protein. The coronavirus spike protein can now be recognized by our immune system and we can produce antibodies against it. This means that in the future, if we were exposed to the virus, our body can recognize the virus and destroy it super quickly. This is summarized below:



Viral Vector Vaccine Efforts (AstraZenica and Johnson&Johnson):

These vaccines work in very similar ways to the mRNA vaccines described before, except they use inert viruses to give the mRNA to our cells. These viruses ARE NOT Covid19, they are simply modified so that they will insert the gene to make the coronavirus spike protein so that the body can recognize the spike protein and destroy the actual coronavirus if we get exposed to it in the future in the exact same way as the mRNA vaccines. The only difference between the two vaccines is that one uses the mRNA by itself while the other uses an inert virus to give our cells the mRNA. Within the viral vector camp, the only difference between the Astrazenica vaccine and the Johnson&Johnson vaccine is the number of shots needed for each. Astrazenica requires two shots while Johnson&Johnson only requires one.

Now to address some concerns that I have heard from many people:

Concern: We don't know the long-term side effects of the vaccine

Answer: The mRNA given to our cells to make the spike protein is quickly destroyed by the cell. The chances of long-term side effects would therefore be very low and they would likely be seen within a few weeks after vaccination if at all. Additionally, mRNA vaccines against SARS have been looked at since 2003 and people volunteered to get these vaccines against SARS at that time. So far, 18 years later, there has been no long-term side effects.

Concern: Won't this vaccine give me the coronavirus?

Answer: No, the vaccine will not give you the coronavirus. None of the four major vaccines administer Covid19 to patients. Pfizer and Moderna give the mRNA that codes for just the spike protein of the coronavirus, not the virus itself. The Astrazenica and Johnson&Johnson vaccine

give an inert virus that is completely different from Covid19. Therefore, there is NO scientific way to get Covid19 from ANY of the vaccines.

Concern: I heard on the news that someone got the Covid vaccine and they tested positive for Covid later on. Does that mean the vaccine does not work?

Answer: No. At the time of this writing, people who received the vaccines have only received the first shot, which provides some protection, but not full protection from the virus. Additionally, vaccines are inherently proactive, meaning that they are only worth something if you do not currently have the virus when the shot is administered. It is possible that the person who tested positive had the virus prior to getting the vaccine but never got tested so they didn't know until just now. That DOES NOT mean the vaccine doesn't work, it just means that they already had the virus before getting the vaccine, thus rendering the vaccine null and void.

Concern: Why do I need to get two shots? Shouldn't one shot be enough?

Answer: It depends on the vaccine you are referring to. Johnson&Johnson is the only vaccine that requires one shot to grant immunity to the virus. The other vaccines require two shots to to maximize the immunity the patient gets. Effectively, the first shot is the sampler shot and provides some modest immunity to the virus and the second shot is the real deal and will provide close to full immunity in most people. They do not combine the two shots because they find the maximum immunity if they do the two separately.

Concern: Why do these vaccines cause fatigue and horrible side effects?

Answer: The vaccine causes fatigue and flu-like symptoms because the spike protein is being recognized and destroyed by the body. Whenever there is a foreign object in the body, the immune system responds and causes inflammation, fatigue, etc. This is not just with this vaccine, but in any scenario where your body is mounting a full on war with a foreign entity. In other words, you get fatigue because the vaccine is working.

Concern: How were they able to rush this vaccine so fast? I don't trust it.

Answer: These vaccines were produced at a very fast and even unprecedented rate. I told my mom that these wouldn't be ready until probably late 2021 at the earliest and I was wrong, so I am right there with you. However, we have to consider several factors that are different about this vaccine compared to previous vaccine efforts

1. Funding and removing red tape
2. International support
3. Technology
4. Financial incentive
5. Previous knowledge on similar viruses

The Covid19 virus has infected so many people and has quickly become an international concern. This is very different from Ebola, SARS, and MERS, which quickly either died out or were contained in one region more or less. This has caused governments such as the US to heavily fund vaccine development efforts and caused Operation Warp Speed to speed up vaccine

development by removing red tape that stalled previous vaccine efforts. For anyone who has worked for the government before or any extremely large corporate enterprise, red tape can stall things for months if not years.

Because of the severity of the problem, scientists all around the world are working on this issue rather than a select few. This greatly speeds up science regardless of the subject matter and is what has allowed so many great advancements in the past three decades.

Now in 2020, we have a very good understanding of virology and have the technology to quickly sequence the virus and to engineer inert viruses to administer the mRNA that codes for the spike protein. We have unprecedented technology at our fingertips and therefore all other vaccine development efforts pale in comparison.

Companies are also extremely invested in finding a vaccine because the market for it is absolutely huge, therefore, private enterprise is incentivized to collaborate with anyone and everyone to help find a vaccine as soon as possible.

Adding on to all of the factors above is the fact that Covid19 is EXTREMELY similar to SARS, which we have been studying since 2003!

Concern: Will I be able to afford it?

Answer: Many major pharmacies such as Rite-Aid have said that the Covid vaccine will be free of charge. This is because of the funding thing I mentioned before, the government has effectively bought millions of vaccines in advance using our tax dollars, so the end result is that we can get the vaccine for free.

Concern: I heard some very smart people like Robert F. Kennedy Jr. say that the vaccine will alter my DNA. Is that true?

Answer: Absolutely not! All the vaccines that are administered to patients are giving them mRNA. mRNA DOES NOT ALTER DNA!!! Remember the central principle of cell biology: DNA to RNA to protein. People who say that the vaccine will alter your DNA are saying RNA goes to DNA, which is completely backwards. That and Robert F. Kennedy Jr. is a lawyer with no science background and is a known antivaxxer; he likely does not understand how the vaccine works.

Concern: Now we have this new coronavirus variant from England, will this affect how well the vaccine works?

Answer: This is a legitimate question that I had to ask myself too, so I understand completely why this is a concern. Most news articles just say “there is no evidence to suggest that it does”, which I find very intellectually unsatisfying and a bit shady, so let’s break it down. When the immune system recognizes the spike protein, it recognizes several spots on that protein. Hundreds of spots called epitopes. If only one of those epitopes changes because of the mutation then the vaccine will still work because the other epitopes did not change. So far, the data suggests that the mutation present in the coronavirus variant from England does not change the spike protein sufficiently to prevent the immune system from recognizing it. Therefore, if the

immune system recognizes the spike protein still then it will still attack the coronavirus regardless of what variant it is.

Concern: What does all of these clinical trial stages mean? I don't understand how Operation Warp Speed allowed stages 1-3 to be combined together.

Each stage is basically a checkpoint that is designed to extract certain information from the drug or vaccine in question. Stage 1 is typically for short-term safety and only involves about 20 otherwise healthy, young adult patients who volunteer to be involved in the study. Stage 2 is typically for dosing and continual safety testing. Because stage 1 already showed short-term safety, stage 2 typically has a few hundred or even a thousand patients ranging in health and age. Stage 3 is the final test before FDA approval and is generally for longer-term safety, clinical efficacy, and finalizing dosage. This stage typically has several thousand participants. After each stage, the FDA or other governmental organizations have to approve the drug or vaccine to move on to the next stage or request additional data, etc. In reality, these distinctions are more for organization and are fraught with bureaucratic red tape, therefore Operation Warp Speed was approved for coronavirus vaccine efforts. The new system combined stages 1-3 with periodic checkpoints and still had the safety testing and clinical efficacy data collection as normal clinical trials, however, the red tape was largely removed and therefore they combined stages 1-3 together to streamline the process to get us the vaccine faster without compromising safety.

Focus 11: DNA Replication

Chapter 54: Initiation of DNA Replication in Prokaryotes

When prokaryotic cells need to divide, they must replicate their DNA. All organisms replicate their DNA in a semi-conservative fashion, meaning that each parent strand is used as a template for replication. Here, we will discuss how DNA replication starts in prokaryotes and discuss the major proteins involved in allowing this to occur.

First, we should start off by defining our major players in DNA replication:

- Helicase (DnaB) = unwind double helix
- Topoisomerase (DNA gyrase) = removes supercoils ahead of replication fork
- Single-stranded binding proteins = stabilizes single stranded DNA during replication
- DNA polymerase III = major DNA polymerase that synthesizes both leading and lagging strands
- DNA polymerase I = removes Okazaki fragments
- DNA ligase = seals up nicked DNA after polymerase I does its thing
- Primase (DnaG) = Adds RNA primer to allow elongation
- DnaC = carries helicase to replication fork
- DnaA = Initiates replication by unwinding DNA

This will be our definition chart moving forward and is essentially the gospel of DNA replication.

In prokaryotes, DNA is generally circular and contains a singular origin of replication called ori. This ori contains several R and I sites to which DnaA will bind. DnaA binding causes a conformational change in the DNA double helix and this causes a nearby A=T rich region called the DNA unwinding element to unwind, this is unwound region is called the replication fork. Once the DNA unwinding element is opened up, DnaC will carry helicase to the replication fork and put one helicase on either side of it at the expense of ATP.

Once helicase is placed on both sides of the replication fork, DNA polymerase III and primase are linked to it. This allows elongation to occur in the future and this induces DnaA to leave the R and I sites after it hydrolyzes ATP. Prokaryotic DNA can only replicate if ATP-DnaA is bound and therefore this step allows the cell to regulate how often it replicates its DNA (MAS**C**). Replication initiation is also regulated by methylation. DnaA will only recognize and bind to methylated DNA, however, newly made DNA is not methylated until some time after replication. Therefore, the cell can control the maximum rate of replication through methylation and amount of ATP-bound DnaA. Eventually, the newly synthesized daughter strands will be methylated at adenine bases using DAM methylase. Overall, this limits bacterial replication to once every 20 minutes (MAS**C**).

Practice Questions:

1. Why is the DNA unwinding element A=T rich? Why could it not be G C rich? Explain.

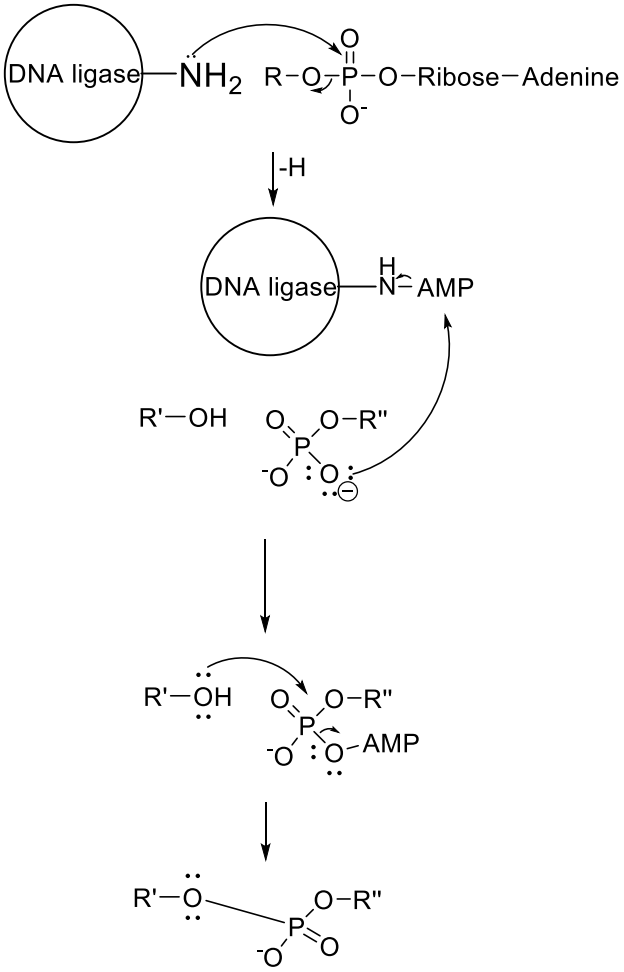
2. Suppose DAM methylase is mutated so that it is constantly on and DnaA is mutated so that it can release from the R and I sites without hydrolyzing ATP. What effect would this have on the amount of time bacteria could replicate their DNA? Explain.

Answer:

1. The DNA unwinding element is A=T rich because A and T base pairs only have 2 hydrogen bonds binding them together whereas G and C base pairs have 3. Therefore, to facilitate unwinding the DNA, it is easier and more efficient if the DNA unwinding element is A=T rich (MAS**C**).
2. If DAM methylase is perpetually on and DnaA can release from R and I sites at will, then the bacteria would be able to replicate its DNA more frequently and therefore could divide more rapidly. There would no longer be the 20 minute refractory period because DnaA will be able to bind the Ori all the time since the daughter strands will be immediately methylated by DAM methylase. DnaA will also be able to leave the R and I sites without hydrolyzing ATP to ADP and therefore that level of control is lost too.

Chapter 55: Elongation during DNA Replication in Prokaryotes

Once the helicase, DNA polymerase III, and primase are on both sides of the replication fork, elongation can occur. The elongation process starts with an RNA primer added by primase to both sides of the replication fork. The DNA polymerase III can only add nucleotides from the 5' to 3' direction because the 3' carbon has the OH group that it uses to extend the chain. However, one strand at each replication fork does not go 5' to 3'. The chain that does not go from 5' to 3' is known as the lagging strand and it is this strand that needs several primers and needs to be synthesized in small portions that will ultimately be connected together using Pol I and ligase. The small segments of DNA that are present on the lagging strand are called Okazaki fragments. The leading strand is much simpler to synthesize, it only requires one primer and the DNA polymerase III can continually add new nucleotides to the chain because it is going in the direction of the replication fork. The lagging strand is synthesized using the exact same DNA polymerase as the leading strand. This is done by bending the DNA, adding more primers, and using one beta clamp for each primer added. Bending the DNA allows the polymerase to add nucleotides to the lagging strand without stalling the replication fork. Each primer allows the DNA polymerase to start adding nucleotides at those positions and the beta clamps are used to stabilize the polymerase and keep it on the desired DNA segment. This is called the slide-trombone model of DNA replication. With each new primer of the lagging strand, there must be a new beta clamp and each beta clamp costs 3 ATP to add. Once all of these Okazaki fragments are made, DNA polymerase I uses its "bulldozer" functionality (5'-3' exonuclease) to remove nucleotides in front of it and replace the RNA primers with DNA nucleotides. The polymerase stops just before the start of the next DNA nucleotide sequence, creating a "nick". These nicks are sealed up by DNA ligase, which uses ATP or NAD^+ to seal the nucleotide gap. The mechanism of DNA ligase is shown below:

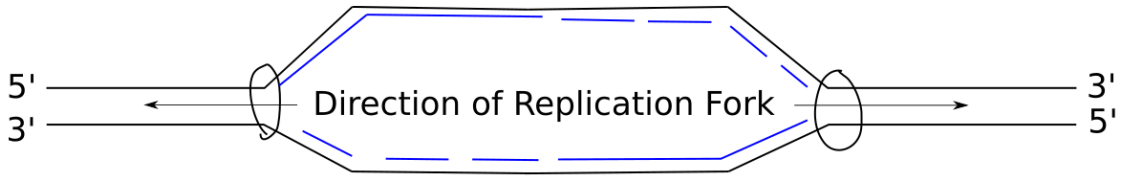


Here, the AMP is acting as the leaving group and the 3' OH of the left hand nucleotide denoted R'-OH is able to now attack the 5' PO_4^{2-} of the right hand nucleotide. This is required because initially, the only good leaving group on the 5' nucleotide is the phosphate, thus adding the AMP allows the OH group to attack the phosphate and displace the AMP, rather than attacking the carbon and displacing the phosphate. This elongation process continues until Ter/Tus sites are reached on the opposite side of the circular DNA molecule. Once one side of the replication fork meets the Ter sequence, the Tus protein prevents that side from moving. Eventually, the other side collides with the stalled replication fork and replication is over and DNA topoisomerase IV separates the parent and daughter DNA molecules.

Practice Questions:

1. Compare and contrast leading and lagging strand synthesis.
2. Label the following diagram of a replication fork and explain how you know which side is the leading and lagging strand:

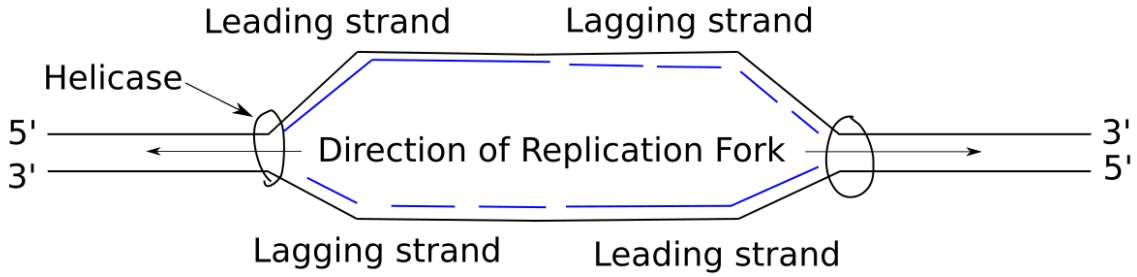
Code:
Black = Original DNA
Blue = New DNA



3. DNA polymerase III has a 3'-5' exonuclease that allows it to correct the mistakes it makes during replication. What would happen to bacteria if this 3'-5' exonuclease was removed?

Answers:

1. Leading strand synthesis requires only one primer, one beta clamp, and creates one continuous DNA strand. Lagging strand synthesis requires multiple primers, several beta clamps, and creates several discontinuous DNA strands called Okazaki fragments. The lagging strand requires much more ATP to synthesize because each Okazaki fragment needs to be sealed using DNA ligase and each beta clamp requires 3 ATP to add onto the DNA strand.
2. Code:
Black = Original DNA
Blue = New DNA



The leading strand is synthesized continuously from 5'-3'. Remember that the nucleotides are added antiparallel. On the left hand side, the black is 5' on the left and therefore must be 3' for blue. Since the 3' blue side is moving with the replication fork, that must be the leading strand for that side of the replication fork. The lagging strand is synthesized discontinuously also from 5'-3'. Again looking at the left hand side, 3' black means 5' blue and therefore the nucleotides are being added in the opposite direction of the replication fork. Thus the bottom left is the lagging strand. The enzyme that is

responsible for breaking the DNA double helix is helicase and therefore the circle on either side of the replication fork is helicase.

3. If the DNA polymerase III 3'-5' exonuclease did not work, then the DNA polymerase would be more error prone and therefore more mutations would result during DNA replication.

Chapter 56: Comparison between Replication in Prokaryotes and Eukaryotes

Bacteria and prokaryotes have a much less sophisticated approach to replication compared to Eukaryotes. Here, we will compare and contrast replication in prokaryotes and eukaryotes.

The first major difference between prokaryotic and eukaryotic replication is the regulation of it. In prokaryotes, the regulation is based solely on DnaA binding. If DnaA was bound to ATP, then it could bind the R and I sites of the origin of replication, Ori IF those nucleotides were methylated by DAM methylase first. Therefore, regulation was very much DnaA-centric in prokaryotes. Eukaryotes have a similar system of regulating when their DNA is replicated, however, it functions within the broader context of the cell cycle, not just DNA replication. This system focuses on the levels of specific cyclins and cyclin-dependent kinases (CDKs). These CDKs are themselves regulated by DNA-damage sensing proteins such as p53 that will prevent replication of damaged DNA so that the genome remains intact and cancer does not develop. In this way, Eukaryotic cells have a more robust way of regulating the cell cycle and DNA replication overall (MAS**C**).

The second major difference is the initiation process. In prokaryotes, DnaA binds the R and I sites of Ori. There was only one Ori sequence in the entire prokaryotic genome. This is very different in eukaryotes. In eukaryotes, there are several hundred autonomously replicating sequences (ARS) and the entire genome is replicated just once per cell cycle. Rather than DnaA, eukaryotes have a more robust protein to initiate replication called the origin recognition complex (ORC).

The third major difference is rate and the way the leading and lagging strands are synthesized. In prokaryotes, the DNA polymerase is 20x faster than the eukaryotic DNA polymerase. Moreover, in prokaryotes, DNA polymerase III synthesized BOTH leading AND lagging strands, however, in eukaryotes, there is a dedicated DNA polymerase for both strands (DNA polymerase delta for lagging strand and DNA polymerase epsilon for leading strand).

Finally, the fourth major difference is how the replication process terminates. In prokaryotes, one side of the replication fork gets stalled because of the Ter/Tus sites opposite Ori, then the other side collides with the stalled fork to terminate replication. This is a byproduct of prokaryotic DNA being circular in nature. Eukaryotic DNA on the other hand is linear, therefore, replication terminates simply when the end of the parent strand is reached.

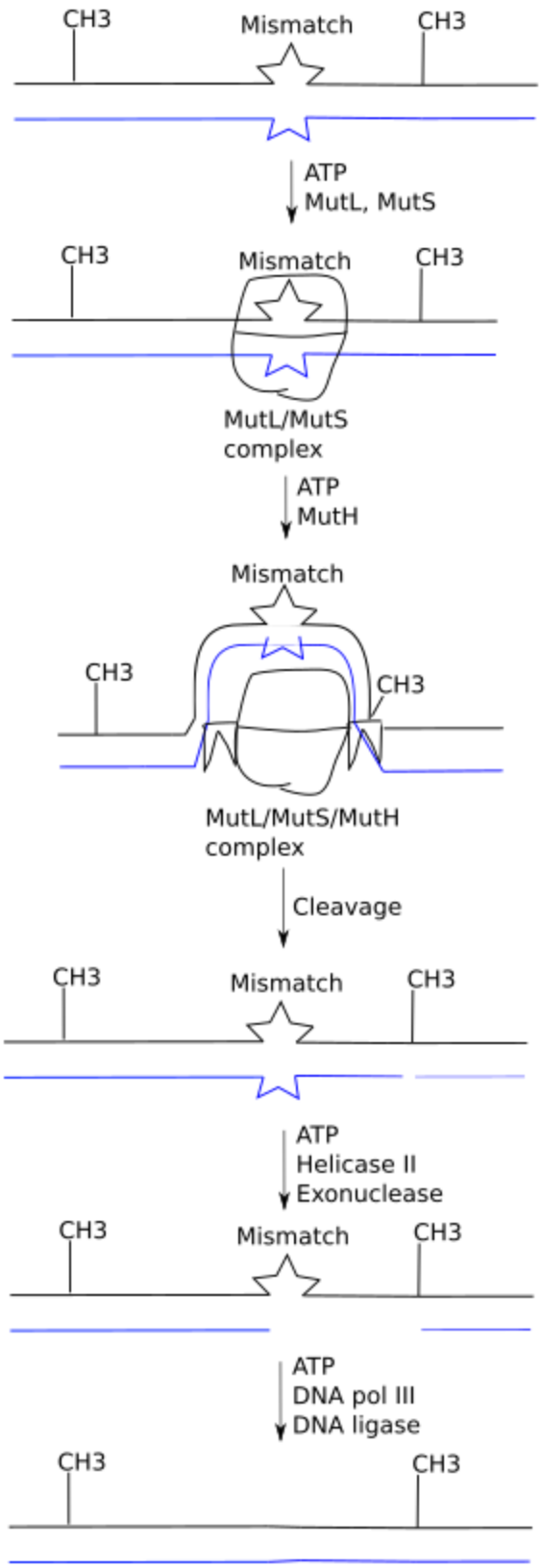
Focus 12: DNA Repair

Chapter 57: Mismatch Repair

If after replication of the DNA, the polymerase still made some errors that it did not resolve, the cell can do what is called mismatch repair. This occurs when the wrong nucleotide was added to the daughter strand. But how does the cell know which strand is the new one? Recall that the daughter strand is not initially methylated, therefore, all the possible errors from the polymerase must occur on the UNmethylated strand. This is the logic that mismatch repair takes advantage of in its mechanism. There are three major players in mismatch repair: MutL, MutS, and MutH. MutL and MutS always work in pairs and they are used to recognize mismatched base pairs, MutH on the other hand is responsible for cutting the unmethylated daughter strand to allow the mismatch to be removed. Here is effectively the entire process in a nutshell:

1. MutL and MutS hydrolyze ATP to bind the mismatched region of the DNA
2. The DNA threads through the MutL/MutS complex until a methylated A base is found
3. MutH hydrolyzes ATP and nicks the unmethylated strand
4. DNA helicase II and exonuclease are used to remove DNA from one methylated A to the mistmatch site at the expense of ATP.
5. DNA polymerase III comes in to add the correct nucleotides
6. DNA ligase seals the gap at the expense of ATP

This is shown visually below:


CH3
Mismatch
CH3
ATP
MutL, MutS
CH3
Mismatch
CH3
MutL/MutS
complex
ATP
MutH
Mismatch
CH3
CH3
MutL/MutS/MutH
complex
Cleavage
CH3
Mismatch
CH3
ATP
Helicase II
Exonuclease
CH3
Mismatch
CH3
ATP
DNA pol III
DNA ligase
CH3
CH3

Practice Questions:

1. How would the mismatch repair system be affected if the DAM methylase was mutated such that it was continuously active? What would the end result of this mutation have on the stability of the genome? Explain.
2. Suppose MutH was mutated such that it can bind MutL and MutS, but cannot cleave DNA. How would this affect DNA repair? Explain.

Answers:

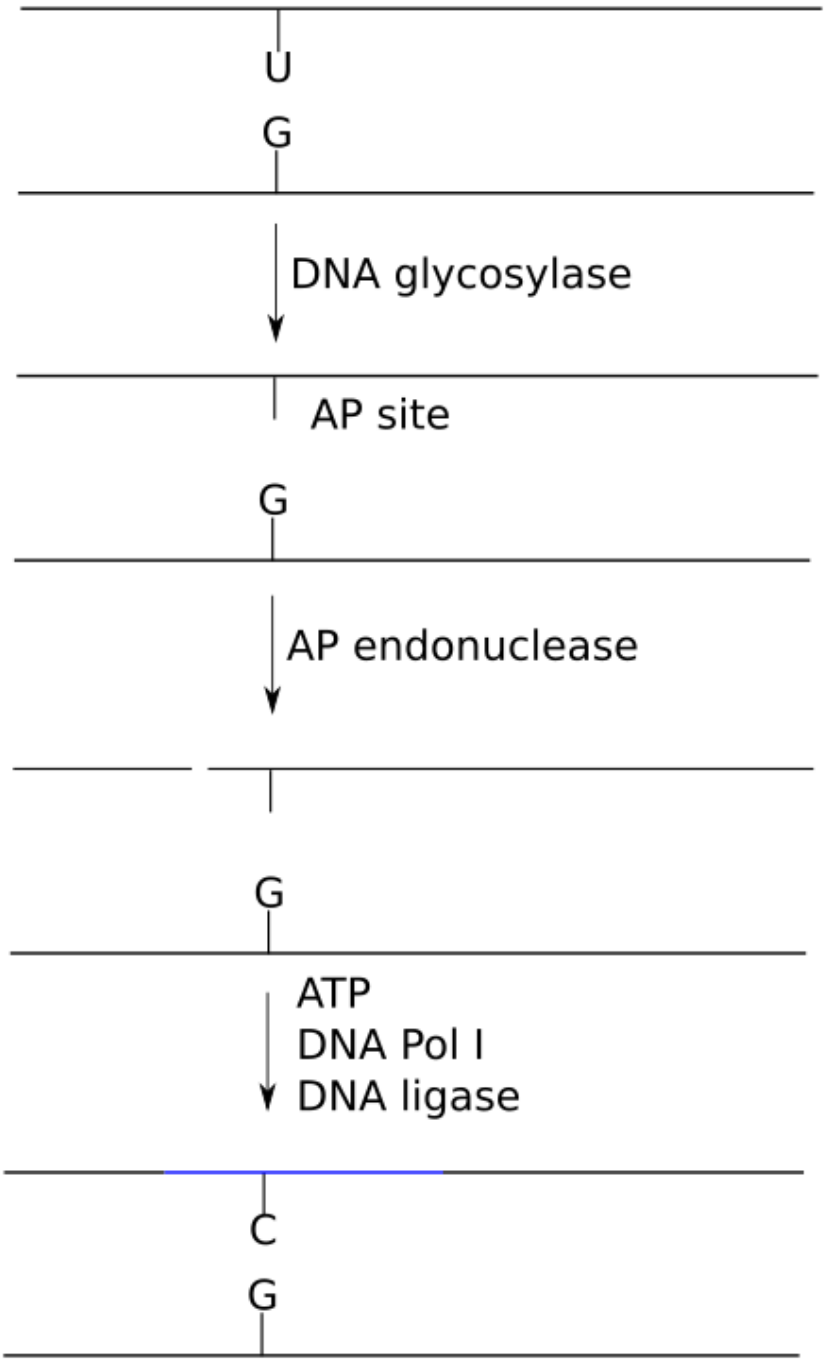
1. If DAM methylase was always on, then it is possible that mismatch repair would not occur since the daughter strand would be immediately methylated during replication. Because mismatch repair would not occur, the stability of the genome would be compromised. This safeguard against mutations would be missing.
2. If MutH cannot cleave DNA, then the MutL/MutS/MutH complex would not dissociate from the DNA and cannot initiate the mismatch repair system via cleavage. Because of this, mutations would be more likely and the cell would likely not divide as rapidly; the cell would be stuck with a bunch of MutL/MutS/MutH complexes on its DNA.

Chapter 58: Base Excision Repair

Base excision repair occurs when the base of a nucleotide is altered such that it is no longer a DNA nucleotide (C to U for example) or it is a base that should not belong in either DNA or RNA (such as covalently modified bases). Base excision repair is done using four major players:

1. DNA glycosylase: removes abnormal base
2. AP endonuclease: nicks DNA where the abnormal base was
3. DNA polymerase I: replaces abnormal base and nucleotides surrounding it with correct nucleotides, leaving a nick at the end
4. DNA ligase: seals nick

This works because the DNA polymerase I has the “bulldozer” functionality aka its 3’-5’ exonuclease. Therefore, it can remove nucleotides in front of the abnormal base and replace them with correct nucleotides. Remember, DNA polymerase III cannot do this because it DOES NOT have the “bulldozer” ability. In mismatch repair, that didn’t matter because in that process, we used exonucleases and THEN the DNA polymerase. Here, we are combining those two steps by using DNA polymerase I. This is shown visually below:



Practice Questions:

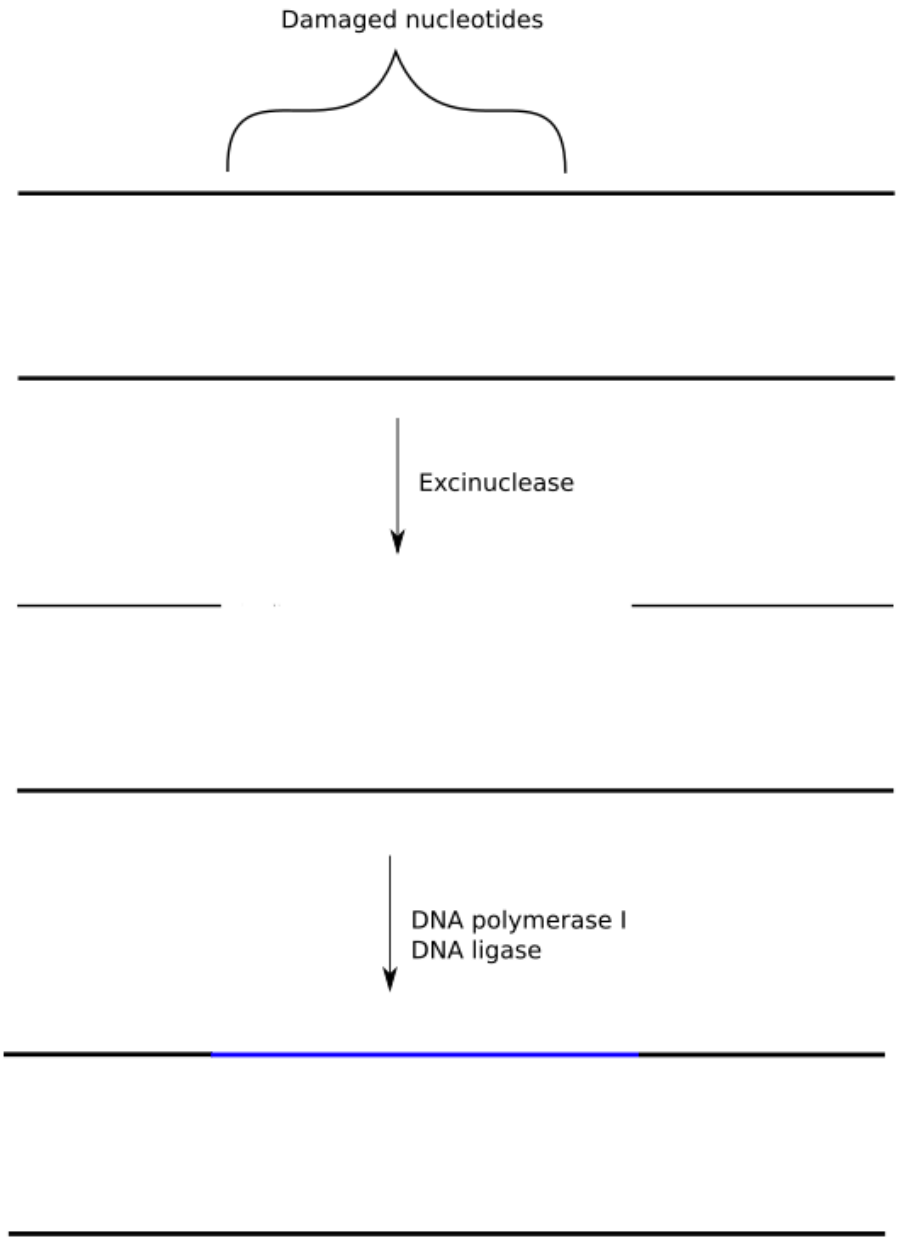
1. What would the effect on base excision repair be if DNA polymerase I was mutated such that its bulldozer no longer works? Explain.
2. Which form of DNA repair costs more energy, base excision or mismatch repair? Explain.

Answers:

1. If DNA pol I was unable to use its bulldozer, then base excision repair would no longer work and the DNA would be permanently nicked by the AP endonuclease, likely causing widespread damage to the genome.
2. Mismatch repair costs more ATP because it uses DNA helicase II, MutL/MutS, MutH, and DNA ligase to repair DNA damage, each of which costs at least 1 ATP each. By comparison, base excision repair only requires 1 ATP for the DNA ligase step.

Chapter 59: Nucleotide Excision Repair

When the DNA has a lot of consecutively damaged nucleotides, the cell does nucleotide excision repair. This type of repair is actually extremely simple. It uses a special kind of nuclease called an excinuclease that will cleave the DNA backbone at two places simultaneously and in E. Coli typically remove 12-13 nucleotides at a time. Once these nucleotides are removed, DNA polymerase I and ligase are used to replace the damaged nucleotides with the correct ones and the gap is sealed. This is shown visually below:



In humans, the same thing happens except a large nucleotide fragment is taken (27-29) and the gap is filled using DNA polymerase epsilon and DNA ligase.

Practice Questions:

1. What is more energetically expensive, nucleotide excision repair or mismatch repair?
2. What would happen if the excinuclease had one of its catalytic domains mutated such that it no longer worked? How would this affect nucleotide excision repair?

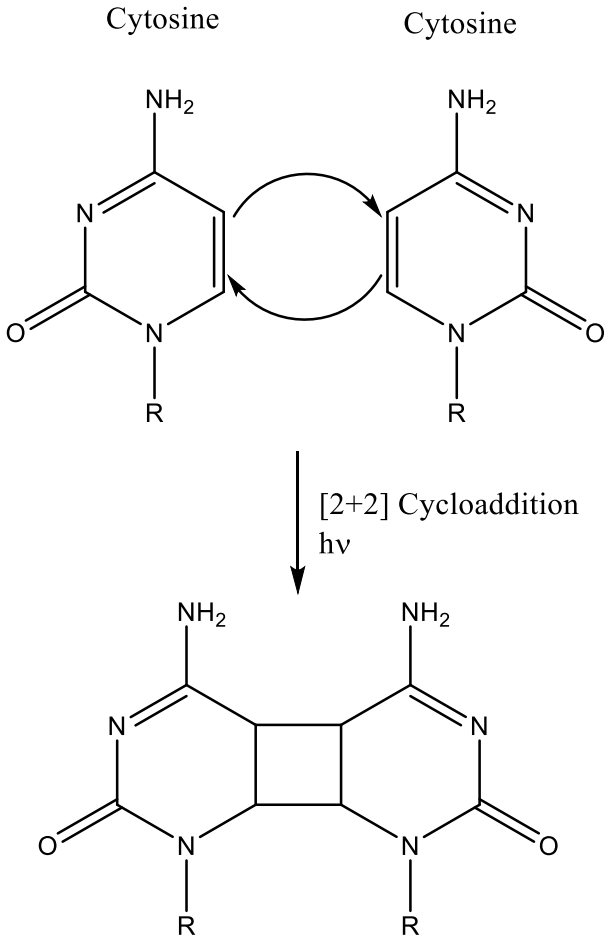
Answers:

1. Mismatch repair because of the energy investment to create the MutL/MutS/MutH complex and because more nucleotides are typically removed in mismatch repair.
2. If the excinuclease had one of its catalytic domains deactivated, it would only nick the DNA rather than remove the entire segment, therefore nucleotide excision repair would be compromised and the DNA would unfurl to release topological strain.

Chapter 60: Direct Repair (Prokaryotes only)

It is well known that many of the DNA bases can be covalently modified in the presence of mutagenic chemicals and UV light. In bacteria, several enzymes are involved with directly repairing each of the common modified bases. Here, we will discuss some of the common mutations that occur and how bacteria use special single-turnover enzymes to fix these errors in the genetic code.

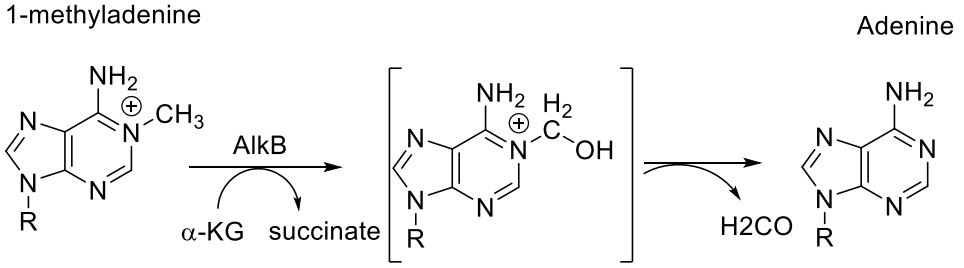
One of the most frequently encountered DNA errors is pyrimidine dimers, which occurs because of an internal [2+2] cycloaddition reaction between nearby pyrimidine bases. The mechanism for this is shown below:

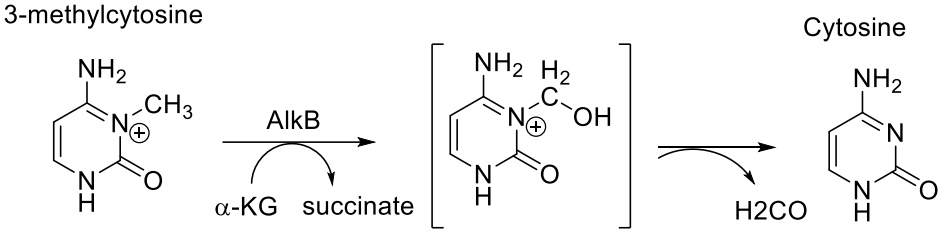


Mutagenic cytosine dimer

These pyrimidine dimers are resolved using an enzyme called photolyase, which uses a MTHFpolyGlu cofactor and FADH to repair pyrimidine dimers back to their initial state. This mechanism proceeds through free radicals, the details are not important to discuss, I leave that to your own search.

The next most common mutation that occurs is methylated bases, specifically 1-methyladenine and 3-methylcytosine. These methyl groups are removed using AlkB, which is a metalloenzyme that uses Fe^{2+} and α-ketogluterate to remove these methyl groups and produce succinate and formaldehyde in the process. This process is shown below:





And finally, the third most common mutation is O^6-methylguanine. This mutation is dangerous because O^6-methylguanine will base pair with thymine and therefore could cause a G→A mutation during replication. Because of this very dangerous mutagenic potential, bacteria have evolved a specific enzyme that will recognize and demethylate O^6-methylguanine.

Practice Questions:

1. Why do you suspect that 1-methyladenine and 3-methylcytosine are such harmful mutagenic bases? Explain.
2. What do you suspect would happen to the rate of mutation if no Fe^{2+} was present in the bacterial cell and there was no means of obtaining exogenous Fe^{2+}? Explain.

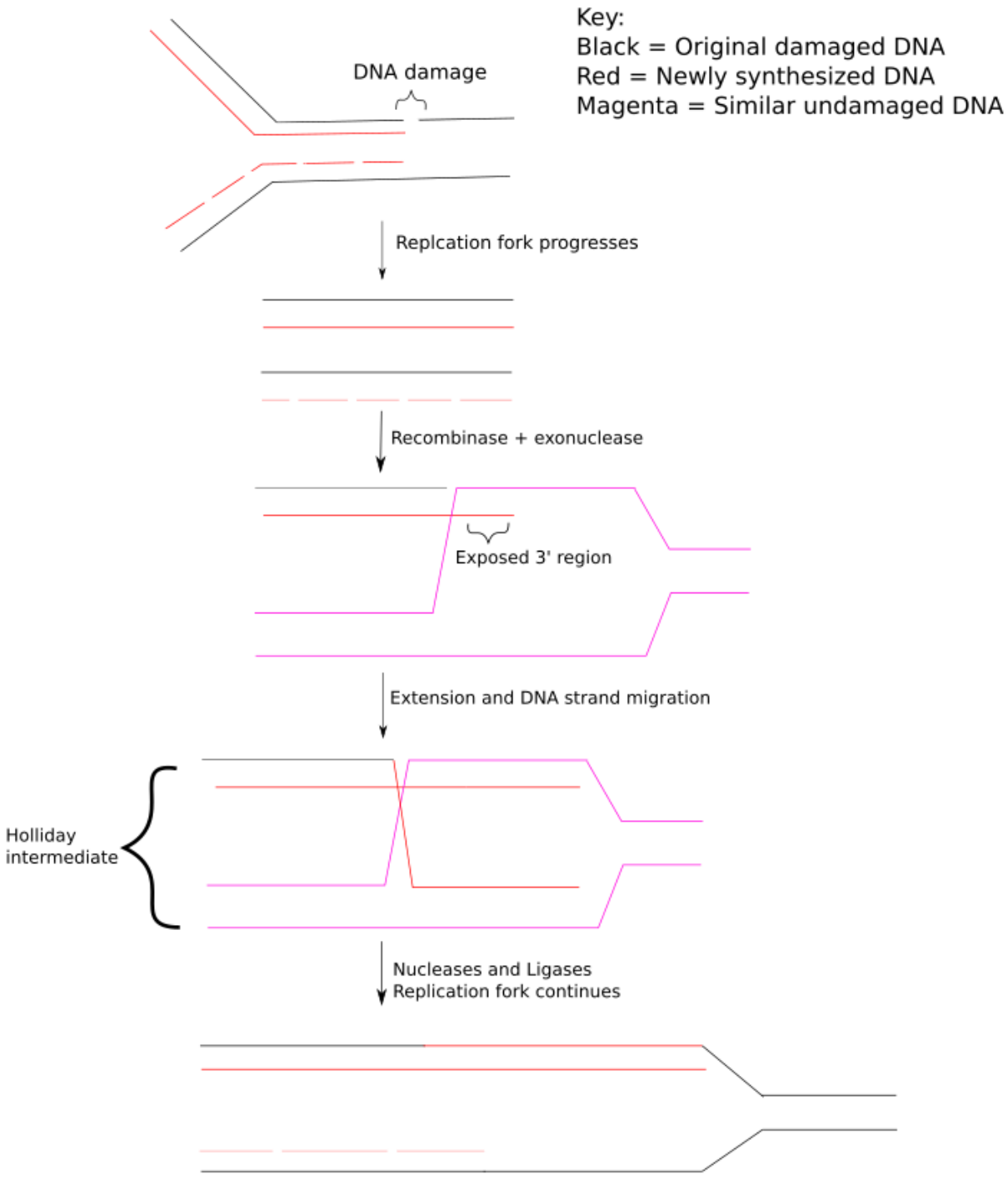
Answers:

1. All bases will pair with each other on the premise of optimal intermolecular forces that they can obtain by pairing up (**M**ASC). Adenine usually pairs with thymine because they form two strong hydrogen bonds with each other and cytosine usually pairs with guanine because they form three strong hydrogen bonds with each other. Adding methyl groups to the 1 position of adenine and the 3 position of cytosine disrupts this hydrogen bonding and prevents these bases from pairing with their partners.
2. If there was no Fe^{2+} present in the cell or in the surrounding environment, then the AlkB enzyme would lack its necessary metallic cofactor Fe^{2+} that it needs to oxidize and subsequently demethylate 1-methyladenine and 3-methylcytosine.

Chapter 61: Homologous Recombination

Sometimes DNA damage is so extensive and somehow gets through all of the previous repair mechanisms that it is still present during replication. These extensive DNA lesions can get fixed in two distinct ways: homologous recombination or trans lesion DNA repair. In this chapter, we will cover homologous recombination and in the next, we will cover trans lesion DNA repair. Both of these mechanisms of DNA repair are united by the fact that they have not undamaged DNA template to use on the same chromosome.

If there is a highly similar region on another chromosome to be used as a template, homologous recombination can be used to fix the lesions and that is where our story begins. In homologous recombination, a nearby chromosome with undamaged DNA invades the DNA double helix near the DNA lesion. This undamaged DNA is separated into two strands, each of which acts as a template for DNA replication on the damaged DNA chromosome during replication. This is facilitated by a recombinase enzyme that assists in creating a Holliday intermediate. The Holliday intermediate is an X-shaped DNA structure that allows the damaged and undamaged DNA to combine together and allow replication to continue after nucleases and ligases restore the structure of the replication fork on both chromosomes. This is shown visually below, keep in mind that we focus solely on the leading strand after the second step and then we bring back the lagging strand at the very end:



Practice Questions:

1. What would happen to the organism if the recombinase was mutated such that it did not work?

Answers:

1. If the recombinase did not work, then the organism would likely have to do trans lesion DNA repair rather than homologous recombination. Otherwise, the organism cannot replicate its DNA and would likely die out.

Chapter 62: Trans Lesion DNA Repair

When there is no nearby similar chromosome to use as a template for unresolved DNA lesions, the cell has to go to its last resort: trans lesion DNA repair. This mechanism of DNA repair uses an error-prone polymerase, DNA polymerase V, which is created from UmuD' and UmuC binding to RecA. This DNA polymerase V is an extremely error-prone polymerase that is only used for the region of extensive DNA damage. This DNA polymerase just puts nucleotides in the region of DNA damage to allow the replication process to continue with no regard to accuracy. This can kill some cells, but some cells will inevitably survive by random luck and therefore its better than doing absolutely nothing, which would guarantee cell death. In eukaryotes, the process is more sophisticated and there are specific error-prone polymerases that will recognize and replace specific forms of DNA lesions such as pyrimidine dimers.

Practice Questions:

1. When describing DNA polymerases, there are two major parameters that are important: processivity and fidelity. Processivity is the number of nucleotides added before the polymerase leaves the DNA while fidelity is the error rate of the polymerase. Knowing this information, what do you suspect the processivity and fidelity of DNA polymerase V is relative to DNA polymerase III? Explain.
2. The Ames test is a means of determining the mutagenic potential of a chemical substance. In the Ames test, Salmonella that is engineered to not be able to make their own histidine are placed in a medium that lacks histidine and contains a certain suspected mutagen. Colonies will only grow if the Salmonella mutates such that they can now make their own histidine, which is an essential nutrient. This mutation is more likely to happen the stronger the mutagen is and therefore more colonies forming indicates a stronger mutagen. Explain how a chemical substance can cause mutations in the genetic code and how those mutations may cause trans lesion DNA repair. Tie this all in with the Ames test.

Answers:

1. DNA polymerase V is only meant to be used over a short region of damaged DNA and is known to be error-prone. Thus DNA polymerase V likely has lower processivity and lower fidelity compared to DNA polymerase III.
2. Mutagens will cause mutations in the DNA, such as alkylating certain bases. If these bases are not corrected before replication, they will stall the replication fork. If there is a similar chromosome nearby then the cell can do accurate homologous recombination, however, if no such chromosome exists, then trans lesion DNA repair occurs that causes a wide variety of other mutations. These mutations could occur in the region of the genome that encode for the histidine synthesis pathway and therefore could allow the cell to, by chance, mutate to allow it to make histidine again. These new mutated bacteria can now make their own histidine and grow, this will be favored by natural selection and the other bacteria will be outcompeted and die off. Thus colonies of these bacteria will be the most prominent and be visible on the Petri dish.

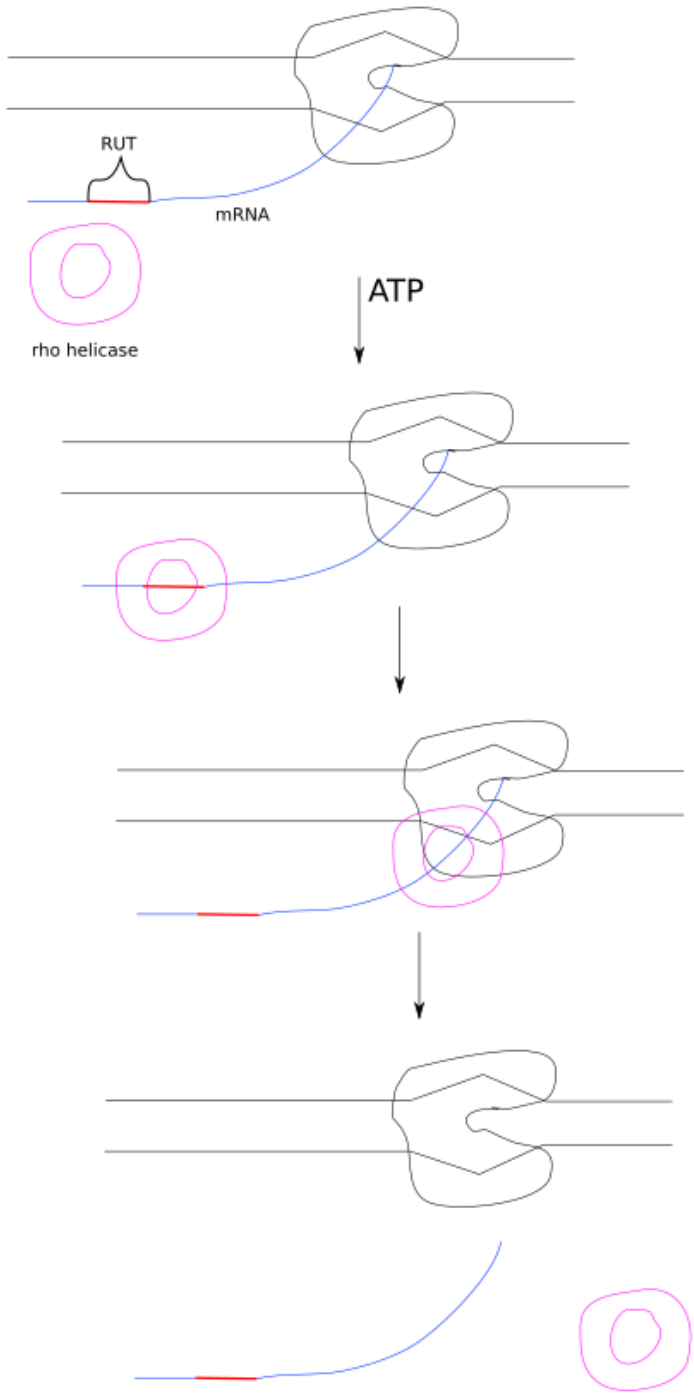
Focus 13: Transcription and Translation

Chapter 63: Transcription in Prokaryotes

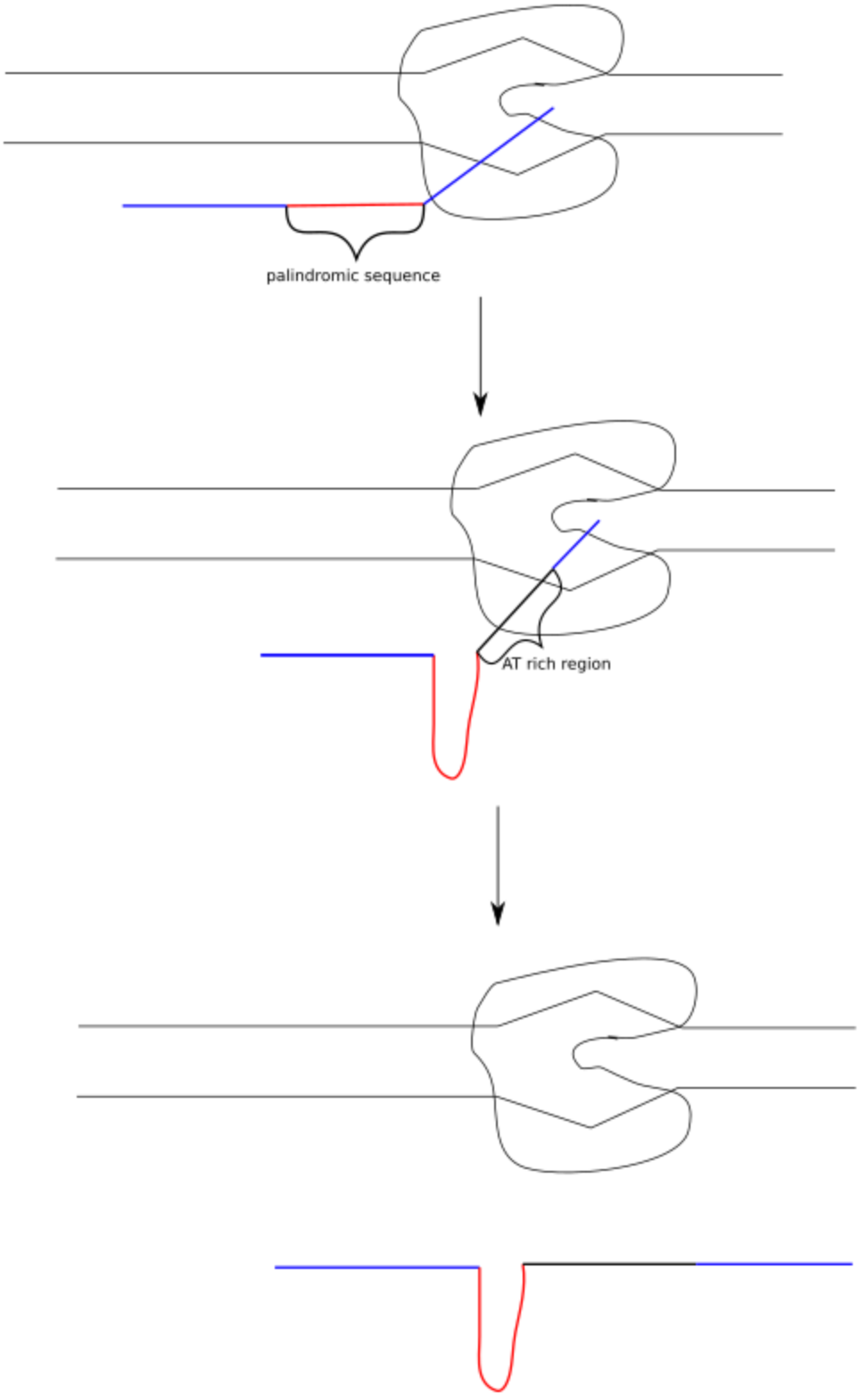
Transcription is how DNA is made into mRNA that can leave the nucleus and ultimately be made into proteins that have some important function. Because of how important this is, there has been a lot of research into how this process starts and the regulatory elements of transcription (details to come in the next focus). Here, we will discuss how transcription occurs in prokaryotes and give a general overview of the genomic landscape for protein-encoding genes.

All protein-encoding genes have promoter regions that contain the -10 and -35 regions. These regions are so named because they are 10 and 35 nucleotides away from the start of the protein-encoding portion of the gene. The -10 region is commonly TATAAT while the -35 region is typically TTGACA. For any gene to be transcribed, the RNA polymerase needs to bind the -10 and -35 regions of the promoter using its σ (sigma) subunit. Once the RNA polymerase docks onto the -10 and -35 regions, the α (alpha) subunit will bind the upstream promoter elements (UP elements) and transcription of the gene will initiate. The strength of RNA polymerase binding governs the strength of the promoter and how fast the gene can be transcribed. Stronger promoter = stronger RNA polymerase binding = faster gene transcription. When the RNA polymerase finally leaves the promoter region, the σ subunit is replaced with NusA and complementary nucleotides are added to a growing strand of mRNA in the elongation phase using the same mechanism as DNA polymerase III.

When the entire gene is transcribed, there are two ways that transcription can terminate: the ρ-dependent and ρ-independent pathways. In the ρ-dependent pathway, a special helicase called the ρ helicase will recognize and bind to a rut site on the mRNA transcript at the expense of ATP. This ρ helicase will slide down the mRNA until it reaches the RNA polymerase, where it separates the RNA polymerase from the mRNA. This is shown below:



The ρ-independent pathway is a bit more complicated, but not too bad either. It occurs in two stages, first a hairpin loop forms in the mRNA using a palindromic sequence such as GGGGGAACCCCC, then an A=T rich region occurs that allows the mRNA to dissociate from the RNA polymerase and the DNA. This is shown below:



Practice Questions:

1. Why do you suspect that the -10 region is AT rich?

2. What do you suspect would happen to the ρ-independent pathway if the hairpin loop region was AT rich as opposed to GC rich? Explain.
3. What do you suspect would happen to the ρ-independent pathway if the AT rich region was instead GC rich? Explain.
4. What kinds of genes do you suspect would have very strong promoters?

Answers:

1. The -10 region is AT rich because before transcription can happen, the double helix needs to be separated. It is easier to separate AT rich regions because the AT base pair is held together by 2 hydrogen bonds rather than 3 for the GC base pairs; therefore, it is logic that the regions near the start of the protein-encoding region would be AT rich.
2. If the hairpin loop was AT rich and not GC rich, then the hairpin would not form as easily and therefore termination would prove more difficult via the ρ-independent pathway.
3. If the AT rich region after the hairpin is instead GC rich, then the mRNA would not want to dissociate from the DNA (more hydrogen bonding holding it together) and therefore it would be more difficult to terminate transcription.
4. Genes that are absolutely necessary for survival are more likely to have strong promoters because these genes need to be continuously expressed for the cell to survive.

Chapter 64: RNA Processing in Eukaryotes for mRNA, rRNA, and tRNA

In this chapter, we will discuss how RNA in eukaryotes is processed to make mature RNA molecules that can be used to make proteins or in catalytic reactions such as in the ribosome. The three types of RNA that we will discuss in this chapter is mRNA, rRNA, and tRNA. The roles of these different types of RNAs are listed below:

mRNA = make functional proteins

rRNA = make ribosomes

tRNA = aid in translating mRNA to proteins

We will discuss each of these types of RNA in turn, starting with mRNA, since this is arguably the most important.

Broadly speaking, there are three things that need to occur before eukaryotic mRNA can be moved from the nucleus to the cytoplasm to be translated into proteins:

1. The introns in the mRNA transcript need to be removed
2. The 5' end needs to get capped with a guanine residue
3. The 3' end needs have several hundred adenine residues.

Introns are essentially "junk" RNA that need to be removed before functional proteins can be made from the mRNA molecule. THESE DO NOT EXIST IN PROKARYOTES! There are three different types of introns, each of which will be removed from the mRNA in different ways, these are the group 1, group 2, and spliceosomal introns. We will cover each of these in turn starting with group 1 introns. These introns, along with group 2, are self-splicing introns meaning that they DO NOT require energy to be removed. In other words, they will spontaneously remove themselves from the mRNA transcript. Group 1 introns leave as a linear RNA molecule while group 2 introns leave as a circular RNA molecule called a lariat. In both cases, two uridine residues are connected together and the intron is removed. The key difference between the two types of introns is that in group 1, the original nucleophile is a separate molecule, typically a G nucleotide while in group 2, an intramolecular attack occurs. That is why we get a linear molecule in group 1 but a circular one is group 2. Recall from orgo that intramolecular attacks ALWAYS give cyclic compounds at the end. The majority of introns, however, are spliceosomal introns and these DO cost energy to remove. Spliceosomal introns, as the name suggests, are removed by a spliceosome. This spliceosome is comprised of small nuclear RNA molecules (snRNA) and there are 5 that are important for removing specific introns: U1, U2, U4, U5, and U6. They skip U3 because they want to make your life harder. Spliceosomes will typically recognize a GU residue at the 5' end and an AG residue at the 3' end. U1 will recognize and bind to the 5' site and U2 will do the same for the 3' site at the expense of ATP. In this way, the U1 and U2 snRNA molecules define which region of the mRNA will be removed. When these two snRNA molecules bind the 3' and 5' intron sites, an A residue is activated and becomes a better nucleophile. This will allow the formation of a lariat down the line, once the spliceosome is primed to cut the intron out. Once the beginning and end site are defined by the U1 and U2 snRNA molecules, U4, U5, and U6 are recruited to the intron

at the expense of ATP to form the inactive spliceosome. Think of these as like power rangers that are coming together to become super powerful. Once the inactive spliceosome is made, it is activated using ATP and the intron is removed as a circular lariat. That was a lot of information, so let's break it down to a list:

1. U1 and U2 mark the start and end of the intron using ATP
2. U4, 5, and 6 combine with U1 and 2 to form the inactive spliceosome using ATP
3. The spliceosome is activated using ATP
4. Intron is removed as a lariat

Once all of the introns are removed from the mRNA, the mRNA will receive a "G cap", this cap is used to help protect the mRNA from nucleases that would break it down and destroy it before it can be translated into protein and, in eukaryotes, the cap will be bound to initiate translation. This G cap consists of a 7-methylguanosine residue and is added to the last 5' nucleotide using a 5',5'-triphosphate linkage.

Lastly, the polyA tail is added to the mRNA towards the end of transcription. This process is initiated when the RNA polymerase synthesizes RNA beyond a specific cleavage sequence. This specific cleavage sequence is bound by a polyadenylate polymerase (which will add the polyA tail) and a nuclease, which will remove the 3' end of the mRNA that was passed the cleavage sequence. Once the nuclease removes part of the 3' end, the polyadenylate polymerase adds the polyA tail. Badabing badaboom.

Now we will discuss how rRNA and tRNA is processed. These are modified in the exact same ways because they are synthesized together as one "bundle" RNA molecule and therefore, they will be discussed together. First, the bases in the RNA transcript are modified into "strange" bases like inosine (I), pseudouridine (Ψ), and dihydrouridine (D) to name a few. Then the RNA is cleaved and exonucleases are used to remove some of the non-coding regions of the transcript.

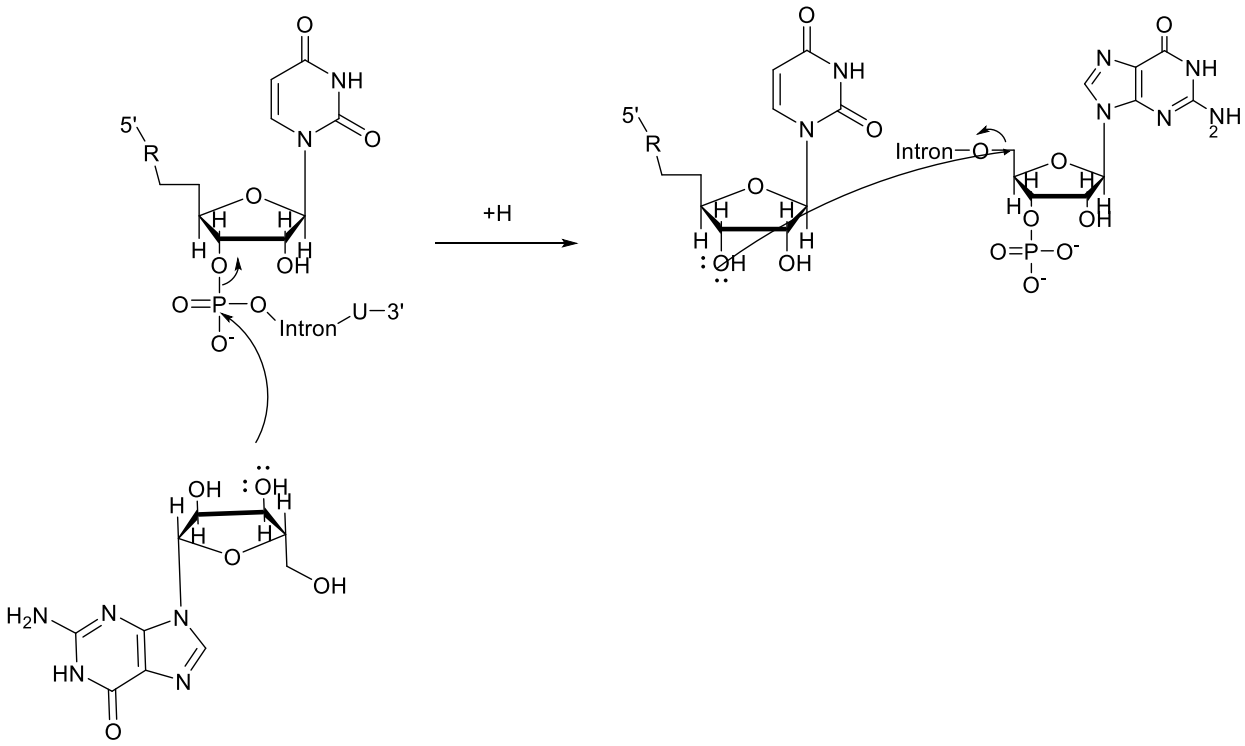
Practice Questions:

1. It is often observed that one gene can code for many proteins. Explain how this is possible considering what you know about the splicing mechanism.
2. One of the recurring themes involving nucleotide chemistry is that OH groups are nucleophilic. Understanding that concept, draw a plausible mechanism for group 1 introns.
3. One of the key concepts of spliceosomal introns is that assembling the spliceosome costs a lot of energy, but actually cutting the intron out requires none. Explain why this is so on the premise of thermodynamics and structure.

Answers:

1. One gene can code for many proteins because of alternative splicing patterns. Recognize that the recognition sites for U1 and U2 are extremely nonspecific (GU and AG

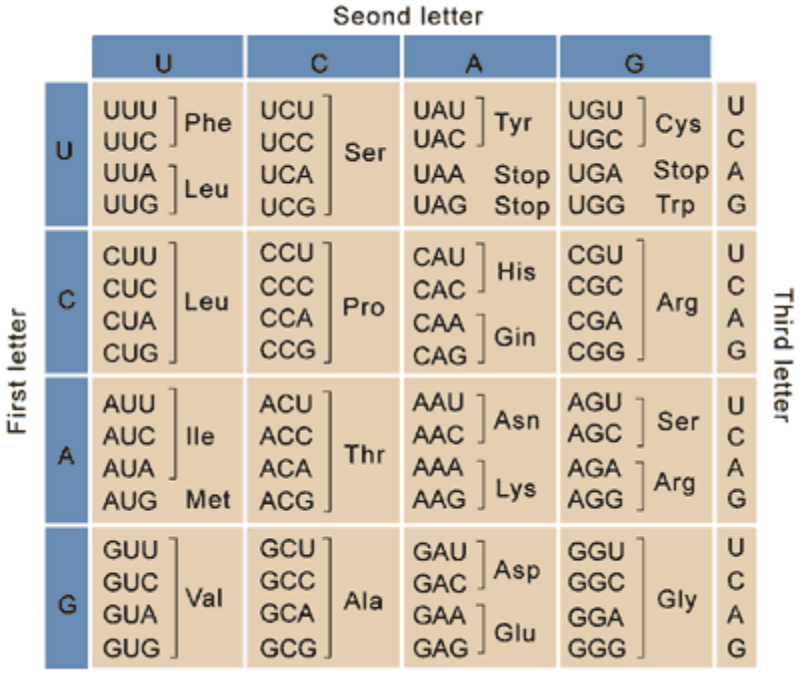
respectively). These can occur at any number of places in the gene and therefore depending on what the cell needs to express, different mature transcripts can arise from the same base mRNA molecule.

2. 

3. Cutting the intron requires no energy because when the intron is cut, one phosphodiester bond is replaced with another, thus there is no net change in phosphodiester bond and no energy is required.

Chapter 65: Genetic Code

Once the mature mRNA moves from the nucleus to the cytoplasm, ribosomes are used to translate the nucleotide code into amino acids in a process aptly named "translation." But, how does the cell know what sequence of nucleotides, called codons, corresponds to which amino acid? That is where tRNA comes into play, through an ingenuous system, life has evolved to make tRNA molecules that act as "adapters" and connect the nucleic acid world to the protein world. Each tRNA has an anticodon that will hydrogen bond with its complementary codon in the mRNA. This tRNA also carries with it a specific amino acid that it will add to the growing peptide chain. In this way, the anticodon allows the tRNA to recognize a specific sequence of RNA and the amino acid allows the cell to associate specific codons with specific amino acids in what is now known as the "genetic code". The genetic code is comprised of words called codons. Each word is made of three letters each and these letters can be any one of the four nucleotide bases (A, C, G, and U). The genetic code is shown below in all its glory:[15]



There are four special codons in the genetic code. These are the start and stop codons. The start codon is ALWAYS AUG, which codes for methionine and the stop codons are UAA, UAG, and UGA. I remember the stop codons by using "You are annoying, You are gone, and You go away."

When proteins are made from mRNA, the ribosome reads the transcript in the same direction that it is synthesized, in other words 5'→3'. For example, an mRNA molecule that reads:

5'-AUGCCGUGGUAG-3'

Would be read as

AUG – CCG – UGG – UAG

And therefore be translated as

Met-Pro-Trp

Something important to note is that many amino acids other than methionine and tryptophan are coded for by SEVERAL codons, rather than just one. This is a byproduct of the genetic code having codons with three letters. This is known as degeneracy. Degeneracy is super important because it allows for the genetic code to be more robust and resistant to mutation.

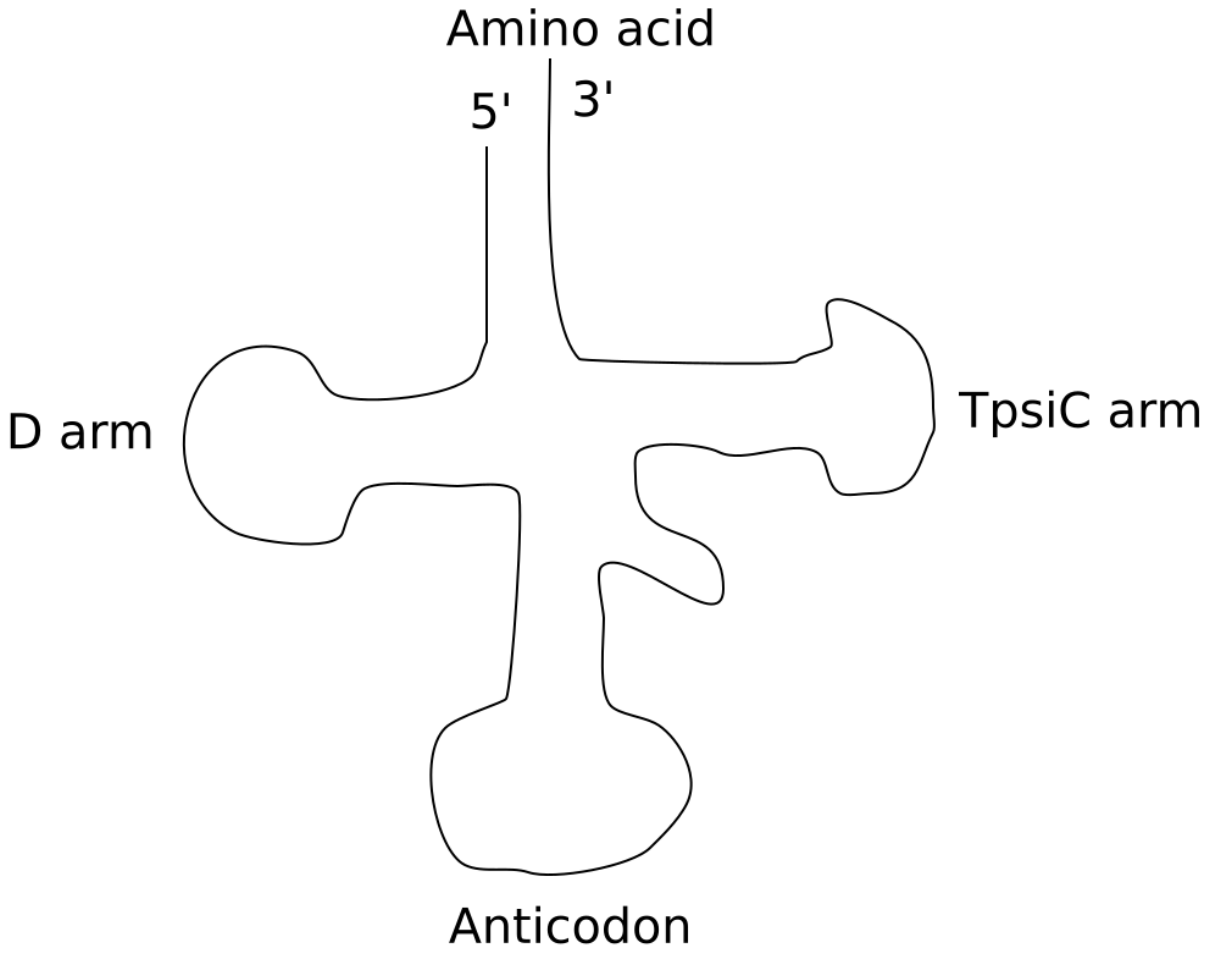
In addition to the genetic code being degenerate, the last base is often not important for determining the identity of the amino acid that is added. This has led many biologists to refer to the last letter as the "wobble base." The third base of the codon can form non-traditional base pairs with U, G, and I in the anticodon of the tRNA. U in the last base of the anticodon can pair with G or A, G can pair with C or U, and I can pair with U, C, and A.

Degeneracy and the wobble base allows for the genetic code to be resistant to mutation. Mutations in the first base typically cause one amino acid to substitute for a similar amino acid. One example of this is CUU (Leu) to AUU (Ile). Mutations in the third base often are "silent" mutations because they do not affect the protein structure at all (CUU to CUC for example are both Leu). It is only really mutations in the second base that cause the most damage, UUU to UCU for example would change phenylalanine to serine. This is a huge difference because phenylalanine is a large, bulky, nonpolar, and aromatic amino acid while serine is a small, polar, and aliphatic amino acid. Degeneracy and wobble bases also allow for codon bias to occur in organisms. This is simply the tendency for an organism to use specific codons for specific amino acids. For example, Leu has 6 codons that code for it, but not all 6 are made by the cell. E. coli, for example by chose to only use 4 of them and that is totally fine because their genome may never use two of them and thus it would be a waste to make those tRNAs (MAS**C**).

Now we can discuss how these activated tRNA molecules are made. Originally, tRNA molecules are not bound to amino acids. Instead, these tRNA molecules need to be activated using ATP and AARS to link the 3' end of the tRNA with the carboxyl group of the amino acid of interest. This process is referred to as "charging" the tRNAs and this occurs in two steps:

1. Adenylating the carboxyl group of the target amino acid
2. Using the 2' or 3' OH group of the 3' adenine residue (for class I and II AARS respectively) to attack the amino acid and kick off AMP.

In the case of class I AARS, the amino acid is eventually moved from the 2'OH to the 3'OH via a transesterification reaction. Thus, making activated tRNA molecules cost effectively 2 ATP equivalents. A general structure for a tRNA is shown below:



These class I and class II AARS have to be very specific for both the amino acid AND the tRNA that they are linking together. This is frequently thought of as the “second genetic code” because AARS chemistry is incredibly important for understanding how proteins are made in cells. These AARS generally recognize the tRNA based off the anticodon region and unique features present in the D arm and 5’ or 3’ regions. The very first tRNA that is used is a special tRNA, in bacteria, this is a formylmethionine tRNA and in eukaryotes it is a special methionine tRNA that is different from the regular methionine tRNA used after the first codon. The stop codons are also unique because the tRNA that recognizes the stop codons will not be charged with an amino acid, but will instead recruit specific release factors that will terminate translation of the protein. These release factors are specific for each type of stop codon.

Here are the key takeaways from the genetic code chapter:

1. tRNAs are used to associate specific genetic codes with specific amino acids
2. tRNAs must be charged using AARS at the expense of 2 ATP equivalents
3. The genetic code is degenerate and the third base is often not important for defining the identity of the amino acid added to the peptide chain
4. Not all codons are created equal and organisms often have preference for which codons they prefer. This is termed codon bias.

5. mRNA molecules are read by ribosomes in the 5’ to 3’ direction

Practice Questions:

1. Suppose this DNA sequence is going to be transcribed and subsequently translated into proteins. Predict the protein sequence that this DNA would encode:
   3’-TACCCGCGAGCGGTAATC-5’
2. How many ATP equivalents are consumed by charging tRNAs for a peptide that is 200 amino acids long?
3. Suppose you want to incorporate an unnatural amino acid at position 12 of a short polypeptide chain. Design an experiment that would allow you to synthesize a peptide that has an unnatural amino acid specifically in that one position, leaving the rest alone. Assume that you can synthesize any mRNA molecule that you wish and that this can be done without any cells but just ribosomes and the necessary amino acids, tRNAs, AARS, and release factors.

Answers:

1. DNA: 3’- TAC-CCG-CGA-GCG-GTA-ATC-5’
   mRNA: 5’-AUG-GGC-GCU-CGC-CAU-UAG-3’
   Protein: Met-Gly-Ala-Arg-His
2. For each tRNA that is charged, 2 ATP equivalents are consumed, therefore 2(200) = 400 ATP equivalents consumed to charge 200 tRNAs.
3. To specifically change the amino acid at position 12, we would first have to synthesize the mRNA molecule that codes the original protein of interest with edits to the 12^{th} codon. This codon needs to be changed such that it is a stop codon that is NOT the same as the actual stop codon for the peptide (i.e. if it is UAG that stops the peptide, this needs to be UAA or UGA). Then we need to design an AARS protein that will recognize the tRNA for that 12^{th} codon and the unnatural amino acid of interest. This will allow us to charge the tRNA at that position with the unnatural amino acid and it would be incorporated using the same system we discussed before. Then just let the ribosome and everything do their thing, but only supply the release factor for the actual stop codon, in this case the release factor for stop codon UAG.

Chapter 66: Translation in Prokaryotes

Translation is likely one of the most important processes in biological systems and it is highly sophisticated. Here, we will discuss the translation system present in prokaryotes, the system in eukaryotes is similar except the mRNA is circularized and the ribosomes are bigger and more "intelligent". First, we will start with a macroscopic overview of translation and then delve into the specifics of the process.

Translation occurs in three phases:

1. Initiation
2. Elongation
3. Termination

Sound familiar? It should! This is the same kind of phases that were present in DNA replication, PCR, AND transcription. But before we get into these phases, we first need to discuss the machine that drives this entire process, the ribosome.

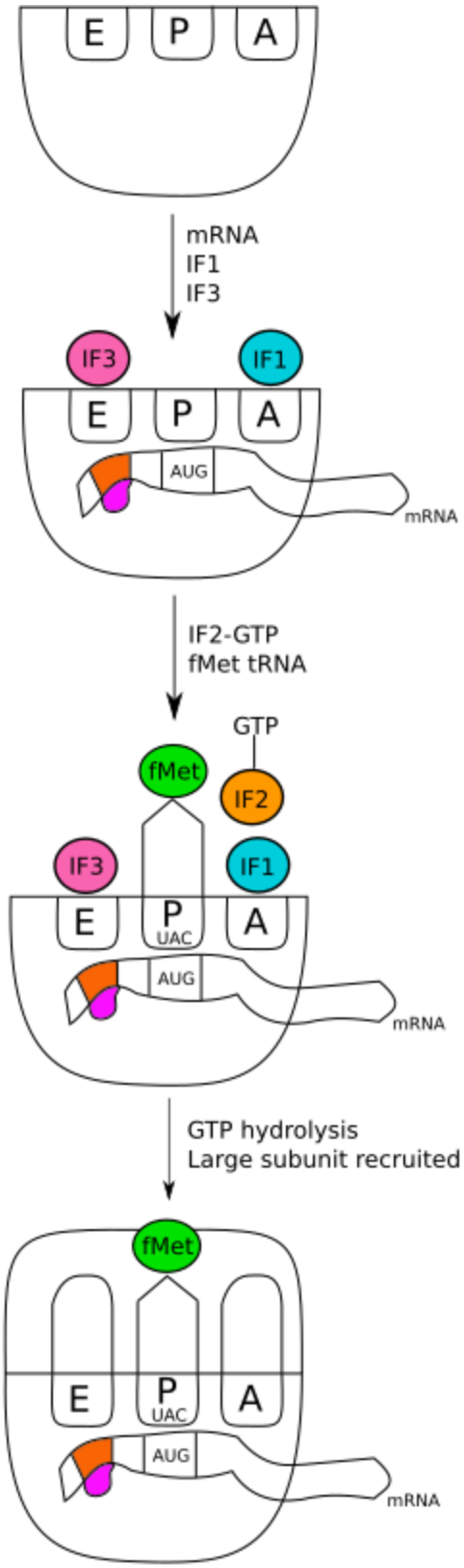
The ribosome is split up into two different subunits, the large and the small subunit. The small subunit contains the 16S rRNA that is used to recognize a special sequence on the mRNA that allows the AUG codon to be the first codon fed into the translation machinery. This sequence is the Shine Delgarno Sequence and it is a feature that is specific for prokaryotes. This sequence aligns the mRNA correctly so that the ribosome will start at AUG and read the transcript from there. It basically allows the ribosome to read the transcript from the beginning rather than the middle or the end. The large subunit on the other hand has the 5S and 23S rRNAs that allow for elongation to occur and catalyze the transpeptidation reaction. When the two subunits come together, the ribosome is fully functional and had three domains: A, P, and E domains. They have the following functions:

A = entering amino-acyl tRNAs go here

P = amino-acyl tRNAs are combined together to make one peptide chain

E = uncharged tRNAs exit here

Translation starts with initiation factors 1 and 3 (IF1 and IF3) binding to the small subunit of the ribosome, this recruits the mRNA. The mRNA is positioned using the Shine-Dalgarno sequence that binds the 16S rRNA. This is essentially the "priming" step for the initiation phase of translation. This is the priming step because now the small ribosomal subunit is positioned to start protein synthesis starting with the first formylmethionine amino acid. Now that the ribosome is primed to add the first amino acid, IF2-GTP and the first tRNA bind to the small ribosomal subunit. The GTP of IF2 is then hydrolyzed and this releases all of the initiation factors, which are replaced with the large subunit. Now the ribosome is fully functional and can be used to extend the protein chain because the first tRNA is in the P site, leaving the A site open for any incoming charged tRNAs. This is depicted visually below:


E
P
A
mRNA
IF1
IF3
IF3
IF1
E
P
A
AUG
mRNA
IF2-GTP
fMet tRNA
GTP
fMet
IF2
IF3
IF1
E
P
UAC
A
AUG
mRNA
GTP hydrolysis
Large subunit recruited
fMet
E
P
UAC
A
AUG
mRNA

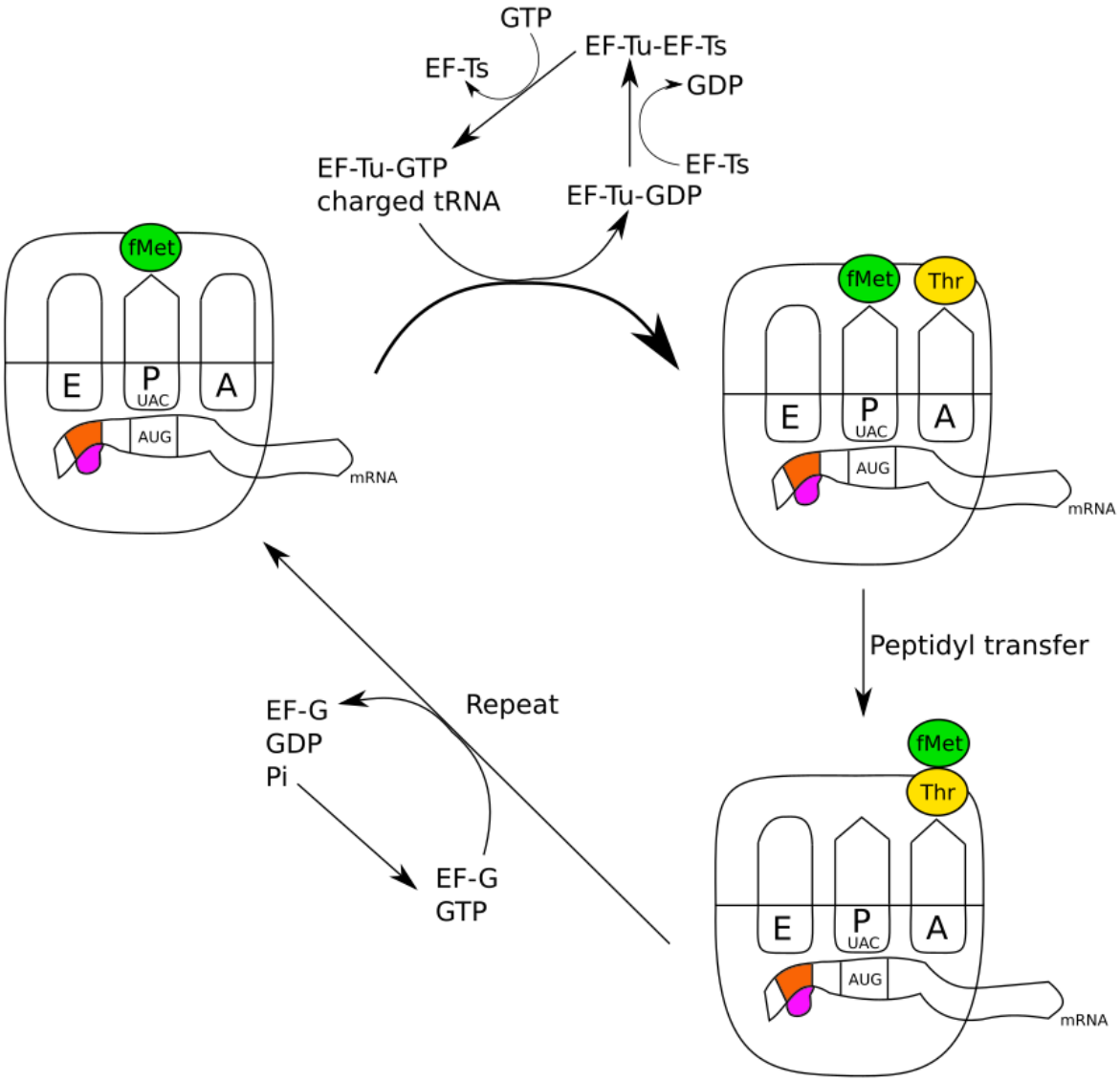
Once the ribosome is completely assembled and the first tRNA is in the P site, the elongation phase can proceed. Just as before, we have to go over the main goals that the elongation phase will set out to accomplish:

1. Charged tRNAs need to enter the A site
2. The amino acid from the A site needs to be added to the growing peptide chain
3. The tRNA in the P site needs to exit out of the E site and make way for the tRNA in the A site
4. This needs to be a cycle for all tRNAs in the future

Let's start with how new tRNAs will enter the A site. This is accomplished using the EF-Tu and EF-Ts cycle. Both of these are elongation factors that have GTP bound to them. The EF-Tu will bind the charged tRNA and drop it off at the A site at the expense of GTP. I remember this because EF-Tu brings the tRNA TO the A site. But now the EF-Tu doesn't have a GTP anymore, so EF-Ts will replace the GDP bonded to EF-Tu. EF-Ts is simply a placeholder for GTP, so that when GTP is present, it will dip the heck out of there and get replaced with GTP. In this way, the cell can recycle the active EF-Tu if there is GTP present (MAS**C**).

Once the charged tRNA is dropped off to the A site by EF-Tu, the tRNA in the P site will transfer its peptide chain to the amino acid on the charged tRNA via a peptidyl transferase reaction that is catalyzed by the 23S rRNA of the large subunit. Now, the peptide chain is attached to the tRNA in the A site and the P site tRNA is "uncharged" since it has no amino acid attached to it.

The uncharged tRNA in the P site needs to get released back into the cytoplasm to make room for the peptidyl tRNA in the A site. To do this, EF-G is used at the expense of GTP. I remember this because G stands for Gone. Now that the uncharged tRNA is gone, the peptidyl tRNA will move over to the P site and the cycle can continue. This is shown below:



Thus for every amino acid added to the peptide chain, 2 GTP is used to add it during elongation.

Once the stop codon is reached in the A site, release factors are recruited to the ribosome and these will accomplish three things:

1. Break off the mature peptide from the tRNA
2. Release tRNA and peptide from the ribosome
3. Dissociate ribosome into its subunits

To break off the mature peptide from the tRNA, the release factor will bind the A site and hydrolyze off the peptide from the tRNA in the P site. Then to release the tRNA, release factor, and to dissociate the ribosome, EF-G will hydrolyze its GTP and IF-3 will bind the small subunit to prevent reassembly of the ribosome.

Thus the total energetic cost to make a peptide of N amino acids long would be:

Initiation = 1 GTP

Elongation = 2 GTP times N-1

Termination = 1 GTP

Charging = 2 ATP times N

Thus the total ATP count would be:

$$Total\ Energy = ATP_{charging} + GTP_{initiation} + GTP_{elongation} + GTP_{termination}$$

$$Total\ Energy = 2N + 1 + 2(N - 1) + 1$$

$$Total\ Energy = 4N$$

Remember that the first amino acid is NOT subject to the same elongation steps as the rest of them, that is why the elongation GTP is represented as 2(N-1) since we do not use 2 GTP for the first amino acid that is added. We do still charge the first tRNA and therefore the charging step is simply multiplied by N.

Here are the key takeaways:

1. It costs 4N ATP equivalents to synthesize a peptide of N amino acids taking into consideration the charging, initiation, elongation, and termination steps of translation.
2. The small subunit of the ribosome recognizes and binds the mRNA while the large subunit catalyzes the peptidyl transferase reaction
3. tRNAs other than the first one always enter the A site then go to the P site and then leave through the E site.
4. GTP powers translation while ATP powers charging tRNAs

Practice Questions:

1. A student is attempting to do cell-free protein synthesis using ribosomes, tRNA, AARS, initiation factors, release factors, elongation factors, ATP, GTP, and mRNA. They find that over time, no more protein is synthesized, what do you suspect to be the reason for this? Assume that the only things in the reaction mixture are the things that were specified in the preface to the question. How could you solve this issue?
2. How many ATP equivalents would it take to synthesize a 1422 amino acid long protein?
3. In the initiation diagram, what is the magenta colored region? What is the orange colored region of the mRNA? What does the orange region do?
4. Suppose you took the gene that codes for human insulin and attempted to transform bacteria to allow them to make it. What problems could possible arise from this if you don't do it carefully? Suppose you found that the protein that the bacteria express is NOT

insulin, but instead contains extra amino acids. Explain this observation and determine how you can fix this issue.

Answers:

1. Over time, no protein is synthesized before there is no GTP or ATP producing system in the cell-free synthesis we described. Eventually all of the GTP and ATP will run out and there will be nothing that can be used to make EF-Tu-GTP or EG-GTP or to charge the tRNAs. Therefore protein synthesis will inevitably stop. The best way to fix this is to include glycolytic and Krebs cycle enzymes and substrates into the reaction mixture. These will allow the production of both ATP and GTP and can help sustain the reaction for longer.
2. 1422(4) = 5688 ATP equivalents
3. The magenta colored region is the 13S rRNA and the orange colored region is the Shine Dalgarno sequence that helps to line up the mRNA so that the first codon is the start codon (AUG).
4. Humans are multicellular eukaryotic organisms, therefore, the human insulin gene HAS introns that need to be removed, but he bacteria doesn't know that because in bacteria there are no introns. If you just ligated the human insulin gene into an expression plasmid and then transformed E. Coli, you wouldn't make insulin because the introns will NOT be spliced out of the mRNA. The end result would be extra amino acids being present in the insulin structure and as we know, structure informs function (MA**S**C). Thus the E. Coli that you transformed would be producing a lot of garbage protein that has no biological effect. To fix this, you would have to make a cDNA copy of the mRNA coding for human insulin using reverse transcriptase and then ligate that to an expression plasmid. This expression plasmid would then have the DNA that codes for the true human insulin and the bacteria would then produce the right functional protein. If protein synthesis is slow, this is likely due to differences in codon bias. E. Coli don't like using the same codons that humans do, so the expression plasmid may need to be optimized to avoid the E. Coli having to search for a charged tRNA that matches the codon in the mRNA it barely uses. This would be just trial and error (I think).

Focus 14: Gene Regulation and Epigenetics

Chapter 67: Histone Structure and Function

If the DNA in one cell was completely unraveled and straightened, it would be 6 feet long (2 meters), but obviously our cells are super duper tiny. That begs the question, how does all that DNA fit into that tiny cell? Well it turns out that we have these special helper proteins called histones and these histones are incredibly important for gene regulation and can promote cancer if they are not working properly. In this chapter, we will discuss the structure of histones, how they function to promote proper gene regulation, and how aberrant modification of histones can lead to cancer.

First, let's start by discussing the structure of a histone. Many histone proteins contain highly basic residues such as arginines and lysines and these side chains are pointed radially outward from the center of the histone. Histones are primarily cylindrical and this allows them to wrap around DNA. You can essentially think of the DNA as the string and the histone as the yo-yo. In eukaryotes, there are five main types of histones, H1, H2A, H2B, H3, and H4. These histones come together to form an octamer that the DNA will wrap around. These histone octamers will form long lines of octamers along which DNA can wrap around continuously. In between the histone octamers are the basic residues that can be covalently modified. These covalent modifications can be extremely extensive, though the most common forms of modification are methylation and acetylation.

Histones serve three key functions in eukaryotes:

1. It condenses DNA down so that it can fit inside the cell
2. It protects DNA from nucleases that would destroy it ordinarily
3. It can regulate gene expression by tightening or loosening its hold on DNA

The first two are relatively straightforward. These are the result of anything binding the DNA. The last function is a bit more nuanced and so we will cover it in-depth now. Histone modification plays a large role in determining gene expression, there are many different types of modifications, but we will just discuss the two major ones which are methylation and acetylation. These both occur on the amino-terminal ends of the amino acid residues that are facing radially outward from the center of the histone. When a histone is methylated, the amino group replaces its H's with CH_3 groups. These CH_3 groups make it more difficult for the amino group to give up its proton and become neutrally charged. Because of this, methylation generally increases the forces of attraction between the histone and the DNA it is bound to. This is because DNA has a phosphate backbone and therefore it is negatively charged. As we know opposite charges attract and therefore any modification that would make it harder to lose that positive charge would increase attraction between the histone and DNA. Because methylation increases the strength of the attraction between DNA and histones, it will generally decrease gene expression. This is because in order for a gene to be transcribed and translated, the DNA needs to be exposed and not bound to the histone. Acetylation has the opposite effect, it will make it EASIER to give up the hydrogen due to more resonance forms being available to the molecule when it is neutral. Recall carbonyl groups are ALWAYS electron-withdrawing groups and these groups ALWAYS increase acidity. Thus methylation will decrease gene expression while acetylation will increase

gene expression. This turns out to be incredibly important in a wide variety of cancers. Histone aceylation of H4 at lysine 16 is linked to cancer development in renal, breast, prostate, lung, and liver cancers.

Practice Questions:

1. Explain how histone acetylation or deacetylation leads to cancer development.
2. What would happen to histone-DNA binding affinity if the amino group was fully methylated i.e. $N(CH_3)_3$? Explain.

Answers:

1. Many genes are encoded by the human genome, some genes will activate DNA damage repair systems and others will break those proteins down. Cancer thrives when mutations run rampant and allow rapid cell growth, thus if a histone that represses DNA repair gets acetylated, then the cell can grow faster since there is no safeguard against DNA damage. Likewise, if the histone activated DNA repair, then deacetylation would downregulate that process and allow the cell to divide faster, thus leading to cancer.
2. If the amino group is fully methylated then it is not possible to remove any protons from it (it has none) and thus it would be permanently positively charged. Because of this permanent positive charge, the histone would be substantially more attracted to the DNA and therefore would repress the gene more strongly.

Chapter 68: Lac Operon

There are some genes that will always be expressed because they are necessary for survival in all situations, however, many genes only need to be turned on in specific circumstances (MAS**C**). In this chapter, we will discuss the Lac operon and how bacteria can control when lactose is broken down and how much these lactose-metabolizing enzymes are expressed depending on the situation.

When genes are regulated, they generally have an operon. An operon is a long sequence of DNA that has several elements:

1. The protein-encoding gene
2. The promoter
3. The operator or enhancer

Within the operator, there are sites for repressors to bind and within the enhancer there are sites for the activators to bind. The lac operon has an operator, so we will discuss operators explicitly in this chapter. Repressors are large proteins that will bind to the operator and stop expression of the protein-encoding gene. The affinity this repressor has for the operator can be altered by the presence of effector molecules. Co-repressors will increase the affinity while inducers will decrease it. The default state of the lac operon gene is "on", however, the repressor is bound by default to the operator. Thus the lac operon is said to be an inducible gene because an inducer can promote expression of the gene. In the next chapter, we will discuss the Trp operon and that is an example of a repressible gene, where the repressor is not bound and needs to become bound when a corepressor comes in.

Now that we have the definitions out of the way, we can delve into the specifics of the lac operon itself. There are three genes that the lac operon controls: lacZ, lacY, and lacA. LacZ is responsible for cleaving lactose into glucose and galactose and for making allolactose, the inducer for the lac operon. LacY is responsible for transporting lactose into the cell, it is effectively a lactose transport protein. LacA prevents the cell from getting toxins from the cell. The main ones that we will focus on in this chapter is lacZ and lacY because those are the more important ones. LacA is cool too I guess but it isn't the main event. These genes are all important for lactose metabolism, however, lactose is not the first choice for the cell to eat. Glucose is WAYYY more preferred from a cellular perspective and therefore, logically the lac genes should NOT be expressed when glucose is high. Likewise, there is no reason to make these proteins if lactose is not present, so overall the lac operon will be most active when glucose is low and lactose is high (MAS**C**). That is going to be the guiding principle behind how this system is regulated. Ordinarily, even in the presence of glucose, there will be some low level of gene expression. This means that if the cell senses lactose in the environment, it can move it into the cell and metabolize it using lacZ. LacZ, in addition to metabolizing lactose, will also isomerize it to allolactose. This allolactose will basically tell the cell "hey, buddy, we got a lot of lactose in the environment, we should eat it for energy!" In this way, it acts as an inducer and it loosens the grip that the repressor protein has on the lac operon and thus gene expression of the lac genes will increase. Separate from the lac genes, there is another region of the operon, lacI, which

codes for the repressor protein. This region is ALWAYS expressed, so even if the expression of lac genes is high, some repressor will still be made. In this way, the cell constantly makes repressors and stops the cell from wasting resources making lactose-metabolizing proteins when it does not need it (MAS**C**).

However, it is not just lactose levels that affects lac gene expression, but also glucose levels. Remember that the lac genes need to be expressed when lactose is high AND glucose is low. Now we will talk about how glucose levels affect lac expression. When glucose is low, the cell will make more cAMP, which signals to the cell that it needs energy and it needs energy badly. When cAMP levels are high enough, it forms a complex with CRP and this CRP-cAMP complex will bind near the lac promoter and stimulate expression to an extremely large extent (50x more than normal). This large increase in expression only occurs when the repressor has left the DNA aka lactose is present and the inducer is made. This makes sense because if lactose is not present, it doesn't matter if the cell needs energy because if lactose isn't there you can't use it for energy anyway (MAS**C**).

Here are the key takeaways from the lac operon chapter:

1. The lac operon regulates expression of lacZ, lacY, and lacA. It is most active when lactose is present and glucose is absent.
2. Lactose being present causes lacZ to metabolize lactose and to make allolactose, which will remove the repressor from the lac operon
3. Glucose being low makes cAMP, which forms a complex with CRP and stimulates expression of the lac genes if the repressor is not bound.

Practice Questions:

1. Suppose you had a mutation in lacZ that prevented it from making allolactose. Explain what effect this would have on the regulation of the lac operon in a situation of high lactose, low glucose; low lactose, high glucose; high lactose, high glucose; etc.
2. The lac repressor works by creating a loop between either the operator and the promoter or the operator and the lacZ-encoding gene. Explain how these two things prevent expression of the lac genes.
3. Suppose you have three cell cultures, one in which lactose was low and glucose was high, lactose was high and glucose was low, and the last in which both are high. Which culture would have the highest expression of lac genes, explain.

Answers:

1. If lacZ did not make allolactose, the repressor will stay bound to the operator regardless of glucose or lactose levels. This means that even when glucose is low and the CRP-cAMP complex is binding the promoter, it cannot stimulate expression of the lac genes to an appreciable extent.
2. If the DNA is looped around the promoter region, then the RNA polymerase cannot bind the promoter and therefore transcription cannot be initiated. If the DNA is looped around

the lacZ gene, then the RNA polymerase cannot access the lacZ gene and therefore allolactose cannot be made and the repressor will stay bound to the operator.

3. The cell culture that has low glucose and high lactose would have the highest expression of the lac genes. Under conditions of high lactose, the lacZ protein will make a lot of allolactose, which will prevent the repressor from binding the operator and thus expression should increase. Under conditions of low glucose, the cAMP will complex with CRP and bind the promoter region, stimulating the expression of lac genes.

Chapter 69: Trp Operon

The Trp operon is a bit different from the lac operon because tryptophan is incredibly important to the cell in most situations, whereas lactose is only important when glucose is not available and lactose is the only reliable source of energy. Because of this key difference, the repressor is NOT bound to the Trp operator by default, instead it needs to be activated by a corepressor before it can be repress expression of the tryptophan synthesis genes. Remember, the central goal of the trp genes are to make tryptophan, therefore the genes that the trp operon controls are generally active UNLESS tryptophan levels are high, in which case the tryptophan will act as a corepressor itself and halt its own synthesis (MAS**C**). This way, the cell doesn't make tryptophan synthesis enzymes when it already has an excess of tryptophan.

There are two ways that the trp operon is regulated: attenuation sequence and repression. Within the trp operon, there is a leader sequence that is always transcribed and has an attenuator sequence at the beginning. In bacteria, translation and transcription occur at the same time, this allows for bacteria to have this unique attenuator method of regulating gene expression. This attenuator has two consecutive UGG tryptophan codons. These two codons are strategically placed one after the other so that when the mRNA is translated into protein, the tryptophan tRNAs need to be used consecutively. Thus, if tryptophan is in low supply, it is difficult to have many charged tryptophan tRNAs and to have them in such quick succession, therefore the ribosome will stall temporarily at that attenuator sequence and the entire gene will be translated into protein. If tryptophan is in high concentration, then translation will be rapid and the ribosome will bump into a 3' hairpin structure in the mRNA and cause ρ-independent termination. Thus under conditions of high tryptophan, the attenuator will not stall the ribosome and instead the ribosome will cause ρ-independent termination of transcription and thus all of the tryptophan synthesis enzymes will not be made. In the case of repression, tryptophan will act as a corepressor and cause the repressor protein to bind the operator and shut off transcription and translation of tryptophan synthesizing enzymes.

Here are the key takeaways:

1. The trp operon controls expression of tryptophan synthesizing enzymes.
2. The trp genes will be most active when tryptophan levels are low and will be turned off when tryptophan is in excess
3. The trp operon is regulated by a repressor protein and by an attenuator sequence
4. When tryptophan is low, the ribosome is stalled and the hairpin loop will go away causing the entire trp operon to be expressed. When tryptophan is high, the ribosome is not stalled and will crash into the hairpin structure, causing ρ-independent termination.

Practice Questions:

1. Suppose that the tryptophan repressor is mutated such that it has a higher affinity for tryptophan. Explain what effect this would have on the trp operon.
2. Suppose that the hairpin region of the attenuator is mutated such that it has more AT bases rather than GC, would the attenuator work better or worse? Explain.
3. Would you describe the trp operon as repressible or inducible? Explain.

4. Compare and contrast the trp and lac operons.

Answers:

1. If the tryptophan repressor has a higher affinity for tryptophan, then it would therefore have a higher affinity for DNA because the tryptophan is the corepressor, thus the trp operon would be prematurely turned off in this scenario.
2. If the hairpin region has more AT bases than GC then the hairpin would not be as thermodynamically stable and thus would collapse more easily. The end result is that it is more possible that the ribosome would not crash into it and thus not promote ρ-independent termination. Overall the trp operon would be more active than usual because the attenuator would be less effective.
3. The trp operon is a repressible operon because it does not have its repressor bound by default and instead needs to have a corepressor molecule to repress the expression of the trp genes.
4. The trp operon is a repressible operon that has its default state being on (default here means when no repressor is bound). The trp operon encodes genes that are necessary for cell survival most of the time. The trp operon has an attenuator to prevent gene expression and uses a repressor protein. The lac operon on the other hand is an inducible operon, encodes genes that are not usually necessary for survival, and does not use an attenuator to prevent gene expression. The lac operon's unique feature is the CRP-cAMP complex that will stimulate gene expression under low glucose conditions. Both operons use a repressor protein and both of their default states are on if the repressor is not bound.

References

1. https://www.rcsb.org/structure/2BVC (Accessed January 6, 2021)
2. https://www.rcsb.org/structure/6VYB (Accessed January 6, 2021)
3. https://www.researchgate.net/publication/279498023_Biotechnology_A_problem_approach (Accessed January 6, 2021)
4. https://www.amherst.edu/system/files/media/0438/Protein%252520Structure%252520F11.pdf (Accessed January 6, 2021)
5. Nelson, DL; Cox, MM Lehninger Principles of Biochemistry 7th Edition W.H. Freeman and Company Macmillan Learning New York 2017
6. https://biologydictionary.net/cell-membrane/ (Accessed January 6, 2021)
7. https://academic.oup.com/chemse/article/27/5/453/309085 (Accessed January 6, 2021)
8. https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_For_All_(Ahern_Rajagopal_and_Tan)/08%3A_Basic_Techniques/8.07%3A_Polymerase_Chain_Reaction_(PCR) (Accessed January 6, 2021)
9. https://en.wikipedia.org/wiki/Reverse_transcription_polymerase_chain_reaction (Accessed January 6, 2021)
10. https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/restriction-enzymes-dna-ligase (Accessed January 6, 2021)
11. https://www.researchgate.net/publication/233534974_Contamination_of_barley_seeds_with_Fusarium_species_and_their_toxins_in_Spain_An_integrated_approach (Accessed January 6, 2021)
12. Doudna, J; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014** 346 1077-1086.
13. He, Jun, Lijun Hu, Xiaojun Huang, Chenran Wang, Zhimin Zhang, Ying Wang, Dongmei Zhang, and Wencai Ye. "Potential of Coronavirus 3C-like Protease Inhibitors for the Development of New Anti-SARS-CoV-2 Drugs: Insights from Structures of Protease and Inhibitors." *International Journal of Antimicrobial Agents* 56.2 (**2020**): 106055. Print.
14. https://www.nejm.org/doi/full/10.1056/NEJMoa2007764 (Accessed January 6, 2021)
15. https://bio.libretexts.org/Courses/Community_College_of_Vermont/Human_Biology_(Gabor_Gyurkovics)/15%3A_Cell_Reproduction/15.05%3A_Genetic_Code (Accessed January 6, 2021)
16.