BIOCHEMISTRY FREE & EASY

Kevin Ahern & Indira Rajagopal
Oregon State University

LibreTexts™

Oregon State University
Biochemistry Free & Easy

Kevin Ahern & Indira Rajagopal

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

TABLE OF CONTENTS

5: Flow of Genetic Information

6: Metabolism I - Oxidative/Reductive Processes

7: Metabolism II

8: Signaling

9: Techniques

10: Putting It All Together

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Cells, Water, and Buffers

In this chapter we introduce the subject and talk about the scientific aspects of the most important and most abundant liquid on the face of Earth - water.

1.1: Introduction

Biochemistry is a relatively young science, but the rate of its expansion has been truly impressive. This rapid pace of discoveries, which shows no signs of slowing, is reflected in the steady increase in the size of biochemistry text books, most of which top a thousand pages and undergo revisions every couple of years to incorporate new findings. These full-scale texts offer an enormous amount of information and serve as invaluable resources. Those who need the greater level of detail and broader coverage that these books provide have many choices available in any good bookstore.

As certified (some might say, certifiable) biochemistry nerds and unrepentant lovers of corny jokes, we firmly believe that students can have fun while learning the subject. Toward this end, we have sprinkled each chapter with rhymes and songs that we hope will have you learning biochemistry happily. The format of the book as available for the iPad, allows readers to click on figures to enlarge them, watch video lectures relevant to each topic, listen to the songs in the book, like the one above, and link out to the internet to find more information simply by clicking on any term. If you are using a PDF version of this book, you will still be able to use the links to the video lectures. Also, though you cannot listen to the songs by clicking on them in the PDF version, you can download them HERE. We hope you find these features useful and that they help you learn biochemistry.

1.2: Cells- The Bio of Biochemistry

Biochemistry happens inside organisms and possibly, the most obvious thing about living organisms is their astounding diversity. If living things are so varied, it seems reasonable to ask whether their chemistry is, too. The invention of the microscope opened up a whole new world of microscopic organisms while also providing the first clue that living organisms had something in common-all living things are made up of cells. Some cells are “lone rangers” in the form of unicellular entities, such as bacteria and some protists. Cells are also the building blocks of more complex organisms (like humans, wombats, and turnips).

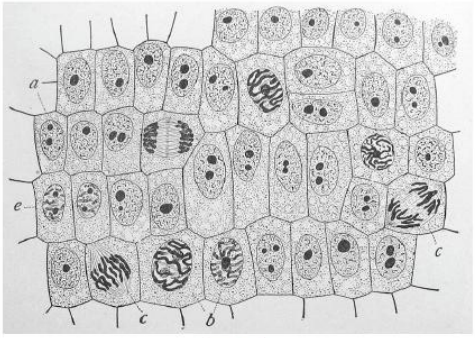

Figure 1.2.1: *Onion cells.*

As increasingly powerful microscopes became available, it was possible to discern that all cells fell into one of two types- those with a nucleus and other sub-cellular compartments like mitochondria and lysosomes, termed eukaryotes, and those that lack such internal compartmentation, the prokaryotes. Some eukaryotes, such as yeast, are unicellular, while others, including animals and plants are multicellular. The prokaryotes may be divided into two very broad categories, the bacteria and the archaeans.

One can find living cells almost everywhere on earth - in thermal vents on the ocean floor, on the surface of your tongue and even in the frozen wastes of the Antarctic. Some cells may have even survived over two years on the moon. Yet, despite their diversity of appearance, habitat, and genetic composition, cells are not as different from each other as you might expect. At the biochemical level, it turns out that all cells are more alike than they are different. A great simplifying feature of biochemistry is that many of the reactions are universal, occurring in all cells. For example, most bacteria process glucose in the same 10-step pathway that plant, animal, and fungal cells do. The genetic code that specifies the amino acids encoded by a nucleic acid sequence is interpreted almost identically by all living cells, as well. Thus, the biochemical spectrum of life is (mercifully) not nearly as broad or as complicated as he evolutionary spectrum. Where cells differ significantly in processes/reactions, we will note these differences.

Figure 1.2.2: *Extremophiles. (CC-SA-BY 3.0; Wikipedia)*

1.3: Water, Water Everywhere

Click here for Kevin's Introductory Lecture on Youtube

Vital for life, water is by far the most abundant component of every cell. To understand life, we must, therefore, understand the basics of water, because everything that happens in cells, even reactions buried deep inside enzymes, away from water, is influenced by water's chemistry.

Figure 1.3.1: *Water.*

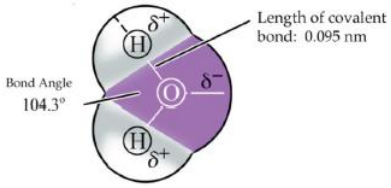

Figure 1.3.2: *Water molecule.*

We start with simple properties. The molecule has a sort of wide 'V' shape (the H-O-H angle is 104°) with uneven sharing of electrons between the oxygen and the hydrogens. The hydrogens, as a result, are described as having a partial positive charge and the oxygen has a partial negative charge. These tiny partial charges allow the formation of what are described as hydrogen bonds, which occur when the partial positive charge of one atom is attracted to the partial negative of another. In water, that means the hydrogen of one water molecule will be attracted to the oxygen of another. Hydrogen bonds play essential roles in proteins, DNA, and RNA, as well, as we shall see.

1.4: Buffers Keep the Cellular Environment Stable

Water can ionize to a slight extent (10^{-7} M is about 6 molecules per 100 million of pure water) to form H^+ (proton) and OH^- (hydroxide). We measure the proton concentration of a solution with pH, which we define as the negative log of the proton concentration.

$$pH = -\log[H^+] \tag{1.4.1}$$

If the proton concentration, $[H^+] = 10^{-7}$ M, then the pH is 7. We could just as easily measure the hydroxide concentration with the pOH by the parallel equation, $pOH = -\log[OH^-]$ In pure water, dissociation of a proton from it creates a hydroxide, so the pOH of pure water is 7, as well. This also means that

$$pH + pOH = 14 \tag{1.4.2}$$

Now, because protons and hydroxides can combine to form water, a large amount of one will cause there to be a small amount of the other. Why is this the case? In simple terms, if I dump 0.1 moles of H^+ into a pure water solution, the high proton concentration will react with the relatively small amount of hydroxides to create water, thus reducing hydroxides. Similarly, if I dump excess hydroxide (as $NaOH$, for example) into pure water, the proton concentration falls for the same reason.

Weak Acid pK_a values

Name	HA	A^-	pK_a
Pyruvic acid	$CH_3COCOOH$	$CH_3C(=O)COO^-$	2.50
Formic acid	$HCOOH$	$HCOO^-$	3.75
Lactic acid	$CH_3CHOHCOOH$	$CH_3CH(OH)COO^-$	3.86
Acetic acid	CH_3COOH	CH_3COO^-	4.76
Oxalic acid (1)	$HOOC—COOH$	$HOOC—COO^-$	1.23
Oxalic acid (2)	$HOOC—COO^-$	$^-OOC—COO^-$	4.19
Carbonic acid (1)	H_2CO_3	HCO_3^-	6.37
Carbonic acid (2)	HCO_3^-	CO_3^{2-}	10.20
Malonic acid (1)	$HOOC—CH_2—COOH$	$HOOC—CH_2—COO^-$	2.83
Malonic acid (2)	$HOOC—CH_2—COO^-$	$^-OOC—CH_2—COO^-$	5.69
Malic acid (1)	$HOOC—CH_2—CHOH—COOH$	$HOOC—CH_2—CHOH—COO^-$	3.40
Malic acid (2)	$HOOC—CH_2—CHOH—COO^-$	$^-OOC—CH_2—CHOH—COO^-$	5.26
Succinic acid (1)	$HOOC—CH_2—CH_2—COOH$	$HOOC—CH_2—CH_2—COO^-$	4.21
Succinic acid (2)	$HOOC—CH_2—CH_2—COO^-$	$^-OOC—CH_2—CH_2—COO^-$	5.63
Phosphoric acid (1)	H_3PO_4	$H_2PO^-_4$	2.14
Phosphoric acid (2)	$H_2PO_4^-$	HPO_4^{2-}	7.20
Phosphoric acid (3)	HPO_4^{2-}	PO_4^{3-}	12.40

Figure 1.4.1: *Chart of pKa values for weak acids.*

Chemists use the term “acid” to refer to a substance which has protons that can dissociate (come off) when dissolved in water. They use the term “base” to refer to a substance that can absorb protons when dissolved in water. Both acids and bases come in strong and weak forms. Strong acids, such as HCl, dissociate completely in water. If we add 0.1 moles of HCl to a solution to make a liter, it will have 0.1 moles of H^+ and 0.1 moles of Cl-. There will be no remaining HCl when this happens. A strong base like NaOH also dissociates completely into Na^+ and OH^-.

Weak acids and bases differ from their strong counterparts. When you put one mole of acetic acid (HAc) into pure water, only about 4 in 1000 HAc molecules dissociate into H^+ and Ac^-. Thus, if I start with 1000 HAc, I will end up with 996 HAc and 4 each of H^+ and Ac^-.

Clearly, weak acids are very different from strong acids. Weak bases behave similarly, except that they accept protons, rather than donate them.

You may wonder why we care about weak acids.You may never have thought much of weak acids when you were in General Chemistry. Your instructor described them as buffers and you probably dutifully memorized the fact that “buffers are substances that resist change in pH” without really learning what it meant. We will not allow that to happen here.

Weak acids are critical for life because their affinity for protons causes them to behave like a UPS. We're not referring to the UPS that is the United Parcel Service®, but instead, to the encased battery backup systems for computers called Uninterruptible Power Supplies that kick on to keep a computer running during a power failure. Your laptop battery is a UPS, for example. We can think of weak acids as **Uninterruptible Proton Suppliers** within certain pH ranges, providing (or absorbing) protons as needed.Weak acids thus help to keep the H+ concentration (and thus the pH) of the solution they are in relatively constant.

Consider the acetic acid (acetate) system. Here is what happens when HAc dissociates

$$HAc \rightleftharpoons H^+ + Ac^- \tag{1.4.3}$$

As noted, about 4 in 1000 HAc molecules come apart. However, what if one started adding hydroxyl ions (by adding a strong base like $NaOH$) to the solution with the HAc in it? As the added OH ions reacted with the H^+ ions to make water, the concentration of H^+ ions would go down and the pH would go up. However, in contrast to the situation with a solution of pure water, there is a backup source of H^+ available in the form of HAc. Here is where the UPS function kicks in. As protons are taken away by the added hydroxyl ions (making water), they are partly replaced by protons from the HAc. This is why a weak acid is a buffer. It resists changes in pH by releasing protons to compensate for those "used up" in reacting with the hydroxyl ions.

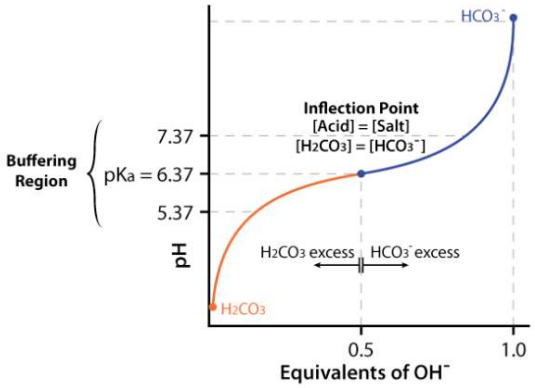

Figure 1.4.2: *An example buffer system - carbonic acid/bicarbonate. Note that near pH 6.73, addition of OH^- results in only small pH changes. This is what a buffer does - resists changes in pH over certain ranges.*

1.5: Henderson-Hasselbalch Approximation

It is useful to be able to predict the response of the HAc system to changes in H^+ concentration. The Henderson-Hasselbalch equation defines the relationship between pH and the ratio of Ac^- and HAc. It is as follows

$$pH = pK_a + \log\left(\frac{[Ac^-]}{[HAc]}\right) \tag{1.5.1}$$

This simple equation defines the relationship between the pH of a solution and the ratio of Ac- and HAc in it. The new term, called the pKa, is defined as

$$\mathrm{pKa} = -\log \mathrm{Ka} \tag{1.5.2}$$

just as

$$\mathrm{pH} = -\log[\mathrm{H}^+] \tag{1.5.3}$$

The Ka is the acid dissociation constant and is a measure of the strength of an acid. For a general acid, HA, which dissociates as

$$\mathrm{HA} \rightleftharpoons \mathrm{H}^+ + \mathrm{A}^- \tag{1.5.4}$$

$$\mathrm{Ka} = [\mathrm{H}^+][\mathrm{A}^-]/[\mathrm{HA}] \tag{1.5.5}$$

Thus, the stronger the acid, the more protons that will dissociate from it and the larger the value its Ka will have. Large values of Ka translate to lower values of pKa. As a result, the lower the pKa value is for a given acid, the stronger the acid is.

Please note that pKais a constant for a given acid. The pKafor acetic acid is 4.76. By comparison, the pKafor formic acid is 3.75. Formic acid is therefore a stronger acid than acetic acid. A stronger acid will have more protons dissociated at a given pH than a weaker acid.

Now, how does this translate into stabilizing pH? The previous figure shows a titration curve. In this curve, the titration begins with the conditions at the lower left (very low pH). At a this pH, the HAc form predominates, but as more and more OH- is added (moving to the right), the pH goes up, the amount of Ac- goes up and (correspondingly), the amount of HAc goes down. Notice that the curve “flattens” near the pKa (4.76).

What this tells us is that the pH is not changing much (not going up as fast) as it did earlier when the same amount of hydroxide was added. The system is resisting a change in pH (not stopping the change, but slowing it) in the region of about one pH unit above and one pH unit below the pKa. Thus, the buffering region of the acetic acid/acetate buffer is from about 3.76 to 5.76. It is maximally strong at a pH of 4.76.

Now it starts to become apparent how the buffer works. HA can donate protons when extras are needed (such as when OH^- is added to the solution. Similarly, A- can accept protons when extra H^+ are added to the solution (adding HCl, for example). The maximum ability to donate or accept protons comes when

$$[\mathrm{A}^-] = [\mathrm{HA}] \tag{1.5.6}$$

To understand how well a buffer protects against changes in pH, consider the effect of adding .01 moles of HCl to 1.0 liter of pure water (no volume change) at pH 7, compared to adding it to 1.0 liter of a 1M acetate buffer at pH 4.76. Since HCl completely dissociates, in 0.01M (10^{-2} M) HCl you will have 0.01M H^+. For the pure water, the pH drops from 7.0 down to 2.0 (pH = -log(0.01M)).

By contrast, the acetate buffer’s pH is 4.74. Thus, the pure water solution sees its pH fall from 7 to 2 (5 pH units), whereas the buffered solution saw its pH drop from 4.76 to 4.74 (0.02 pH units). Clearly, the buffer minimizes the impact of the added protons compared to the pure water.

It is important to note that buffers have capacities limited by their concentration. Let’s imagine that in the previous paragraph, we had added the 0.01 moles HCl to an acetate buffer that had a concentration of 0.01M and equal amounts of Ac- and HAc. When we try to do the math in parallel to the previous calculation, we see that there are 0.01M protons, but only 0.005M A- to absorb them. We could imagine that 0.005M of the protons would be absorbed, but that would still leave 0.005M of protons unbuffered. Thus, the pH of this solution would be approximately

$$\mathrm{pH} = -\log 0.005\mathrm{M} = 2.30 \tag{1.5.7}$$

Exceeding buffering capacity dropped the pH significantly compared to adding the same amount of protons to a 1M acetate buffer. Consequently, when considering buffers, it is important to recognize that their concentration sets their limits. Another limit is the pH range in which one hopes to control proton concentration.

Now, what happens if a molecule has two (or more) ionizable groups? It turns out, not surprisingly, that each group will have its own pKa and, as a consequence, will tend to ionize at different pH values. The figure above right shows the titration curve for a simple amino acid, alanine. Note that instead of a single flattening of the curve, as was seen for acetic acid, alanine displays two such regions. These are individual buffering regions,each centered on the respective pKa values for the carboxyl group and the amino group.

If we think about alanine, it can have three possible charges: +1 (alpha carboxyl group and alpha amino group each has a proton), 0 (alpha carboxyl group missing a proton and alpha amino group has a proton) and -1 (alpha carboxyl group and alpha amino group each lacking a proton).

How does one predict the charge at a given pH for an amino acid? A good rule of thumb for estimating charge is that if the pH is more than one unit below the pKa for a group (carboxyl or amino), the proton is on. If the pH is more than one unit above the pKa for the group, the proton is off. If the pH is NOT more than one or less than one pH unit from the pKa, this simple assumption will not work.

Further, it is important to recognize that these rules of thumb are estimates only. The pI (pH at which the charge of a molecule is zero) is an exact value calculated as the average of the two pKa values on either side of the zero region. It is calculated at the average of the two pKa values around the point where the charge of the molecule is zero.

CHAPTER OVERVIEW

2: Energy

Living organisms are made up of cells, and cells contain many biochemical components such as proteins, lipids, and carbohydrates. But, living cells are not random collections of these molecules. They are extraordinarily organized or "ordered". By contrast, in the nonliving world, there is a universal tendency to increasing disorder. Maintaining and creating order in cells takes the input of energy. Without energy, life is not possible. It is therefore important that we consider energy first in our attempt to understand biochemistry. Where does energy come from? Photosynthetic organisms can capture energy from the sun, converting it to chemical forms usable by cells. Heterotrophic organisms like ourselves get our energy from the food we eat. How do we extract the energy from the food we eat?

2.1: Oxidative Energy

Living organisms are made up of cells, and cells contain many biochemical components such as proteins, lipids, and carbohydrates. But, living cells are not random collections of these molecules. They are extraordinarily organized or "ordered". By contrast, in the nonliving world, there is a universal tendency to increasing disorder. Maintaining and creating order in cells takes the input of energy.

Without energy, life is not possible. It is therefore important that we consider energy first in our attempt to understand biochemistry. Where does energy come from? Photosynthetic organisms can capture energy from the sun, converting it to chemical forms usable by cells. Heterotrophic organisms like ourselves get our energy from the food we eat. How do we extract the energy from the food we eat?

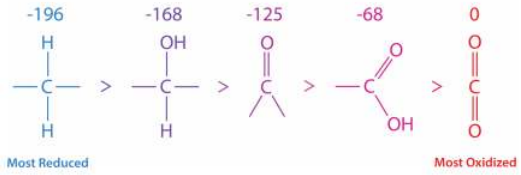

Figure 2.2.1: *Free energies of oxidation in KJ/mol*

In this series, the most reduced form of carbon is on the left. The energy of oxidation of each form is shown above it. Fatty acids are more reduced overall than sugars. This can also be seen by their formulas.

Palmitic acid = $C_{16}H_{34}O_2$

Glucose = $C_6H_{12}O_6$

Palmitic acid only contains two oxygens per sixteen carbons, whereas glucose has six oxygen atoms per six carbons.

Consequently, when palmitic acid is fully oxidized, it generates more ATP per carbon (128/16) than glucose (38/6). It is because of this that we use fat as our primary energy storage material.

2.2: Oxidation vs Reduction in Metabolism

Biochemical processes that break things down from larger to smaller are called catabolic processes. Catabolic processes are often oxidative in nature and energy releasing. Some, but not all of that energy is captured as ATP. If not all of the energy is captured as ATP, what happens to the rest of it? The answer is simple. It is released as heat and it is for this reason that we get hot when we exercise. By contrast, synthesizing large molecules from smaller ones (for example, making proteins from amino acids) is referred to as anabolism. Anabolic processes are often reductive in nature and require energy input. By themselves, they would not occur, as they are reversing oxidation and decreasing entropy (making many small things into a larger one). To overcome this energy 'barrier', cells must expend energy. For example, if one wishes to reduce CO2 to carbohydrate, energy must be used to do so. Plants do this during the dark reactions of photosynthesis. The energy source for the reduction is ultimately the sun. The electrons for the reduction ultimately come from water, and the CO2 comes from the atmosphere and gets incorporated into a sugar.

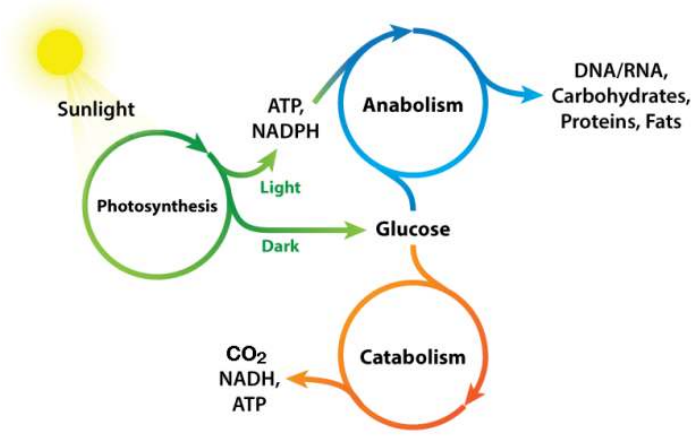

Figure 2.3.1: *Biological energy movement.*

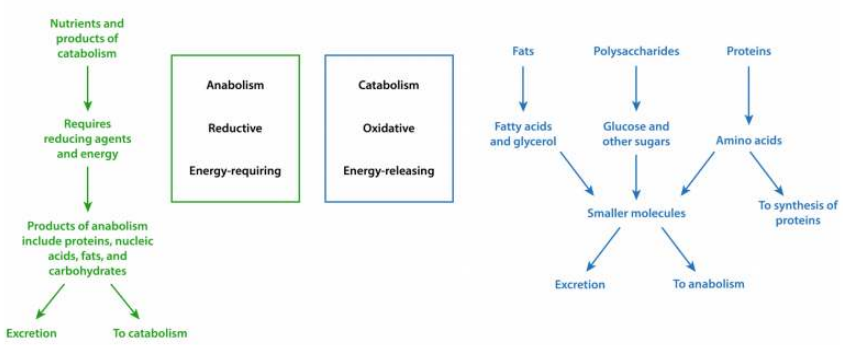

Figure 2.3.2: *Anabolic vs. catabolic processes.*

2.3: Energy Coupling

The addition of phosphate to a sugar is a common reaction that occurs in a cell. By itself, this process is not very energetically favorable (that is, it needs an input of energy to occur). Cells overcome this energy obstacle by using ATP to “drive” the reaction. The energy needed to drive reactions is harvested in very controlled conditions in the confines of an enzyme. This involves a process called ‘coupling’. In coupled reactions, an enzyme binds both a high energy molecule (usually ATP) and the other molecule(s) involved in the reaction. Hydrolysis of ATP provides energy for the enzyme to stimulate the reaction on the other substance(s).

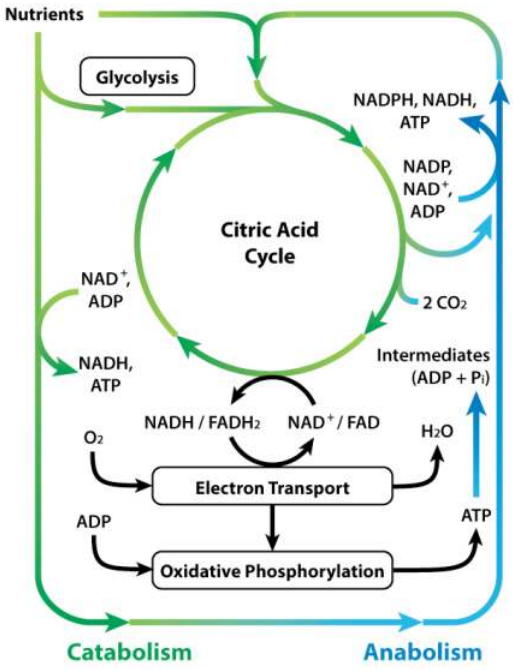

Figure 2.4.1: *Citric acid cycle.*

Hexokinase, for example, catalyzes the phosphorylation of glucose to form glucose-6-phosphate. In the absence of ATP, the reaction has a fairly positive ΔG°’ (described later), but hydrolysis of ATP provides excess energy, giving the coupled reaction a fairly negative ΔG°’ value.

2.4: Entropy and Energy

Most students who have had some chemistry know about the principle of the Second Law of Thermodynamics with respect to increasing disorder of a system. Cells are very organized or ordered structures, leading some to mistakenly conclude that life somehow violates the second law. In fact, that notion is incorrect. The second law doesn't say that entropy always increases, just that, left alone, it tends to do so, in an isolated system. Cells are not isolated systems, in that they obtain energy, either from the sun, if they are autotrophic, or food, if they are heterotrophic. To counter the universal tendency towards disorder on a local scale requires energy. As an example, take a fresh deck of cards which is neatly aligned with Ace-King-Queen . . . . 4,3,2 for each suit. Throw the deck into the air, letting the cards scatter. When you pick them up, they will be more disordered than when they started.

However, if you spend a few minutes (and expend a bit of energy), you can reorganize the same deck back to its previous, organized state. If entropy always increased everywhere, you could not do this. However, with the input of energy, you overcame the disorder. The cost of fighting disorder is energy.

There are, of course, other reasons that organisms need energy. Muscular contraction, synthesis of molecules, neurotransmission, signaling, thermoregulation, and subcellular movements are examples. Where does this energy come from? The currencies of energy are generally high-energy phosphate-containing molecules. ATP is the best known and most abundant, but GTP is also an important energy source (required for protein synthesis). CTP is involved in synthesis of glycerophospholipids and UTP is used for synthesis of glycogen. In each of these cases, the energy is in the form of potential chemical energy stored in the multi-phosphate bonds. Hydrolyzing those bonds releases the energy in them.

Of the triphosphates, ATP is the primary energy source, acting to facilitate the synthesis of the others by action of the enzyme NDPK. ATP is made by three distinct types of phosphorylation – oxidative phosphorylation (in mitochondria), photophosphorylation (in chloroplasts of plants), and substrate level phosphorylation (in enzymatically catalyzed reactions).

2.5: Gibbs Free Energy

Most of the time, ATP is the “storage battery” of cells (See also ‘Molecular Battery Backups for Muscles below). In order to understand how energy is captured, we must first understand Gibbs free energy and in doing so, we begin to see the role of energy in determining the directions chemical reactions take. Wikipedia defines Gibbs free energy as “a thermodynamic potential that measures the "useful" or process-initiating work obtainable from an isothermal, isobaric thermodynamic system,” and further points out that it is “the maximum amount of non-expansion work that can be extracted from a closed system; this maximum can be attained only in a completely reversible process.”

Mathematically, the Gibbs free energy is given as

$$G = H–TS \tag{2.5.1}$$

where H is the enthalpy, T is the temperature in Kelvin, and S is the entropy.

At standard temperature and pressure, every system seeks to achieve a minimum of free energy. Thus, increasing entropy will reduce Gibbs free energy. Similarly, if excess heat is available (reducing the enthalpy), the free energy can also be reduced. Cells must work within the laws of thermodynamics, as noted, so all of their biochemical reactions, too, have limitations. Now we shall consider energy in the cell. The change in Gibbs free energy (ΔG) for a reaction is crucial, for it, and it alone, determines whether or not a reaction goes forward.

$$\Delta G = \Delta H–T\Delta S, \tag{2.5.2}$$

There are three cases

- ΔG < 0 - the reaction proceeds as written
- ΔG = 0 - the reaction is at equilibrium
- ΔG > 0 - the reaction runs in reverse

For a reaction aA <=> bB (where ‘a’ and ‘b’ are integers and A and B are molecules) at pH 7, ΔG can be determined by the following equation,

$$\Delta G = \Delta G^{\circ\prime} + RT \ln \frac{([B]^b}{[A]^a} \tag{2.5.3}$$

For multiple substrate reactions, such as aA + cC <=> bB + dD

$$\Delta G = \Delta G^{\circ\prime} + RT \ln \frac{[B]^b[D]^d}{[A]^a[C]^c} \tag{2.5.4}$$

The ΔG°’ term is called the change in Standard Gibbs Free energy, which is the change in energy that occurs when all of the products and reactants are at standard conditions and the pH is 7.0. It is a constant for a given reaction.

In simple terms, if we collect all of the terms of the numerator together and call them {Products} and all of the terms of the denominator together and call them {Reactants},

$$\Delta G = \Delta G^{\circ\prime} + RT \ln \frac{Products}{Reactants} \tag{2.5.5}$$

For most biological systems, the temperature, T, is a constant for a given reaction. Since ΔG°’ is also a constant for a given reaction, the ΔG is changed almost exclusively as the ratio of {Products}/ {Reactants} changes. If one starts out at standard conditions, where everything except protons is at 1M, the RTln({Products}/{Reactants}) term is zero, so the ΔG°’ term determines the direction the reaction will take. This is why people say that a negative ΔG°’ indicates an energetically favorable reaction, whereas a positive ΔG°’ corresponds to an unfavorable one.

Increasing the ratio of {Products}/{Reactants} causes the value of the natural log (ln) term to become more positive (less negative), thus making the value of ΔG more positive. Conversely, as the ratio of {Products}/{Reactants} decreases, the value of the natural log term becomes less positive (more negative), thus making the value of ΔG more negative.

Intuitively, this makes sense and is consistent with **Le Chatelier's principle** – a system responds to stress by acting to alleviate the stress. If we examine the ΔG for a reaction in a closed system, we see that it will always move to a value of zero (equilibrium), no matter whether it starts with a positive or negative value.

Another type of free energy available to cells is that generated by electrical potential. For example, mitochondria and chloroplasts partly use Coulombic energy (based on charge) from a proton gradient across their membranes to provide the necessary energy for the synthesis of ATP. Similar energies drive the transmission of nerve signals (differential distribution of sodium and potassium) and the movement of some molecules in secondary active transport processes across membranes (e.g., H+ differential driving the movement of lactose). From the Gibbs free energy change equation,

$$\Delta G = \Delta H – T\Delta S \tag{2.5.6}$$

it should be noted that an increase in entropy will help contribute to a decrease in ΔG. This happens, for example when a large molecule is being broken into smaller pieces or when the rearrangement of a molecule increases the disorder of molecules around it. The latter situation arises in the hydrophobic effect, which helps drive the folding of proteins.

2.6: Cellular Phosphorylations

Formation of triphosphates is essential to meet the cell's immediate energy needs for synthesis, motion, and signaling. In a given day, an average human being uses more than their body weight in triphosphates. Since triphosphates are the "currency" that meet immediate needs of the cell, it is important to understand how triphosphates are made. There are three phosphorylation mechanisms – 1) substrate level; 2) oxidative; and 3) photophosphorylation. We consider them here individually.

Substrate Level Phosphorylation

The easiest type of phosphorylation to understand is that which occurs at the substrate level. This type of phosphorylation involves the direct synthesis of ATP from ADP and a reactive intermediate, typically a high energy phosphate-containing molecule. Substrate level phosphorylation is a relatively minor contributor to the total synthesis of triphosphates by cells. An example substrate phosphorylation comes from glycolysis.

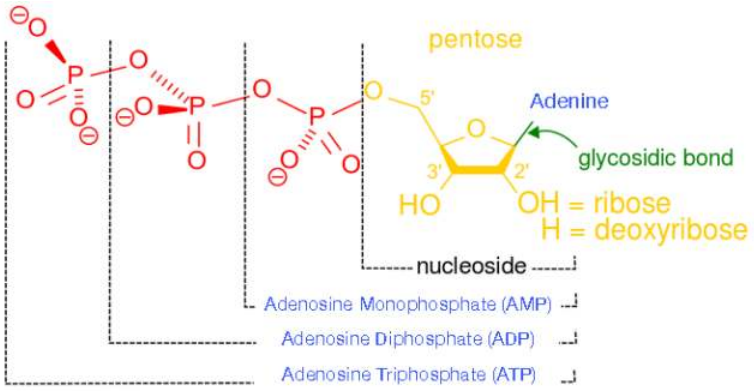

Figure **2.6.1**: *Adenine nucleotides.*

$$\text{Phosphoenolpyruvate (PEP)} + \text{ADP} \rightleftharpoons \text{Pyuvate} + \text{ATP} \tag{2.6.1}$$

This reaction has a very negative $\Delta G^{o'}$ (-31.4 kJ/mol), indicating that the PEP contains more energy than ATP, thus energetically favoring ATP's synthesis. Other triphosphates can be made by substrate level phosphorylation, as well. For example, GTP can be synthesized by the following citric acid cycle reaction

$$\text{Succinyl}-\text{CoA} + \text{GDP} + \text{P}_\text{i} \rightleftharpoons \text{Succinate} + \text{GTP} + \text{CoA}-\text{SH} \tag{2.6.2}$$

Triphosphates can be interchanged readily in substrate level phosphorylations catalyzed by the enzyme Nucleoside Diphosphate Kinase (NDPK). A generalized form of the reactions catalyzed by this enzyme is as follows:

$$\text{XTP} + \text{YDP} \rightleftharpoons \text{XDP} + \text{YTP} \tag{2.6.3}$$

where **X** can be adenosine, cytidine, uridine, thymidine, or guanosine and **Y** can be any of these as well. Last, an unusual way of synthesizing ATP by substrate level phosphorylation is that catalyzed by adenylate kinase

$$2\,\text{ADP} \rightleftharpoons \text{ATP} + \text{AMP} \tag{2.6.4}$$

This reaction is an important means of generating ATP when the cell doesn't have other sources of energy. The accumulation of AMP resulting from this reaction activates enzymes, such as phosphofructokinase, of glycolysis, that will catalyze reactions to give the cell additional, needed energy.

Electron Transport/Oxidative Phosphorylation

Mitocohondria are called the power plants of the cell because most of a cell's ATP is produced there, in a process referred to as oxidative phosphorylation. The mechanism by which ATP is made in oxidative phosphorylation is one of the most interesting processes in all of biology. It has three primary considerations. The first is electrical – electrons from reduced energy carriers, such as NADH and FADH2, enter an electron transport system via protein complexes containing iron. As seen in the figure on the following page,electrons move from one complex to the next, not unlike the way they might move through an electrical circuit.

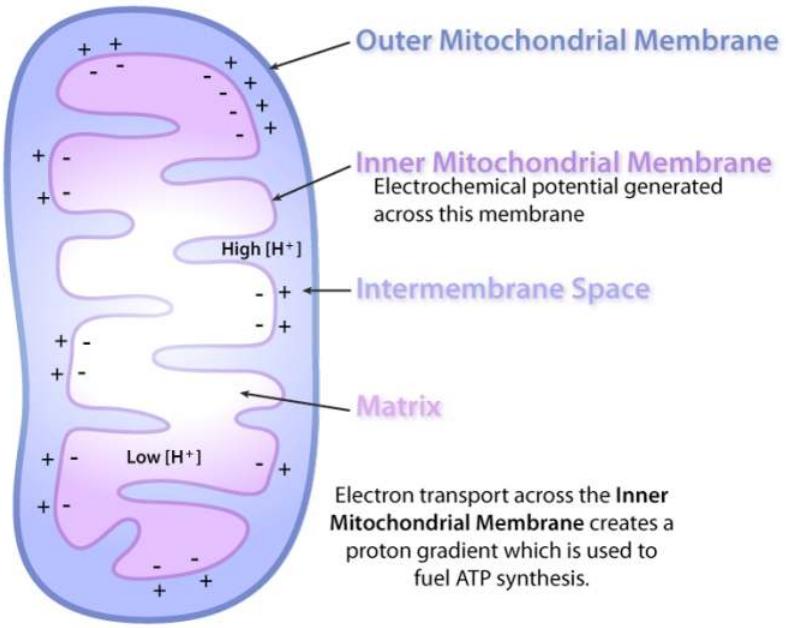

Figure 2.6.2: Mitochondria.

The next consideration arises as a secondary phenomenon. When electrons pass through complexes I, III, and IV, protons are moved from the mitochondrial matrix (inside of mitochondrion) and deposited in the intermembrane space (between the inner and outer membranes of the mitochondrion). The effect of this redistribution is to increase the electrical and chemical potential across the membrane. Students may think of the process as "charging the battery."

Just like a charged battery, the potential arising from the proton differential across the membrane can be used to do things. This is the third consideration. In the mitochondrion, the "thing" that the proton gradient does is create ATP from ADP and Pi (inorganic phosphate). This process requires energy and is accomplished by movement of protons through a protein complex in the inner mitochondrial membrane. The protein complex is an enzyme that has several names, including Complex V, PTAS (Proton Translocating ATP Synthase), and ATP Synthase. Central to its function is the movement of protons through it (from outside back into the matrix). Protons will only move through ATP Synthase if their concentration is greater outside the inner membrane than in the matrix.

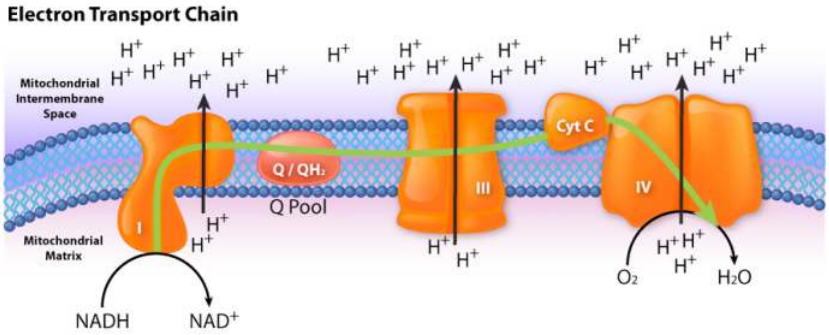

Figure 2.6.3: *Electron transport starting with complex I.*

In summary, the electron transport system charges the battery for oxidative phosphorylation by pumping protons out of the mitochondrion. The intact inner membrane of the mitochondrion keeps the protons out, except for those that re-enter through ATP Synthase. The ATP Synthase allows protons to re-enter the mitochondrial matrix and harvests their energy to make ATP.

ATP Synthase

The ATP Synthase itself is an amazing nanomachine that makes ATP using a gradient of protons flowing through it from the intermembrane space back into the matrix. It is not easy to depict in a single image what the synthase does. The figure at the right

illustrates the multi-subunit nature of this membrane protein, which acts like a turbine at a hydroelectric dam. The movement of protons through the ATP Synthase causes it to spin like a turbine, and the spinning is necessary for making ATP.

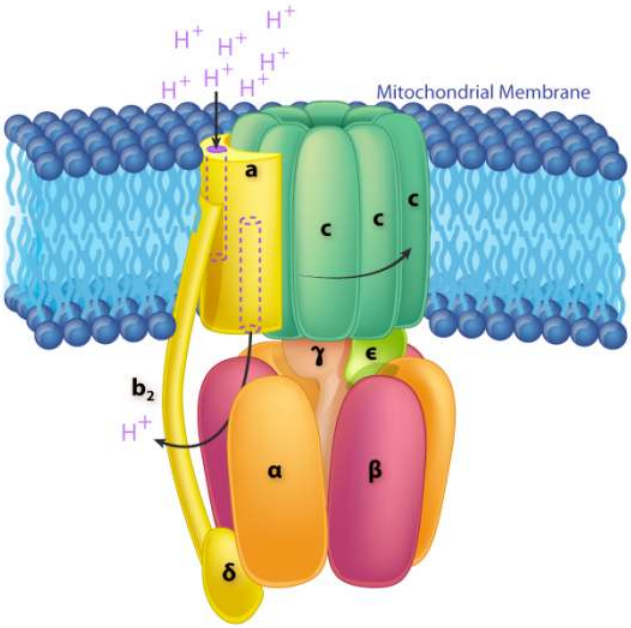

Figure **2.6.4**: *ATP Synthase (complex V).*

In ATP Synthase, the spinning component is the membrane portion (c ring) of the F0 stalk. The c ring proteins are linked to the gamma-epsilon stalk, which projects into the F1 head of the mushroom structure. The F1 head contains the catalytic ability to make ATP. The F1 head is hexameric in structure with paired alpha and beta proteins arranged in a trimer of dimers. Movement of the gamma protein inside the alpha-beta trimer causes each set of beta proteins to change structure slightly into three different forms called Loose, Tight, and Open (L,T,O). Each of these forms has a function. The Loose form binds **ADP** and $\mathbf{P_i}$. The tight form "squeezes" them together to form the ATP. The open form releases the ATP into the mitochondrial matrix. Thus,as a result of the proton excess in the intermembrane space, ATP is made.

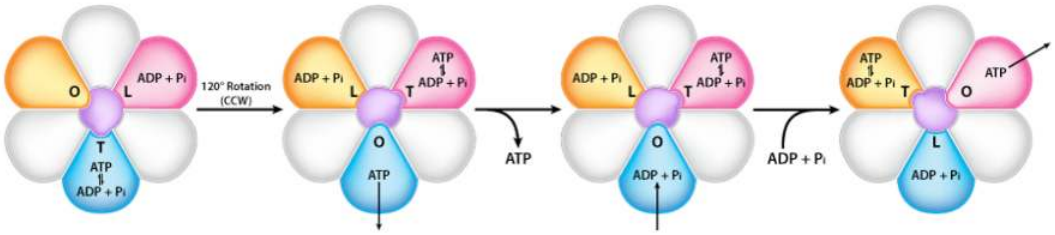

Figure **2.6.5**: *Three states of ATP Synthase.*

Photophosphorylation

The third type of phosphorylation to make ATP is found only in cells that carry out photosynthesis. This process is similar to oxidative phosphorylation in several ways. A primary difference is the ultimate source of the energy for ATP synthesis. In oxidative phosphorylation, the energy comes from electrons produced by oxidation of biological molecules. In the case of photosynthesis, the energy comes from the light of the sun.

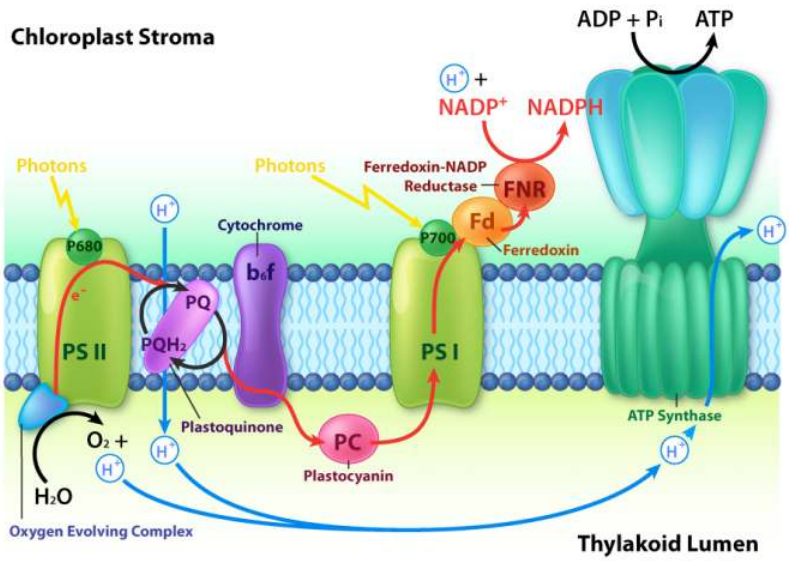

Figure **2.6.6**: Photophosphorylation.

Photons from the sun interact with chlorophyll molecules in reaction centers in the chloroplasts of plants or membranes of photosynthetic bacteria. A schematic of the process is shown above. The similarities of photophosphorylation to oxidative phosphorylation include:

- an electron transport chain
- creation of a proton gradient
- harvesting energy of the proton gradient by making ATP

with the help of an ATP synthase. Some of the differences include:

- the source of the electrons – $\mathbf{H_2O}$ for photosynthesis versus $\mathbf{NADH/FADH_2}$ for oxidative phosphorylation
- direction of proton pumping – into the thylakoid space of the chloroplasts versus outside the matrix of the mitochondrion
- movement of protons during ATP synthesis – out of the thylakoid space in photosynthesis versus into the mitochondrial matrix
- nature of the terminal electron acceptor – $\mathbf{NADP^+}$ in photosynthesis versus $\mathbf{O_2}$ in oxidative phosphorylation.

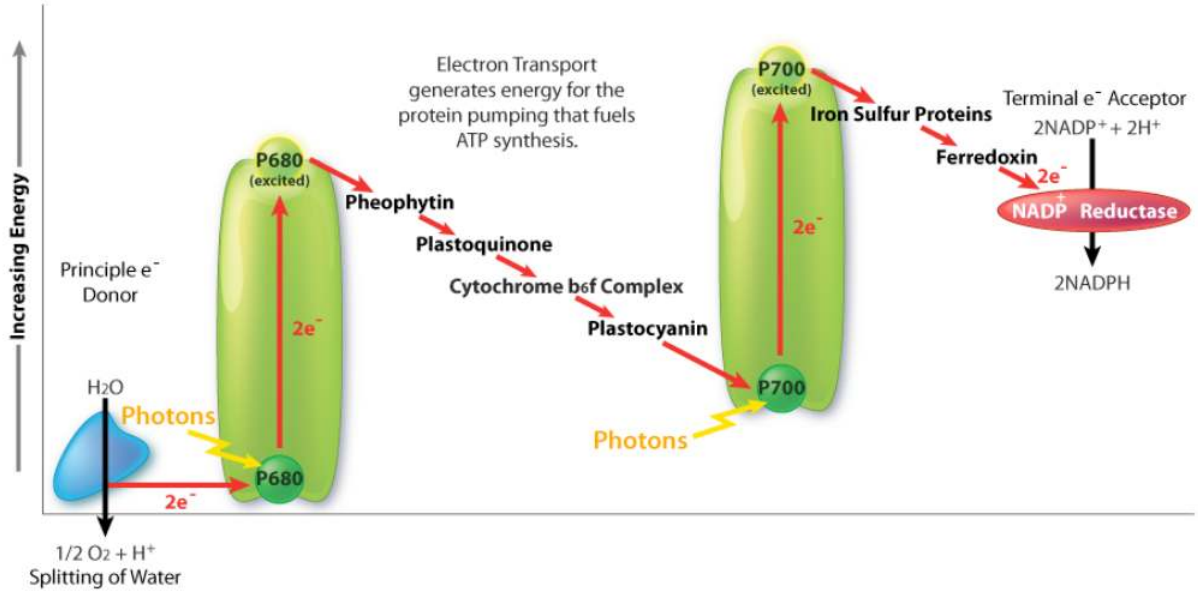

Figure **2.6.7**: Electron movement in photosynthesis.

Electron Transport in Chloroplasts vs. Mitochondria

In some ways, the movement of electrons in chloroplasts during photosynthesis is opposite that of electron transport in mitochondria. In photosynthesis, water is the source of electrons and their final destination is $NADPH$. In mitochondria, $NADH$/ $FADH_2$ are electron sources and H_2O is their final destination.

How do biological systems get electrons to go both ways? It would seem to be the equivalent of going to and from a particular place while always going downhill, since electrons will move according to potential. The answer is the captured energy of the photons, which elevates electrons in photosynthesis to an energy where they move "downhill" to their $NADPH$ destination in a Z-shaped scheme. The movement of electrons through this scheme in plants requires energy from photons in two places to "lift" the energy of the electrons sufficiently. Last, it should be noted that photosynthesis actually has two phases, referred to as the *light cycle* (described above) and the *dark cycle*, which is a set of chemical reactions that captures CO_2 from the atmosphere and "fixes" it, ultimately into glucose. The dark cycle is also referred to as the Calvin Cycle.

2.7: Energy Efficiency

Cells are not 100% efficient in energy use; nothing that we know of is. Consequently, cells do not get as much energy out of catabolic processes as they put into anabolic processes. A good example is the synthesis and breakdown of glucose, something liver cells are frequently doing. The complete conversion of glucose to pyruvate in glycolysis (catabolism) yields two pyruvates plus 2 NADH plus 2 ATPs.

Conversely, the complete conversion of two pyruvates into glucose by gluconeogenesis (anabolism) requires 4 ATPs, 2 NADH, and 2 GTPs. Since the energy of GTP is essentially equal to that of ATP, gluconeogenesis requires a net of 4 ATPs more than glycolysis yields. This difference must be made up in order for the organism to balance everything. It is for this reason that we eat. In addition, the inefficiency of our capture of energy in reactions results in the production of heat and helps to keep us warm.

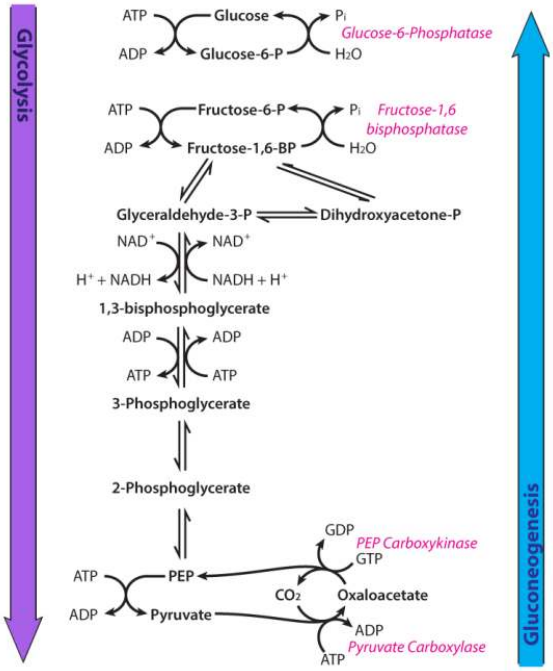

Figure 2.8.1: *Glycolysis and gluconeogenesis.*

2.8: Metabolic Controls of Energy

It is also noteworthy that cells do not usually have both catabolic and anabolic processes for the same molecules (for example, breakdown of glucose and synthesis of glucose, shown on the previous page) occurring simultaneously inside of them because the cell would see no net production of anything but heat and a loss of ATPs with each turn of the cycle. Such cycles are called futile cycles and cells have controls in place to limit the extent to which they occur. Since futile cycles can, in fact, yield heat, they are sources of heat in some types of tissue.

2.9: Molecular Backups for Muscles

For plants, the needs for energy are different than for animals. Plants do not need to access energy sources as rapidly as animals do, nor do they have to maintain a constant internal temperature. Plants can neither flee predators, nor chase prey. These needs of animals are much more immediate and require that energy stores be accessible on demand. Muscles, of course, enable the motion of animals and the energy required for muscle contraction is ATP. To have stores of energy readily available, muscles have, in addition to ATP, creatine phosphate and glycogen for quick release of glucose from glycogen. The synthesis of creatine phosphate is a prime example of the effects of concentration on the synthesis of high energy molecules. For example, creatine phosphate has an energy of hydrolysis of -43.1 kJ/mol whereas ATP has an energy of hydrolysis of -30.5 kJ/mol Creatine phosphate, however, is made from creatine and ATP in the reaction shown below. How is this possible?

Creatine + ATP <=> Creatine phosphate + ADP

The ΔG°' of this reaction is +12.6 kJ/mol, reflecting the energies noted above. In a resting muscle cell, ATP is abundant and ADP is low, driving the reaction to the right, creating creatine phosphate. When muscular contraction commences, ATP levels fall and ADP levels climb. The above reaction then reverses and proceeds to synthesize ATP immediately. Thus creatine phosphate acts like a battery, storing energy when ATP levels are high and releasing it almost instantaneously to create ATP when its levels fall.

2.10: Summary

In summary, energy is needed for cells to perform the functions that they must carry out in order to stay alive. At its most basic level, this means fighting a continual battle with entropy, but it is not the only need for energy that cells have.

CHAPTER OVERVIEW

3: Structure & Function

Function flows from structure. In order to understand the function of biomolecules, we must first understand their structures.

Thumbanil: An antibody molecule. The two heavy chains are colored red and blue and the two light chains green and yellow. (Public Domain; TimVickers).

3.1: Introduction to Structure & Function

If we hope to understand function in biological systems, we must first understand structure. At a simple level, we can divide molecules up according to their affinities for water – hydrophobic (limited solubility in water), hydrophilic (soluble in water) and amphiphilic (have characteristics of both hydrophobicity and hydrophilicity). Hydrophobicity in biological molecules arises largely because carbon-hydrogen bonds have electrons that are fairly evenly shared (not unlike carbon-carbon bonds). By contrast, the electrons between the oxygen and hydrogen of water are not equally shared. Oxygen has a greater electronegativity, so it holds them closer than hydrogen does. As a consequence, oxygen has what we call a partial negative charge and hydrogen has a partial positive charge.

Virtually all of life on Earth is built upon the biochemistry that arises from the molecular properties described in the preceding paragraph. The biomolecules referred to as lipids are largely water insoluble because they have predominantly carbon-hydrogen bonds with few ionic or hydrogen bond characteristics.

3.2: Building Blocks

Biological macromolecules are all polymers of a sort, even fats, in which the fatty acids can be thought of as polymers of carbon. (We will consider fatty compounds - fats, glycerophospholipids, sphingolipids, isoprenoids/terpenoids separately). The remaining categories of biological macromolecules include proteins, nucleic acids, and polysaccharides. The building blocks of these, respectively, are amino acids, nucleotides, and monosaccharides (sugars). Of these, the most diverse collection of chemical properties is found among the amino acids.

3.3: Proteins

Whereas nucleotides all are water soluble and have the same basic composition (sugar, base, phosphate) and the sugars also are water soluble and mostly contain 5 or 6 carbons (a few exceptions), the amino acids (general structure below) are structurally and chemically diverse.

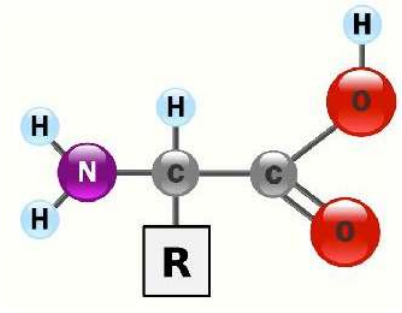

Figure 3.3.1: Amino Acid Schematic

Though all of the amino acids are, in fact, soluble in water, the interactions of their side chains with water differ significantly. This is important, because it is only in the side chains (R-groups) that amino acids differ from each other. Based on side chains, we can group the 20 amino acids found in proteins as follows:

- Aromatic (phenylalanine, tyrosine, tryptophan)
- Aliphatic (leucine, isoleucine, alanine, methionine, valine)
- Hydroxyl/Sulfhydryl (threonine, serine, tyrosine, cysteine)
- Carboxyamide (glutamine, asparagine)
- R-Acids (glutamic acid, aspartic acid)
- R-Amines (lysine, histidine, arginine)
- Odd (glycine, proline)

Note that tyrosine has a hydroxyl group and fits into two categories. Note also that biochemistry books vary in how they organize amino acids into categories. Amino acids are joined to each other by peptide bonds. This introduces a slight simplifying aspect to the structure of proteins – one need only consider the positioning of the R-groups around each peptide bond when determining protein structure schematically. Proteins that are in aqueous environments, such as the cytoplasm of the cell, have their amino acids arranged so that those with hydrophilic side chains (such as threonine or lysine) predominate on the exterior of the protein so as to interact with water. The hydrophobic amino acids in these proteins are found predominantly on the interior. When one examines the structure of proteins in non-aqueous environments, such as the interior of a lipid bilayer, the arrangement is flipped – hydrophobics predominate on the outside where they can interact with the hydrophobic side chains of membrane fatty acids and the hydrophilic amino acids are arranged anyplace where they can contact water. For a protein like porin, which provides an interior channel through which water can pass, this is where the hydrophilics are found. For transmembrane proteins, which project through both sides of the membrane, the hydrophilics are found at each point where the polypeptide chain emerges from the membrane.

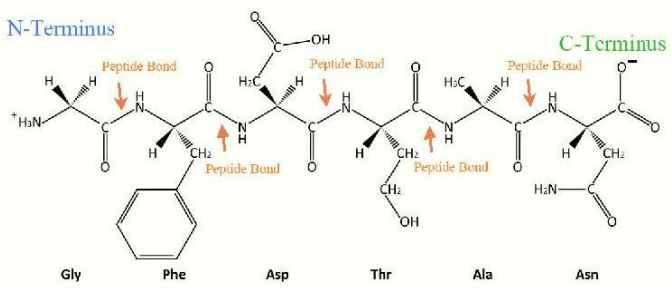

Figure 3.3.2: An image of the primary structure of a protein.

Primary Structure

How do proteins obtain such arrangements of amino acids? As we shall see, the structures of all proteins ultimately arise from their amino acid sequences. The amino acid sequence is referred to as the primary structure and changes in it can affect every other level of structure as well as the properties of a protein. The primary structure of a protein arrived at its current state as a result of mutation and selection over evolutionary time. On a more immediate time scale, 3D protein structure arises as a result of a phenomenon called folding. Protein folding results from three different structural elements beyond primary structure. They are referred to as secondary, tertiary, and quaternary structures, each arising from interactions between progressively more distant amino acids in the primary structure.

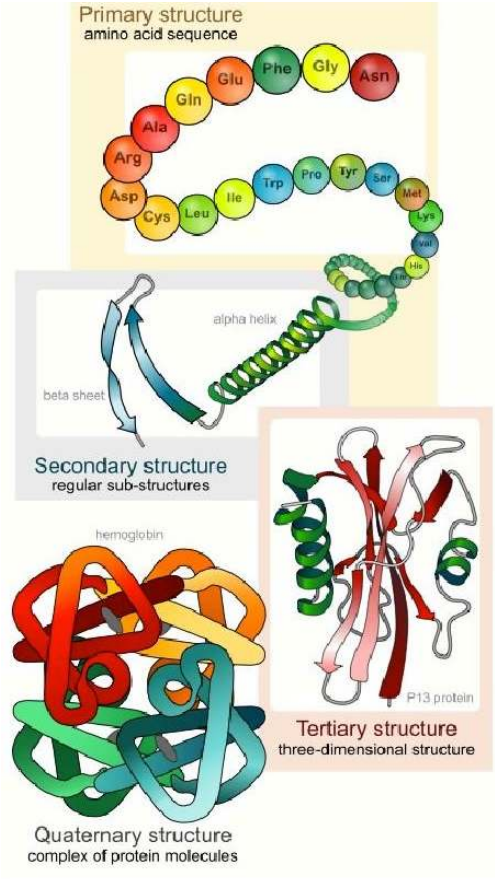

Figure **3.3.3**: *Schematic of a protein.*

Secondary Structure

Interactions between amino acids within about ten units of each other give rise to regular repeating structures. These secondary structures include the well known alpha-helix and beta strands. Both were predicted by Linus Pauling, Robert Corey, and Herman Branson in 1951. Each structure has unique features. We use the terms rise, repeat, and pitch to describe the parameters of a helix. The repeat is the number of residues in a helix before it begins to repeat itself. The rise is the distance the helix elevates with addition of each residue. The pitch is the distance between the turns of the helix.

Alpha Helix

The alpha helix (Figure **3.3.3**) forms as the result of interactions between amino acids separated by four residues. Interestingly, the side chains of the amino acids in an alpha helix are all pointed outwards from the axis of the helix. Alpha helices have a repeat of 3.6 amino acid residues per turn of the helix, meaning that four turns of the helix have approximately 14 amino acid residues. Hydrogen bonds occur between the C=O of one amino acid and the N-H of another amino acid four residues distant and these help to stabilize the structure (note that the C=O and N-H involved are part of the polypeptide backbone, not the R-groups). Some amino

acids have high helix forming tendencies. They include methionine, alanine, leucine, uncharged glutamate, and lysine. Others, such as proline, glycine, and negatively charged aspartate, disfavor its formation.

Beta Strands

Beta strands are the most fundamental helix, having essentially a 2D backbone of 'fold' like those of the pleats of a curtain. Indeed, beta strands can be arranged together to form what are called beta sheets. Other regular structures are also known. What determines whether a given stretch of a protein is in a helical or other structure? Here is where the shape and chemistry of the side chains play a role.

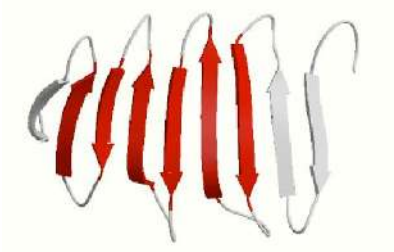

Figure **3.3.4**: *A beta sheet.*

Fibrous Proteins

Not all proteins have significant amounts of tertiary or quaternary structure. (As we shall see, these last two levels of structure arise from 'bends' in polypeptide chains and interactions between separate polypeptide chains, respectively.)

Alpha keratin, for example, is what we refer to as a fibrous protein (also called scleroprotein). Alpha keratin has primary structure and secondary structure, but little tertiary or quaternary structure.

Consequently, alpha keratin exists mostly as long fibers, such as are found in hair. Beta-keratin is a harder fibrous protein found in nails, scales, and claws. It is made up mostly of beta sheets. Proline, which is the least flexible amino acid, due to attachment of the side chain to the alpha-amino grip, is less likely to be found in alpha helices, but curiously it is found abundantly in the fibrous protein known as collagen. Collagen (previous page) is the most abundant protein in the human body and is the 'glue' that literally sticks us together. How does the inflexibility of proline permit it to be in a helix? The answer is probably the parallel abundance in collagen of glycine, which contains the smallest side group and therefore has the greatest flexibility.

As an interesting sidelight of the presence of proline in collagen is the chemical modification of prolines, by the addition of hydroxyl groups, after the protein is made. Such 'post-translational modifications' are not uncommon. Threonine, serine, and tyrosine frequently have their hydroxyl side-chains phosphorylated. Lysines in collagen too are hydroxylated post-translationally. The hydroxylated prolines and lysines play a role in the formation of interchain hydrogen bonds and crosslinking of triple helices during the assembly of collagen fibrils. These bonds provide structural integrity to the collagen. The enzymes that add hydroxyls to proline and lysine require vitamin C (ascorbic acid) for their activity. Lack of vitamin C leads to the production of weakened collagen fibrils, resulting in a condition called scurvy.

The carbonyl oxygen of the peptide bond can exist in resonance with the C-N bond, giving the peptide bond characteristics of a double bond and imposing limitations for rotation around it. If we treat the peptide bond as a double bond, then the arrangements of adjacent carbon bonds around it can be thought of as being in the cis or trans configurations. In proteins, not surprisingly, the preferred arrangement of these groups is strongly trans (1000/1). Of the 20 amino acids, the one that favors peptide bonds in the cis configuration most commonly is proline, but even for proline, the trans isomer is strongly preferred.

Figure 3.2.5: *Collagen.*

Ramachandran Plots

Another consequence of considering the peptide bond as a double bond is that it reduces the number of variable rotational angles of the polypeptide backbone. The terms phi and psi refer to rotational angles about the bonds between the N-alpha carbon and alpha carbon-carbonyl carbon respectively (previous page). Given the bulkiness of R-groups, the phenomenon of steric hindrance and the tendency of close side chains to interact with each other, one might expect to find a bias in the values of phi and psi. Indeed, that is exactly what is observed. Dr. G.N. Ramachandran proposed such a result and, in a plot that bears his name, depicted the theoretical

likelihood of each angle appearing in a polypeptide. More recent observations of actual phi and psi angles in data from the PDB protein database bear out Dr. Ramachandran's predictions. In the plot above, beta strands fit nicely in the darker blue section at the top and alpha helices fit in the yellow section near the middle.

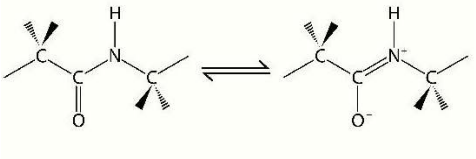

Figure **3.3.6**: Peptide bond resonance.

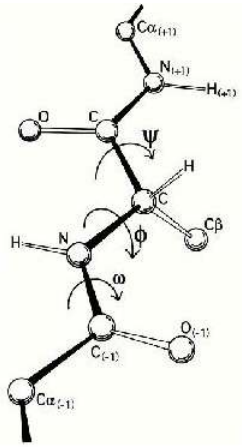

Figure **3.3.7**: Phi and psi angles.

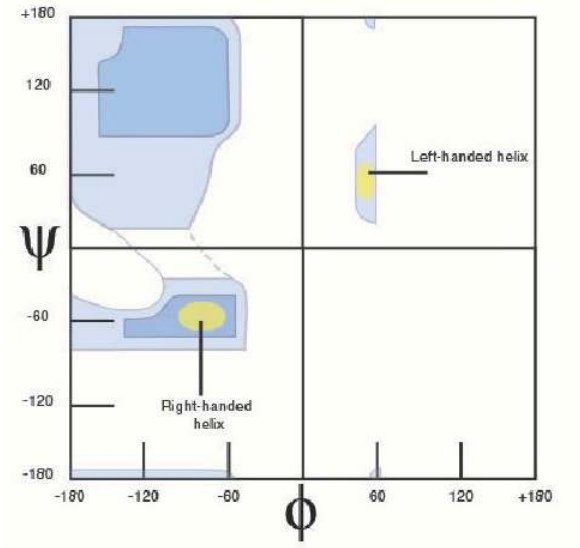

Figure **3.3.8**: Ramachandran Plot.

Tertiary Structure

In contrast to secondary structures, which arise from interactions between amino acids close in primary structure, tertiary structure arises from interactions between amino acids more distant in primary structure. Such interactions are not possible in an endlessly stretching fiber because each amino acid placed between two amino acids causes them to be moved farther away from each other in what is essentially the two dimensions of a secondary structure. For distant amino acids to interact, they must be brought into closer proximity and this requires bending and folding of the polypeptide chain. Proteins with such structures are referred to as ‘globular’ and they are, by far, the most abundant class of proteins. Indeed, it is in globular proteins that we have the most vivid images of the results of folding. “Folds" in polypeptides arise as a result of ‘bends’ between regions of secondary structure (such as alpha helix or beta strands). Such structures may be preferred due to incompatibility of a given amino acid side chain for a secondary structure

formed by the amino acids preceding it. Bends occur commonly in proteins and proline is often implicated. Bends do not have the predictable geometry of alpha helices or beta strands and are often referred to as random coils. Thus, even though protein structure can be described easily as regions of secondary structure separated by bends, the variability of bend structures makes prediction of tertiary structure from amino acid sequence enormously more difficult than identifying/predicting regions of secondary structure.

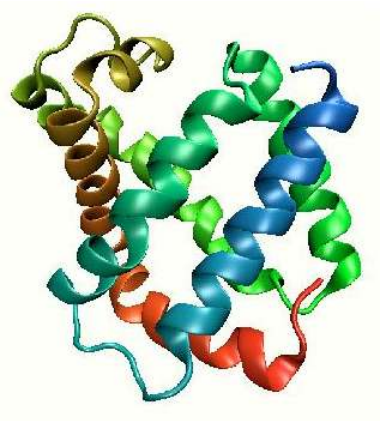

Figure **3.3.9**: *The structure of Myoglobin*

Hydrophobic Effect

It is at the level of tertiary structure that the characteristic arrangement of hydrophobic and hydrophilic amino acids in a protein occurs. In an aqueous environment, for a protein to remain soluble, it must have favorable interactions with the water around it, hence, the positioning of hydrophilic amino acids externally. Another impetus for the folding phenomenon is a bit harder to understand. It is known as the hydrophobic effect. At a chemical level, it makes sense – hydrophobic amino acids will 'prefer' to interact with each other internally and away from water. The driving force for this phenomenon, though, is a bit more conceptually difficult. Consider a bottle containing oil and water. As everyone knows, the two liquids will not mix and instead will form separate layers. A reasonable question might be why they do this instead of one existing as tiny globules inside of the other. The answer to that question, as well as the positioning of hydrophobic amino acids in the interior of water soluble proteins, is the hydrophobic effect. To understand the hydrophobic effect, perform the followingexperiment – take the water-oil mixture and shake it vigorously. This will force the layers to mix and one will observe that tiny globules of both water and oil can, in fact, be found initially in the layer of each. Over time, though, the tiny globules break up and merge with the appropriate layer. This is due to the phenomenon of entropy and consideration of surface area. First, the sum of the surface area of the embedded tiny globs is far greater than the area of the region between the two layers after mixing is over. The smaller the globs, the more the surface area of interaction between the oil and the water. The minimum possible surface area of interaction occurs when there are no globs at all – just two layers and nothing else.

Figure **3.3.10**: *A hydrophobic effect causes the water on this leaf to assume a spherical shape.*

How does this relate to entropy? Interactions between the water-hydrophobic layers causes the molecules at the interface to arrange themselves precisely/regularly so as to minimize their interactions. Ordering thus occurs at the layer interfaces. The maximum amount of ordering occurs when the maximum surface areas of oil and water interact. Small globules give rise to more exposed surface area between the water and hydrophobic layers and, as a consequence, more ordering. Since entropy in a closed system tends to increase, it will tend to reduce the amount of ordering, if left alone. Thus, one can increase the ordering on a nanoscopic

scale (forming globules) by applying energy in the form of shaking. When left alone, however, the system will increase its disorder by reducing the interactions between hydrophobic groups and hydrophilic ones.

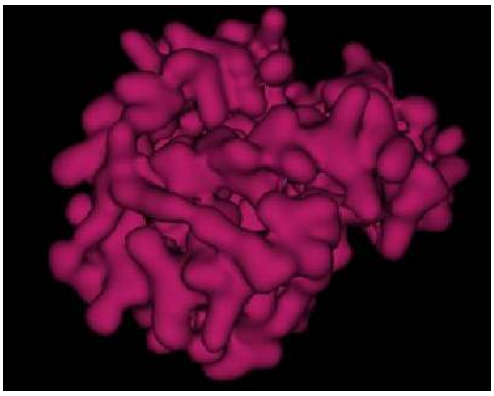

Figure **3.3.11**: Oxygenated Myoglobin.

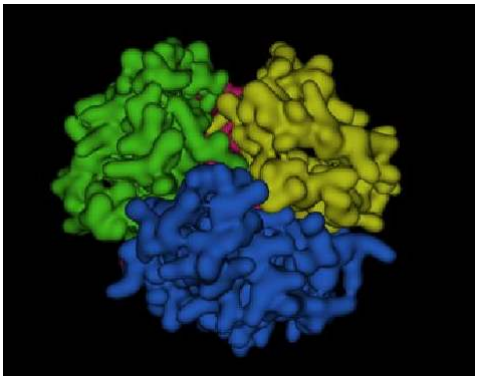

Figure **3.3.12**: *Hemoglobin in the absence of oxygen.*

In the oil water mixture, this causes the tiny globs to break up and produce the two layers we are familiar with because this is the minimum surface area that can be made between the two layers and thus the least ordering. In proteins, hydrophobic amino acid side chains are 'shielded' from water by placement internal to the protein, thus also reducing interfaces between hydrophobic residues and water. In both cases, entropy is increased, due to the reduced organization of the layers. Once formed, the interactions between the hydrophobic amino acid side chains helps to stabilize the overall protein structure.

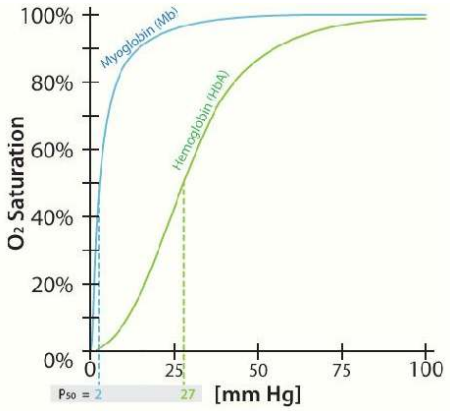

Figure **3.3.13**

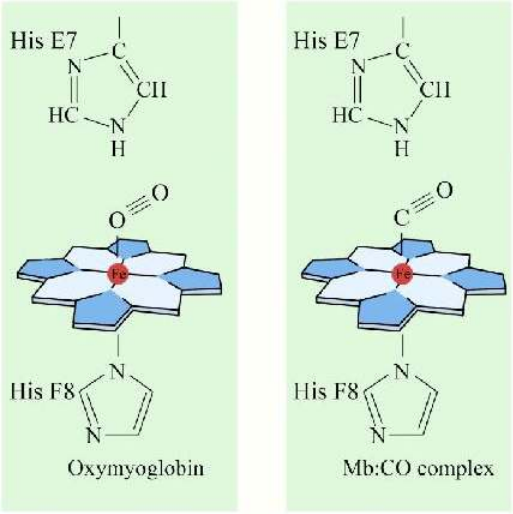

Figure 3.3.14: *Oxygen and Carbon Monoxide binding to heme.*

Quaternary Structure

The last level of protein structure we will consider is that of quaternary structure. In order to have quaternary structure, a protein must have multiple polypeptide subunits because the structure involves the arrangement of those subunits with respect to each other. Consider hemoglobin, the oxygen-carrying protein of our blood. It contains two identical subunits known as alpha and two other identical ones known as beta. These are arranged together in a fashion as shown on the previous page. By contrast, the related oxygen storage protein known as myoglobin only contains a single subunit. Hemoglobin has quaternary structure, but myoglobin does not. Multiple subunit proteins are common in cells and they give rise to very useful properties not found in single subunit proteins. In the case of hemoglobin, the multiple subunits confer the property of cooperativity – variable affinity for oxygen depending on the latter's concentration. In the case of enzymes, it can impart allosterism – the ability to have the activity of the enzyme altered by interaction with an effector molecule. We will discuss allosterism in detail in the next chapter.

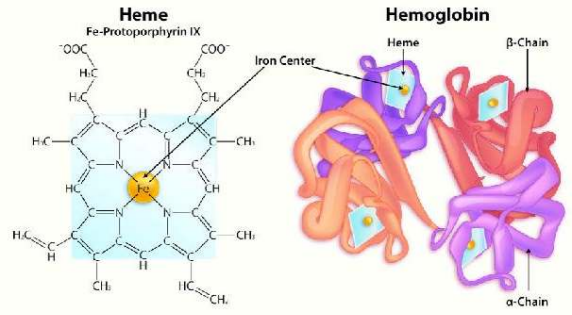

Figure 3.3.15: *Hemoglobin.*

Other Protein Structural Features

Not everything found in a protein is an amino acid. Proteins frequently have other chemical groups, known as prosthetic groups, bound to them, that are necessary for the function of a protein. Examples include the porphyrin ring of heme in myoglobin and hemoglobin that carries an iron so that oxygen can be bound. Metals are frequently employed by enzymes in their catalysis. Several vitamins (referred to as coenzymes), such as thiamine (B1) and riboflavin(B2) are modified and chemically bound to enzymes to help them perform specific catalytic functions.

Cooperativity

An interesting and important aspect of some proteins is the phenomenon of cooperativity. Cooperativity refers to the fact that binding of one ligand molecule by a protein favors the binding of additional molecules of the same type. Hemoglobin, for example, exhibits cooperativity when the binding of an oxygen molecule by the iron of the heme group in one of the four subunits causes a

slight conformation change in the subunit. This happens because the heme iron is attached to a histidine side chain and binding of oxygen 'lifts' the iron along with the histidine ring (also known as the imidazole ring).

Since each hemoglobin subunit interacts with and influences the other subunits, they too are induced to change shape slightly when the first subunit binds to oxygen (a transition described as going from the T-state to the R-state). These shape changes favor each of the remaining subunits binding oxygen, as well. This is very important in the lungs where oxygen is picked up by hemoglobin, because the binding of the first oxygen molecule facilitates the rapid uptake of more oxygen molecules. In the tissues, where the oxygen concentration is lower, the oxygen leaves hemoglobin and the proteins .ips from the R-state back to the T-state.

Cooperativity is only one of many fascinating structural aspects of hemoglobin that help the body to receive oxygen where it is needed and pick it up where it is abundant. Hemoglobin also assists in the transport of the product of cellular respiration (carbon dioxide) from the tissues producing it to the lungs where it is exhaled. Let us consider these individually.

Bohr Effect

The Bohr Effect was first described over 100 years ago by Christian Bohr. Shown graphically (above left), the observed effect is that hemoglobin's affinity for oxygen decreases as the pH decreases and/or as the concentration of carbon dioxide increases. Binding of the protons by histidine helps to facilitate structural changes in the protein and also with the uptake of carbon dioxide. Physiologically, this has great significance because actively respiring tissues (such as contracting muscles) require oxygen and release protons and carbon dioxide. The higher the concentration of protons and carbon dioxide, the more oxygen is released to feed the tissues that need it most.

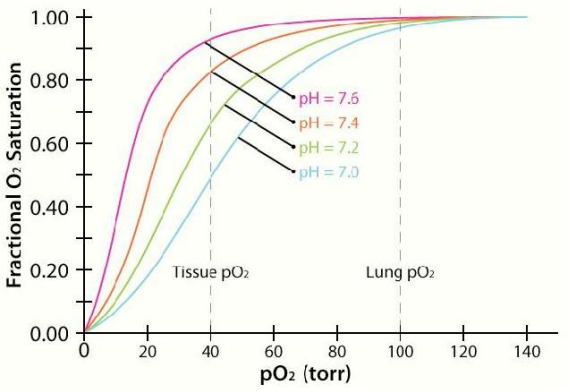

Figure **3.3.16**: *The Bohr Effect.*

2,3 BPG

Another molecule affecting the release of oxygen by hemoglobin is 2,3 bisphosphoglycerate (also called 2,3 BPG or just BPG). Like protons and carbon dioxide, 2,3 BPG is produced by actively respiring tissues, as a byproduct of glucose metabolism. The 2,3 BPG molecule fits into the 'hole of the donut' of adult hemoglobin. Such binding of 2,3 BPG favors the T (tight) state of hemoglobin, which has a reduced affinity for oxygen. In the absence of 2,3 BPG, hemoglobin can exist in the R (relaxed) state, which has a high affinity for oxygen.

Fetal Hemoglobin

Adult hemoglobin releases oxygen when it binds 2,3 BPG. This is in contrast to fetal hemoglobin, which has a slightly different configuration (α2Γ2) than adult hemoglobin (α2ß). Fetal hemoglobin has a greater affinity for oxygen than maternal hemoglobin, allowing the fetus to obtain oxygen effectively from the mother's blood. Part of the reason for fetal hemoglobin's greater affinity for oxygen is that it doesn't bind 2,3 BPG.

Another significant fact about 2,3 BPG is that its concentration is higher in the blood of smokers than it is of non-smokers. Consequently, hemoglobin in a smoker's blood spends more time in the T state than the R state. That is a problem when it is in the lungs, where being in the R state is necessary to maximally load the hemoglobin with oxygen. A high blood level of 2,3 BPG is one of the reasons smokers have trouble breathing when they exercise – they have reduced oxygen carrying capacity.

Last, though it is not related directly to 2,3 BPG, smokers have another reason why their oxygen carrying capacity is lower than that of non-smokers. Cigarette smoke contains carbon monoxide and this molecule, which has almost identical dimensions to molecular oxygen, competes effectively with oxygen for binding to the iron atom of heme. Part of carbon monoxide's toxicity is due to its ability to bind hemoglobin and prevent oxygen from binding.

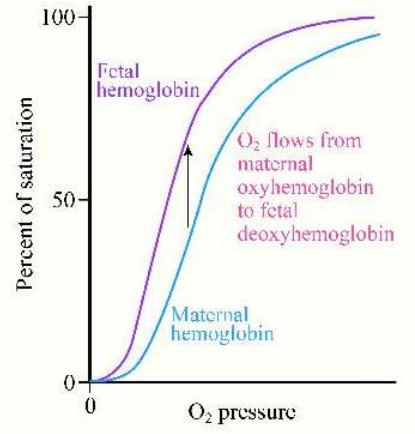

Figure **3.3.17***: Fetal hemoglobin binding of oxygen.*

Denaturation

For proteins, function is dependent on precise structure. Loss of the precise, folded structure of a protein is known as denaturation and is usually accompanied by loss of function. Anyone who has ever worked to purify an enzyme knows how easy it is for one to lose its activity. A few enzymes, such as ribonuclease, are remarkably stable under even very harsh conditions. For most others, a small temperature or pH change can drastically affect activity. The reasons for these differences vary, but relate to 1) the strength of the forces holding the structure together and 2) the ability of a protein to refold itself after being denatured. Let us consider these separately below.

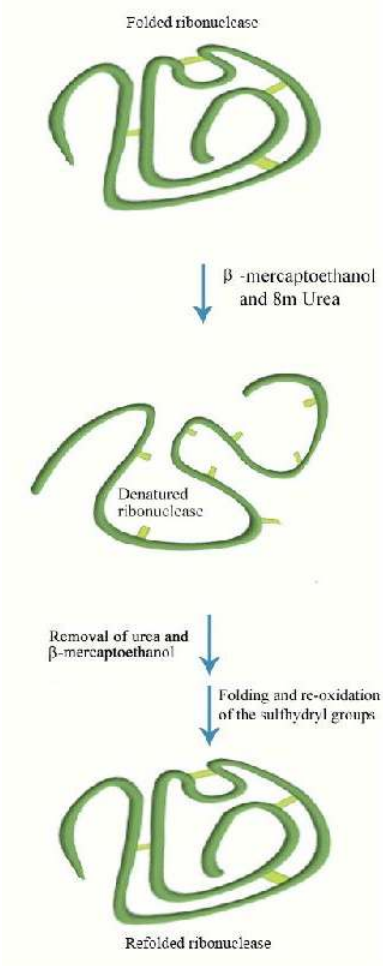

Figure **3.3.18**: *Denaturation and renaturation of ribonuclease.*

Forces Stabilizing Structures

Amino acids are linked one to the other by peptide bonds. These covalent bonds are extraordinarily stable at neutral pHs, but can be broken by hydrolysis with heat under acidic conditions. Peptide bonds, however, only stabilize primary structure and, in fact, are the only relevant force responsible for it. Secondary structure, on the other hand, is generally stabilized by weaker forces, including hydrogen bonds. Hydrogen bonds are readily disrupted by heat, urea, or guanidinium chloride.

Forces stabilizing tertiary structure include ionic interactions, disulfide bonds, hydrophobic interactions, metallic bonds, and hydrogen
bonds. Of these, the ionic interactions are most sensitive to pH changes. Hydrophobic bonds are most sensitive to detergents. Thus, washing one's hands helps to kill bacteria by denaturing critical proteins they need to survive. Metallic bonds are sensitive to oxidation/reduction. Breaking disulfide bonds requires either a strong oxidizing agent, such as performic acid or a strong reducing agent
on another disulfide, such as mercaptoethanol or dithiothreitol.

Quaternary structures are stabilized by the same forces as tertiary structure and have the same sensitivities.

Refolding Denatured Proteins

All of the information for protein folding is contained in the primary structure of the protein. It may seem curious then that most proteins do not refold into their proper, fully active form after they have been denatured and the denaturant is removed. A few do, in fact, refold correctly under these circumstances. A good example is bovine ribonuclease (also called RNase). Its catalytic activity is very resistant to heat and urea. However, if one treats the enzyme with mercaptoethanol (which breaks disulfide bonds) prior to urea treatment and

heating, activity is lost, indicating that the covalent disulfide bonds help stabilize the overall enzyme structure. If one allows the enzyme
mixture to cool back down to room temperature, over time some enzyme activity reappears, indicating that ribonuclease can re-fold under the proper conditions.

Irreversible Denaturation

Most enzymes, however, do not behave like ribonuclease. Once denatured, their activity cannot be recovered to any significant extent. This may seem to contradict the idea that folding information is inherent to the sequence of amino acids in the protein. It does not. The reason most enzymes can't refold properly is due to two phenomena. First, normal folding may occur as proteins are being made. Interactions among amino acids early in the synthesis are not "confused" by interactions with amino acids later in the synthesis because those amino acids aren't present as protein synthesis starts. In many cases, the proper folding of newly made polypeptides is also assisted by special proteins called chaperones. Chaperones bind to newly made proteins, preventing interactions that might result in misfolding. Thus, early folding and the assistance of chaperones eliminate some potential "wrong-folding" interactions that can occur if the entire sequence was present when folding started.

Denatured full-length polypeptides have many more potential wrong folds that can occur. A second reason most proteins don't refold properly after denaturation is probably that folding, like any other natural phenomenon, is driven by energy minimization. Though the folded structure may have a low energy, the path leading to it may not be all downhill. Like a chemical reaction that has energies of activation that must be overcome for the reaction to occur, folding likely has peaks and valleys of energy that do not automatically lead directly to the proper fold. Again, folding during synthesis leads the protein along a better-defined path through the energy maze of folding that denatured full-length proteins can't navigate.

Prions and Misfolding

Folding and the stability of folded proteins is an important consideration for so-called "infectious" proteins known as prions. These mysterious proteins, which are implicated in diseases, such as mad cow disease and the related human condition known as Creutzfeldt-Jakob disease, result from the improper folding of a brain protein known as PrP. The misfolded protein has two important properties that lead to the disease. First, it tends to aggregrate into large complexes called amyloid plaques that damage/destroy nerve cells in the brain, leading ultimately to dementia and loss of brain function.

Second, and probably worse, the misfolded protein "induces" other copies of the same protein to misfold as well. Thus, a misfolded protein acts something like a catalytic center and the disease progresses rapidly. The question arises as to how the PrP protein misfolds to begin with, but the answer to this is not clear. There are suggestions that exposure in the diet to misfolded proteins may be a factor, but this is disputed. An outbreak of mad cow disease in Britain in the 1980s was followed by a rise in the incidence of a rare form of human Creutzfeld-Jakob disease called variant CJD (v-CJD), lending some credence to the hypothesis. It is possible that misfolding of many proteins occurs sporadically without consequence or observation, but if PrP misfolds, the results are readily apparent. Thus, Creutzfeld-Jakob disease may ultimately give insights into the folding process itself.

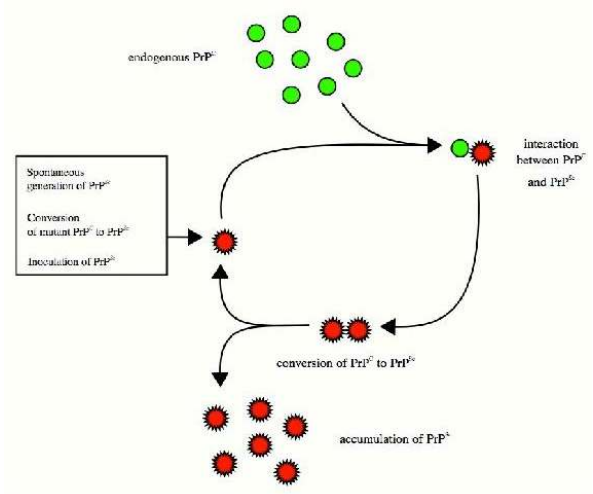

Figure **3.3.19**: *Prion protein misfolding.*

3.4: Nucleic Acids

Determination of the structure of the most common form of DNA, known as the B form, was one of the most important scientific advances of the 20th century. Using data from Rosalind Franklin, James Watson and Francis Crick initiated the modern era of molecular biology with their paper in the April 25, 1953 issue of Nature. Arguably, that single page paper has had more scientific impact per word than any other research article ever published. Today, every high school biology student knows the double helical structure in which G pairs with C and A pairs with T. The DNA molecule is a polymer of nucleoside monophosphates with phosphodiester bonds between the phosphate and the 5' end of one deoxyribose and the 3' end of the next one. In the B form the DNA helix has a repeat of 10.5 base pairs per turn, with sugars and phosphate forming the covalent "backbone" of the molecule and the adenine, guanine, cytosine, and thymine bases oriented in the middle where they form the now familiar base-pairs that look like the rungs of a ladder.

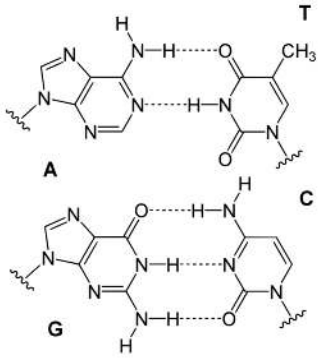

Figure **3.4.1***: A-T Base pairs (top) and G-C base pairs (bottom).*

Hydrogen bonds help to hold the base pairs together, with two hydrogen bonds per A-T pair and three hydrogen bonds per G-C pair. The two strands of a DNA duplex run in opposite directions. The 5' end of one strand is paired with the 3' end of the other strand and vice-versa at the other end of the duplex. The B form of DNA has a prominent major groove and a minor groove tracing the path of the helix (shown at left). Proteins, such as transcription factors bind in these grooves and access the hydrogen bonds of the base pairs to "read" the sequence therein.

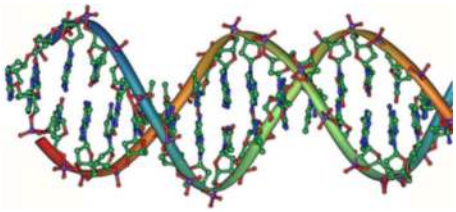

Figure **3.4.2**: *DNA double helix.*

Other forms of DNA besides the B form are known. One of these, the 'A' form, was identified by Rosalind Franklin in the same issue of Nature as Watson and Crick's paper. Though the A structure is a relatively minor form of DNA and resembles the B form, it turns out to be important in the duplex form of RNA and in RNA-DNA hybrids. Both the A form and the B form of DNA have the helix oriented in what is termed the right-handed form.

These stand in contrast to another form of DNA, known as the Z form. Z-DNA, as it is known, has the same base-pairing rules as the B and A forms, but instead has the helices twisted in the opposite direction, making a left-handed helix (Figure **3.4.3**). The Z form has a sort of zig-zag shape, giving to the name Z DNA. In addition, the helix is rather stretched out compared to the A and B forms. Why are there different forms of DNA. The answer relates to both superhelical tension and sequence bias. Sequence bias means that certain sequences tend to favor the "flipping" of B form DNA into other forms. Z DNA forms are favored by long stretches of alternating Gs and Cs.

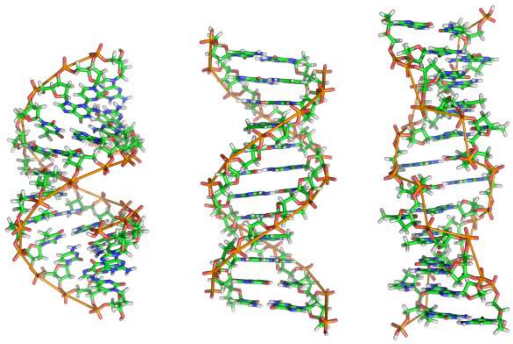

Figure **3.4.3**: *A, B, and Z forms of DNA*

Superhelicity

Short stretches of linear DNA duplexes exist in the B form and have 10.5 base pairs per turn. Double helices of DNA in the cell can vary in the number of base pairs per turn they contain. There are several reasons for this. For example, during DNA replication, strands of DNA at the site of replication get unwound at the rate of 6000 rpm by an enzyme called helicase. The effect of such local unwinding at one place in a DNA has the effect increasing the winding ahead of it. Unrelieved, such 'tension' in a DNA duplex can result in structural obstacles to replication.

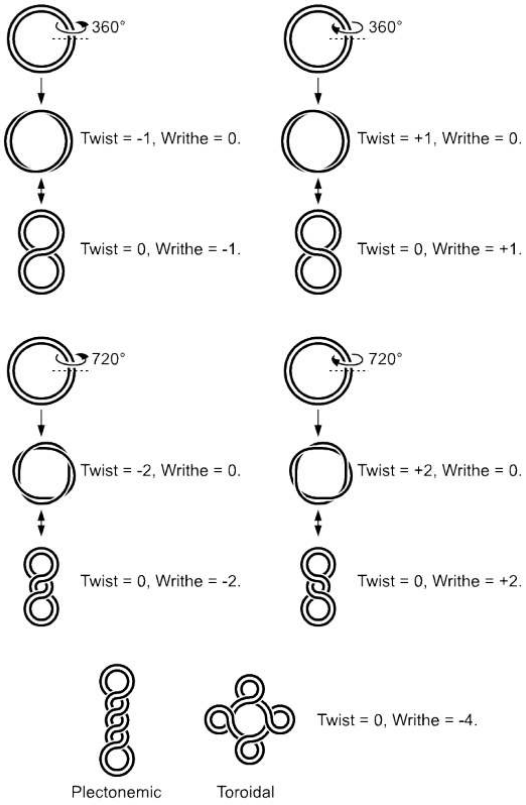

Figure **3.4.4**: *Topoisomers of DNA*

Such adjustments can occur in three ways. First, tension can provide the energy for 'flipping' DNA structure. Z-DNA can arise as a means of relieving the tension. Second, DNA can 'supercoil' to relieve the tension. In this method, the strands of the duplex can cross each other repeatedly, much like a rubber band will coil up if one holds one section in place and twists another part of it. Third, enzymes called topoisomerases can act to relieve or, in some cases, increase the tension by adding or removing twists in the DNA.

RNA Structures

With respect to structure, RNAs are more varied than their DNA cousins. Created by copying regions of DNA, cellular RNAs are synthesized as single strands, but they often have self-complementary regions leading to “fold-backs" containing duplex regions. The structure of tRNAs and rRNAs are excellent examples. The base-pairing rules of DNA are the same in RNA (with U in RNA replacing the T from DNA), but in addition, base pairing between G and U can also occur in RNA. This latter fact leads to many more possible duplex regions in RNA that can exist compared to single strands of DNA.

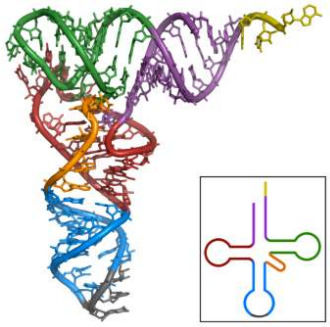

Figure **3.4.5**: *3D structure of tRNA (left) and 2D projection (right).*

RNA structure, like protein structure, has importance, in some cases, for catalytic function. Like random coils in proteins that give rise to tertiary structure, single-stranded regions of RNA that link duplex regions give these molecules a tertiary structure, as well. Catalytic RNAs, called ribozymes, catalyze important cellular reactions, including the formation of peptide bonds. DNA, which is usually present in cells in strictly duplex forms (no tertiary structure, per se), is not known to be involved in catalysis.

RNA structures are important for reasons other than catalysis. The 3D arrangement of tRNAs is important for enzymes that attach amino acids to them to do so properly. Small RNAs called siRNAs found in the nucleus of cells appear to play roles in both gene regulation and in cellular defenses against viruses. The key to the mechanisms of these actions is the formation of short fold-back RNA structures that are recognized by cellular proteins and then chopped into smaller units. One strand is copied and used to base pair with specific mRNAs to prevent the synthesis of proteins from them.

Denaturing Nucleic Acids

Like proteins, nucleic acids can be denatured. Forces holding duplexes together include hydrogen bonds between the bases of each strand that, like the hydrogen bonds in proteins, can be broken with heat or urea. (Another important stabilizing force for DNA arises from the stacking interactions between the bases in a strand.) Single strands absorb light at 260 nm more strongly than double strands (hyperchromic effect), allowing one to easily follow denaturation. For DNA, strand separation and strand hybridization are important aspects of the technique known as the polymerase chain reaction (PCR). Strand separation of DNA duplexes is accomplished in the method by heating them to boiling. Hybridization is an important aspect of the method that requires complementary single strands to “find" each other and form a duplex. Thus, DNAs (and RNAs too) can renature readily, unlike most proteins. Considerations for efficient hybridization (also called annealing) include temperature, salt concentration, strand concentration, and magnesium ion levels.

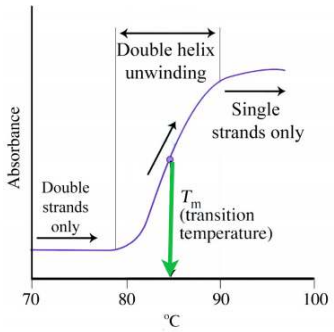

Figure **3.4.6**: *Hyperchromic effect.*

3.5: Carbohydrates

The last class of macromolecules we will consider structurally here is the carbohydrates. Built of sugars or modified sugars, carbohydrates have several important functions, including structural integrity, cellular identification, and energy storage.

Monosaccharides

Simple sugars, also known as monosaccharides, can generally be written in the form $C_x(H_2O)_x$. It is for this reason they are referred to as carbo-hydrates. By convention, the letters 'ose' at the end of a biochemical name flags a molecule as a sugar. Thus, there are glucose, galactose, sucrose, and many other '-oses'. Other descriptive nomenclature involves use of a prefix that tells how many carbons the sugar contains. For example, glucose, which contains six carbons, is described as a hexose. The following list shows the prefixes for numbers of carbons in a sugar:

- Tri- = 3
- Tetr- = 4
- Pent- = 5
- Hex- =6
- Hept- = 7
- Oct- = 8

Other prefixes identify whether the sugar contains an aldehyde group (aldo-) or a ketone (keto) group. Prefixes may be combined. Glucose, which contains an aldehyde group, can be described as an aldo-hexose. The list that follows gives some common sugars and some descriptors.

- Ribose = aldo-pentose
- Glucose = aldo-hexose
- Galactose = aldo-hexose
- Mannose = aldo-hexose
- Glyceraldehyde = aldo-triose
- Erythrose – aldo-tetrose
- Fructose = keto-hexose
- Ribulose = keto-pentose
- Sedoheptulose = keto-heptose
- Dihydroxyacetone = keto-triose

Stereoisomer Nomenclature

Sugars of a given category (hexoses, for example) differ from each other in the stereoisomeric configuration of their carbons. Two sugars having the same number of carbons (hexoses, for example) and the same chemical form (aldoses, for example), but differing in the stereoisomeric configuration of their carbons are called diastereomers. Biochemists use D and L nomenclature to describe sugars, as explained below.

D-sugars predominate in nature, though L-forms of some sugars, such as fucose, do exist. The D and L designation is a bit more complicated than it would appear on the surface. To determine if a sugar is a D-sugar or an L-sugar, one simply examines the configuration of the highest numbered asymmetric carbon. If the hydroxyl is written to the right, it is a D-sugar. If the hydroxyl is on the left, it is an L-sugar. That part is simple. The confusion about D and L arises because L sugars of a given name (glucose, for example) are mirror images of D sugars of the same name. The figure on the previous page shows the structure of D- and L-glucose. Notice that D-glucose is not converted into L-glucose simply by .ipping the configuration of the fifth carbon in the molecule. There is another name for sugars that are mirror images of each other. They are called enantiomers. Thus, L-glucose and D-glucose are enantiomers, but D-Erythrose and D-Threose are diastereomers.

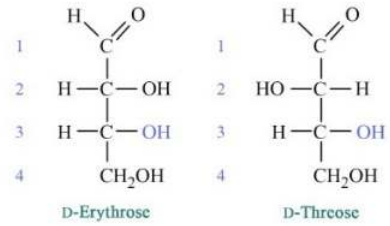

Figure 3.5.1: *Diastereomers.*

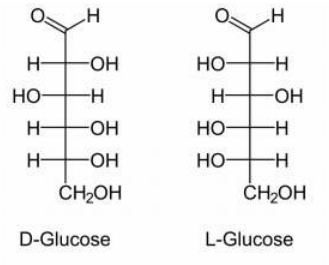

Figure 3.5.2: *Enantiomers.*

Sugars of 5-7 carbons can fairly easily form ring structures (called Haworth structures). For aldoses like glucose, this involves formation of a hemi-acetal. For ketoses like fructose, it involves formation of a hemi-ketal. The bottom line for both is that the oxygen that was part of the aldehyde or the ketone group is the one that becomes a part of the ring. More important than the oxygen, though, is the fact that the carbon attached to it (carbon #1 in aldoses or #2 in ketoses) becomes asymmetric as a byproduct of the cyclization. This new asymmetric carbon is called the anomeric carbon and it has two possible configurations, called alpha and beta.

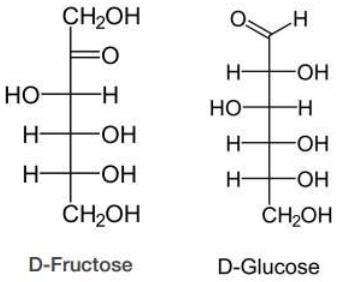

Figure 3.5.3: *Ketose and Aldose.*

A solution of glucose will contain a mixture of alpha and beta forms. Whether the alpha or the beta arises upon cyclization is partly determined by geometry and partly random. Thus, one can find a bias for one form, but usually not that form exclusively. A given molecule of sugar will flip between alpha and beta over time. A requirement for this is that the hydroxyl on the anomeric carbon is unaltered, thus facilitating flipping back to the straight chain form followed by recyclization. If the hydroxyl becomes chemically altered in any way (for example, replacement of its hydrogen by a methyl group), a glycoside is formed. Glycosides are locked in the same alpha or beta configuration they were in when the modification was made. Glycosides are commonly found in nature. Sucrose, for example , is a di-glycoside – both the glucose and the fructose have had their anomeric hydroxyls altered by being joined together.

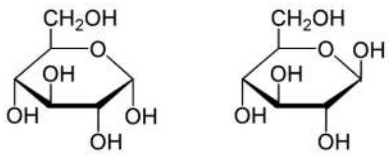

Figure 3.5.4: *alpha-D Glucose (left) and beta-D Glucose (right)*

The last considerations for sugars relative to their structure are their chemical reactivity and modification. The aldehyde group of aldoses is susceptible to oxidation, whereas ketoses are less so. Sugars that are readily oxidized are called ‘ reducing sugars ’ because their oxidation causes other reacting molecules to be reduced. Reducing sugars can easily be identified in a chemical test. Chemical modification of sugars occurs readily in cells. As we will see, phosphorylation of sugars occurs routinely during metabolism. Oxidation of sugars to create carboxyl groups also can occur. Reduction of aldehyde/ketone groups of sugars creates what are called sugar alcohols, and other modifications, such as addition of sulfates and amines also readily occur.

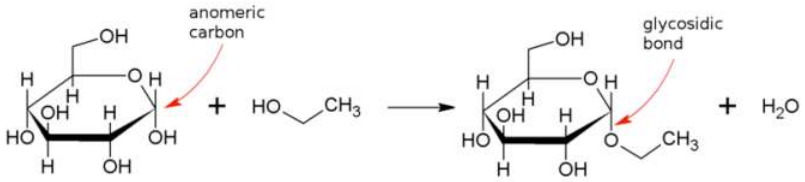

Figure **3.5.5**: Glycoside formation

Boat/Chair Conformations

Independent of stereoisomerization, sugars in ring form of a given type (such as glucose) can “twist" themselves into alternative conformations called boat and chair. Note that this rearrangement does not change the relative positions of hydroxyl groups. All that has changed is the shape of the molecule. As shown for glucose, one can see that the beta-hydroxyl of glucose is closer to the CH_2OH (carbon #6) in the boat form than it is in the chair form. Steric hindrance can be a factor in favoring one configuration over another.

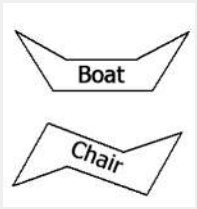

Figure **3.5.6**: *Boat and Chair*

Disaccharides

Sugars are readily joined together (and broken apart) in cells. Sucrose (Figure **3.5.7**), which is common table sugar, is made by joining the anomeric hydroxyl of alpha-D-glucose to the anomeric hydroxyl of beta-D-fructose. Not all disaccharides join the anomeric hydroxyls of both sugars. For example, lactose (milk sugar) is made by linking the anomeric hydroxyl of galactose in the beta configuration to the hydroxyl of carbon #4 of glucose.

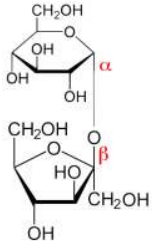

Figure **3.5.7**: Sucrose

Oligosaccharides

The term ‘oligosaccharide’ is used to describe polymers of sugars of 5-15 units, typically. Oligosaccharides are not commonly found free in cells, but instead are found covalently attached to proteins, which are then said to be glycosylated. Oligosaccharides attached to proteins may be **N-linked** (through asparagine) or **O-linked** (though serine or threonine). O-linked sugars are added only in the Golgi apparatus while N-linked sugars are attached starting in the endoplasmic reticulum and then completed in the Golgi.

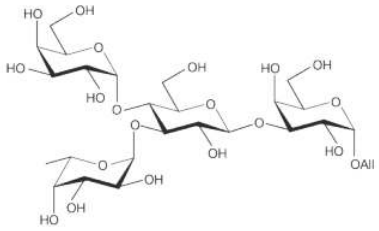

Figure **3.5.8**: *An oligosaccharide.*

Oligosaccharides often function as identity markers, both of cells and proteins. On the cell surface, glycoproteins with distinctive oligosaccharides attached establish the identity of each cell. The types of oligosaccharides found on the surface of blood cells is a determinant of blood type. The oligosaccharides that are attached to proteins may also determine their cellular destinations. Improper glycosylation or sugar modification patterns can result in the failure of proteins to reach the correct cellular compartment. For example, inclusion cell (I-cell) disease arises from a defective phosphotransferase in the Golgi. This enzyme normally catalyzes the addition of a phosphate to a mannose sugar attached to a protein destined for the lysosome. In the absence of a functioning enzyme, the unphosphorylated glycoprotein never makes it to the lysosome and is instead exported out of the cell where it accumulates in the blood and is excreted in the urine. Individuals with I-cell disease suffer developmental delays, abnormal skeletal development, and restricted joint movement.

Polysaccharides

Polysaccharides, as their name implies, are made by joining together many sugars. The functions for polysaccharides are varied. They include energy storage, structural strength, and lubrication. Polysaccharides involved in energy storage include the plant polysaccharides, amylose and amylopectin. The polysaccharide involved in energy storage in animals is called glycogen and it is mostly found in the muscles and liver.

Amylose/Amylopectin

Amylose is the simplest of the polysaccharides, being comprised solely of glucose units joined in an alpha 1-4 linkage. Amylose is broken down by the enzyme alpha-amylase, found in saliva. Amylopectin is related to amylose in being composed only of glucose, but it differs in how the glucose units are joined together. Alpha 1-4 linkages predominate, but every 30-50 residues, a ‘branch’ arises from an alpha 1-6 linkage. Branches make the structure of amylopectin more complex than that of amylose.

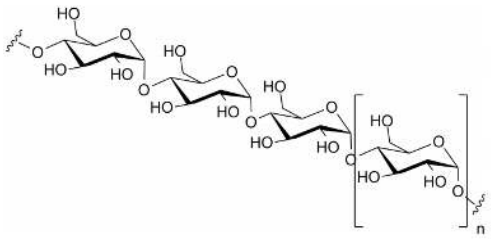

Figure **3.5.9**: Amylose.

Glycogen

Glycogen is a polysaccharide that is physically related to amylopectin in being built only of glucose and in having a mix of alpha 1-4 and alpha 1-6 bonds. Glycogen, however, has many more alpha 1-6 branches than amylopectin, with such bonds occurring about every 10 residues. One might wonder why such branching occurs more abundantly in animals than in plants. A plausible explanation is based on the method by which these molecules are broken down. The breakdown of these polysaccharides is catalyzed by enzymes, known as phosphorylases, that clip glucose residues from the ends of glycogen chains and attach a phosphate to them in the process, producing glucose-1-phosphate. More highly branched polysaccharides have more ends to clip, and this translates to more glucose-1-phosphates that can be removed simultaneously by numerous phosphorylases. Since glucose is used for energy by muscles, glucose concentrations can be increased faster the more branched the glycogen is. Plants, which are immobile do not have needs for such immediate release of glucose and thus have less need for highly branched polysaccharides.

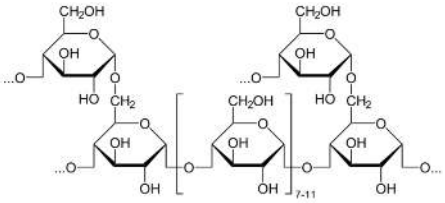

Figure **3.5.10**: *Glycogen Structure*

Cellulose

Another important polysaccharide containing only glucose is cellulose. It is a polymer of glucose used to give plant cell walls structural integrity and has the individual units joined solely in a beta 1-4 configuration. That simple structural change makes a radical diff erence in its digestibility. Humans are unable to break down cellulose and it passes through the digestive system as roughage. Ruminant animals, such as cattle, however have bacteria in their rumens that contain the enzyme cellulase. It breaks the beta 1-4 links of the glucoses in cellulose to release the sugars for energy.

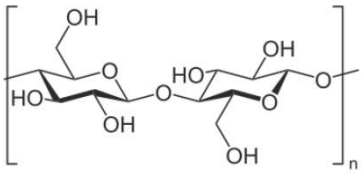

Figure **3.5.11**: *The repeating structure of cellulose.*

Another polysaccharide used for structural integrity is known as chitin. Chitin makes up the exoskeleton of insects and is a polymer of a modified form of glucose known as N-acetyl-glucosamine.

Glycosaminoglycans

Yet another category of polysaccharides are the glycosaminoglycans (also called mucopolysaccharides), some examples of which include keratan sulfate, heparin, hyaluronic acid (right), and chondroitin sulfate. The polysaccharide compounds are linked to proteins, but differ from glycoproteins in having a much larger contingent of sugar residues and, further, the sugars are considerably more chemically modified. Each of them contains a repeating unit of a disaccharide that contains at least one negatively charged residue. The result is a polyanionic substance that, in its interactions with water, makes for a “slimy" feel. Glycosaminoglycans are found in snot, and in synovial fluid, which lubricates joints. Heparin is a glycosaminoglycan that helps to prevent blood from clotting.

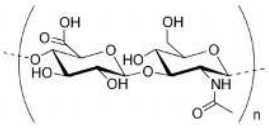

Figure **3.5.12**: *The repeating unit of a glycosaminoglycan structure.*

3.6: Lipids and Membranes

Lipids are a broad class of molecules that all share the characteristic that they have at least a portion of them that is hydrophobic. The class of molecules includes fats, oils (and their substituent fatty acids), steroids, fat-soluble vitamins, prostaglandins, glycerophospholipids, and sphingolipids. Interestingly, each of these can be derived from acetyl-CoA.

Fatty Acids

Arguably, the most important lipids in our cells are the fatty acids, because they are components of all of the other lipids, except some of the steroids and fat-soluble vitamins. Consisting of a carboxyl group linked to a long aliphatic tail, fatty acids are described as either saturated (no double bonds) or unsaturated (one or more double bonds). Fatty acids with more than one double bond are described as polyunsaturated. Increasing the amount of unsaturated fatty acids (and the amount of unsaturation in a given fatty acid) in a fat decreases its melting temperature. This is also a factor in membrane fluidity. If the melting temperature of a fat is decreased sufficiently so that it is a liquid at room temperature, it is referred to as an oil. It is worth noting that organisms like fish, which live in cool environments, have fats with more unsaturation. This is why fish oil is a rich source of polyunsaturated fatty acids.

	Number of Carbon Atoms	Unsaturation	Formula	Melting Point (°C)
Palmitoleic	16	$16:1-\Delta^{9}$	$CH_3(CH_2)_5CH{=}CH(CH_2)_7CO_2H$	-0.5
Oleic	18	$18:1-\Delta^{9}$	$CH_3(CH_2)_7CH{=}CH(CH_2)_7CO_2H$	16
Linoleic	18	$18:2-\Delta^{9,12}$	$CH_3(CH_2)_4CH{=}CH(CH_2)CH{=}CH(CH_2)_7CO_2H$	-5
Linolenic	18	$18:3-\Delta^{9,12,15}$	$CH_3(CH_2CH{=}CH)_3(CH_2)_7CO_2H$	-11
Arachidonic	20	$20:4-\Delta^{5,8,11,14}$	$CH_3(CH_2)_4CH{=}CH(CH_2)_4(CH_2)_2CO_2H$	-50

Figure 3.5.1: *Important unsaturated fatty acids.*

Biochemically, the double bonds found in fatty acids are predominantly in the cis configuration. So-called trans fats arise as a chemical by-product of partial hydrogenation of vegetable oil (small amounts of trans fats also occur naturally). In humans, consumption of trans fats raises low density lipoprotein (LDL) levels and lowers high density lipoprotein (HDL) levels. Each is thought to contribute to the risk of developing coronary artery disease. The most common fatty acids in our body include palmitate, stearate, oleate, linolenate, linoleate, and arachidonate. Fatty acids are numbered by two completely different schemes. The delta numbering scheme has the carboxyl group as #1, whereas the omega number scheme starts at the other end of the fatty acid with the methyl group as #1. Fatty acids are described as essential if they must be in the diet (can't be synthesized by the organism). Animals, including humans, cannot synthesize fatty acids with double bonds beyond position delta 9, so linoleic and linolenic acids are considered essential in these organisms.

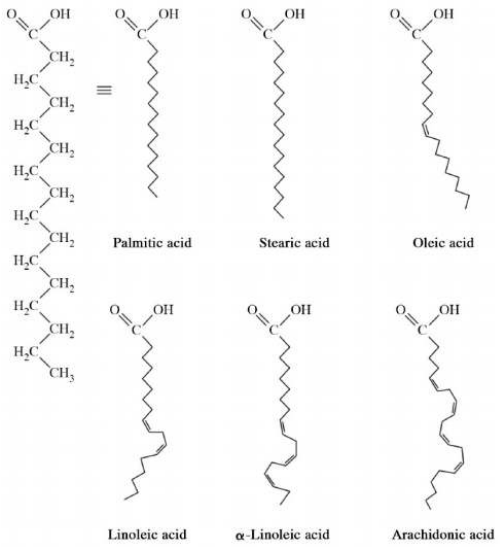

Figure 3.5.2: *Fatty acids.*

In animal cells, fats are the primary energy storage forms. They are also known as triacylglycerols, since they consist of a glycerol molecule esterified to three fatty acids. Fats are synthesized by replacing the phosphate on phosphatidic acid with a fatty acid. Fats

are stored in the body in specialized cells known as adipocytes. Enzymes known as lipases release fatty acids from fats by hydrolysis reactions. Of the various lipases acting on fat, the one that acts first, triacylglycerol lipase, is regulated hormonally.

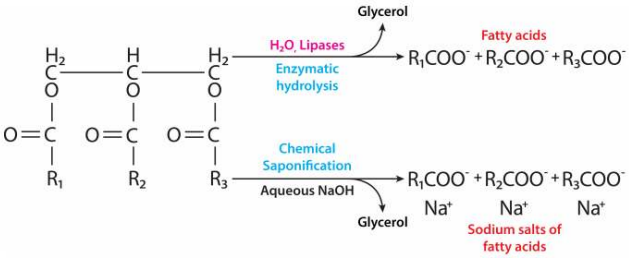

Figure 3.5.3: *Hydrolysis of fat.*

Membrane Lipids

The predominant lipids found in membranes are glycerophospholipids (phosphoglycerides) and sphingolipids. The former are related to fats structurally as both are derived from phosphatidic acid. Phosphatidic acid is a simple glycerophospholipid that is usually converted into phosphatidyl compounds. These are made by esterifying various groups, such as ethanolamine, serine, choline, inositol, and others to the phosphate. All of these compounds form lipid bilayers in aqueous solution, due to the amphiphilic nature of their structure.

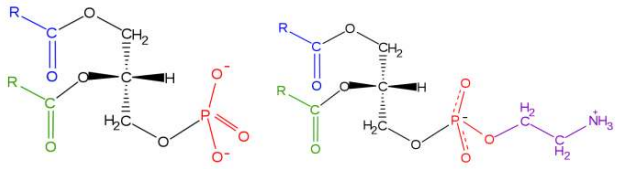

Figure 3.5.4: *Phosphatic Acid (Left) and Phosphatidyl Ethanolamine (Right)*

Though structurally similar to glycerophospholipids, sphingolipids are synthesized completely independently of them, starting with palmitic acid and the amino acid serine. The figure on the right shows the structure of several sphingolipids. LIke the glycerophospholipids, sphingolipids are amphiphilic, but unlike them, they may have simple (in cerebrosides) or complex (in gangliosides) carbohydrates attached at one end. Most sphingolipids, except sphingomyelin, do not contain phosphate.

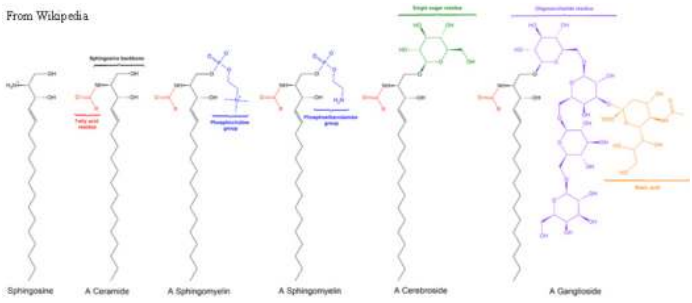

Figure 3.5.5: *The structures of several sphingolipids.*

Steroids, such as cholesterol are also found in membranes. Cholesterol, in particular, may play an important role in membrane fluidity. Membranes can be thought of a being more “frozen" or more “fluid." Fluidity is important for cellular membranes. When heated, membranes move from a more “frozen" character to that of a more “fluid" one as the temperature rises. The mid-point of this transition, referred to as the Tm, is influenced by the fatty acid composition of the lipid bilayer compounds. Longer and more saturated fatty acids will favor higher Tm values, whereas unsaturation and short fatty acids will favor lower Tm values. Interestingly, cholesterol does not change the Tm value, but instead widens the transition range between frozen and fluid forms of the membrane.

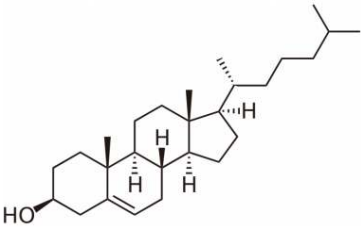

Figure 3.5.6: *Cholesterol*

Lipid Bilayers

The membrane around cells contains many components, including cholesterol, proteins, glycolipids, glycerophospholipids and sphingolipids. The last two of these will, in water, form what is called a lipid bilayer, which serves as a boundary for the cell that is largely impermeable to the movement of most materials across it. With the notable exceptions of water, carbon dioxide, carbon monoxide, and oxygen, most polar/ionic compounds require transport proteins to help them to efficiently navigate across the bilayer. The orderly movement of these compounds is critical for the cell to be able to 1) get food for energy; 2) export materials; 3) maintain osmotic balance; 4) create gradients for secondary transport; 5) provide electromotive force for nerve signaling; and 6) store energy in electrochemical gradients for ATP production (oxidative phosphorylation or photosynthesis). In some cases, energy is required to move the substances (active transport). In other cases, no external energy is required and they move by diffusion through specific cellular channels.

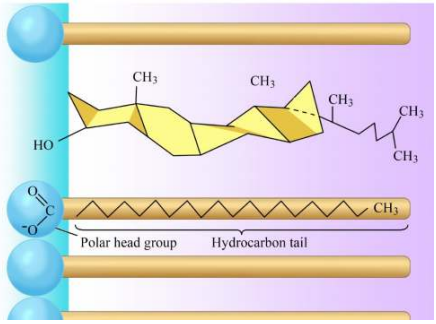

Figure 3.5.7: *Cholesterol in a lipid bilayer.*

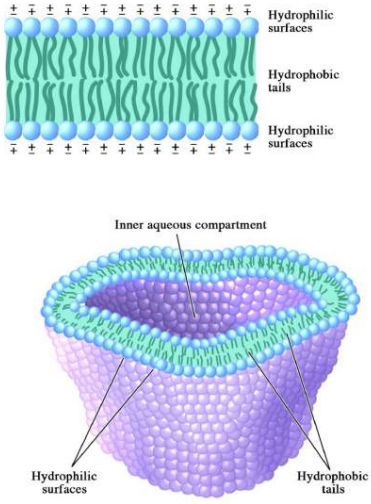

Figure 3.5.8: Lipid bilayer structure.

The spontaneous ability of these compounds to form lipid bilayers is exploited in the formation of artificial membranous structures called liposomes. Liposomes have some uses in delivering their contents into cells via membrane fusion.

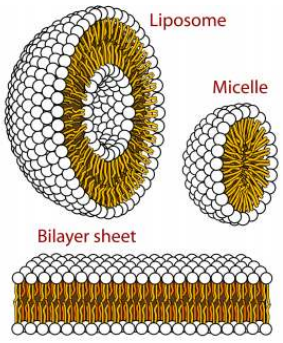

Figure 3.5.9: *Liposome, Micelle, and Bilayer Sheet.*

Membrane Proteins

Other significant components of cellular membranes include proteins. We can put them into several categories. Integral membrane proteins are embedded in the membrane and project through both sides of the lipid bilayer. Peripheral membrane proteins are embedded in or tightly associated with part of the bilayer, but do not project completely through both sides. Associated membrane proteins are found near membranes, but may not be embedded in them. Their association may arise as a result of interaction with other proteins or molecules in the lipid bilayer. Anchored membrane proteins are not themselves embedded in the lipid bilayer, but instead are attached to a molecule (typically a fatty acid) that is embedded in the membrane.

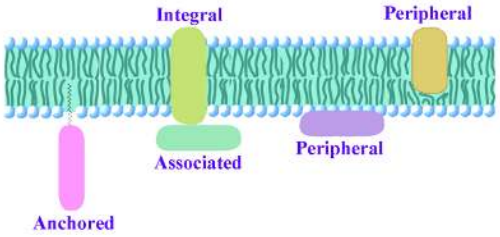

Figure 3.5.10: Types of membrane proteins.

The geometry of the lipid bilayer is such that is hydrophobic on its interior and hydrophilic on the exterior. Such properties also dictate the amino acid side chains of proteins that interact with the bilayer. For most membrane proteins, the polar amino acids are found where the protein projects through the bilayer (interacting with aqueous/polar substances) and the non-polar amino acids are embedded within the non-polar portion of the bilayer containing the fatty acid tails.

Glycolipids and glycoproteins play important roles in cellular identification. Blood types, for example, differ from each other in the structure of the carbohydrate chains projecting out from the surface of the glycoprotein in their membranes.

Cells have hundreds of membrane proteins and the protein composition of a membrane varies with its function and location. Mitochondrial membranes are among the most densely packed with proteins. The plasma membrane has a large number of integral proteins involved in communicating information across the membrane (signaling) or in transporting materials into the cell.

Membrane Transport

Materials, such as food and waste must be moved across a cell's lipid bilayer. There are two means of accomplishing this - passive processes and active processes. Passive processes have as their sole driving force the process of diffusion. In these systems, molecules always move from a higher concentration to a lower concentration. These can occur directly across a membrane (water, oxygen, carbon dioxide, and carbon monoxide) or through special transport proteins (glucose transport proteins of red blood cells, for example). In each case, no cellular energy is expended in the movement of the molecules. On the other hand, active processes require energy to accomplish such transport. A common energy source is ATP (see Na+/K+ ATPase), but many other energy sources are employed. For example, the sodium-glucose transporter uses a sodium gradient as a force for actively transporting glucose into a cell. Thus, it is important to know that not all active transport uses ATP energy. Proteins, such as the sodium-glucose transporter that move two molecules in the same direction across the membrane are called symporters (also called synporters). If the action of a protein in moving ions across a membrane results in a change in charge, the protein is described as electrogenic and if there is no change in charge the protein is described as electro-neutral.

Sodium-Potassium ATPase

Another important integral membrane protein is the Na+/K+ ATPase (Figure 3.5.11), which transports sodium ions out of the cell and potassium ions into the cell. The protein, which is described as an anti-port (molecules moved in opposite directions across the membrane) uses the energy of ATP to create ion gradients that are important both in maintaining cellular osmotic pressure and (in nerve cells) for creating the ion gradients necessary for signal transmission. The transport system moves three atoms of sodium out of the cell and two atoms of potassium into the cell for each ATP hydrolyzed.

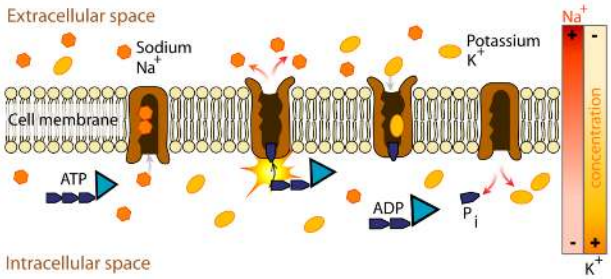

Figure 3.5.11: *Sodium-Potassium ATPase.*

Bacteriorhodopsin

An interesting integral membrane protein is bacteriorhodopsin. The protein has three identical polypeptide chains, each rotated by 120 degrees relative to the others. Each chain has seven transmembrane alpha helices and contains one molecule of retinal (Vitamin A) buried deep within each cavity (shown in purple in lower figure at left). Vitamin A is light sensitive an isomerizes rapidly between a cis and a trans form in the presence of light. The changing conformation of the vitamin A is used to transport protons through the protein and out of the bacterium, creating a proton gradient across the cell membrane, which is used ultimately to make ATP. It is not too difficult to imagine engineering an organism (say a transparent fish) to contain bacteriorhodopsin in its mitochondrial inner membrane. When light is shone upon it, the bacteriorhodopsin could be used to generate a proton gradient (much like electron transport does) and power oxidative phosphorylation. Such a fish would be partly photosynthetic in that it would be deriving energy from light, but would differ from plants in being unable to assimilate carbon dioxide in a series of “dark reactions."

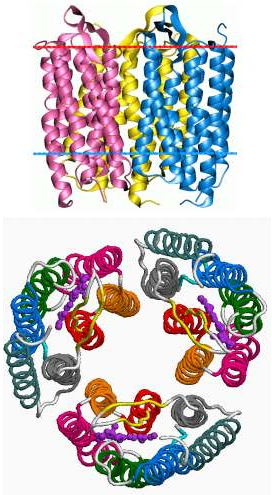

Figure 3.5.12: *Bacteriorhodopsin*

Fat Soluble Vitamins

Other lipids of note include the fat-soluble vitamins - A, D, E, and K. Vitamin A comes in three primary chemical forms, retinol (storage in liver), retinal (role in vision), and retinoic acid (roles in growth and development). Vitamin D (cholecalciferol) plays important roles in the intestinal absorption of calcium and phosphate and thus in healthy bones. Derived from ultimately from cholesterol, the compound can be synthesized in a reaction catalyzed by ultraviolet light. Vitamin E (tocopherol) is the vitamin about which the least is known. It consists of a group of eight fat-soluble compounds of which the alpha-isomer has the most

biological activity. Vitamin K (the name comes from the German for coagulation vitamin) is essential for blood clotting. It is used as a co-factor for the enzyme that modifies prothrombin to increase its affinity for calcium, allowing it to be positioned closer to the site of a wound.

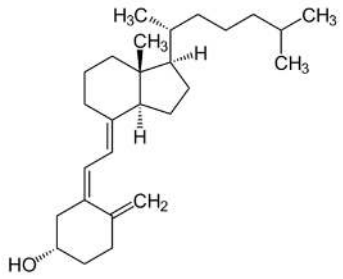

Figure 3.5.13: *Vitamin D*

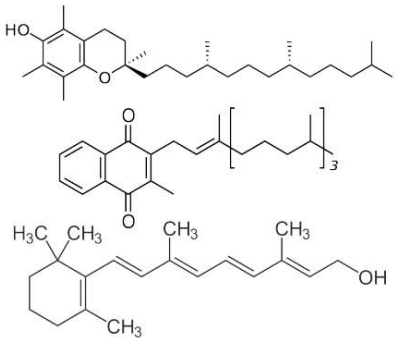

Figure 3.5.14: *Top to bottom - Vitamins E, K, and A*

CHAPTER OVERVIEW

4: Catalysis

In living systems, speed is everything. Providing the reaction speeds necessary to support life are the catalysts, mostly in the form of enzymes.

4.1: Introduction to Catalysis

If there is a magical component to life, an argument can surely be made for it being catalysis. Thanks to catalysis, reactions that could take hundreds of years to complete in the “real world," occur in seconds in the presence of a catalyst. Chemical catalysts, like platinum, speed reactions, but enzymes (which are simply super-catalysts with a twist) put chemical catalysts to shame. To understand enzymatic catalysis, we must first understand energy. In Chapter 2, we noted the tendency for processes to move in the direction of lower energy. Chemical reactions follow this universal trend, but they often have a barrier in place that must be overcome. The secret to catalytic action is reducing the magnitude of that barrier, as we shall see.

Enzyme	Nonenzymatic Half-Life	Uncatalyzed Rate (k_{un} s^{-1})	Catalyzed Rate (k_{cat} s^{-1})	Rate Enhancement (k_{cat} s^{-1}/k_{un} s^{-1})
OMP decarboxylase	78,000,000 years	28×10^{-16}	39	1.4×10^{17}
Staphylococcal nuclease	130,000 years	1.7×10^{-13}	95	5.6×10^{14}
Carboxypeptidase A	7.3 years	3.0×10^{-9}	578	1.9×10^{11}
Ketosteroid isomerase	7 weeks	1.7×10^{-7}	66,000	3.9×10^{11}
Triose phosphate isomerase	1.9 days	4.3×10^{-6}	4,300	1.0×10^{9}
Chorismate mutase	7.4 hours	2.6×10^{-5}	50	1.9×10^{6}
Carbonic anhydrase	5 seconds	1.3×10^{-1}	1×10^{6}	7.7×10^{6}

Figure 4.0.1: *Enzyme Rate Enhancements*

4.2: Activation Energy

Figure 4.1.1 schematically depicts the energy changes that occur during the progression of a simple reaction. In the figure, the energy differences during the reaction are compared for a catalyzed (plot on the right) and an uncatalyzed reaction (plot on the left). Notice that the reactants start at the same energy level for both conditions and that the products end at the same energy for both as well. Thus, the difference in energy between the energy of the ending compounds and the starting compounds is the same in both cases. This is the first important rule to understand any kind of catalysis – catalysts do not change the overall energy of a reaction. Given enough time, a non-catalyzed reaction will get to the same equilibrium as a catalyzed one.

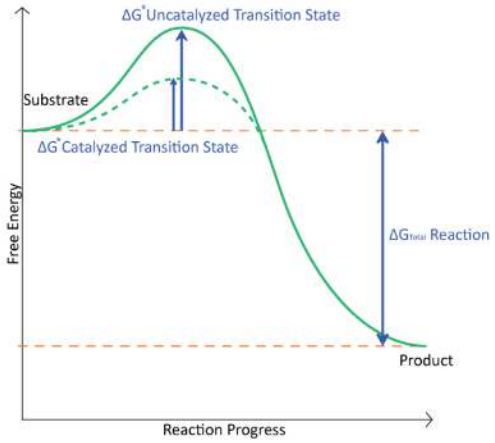

Figure 4.1.1: *Energetic considerations of catalysis*

Another feature to note about catalyzed reactions is the reduced energy barrier (also called the activation energy or free energy of activation) to reach the transition state of the catalyzed reaction. This is the second important point about catalyzed reactions – catalysts work by lowering activation energies of reactions and thus molecules more easily reach the energy necessary to get to the point where the reaction occurs. Note that these reactions are reversible. The extent to which they will proceed is a function of the size of the energy difference between the product and reactant states. The lower the energy of the products compared to the reactants, the larger the percentage of molecules that will be present as products at equilibrium. At equilibrium, of course, no change in concentration of reactants and products occurs because at this point, the forward and reverse reaction rates are the same.

4.3: General Mechanisms of Action

As noted above, enzymes are orders of magnitude more effective (faster) than chemical catalysts. The secret of their success lies in a fundamental difference in their mechanisms of action. Every chemistry student has had hammered into their heads the fact that a catalyst speeds a reaction without being consumed by it. In other words, the catalyst ends up after a reaction just the way it started so it can catalyze other reactions, as well. Enzymes share this property, but in the middle, during the catalytic action, an enzyme is transiently changed. Such changes may be subtle electronic ones or more significant covalent modifications. It is also important to recognize that enzymes are not fixed, rigid structures, but rather are flexible. Flexibility allows movement and movement facilitates alteration of electronic environments necessary for catalysis. Enzymes are, thus, much more effcient than rigid chemical catalysts as a result of their abilities to facilitate the changes necessary to optimize the catalytic process.

4.4: Substrate Binding

Another important difference between the mechanism of action of an enzyme and a chemical catalyst is that an enzyme has binding sites that not only 'grab' the substrate (molecule involved in the reaction being catalyzed), but also place it in a position to be electronically induced to react, either within itself or with another substrate. The enzyme itself may play a role in the electronic induction or the induction may occur as a result of substrates being placed in very close proximity to each other. Chemical catalysts have no such ability to bind substrates and are dependent upon them colliding in the right orientation at or near their surfaces.

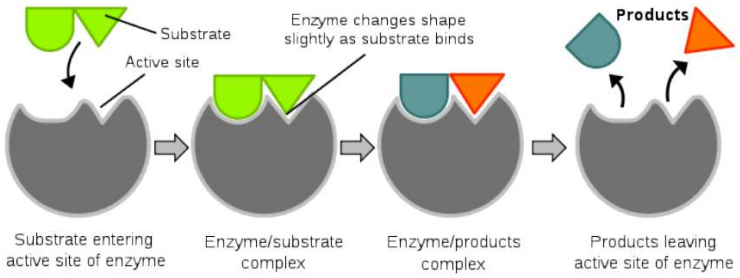

Figure 4.3.1: *Mechanism of induced fit*

4.5: Enzyme Flexibility

As mentioned earlier, a difference between an enzyme and a chemical catalyst is that an enzyme is flexible. Its slight changes in shape (often arising from the binding of the substrate itself) help to position substrates for reaction after they bind. These changes in shape are explained, in part, by Koshland's Induced Fit Model of Catalysis, which illustrates that not only do enzymes change substrates, but that substrates also transiently change enzymes. At the end of the catalysis, the enzyme is returned to its original state. Enzyme flexibility also is important for control of enzyme activity. Two distinct structures are typically described– the T (tight) state, which is a lower activity state and the R (relaxed) state, which has greater activity.

4.6: Active Site

Reactions in enzymes are catalyzed at a specific location known as the ‘ active site ’. Substrate binding sites are located in close physical proximity to the active site and oriented to provide access for the relevant portion of the molecule to the electronic environment of the enzyme where catalysis is initiated.

4.7: Chymotrypsin

Consider the mechanism of catalysis of the enzyme known as chymotrypsin. Found in our digestive system, chymotrypsin's catalytic action is cleaving peptide bonds in proteins and it uses the side chain of a serine in its mechanism of catalysis. Many other protein- cutting enzymes employ a very similar mechanism and they are known collectively as serine proteases. As aprotease, it acts fairly specifically, cutting not all peptide bonds, but only those that are adjacent to specific amino acids in the protein. One of the amino acids it cuts adjacent to is phenylalanine. The enzyme's action occurs in two phases – a fast phase that occurs first and a slower phase that follows. The enzyme has a substrate binding site that includes a region of the enzyme known as the S1 pocket. Let us step through the mechanism by which chymotrypsin cuts adjacent to phenylalanine.

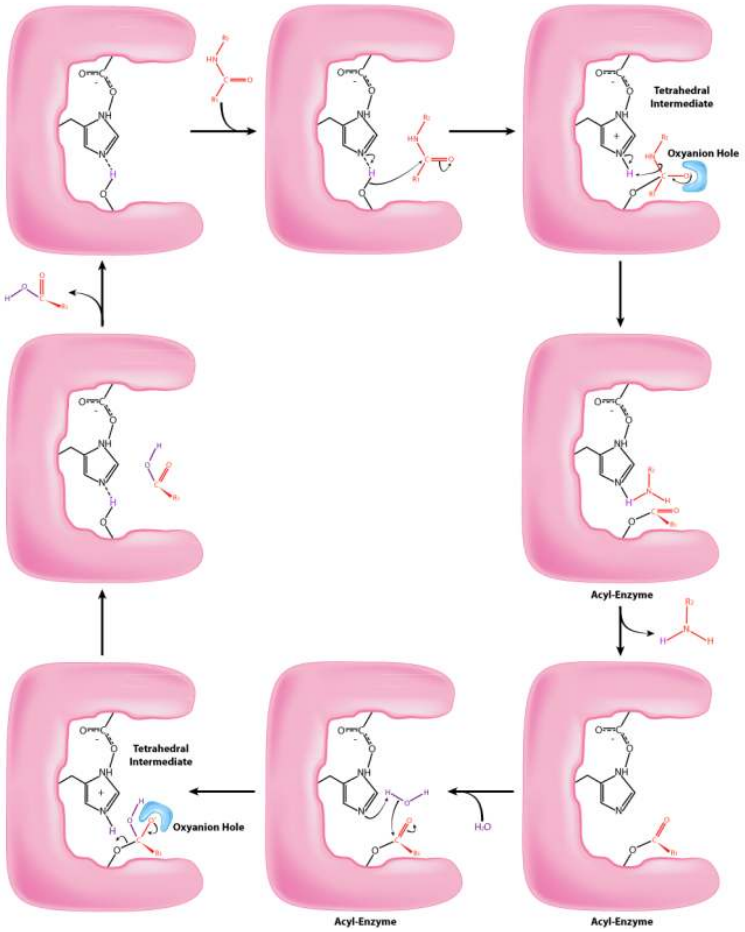

Figure 4.6.1: *Serine protease mechanism*

The process starts with the binding of the substrate in the S1 pocket. The S1 pocket in chymotrypsin has a hydrophobic hole in which the substrate is bound. Preferred substrates will include amino acid side chains that are hydrophobic, like phenylalanine. If an ionized side chain, like that of glutamic acid binds in the S1 pocket, it will quickly exit, much like water would avoid an oily interior. When the proper substrate binds, it stays and its presence induces an ever so slight shift in the shape of the enzyme. This subtle shape change on the binding of the proper substrate starts the steps of the catalysis and is the reason that the enzyme shows specificity for cutting at specific enzyme positions in the target protein. Only amino acids with the side chains that interact well with the S1 pocket start the catalytic wheels turning.

The slight changes in shape of the enzyme upon binding of the proper substrate cause changes in the positioning of three amino acids (aspartic acid, histidine, and serine) in the active site known as the catalytic triad, during the second step of the catalytic action. The shift of the negatively charged aspartic acid towards the electron rich histidine ring favors the abstraction of a proton by the histidine from the hydroxyl group on the side chain of serine, resulting in production of a very reactive alkoxide ion in the active site. Since the active site at this point also contains the polypeptide chain positioned with the phenylalanine side chain embedded in the S1 pocket, the alkoxide ion performs a nucleophilic attack on the peptide bond on the carboxyl side of

phenylalanine sitting in the active site. This reaction, which is the third step of catalysis, breaks the bond and causes two things to happen. First, one end of the original polypeptide is freed and exits the active site. The second is that the end containing the phenylalanine is covalently linked to the oxygen of the serine side chain. At this point we have completed the first (fast) phase of the catalysis.

The second phase of the catalysis by chymotrypsin is slower. It requires that the covalent bond between phenylalanine and serine's oxygen be broken so the peptide can be released and the enzyme can return to its original state. The process starts with entry of water into the active site. Water is attacked in a fashion similar to that of the serine side chain in the first phase, creating a reactive hydroxyl group that performs a nucleophilic attack on the phenylalanine-serine bond, releasing it and replacing the proton on serine. The second peptide is released in the process and the reaction is complete with the enzyme back in its original state.

4.8: Enzyme Parameters

Scientists spend a considerable amount of time characterizing enzymes. To understand how they do this and what the characterizations tell us, we must first understand a few parameters. Imagine I wished to study the reaction catalyzed by an enzyme I have just isolated. I would be interested to understand how fast the enzyme works and how much affinity the enzyme has for its substrate(s).

To perform this analysis, I would perform the following experiment. Into 20 different tubes, I would put enzyme buffer (to keep the enzyme stable), the same amount of enzyme, and then a different amount of substrate in each tube, ranging from tiny amounts in the first tubes to very large amounts in the last tubes. I would let the reaction proceed for a fixed, short amount of time and then I would measure the amount of product contained in each tube. For each reaction, I would determine the velocity of the reaction as the concentration of product found in each tube divided by the time. I would then plot the data on a graph using velocity on the Y-axis and the concentration of substrate on the X-axis.

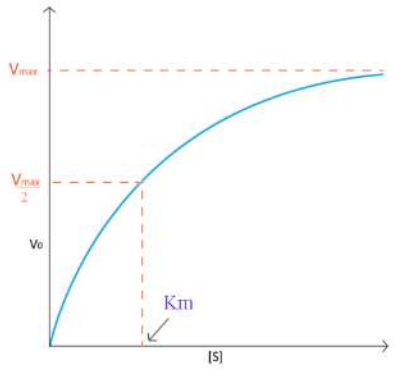

Figure 4.8.1: Velocity vs Substrate concentration plot

Typically, I would generate a curve like that shown on Figure 4.8.1. Notice how the velocity increase is almost linear in the tubes with the lowest amounts of substrate. This indicates that substrate is limiting and the enzyme converts it into product as soon as it can bind it. As the substrate concentration increases, however, the velocity of the reaction in tubes with higher substrate concentration ceases to increase linearly and instead begins to flatten out, indicating that as the substrate concentration gets higher and higher, the enzyme has a harder time keeping up to convert the substrate to product. What is happening is the enzyme is becoming saturated with substrate at higher concentrations of the latter. Not surprisingly, when the enzyme becomes completely saturated with substrate, it will not have to wait for substrate to diffuse to it and will therefore be operating at maximum velocity.

Turnover Number

On a plot of Velocity versus Substrate Concentration (V vs. $[S]$), the maximum velocity (known as $\mathbf{V_{max}}$) is the value on the Y axis that the curve asymptotically approaches. It should be noted that the value of $\mathbf{V_{max}}$ depends on the amount of enzyme used in a reaction. Double the amount of enzyme, double the $\mathbf{V_{max}}$. If one wanted to compare the velocities of two different enzymes, it would be necessary to use the same amounts of enzyme in the different reactions they catalyze. It is desirable to have a measure of velocity that is independent of enzyme concentration. For this, we define the value $\mathbf{K_{cat}}$, also known as the **turnover number**. Mathematically,

$$\mathbf{K_{cat}} = \frac{\mathbf{V_{max}}}{[\mathbf{Enzyme}]} \tag{4.7.1}$$

To determine $\mathbf{K_{cat}}$, one must obviously know the $\mathbf{V_{max}}$ at a particular concentration of enzyme, but the beauty of the term is that it is a measure of velocity independent of enzyme concentration, thanks to the term in the denominator. Kcat is thus a constant for an enzyme under given conditions. The units of $\mathbf{K_{cat}}$ are $\mathbf{time^{-1}}$. An example would be 35/second. This would mean that each molecule of enzyme is catalyzing the formation of 35 molecules of product every second. While that might seem like a high value, there are enzymes known (carbonic anhydrase, for example) that have $\mathbf{K_{cat}}$ values of $\mathbf{10^6}$/second. This astonishing number illustrates clearly why enzymes seem almost magical in their action.

Enzyme	Turnover Number (per second)
Carbonic anhydrase	600,000
3-Ketoesteroid isomerase	280,000
Acetylcholinesterase	25,000
Penicillinase	2,000
Lactate dehydrogenase	1,000
Chymostrypsin	100
DNA Polymerase I	15
Tryptophan synthetase	2
Lysozyme	0.5

Figure 4.8.2: Common turnover numbers

Michaelis constant

Another parameter of an enzyme that is useful is known as $\mathbf{K_M}$, the **Michaelis constant**. What it measures, in simple terms, is the affinity an enzyme has for its substrate. Affinities of enzymes for substrates vary considerably, so knowing $\mathbf{K_M}$ helps us to understand how well an enzyme is suited to the substrate being used. Measurement of KM depends on the measurement of V_{max}. On a V vs. [S] plot, $\mathbf{K_M}$ is determined as the x value that give $\mathbf{V_{max}/2}$. A common mistake students make in describing V_{max} is saying that $\mathbf{K_M = V_{max}/2}$. This is, of course not true. $\mathbf{K_M}$ is a substrate concentration and is the amount of substrate it takes for an enzyme to reach $V_{max}/2$. On the other hand $V_{max}/2$ is a velocity and is nothing more than that. The value of $\mathbf{K_M}$ is inversely related to the affinity of the enzyme for its substrate. High values of $\mathbf{K_M}$ correspond to low enzyme affinity for substrate (it takes more substrate to get to V_{max}). Low $\mathbf{K_M}$ values for an enzyme correspond to high affinity for substrate.

4.9: Perfect Enzymes

Now, if we think about what an ideal enzyme might be, it would be one that has a very high velocity and a very high affinity for its substrate. That is, it wouldn't take much substrate to get to $V_{max}/2$ and the K_{cat} would be very high. Such enzymes would have values of K_{cat}/K_M that are maximum. Interestingly, there are several enzymes that have this property and their maximal values are all approximately the same. Such enzymes are referred to as being "perfect" because they have reached the maximum possible value. Why should there be a maximum possible value of K_{cat}/K_M. The answer is that movement of substrate to the enzyme becomes the limiting factor for perfect enzymes. Movement of substrate by diffusion in water has a fixed rate and that limitation ultimately determines how fast the enzyme can work. In a macroscopic world analogy, factories can't make products faster than suppliers can deliver materials. It is safe to say for a perfect enzyme that the only limit it has is the rate of substrate diffusion in water.

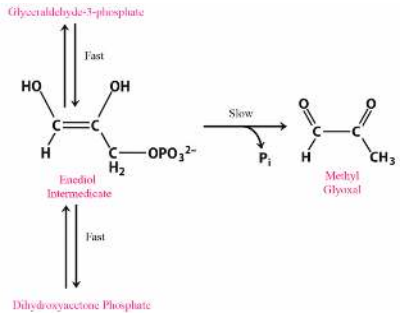

Figure 4.8.1: *Avoidance of formation of an unstable intermediate in triose phosphate isomerase*

Given the "magic" of enzymes alluded to earlier, it might seem that all enzymes should have evolved to be "perfect." There are very good reasons why most of them have not. Speed can be a dangerous thing. The faster a reaction proceeds in catalysis by an enzyme, the harder it is to control. As we all know from learning to drive, speeding causes accident. Just as drivers need to have speed limits for operating automobiles, so too must cells exert some control on the 'throttle' of their enzymes. In view of this, one might wonder then why any cells have evolved any enzymes to perfection. There is no single answer to the question, but a common one is illustrated by the perfect enzyme known as triose phosphate isomerase (TPI), which catalyzes a reaction in glycolysis (figure on previous page). The enzyme appears to have been selected for this ability because at lower velocities, there is breakdown of an unstable enediol intermediate that then readily forms methyl glyoxal, a cytotoxic compound. Speeding up the reaction provides less opportunity for the unstable intermediate to accumulate and fewer undesirable byproducts are made.

4.10: Lineweaver-Burk Plots

The study of enzyme kinetics is typically the most math intensive component of biochemistry and one of the most daunting aspects of the subject for many students. Although attempts are made to simplify the mathematical considerations, sometimes they only serve to confuse or frustrate students. Such is the case with modified enzyme plots, such a Lineweaver-Burk (Figure 4.9.1). Indeed, when presented by professors as simply another thing to memorize, who can blame students. In reality, both of these plots are aimed at simplifying the determination of parameters, such as K_M and V_{max}. In making either of these modified plots, it is important to recognize that the same data is used as in making a V vs. $[S]$ plot. The data are simply manipulated to make the plotting easier.

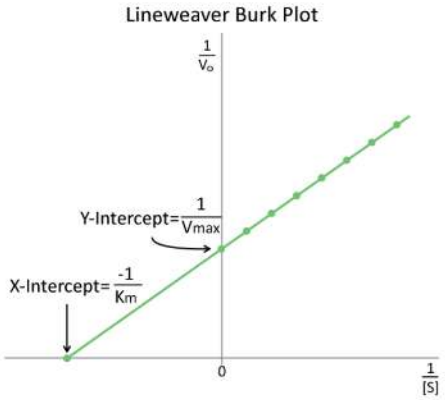

Figure 4.9.1: *Line-Weaver Burk Plot*

For a Lineweaver-Burk, the manipulation is using the reciprocal of the values of both the velocity and the substrate concentration. The inverted values are then plotted on a graph as $1/V$ vs. $1/[S]$. Because of these inversions, Lineweaver-Burk plots are commonly referred to as 'double-reciprocal' plots. As can be seen at left, the value of K_M on a Lineweaver Burk plot is easily determined as the negative reciprocal of the x-intercept , whereas the V_{max} is the inverse of the y-intercept. Other related manipulation of kinetic data include Eadie-Hofstee diagrams, which plot V vs V/[S] and give V_{max} as the Y-axis intercept with the slope of the line being $-K_M$.

4.11: Enzyme Inhibition

Inhibition of specific enzymes by drugs can be medically useful. Understanding the mechanisms of enzyme inhibition is therefore of considerable importance. We will discuss four types of enzyme inhibition – competitive, non- competitive, uncompetitive, and suicide. Of these, the first three types are reversible. The last one is not.

Competitive Inhibition

Probably the easiest type of enzyme inhibition to understand is competitive inhibition and it is the one most commonly exploited pharmaceutically. Molecules that are competitive inhibitors of enzymes resemble one of the normal substrates of an enzyme. An example is methotrexate, which resembles the folate substrate of the enzyme dihydrofolate reductase (DHFR). This enzyme normally catalyzes the reduction of folate, an important reaction in the metabolism of nucleotides. When the drug methotrexate is present, some of the enzyme binds to it instead of to folate and during the time methotrexate is bound, the enzyme is inactive and unable to bind folate. Thus, the enzyme is inhibited. Notably, the binding site on DHFR for methotrexate is the active site, the same place that folate would normally bind. As a result, methotrexate 'competes' with folate for binding to the enzyme. The more methotrexate there is, the more effectively it competes with folate for the enzyme's active site. Conversely, the more folate there is, the less of an effect methotrexate has on the enzyme because folate outcompetes it.

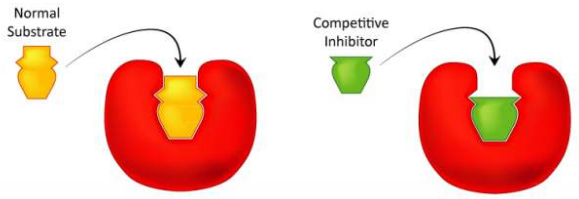

Figure 4.10.1: *Competitive Inhibition*

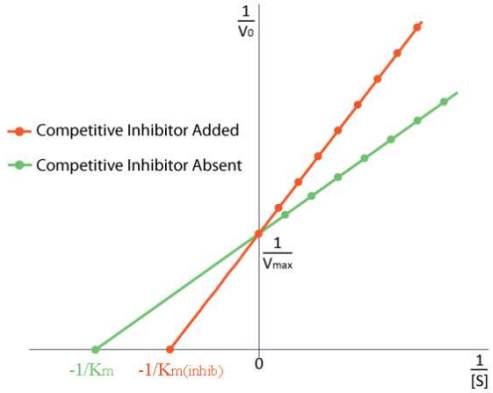

Figure 4.10.2: *Line-Weaver Burk Plot of competitive inhibition*

No Effect On V_{MAX}

How do we study competitive inhibition. It is typically done as follows. First one performs a set of V vs. [S] reactions without inhibitor (20 or so tubes, with buffer and constant amounts of enzyme, varying amounts of substrate, equal reaction times). V vs. [S] is plotted, as well as 1/V vs. 1/[S], if desired. Next, a second set of reactions is performed in the same manner as before, except that a fixed amount of the methotrexate inhibitor is added to each tube. At low concentrations of substrate, the inhibitor competes for the enzyme effectively, but at high concentrations of substrate, the inhibitor will have a much reduced effect, since the substrate outcompetes it, due to its higher concentration (remember that the inhibitor is at fixed concentration). Graphically, the results of these experiments are shown above. Notice that at high substrate concentrations, the competitive inhibitor has essentially no effect, causing the Vmax for the enzyme to remain unchanged. To reiterate, this is due to the fact that at high substrate concentrations, the inhibitor doesn't compete well. However, at lower substrate concentrations it does.

Increased KM

Note that the apparent KM of the enzyme for the substrate increases (-1/KM gets closer to zero - red line above) when the inhibitor is present, thus illustrating the better competition of the inhibitor at lower substrate concentrations. It may not be obvious why we call the changed KM the apparent KM of the enzyme. The reason is that the inhibitor doesn't actually change the enzyme's affinity for the folate substrate. It only appears to do so. This is because of the way that competitive inhibition works. When the competitive inhibitor binds the enzyme, it is effectively 'taken out of action.' Inactive enzymes have NO affinity for substrate and no activity either. We can't measure KM for an inactive enzyme.

The enzyme molecules that are not bound by methotrexate can, in fact, bind folate and are active. Methotrexate has no effect on them and their KM values are unchanged. Why then, does KM appear higher in the presence of a competitive inhibitor. The reason is that the competitive inhibitor is reducing the amount of active enzyme at lower concentrations of substrate. When the amount of enzyme is reduced, one must have more substrate to supply the reduced amount of enzyme sufficiently to get to Vmax/2.

It is worth noting that in competitive inhibition, the percentage of inactive enzyme changes drastically over the range of [S] values used. To start, at low [S] values, the greatest percentage of the enzyme is inhibited. At high [S], no significant percentage of enzyme is inhibited. This is not always the case, as we shall see in non-competitive inhibition.

Non-Competitive Inhibition

A second type of inhibition employs inhibitors that do not resemble the substrate and bind not to the active site, but rather to a separate site on the enzyme (rectangular site below). The effect of binding a non-competitive inhibitor is significantly different from binding a competitive inhibitor because there is no competition. In the case of competitive inhibition, the effect of the inhibitor could be reduced and eventually overwhelmed with increasing amounts of substrate. This was because increasing substrate made increasing percentages of the enzyme active. With non-competitive inhibition, increasing the amount of substrate has no effect on the percentage of enzyme that is active. Indeed, in non-competitive inhibition, the percentage of enzyme inhibited remains the same through all ranges of [S].

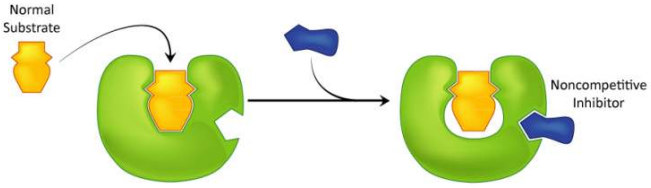

Figure 4.10.3: *Non-competitive Inhibition*

This means, then, that non-competitive inhibition effectively reduces the amount of enzyme by the same fixed amount in a typical experiment at every substrate concentration used The effect of this inhibition is shown above. As you can see, Vmax is reduced in non-competitive inhibition compared to uninhibited reactions. This makes sense if we remember that Vmax is dependent on the amount of enzyme present. Reducing the amount of enzyme present reduces Vmax. In competitive inhibition, this doesn't occur detectably, because at high substrate concentrations, there is essentially 100% of the enzyme active and the Vmax appears not to change. Additionally, KM for non-competitively inhibited reactions does not change from that of uninhibited reactions. This is because, as noted previously, one can only measure the KM of active enzymes and KM is a constant for a given enzyme.

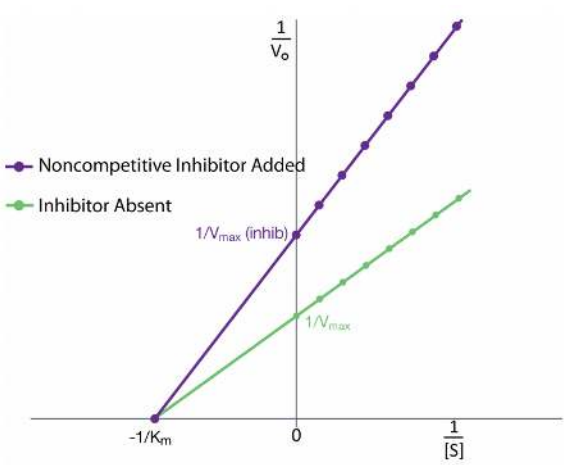

Figure 4.10.4: *Line-Weaver Burk Plot of noncompetitive inhibition*

Uncompetitive Inhibition

A third type of enzymatic inhibition is that of uncompetitive inhibition, which has the odd property of a reduced Vmax as well as a reduced KM. The explanation for these seemingly odd results is rooted in the fact that the uncompetitive inhibitor binds only to the enzyme-substrate (ES) complex. The inhibitor-bound complex forms mostly under concentrations of high substrate and the ES-I complex cannot release product while the inhibitor is bound, thus explaining the reduced Vmax.

The reduced KM is a bit harder to conceptualize. The answer lies in the fact that the inhibitor-bound complex effectively reduces the concentration of the ES complex. By Le Chatelier's Principle, a shift occurs to form additional ES complex, resulting in less free enzyme and more enzyme in the forms ES and ESI (ES with inhibitor). Decreases in free enzyme correspond to an enzyme with greater affinity for its substrate. Thus, paradoxically, uncompetitive inhibition both decreases Vmax and increases an enzyme's affinity for its substrate.

Suicide Inhibition

In contrast to the first three types of inhibition, which involve reversible binding of the inhibitor to the enzyme, suicide inhibition is irreversible because the inhibitor becomes covalently bound to the enzyme during the inhibition and thus cannot be removed. Suicide inhibition rather closely resembles competitive inhibition because the inhibitor generally resembles the substrate and binds to the active site of the enzyme. The primary difference is that the suicide inhibitor is chemically reactive in the active site and makes a bond with it that precludes its removal. Such a mechanism is that employed by penicillin (Figure 4.10.5), which covalently links to the bacterial enzyme, D-D transpeptidase and stops it from functioning. Since the normal function of the enzyme is to make a bond necessary for the peptido-glycan complex of the bacterial cell wall, the cell wall cannot properly form and bacteria cannot reproduce. If one were to measure the kinetics of suicide inhibitors under conditions where there was more enzyme than inhibitor, they would resemble non-competitive inhibition's kinetics because both involve reducing the amount of active enzyme by a fixed amount in a set of reactions.

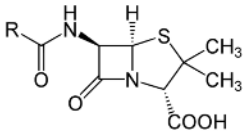

Figure 4.10.5: *Penicillin*

4.12: Control of Enzymes

It is appropriate that we talk at this point about mechanisms cells use to control enzymes. There are four general methods that are employed. They include

1. allosterism
2. covalent modification
3. access to substrate
4. control of enzyme synthesis/breakdown

Some enzymes are controlled by more than one of these methods.

Allosterism

The term allosterism refers to the fact that the activity of certain enzymes can be affected by the binding of small molecules to the enzyme. In allostery, the molecules that are binding are non-substrate molecules that bind at a place on the enzyme other than the active site.

An excellent example of allosteric control is the regulation of HMG-CoA reductase, which catalyzes an important reaction in the pathway leading to the synthesis of cholesterol. Binding of cholesterol to the enzyme reduces the enzyme's activity significantly. Cholesterol is not a substrate for the enzyme, but, notably, is the end-product of the pathway that HMG-CoA catalyzes a reaction in. When enzymes are inhibited by an end-product of the pathway in which they participate, they are said to be feedback inhibited.

Feedback inhibition always operates by allosterism and further, provides important and efficient control of an entire pathway. By inhibiting an early enzyme in a pathway, the flow of materials for the entire pathway is stopped or reduced, assuming there are not alternate supply methods. In the cholesterol biosynthesis pathway, stopping this one enzyme has the effect of shutting off (or at least slowing down) the entire pathway.

Another excellent example is the enzyme aspartate transcarbamoylase (ATCase), which catalyzes an early reaction in the synthesis of pyrimidine nucleotides. This enzyme has two allosteric effectors, ATP and CTP, that are not substrates and that bind at a regulatory site on the enzyme that is apart from the catalytic, active site. CTP, which is the end-product of the pathway, is a feedback inhibitor of the enzyme. ATP, on the other hand, acts to activate the enzyme when it binds to it.

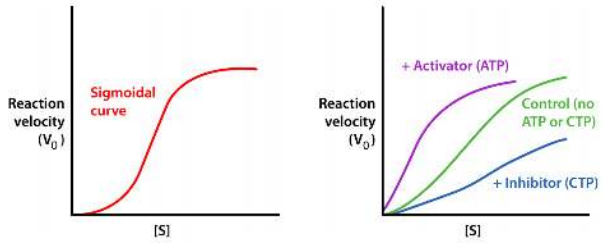

Figure 4.11.1: *Allosteric effects of ATCase*

Allosterically, regulation of these enzymes works by inducing different physical states (shapes, as it were) that affect their ability to bind to substrate. When an enzyme is inhibited by binding an effector, it is converted to the T (also called tight) state, it has a reduced affinity for substrate and it is through this means that the reaction is slowed. On the other hand, when an enzyme is activated by effector binding, it converts to the R (relaxed) state and binds substrate much more readily. When no effector is present, the enzyme may be in a mixture of T and R state. The V vs. S plot of allosteric enzymes resembles the oxygen binding curve of hemoglobin (see HERE). Even though hemoglobin is not an enzyme and is thus not catalyzing a reaction, the similarity of the plots is not coincidental. In both cases, the binding of an external molecule is being measured – directly by the hemoglobin plot and indirectly by the enzyme plot, since substrate binding is a factor in enzyme reaction velocity.

Covalent Control of Enzymes

Some enzymes are synthesized in a completely inactive form and their activation requires covalent bonds in them to be cleaved. Such inactive forms of enzymes are called zymogens. Examples include the proteins involved in blood clotting and proteolytic enzymes of the digestive system, such as trypsin, chymotrypsin, and others. The zymogenic forms of these enzymes are known as

trypsinogen and chymotrypsinogen, respectively. Synthesizing some enzymes in an inactive form makes very good sense when an enyzme's activity might be harmful to the tissue where they are being made. For example, the painful condition known as pancreatitis arises when digestive enzymes made in the pancreas are activated too soon and end up attacking the pancreas.

Blood clotting involves polymerization of a protein known as fibrin. Since random formation of fibrin is extremely hazardous (heart attack/stroke), the body synthesizes fibrin as a zymogen (fibrinogen) and its activation results from a "cascade" of activations of proteases that arise when a signal is received from a wound. Similarly, removal of fibrin clots is also controlled by a zymogen (plasminogen), since random clot removal would also be hazardous.

Another common mechanism for control of enzyme activity by covalent modification is phosphorylation. The phosphorylation of enzymes (on the side chains of serine, threonine or tyrosine residues) is carried out by protein kinases. Enzymes activated by phosphorylation can be regulated by the addition of phosphate groups by kinases or their removal by phosphatases.

Other Controls of Enzymes

Other means of controlling enzymes relate to access to substrate (substrate-level control) and control of enzyme synthesis. Hexokinase is an enzyme that is largely regulated by availability of its substrate, glucose. When glucose concentration is low, the product of the enzyme's catalysis, glucose-6-phosphate, accumulates and inhibits the enzyme's function.

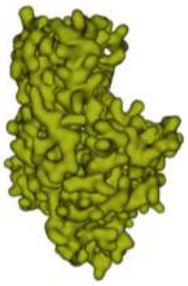

Figure 4.11.2: *Hexokinase - not bound to substrate*

Regulation of enzymes by controlling their synthesis is covered later in the book in the discussion relating to control of gene expression.

4.13: Ribozymes

Proteins do not have a monopoly on acting as biological catalysts. Certain RNA molecules are also capable of speeding reactions. The most famous of these molecules was discovered by Tom Cech in the early 1980s. Studying excision of an intron in Tetrahymena, Cech was puzzled at his inability to find any proteins catalyzing the process. Ultimately, the catalysis was recognized as coming from the intron itself. It was a self-splicing RNA and since then, many other examples of catalytic RNAs capable of cutting other RNAs have been found.

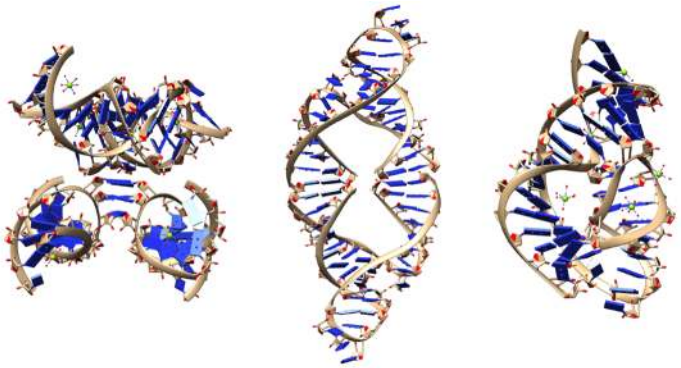

Figure 4.12.1: Image showing the diversity of ribozyme structures. From left to right: leadzyme, hammerhead ribozyme, twister ribozyme. from Wikipedia (CC-BY-SA-4.0 and credit: Lucasharr).

Ribozymes, however, are not rarities of nature. The protein- making ribosomes of cells are essentially giant ribozymes. The 23S rRNA of the prokaryotic ribosome and the 28S rRNA of the eukaryotic ribosome catalyze the formation of peptide bonds. Ribozymes are also important in our understanding of the evolution of life on Earth. They have been shown to be capable via selection to evolve self-replication. Indeed, ribozymes actually answer a chicken/egg dilemma - which came first, enzymes that do the work of the cell or nucleic acids that carry the information required to produce the enzymes. As both carriers of genetic information and catalysts, ribozymes are likely both the chicken and the egg in the origin of life.

CHAPTER OVERVIEW

5: Flow of Genetic Information

As the cell's so-called blueprint, DNA must be copied to pass on to new cells and its integrity safeguarded. The information in the DNA must also be accessed and transcribed to make the RNA instructions that direct the synthesis of proteins.

5.1: DNA Replication

The only way to make new cells is by the division of pre-existing cells. This means that all organisms depend on cell division for their continued existence. DNA, as you know, carries the genetic information that each cell needs. Each time a cell divides, all of its DNA must be copied faithfully so that a copy of this information can be passed on to the daughter cell. This process is called DNA replication. Before examining the actual process of DNA replication, it is useful to think about what it takes to accomplish this task successfully. Consider the challenges facing a cell in this process:

- The sheer number of nucleotides to be copied is enormous: e.g., in human cells, on the order of several billion.
- A double-helical parental DNA molecule must be unwound to expose single strands of DNA that can serve as templates for the synthesis of new DNA strands.
- This unwinding must be accomplished without introducing significant topological distortion into the molecule.
- The unwound single strands of DNA must be kept from coming back together long enough for the new strands to be synthesized.
- DNA polymerases cannot begin synthesis of a new DNA strand de novo and require a free 3' OH to which they can add DNA nucleotides.
- DNA polymerases can only extend a strand in the 5' to 3' direction. The 5' to 3' extension of both new strands at a single replication fork means that one of the strands is made in pieces.
- The use of RNA primers requires that the RNA nucleotides must be removed and replaced with DNA nucleotides and the resulting DNA fragments must be joined.
- Ensuring accuracy in the copying of so much information.

With this in mind, we can begin to examine how cells deal with each of these challenges. Our understanding of the process of DNA replication is derived from studies using bacteria, yeast, and other systems, such as Xenopus eggs. These investigations have revealed that DNA replication is carried out by the action of a large number of proteins that act together as a complex protein machine called the replisome. Numerous proteins involved in replication have been identified and characterized, including multiple different DNA polymerases in both prokaryotes and eukaryotes. Although the specific proteins involved are different in bacteria and eukaryotes, it is useful to understand the basic considerations that are relevant in all cells, before attempting to address the details of each system.

A generalized account of the steps in DNA replication is presented below, focused on the challenges mentioned above.

- The sheer number of nucleotides to be copied is enormous.

For example, in human cells, the number of nucleotides to be copied is on the order of several billion. Even in bacteria, the number is in the millions. Cells, whether bacterial or eukaryotic, have to replicate all of their DNA before they can divide. In cells like our own, the vast amount of DNA is broken up into many chromosomes, each of which is composed of a linear strand of DNA. In cells like those of E. coli, there is a single circular chromosome.

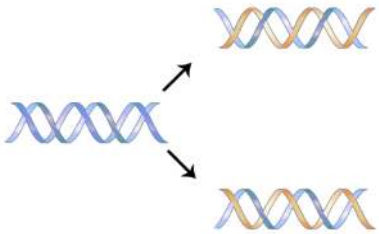

Figure 5.1.1: *Semi-conservative DNA replication*

In either situation, DNA replication is initiated at sites called origins of replication. These are regions of the DNA molecule that are recognized by special origin recognition proteins that bind the DNA. The binding of these proteins helps open up a region of single-stranded DNA where the synthesis of new DNA can begin. In the case of E. coli, there is a single origin of replication on its circular chromosome. In eukaryotic cells, there may be many thousands of origins of replication, with each chromosome having hundreds. DNA replication is thus initiated at multiple points along each chromosome in eukaryotes as shown in Figure 5.1.3. Electron micrographs of replicating DNA from eukaryotic cells show many replication “bubbles" on a single chromosome.

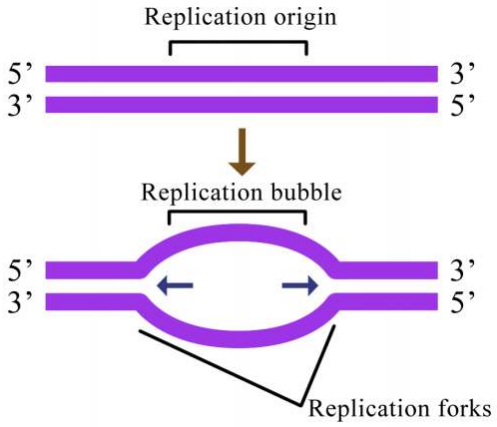

Figure 5.1.2: *Image of a replication bubble*

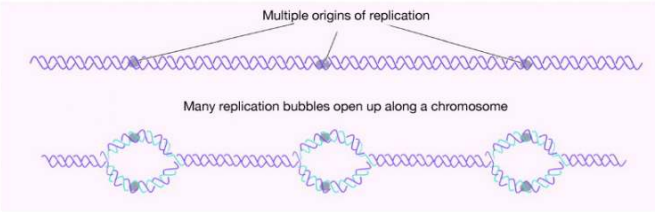

Figure 5.1.3: *Multiple replication bubbles*

This makes sense in light of the large amount of DNA that there is to be copied in cells like our own, where beginning at one end of each chromosome and replicating all the way through to the other end from a single origin would simply take too long. This is despite the fact that the DNA polymerases in human cells are capable of building new DNA strands at the very respectable rate of about 50 nucleotides per second!

- A double-helical parental molecule must be unwound to expose single strands of DNA that can serve as templates for the synthesis of new DNA strands.

Once a small region of the DNA is opened up at each origin of replication, the DNA helix must be unwound to allow replication to proceed. How are the strands of the parental DNA double helix separated? The unwinding of the DNA helix requires the action of an enzyme called helicase. Helicase uses the energy released when ATP is hydrolyzed to unwind the DNA helix. Note that each replication bubble is made up of two replication forks that "move" or open up, in opposite directions. At each replication fork, the parental DNA strands must be unwound to expose new sections of single-stranded template.

- This unwinding must be accomplished without introducing topological distortion into the molecule.

What is the effect of unwinding one region of the double helix? Unwinding the helix locally causes over-winding or topological distortion of the DNA ahead of the unwound region. The DNA ahead of the unwound helix has to rotate, or it will get twisted on itself. How is this problem solved? Enzymes called topoisomerases can relieve the topological stress caused by local unwinding of the double helix. They do this by cutting the DNA and allowing the strands to swivel around each other to release the tension before rejoining the ends. In E. coli, the topoisomerase that performs this function is called gyrase.

- The unwound single strands of DNA must be kept from coming back together long enough for the new strands to be synthesized.

Once the two strands of the parental DNA molecule are separated, they must be prevented from going back together to form double-stranded DNA. To ensure that unwound regions of the parental DNA remain single-stranded and available for copying, the separated strands of the parental DNA are bound by many molecules of a protein called single-strand DNA binding protein (SSB).

- DNA polymerases cannot begin synthesis of a new DNA strand de novo and require a free 3' OH to which they can add DNA nucleotides.

Although single-stranded parental DNA is now available for copying, DNA polymerases cannot begin synthesis of a complementary strand de novo. This is because all DNA polymerases can only add new nucleotides to the 3' end of a pre- existing chain. This means that some enzyme other than a DNA polymerase must first make a small region of nucleic acid, complementary to the parental strand, that can provide a free 3' OH to which DNA polymerase can add a deoxyribonucleotide.

This task is accomplished by an enzyme called a *primase*, which assembles a short stretch of RNA, called the primer, across from the parental DNA template. This provides a short base-paired region with a free 3'OH group to which DNA polymerase can add the first new DNA nucleotide (see figure on previous page). Once a primer provides a free 3'OH for extension, other proteins get into the act. These proteins are involved in loading the DNA polymerase onto the primed template and help to keep it attached to the DNA once it's on.

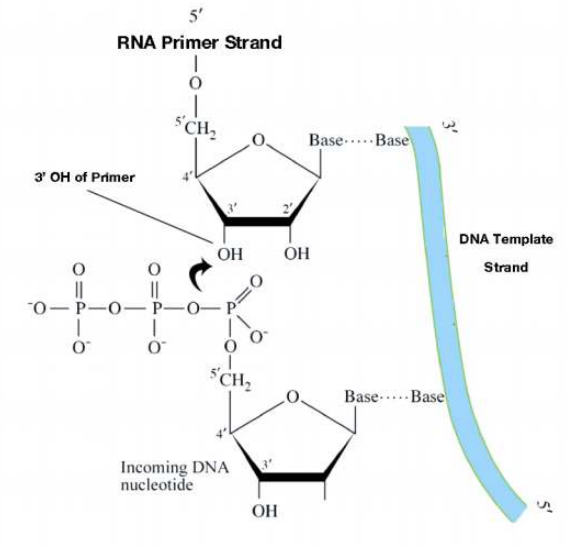

Figure 5.1.4: *Addition of a nucleotide to a growing strand*

The first of these is the clamp loader. As its name suggests, the clamp loader helps to load a protein complex called the sliding clamp onto the DNA at the replication fork. The sliding clamp is then joined by the DNA Polymerase. The function of the sliding clamp is to increase the processivity of the DNA polymerase. This is a fancy way of saying that it keeps the polymerase associated with the replication fork by preventing it from falling off - in fact, the sliding clamp has been described as a seat-belt for the DNA polymerase.

The DNA polymerase is now poised to start synthesis of the new DNA strand (in E. coli, the primary replicative polymerase is called DNA polymerase III). As you already know, the synthesis of new DNA is accomplished by the addition of new nucleotides complementary to those on the parental strand. DNA polymerase catalyzes the reaction by which an incoming deoxyribonucleotide is added onto the 3' end of the previous nucleotide, starting with the 3'OH on the end of the RNA primer.

The 5' phosphate on each incoming nucleotide is joined by the DNA polymerase to the 3' OH on the end of the growing nucleic acid chain. As we already noted, the new DNA strands are synthesized by the addition of DNA nucleotides to the end of an RNA primer. The new DNA molecule thus has a short piece of RNA at the beginning.

- DNA polymerases can only extend a strand in the 5' to 3' direction. The 5' to 3' growth of both new strands means that one of the strands is made in pieces.

We have noted that DNA polymerase can only build a new DNA strand in the 5' to 3' direction. We also know that the two parental strands of DNA are antiparallel. This means that at each replication fork, one new strand, called the leading strand can be synthesized continuously in the 5' to 3' direction because it is being made in the same direction that the replication fork is opening up.

The synthesis of the other new strand, called the lagging strand, requires that multiple RNA primers must be laid down and the new DNA be made in many short pieces that are later joined.These short nucleic acid pieces, each composed of a small stretch of RNA primer and about 1000-2000 DNA nucleotides, are called Okazaki fragments, for Reiji Okazaki, the scientist who first demonstrated their existence.

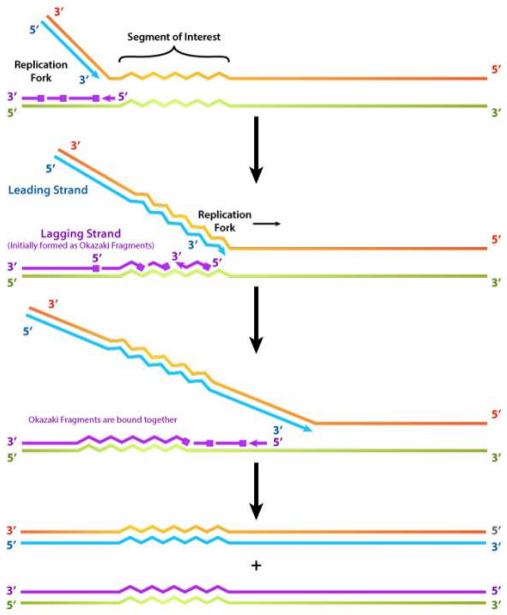

Figure 5.1.5: *Leading and lagging strand replication*

- The use of RNA primers requires that the RNA nucleotides must be removed and replaced with DNA nucleotides.
- We have seen that each newly synthesized piece of DNA starts out with an RNA primer, effectively making a new nucleic acid strand that is part RNA and part DNA. The finished DNA strand cannot be allowed to have pieces of RNA attached. o the RNA nucleotides are removed and the gaps are filled in with DNA nucleotides (by DNA polymerase I in E. coli). The DNA pieces are then joined together by the enzyme DNA ligase.

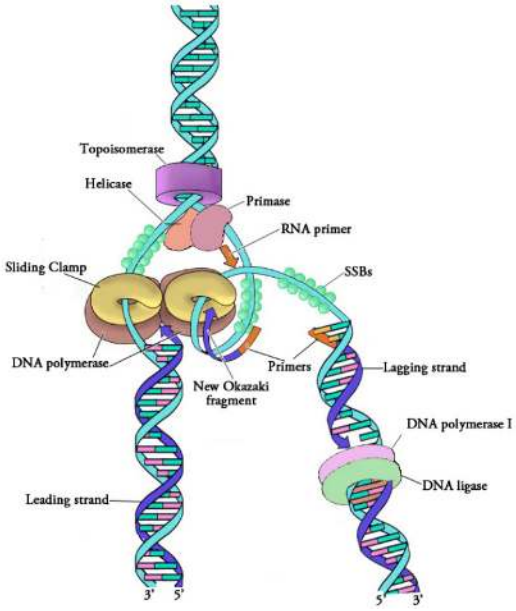

Figure 5.1.6: *Proteins at a replication fork*

The steps outlined above essentially complete the process of DNA replication. Figure 5.1.6 shows a replication fork, complete with the associated proteins that form the replisome.

- Ensuring accuracy in the copying of so much information

How accurate is the copying of information in the DNA by DNA polymerase? As you are aware, changes in DNA sequence (mutations) can change the amino acid sequence of the encoded proteins and that this is often, though not always, deleterious to the

functioning of the organism. When billions of bases in DNA are copied during replication, how do cells ensure that the newly synthesized DNA is a faithful copy of the original information?

DNA polymerases, as we have noted earlier, work fast (averaging 50 bases a second in human cells and up to 20 times faster in E. coli). Yet, both human and bacterial cells seem to replicate their DNA quite accurately. This is because of the proof-reading function of DNA polymerases. The proof-reading function of a DNA polymerase enables the polymerase to detect when the wrong base has been inserted across from a template strand, back up and remove the mistakenly inserted base. This is possible because the polymerase is a dual-function enzyme. It can extend a DNA chain by virtue of its 5' to 3' polymerase activity but it can also backtrack and remove the last inserted base because it has a 3' to 5' exonuclease activity (an exonuclease is an enzyme that removes bases, one by one, from the ends of nucleic acids). The exonuclease activity of the DNA polymerase allows it to excise a wrongly inserted base, after which the polymerase activity inserts the correct base and proceeds with extending the strand.

In other words, DNA polymerase is monitoring its own accuracy (also termed its fidelity) as it makes new DNA, correcting mistakes immediately before moving on to add the next base. This mechanism, which operates during DNA replication, corrects many errors as they occur, reducing by about 100-fold the mistakes made when DNA is copied.

5.2: DNA Repair

Maintaining the Integrity of the Cell's Information: DNA Repair

In the last section we considered the ways in which cells deal with the challenges associated with replicating their DNA, a vital process for all cells. It is evident that if DNA is the master copy of instructions for an organism, then it is important not to make mistakes when copying the DNA to pass on to new cells. Although proofreading by DNA polymerases greatly increases the accuracy of replication, there are additional mechanisms in cells to further ensure that newly replicated DNA is a faithful copy of the original, and also to repair damage to DNA during the normal life of a cell.

All DNA suffers damage over time, from exposure to ultraviolet and other radiation, as well as from various chemicals in the environment. Even chemical reactions naturally occurring within cells can give rise to compounds that can damage DNA. As you already know, even minor changes in DNA sequence, such as point mutations can sometimes have far-reaching consequences. Likewise, unrepaired damage caused by radiation, environmental chemicals or even normal cellular chemistry can interfere with the accurate transmission of information in DNA. Maintaining the integrity of the cell's "blueprint" is of vital importance and this is reflected in the numerous mechanisms that exist to repair mistakes and damage in DNA.

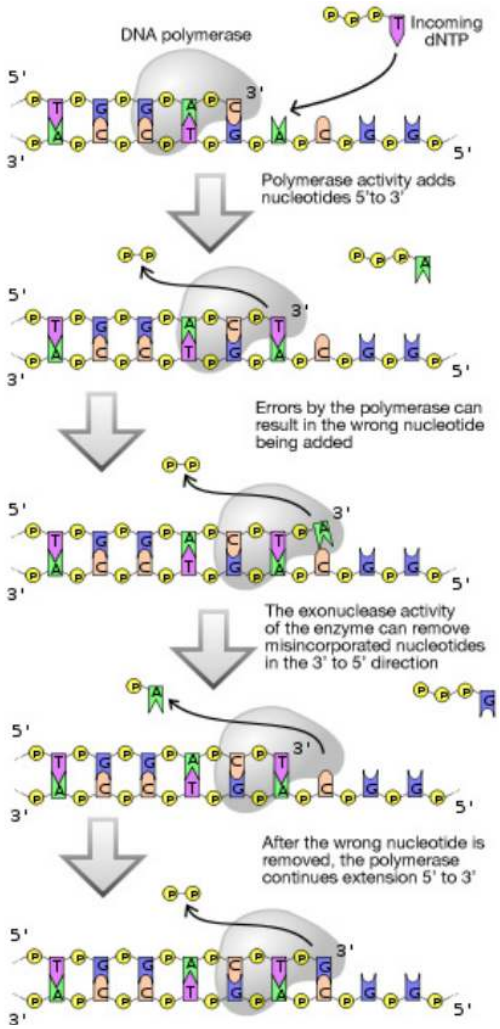

Figure **5.2.1**: *Example of DNA correction*

Post-Replicative Mismatch Repair

We earlier discussed proof-reading by DNA polymerases during replication. Does proofreading eliminate all errors made during replication. No. While proof-reading significantly reduces the error rate, not all mistakes are fixed on the fly by DNA polymerases. What mechanisms exist to correct the replication errors that are missed by the proof-reading function of DNA polymerases.

Errors that slip by proofreading during replication can be corrected by a mechanism called mismatch repair. While the error rate of DNA replication is about one in 10^7 nucleotides in the absence of mismatch repair, this is further reduced a hundred-fold to one in \)10^9\) nucleotides when mismatch repair is functional.

What are the tasks that a mismatch repair system faces. It must:

- Scan newly made DNA to see if there are any mispaired bases (e.g., a G paired to a T)
- Identify and cut out the region of the mismatch.
- Correctly fill in the gap created by the excision of the mismatch region.

Importantly, the mismatch repair system must have a means to distinguish the newly made DNA strand from the template strand, if replication errors are to be fixed correctly. In other words, when the mismatch repair system encounters an A-G mispair, for example, it must know whether the A should be removed and replaced with a C or if the G should be removed and replaced with a T.

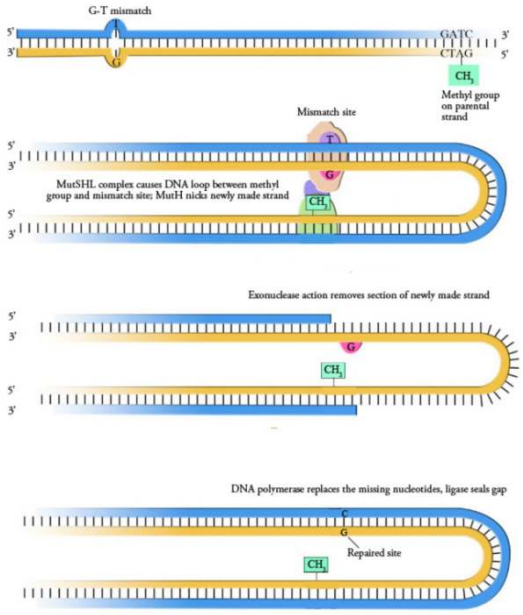

Figure 5.2.1: *Mismatch repair*

Mismatch repair has been well studied in bacteria, and the proteins involved have been identified. Eukaryotes have a mismatch repair system that repairs not only single base mismatches but also insertions and deletions. In bacteria, mismatch repair proteins are encoded by a group of genes collectively known as the mut genes. Some of the most important components of the mismatch repair machinery are the proteins MutS, L and H. MutS acts to recognize the mismatch, while MutL and MutH are recruited to the mismatch site by the binding of Mut S, to help cut out the region containing the mismatch. A DNA polymerase and ligase fill in the gap and join the ends, respectively.

But how does the mismatch repair system distinguish between the original and the new strands of DNA? In bacteria, the existence of a system that methylates the DNA at GATC sequences is the solution to this problem. *E.coli* has an enzyme that adds methyl groups on the to adenines in GATC sequences. Newly replicated DNA lacks thismethylation and thus, can be distinguished from the template strand, which is methylated. In Figure 5.2.2, the template strand shown in yellow is methylated at GATC sequences. The mismatch repair proteins selectively replace the strand lacking methylation, shown in blue in the figure, thus ensuring that it is mistakes in the newly made strand that are removed and replaced. Because methylation is the criterion that enables the mismatch repair system to choose the strand that is repaired, the bacterial mismatch repair system is described as being methyl-directed.

Eukaryotic cells do not use this mechanism to distinguish the new strand from the template, and it is not yet understood how the mismatch repair system in eukaryotes "knows" which strand to repair.

Systems to Repair Damage to DNA

In the preceding section we discussed mistakes made when DNA is copied, where the wrong base is inserted during synthesis of the new strand. But even DNA that is not being replicated can get damaged or mutated. These sorts of damage are not associated with DNA replication, rather they can occur at any time.

What causes damage to DNA? Some major causes of DNA damage are:

- Radiation (e.g., UV rays in sunlight, in tanning booths)
- Exposure to damaging chemicals (such as benzopyrene in car exhaust and cigarette smoke)
- Chemical reactions within the cell (such as the deamination of cytosine to give uracil).

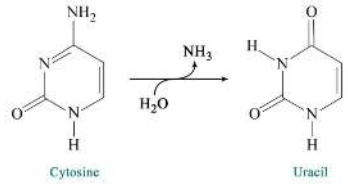

Figure **5.2.3**: *Deamination of cytosine*

This means the DNA in your cells is vulnerable to damage simply from normal sorts of actions, such as walking outdoors, being in traffic, or from the chemical transformations occurring in every cell as part of its everyday activities. (Naturally, the damage is much worse in situations where exposure to radiation or damaging chemicals is greater, such as when people repeatedly use tanning beds or smoke.)

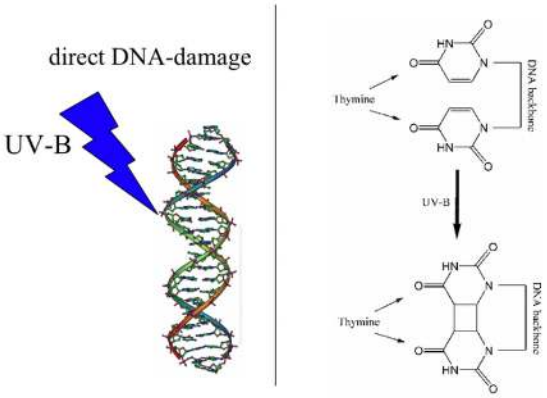

Figure **5.2.4**: *Thymine dimer formation*

What kinds of damage do these agents cause? Radiation can cause different kinds of damage to DNA. Sometimes, as with much of the damage done by UV rays, two adjacent pyrimidine bases in the DNA will be cross-linked to form pyrimidine dimers (note that we are talking about two neighboring pyrimidine bases on the same strand of DNA). This is illustrated in the figure on the previous page where two adjacent thymines on a single DNA strand are cross-linked to form a thymine dimer. Radiation can also cause breaks in the DNA backbone.

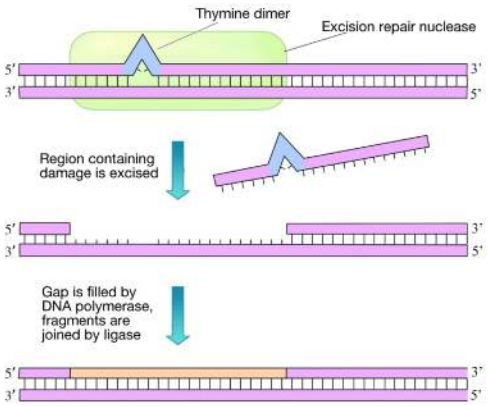

Figure 5.2.5: *Thymine dimer removal*

Chemicals like benzopyrene can attach themselves to bases, forming bulky DNA adducts in which large chemical groups are linked to bases in the DNA. The formation of chemical adducts can physically distort the DNA helix, making it hard for DNA and RNA polymerases to copy those regions of DNA.

Chemical reactions occurring within cells can cause cytosines in DNA to be deaminated to uracil, as shown in Figure 5.2.3.

Other sorts of damage in this category include the formation of oxidized bases like 8-oxo-guanine. These do not actually change the physical structure of the DNA helix, but they can cause problems because uracil and 8-oxo-guanine pair with different bases than the original cytosine or guanine, leading to mutations on the next round of replication.

How do cells repair such damage? Cells have several ways to remove the sorts of damage described above, with excision repair being a common strategy. Excision repair is a general term for the cutting out and re-synthesis of the damaged region of the DNA. There are a couple of varieties of excision repair:

Nucleotide Excision Repair (NER)

This system fixes damage by chemicals as well as UV damage. As shown in the figure on the previous page, in nucleotide excision repair, the damage is recognized and a cut is made on either side of the damaged region by an enzyme called an excinuclease (shown in green). A short portion of the DNA strand containing the damage is then removed and a DNA polymerase fills in the gap with the appropriate nucleotides. The newly made DNA is joined to the rest of the DNA backbone by the enzyme DNA ligase. In E. coli, NER is carried out by a group of proteins encoded by the uvrABC genes. As you can see, NER is similar, in principle, to mismatch repair. However, in NER, the distortion of the helix, caused by the DNA damage, clearly indicates which strand of the DNA needs to be removed and replaced.

Base Excision Repair (BER)

BER deals with situations like the deamination of cytosine to uracil. As noted earlier, cytosines in DNA sometimes undergo deamination to form the base uracil.

Because cytosines pair with guanines and uracils pair with adenine, the conversion of cytosine to uracil in the DNA would lead to the insertion of an A in the newly replicated strand instead of the G that should have gone in across from a C. To prevent this from happening, uracils are removed from DNA by base excision repair.

In base excision repair, a single base is first removed from the DNA, followed by removal of a region of the DNA surrounding the missing base. The gap is then repaired.

The removal of uracil from DNA is accomplished by the enzyme uracil DNA glycosylase, which breaks the bond between the uracil and the sugar in the nucleotide.

The removal of the uracil base creates a gap called an apyrimidinic site (AP site). The presence of the AP site triggers the activity of an AP endonuclease that cuts the DNA backbone.

A short region of the DNA surrounding the site of the original uracil is then removed and replaced.

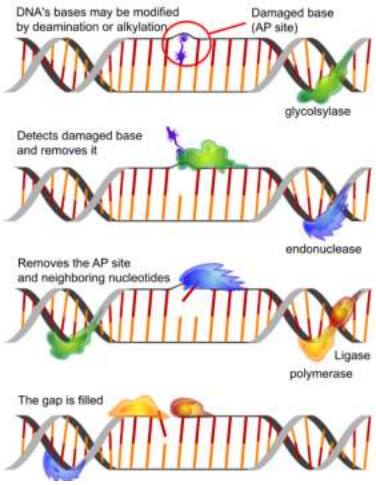

Figure **5.2.6**: *Base excision repair*

5.3: Transcription

In the preceding sections, we have discussed the replication of the cell's DNA and the mechanisms by which the integrity of the genetic information is carefully maintained. What do cells do with this information? How does the sequence in DNA control what happens in a cell? If DNA is a giant instruction book containing all of the cell's "knowledge" that is copied and passed down from generation to generation, what are the instructions for? And how do cells use these instructions to make what they need?

You have learned in introductory biology courses that genes, which are instructions for making proteins, are made of DNA. You also know that information in genes is copied into temporary instructions called messenger RNAs that direct the synthesis of specific proteins. This description of flow of information from DNA to RNA to protein, shown on the previous page, is often called the Central Dogma of molecular biology and is a good starting point for an examination of how cells use the information in DNA.

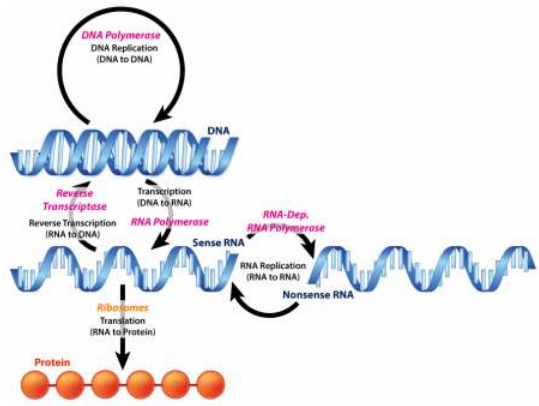

Figure 5.3.1: *Central dogma*

Consider that all of the cells in a multicellular organism have arisen by division from a single fertilized egg and therefore, all have the same DNA. Division of that original fertilized egg produces, in the case of humans, over a trillion cells, by the time a baby is produced from that egg (that's a lot of DNA replication!). Yet, we also know that a baby is not a giant ball of a trillion identical cells, but has the many different kinds of cells that make up tissues like skin and muscle and bone and nerves. How did cells that have identical DNA turn out so different.

The answer lies in gene expression, which is the process by which the information in DNA is used. Although all the cells in a baby have the same DNA, each different cell type uses a different subset of the genes in that DNA to direct the synthesis of a distinctive set of RNAs and proteins. The first step in gene expression is transcription, which we will examine next.

What is transcription? Transcription is the process of copying information from DNA sequences into RNA sequences. This process is also known as DNA-dependent RNA synthesis. When a sequence of DNA is transcribed, only one of the two DNA strands is copied into RNA.

General Features of Transcription

- The process of making RNA copies using a DNA as template is known as transcription
- The process of transcription produces all sorts of RNAs (mRNA, tRNA, rRNA, etc)
- One strand of the DNA serves as a template for the synthesis of RNA
- Enzymes that synthesize RNA are called RNA polymerases
- RNA polymerases synthesize RNA in the 5' to 3' direction
- RNA polymerases do not need a primer
- RNA polymerases uses rNTPs (ATP,GTP,UTP and CTP) to build the new RNA strand
- RNA polymerases bind at specific DNA sequences called promoters to start transcription
- RNA polymerases stop RNA synthesis when they reach sequences called terminators

Figure 5.3.2: *General Features of Transcription*

But, apart from copying one, rather than both strands of DNA, how is transcription different from replication of DNA. DNA replication serves to copy all the genetic material of the cell and occurs before a cell divides, so that a full copy of the cell's genetic information can be passed on to the daughter cell. Transcription, by contrast, copies short stretches of the coding regions of DNA to

make RNA. Different genes may be copied into RNA at different times in the cell's lifecycle. RNAs are, so to speak, temporary copies of instructions of the information in DNA and different sets of instructions are copied for use at different times.

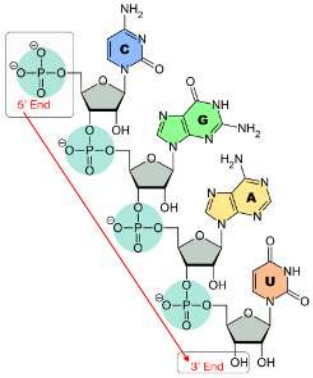

Figure 5.3.3: *RNA structure*

Cells make several different kinds of RNA:

- mRNAs that code for proteins
- rRNAS that form part of ribosomes
- tRNAs that serve as adaptors between mRNA and amino acids during translation
- Micro RNAs that regulate gene expression
- Other small RNAs that have a variety of functions.

Building an RNA strand is very similar to building a DNA strand. This is not surprising, knowing that DNA and RNA are very similar molecules. What enzyme carries out transcription? Transcription is catalyzed by the enzyme RNA Polymerase. "RNA polymerase" is a general term for an enzyme that makes RNA. There are many different RNA polymerases.

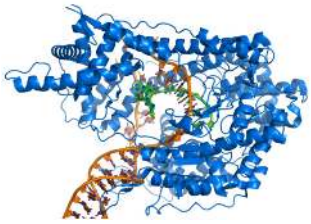

Figure 5.3.4: *T7 RNA Polymerase (blue) making RNA (green) using DNA template (brown)*

Like DNA polymerases, RNA polymerases synthesize new strands only in the 5' to 3' direction, but because they are making RNA, they use ribonucleotides (i.e., RNA nucleotides) rather than deoxyribonucleotides. Ribonucleotides are joined in exactly the same way as deoxyribonucleotides, which is to say that the 3'OH of the last nucleotide on the growing chain is joined to the 5' phosphate on the incoming nucleotide.

One important difference between DNA polymerases and RNA polymerases is that the latter do not require a primer to start making RNA. Once RNA polymerases are in the right place to start copying DNA, they just begin making RNA by stringing together RNA nucleotides complementary to the DNA template.

This, of course, brings us to an obvious question- how do RNA polymerases "know" where to start copying on the DNA. Unlike the situation in replication, where every nucleotide of the parental DNA must eventually be copied, transcription, as we have already noted, only copies selected genes into RNA at any given time.

Consider the challenge here: in a human cell, there are approximately 6 billion basepairs of DNA. Most of this is non- coding DNA, meaning that it won't need to be transcribed. The small percentage of the genome that is made up of coding sequences still amounts to between 20,000 and 30,000 genes in each cell. Of these genes, only a small number will need to be expressed at any given time.

What indicates to an RNA polymerase where to start copying DNA to make a transcript? Signals in DNA indicate to RNA polymerase where it should start (and end) transcription. These signals are special sequences in DNA that are recognized by the

RNA polymerase or by proteins that help RNA polymerase determine where it should bind the DNA to start transcription. A DNA sequence at which the RNA polymerase binds to start transcription is called a promoter.

A promoter is generally situated upstream of the gene that it controls. What this means is that on the DNA strand that the gene is on, the promoter sequence is "before" the gene. Remember that, by convention, DNA sequences are read from 5' to 3'. So the promoter lies 5' to the start point of transcription.

Also notice that the promoter is said to "control" the gene it is associated with. This is because expression of the gene is dependent on the binding of RNA polymerase to the promoter sequence to begin transcription. If the RNA polymerase and its helper proteins do not bind the promoter, the gene cannot be transcribed and it will therefore, not be expressed.

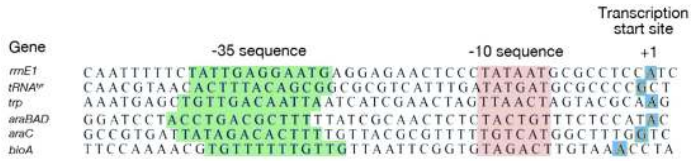

Figure 5.3.5: *Promoter Sequences*

What is special about a promoter sequence? In an effort to answer this question, scientists looked at many genes and their surrounding sequences. It makes sense that because the same RNA polymerase has to bind to many different promoters, the promoters should have some similarities in their sequences. Sure enough, common sequence patterns were seen to be present in many promoters. We will first take a look at prokaryotic promoters. When prokaryotic genes were examined, the following features commonly emerged (Figure 5.3.5):

- A transcription start site (this the base in the DNA across from which the first RNA nucleotide is paired).
- A -10 sequence: this is a 6 bp region centered about 10 bp upstream of the start site. The consensus sequence at this position is TATAAT. In other words, if you count back from the transcription start site, which by convention, is called the +1, the sequence found at -10 in the majority of promoters studied is TATAAT).
- A -35 sequence: this is a sequence at about 35 basepairs upstream from the start of transcription. The consensus sequence at this position is TTGACA.

What is the significance of these sequences? It turns out that the sequences at -10 and -35 are recognized and bound by a subunit of prokaryotic RNA polymerase before transcription can begin.

Figure 5.3.6: *Promoter sequence elements*

The RNA polymerase of E. coli, for example, has a subunit called the sigma subunit (or sigma factor) in addition to the core polymerase, which is the part of the enzyme that actually makes RNA. Together, the sigma subunit and core polymerase make up what is termed the RNA polymerase holoenzyme. The sigma subunit of the polymerase (shown in brown in Figure 5.3.7) can recognize and bind to the -10 and -35 sequences in the promoter, thus positioning the RNA polymerase (shown in green) at the right place to initiate transcription. Once transcription begins, the core polymerase and the sigma subunit separate, with the core polymerase continuing RNA synthesis and the sigma subunit wandering off to escort another core polymerase molecule to a promoter. The sigma subunit can be thought of as a sort of usher that leads the polymerase to its "seat" on the promoter.

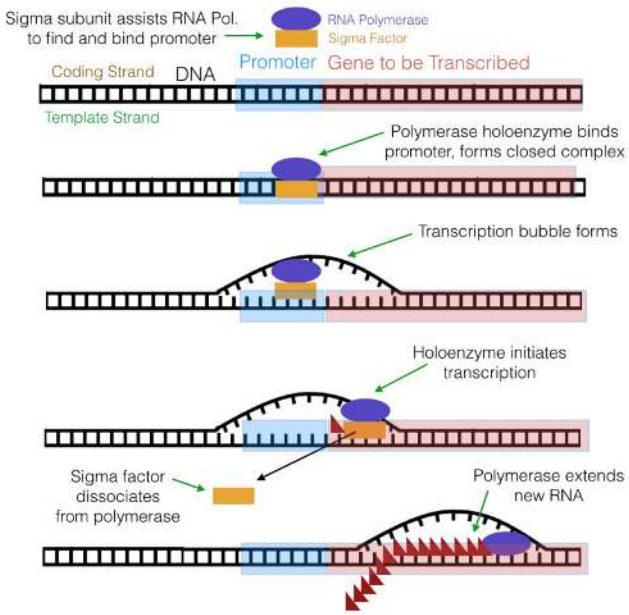

Figure 5.3.7: *Transcription initiation in E. coli*

As already mentioned, an RNA chain, complementary to the DNA template, is built by the RNA polymerase by the joining of the 5' phosphate of an incoming ribonucleotide to the 3'OH on the last nucleotide of the growing RNA strand. How does the polymerase know where to stop? A sequence of nucleotides called the terminator is the signal to the RNA polymerase to stop transcription and dissociate from the template.

Although the process of RNA synthesis is the same in eukaryotes as in prokaryotes, there are some additional issues to keep in mind in eukaryotes. One is that in eukaryotes, the DNA template exists as chromatin, where the DNA is tightly associated with histones and other proteins. The "packaging" of the DNA must therefore be opened up to allow the RNA polymerase access to the template in the region to be transcribed.

A second difference is that eukaryotes have multiple RNA polymerases, not one as in bacterial cells. The different polymerases transcribe different genes. For example, RNA polymerase I transcribes the ribosomal RNA genes, while RNA polymerase III copies tRNA genes. The RNA polymerase we will focus on most is RNA polymerase II, which transcribes protein-coding genes to make mRNAs.

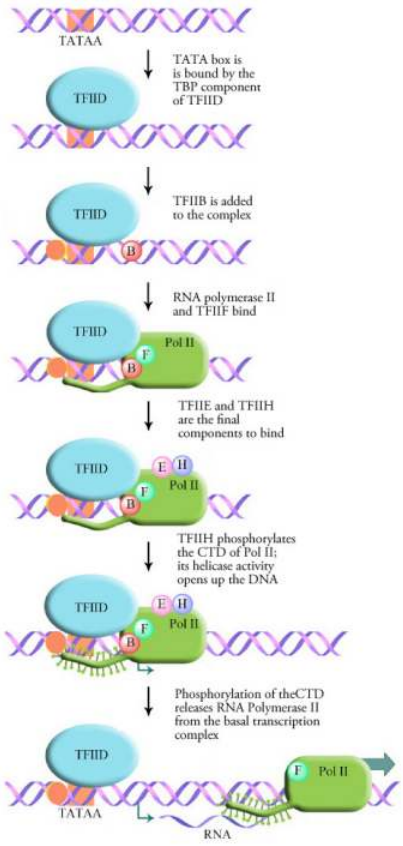

Figure 5.3.8: *Assembly of basal transcription complex and initiation of transcription*

All three eukaryotic RNA polymerases need additional proteins to help them get transcription started. In prokaryotes, RNA polymerase by itself can initiate transcription (remember that the sigma subunit is a subunit of the prokaryotic RNA polymerase). The additional proteins needed by eukaryotic RNA polymerases are referred to as transcription factors. We will see below that there are various categories of transcription factors.

Finally, in eukaryotic cells, transcription is separated in space and time from translation. Transcription happens in the nucleus, and the mRNAs produced are processed further before they are sent into the cytoplasm. Protein synthesis (translation) happens in the cytoplasm. In prokaryotic cells, mRNAs can be translated as they are coming off the DNA template, and because there is no nucleus, transcription and protein synthesis occur in a single cellular compartment.

Like genes in prokaryotes, eukaryotic genes also have promoters. Eukaryotic promoters commonly have a TATA box, a sequence about 25 basepairs upstream of the start of transcription that is recognized and bound by proteins that help the RNA polymerase to position itself correctly to begin transcription. (Some eukaryotic promoters lack TATA boxes, and have, instead, other recognition sequences to help the RNA polymerase find the spot on the DNA where it spot on the DNA where it binds and initiates transcription.)

We noted earlier that eukaryotic RNA polymerases need additional proteins to bind promoters and start transcription. What are these additional proteins that are needed to start transcription? General transcription factors are proteins that help eukaryotic RNA polymerases find transcription start sites and initiate RNA synthesis. We will focus on the transcription factors that assist RNA polymerase II. These transcription factors are named TFIIA, TFIIB and so on (TF= transcription factor, II=RNA polymerase II, and the letters distinguish individual transcription factors).

Transcription in eukaryotes requires the general transcription factors and the RNA polymerase to form a complex at the TATA box called the basal transcription complex or transcription initiation complex. This is the minimum requirement for any gene to be transcribed. The first step in the formation of this complex is the binding of the TATA box by a transcription factor called the TATA Binding Protein or TBP. Binding of the TBP causes the DNA to bend at this spot and take on a structure that is suitable for the

binding of additional transcription factors and RNA polymerase. As shown in the figure at left, a number of different general transcription factors, together with RNA polymerase (Pol II) form a complex at the TATA box.

The final step in the assembly of the basal transcription complex is the binding of a general transcription factor called TFIIH. TFIIH is a multifunctional protein that has helicase activity (i.e., it is capable of opening up a DNA double helix) as well as kinase activity. The kinase activity of TFIIH adds a phosphate onto the C-terminal domain (CTD) of the RNA polymerase. This phosphorylation appears to be the signal that releases the RNA polymerase from the basal transcription complex and allows it to move forward and begin transcription.

5.4: Regulation of Transcription

The processes described above are required whenever any gene is transcribed. But what determines which genes are transcribed at a given time. What are the molecular switches that turn transcription on or off? Although there are entire books written on this one topic, the basic mechanism by which transcription is regulated depends on highly specific interactions between transcription regulating proteins and regulatory sequences on DNA.

We know that promoters indicate where transcription begins, but what determines that a given gene will be transcribed? In addition to the promoter sequences required for transcription initiation, genes have additional regulatory sequences (sequences of DNA on the same DNA molecule as the gene) that control when a gene is transcribed. Regulatory sequences are bound tightly and specifically by transcriptional regulators, proteins that can recognize DNA sequences and bind to them. The binding of such proteins to the DNA can regulate transcription by preventing or increasing transcription from a particular promoter.

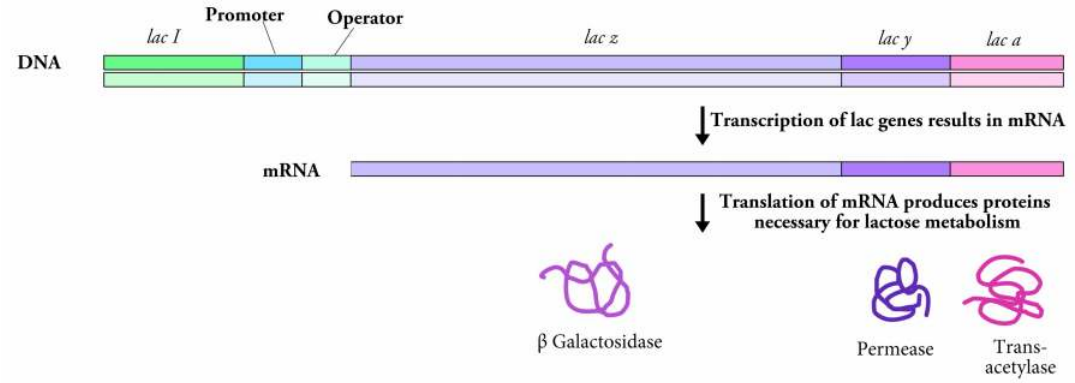

Figure 5.4.1: *The genes lac z, lac y, and lac a are all under the control of a single promotor in the lac operon*

Regulation in Prokaryotes

Let us first consider an example from prokaryotes. In bacteria, genes are often clustered in groups, such that genes that need to be expressed at the same time are next to each other and all of them are controlled as a single unit by the same promoter. The lac operon, shown in Figure 5.4.2, is one such group of genes that encode proteins needed for the uptake and breakdown of the sugar lactose. The three genes of the lac operon, lac z, lac y and lac a are controlled by a single promoter.

Bacterial cells generally prefer to use glucose for their energy needs, but if glucose is unavailable, and lactose is present, the bacteria will take up lactose and break it down for energy. Since the proteins for taking up and breaking down lactose are only needed when glucose is absent and lactose is available, the bacterial cells need a way to express the genes of the lac operon only under those conditions. At times when lactose is absent, the cells do not need to express these genes.

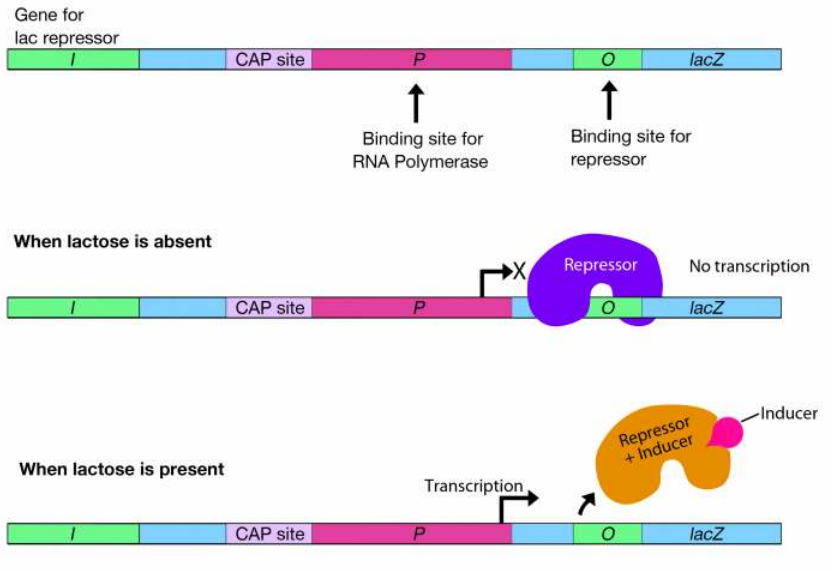

Figure 5.4.2: *Lac operon regulation*

How do bacteria achieve this? Transcription of the lac cluster of genes is primarily controlled by a repressor protein that binds to a region of the DNA just downstream of the -10 sequence of the lac promoter. Recall that the promoter is where the RNA polymerase must bind to begin transcription. The place where the repressor is bound is called the operator (labeled O in the figure). When the repressor is bound at this position, it physically blocks the RNA polymerase from transcribing the genes, just as a vehicle blocking your driveway would prevent you from pulling out.

Obviously, if you want to leave, the vehicle that is blocking your path must be removed. Likewise, in order for transcription to occur, the repressor must be removed from the operator to clear the path for RNA polymerase. How is the repressor removed?

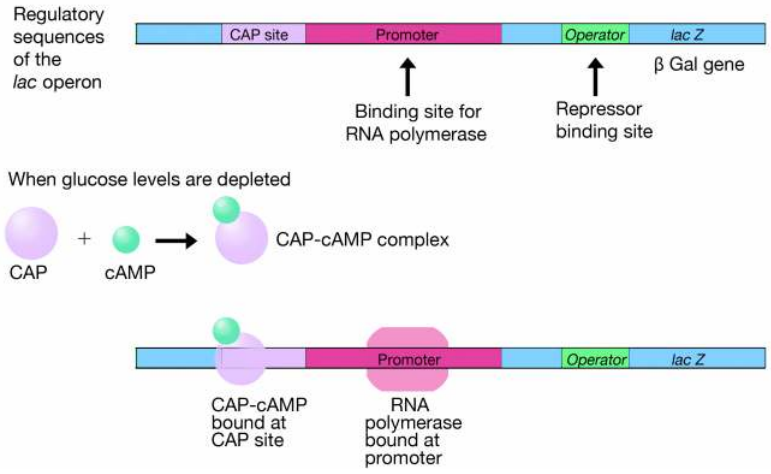

Figure 5.4.3: *Lac operon activation*

When the sugar lactose is present, it binds to the repressor, changing its conformation so that it no longer binds to the operator. When the repressor is no longer bound at the operator, the "road-block" in front of the RNA polymerase is removed, permitting the transcription of the genes of the lac operon.

Because the binding of the lactose induces the expression of the genes in the lac operon, lactose is called an inducer. (Technically, the inducer is allolactose, a molecule made from lactose by the cell, but the principle is the same.)

What makes this an especially effective control system is that the genes of the lac operon encode proteins that break down lactose. Turning on these genes requires lactose to be present. Once the lactose is broken down, the repressor binds to the operator once more and the lac genes are no longer expressed. This allows the genes to be expressed only when they are needed.

But how do glucose levels affect the expression of the lac genes? We noted earlier that if glucose was present, lactose would not be used. A second level of control is exerted by a protein called CAP that binds to a site adjacent to the promoter and recruits RNA polymerase to bind the lac promoter. When glucose is depleted, there is an increase in levels of cAMP which binds to CAP. The CAP cAMP complex then binds the CAP site, as shown in Figure 5.4.3. The combination of CAP binding and the lac repressor dissociating from the operator when lactose levels are high ensures high levels of transcription of the lac operon just when it is most needed. The CAP protein binding may be thought of as a green light for the RNA polymerase, while the removal of repressor is like the lifting of a barricade in front of it. When both conditions are met, the RNA polymerase transcribes the downstream genes.

The lac operon we have just described is a set of genes that are expressed only under the specific conditions of glucose depletion and lactose availability. Other genes may be expressed unless a particular condition is met. An example of this is the trp operon in bacterial cells, which encodes enzymes necessary for the synthesis of the amino acid tryptophan. These genes are expressed at all times, except when tryptophan is available from the cell's surroundings. This means that these genes must be prevented from being expressed in the presence of tryptophan. This is achieved by having a repressor protein that will bind to the operator only in the presence of tryptophan.

Regulation in Eukaryotes

Transcription in eukaryotes is also regulated by the binding of proteins to specific DNA sequences, but with some differences from the simple schemes outlined above. For most eukaryotic genes, general transcription factors and RNA polymerase (i.e., the basal transcription complex) are necessary, but not sufficient, for high levels of transcription.

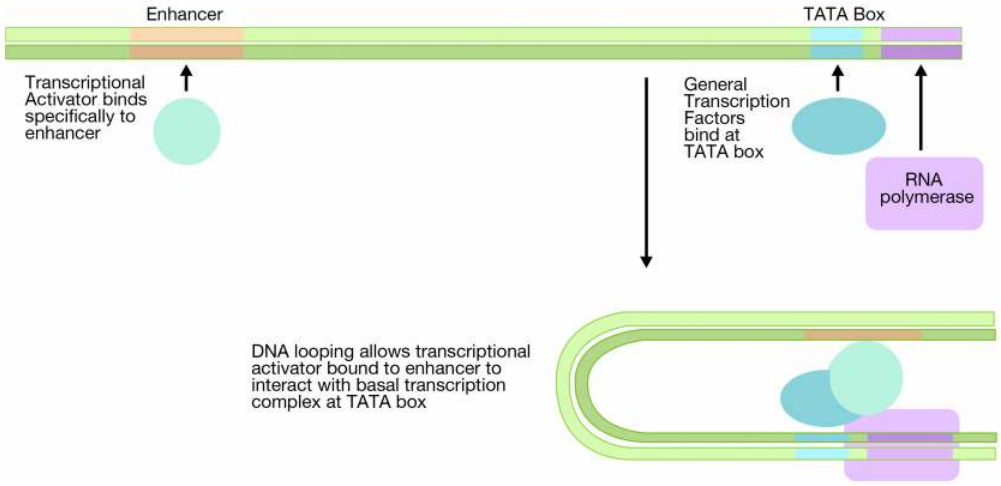

Figure 5.4.4: *Enhancer Mechanism*

In eukaryotes, additional regulatory sequences called enhancers and the proteins that bind to the enhancers are needed to achieve high levels of transcription. Enhancers are DNA sequences that regulate the transcription of genes. Unlike prokaryotic regulatory sequences, enhancers don't need to be next to the gene they control. Often they are many kilobases away on the DNA. As the name suggests, enhancers can enhance (increase) transcription of a particular gene.

How can a DNA sequence far from the gene being transcribed affect the level of its transcription?

Enhancers work by binding proteins (transcriptional activators) that can, in turn, interact with the proteins bound at the promoter. The enhancer region of the DNA, with its associated transcriptional activator(s) can come in contact with the basal transcription complex that is bound at a distant TATA box by looping of the DNA (previous page). This allows the protein bound at the enhancer to make contact with the proteins in the basal transcription complex.

One way that the transcriptional activator bound to the enhancer increases the transcription from a distant promoter is that it increases the frequency and efficiency with which the basal transcription complex is formed at the promoter.

Another mechanism by which proteins bound at the enhancer can affect transcription is by recruiting to the promoter other proteins that can modify the structure of the chromatin in that region. As we noted earlier, in eukaryotes, DNA is packaged with proteins to

form chromatin. When the DNA is tightly associated with these proteins, it is difficult to access for transcription. So proteins that can make the DNA more accessible to the transcription machinery can also play a role in the extent to which transcription occurs.

In addition to enhancers, there are also negative regulatory sequences called silencers. Such regulatory sequences bind to transcriptional repressor proteins. Transcriptional activators and repressors are modular proteins- they have a part that binds DNA and a part that activates or represses transcription by interacting with the basal transcription complex.

5.5: RNA Processing

So far, we have looked at the mechanism by which the information in genes (DNA) is transcribed into RNA. The newly made RNA, also known as the primary transcript (the product of transcription is known as a transcript) is further processed before it is functional. Both prokaryotes and eukaryotes process their ribosomal and transfer RNAs.

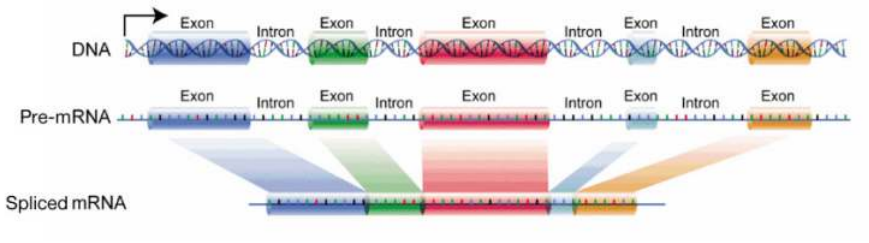

Figure 5.5.1: *RNA splicing*

The major difference in RNA processing, however, between prokaryotes and eukaryotes, is in the processing of messenger RNAs. We will focus on the processing of mRNAs in this discussion. You will recall that in bacterial cells, the mRNA is translated directly as it comes off the DNA template. In eukaryotic cells, RNA synthesis, which occurs in the nucleus, is separated from the protein synthesis machinery, which is in the cytoplasm. In addition, eukaryotic genes have introns, noncoding regions that interrupt the gene's coding sequence. The mRNA copied from genes containing introns will also therefore have regions that interrupt the information in the gene. These regions must be removed before the mRNA is sent out of the nucleus to be used to direct protein synthesis. The process of removing the introns and rejoining the coding sections or exons, of the mRNA, is called splicing. Once the mRNA has been capped, spliced and had a polyA tail added, it is sent from the nucleus into the cytoplasm for translation.

The initial product of transcription of a protein coding gene is called the pre-mRNA (or primary transcript). After it has been processed and is ready to be exported from the nucleus, it is called the mature mRNA or processed mRNA.

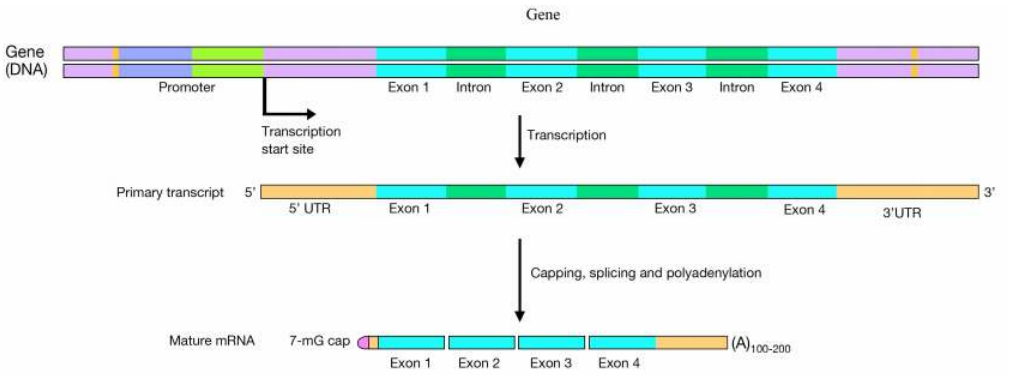

Figure 5.5.2: *Steps in processing eukaryotic messenger RNAs*

What are the processing steps for messenger RNAs?
In eukaryotic cells, pre-mRNAs undergo three main processing steps:

- Capping at the 5' end
- Addition of a polyA tail at the 3' end. and
- Splicing to remove introns

In the capping step of mRNA processing, a 7-methyl guanosine (shown at left) is added at the 5' end of the mRNA. The cap protects the 5' end of the mRNA from degradation by nucleases and also helps to position the mRNA correctly on the ribosomes during protein synthesis.

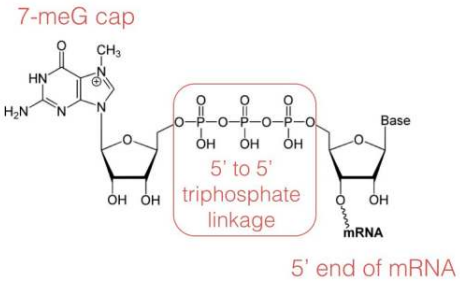

Figure 5.5.3: *mRNA capping structure*

The 3' end of a eukaryotic mRNA is first trimmed, then an enzyme called PolyA Polymerase adds a "tail" of about 200 ‘A’ nucleotides to the 3' end. There is evidence that the polyA tail plays a role in efficient translation of the mRNA, as well as in the stability of the mRNA. The cap and the polyA tail on an mRNA are also indications that the mRNA is complete (i.e., not defective). Introns are removed from the pre-mRNA by the activity of a complex called the spliceosome. The spliceosome is made up of proteins and small RNAs that are associated to form protein-RNA enzymes called small nuclear ribonucleoproteins or snRNPs (pronounced SNURPS). The splicing machinery must be able to recognize splice junctions (i.e., the end of each exon and the start of the next) in order to correctly cut out the introns and join the exons to make the mature, spliced mRNA.

What signals indicate where an intron starts and ends? The base sequence at the start (5' or left end, also called the donor site) of an intron is GU while the sequence at the 3' or right end (a.k.a. acceptor site) is AG. There is also a third important sequence within the intron, called a branch point, that is important for splicing.

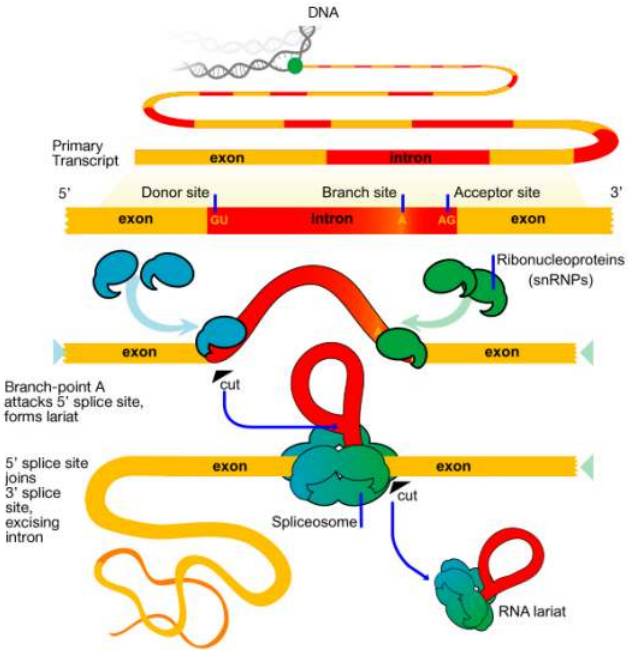

Figure 5.5.4: *Splicing*

There are two main steps in splicing:

- In the first step, the pre-mRNA is cut at the 5' splice site (the junction of the 5' exon and the intron). The 5' end of the intron then is joined to the branch point within the intron. This generates the lariat-shaped molecule characteristic of the splicing process
- In the second step, the 3' splice site is cut, and the two exons are joined together, and the intron is released.

Many pre-mRNAs have a large number of exons that can be spliced together in different combinations to generate different mature mRNAs. This is called alternative splicing, and allows the production of many different proteins using relatively few genes, since a single RNA can, by combining different exons during splicing, create many different protein coding messages. Because of alternative splicing, each gene in our DNA gives rise, on average, to three different proteins. Once protein coding messages have

been processed by capping, splicing and addition of a poly A tail, they are transported out of the nucleus to be translated in the cytoplasm.

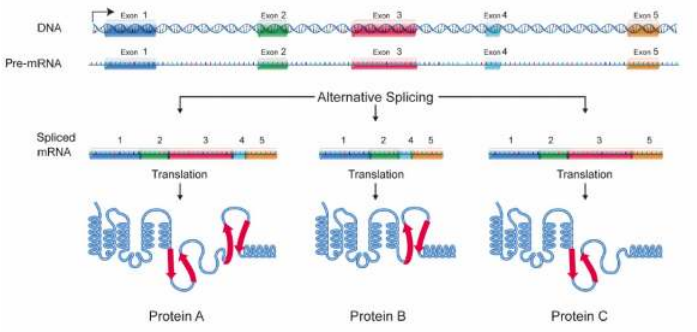

Figure 5.5.5: *Splicing and protein diversity*

5.6: Translation

Translation is the process by which information in mRNAs is used to direct the synthesis of proteins. As you have learned in introductory biology, in eukaryotic cells, this process is carried out in the cytoplasm of the cell, by large RNA-protein machines called ribosomes. Ribosomes contain ribosomal RNAs (rRNAs) and proteins. The proteins and rRNAs are organized into two subunits, a large and a small. The large subunit has an enzymatic activity, known as a peptidyl transferase, that makes the peptide bonds that join amino acids to make a polypeptide. The small and large subunits assemble on the mRNA at its 5'end to initiate translation. Ribosomes function by binding to mRNAs and holding them in a way that allows the amino acids encoded by the RNA to be joined sequentially to form a polypeptide.

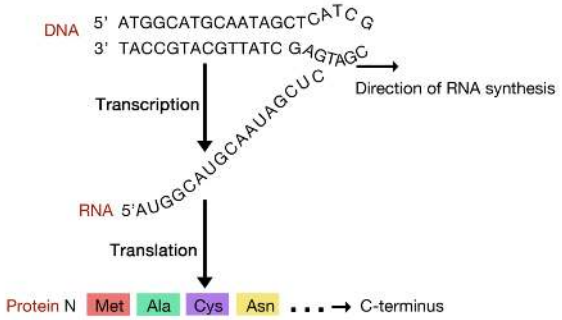

Figure 5.6.1: *Coupled transcription and translation in prokaryotes*

The sequence of an mRNA directly specifies the sequence of amino acids in the protein it encodes. Each amino acid in the protein is specified by a sequence of 3 bases called a codon in the mRNA. For example, the amino acid tryptophan is encoded by the sequence 5'UGG3' on an mRNA. Given that there are 4 bases in RNA, the number of different 3-base combinations that are possible is $\mathbf{4^3}$, or 64. There are, however, only 20 amino acids that are used in building proteins. This discrepancy in the number of possible codons and the actual number of amino acids they specify is explained by the fact that the same amino acid may be specified by more than one codon. In fact, with the exception of the amino acids methionine and tryptophan, all the other amino acids are encoded by multiple codons. The figure above shows the codons that are used for each of the twenty amino acids.

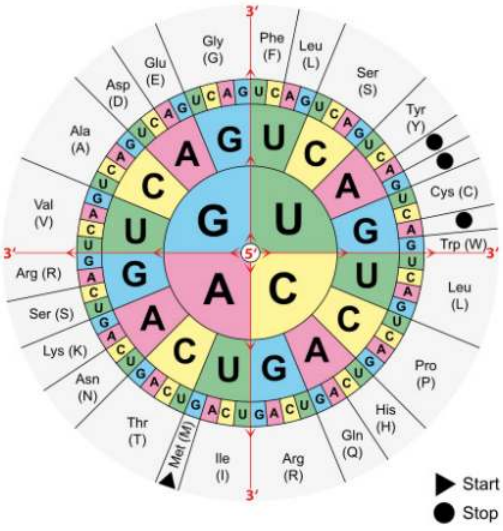

Figure 5.6.2: *The genetic code*

Three of the 64 codons are known as termination or *stop codons* and as their name suggests, indicate the end of a protein coding sequence. The codon for methionine, AUG, is used as the start, or initiation, codon.

This ingenious system is used to direct the assembly of a protein in the same way that you might string together colored beads in a particular order using instructions that used symbols like 111 for a red bead, followed by 222 for a green bead, 333 for yellow, and so on, till you came to 000, indicating that you should stop stringing beads.

While the ribosomes are literally the protein factories that join amino acids together using the instructions in mRNAs, another class of RNA molecules, the transfer RNAs (tRNAs) are also needed for translation. Transfer RNAs (see figure, left) are small RNA molecules, about 75-80 nucleotides long, that function to 'interpret' the instructions in the mRNA during protein synthesis. In terms of the bead analogy above, someone, or something, has to be able to bring a red bead in when the instructions indicate 111, and a green bead when the instructions say 222. Unlike a human, who can choose a red bead when 111 is present in the instructions, neither ribosomes nor tRNAs can think. The system, therefore, relies, like so many processes in cells, solely on molecular recognition.

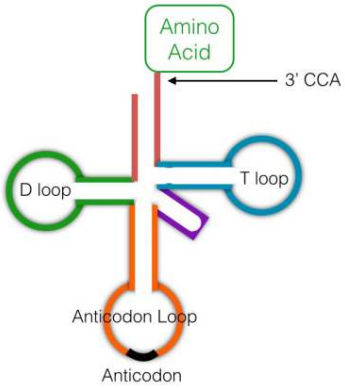

Figure 5.6.3: *tRNA structure*

A given transfer RNA is specific for a particular amino acid. It is linked covalently to this amino acid at its 3' end by an enzyme called aminoacyl tRNA synthetase. There is an aminoacyl tRNA synthetase specific for each amino acid. A tRNA with an amino acid attached to it is said to be charged. Another region of the tRNA has a sequence of 3 bases, the anticodon, that is complementary to the codon for the amino acid it is carrying. When the tRNA encounters the codon for its amino acid on the messenger RNA, the anticodon will base-pair with the codon, and the amino acid attached to it will be brought in to the ribosome to be added on to the growing protein chain.

With an idea of the various components necessary for translation we can now take a look at the process of protein synthesis. The main steps in the process are similar in prokaryotes and eukaryotes. As we already noted, ribosomes bind to mRNAs and facilitate the interaction between the codons in the mRNA and the anticodons on charged tRNAs.

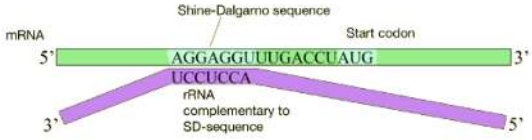

Figure 5.6.4: *mRNA alignment by Shine-Dalgamo sequence*

In bacterial cells, translation is coupled with transcription and begins even before the mRNA has been completely synthesized. How does the ribosome recognize and bind to the mRNA. Many bacterial mRNAs carry a short purine-rich sequence known as the Shine-Dalgarno site upstream of the AUG start codon, as shown in the figure below. This sequence is recognized and bound by a complementary sequence in the 16S rRNA that is part of the small ribosomal subunit as shown above. Because the Shine-Dalgarno site serves to recruit and bind the ribosome, it is also referred to as the Ribosome Binding Site or RBS.

Figure 5.6.5: *Shine-Dalgamo sequences*

A variation of this process of ribosome assembly operates in eukaryotic cells. We already know that in eukaryotic cells, processed mRNAs are sent from the nucleus to the cytoplasm.

The small and large subunits of ribosomes, each composed of characteristic rRNAs and proteins are found in the cytoplasm and assemble on mRNAs to form complete ribosomes that carry out translation.

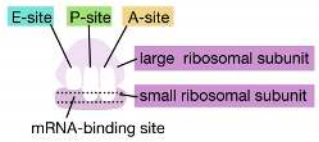

Figure 5.6.6: *A, P, and E sites of the ribosome*

Protein synthesis in eukaryotes starts with the binding of the small subunit of the ribosome to the 5' end of the mRNA. The assembly of the translation machinery begins with the binding of the small ribosomal subunit to the 7-methyl guanosine cap on the 5'end of an mRNA. Meanwhile, the initiator tRNA pairs with the start codon. (Recall that the start codon is AUG, and codes for methionine. The initiator tRNA carries the amino acid methionine). The large subunit of the ribosome then joins the complex, which is now ready to start protein synthesis.

Ribosomes have two sites for binding charged tRNAs, each of which is positioned to make two adjacent codons on the mRNA available for binding by tRNAs. The initiation codon occupies the first of these ribosomal sites, the P-site. The anticodon complementary to this is on the initiator tRNA, which brings in the first amino acid of the protein. This initial phase of translation is called initiation and requires the help of protein factors called initiation factors.

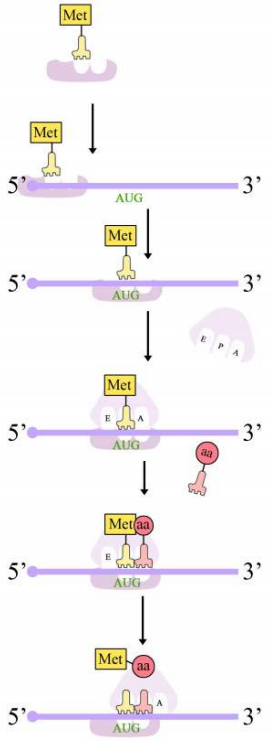

Figure 5.6.7: *Initiation of translation*

The second codon of the mRNA is positioned adjacent to the second site on the ribosome, the A site. This is where the tRNA carrying the amino acid specified by the second codon binds. The binding of aminoacyl tRNA to the A-site is mediated by proteins called elongation factors and requires the input of energy. Once the appropriate charged tRNAS have "docked" on the codons by base-pairing between the anticodon on the tRNA and the codon on the mRNA, the ribosome joins the amino acids carried by the two tRNAs by making a peptide bond (see figure at right).

Interestingly, the formation of the peptide bond is catalyzed by a catalytic RNA (the 23S rRNA in prokaryotes) rather than by a protein enzyme.

This and subsequent steps in the synthesis of the polypeptide are called the elongation phase of translation. Once the first two amino acids are linked , the first tRNA dissociates, and moves out of the P-site and into the E, or Exit site. The second tRNA then moves into the P-site, vacating the A-site for the tRNA corresponding to the next codon.

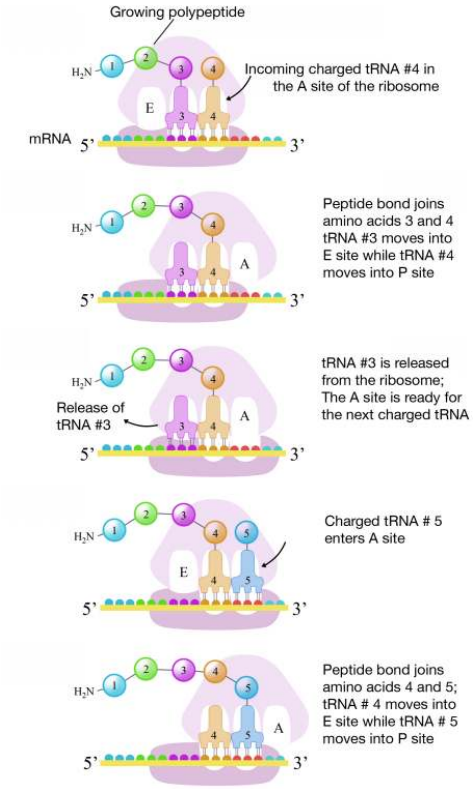

Figure 5.6.8: *Elongation of translation*

The process repeats till the stop codon is in the A-site. At this point, a release factor binds at the A-site, adds a water molecule to the polypeptide at the P-site, and releases the completed polypeptide from the ribosome, which itself, then dissociates into subunits.

As described in Chapter 3, polypeptides made in this way are then folded into their three dimensional shapes, post-translationally modified and delivered to the appropriate cellular compartments to carry out their functions.

CHAPTER OVERVIEW

6: Metabolism I - Oxidative/Reductive Processes

The cost of living is energy and the producers and consumers of energy in the cell are the chemical reactions known collectively as metabolism. Metabolic processes are governed by the same laws of energy as the rest of the universe, so they must be viewed in the light of Gibbs free energy. For the most part, the drivers of changes in Gibbs free energy are changes in concentration of reactants and products but for some reactions, the concentration changes required to run a reaction in the desired direction are not practical. In such cases, cells may use alternative strategies, such as energy coupling reactions (combining an energetically unfavorable reaction with a favorable one, such as the hydrolysis of ATP) to help “drive" the unfavorable reaction. In other cases, cells use alternate pathways around energetically unfavorable reactions.

Depending on your mathematical perspective, life is the sum of the product of the biochemical reactions that occur in cells. The collection of these reactions is known as metabolism. We break the subject into two broad areas: 1) oxidative/reductive metabolism and 2) pathways that involve little oxidation/reduction. This chapter deals with the former.

6.1: Definitions

We start by defining a few terms. Anabolic processes refer to collections of biochemical reactions that make bigger molecules from smaller ones. Examples include the synthesis of fatty acids from acetyl-CoA, of proteins from amino acids, of complex carbohydrates from simple sugars, and of nucleic acids from nucleotides. Just as any construction project requires energy, so, too, do anabolic processes require input of energy. Anabolic processes tend to be reductive in nature, in contrast to catabolic processes, which are oxidative. Not all anabolic processes are reductive, though. Protein synthesis and nucleic acid synthesis do not involve reduction, though the synthesis of amino acids and nucleotides does.

Catabolic processes are the primary sources of energy for heterotrophic organisms and they ultimately power the anabolic processes. Examples include glycolysis (breakdown of glucose), the citric acid cycle, and fatty acid oxidation. Reductive processes require electron sources, such as NADPH, NADH, or $\mathbf{FADH_2}$. Oxidative processes require electron carriers, such as $\mathbf{NAD^+}$, $\mathbf{NADP^+}$, or FAD. Catabolic processes are ultimately the source of ATP energy in cells, but the vast majority of ATP i
heterotrophic organisms is not made in directly in these reactions. Instead, the electrons released by oxidation are collected by electron carriers which donate them, in the mitochondria, to complexes that make ATP (ultimately) by oxidative phosphorylation.

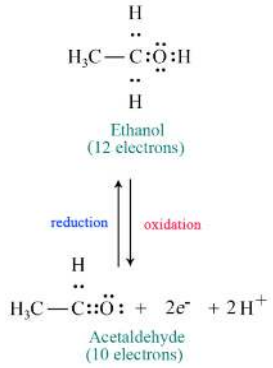

Figure 6.1.1: *Redox Reactions*

In our tour of metabolism, we will tackle in this chapter processes that are the most oxidative/reductive in nature and in the following chapter those pathways that involve less reduction/oxidation. The aim in this coverage is not to go through the step- by-step reactions of the pathway, but rather to focus on control points, interesting enzymes, molecules common between pathways, and how the metabolic pathways meet the organism's needs.

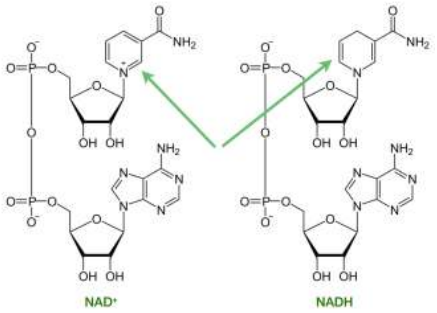

Figure 6.1.2: *NAD+ and NADH*

6.2: Perspectives

We can view metabolism at several levels. At the highest level, we have nutrients, such as sugars, fatty acids and amino acids entering cells and carbon dioxide and other waste products (such as urea) exiting. Cells use the incoming materials for energy and substance to synthesize sugars, nucleotides, and other amino acids as building blocks for the carbohydrates, nucleic acids, fatty compounds, and proteins necessary for life. As we zoom in, we can imagine pathways made up of reactions for breakdown and synthesis of each of these compounds. The figure at left shows such a simple schematic and how the pathways are not isolated from each other – molecular products of one are substrates for another. At a deeper level, we can study individual reactions and discover the enormous complexity and commonality of metabolic reactions.

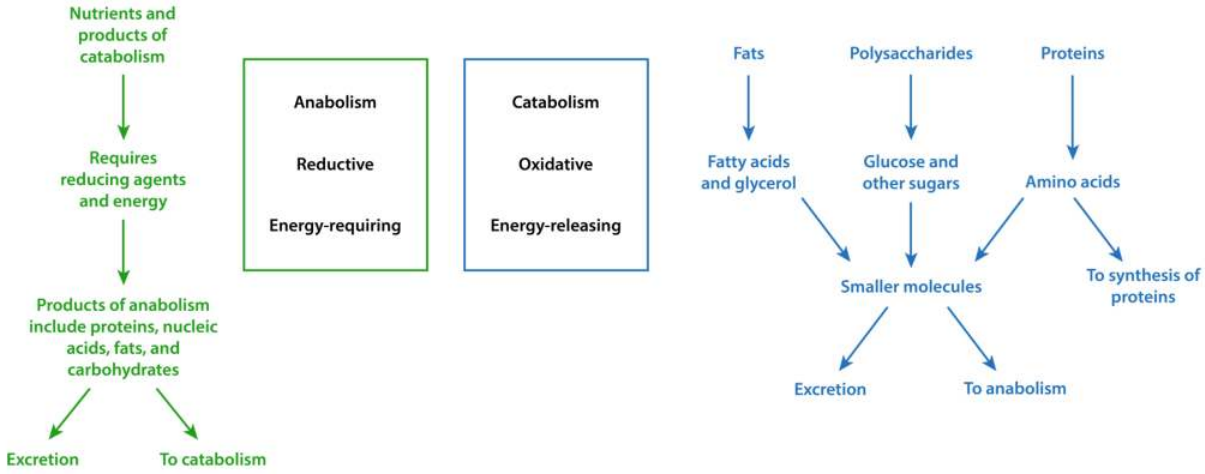

Figure 6.2.1: *Anabolic vs Catabolic Processes*

In studying metabolism, we recognize that metabolic pathways are manmade concepts with artificial boundaries. Students commonly think of the molecules in the pathways being tied exclusively to those individual pathways, but with the exception of reactions that have physical barriers (such as those occurring within an organelle), metabolic pathways have many common intermediates used in multiple reactions occurring in the same location at the same time and thus cannot be ascribed to any one pathway. The best we can do is understand general directions of pathways in cells.

6.3: Glycolysis

Glycolysis, which literally means “breakdown of sugar," is a catabolic process in which six-carbon sugars (hexoses) are oxidized and broken down into pyruvate molecules. The corresponding anabolic pathway by which glucose is synthesized is termed gluconeogenesis. Both glycolysis and gluconeogenesis are not major oxidative/reductive processes by themselves, with one step in each one involving loss/gain of electrons, but the product of glycolysis, pyruvate, can be completely oxidized to carbon dioxide. Indeed, without production of pyruvate from glucose in glycolysis, a major energy source for the cell is not available. By contrast, gluconeogenesis can synthesize glucose reductively from very simple materials, such as pyruvate and acetyl-CoA/ glyoxylate (at least in plants). For these reasons we include these pathways in the red/ox collection.

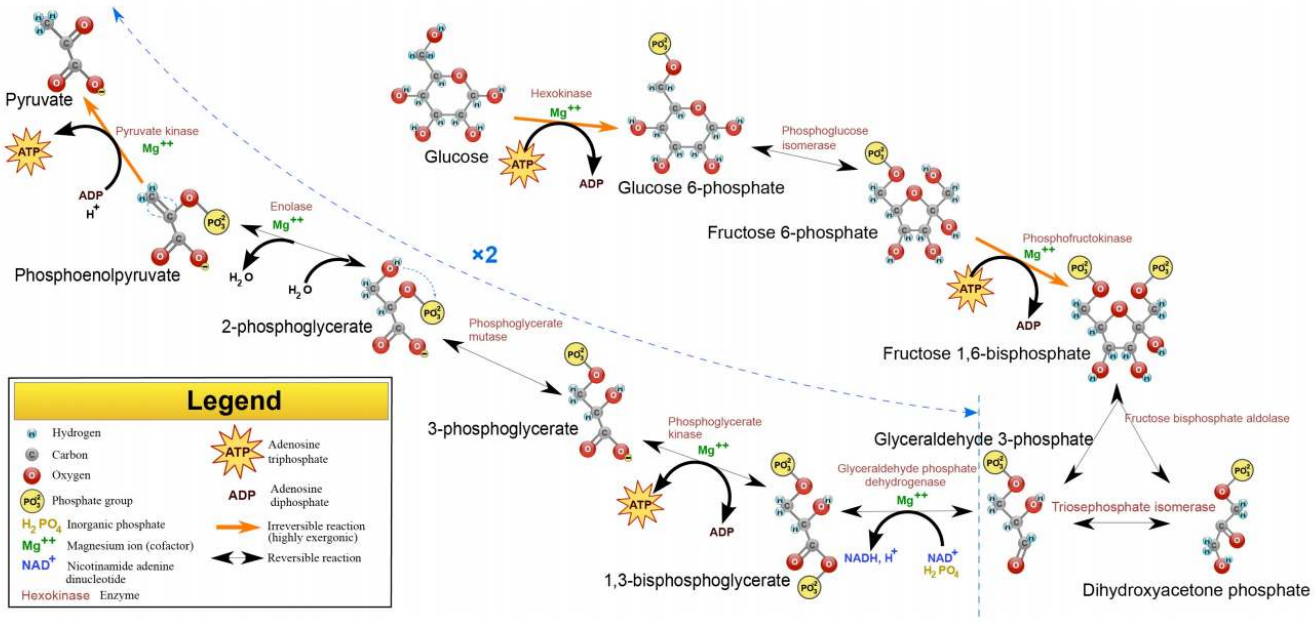

Figure 6.3.1: *The Reactions of Glycolysis*

Glucose is the most abundant hexose in nature and is the one people typically associate with glycolysis, but fructose (in the form of fructose-6-phosphate) is metabolized in the cell and galactose can easily be converted into glucose for catabolism in the pathway as well. The end metabolic products of the pathway are two molecules of ATP, two molecules of NADH and two molecules of pyruvate, which, in turn, can be oxidized further in citric acid cycle.

Intermediates

Glucose and fructose are the sugar ‘funnels’ serving as entry points to the glycolytic pathway. Other sugars must be converted to either of these forms to be directly metabolized. Some pathways, including the Calvin Cycle and the Pentose Phosphate Pathway (PPP, see below) contain intermediates in common with glycolysis, so in that sense, almost any cellular sugar can be metabolized here. Intermediates of glycolysis that are common to other pathways include glucose-6-phosphate (PPP, glycogen metabolism), F6P (PPP), G3P (Calvin, PPP), DHAP (PPP, glycerol metabolism, Calvin), 3PG (Calvin, PPP), PEP (C4 plant metabolism, Calvin), and pyruvate (fermentation, acetyl-CoA genesis, amino acid metabolism).

Reactions

The pathway of glycolysis begins with two inputs of energy. First, glucose gets a phosphate from ATP to make glucose-6-phosphate (G6P) and later fructose-6-phosphate (F6P) gets another phosphate from ATP to make fructose-1,6-bisphosphate (F1,6BP). With the pump thus primed, the pathway proceeds first to split the F1,6BP into two 3-carbon intermediates. Later the only oxidation step in the entire pathway occurs. In that reaction, glyceraldehyde-3-phosphate (G3P) is oxidized and a phosphate is added, creating 1,3-bisphosphoglycerate (1,3 BPG).

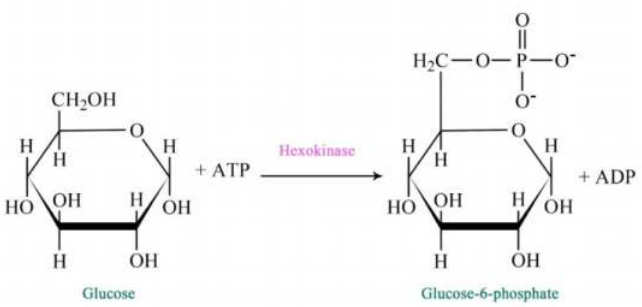

Figure 6.3.2: *Step 1 of Glycolysis*

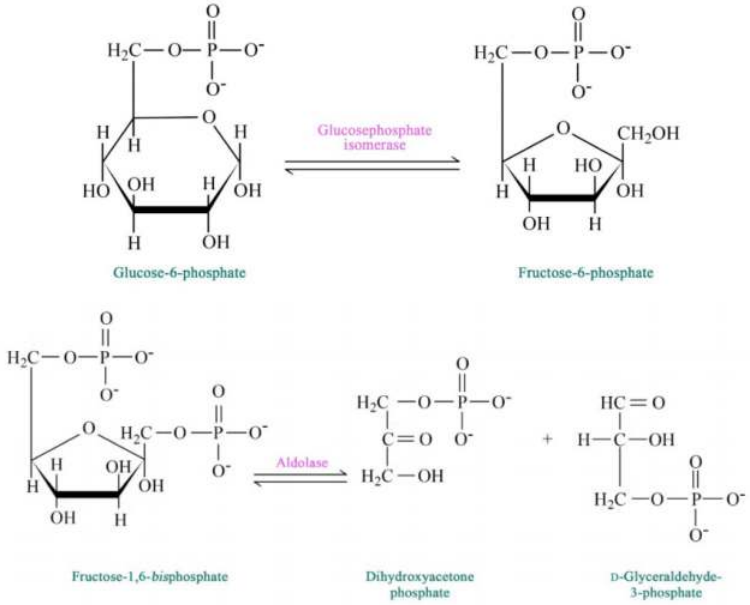

Figure 6.3.3: *Step 2 and 4 of Glycolysis*

The addition of the phosphate sometimes conceals the oxidation that occurred. G3P was an aldehyde. 1,3 BGP is an acid esterified to a phosphate. The two phosphates in the tiny 1,3BPG molecule repel each other and give the molecule high energy. It uses this energy to phosphorylate ADP to make ATP.

Since there are two 1,3 BPGs produced for every glucose, the two ATP produced replenish the two ATPs used to start the cycle.

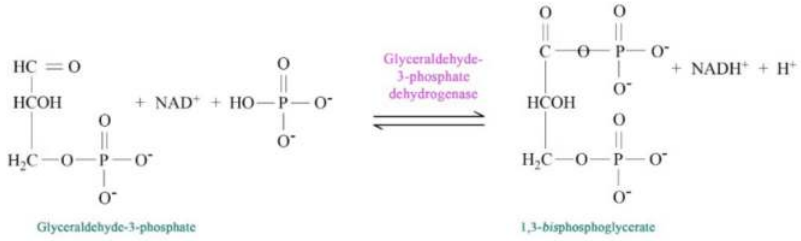

Figure 6.3.4: *Step 5 of Glycolysis*

The synthesis of ATP directly from a metabolic reaction is known as substrate level phosphorylation, though it is not a significant source of ATP. Glycolysis has two reactions during which substrate-level phosphorylation occurs.

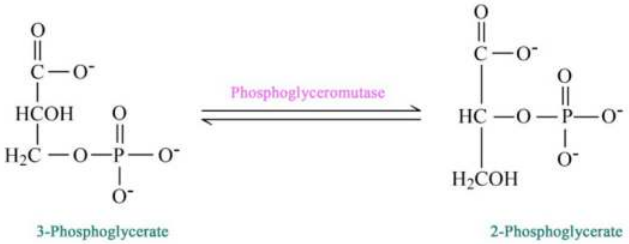

Figure 6.3.5: *Step 8 of Glycolysis*

The transfer of phosphate from 1,3BPG to ATP creates 3-phosphoglycerate (3-PG). Conversion of 3-PG to 2-PG occurs by an important mechanism. An intermediate in the reaction (catalyzed by phosphogly cerate mutase) is 2,3 BPG. This intermediate, which is stable, is released with low frequency by the enzyme instead of being converted to 2-PG. 2,3BPG is important because it binds to hemoglobin and stimulates release of oxygen. Thus, cells which are metabolizing glucose rapidly release more 2,3BPG and, as a result, stimulate release of more oxygen, supporting their needs.

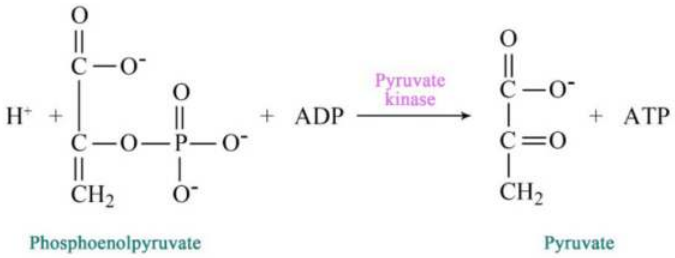

Figure 6.3.6: *Step 10 of Glycolysis*

2-PG is converted to phosphoenolpyyruvate (PEP) by removal of water, creating a very high energy intermediate. Conversion of PEP to pyruvate is the second substrate level phosphorylation of glycolysis, creating ATP. There is almost enough energy in PEP to stimulate production of a second ATP, but it is not used. Consequently, the energy is lost as heat. If you wonder why you get hot when you exercise, the reaction that converts PEP to pyruvate is a prime culprit.

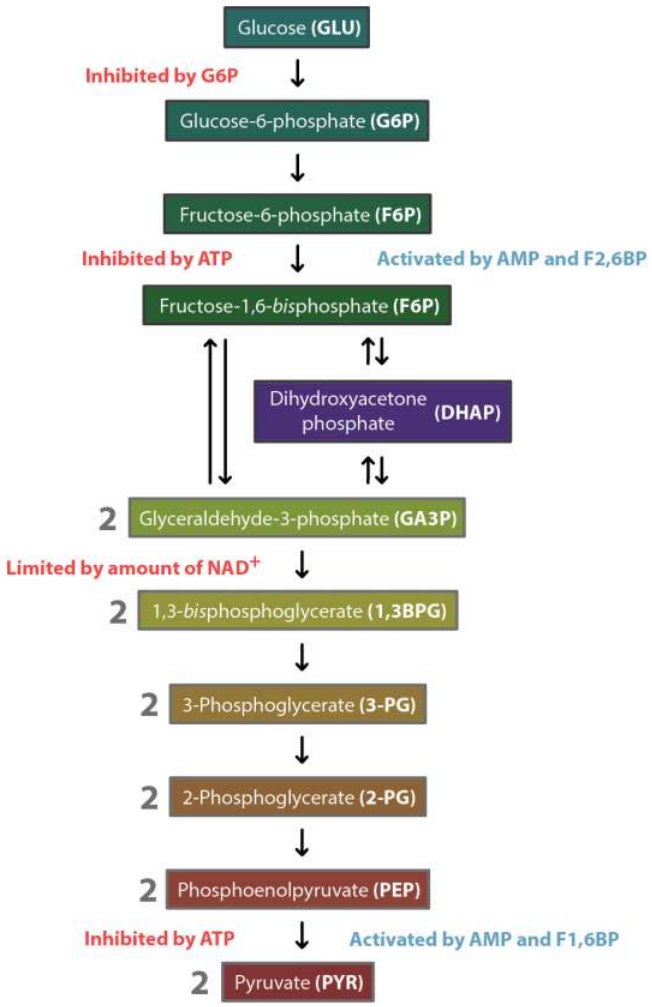

Figure 6.3.7: *Glycolysis Regulation*

Enzymes/Control

Control of glycolysis is unusual for a metabolic pathway, in that regulation occurs at three enzymatic points:

$$\underbrace{\text{Glucose} \rightleftharpoons \text{G}_6\text{P}}_{\text{hexokinase}} \tag{6.3.1}$$

$$\underbrace{\text{F}_6\text{P} \rightleftharpoons \text{F}_1,6\,\text{BP}}_{\text{phosphofructokinase (PFK)}} \tag{6.3.2}$$

and

$$\underbrace{\text{PEP} \rightleftharpoons \text{pyruvate}}_{\text{pyruvate kinase}}. \tag{6.3.3}$$

Glycolysis is regulated in a reciprocal fashion compared to its corresponding anabolic pathway, gluconeogenesis. Reciprocal regulation occurs when the same molecule or treatment (phosphorylation, for example) has opposite effects on catabolic and anabolic pathways. Reciprocal regulation is important when anabolic and corresponding catabolic pathways are occurring in the same cellular location.

As an example, consider regulation of PFK. It is activated by several molecules, most importantly fructose-2,6- bisphosphate (F2,6BP). This molecule has an inhibitory effect on the corresponding gluconeogenesis enzyme, fructose-1,6-bisphosphatase (F1,6BPase).

You might wonder why pyruvate kinase, the last enzyme in the pathway, is regulated. The answer is simple. Pyruvate kinase catalyzes the most energetically rich reaction of glycolysis. The reaction is favored so strongly in the forward direction that cells must do a ‘two-step’ around it in the reverse direction when making glucose. In other words, it takes two enzymes, two reactions, and two triphosphates to go from pyruvate back to PEP in gluconeogenesis. When cells are needing to make glucose, they can’t be sidetracked by having the PEP they have made in gluconeogenesis be converted directly back to pyruvate by pyruvate kinase. Consequently, pyruvate kinase is inhibited during gluconeogenesis, lest a “futile cycle" occur.

Another interesting control mechanism called feedforward activation involves pyruvate kinase. Pyruvate kinase is activated allosterically by F1,6BP. This molecule is a product of the PFK reaction and a substrate for the aldolase reaction. It should be noted that the aldolase reaction is energetically unfavorable (high +ΔG°’), thus allowing F1,6BP to accumulate. When this happens, some of the excess F1,6BP activates pyruvate kinase, which jump-starts the conversion of PEP to pyruvate. The resulting drop in PEP levels has the effect of “pulling" on the reactions preceding pyruvate kinase. As a consequence, the concentrations of G3P and DHAP fall, helping to move the aldolase reaction forward.

Pyruvate Metabolism

As noted, pyruvate produced in glycolysis can be oxidized to acetyl-CoA, which is itself oxidized in the citric acid cycle to carbon dioxide. That is not the only metabolic fate of pyruvate, though.

Pyruvate is a “starting" point for gluconeogenesis, being converted to oxaloacetate in the mitochondrion in the first step. Pyruvate in animals can also be reduced to lactate when oxygen is limiting. This reaction, which requires NADH produces $\mathbf{NAD^+}$ and is critical for generating the latter molecule to keep the glyceraldehyde-3-phosphate dehydrogenase reaction of glycolysis going when there is no oxygen.

Oxygen is necessary for the electron transport system to operate and this, in turn, is what oxidizes NADH to $\mathbf{NAD^+}$. In the absence of oxygen, thus, an alternative means of making $\mathbf{NAD^+}$ is necessary, or else glycolysis will halt. Bacteria and yeast have NADH requiring reactions that regenerate $\mathbf{NAD^+}$ while producing ethanol from pyruvate under anaerobic conditions, instead of lactic acid. Thus, fermentation of pyruvate is necessary to keep glycolysis operating when oxygen is limiting. It is also for these reasons that brewing of beer (using yeast) involves depletion of oxygen and muscles low in oxygen produce lactic acid (animals).

Pyruvate is a precursor of alanine which can be easily synthesized by transfer of a nitrogen from an amine donor, such as glutamic acid. Pyruvate can also be converted into oxaloacetate by carboxylation in the process of gluconeogenesis (see Figure 6.3.8).

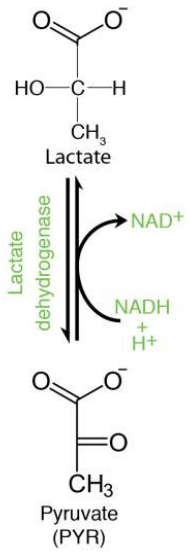

Figure 6.3.8: *Pyruvate Reduction*

The enzymes involved in pyruvate metabolism include pyruvate dehydrogenase (makes acetyl-CoA), lactate dehydrogenase (makes lactate), transaminases (make alanine), an
pyruvate carboxylase (makes oxaloacetate).

6.4: Gluconeogenesis

The anabolic counterpart to glycolysis is gluconeogenesis, which occurs mostly in the cells of the liver and kidney. In seven of the eleven reactions of gluconeogenesis (starting from pyruvate), the same enzymes are used as in glycolysis, but the reaction directions are reversed. Notably, the ΔG values of these reactions in the cell are typically near zero, meaning their direction can be readily controlled by changing substrate and product concentrations.

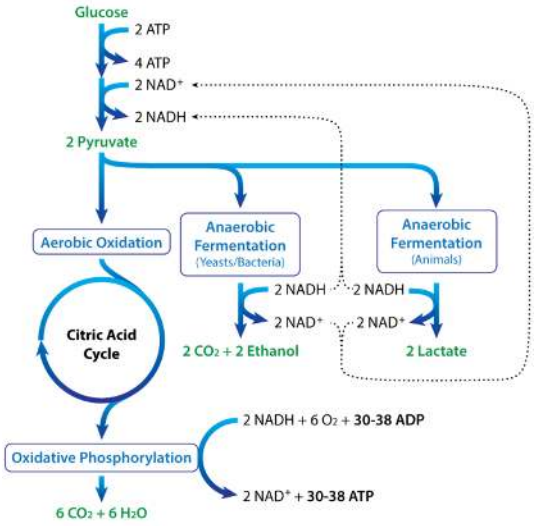

Figure 6.4.1: *Metabolic Redox*

The three regulated enzymes of glycolysis all catalyze reactions whose ΔG values are not close to zero, making manipulation of reaction direction non-trivial. Consequently, cells employ “work-around" reactions catalyzed by four different enzymes to favor gluconeogenesis, when appropriate.

Two of the enzymes (pyruvate carboxylase and PEP carboxykinase -PEPCK) catalyze reactions that bypass pyruvate kinase. F1,6BPase bypasses PFK and G6Pase bypasses hexokinase. Notably, pyruvate carboxylase and G6Pase are found in the mitochondria and endoplasmic reticulum, respectively, whereas the other two are found in the cytoplasm along with all of the enzymes of glycolysis. As a result, all of glycolysis and most of gluconeogenesis occurs in the cytoplasm. Controlling these pathways then becomes of critical importance because cells generally need to minimize the extent to which paired anabolic and catabolic pathways are occurring simultaneously, lest they waste energy and make no tangible product except heat. The mechanisms of controlling these pathways work, in some ways, in opposite fashions, called reciprocal regulation (see above).

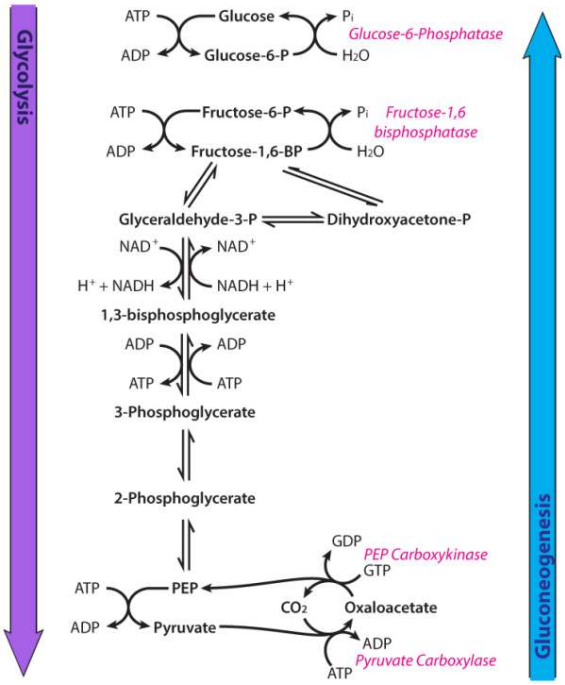

Figure 6.4.2: *Gluconeogensis and Glycolysis*

Besides reciprocal regulation, other mechanisms help control gluconeogenesis. First, PEPCK is controlled largely at the level of synthesis. Overexpression of PEPCK (stimulated by glucagon, glucocorticoids, and cAMP and inhibited by insulin) causes symptoms of diabetes. Pyruvate carboxylase is sequestered in the mitochondrion and is sensitive to acetyl-CoA, which is an allosteric activator. Acetyl-CoA concentrations increase as the citric acid cycle activity decreases. Glucose-6-phosphatase is present in low concentrations in many tissues, but is found most abundantly and importantly in the major gluconeogenic organs – the liver and kidney cortex.

6.5: Citric Acid Cycle

Cori Cycle

With respect to energy, the liver and muscles act complementarily. The liver is the major organ in the body for the synthesis of glucose. Muscles are major users of ATP. Actively exercising muscles generate lactate as a result of running glycolysis faster than the blood can deliver oxygen during periods of heavy exercise. As a consequence, the muscles go anaerobic and produce lactate. This lactate is of no use to muscle cells, so they dump it into the blood. Lactate travels in the blood to the liver, which takes it up and reoxidizes it back to pyruvate, catalyzed by the enzyme lactate dehydrogenase. Pyruvate in the liver is then converted to glucose by gluconeogenesis. The glucose thus made by the liver is dumped into the bloodstream where it is taken up by muscles and used for energy, completing a very important intercellular pathway known as the Cori cycle.

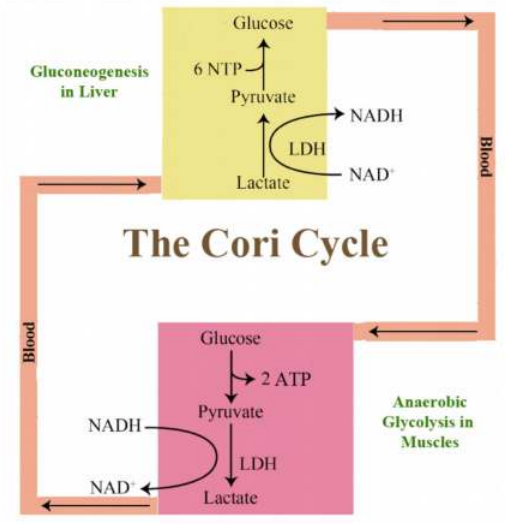

Figure 6.5.1: *The Cori Cycle*

Citric Acid Cycle

The primary catabolic pathway in the body is the citric acid cycle because it is here that oxidation to carbon dioxide occurs for breakdown products of the cell's major building blocks - sugars, fatty acids, amino acids. The pathway is cyclic (Figure 6.5.2) and thus, doesn't really have a starting or ending point. All of the reactions occur in the mitochondrion, though one enzyme is embedded in the organelle's membrane. As needs change, cells may use a subset of the reactions of the cycle to produce a desired molecule rather than to run the entire cycle (Figure 6.5.2).

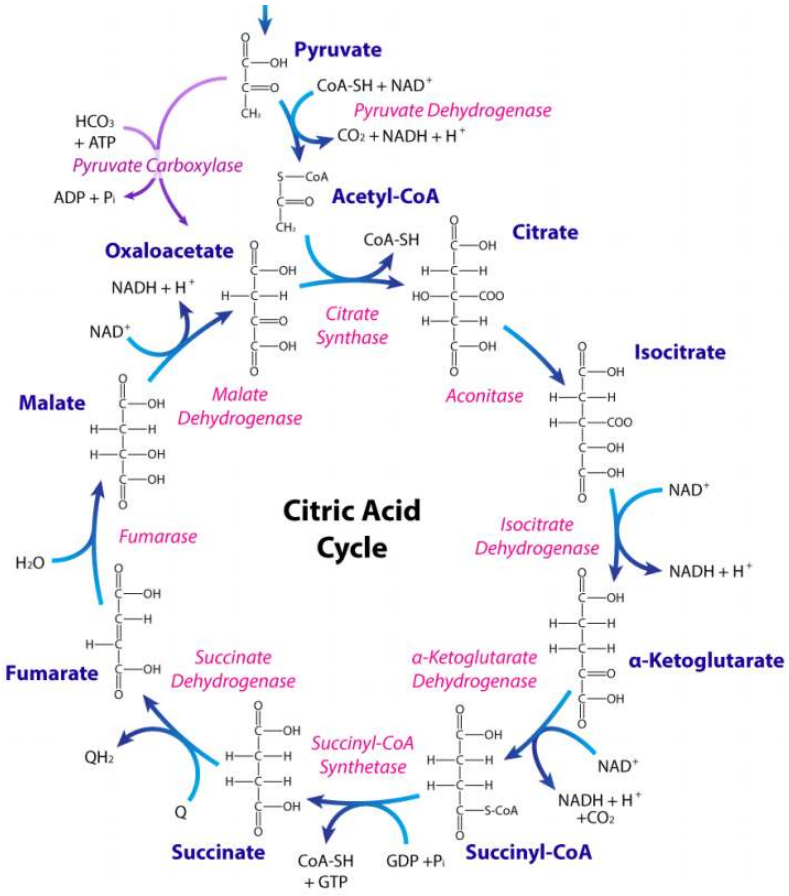

Figure 6.5.2: *The Citric Acid Cycle Reactions*

Focusing on the pathway itself, the traditional point to start discussion is addition of acetyl-CoA to oxaloacetate (OAA) to form citrate. Acetyl-CoA for the pathway can come from a variety of sources. They include pyruvate oxidation (from glycolysis and amino acid metabolism), fatty acid oxidation, and amino acid metabolism. The reaction joining it to OAA is catalyzed by citrate synthase and the $\Delta G^{\circ\prime}$ is fairly negative. This, in turn, helps to "pull" the reaction preceding it in the cycle (catalyzed by malate dehydrogenase).

In the next reaction, citrate is isomerized to isocitrate by action of the enzyme called aconitase. Isocitrate is a branch point in plants and bacteria for the glyoxylate cycle. Oxidative decarboxylation of isocitrate by isocitrate dehydrogenase produces the first NADH and yields alpha-ketoglutarate. This five carbon intermediate is a branch point for synthesis of glutamate. In addition, glutamate can also be made easily into this citric acid cycle intermediate. Decarboxylation of alpha-ketoglutarate yields succinyl-CoA and is catalyzed by alpha- ketoglutarate dehydrogenase. This enzyme is structurally very similar to pyruvate dehydrogenase and employs the same five coenzymes – NAD, FAD, CoASH, TPP, and lipoic acid.

The remainder of the citric acid cycle involves conversion of the four carbon succinyl-CoA into oxaloacetate. Succinyl-CoA is a branch point for the synthesis of heme. Succinyl-CoA is converted to succinate in a reaction catalyzed by succinyl-CoA synthetase (named for the reverse reaction) and a GTP is produced, as well – the only substrate level phosphorylation in the cycle. The energy for the synthesis of the GTP comes from hydrolysis of the high energy thioester bond between succinate and the CoA. Evidence for the high energy of a thioester bond is also evident in the citrate synthase reaction, which is also very energetically favorable. Succinate is also produced by metabolism of odd-chain fatty acids (see below).

Oxidation of succinate occurs in the next step, catalyzed by succinate dehydrogenase. This interesting enzyme both catalyzes this reaction and participates in the electron transport system, funneling electrons from the $\mathbf{FADH_2}$ it gains in the reaction to coenzyme Q. The product of the reaction, fumarate gains a water across its trans double bond in the next reaction, catalyzed by fumarase to form malate . Fumarate is also a byproduct of nucleotide metabolism and of the urea cycle . Malate is important also for transporting electrons across membranes in the malate aspartate shuttle and in ferrying carbon dioxide in C4 plants.

Conversion of malate to OAA is a rare biological oxidation that has a $\Delta G°'$ with a positive value. The reaction product includes NADH and the reaction is 'pulled' by the energetically favorable conversion of OAA to citrate in what was described above as the first reaction of the cycle. OAA intersects other important pathways, including amino acid metabolism (readily converted to aspartic acid), transamination (nitrogen movement) and gluconeogenesis.

6.6: Glyoxylate Pathway

A pathway related to the Citric Acid Cycle (CAC) is the glyoxylate pathway (Figure 6.6.1). This pathway, which overlaps all of the non-decarboxylation reactions of the CAC does not operate in animals, because they lack two enzymes necessary for the pathway – isocitrate lyase and malate synthase. Isocitrate lyase catalyzes the conversion of isocitrate into succinate and glyoxylate. Because of this, all six carbons of the CAC survive and do not end up as carbon dioxide.

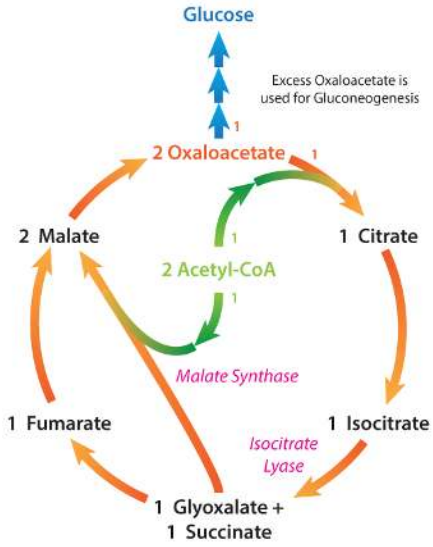

Figure 6.6.1: *The Glyoxalate Cycle*

Succinate continues through the remaining reactions of the CAC to produce oxaloacetate. Glyoxylate combines with another acetyl-CoA (one acetyl-CoA was used to start the cycle) to create malate (catalyzed by malate synthase). Malate can, in turn, be oxidized to oxaloacetate.

It is at this point that the pathway's contrast with the CAC is apparent. After one turn of the CAC, a single oxaloacetate is produced and it balances the single one used in the first reaction of the cycle. Thus, in the CAC, no net production of oxaloacetate is realized. By contrast, at the end of a turn of the glyoxylate cycle, two oxaloacetates are produced, starting with one. The extra oxaloacetate can then be used to make other molecules, including glucose in gluconeogenesis.

Because animals do not run the glyoxylate cycle, they cannot produce glucose from acetyl-CoA in net amounts, but plants and bacteria can. As a result, these organisms can turn acetyl-CoA from fat into glucose, while animals can't. Bypassing the decarboxylations (and substrate level phosphorylation) has its costs, however. Each turn of the glyoxylate cycle produces one FADH and one NADH instead of the three NADHs, one $\mathbf{FADH_2}$, and one GTP made in each turn of the CAC.

6.7: Acetyl-CoA Metabolism

Acetyl-CoA is one of the most “connected" metabolites in biochemistry, appearing in fatty acid oxidation/reduction, pyruvate oxidation, the citric acid cycle, amino acid anabolism/catabolism, ketone body metabolism, steroid/bile acid synthesis, and (by extension from fatty acid metabolism) prostaglandin synthesis. Most of these pathways will be dealt with separately. Here we will cover the last three.

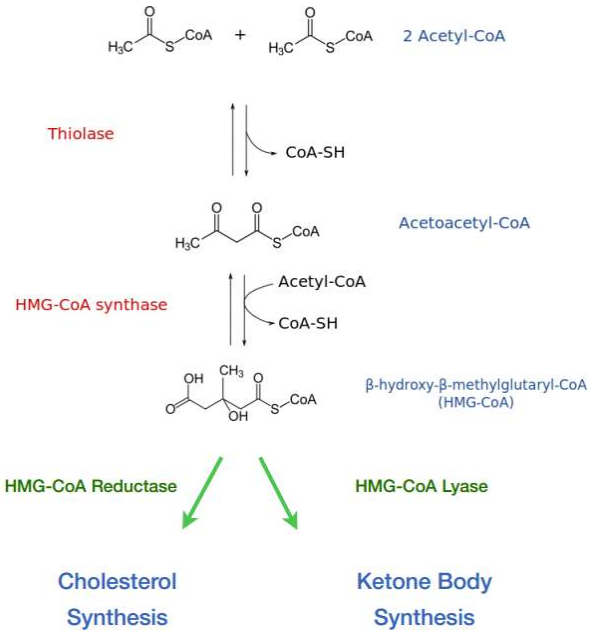

Figure 6.7.1: *Pathways for Cholestrol and Ketone Body Synthesis*

The pathways for ketone body synthesis and cholesterol biosynthesis overlap at the beginning. Each of these starts by combining two acetyl-CoAs together to make acetoacetyl-CoA. Not coincidentally, that is the next to last product of oxidation of fatty acids with even numbers of carbons. In fact, the enzyme that catalyzes the joining is the same as the one that catalyzes its breakage in fatty acid oxidation – thiolase. Thus, these pathways start by reversing the last step of the last round of fatty acid oxidation. Both pathways also include addition of two more carbons from a third acetyl-CoA to form Hydroxy-Methyl-Glutaryl-CoA, or HMG-CoA, as it is more commonly known. It is at this point that the two pathways diverge.

6.8: Cholesterol Metabolism

The cholesterol biosynthesis pathway is a long one and it requires significant amounts of reductive and ATP energy, which is why it is included here. Cholesterol has important roles in the body in membranes. It as also a precursor of steroid hormones and bile acids and its immediate metabolic precursor, 7-dehydrocholesterol, is a precursor of Vitamin D. The pathway leading to cholesterol is known as the isoprenoid pathway and branches of it lead to other molecules including other fat-soluble vitamins.

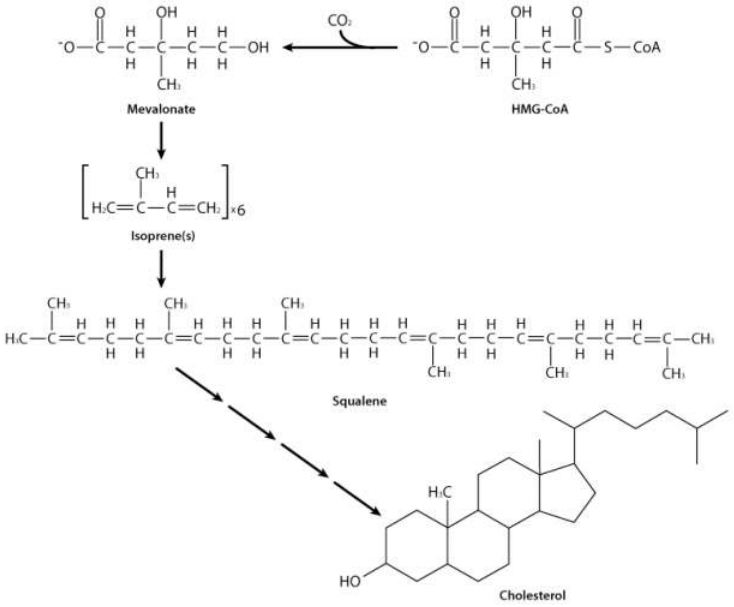

Figure 6.8.1: *The pathway to cholesterol*

From HMG-CoA, the enzyme HMG-CoA reductase catalyzes the formation of mevalonate. The reaction requires NADPH and results in release of coenzyme A and appears to be one of the most important regulatory steps in the synthesis pathway. The enzyme is regulated both by feedback inhibition (cholesterol inhibits it) and by covalent modification (phosphorylation inhibits it). The enzyme's synthesis is also regulated transcriptionally. When cholesterol levels fall, transcription of the gene increases.

Mevalonate gets phosphorylated twice and then decarboxylated to yield the five carbon intermediate known as isopentenyl-pyrophosphate (IPP). IPP is readily converted to dimethylallylpyrophosphate (DMAPP). These two five carbon compounds, also called isoprenes, are the building blocks for the synthesis of cholesterol and related compounds. This pathway is known as the isoprenoid pathway. It proceeds in the direction of cholesterol starting with the joining of IPP and DMAPP to form geranyl-pyrophosphate. Geranyl-pyrophosphate combines with another IPP to make farnesyl-pyrophosphate, a 15-carbon compound. Two farnesyl-pyrophosphates join to create the 30-carbon compound known as squalene. Squalene, in a complicated rearrangement involving reduction and molecular oxygen forms a cyclic intermediate known as lanosterol that resembles cholesterol. Conversion of lanosterol to cholesterol is a lengthy process involving 19 steps that occur in the endoplasmic reticulum.

Branching from cholesterol, one can form *Vitamin D* or the steroid hormones, which include the progestagens, androgens, estrogens, mineralocorticoids, and the glucocorticoids. The branch molecule for all of these is the cholesterol metabolite (and progestagen) known as *pregnenalone*. The progestagens are precursors of all of the other classes.

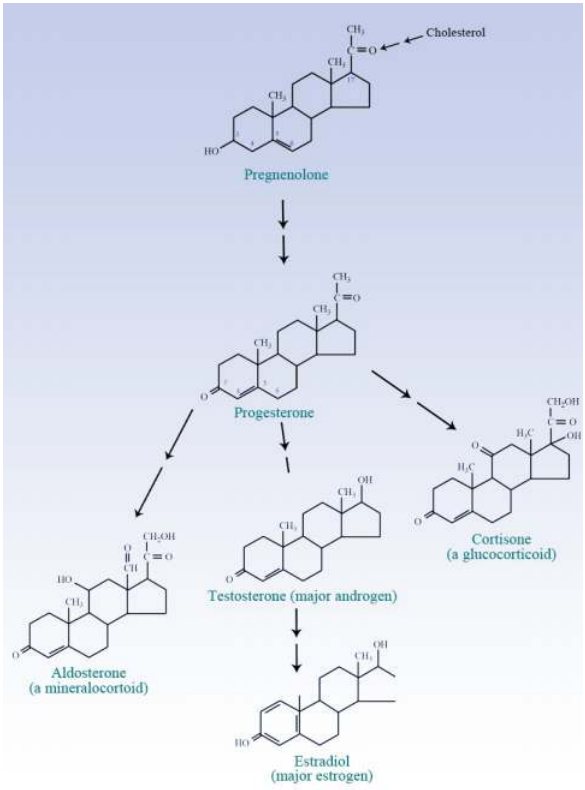

Figure 6.8.2: *Steroid Hormone Synthesis*

The estrogens are derived from the androgens in an interesting reaction that required formation of an aromatic ring. The enzyme catalyzing this reaction is known as an *aromatase* and it is of medical significance. The growth of some tumors is stimulated by estrogens, so aromatase inhibitors are prescribed to prevent the formation of estrogens and slow tumor growth. It is worth noting that synthesis of other fat soluble vitamins and chlorophyll also branches from the isoprenoid synthesis pathway at geranylpyrophosphate. Joining of two geranylgeranylpyrophosphates occurs in plants and bacteria and leads to synthesis of lycopene, which, in turn is a precursor of beta-carotene, the final precursor of Vitamin A. Vitamins E and K, as well as chlorophyll are all also synthesized from geranylgeranylpyrophosphate.

Bile Acid Metabolism

Another pathway from cholesterol leads to the polar bile acids, which are important for the solubilization of fat during digestion. Converting the very non-polar cholesterol to a bile acid involves oxidation of the terminal carbon on the side chain off the rings. Other alterations to increase the polarity of these compounds include hydroxylation of the rings and linkage to other polar compounds.

Common bile acids include cholic acid, chenodeoxycholic acid, glycocholic acid, taurocholic acid, and deoxycholic acid. Another important fact about bile acids is that their synthesis reduces the amount of cholesterol available and promotes uptake of LDLs by the liver. Normally bile acids are recycled efficiently resulting in limited reduction of cholesterol levels. However, inhibitors of the recycling promote reduction of cholesterol levels.

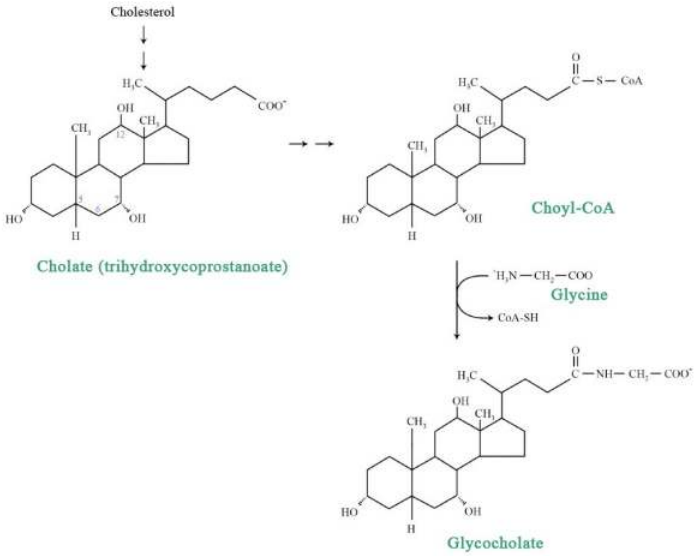

Figure 6.8.3: *Bile Salts*

6.9: Ketone Body Synthesis

In ketone body synthesis, an acetyl-CoA is split off from HMG-CoA, yielding acetoacetate, a four carbon ketone body that is somewhat unstable, chemically. It will decarboxylate spontaneously to some extent to yield acetone. Ketone bodies are made when the blood levels of glucose fall very low. Ketone bodies can be converted to acetyl-CoA, which can be used for ATP synthesis via the citric acid cycle. People who are very hypoglycemic (including some diabetics) will produce ketone bodies and these are often first detected by the smell of acetone on their breath.

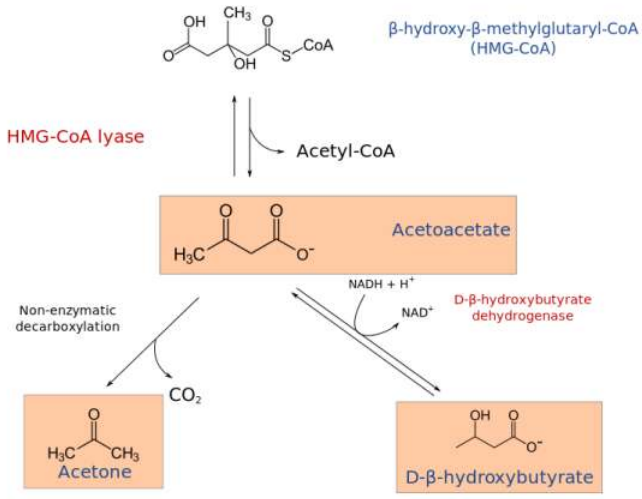

Figure 6.9.1: *Ketone Body Reactions*

Acetone is of virtually no use for energy production since it is not readily converted to acetyl-CoA. Consequently, cells convert acetoacetate into beta-hydroxybutyrate, which is more chemically stable. Though technically not a ketone, beta-hydroxybutyrate is frequently referred to as a ketone body. Upon arrival at a target cell, it can be oxidized back to acetoacetate with conversion to acetyl-CoA. Both acetoacetate and beta-hydroxybutyrate can cross the blood-brain barrier and provide important energy for the brain when glucose is limiting.

6.10: Prostaglandin Synthesis

The pathway for making prostaglandins is an extension of the fatty acid synthesis pathway (Figure 6.10.1). Prostaglandins, molecules associated with localized pain, are synthesized in cells from arachidonic acid (see previous page) which has been cleaved from membrane lipids. The enzyme catalyzing their synthesis is known as prostaglandin synthase, but is more commonly referred to as a cyclooxygenase (or COX) enzyme. Inhibition of the action of this enzyme is a strategy of non- steroidal pain relievers (also called NSAIDs), such as aspirin or ibuprofen. Inhibition of the release of arachidonic acid from membranes is the mechanism of action of steroidal anti-inffammatories, which inhibit the phospholipase $\mathbf{A_2(PLA_2)}$ that catalyzes the cleavage reaction.

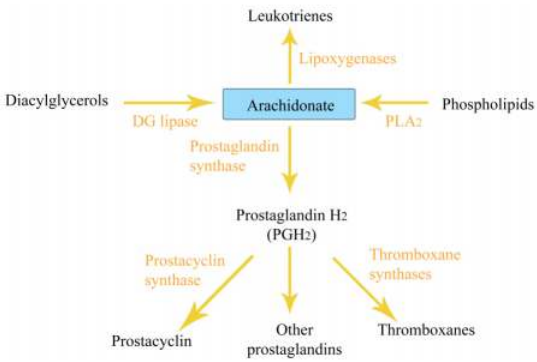

Figure 6.10.1: *Prostaglandin Metabolism*

6.11: Fatty Acid Oxidation

Breakdown of fats yields fatty acids and glycerol. Glycerol can be readily converted to DHAP for oxidation in glycolysis or synthesis into glucose in gluconeogenesis. Fatty acids are broken down in two carbon units of acetyl-CoA. To be oxidized, they must be transported through the cytoplasm attached to coenzyme A and moved into mitochondria. The latter step requires removal of the CoA and attachment of the fatty acid to a molecule of carnitine. The carnitine complex is transported across the inner membrane of the mitochondrion after which the fatty acid is reattached to coenzyme A in the mitochondrial matrix.

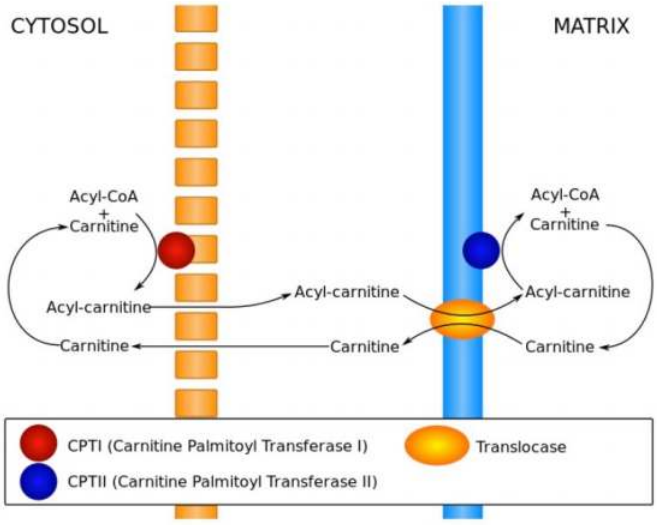

Figure 6.11.1: *Movement of Acyl-CoAs into the Mitochondrial Matrix*

The process of fatty acid oxidation, called beta oxidation, is fairly simple. The reactions all occur between carbons 2 and 3 (with #1 being the one linked to the CoA) and sequentially include the following:

1. dehydrogenation to create $\mathbf{FADH_2}$ and a fatty acyl group with a double bond in the trans configuration;
2. hydration across the double bond to put a hydroxyl group on carbon 3 in the L configuration;
3. oxidation of the hydroxyl group to make a ketone; and
4. thiolytic cleavage to release acetyl-CoA and a fatty acid two carbons shorter than the starting one.

Unsaturated fatty acids complicate the picture a bit (see below), primarily because they have cis bonds, for the most part, if they are of biological origin and these must be converted to the relevant trans intermediate made in step 1. Sometimes the bond must be moved down the chain, as well, in order to be positioned properly. Two enzymes (described below) handle all the necessary isomerizations and moves necessary to oxidize all of the unsaturated fatty acids.

Enzymes of Beta Oxidation

The reactions of fatty acid oxidation are notable in mirroring the oxidations in the latter half of the citric acid cycle – dehydrogenation of succinate to make a transdouble bond (fumarate), hydration across the double bond to make L-malate and oxidation of the hydroxyl to make a ketone (oxaloacetate). Two of the enzymes of beta-oxidation are notable. The first is acyl-CoA dehydrogenase, which catalyzes the initial dehydrogenation and yields FADH2. It comes in three different forms – ones that work on long, medium, or short chain length fatty acids. The first of these is sequestered in the peroxisome of animals whereas the others are found in the mitochondria. Plants and yeast perform beta oxidation exclusively in the peroxisome. The most interesting of the acyl-CoA dehydrogenases is the one that works on medium length fatty acids. This one, which is the one most commonly deficient in animals, has been linked to sudden infant death syndrome.

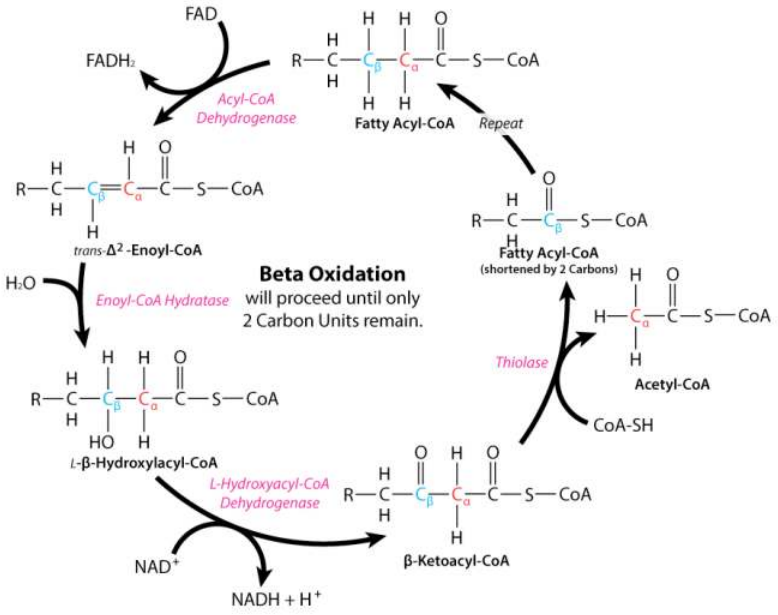

Figure 6.11.2: *Beta Oxidation of Fatty Acids*

Reactions two and three in beta oxidation are catalyzed by enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase, respectively. The latter reaction yields an NADH. The final enzyme of beta oxidation is thiolase and this enzyme is notable in not only catalyzing the formation of acetyl-CoAs in beta oxidation, but also catalyzing the joining of two acetyl-CoAs (essentially the reversal of the last step of beta oxidation) to form acetoacetyl-CoA– essential for the pathways of ketone body synthesis and cholesterol biosynthesis.

Oxidation of Odd-Chain Fatty Acids

Though most fatty acids of biological origin have even numbers of carbons, not all of them do. Oxidation of fatty acids with odd numbers of carbons ultimately produces an intermediate with three carbons called propionyl-CoA, which cannot be oxidized further in the beta-oxidation pathway. Metabolism of this intermediate is odd. Sequentially, the following steps occur:

1. carboxylation to make D-methylmalonyl-CoA;
2. isomerization to L-methylmalonyl-CoA;
3. rearrangement to form succinyl-CoA. The last step of the process utilizes the enzyme methylmalonyl-CoA mutase, which uses the $\mathbf{B_12}$ coenzyme in its catalytic cycle. Succinyl-CoA can then be metabolized in the citric acid cycle.

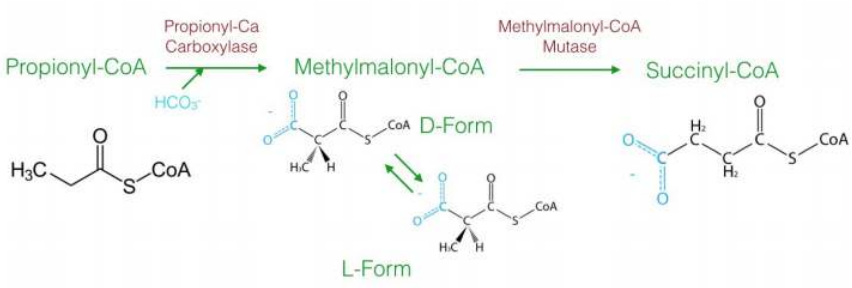

Figure 6.11.3: *Odd-Chain Fatty Acid Oxidation*

Unsaturated Fatty Acid Oxidation

As noted above, oxidation of unsaturated fatty acids requires two additional enzymes to the complement of enzymes for beta oxidation. If the beta oxidation of the fatty acid produces an intermediate with a cis bond between carbons three and four, cis-Δ3-Enoyl-CoA Isomerase will convert the bond to a trans bond between carbons two and three and beta oxidation can proceed as normal.

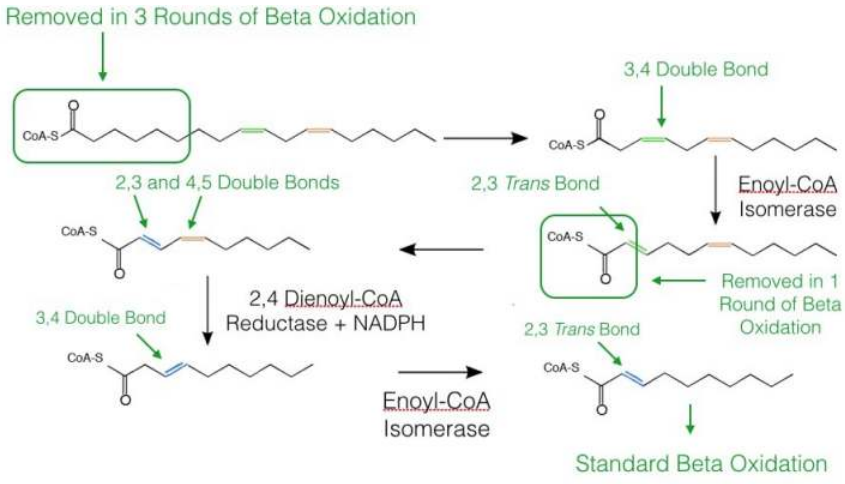

Figure 6.11.4: *Unsaturated Fatty Acid Oxidation*

On the other hand, if beta oxidation produces an intermediate with a cis double bond between carbons four and five, the first step of beta oxidation (dehydrogenation between carbons two and three) occurs to produce an intermediate with a trans double bond between carbons two and three and a cis double bond between carbons four and five. The enzyme 2,4 dienoyl CoA reductase reduces this intermediate (using NADPH) to one with a single cis bond between carbons three and four. This intermediate is then identical to the one acted on by cis-Δ3-Enoyl-CoA Isomerase above, which converts it into a regular beta oxidation intermediate, as noted above.

Alpha Oxidation

Yet another consideration for oxidation of fatty acids is alpha oxidation. This pathway is necessary for catabolism of fatty acids that have branches in their chains. For example, breakdown of chlorophyll's phytol group yields phytanic acid, which undergoes hydroxylation and oxidation on carbon number two (in contrast to carbon three of beta oxidation), followed by decarboxylation and production of a branched intermediate that can be further oxidized by the beta oxidation pathway. Though alpha oxidation is a relatively minor metabolic pathway, the inability to perform the reactions of the pathway leads to Refsum's disease where accumulation of phytanic acid leads to neurological damage.

6.12: Fatty Acid Synthesis

Synthesis of fatty acids occurs in the cytoplasm and endoplasmic reticulum of the cell and is chemically similar to the beta-oxidation process, but with a couple of key differences. The first of these occur in preparing substrates for the reactions that grow the fatty acid. Transport of acetyl-CoA from the mitochondria occurs when it begins to build up. Two molecules can play roles in moving it to the cytoplasm – citrate and acetylcarnitine. Joining of oxaloacetate with acetyl-CoA in the mitochondrion creates citrate which moves across the membrane, followed by action of citrate lyase in the cytoplasm of the cell to release acetyl-CoA and oxaloacetate. Additionally, when free acetyl-CoA accumulates in the mitochondrion, it may combine with carnitine and be transported out to the cytoplasm.

Starting with two acetyl-CoA, one is converted to malonyl-CoA by carboxylation catalyzed by the enzyme acetyl-CoA carboxylase (ACC), the only regulatory enzyme of fatty acid synthesis (Figure **6.12.1**). Next, both molecules have their CoA portions replaced by a carrier protein known as ACP (acyl-carrier protein) to form acetyl-ACP and malonyl-ACP. Joining of a fatty acyl-ACP (in this case, acetyl-ACP) with malonyl-ACP splits out the carboxyl that was added and creates the intermediate at the upper right in the figure at left.

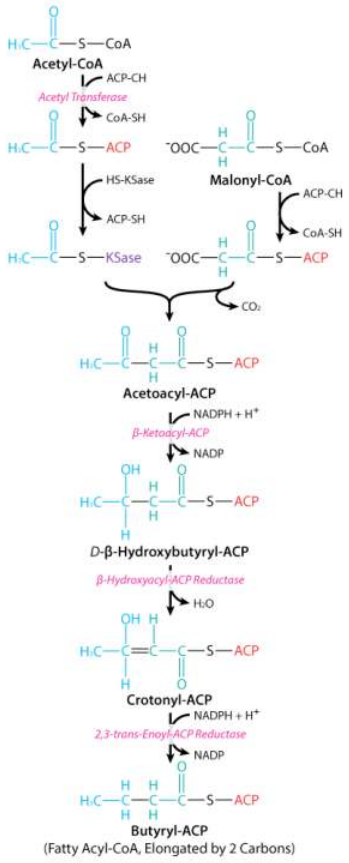

Figure* 6.12.1*: *Fatty Acid Synthesis*

From this point forward, the chemical reactions resemble those of beta oxidation reversed. First, the ketone is reduced to a hydroxyl using NADPH. In contrast to the hydroxylated intermediate of beta oxidation, the beta intermediate here is in the D-configuration. Next, water is removed from carbons 2 and 3 of the hydroxyl intermediate to produce a trans doubled bonded molecule. Last, the double bond is hydrogenated to yield a saturated intermediate. The process cycles with the addition of another malonyl-ACP to the growing chain until ultimately an intermediate with 16 carbons is produced (palmitoyl-CoA). At this point, the cytoplasmic synthesis ceases.

Enzymes of Fatty Acid Synthesis

Acetyl-CoA carboxylase, which catalyzes synthesis of malonyl-CoA, is the only regulated enzyme in fatty acid synthesis. Its regulation involves both allosteric control and covalent modification. The enzyme is known to be phosphorylated by both AMP Kinase and Protein Kinase A. Dephosphorylation is stimulated by phosphatases activated by insulin binding. Dephosphorylation activates the enzyme and favors its assembly into a long polymer, while phosphorylation reverses the process.Citrate acts as an allosteric activator and may also favor polymerization. Palmitoyl-CoA allosterically inactivates it.

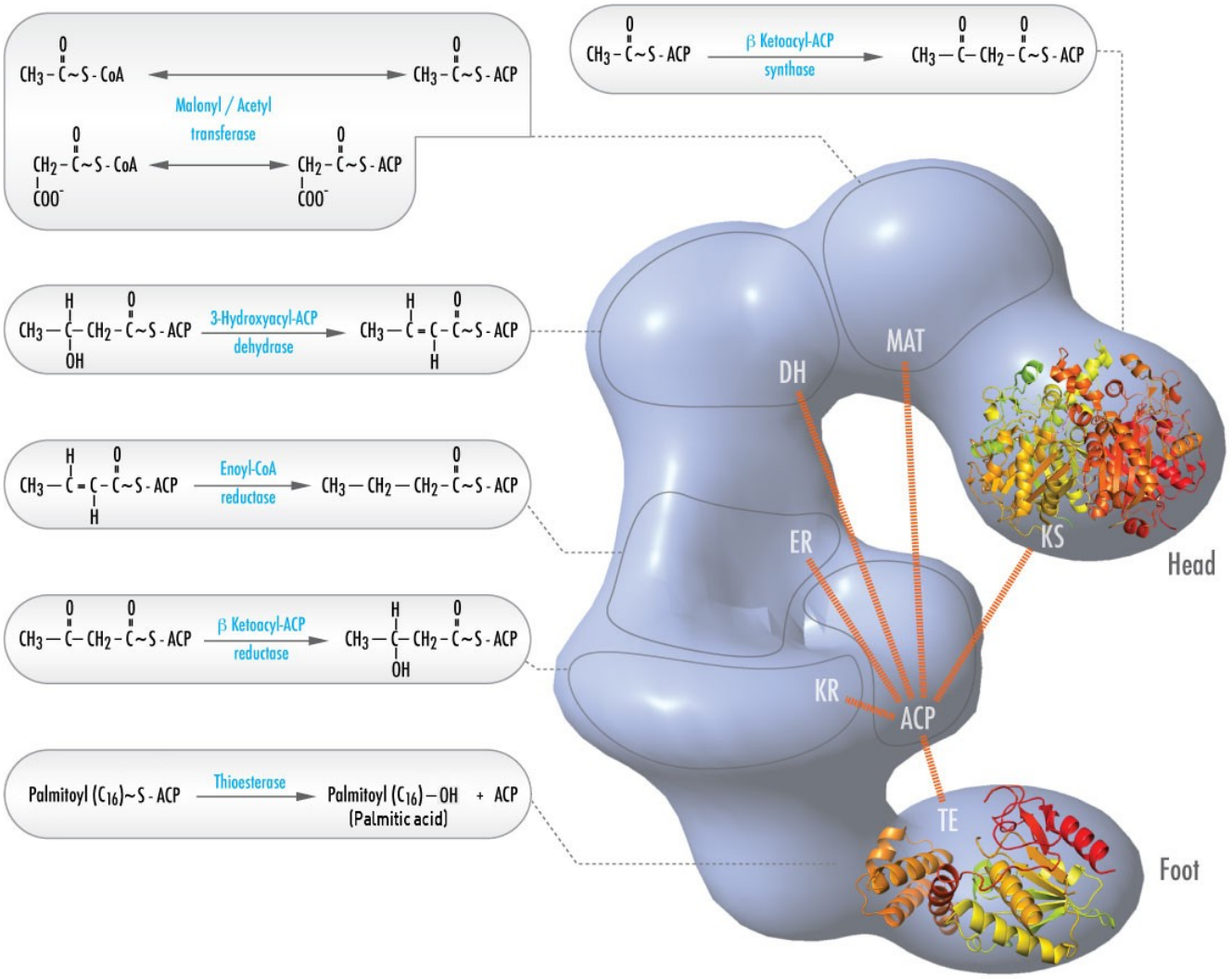

Figure **6.12.2**: Fatty Acid Synthase (FAS) revised model with positions of polypeptides, three catalytic domains and their corresponding reactions, visualization by Kosi Gramatikoff. Note that FAS is only active as a homodimer rather than the monomer pictured. (Public Domain; Kosigrim)

In animals, six different catalytic activities necessary for the remaining catalytic actions to fully make palmitoyl-CoA are contained in a single complex called Fatty Acid Synthase (Figure **6.12.2**). These include transacylases for swapping CoA with ACP on acetyl-CoA and malonyl-CoA; a synthase to catalyze addition of the two carbon unit from the three carbon malonyl-ACP in the first step of the elongation process; a reductase to reduce the ketone; a dehydrase to catalyze removal of water, and a reductase to reduce the trans double bond. In bacteria, these activities are found on separate enzymes and are not part of a complex.

Elongation of Fatty Acids

Elongation to make fatty acids longer than 16 carbons occurs in the endoplasmic reticulum and is catalyzed by enzymes described as elongases. Mitochondria also can elongate fatty acids, but their starting materials are generally shorter than 16 carbons long. The mechanisms in both environments are similar to those in the cytoplasm (a malonyl group is used to add two carbons, for example),

but CoA is attached to the intermediates, not ACP. Further, whereas cytoplasmic synthesis employs the fatty acid synthase complex (Figure **6.12.2**), the enzymes in these organelles are separable and not part of a complex.

Desaturation of Fatty Acids

Fatty acids are synthesized in the saturated form and desaturation occurs later. Enzymes called desaturases catalyze the formation of cis double bonds in mature fatty acids. These enzymes are found in the endoplasmic reticulum. Animals are limited in the desaturated fatty acids they can make, due to an inability to catalyze reactions beyond carbons 9 and 10. Thus, humans can make oleic acid, but cannot synthesis linoleic acid or linolenic acid. Consequently, these two must be provided in the diet and are referred to as essential fatty acids.

6.13: Metabolism of Fat

Breakdown of fat in adipocytes requires catalytic action of three enzymes, hormone sensitive triacylglycerol lipase (called LIPE) to remove the first fatty acid from the fat, diglyceride lipase to remove the second one and monoglyceride lipase to remove the third. Of these, only LIPE is regulated and it appears to be the rate limiting reaction. Synthesis of fat starting with glycerol-3-phosphate requires action of acyl transferase enzymes, such as glycerol-3-phosphate acyl transferase, which catalyze addition of fatty acids to the glycerol backbone.

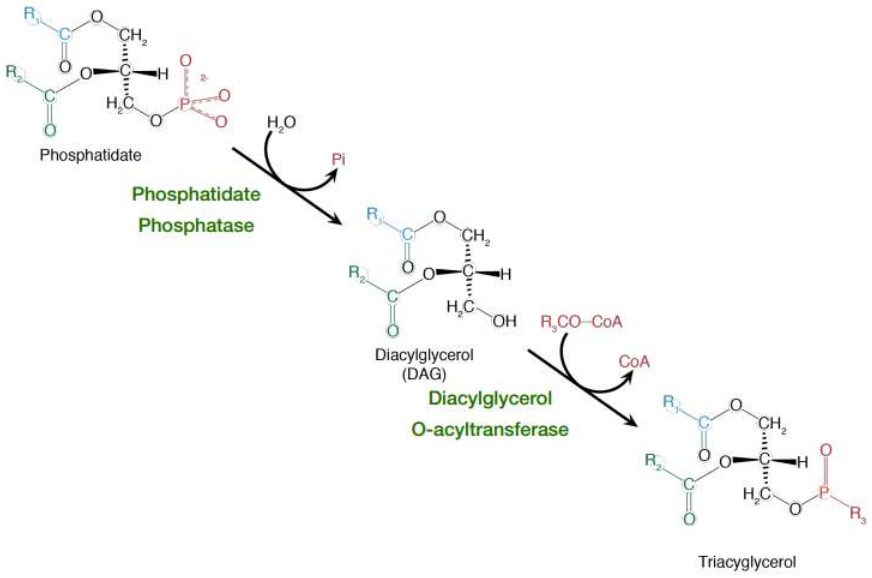

Figure 6.13.1: *Synthesis of Fat*

Interestingly, there appear to be few controls of the metabolism of fatty acids. The primary control of their oxidation is availability. One way to control that is by control of the breakdown of fat. This process, which can be stimulated by the epinephrine kinase cascade, is controlled through LIPE, found in adipocytes (fat-containing cells). Breakdown of fat in apidocytes requires action of three enzymes, each hydrolyzing one fatty acid from the glycerol backbone. As noted earlier, only HSTL, which catalyzes the first hydrolysis, is regulated.

Synthesis of fat requires glycerol-3-phosphate (or DHAP) and three fatty acids. In the first reaction, glycerol-3-phosphate is esterified at position 1 with a fatty acid, followed by a duplicate reaction at position 2 to make phosphatidic acid. This molecule, which is an intermediate in the synthesis of both fats and phosphoglycerides, gets dephosphorylated to form diacylglycerol before the third esterification to make a fat.

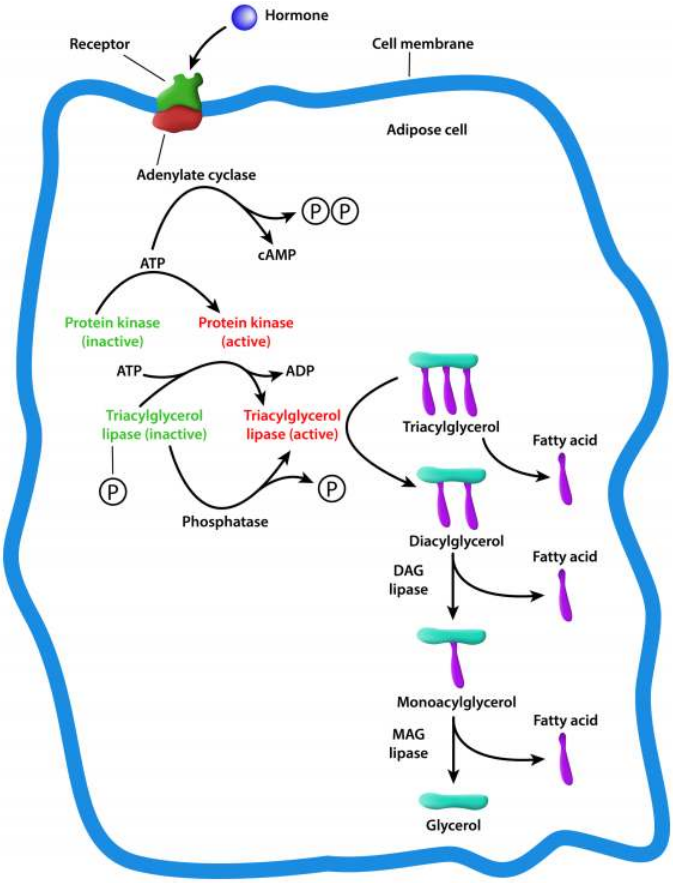

Figure 6.13.2: *Activation of Fat Hydrolysis*

Glycerophospholipid Metabolism

Phosphatidic acid, as noted above, is an important intermediate in the metabolism of glycerophospholipids. These compounds, which are important membrane constituents, can be synthesized in several ways.

6.14: Connections to Other Pathways

There are several connections between metabolism of fats and fatty acids to other metabolic pathways. As noted, phosphatidic acid is an intermediate in the synthesis of triacylglycerols, as well as of other lipids, including phosphoglycerides. Diacylglycerol (DAG), which is an intermediate in fat synthesis, also acts as a messenger in some signaling systems. Fatty acids twenty carbons long based on arachidonic acid (also called eicosanoids) are precursors of the classes of molecules known as leukotrienes and prostaglandins. The latter, in turn, are precursors of the class of molecules known as thromboxanes. The ultimate products of beta oxidation are acetyl-CoA molecules and these can be assembled by the enzyme thiolase to make acetoacetyl-CoA, which is a precursor of both ketone bodies and the isoprenoids, a broad category of compounds that include steroid hormones, cholesterol, bile acids, and the fat soluble vitamins, among others.

CHAPTER OVERVIEW

7: Metabolism II

In the last chapter, we focused on metabolic pathways that played important oxidative/reductive roles relative to cellular energy. In this chapter, the pathways that we cover have lesser roles from an energy perspective, but important roles, nonetheless, in catabolism and anabolism of building blocks of proteins and nucleic acids, nitrogen balance, and sugar balance. In a sense, these might be thought of as the “kitchen sink" pathways, but it should be noted that all cellular pathways are important. In this second section of metabolism, we cover metabolic pathways that do not have a strong emphasis on oxidation/reduction.

7.1: Carbohydrate Storage and Breakdown

Carbohydrates are important cellular energy sources. They provide energy quickly through glycolysis and passing of intermediates to pathways, such as the citric acid cycle, amino acid metabolism (indirectly), and the pentose phosphate pathway. It is important, therefore, to understand how these important molecules are made.

Plants are notable in storing glucose for energy in the form of amylose and amylopectin (see and for structural integrity in the form of cellulose. These structures differ in that cellulose contains glucoses solely joined by beta-1,4 bonds, whereas amylose has only alpha1,4 bonds and amylopectin has alpha 1,4 and alpha 1,6 bonds.Animals store glucose primary in liver and muscle in the form of a compound related to amylopectin known as glycogen. The structural differences between glycogen and amylopectin are solely due to the frequency of the alpha 1,6 branches of glucoses. In glycogen they occur about every 10 residues instead of every 30-50, as in amylopectin.

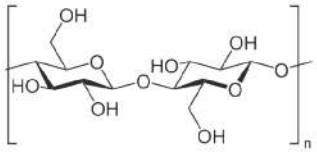

Figure 7.1.1: *The Repeating Structure of Cellulose*

Glycogen provides an additional source of glucose besides that produced via gluconeogenesis. Because glycogen contains so many glucoses, it acts like a battery backup for the body, providing a quick source of glucose when needed and providing a place to store excess glucose when glucose concentrations in the blood rise. The branching of glycogen is an important feature of the molecule metabolically as well. Since glycogen is broken down from the "ends" of the molecule, more branches translate to more ends, and more glucose that can be released at once. Breakdown of glycogen involves

1. release of glucose-1- phosphate (G1P),
2. rearranging the remaining glycogen (as necessary) to permit continued breakdown, and
3. conversion of G1P to G6P for further metabolism. G6P can be 1) broken down in glycolysis, 2) converted to glucose by gluconeogenesis, and 3) oxidized in the pentose phosphate pathway.

Just as in gluconeogenesis, the cell has a separate mechanism for glycogen synthesis that is distinct from glycogen breakdown. As noted previously, this allows the cell to separately control the reactions, avoiding futile cycles, and enabling a process to occur efficiently (synthesis of glycogen) that would not occur if it were simply the reversal of glycogen breakdown.

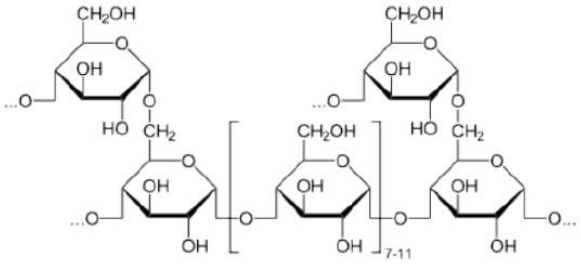

Figure 7.1.2: *The Repeating Unit of Glycogen*

Synthesis of glycogen starts with G1P, which is converted to an 'activated' intermediate, UDP-glucose. This activated intermediate is what 'adds' the glucose to the growing glycogen chain in a reaction catalyzed by the enzyme known as glycogen synthase. Once the glucose is added to glycogen, the glycogen molecule may need to have branches inserted in it by the enzyme known as branching enzyme.

Glycogen Breakdown

Glycogen phosphorylase (sometimes simply called phosphorylase) catalyzes breakdown of glycogen into Glucose-1-Phosphate (G1P). The reaction, (see above right) that produces G1P from glycogen is a phosphorolysis, not a hydrolysis reaction. The distinction is that hydrolysis reactions use water to cleave bigger molecules into smaller ones, but phosphorolysis reactions use

phosphate instead for the same purpose. Note that the phosphate is just that - it does NOT come from ATP. Since ATP is not used to put phosphate on G1P, the reaction saves the cell energy.

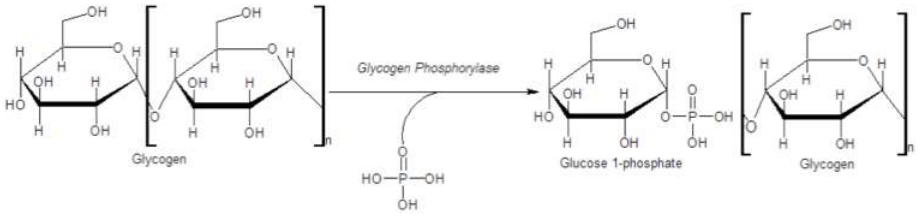

Figure 7.1.3: *Phosphorolysis of Glycogen*

Glycogen phosphorylase will only act on non-reducing ends of a glycogen chain that are at least 5 glucoses away from a branch point. A second enzyme, Glycogen Debranching Enzyme (GDE), is therefore needed to convert alpha(1-6) branches to alpha(1-4) branches. GDE acts on glycogen branches that have reached their limit of hydrolysis with glycogen phosphorylase. GDE acts to transfer a trisaccharide from a 1,6 branch onto an adjacent 1,4 branch, leaving a single glucose at the 1,6 branch. Note that the enzyme also catalyzes the hydrolysis of the remaining glucose at the 1,6 branch point. Thus, the breakdown products from glycogen are G1P and glucose (mostly G1P, however). Glucose can, of course, be converted to Glucose-6-Phosphate (G6P) as the first step in glycolysis by either hexokinase or glucokinase. G1P can be converted to G6P by action of an enzyme called phosphoglucomutase. This reaction is readily reversible, allowing G6P and G1P to be interconverted as the concentration of one or the other increases. This is important, because phosphoglucomutase is needed to form G1P for glycogen biosynthesis.

Regulation of Glycogen Metabolism

Regulation of glycogen metabolism is complex, occurring both allosterically and via hormone-receptor controlled events that result in protein phosphorylation or dephosphorylation. In order to avoid a futile cycle of glycogen synthesis and breakdown simultaneously, cells have evolved an elaborate set of controls that ensure only one pathway is primarily active at a time.

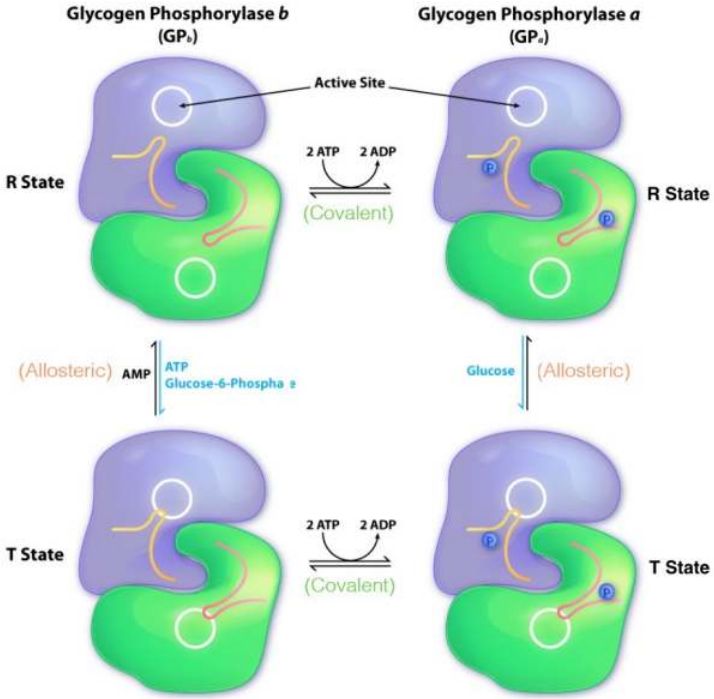

Figure 7.1.4: *Regulation of Glycogen Phosphorylase*

Regulation of glycogen metabolism is managed by the enzymes glycogen phosphorylase and glycogen synthase. Glycogen phosphorylase is regulated by both allosteric factors (ATP, G6P, AMP, and glucose) and by covalent modification (phosphorylation/dephosphorylation). Its regulation is consistent with the energy needs of the cell. High energy substrates (ATP, G6P, glucose) allosterically inhibit GP, while low energy substrates (AMP, others) allosterically activate it.

GPa/GPb Allosteric Regulation

Glycogen phosphorylase exists in two different covalent forms – one form with phosphate (called GPa here) and one form lacking phosphate (GPb here). GPb is converted to GPa by phosphorylation by an enzyme known as phosphorylase kinase. GPa and GPb can each exist in an 'R' state and a 'T' state. For both GPa and GPb, the R state is the more active form of the enzyme. GPa's negative allosteric effector (glucose) is usually not abundant in cells, so GPa does not .ip into the T state often. There is no positive allosteric effector of GPa, so when glucose is absent, GPa automatically flips into the R (more active) state.

GPb can convert from the T state to the GPb R state by binding AMP. Unless a cell is low in energy, AMP concentration is low. Thus GPb is not converted to the R state very often. On the other hand, ATP and/or G6P are usually present at high enough concentration in cells that GPb is readily flipped into the T state.

GPa/GPb Covalent Conversion

Because the relative amounts of GPa and GPb largely govern the overall process of glycogen breakdown, it is important to understand the controls on the enzymes that interconvert GPa and GPb. This is accomplished by the enzyme Phosphorylase Kinase, which transfers phosphates from 2 ATPs to GPb to form GPa. Phosphorylase kinase has two covalent forms – phosphorylated (active) and dephosphorylated (inactive). It is phosphorylated by the enzyme Protein Kinase A (PKA). Another way to activate the enzyme is with calcium. Phosphorylase kinase is dephosphorylated by the same enzyme, phosphoprotein phosphatase, that removes phosphate from GPa.

PKA is activated by cAMP, which is, in turn produced by adenylate cyclase after activation by a G-protein. G-proteins are activated ultimately by binding of ligands to specific 7-TM receptors, also known as G-protein coupled receptors. These are discussed in greater detail in Chapter 8. Common ligands for these receptors include epinephrine (binds beta-adrenergic receptor) and glucagon (binds glucagon receptor). Epinephrine exerts it greatest effects on muscle and glucagon works preferentially on the liver.

Turning Off Glycogen Breakdown

Turning OFF signals is as important, if not more so, than turning them ON. The steps in the glycogen breakdown regulatory pathway can be reversed at several levels. First, the ligand can leave the receptor. Second, the G-proteins have an inherent GTPase activity that serves to turn them off over time. Third, cells have phosphodiesterase (inhibited by caffeine) for breaking down cAMP. Fourth, an enzyme known as phosphoprotein phosphatase can remove phosphates from phosphorylase kinase (inactivating it) AND from GPa, converting it to the much less active GPb.

Glycogen Synthesis

The anabolic pathway contrasting with glycogen breakdown is that of glycogen synthesis. Just as cells reciprocally regulate glycolysis and gluconeogenesis to prevent a futile cycle, so too do cells use reciprocal schemes to regulate glycogen breakdown and synthesis. Let us first consider the steps in glycogen synthesis. 1) Glycogen synthesis from glucose involves phosphorylation to form G6P, and isomerization to form G1P (using phosphoglucomutase common to glycogen breakdown). G1P is reacted with UTP to form UDP-glucose in a reaction catalyzed by UDP-glucose pyrophosphorylase. Glycogen synthase catalyzes synthesis of glycogen by joining carbon #1 of the UDPG-derived glucose onto the carbon #4 of the non-reducing end of a glycogen chain. to form the familiar alpha(1,4) glycogen links. Another product of the reaction is UDP.

It is also worth noting in passing that glycogen synthase will only add glucose units from UDPG onto a preexisting glycogen chain that has at least four glucose residues. Linkage of the first few glucose units to form the minimal "primer" needed for glycogen synthase recognition is catalyzed by a protein called glycogenin, which attaches to the first glucose and catalyzes linkage of the first eight glucoses by alpha(1,4) bonds. 3) The characteristic alpha(1,6) branches of glycogen are the products of an enzyme known as Branching Enzyme. Branching Enzyme breaks alpha(1,4) chains and carries the broken chain to the carbon #6 and forms an alpha(1,6) linkage.

Regulation of Glycogen Synthesis

The regulation of glycogen biosynthesis is reciprocal to that of glycogen breakdown. It also has a cascading covalent modification system similar to the glycogen breakdown system described above. In fact, part of the system is identical to glycogen breakdown. Epinephrine or glucagon signaling can stimulate adenylate cyclase to make cAMP, which activates PKA, which activates phosphorylase kinase.

In glycogen breakdown, phosphorylase kinase phosphorylates GPb to the more active form, GPa. In glycogen synthesis, protein kinase A phosphorylates the active form of glycogen synthase (GSa), and converts it into the usually inactive b form (called GSb). Note the conventions for glycogen synthase and glycogen phosphorylase. For both enzymes, the more active forms are called the 'a' forms (GPa and GSa) and the less active forms are called the 'b' forms (GPb and GSb). The major difference, however, is that GPa has a phosphate, but GSa does not and GPb has no phosphate, but GSb does. Thus phosphorylation and dephosphorylation have opposite effects on the enzymes of glycogen metabolism. This is the hallmark of reciprocal regulation. It is of note that the less active glycogen synthase form, GSb, can be activated by G6P. Recall that G6P had the exactly opposite effect on GPb.

Figure 7.1.5: *Glycogen Synthase*

Glycogen synthase, glycogen phosphorylase (and phosphorylase kinase) can be dephosphorylated by several enzymes called phosphatases. One of these is called Protein Phosphatase and it is activated when insulin binds to a receptor in the cell membrane. It causes PP to be activated, stimulating dephosphorylation, and thus activating glycogen synthesis and inhibiting glycogen breakdown. Again, there is reciprocal regulation of glycogen synthesis and degradation.

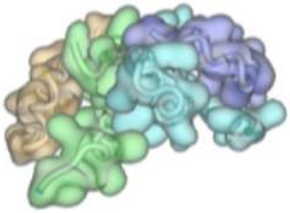

Figure 7.1.6: *The 3D structure of Insulin*

Maintaining Blood Glucose Levels

After a meal, blood glucose levels rise and insulin is released. It simultaneously stimulates uptake of glucose by cells and incorporation of it into glycogen by activation of glycogen synthase and inactivation of glycogen phosphorylase. When blood glucose levels fall, GPa gets activated (stimulating glycogen breakdown to raise blood glucose) and GSb is formed (stopping glycogen synthesis).

7.2: Pentose Phospate Pathway

The Pentose Phosphate Pathway (PPP) is one that many students are confused by. Perhaps the reason for this is that it does not really have a single direction in which it proceeds, as will be apparent below.

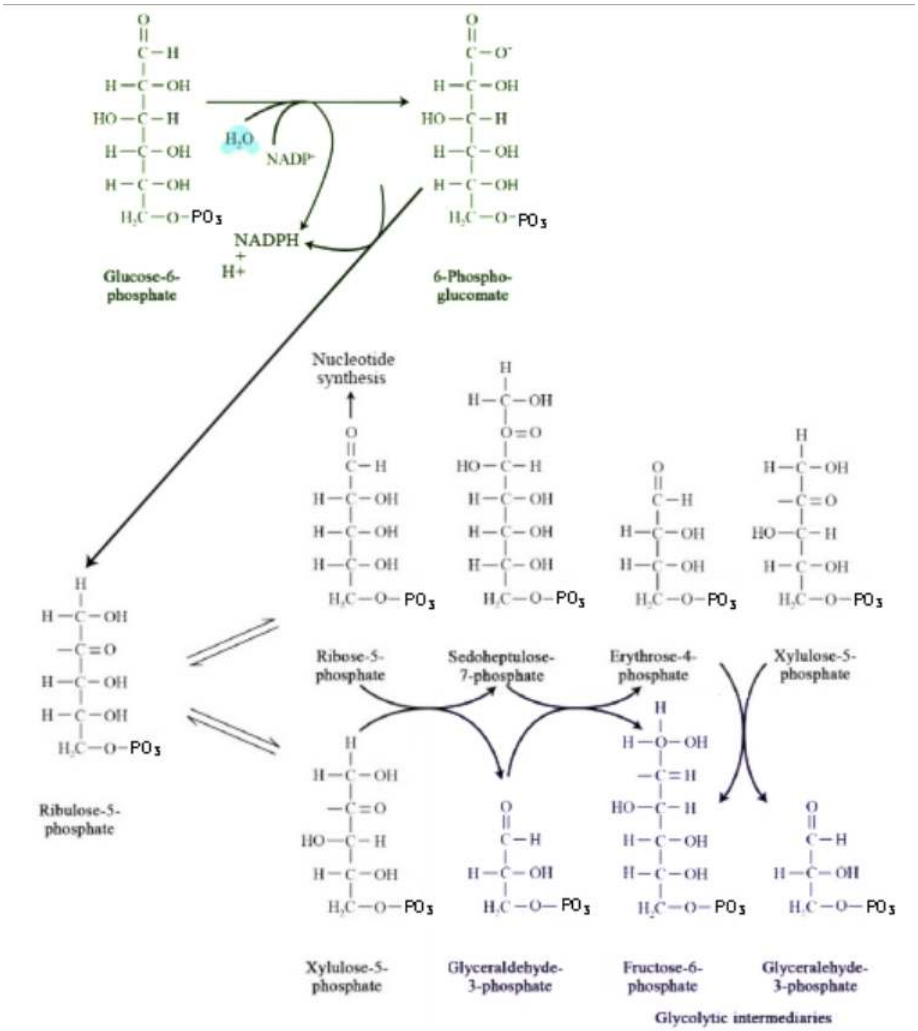

Figure 7.2.1: *Pentose Phosphate Pathway*

Portions of the PPP are similar to the Calvin Cycle of plants, also known as the dark reactions of photosynthesis. We discuss these reactions separately in the next section. The primary functions of the PPP are to produce NADPH (for use in anabolic reductions), ribose-5-phosphate (for making nucleotides), and erythrose-4-phosphate (for making aromatic amino acids). Three molecular intermediates of glycolysis can funnel into PPP (or be used as usual in glycolysis). They include G6P, fructose-6-phosphate (in two places), and glyceraldehyde-3-phosphate (also in two places).

A starting point for the pathway (though there are other entry points) is the oxidative phase. It includes two reactions generating NADPH. In the first of these, oxidation of glucose-6-phosphate (catalyzed by glucose-6-phosphate dehydrogenase), produces NADPH and 6-phosphogluconolactone. 6-phosphogluconolactone spontaneously gains water and loses a proton to become 6-phosphogluconate. Oxidation of this produces ribulose-5-phosphate and another NADPH and releases CO_2. The remaining steps of the pathway are known as the non-oxidative phase and involve interconversion of sugar phosphates.

For example, ribulose-5-phosphate is converted to ribose-5-phosphate (R5P) by the enzyme ribulose-5-phosphate isomerase. Alternatively, ribulose-5-phosphate can be converted to xylulose-5-phosphate (Xu5P). R5P and Xu5P (10 carbons total) can be combined and rearranged by transketolase to produce intermediates with 3 and 7 carbons (glyceraldehyde-3-phosphate and sedoheptulose-7-phosphate, respectively). These last two molecules can, in turn be rearranged by transaldolase into 6 and 4 carbon

sugars (fructose-6-phosphate and erythrose-4-phosphate, respectively). Further, the erythrose-4-phosphate can swap parts with Xu5P to create glyceraldehyde-3-phosphate and fructose-6-phosphate.

It is important to recognize that the PPP pathway is not a “top-down" pathway, with all the intermediates derived from a starting G6P. All of the reactions are reversible, so that, for example, fructose-6-phosphate and glyceraldehyde-3-phosphate from glycolysis can reverse the last reaction of the previous paragraph to provide a means of synthesizing ribose-5-phosphate non-oxidatively. The pathway also provides a mechanism to cells for metabolizing sugars, such as Xu5P and ribulose-5-phosphate. In the bottom line of the pathway, the direction the pathway goes and the intermediates it produces are determined by the needs of, and intermediates available to, the cell.

As noted above, the pathway connects in three places with glycolysis. In non- plant cells, the PPP pathway occurs in the cytoplasm (along with glycolysis), so considerable “intermingling" of intermediates can and does occur. Erythrose-4-phosphate is an important precursor of aromatic amino acids and ribose-5-phosphate is an essential precursor for making nucleotides.

7.3: Calvin Cycle

The Calvin Cycle occurs exclusively in photosynthetic organisms and is the part of photosynthesis referred to as the “Dark Cycle." It is in this part of the process that carbon dioxide is taken from the atmosphere and ultimately built into glucose (or other sugars). Though reduction of carbon dioxide to glucose ultimately requires electrons from twelve molecules of NADPH (and 18 ATPs), it is a bit confusing because one reduction occurs 12 times (1,3 BPG to G3P) to achieve the reduction necessary to make one glucose.

One of the reasons students find the pathway a bit confusing is because the carbon dioxides are absorbed one at a time into six different molecules of ribulose-1,5-bisphosphate (Ru1,5BP). At no point are the six carbons ever together in the same molecule to make a single glucose. Instead, six molecules of Ru1,5BP (30 carbons) gain six more carbons via carbon dioxide and then split into 12 molecules of 3-phosphoglycerate (36 carbons). The gain of six carbons allows two three carbon molecules to be produced in excess for each turn of the cycle. These two molecules molecules are then converted into glucose using the enzymes of gluconeogenesis.

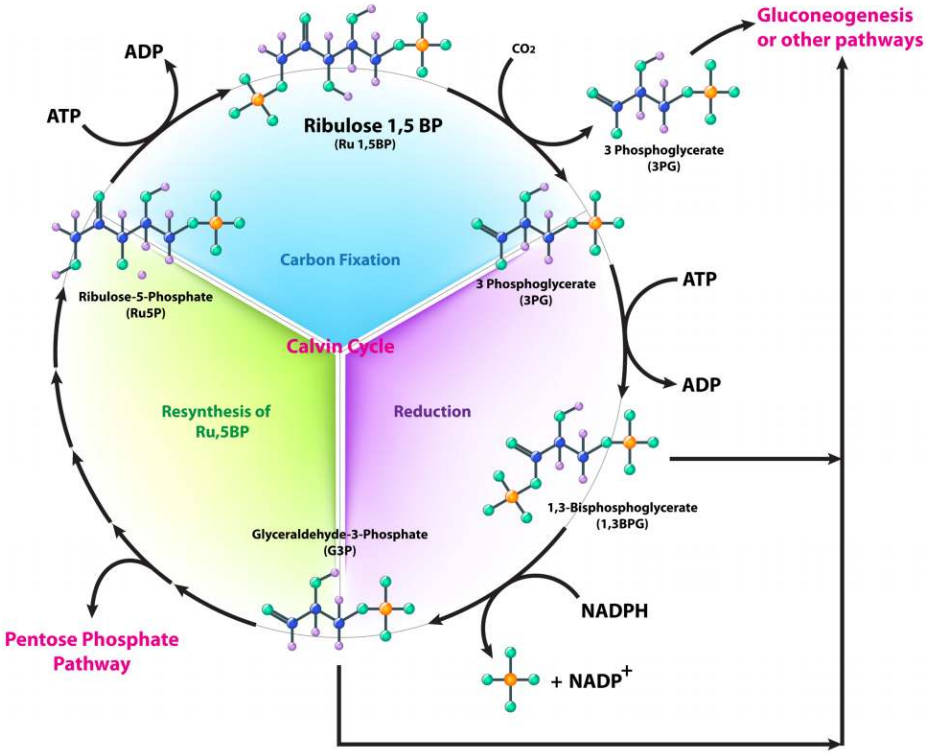

Figure 7.3.1: *The Calvin Cycle*

Like the citric acid cycle, the Calvin Cycle doesn’t really have a starting or ending point, but can we think of the first reaction as the fixation of carbon dioxide to Ru1,5BP. This reaction is catalyzed by the enzyme known as ribulose-1,5bisphosphate carboxylase (RUBISCO). The resulting six carbon intermediate is unstable and each Ru1,5BP is rapidly converted to 3-phosphoglycerate. As noted, if one starts with 6 molecules of Ru1,5BP and makes 12 molecules of 3-PG, the extra 6 carbons that are a part of the cycle can be shunted off as two three-carbon molecules of glyceraldehyde-3-phosphate (GA3P) to gluconeogenesis, leaving behind 10 molecules of GA3P to be reconverted into 6 molecules of Ru1,5BP. That part of the pathway requires multiple steps, but only utilizes two enzymes unique to plants - sedoheptulose-1,7bisphosphatase and phosphoribulokinase. RUBISCO is the third enzyme of the pathway that is unique to plants. All of the other enzymes of the pathway are common to plants and animals and include some found in the pentose phosphate pathway and gluconeogenesis.

7.4: C4 Plants

The Calvin Cycle is the means by which plants assimilate carbon dioxide from the atmosphere, ultimately into glucose. Plants use two general strategies for doing so. The first is employed by plants called C3 plants (most plants) and it simply involves the pathway described above. Another class of plants, called C4 plants employ a novel strategy for concentrating the CO_2 prior to assimilation. C4 plants are generally found in hot, dry environments where conditions favor the wasteful photorespiration reactions of RUBISCO, as well as loss of water. In these plants, carbon dioxide is captured in special mesophyll cells first by phosphoenolpyruvate (PEP) to make oxaloacetate. The oxaloacetate is converted to malate and transported into bundle sheath cells where the carbon dioxide is released and it is captured by ribulose-1,5-bisphosphate, as in C3 plants and the Calvin Cycle proceeds from there. The advantage of this scheme is that it allows concentration of carbon dioxide while minimizing loss of water and photorespiration.

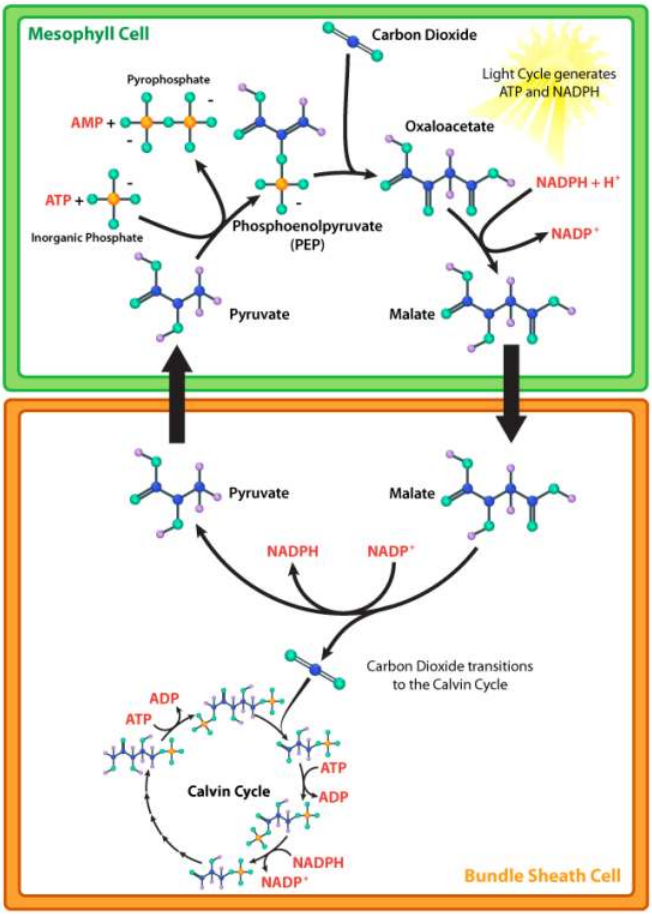

Figure 7.4.1: *C4 Plant Cycle*

7.5: Urea Cycle

Yet another cyclic pathway important in cells is the urea cycle (Figure 7.5.1). With reactions spanning the cytoplasm and the mitochondria, the urea cycle occurs mostly in the liver and kidney. The cycle plays an important role in nitrogen balance in cells and is found in organisms that produce urea as a way to excrete excess amines.

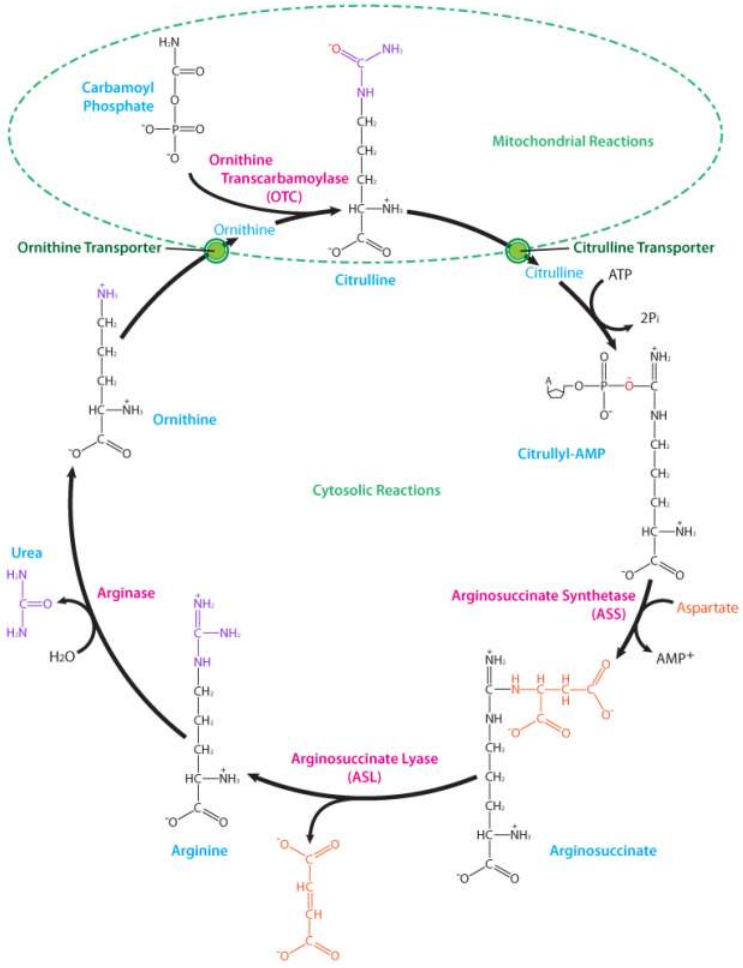

Figure 7.5.1: *The Urea Cycle*

The cycle scavenges free ammonia (as ammonium ion) which is toxic if it accumulates. The capture reaction also requires ATP, and bicarbonate, and the product is carbamoyl phosphate. This molecule is combined with the non-protein amino acid known as ornithine to make another non-protein amino acid known as citrulline. Addition of aspartate to citrulline creates argninosuccinate, which splits off a fumarate, creating arginine (a source of arginine). If arginine is not needed, it can be hydrolyzed to yield urea (excreted) an
ornithine, thus completing the cycle.

The first two reactions described here occur in the mitochondrion and the remaining ones occur in the cytoplasm. Molecules of the urea cycle intersecting other pathways include fumarate (citric acid cycle), aspartate (amino acid metabolism), arginine (amino acid metabolism), and ammonia (amino acid metabolism).

7.6: Nitrogen Fixation

The process of nitrogen fixation is important for life on earth, because atmospheric nitrogen is ultimately the source of amines in proteins and DNA. The enzyme playing an important role in this process is called nitrogenase and it is found in certain types of anaerobic bacteria called *diazotrophs*. Symbiotic relationships between some plants (legumes, for example) and the nitrogen-fixing bacteria provide the plants with access to reduced nitrogen. The overall reduction reaction catalyzed by nitrogenase is

$$\mathrm{N_2 + 6H^+ + 6e^- \rightarrow 2NH_3} \tag{7.6.1}$$

In these reactions, the hydrolysis of 16 ATP is required. The ammonia can be assimilated into glutamate and other molecules. Enzymes performing nitrogenase catalysis are very susceptible to oxygen and must be kept free of it. It is for this reason that most nitrogen-fixing bacteria are anaerobic. Movement of amines through biological systems occurs largely by the process of transmination, discussed below in amino acid metabolism.

7.7: Amino Acid Metabolism

The pathways for the synthesis and degradation of amino acids used in proteins are the most varied among the reactions synthesizing biological building blocks. We start with some terms. First, not all organisms can synthesize all the amino acids they need. Amino acids that an organism cannot synthesize (and therefore must have in their diets) are called *essential amino acids*. The remaining amino acids that the body can synthesize are called non-essential.

Glucogenic and Ketogenic Amino Acids

Glucogenic	Both Glucogenic and Ketogenic	Ketogenic
Aspartate	Isoleucine	Leucine
Asparagine	Phenylalanine	Lysine
Alanine	Tryptophan	
Glycine	Tyrosine	
Serine		
Threonine		
Cysteine		
Glutamate		
Glutamine		
Arginine		
Proline		
Histidine		
Valine		
Methionine		

Figure 7.7.1: *Glucogenic and Ketogenic Amino Acids*

Amino acids are also divided according to the pathways involved in their degradation; there are three general categories. Ones that yield intermediates in the glycolysis pathway are called *glucogenic* and those that yield intermediates of acetyl-CoA or acetoacetate are called *ketogenic*. Those that involve both are called glucogenic and ketogenic.

An important general consideration in amino acid metabolism is that of transamination. In this process, an exchange of amine and oxygen between an amino acid and an alpha-ketoacid occurs (see below)

$$\text{Alpha-ketoacid} + \text{amino acid} \leftrightarrow \text{amino acid} + \text{alpha-ketoacid} \tag{7.7.1}$$

An example reaction follows

$$\text{Pyruvate} + \text{Aspartic acid} \leftrightarrow \text{Alanine} + \text{Oxaloacetate} \tag{7.7.2}$$

This reaction is catalyzed by an enzyme known as a transaminase. Amino acids, such as glutamate, can also gain nitrogen directly from ammonium ion, as shown below

$$\text{Alpha-ketoglutarate} + NH_4^+ \leftrightarrow \text{Glutamate} \tag{7.7.3}$$

This reaction can occur, for example, in nitrifying bacteria, and in places where ammonia waste is produced. Many amino acids can be synthesized from citric acid cycle intermediates. For example, synthesis of the non-essential amino acids occurs as follows: aspartic acid can be made by transamination of oxaloacetate. Glutamate comes from transamination of alpha-ketoglutarate. Pyruvate, as noted, is a precursor of alanine (via transamination). Amino acids that can be made from glutamate include glutamine (by addition of an additional ammonium ion), proline, and arginine, Asparagine is made from aspartate by addition of ammonium ion also. Serine is formed from 3-phosphoglycerate and is itself the precursor of both glycine and cysteine. Cysteine and serine are also made from methionine. Tyrosine is made by hydroxylation of phenylalanine.

7.8: Amino Acid Catabolism

Breakdown of glutamine by glutaminase is a source of ammonium ion in the cell. The other product is glutamate. Glutamate, of course, can be converted by a transamination reaction to alpha-ketoglutarate, which can be oxidized in the citric acid cycle.

- Asparagine can similarly be broken to ammonium and aspartate by asparaginase and aspartate can be converted by transamination to oxaloacetate for oxidation in the citric acid cycle.
- Alanine is converted to pyruvate in a transamination reaction, making it glucogenic.
- Arginine is hydrolzyed in the urea cycle to yield urea and ornithine.
- Proline is catabolized to glutamate in a reversal of its synthesis pathway.
- Serine donates a carbon to form a folate and the other product of the reaction is glycine, which is itself oxidized to carbon dioxide and ammonia. Glycine can also be converted back to serine, which can also be converted back to 3-phosphoglycerate or pyruvate.
- Threonine can be broken down in three pathways, though only two are relevant for humans. One pathway leads to acetyl-CoA and glycine. The other leads to pyruvate.
- Cysteine can be broken down in several ways. The simplest occurs in the liver, where a desulfurase can act on it to yield hydrogen sulfide and pyruvate.
- Methionine can be converted to cysteine for further metabolism. It can be converted to succinyl-CoA for oxidation in the citric acid cycle. It can also be converted to S-Adenosyl-Methionine (SAM), a carbon donor.
- Isoleucine and valine can also be converted to succinyl-CoA after conversion first to propionyl-CoA. Since conversion of propionyl-CoA to succinyl-CoA requires vitamin $\mathbf{B_{12}}$, catabolism of these amino acids also requires the vitamin.
- Phenylalanine is converted during catabolism to tyrosine, which is degraded ultimately to fumarate and acetoacetate. Thus, both of these amino acids are glucogenic and ketogenic. Tyrosine can also be converted to dopamine, norepinephrine, and epinephrine.
- Leucine and lysine can be catabolized to acetoacetate and acetyl-CoA. Lysine is also an important precursor of carnitine.
- Histidine can be catabolized by bacteria in intestines to histamine, which causes construction or dilation of various blood vessels when in excess.
- Tryptophan's catabolism is complex, but can proceed through alanine, acetoacetate and acetyl-CoA

In summary, the following are metabolized to pyruvate – alanine, cysteine, glycine, serine, and threonine

- Oxaloacetate is produced from aspartate and asparagine
- Succinyl-CoA is produced from isoleucine, valine, and methionine
- Alpha-ketoglutarate is produced from arginine, glutamate, glutamine, histidine and proline.
- Phenylalanine and tyrosine are broken down to fumarate and acetoacetate
- Leucine and lysine yield acetoacetate and acetyl-CoA.
- Tryptophan leads to alanine, acetoacetate and acetyl-CoA.

Last, amino acids, besides being incorporated into proteins, serve as precursors of important compounds, including serotonin (from tryptophan), porphyrin heme (from glycine), nitric oxide (from arginine), and nucleotides (from aspartate, glycine, and glutamine).

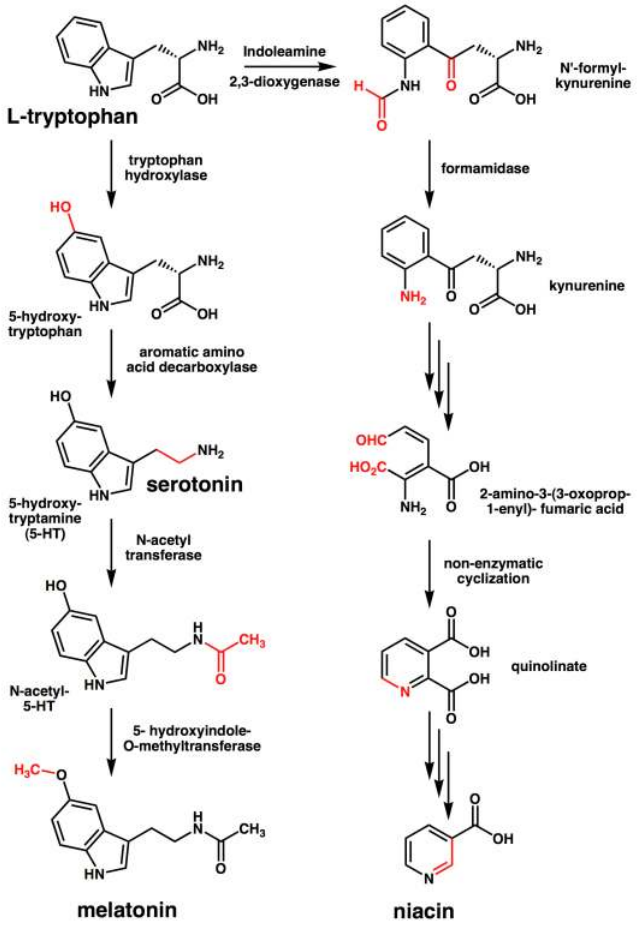

Figure 7.8.1: *Conversion of L-Tryptophan into Serotonin, Melatonin, and Niacin*

7.9: Nucleotide Metabolism

Arguably, the most interesting metabolic pathways from the perspective of regulation are those leading to the synthesis of nucleotides. We shall consider ribonucleotide synthesis from from scratch (de novo synthesis). Deoxyribonucleotide synthesis from ribonucleotides will be considered separately.

Synthesis of ribonucleotides by the de novo method occurs in two pathways – one for purines and one for pyrimidines. What is notable about both of these pathways is that nucleotides are built from very simple building blocks.

7.10: Pyrimidine de novo Biosynthesis

Starting materials for pyrimidine biosynthesis (Figure 7.10.1) include bicarbonate, amine from glutamine, and phosphate from ATP to make carbamoyl-phosphate (similar to the reaction of the urea cycle). Joining of carbamoyl phosphate to aspartic acid (forming carbamoyl aspartate) is catalyzed by the most important regulatory enzyme of the cycle, aspartate transcarbamoylase (also called aspartate carbamoyltransferase or ATCase).

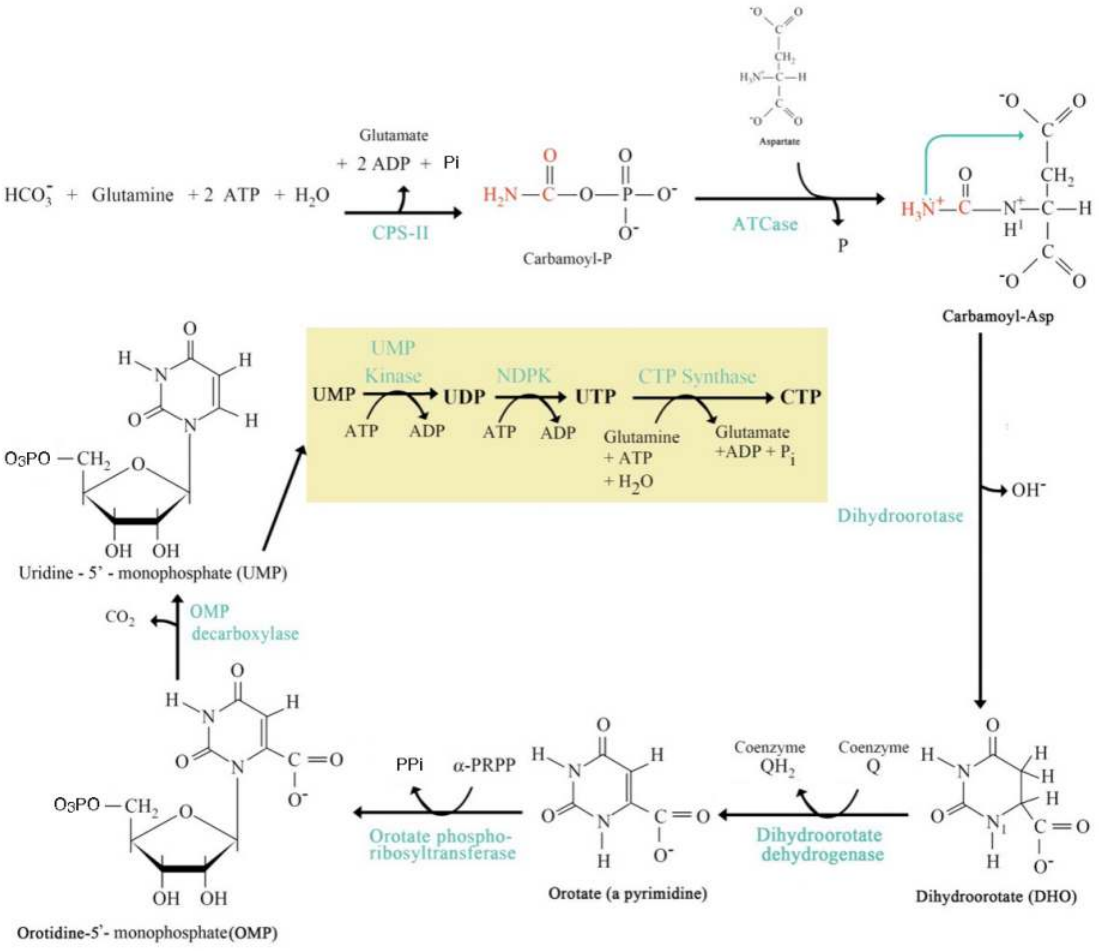

Figure 7.10.1: *De Novo Synthesis of Pyrimidine Nucleotides*

ATCase is regulated by three compounds. One of these (aspartate) is a substrate and it activates the enzyme by binding to the catalytic site and favoring the enzyme's R state. The other two regulators bind to regulatory subunits of the enzyme and either inhibit (CTP) or activate (ATP) the enzyme.

The reaction product, carbamoyl aspartate, is transformed in two reactions to orotic acid, which is, in turn combined with phosphoribosylpyrophosphate PRPP). The product of that reaction, orotidyl monophosphate (OMP) is decarboxylated to form the first pyrimidine nucleotide, UMP. Conversion of UMP to UDP is catalyzed by nucleoside monophosphate kinases (NMPs) and UDP is converted to UTP by nucleoside diphosphokinase (NDPK).

UDP (like all of the nucleoside diphosphates) is a branch point to deoxyribonucleoside diphosphates, catalyzed by ribonucleotide reductases, which are discussed later. UTP is converted to CTP by CTP synthase. This enzyme, which uses an amino group from glutamine for the reaction, serves to balance the relative amounts of CTP and UTP, thanks to inhibition by excess CTP.

7.11: Purine de novo Biosynthesis

Synthesis of purine nucleotides differs fundamentally from that of pyrimidine nucleotides in that the bases are built on the ribose ring. The starting material is ribose 5-phosphate, which is phosphorylated by PRPP synthetase to PRPP using two phosphates from ATP. PRPP amidotransferase catalyzes the transfer of an amine group to PRPP, replacing the pyrophosphate on carbon 1. Thus begins the synthesis of the purine ring.

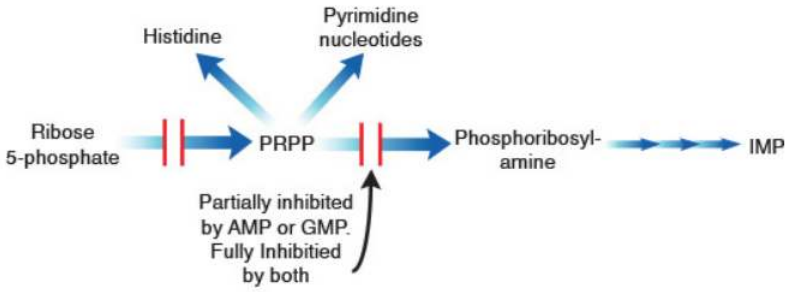

Figure 7.11.1: *Beginning of Purine Metabolism*

PRPP amidotransferase is regulated partly by GMP and partly by AMP. The presence of either of these can reduce the enzyme's activity. Only when both are present is the enzyme fully inactivated. Subsequent reactions include adding glycine, adding carbon (from N 10-formyltetrahydrofolate), adding amine (from glutamine), closing of the first ring, addition of carboxyl (from CO_2), addition of aspartate, loss of fumarate (a net gain of an amine), addition of another carbon (from N_10-formyltetrahydrofolate), and closing of the second ring to form inosine monophosphate (IMP).

IMP is a branch point for the synthesis of the adenine and guanine nucleotides. The pathway leading from IMP to AMP involves addition of amine from asparate and requires energy from GTP. The pathway from IMP to GMP involves an oxidation and addition of an amine from glutamine. It also requires energy from ATP. The pathway leading to GMP is inhibited by its end product and the pathway to AMP is inhibited by its end product.

Thus, balance of the purine nucleotides is achieved from the IMP branch point forward. It is at this point that the significance of the unusual regulation of PRPP amidotransferase becomes apparent. If there is an imbalance of AMP or GMP, the enzyme is slowed, but not stopped, thus allowing the reactions leading to IMP to proceed, albeit slowly. At IMP, the nucleotide in excess feedback inhibits its own synthesis, thus allowing the partner purine nucleotide to be made and balance to be achieved. When both nucleotides are in abundance, then PRPP amidotransferase is fully inhibited and the production of purines is stopped, thus preventing them from over-accumulating.

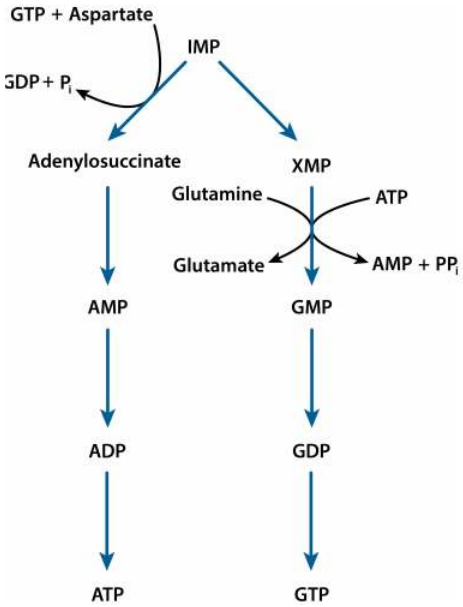

Figure 7.11.2: *Synthesis of ATP and GTP*

7.12: Deoxyribonucleotide de novo Biosynthesis

Synthesis of deoxyribonucleotides de novo requires an interesting enzyme called ribonucleotide reductase (RNR). RNR catalyzes the formation of deoxyribonucleotides from ribonucleotides. The most common form of RNR is the Type I enzyme, whose substrates are ribonucleoside diphosphates (ADP, GDP, CDP, or UDP) and the products are deoxyribonucleoside diphosphates (dADP, dGDP, dCDP, or dUDP). Thymidine nucleotides are synthesized from dUDP. RNR has two pairs of two identical subunits - R1 (large subunit) and R2 (small subunit). R1 has two allosteric binding sites and an active site. R2 forms a tyrosine radical necessary for the reaction mechanism of the enzyme.

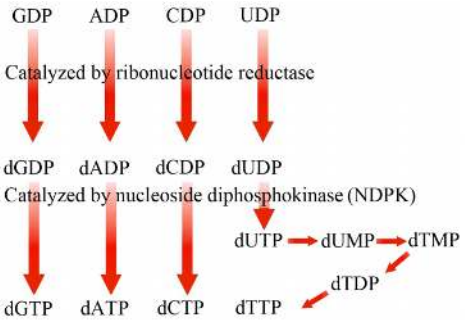

Figure 7.12.1: *dNTP de novo Synthesis*

Because a single enzyme, RNR, is responsible for the synthesis of all four deoxyribonucleotides, it is necessary to have mechanisms to ensure that the enzyme produces the correct amounts of each dNDP. This means that the enzyme must be responsive to the levels of the each deoxynucleotide, selectively making more of those that are in short supply, and preventing synthesis of those that are abundant. These demands are met by having two separate control mechanisms, one that determines which substrate will be acted on, and another that controls the enzyme's catalytic activity.

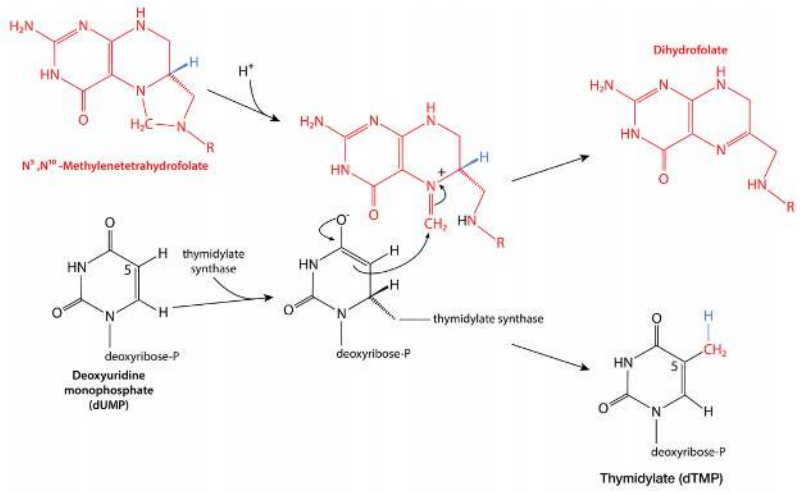

Figure 7.12.2: *Thymidylate (dTMP) Synthesis*

Ribonucleotide reductase is allosterically regulated via two binding sites - a specificity binding site (binds dNTPs and controls which substrates the enzyme binds and thus, which deoxyribonucleotides are made) and an activity binding site (controls whether or not enzyme is active - ATP activates, dATP inactivates).

When a deoxypyrimidine triphosphate, dTTP is abundant, it binds to the specificity site and inhibits binding and reduction of pyrimidine diphosphates (CDP and UDP) but stimulates binding and reduction of GDP by the enzyme. Conversely, binding of the deoxypurine triphosphate, ATP stimulates reduction of pyrimidine diphosphates, CDP and UDP.

Students sometimes confuse the active site of RNR with the activity site. The active site is where the reaction is catalyzed, and could also be called the catalytic site, whereas the activity site is the allosteric binding site for ATP or dATP that controls whether the enzyme is active.

Synthesis of dTTP by the de novo pathway takes a convoluted pathway from dUDP to dUTP to dUMP to dTMP, then dTDP, and finally dTTP. Conversion of dUMP to dTMP, requires a tetrahydrofolate derivative and the enyzme thymidylate synthase. In the process, dihydrofolate is produced and must be converted back to tetrahyrdolate in order to keep nucleotide synthesis occurring. The enzyme involved in the conversion of dihydrofolate to tetrahydrofolate, dihydrofolate reductase (DHFR), is a target of anticancer drugs like methotrexate or aminopterin, which inhibit the enzyme.

CHAPTER OVERVIEW

8: Signaling

Cells must receive and respond to signals from their surroundings. Cellular signals and the pathways through which they are passed on and amplified to produce the desired effects on their targets are the focus of this section.

8.1: Cell Signaling

How do cells receive signals from their environment and how do they communicate among themselves? It is intuitively obvious that even bacterial cells must be able to sense features of their environment, such as the presence of nutrients or toxins, if they are to survive. In addition to being able to receive information from the environment, multicellular organisms must find ways by which their cells can communicate among themselves. Since different cells take on specialized functions in a multicellular organism, they must be able to coordinate activities perfectly like the musicians in an orchestra performing a complicated piece of music. Cells grow, divide, or differentiate in response to specific signals. They may change shape or migrate to another location. At the physiological level, cells in a multicellular organism, must respond to everything from a meal just eaten to injury, threat or the availability of a mate. They must know when to repair damage to DNA, when to undergo apoptosis (programmed cell death) and even when to regenerate a lost limb. A variety of mechanisms have arisen to ensure that cell-cell communication is not only possible, but astonishingly swift, accurate and reliable.

How are signals sent between cells?
Like pretty much everything that happens in cells, signaling is dependent on molecular recognition. The basic principle of cell-cell signaling is simple. A particular kind of molecule, sent by a signaling cell, is recognized and bound by a receptor protein in (or on the surface of) the target cell. The signal molecules are chemically varied- they may be proteins, short peptides, lipids, nucleotides or catecholamines, to name a few. The chemical properties of the signal determine whether its receptors are on the cell surface or intracellular. If the signal is small and hydrophobic it can cross the cell membrane and bind a receptor inside the cell. If, on the other hand, the signal is charged, or very large, it would not be able to diffuse through the plasma membrane. Such signals need receptors on the cell surface, typically transmembrane proteins that have an extracellular portion that binds the signal and an intracellular part that passes on the message within the cell.

Receptors are specific for each type of signal, so each cell has many different kinds of receptors that can recognize and bind the many signals it receives. Because different cells have different sets of receptors, they respond to different signals or combinations of signals. The binding of a signal molecule to a receptor sets off a chain of events in the target cell. These events could cause change in various ways, including, but not limited to, alterations in metabolic pathways or gene expression in the target cell.

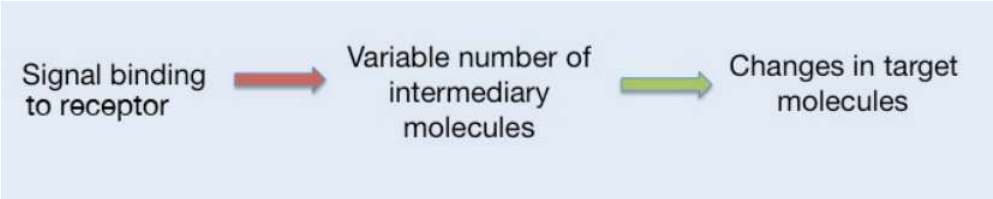

Figure 8.1.1: *Cellular Signaling*

How the binding of a signal to a receptor brings about change in cells is the topic of this section. Although the specific molecular components of the various signal transduction pathways differ, they all have some features in common:

- The binding of a signal to its receptor is usually, though not always, followed by the generation of a new signal(s) within the cell. The process by which the original signal is converted to a different form and passed on within the cell to bring about change is called signal transduction.
- Most signaling pathways have multiple signal transduction steps by which the signal is relayed through a series of molecular messengers that can amplify and distribute the message to various parts of the cell.
- The last of these messengers usually interacts with a target protein(s) and changes its activity, often by phosphorylation.

When a signal sets a particular pathway in motion, it is acting like an ON switch. This means that once the desired result has been obtained, the cell must have a mechanism that acts as an OFF switch.

Understanding this underlying similarity is helpful, because learning the details of the different pathways becomes merely a matter of identifying which molecular component performs a particular function in each individual case. We will consider several different signal transduction pathways, each mediated by a different kind of receptor. The first two examples we will examine are those with the fewest steps between the binding of the signal by a receptor and a cellular response.

8.2: Ligand-gated Ion Channel Receptors

The simplest and fastest of signal pathways is seen in the case of signals whose receptors are gated ion channels. Gated ion channels are made up of multiple transmembrane proteins that create a pore, or channel, in the cell membrane. Depending upon its type, each ion channel is specific to the passage of a particular ionic species. The term "gated" refers to the fact that the ion channel is controlled by a "gate" which must be opened to allow the ions through. The gates are opened by the binding of an incoming signal (ligand) to the receptor, allowing the almost instantaneous passage of millions of ions from one side of the membrane to the other. Changes in the interior environment of the cell are thus brought about in microseconds and in a single step.

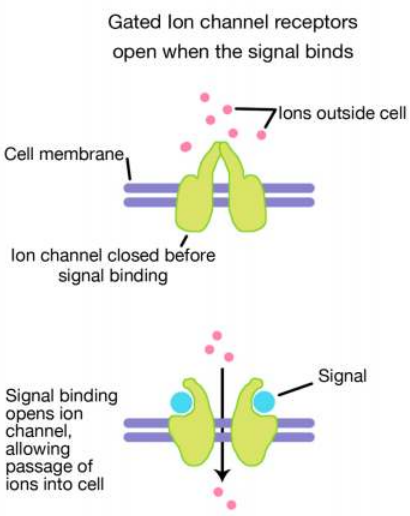

Figure 8.2.1: *Signaling Through Gated Ion Channels*

This type of swift response is seen, for example, in neuromuscular junctions, where muscle cells respond to a message from the neighboring nerve cell. The nerve cell releases a neurotransmitter signal into the synaptic cleft, which is the space between the nerve cell and the muscle cell it is "talking to". Examples of neurotransmitter signal molecules are acetylcholine and serotonin, shown in Figure 8.2.2.

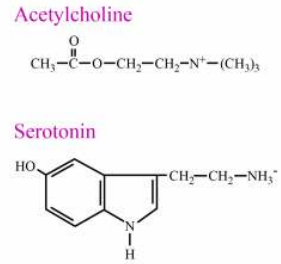

Figure 8.2.2: *Neurotransmitter*

When acetylcholine molecules are released into the synaptic cleft (the space between the pre- and post-synaptic cells) they diffuse rapidly till they reach their receptors on the membrane of the muscle cell. The binding of the acetylcholine to its receptor, an ion channel on the membrane of the muscle cell, causes the gate in the ion channel to open. The resulting ion flow through the channel can immediately change the membrane potential. This, in turn, can trigger other changes in the cell. The speed with which changes are brought about in neurotransmitter signaling is evident when you think about how quickly you remove your hand from a hot surface. Sensory neurons carry information to the brain from your hand on the hot surface and motor neurons signal to your muscles to move the hand, in less time than it took you to read this sentence!

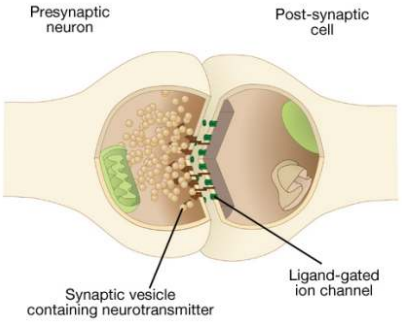

Figure 8.2.3: *Signaling across nerve cells*

8.3: Nuclear Hormone Receptors

Another type of relatively simple, though much slower, signaling is seen in pathways in which the signals are steroid hormones, like estrogen or testosterone, pictured below. Steroid hormones, as you are aware, are related to cholesterol, and as hydrophobic molecules, they are able to cross the cell membrane by themselves. This is unusual, as most signals coming to cells are incapable of crossing the plasma membrane, and thus, must have cell surface receptors.

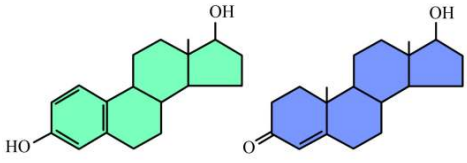

Figure 8.3.1: *Estradiol and Testosterone*

By contrast, steroid hormones have receptors inside the cell (intracellular receptors). Steroid hormone receptors are proteins that belong in a family known as the *nuclear receptors*. Nuclear hormone receptors are proteins with a double life: they are actually dormant transcription regulators. In the absence of signal, these receptors are in the cytoplasm, complexed with other proteins (HSP in Figure 8.3.2) and inactive. When a steroid hormone enters the cell, the nuclear hormone receptor binds the hormone and dissociates from the HSP. The receptors, then, with the hormone bound, translocate into the nucleus.

In the nucleus, Nuclear hormone receptors regulate the transcription of target genes by binding to their regulatory sequences (labeled HRE for hormone- response elements). The binding of the hormone-receptor complex to the regulatory elements of hormone-responsive genes modulates their expression. Because these responses involve gene expression, they are relatively slow. Most other signaling pathways, besides the two we have just discussed, involve multiple steps in which the original signal is passed on and amplified through a number of intermediate steps, before the cell responds to the signal.

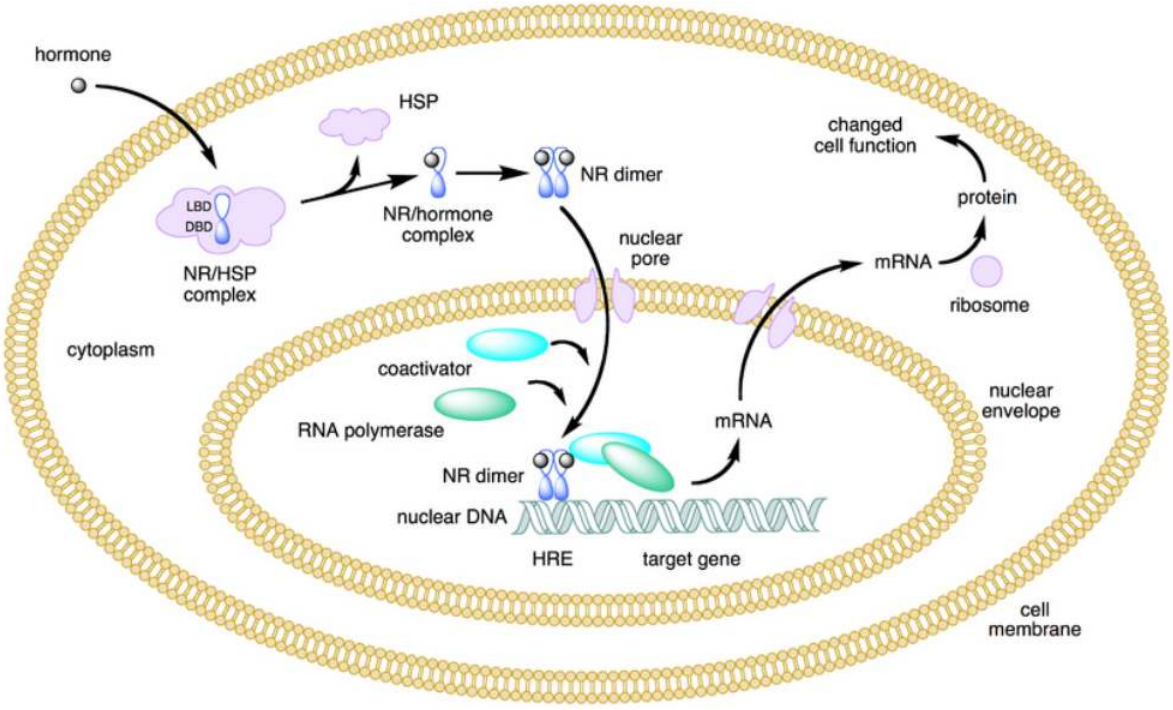

Figure 8.3.2: *Steroid hormones act by modulating expression of hormone-responsive genes*

We will now consider two signaling pathways, each mediated by a major class of cell surface receptor- the G-protein coupled receptors (GPCRs) and the receptor tyrosine kinases (RTKs). While the specific details of the signaling pathways that follow the binding of signals to each of these receptor types are different, it is easier to learn them when you can see what the pathways have

in common, namely, interaction of the signal with a receptor, followed by relaying the signal through a variable number of intermediate molecules, with the last of these molecules interacting with target protein(s) to modify their activity in the cell.

8.4: G-protein Coupled Receptors (GPCRs)

G-protein coupled receptors are involved in responses of cells to many different kinds of signals, from epinephrine, to odors, to light. In fact, a variety of physiological phenomena including vision, taste, smell and the fight-or-flight response are mediated by GPCRs.

What are G-protein coupled receptors?

G-protein coupled receptors are cell surface receptors that pass on the signals that they receive with the help of guanine nucleotide binding proteins (a.k.a. G-proteins). Before thinking any further about the signaling pathways downstream of GPCRs, it is necessary to know a few important facts about these receptors and the G-proteins that assist them. Though there are hundreds of different G-protein coupled receptors, they all have the same basic structure: they all consist of a single polypeptide chain that threads back and forth seven times through the lipid bilayer of the plasma membrane. For this reason, they are sometimes called seven- pass transmembrane (7TM) receptors.

One end of the polypeptide forms the extracellular domain that binds the signal while the other end is in the cytosol of the cell.

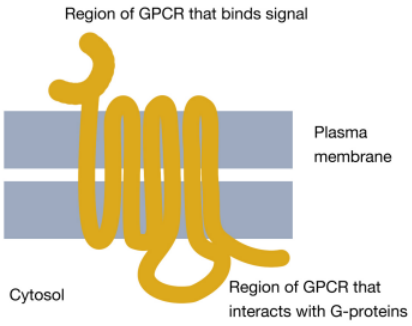

Figure 8.4.1: *G-protein coupled receptor*

When a ligand (signal) binds the extracellular domain of a GPCR, the receptor undergoes a conformational change that allows it to interact with a G-protein that will then pass the signal on to other intermediates in the signaling pathway.

What is a G-protein?

As noted above, a G-protein is a guanine nucleotide-binding protein that can interact with a G-protein linked receptor. G-proteins are associated with the cytosolic side of the plasma membrane, where they are ideally situated to interact with the cytosolic tail of the GPCR, when a signal binds to the GPCR.

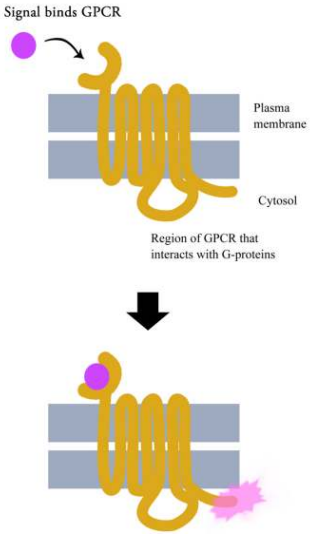

Figure 8.4.2: *G-protein coupled receptor signaling*

There are many different G-proteins, all of which share a characteristic structure- they are composed of three subunits called alpha, beta and gamma (aß.). Because of this, they are sometimes called heterotrimeric G proteins (hetero=different, trimeric= having three parts). The a subunit of such proteins can bind GDP or GTP and is capable of hydrolyzing a GTP molecule bound to it into GDP. In the unstimulated state of the cell, that is, in the absence of a signal bound to the GPCR, the G-proteins are found in the trimeric form (aß. bound together) and the a subunit has a GDP molecule bound to it.

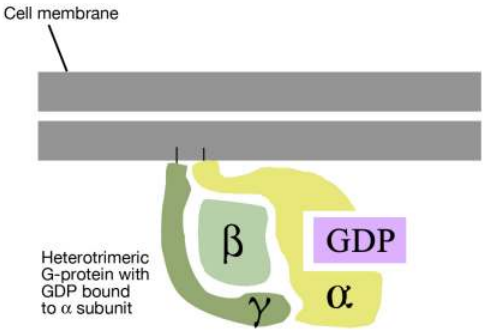

Figure 8.4.3: *G-protein with GDP Bound*

With this background on the structure and general properties of the GPCRs and the G-proteins, we can now look at what happens when a signal arrives at the cell surface and binds to a GPCR. The binding of a signal molecule by the extracellular part of the G-protein linked receptor causes the cytosolic tail of the receptor to interact with, and alter the conformation of, a G-protein. This has two consequences:

- First, the alpha subunit of the G- protein loses its GDP and binds a GTP instead.
- Second, the G-protein breaks up into the GTP-bound a part and the ß. part.

These two parts can diffuse freely along the cytosolic face of the plasma membrane and act upon their targets.

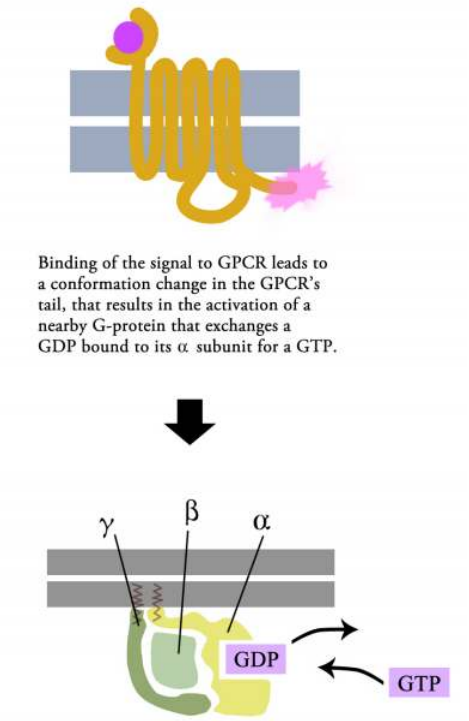

Figure 8.4.4: *G-protein Activation*

What happens when G-proteins interact with their target proteins? That depends on what the target is. G-proteins interact with different kinds of target proteins, of which we will examine two major categories:

Ion Channels

We have earlier seen that some gated ion channels can be opened or closed by the direct binding of neurotransmitters to a receptor that is an ion-channel protein. In other cases, ion channels are regulated by the binding of G-proteins. That is, instead of the signal directly binding to the ion channel, it binds to a GPCR, which activates a G-protein that then binds and opens the ion channel. The change in the distribution of ions across the plasma membrane causes a change in the membrane potential.

Specific Enzymes

The interaction of G-proteins with their target enzymes can regulate the activity of the enzyme, either increasing or decreasing its activity. Often the target enzyme will pass the signal on in another form to another part of the cell. As you might imagine, this kind of response takes a little longer than the kind where an ion channel is opened instantaneously. Two well-studied examples of enzymes whose activity is regulated by a G-protein are adenylate cyclase and phospholipase C. When adenylate cyclase is activated, the molecule cAMP is produced in large amounts.

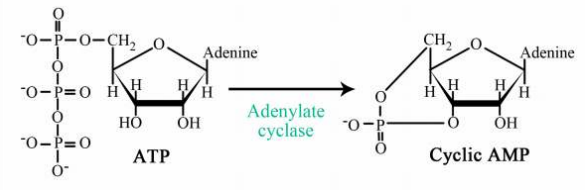

Figure 8.4.5: *Synthesis of cAMP*

When phospholipase C is activated, the molecules inositol trisphosphate (IP3) and diacylglycerol (DAG) are made. cAMP, IP3 and DAG are second messengers, small, diffusible molecules that can "spread the message" brought by the original signal, to other parts of the cell.

In these cases, the binding of a signal to the GPCR activated a G- protein, which in turn, activated an enzyme that makes a second messenger that can amplify the message in the cell. We will first trace the effects of activating adenylate cyclase and the resulting increase in cAMP.

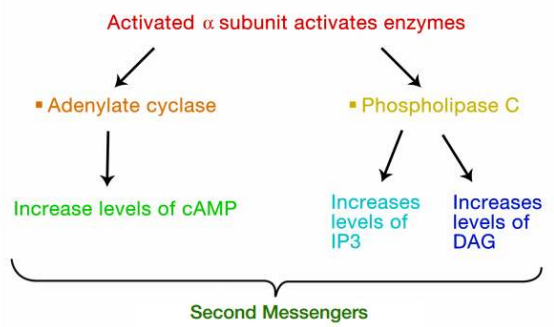

Figure 8.4.6: *Second Messengers*

What is the effect of elevated cAMP levels?

cAMP molecules bind to, and activate an enzyme, protein kinase A (PKA). PKA is composed of two catalytic and two regulatory subunits that are bound tightly together. Upon binding of cAMP the catalytic subunits are released from the regulatory subunits, allowing the enzyme to carry out its function, namely phosphorylating other proteins.

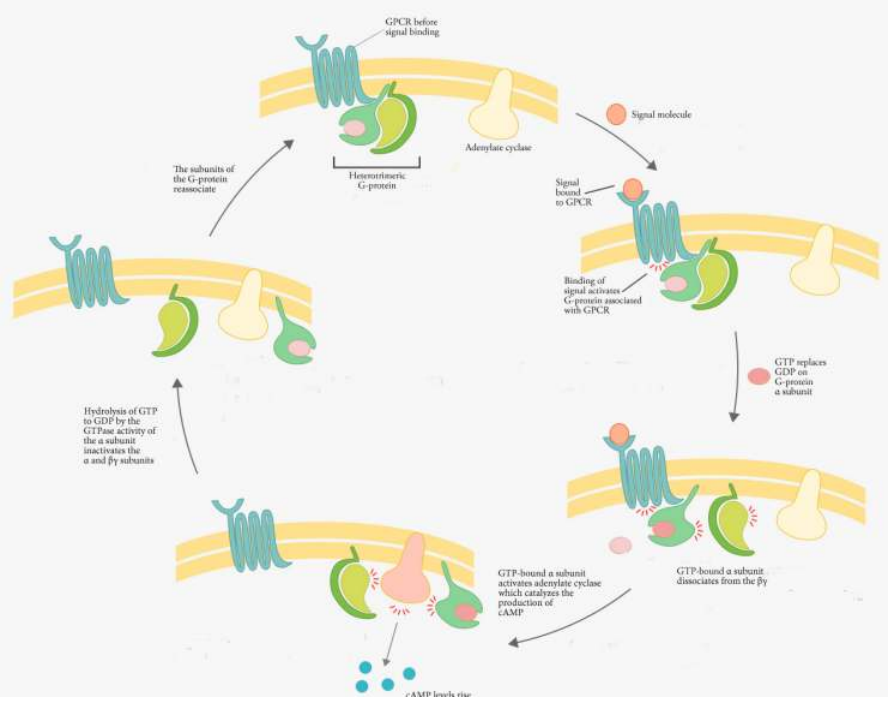

Figure 8.4.7: *G-protein Signaling Cycle*

Thus, cAMP can regulate the activity of PKA, which in turn, by phosphorylating other proteins can change their activity. The targets of PKA may be enzymes that are activated by phosphorylation, or they may be proteins that regulate transcription. The phosphorylation of a transcriptional activator, for example, may cause the activator to bind to a regulatory sequence on DNA and to increase the transcription of the gene it controls. The activation of previously inactive enzymes alters the state of the cell by changing the reactions that are occurring within the cell.

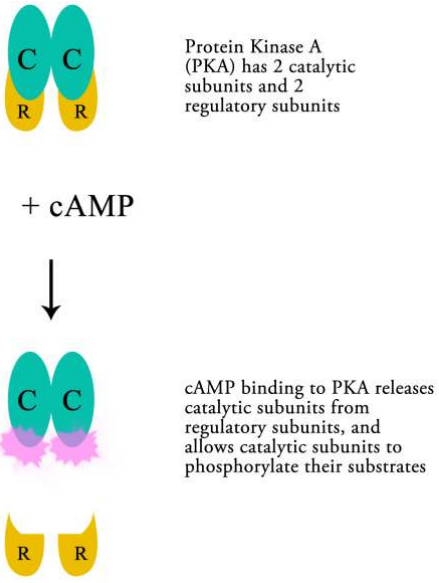

Figure 8.4.8: *Protein Kinase A Activation*

For example, the binding of epinephrine to its receptor on the cell surface, activates, through the action of G-proteins, and subsequent activation of PKA, the phosphorylation of glycogen phosphorylase. The resulting activation of glycogen phosphorylase leads to the breakdown of glycogen, releasing glucose (in the form of glucose-1-phosphate) for use by the cell. Changes in gene expression, likewise, lead to changes in the cell by altering the production of particular proteins in response to the signal.

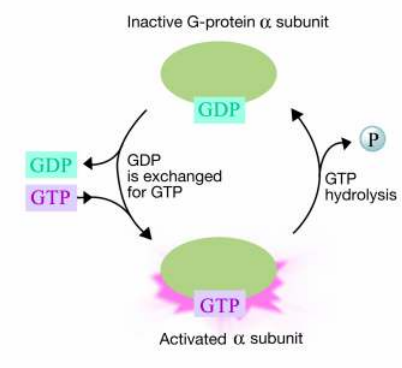

Figure 8.4.9: *G-protein nucleotide swapping*

Although the steps described above seem complicated, they follow the simple pattern outlined at the beginning of this section:

- Binding of signal to receptor
- Several steps where the signal is passed on through intermediate molecules (G-proteins, adenylate cyclase, cAMP, and finally, PKA)
- Phosphorylation of target proteins by the kinase, leading to changes in the cell.

Finally, if the signal binding to the receptor serves as a switch that sets these events in motion, there must be mechanisms to turn the pathway off. The first is at the level of the G-protein. Recall that the alpha subunit of the G-protein is in its free and activated state when it has GTP bound and that it associates with the beta- gamma subunits and has a GDP bound when it is inactive. We also know that the alpha subunit has an activity that enables it to hydrolyze GTP to GDP, as shown in the figure above left. This GTP-hydrolyzing activity makes it possible for the alpha subunit, once it has completed its task, to return to its GDP bound state, re-associate with the beta-gamma part and become inactive again.

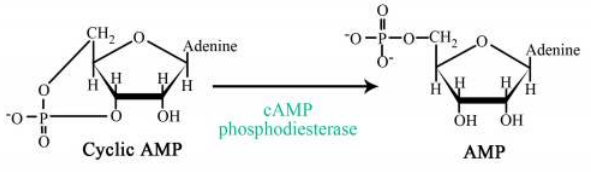

Figure 8.4.10: *cAMP Breakdown*

The second "off switch" is further down the signaling pathway, and controls the level of cAMP. We just noted that cAMP levels increase when adenylate cyclase is activated. When its job is done, cAMP is broken down by an enzyme called phosphodiesterase. When cAMP levels drop, PKA returns to its inactive state, putting a halt to the changes brought about by the activation of adenylate cyclase by an activated G-protein.

Let us now examine the events that follow the activation of Phospholipase C (PLC) by a G-protein. As we noted earlier, the activation of PLC results in the production of the second messengers IP3 and DAG. What do these molecules do?

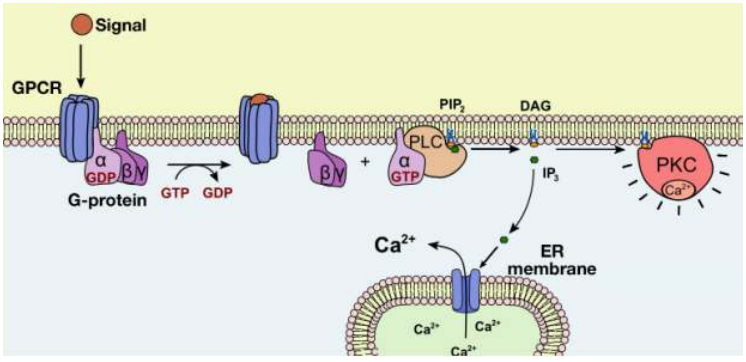

Figure 8.4.11: *Phospholipase C Signaling*

The IP3 and DAG produced by activated phospholipase C work together to activate a protein kinase. First, IP3 diffuses to the endoplasmic reticulum membrane where it binds to gated calcium ion channels. This causes calcium channels in the ER membrane to open and release large amounts of calcium into the cytoplasm from the ER lumen, as shown in the figure below.

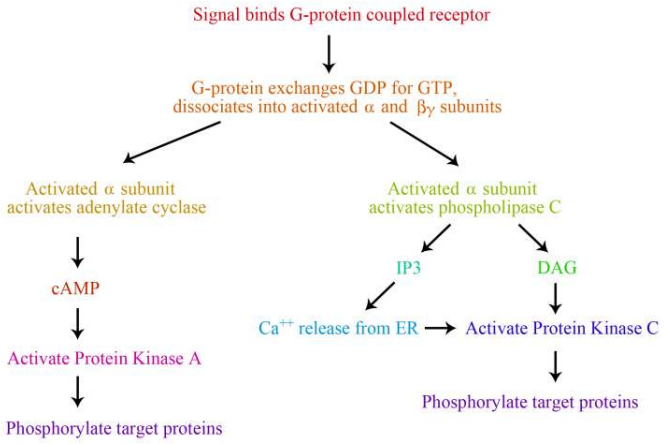

Figure 8.4.12: *Signaling Outcomes*

The increase in cytosolic calcium ion concentration has various effects, one of which is to activate a protein kinase called protein kinase C (C for calcium), together with the DAG made in the earlier step. Like PKA, Protein kinase C phosphorylates a variety of proteins in the cell, altering their activity and thus changing the state of the cell.

The pathways leading to PKC and PKA activation following the binding of a signal to a GPCR are summarized in Figure 8.4.12.

8.5: Receptor Tyrosine Kinases (RTKs)

Receptor tyrosine kinases mediate responses to a large number of signals, including peptide hormones like insulin and growth factors like epidermal growth factor. Like the GPCRs, receptor tyrosine kinases bind a signal, then pass the message on through a series of intracellular molecules, the last of which acts on target proteins to change the state of the cell.

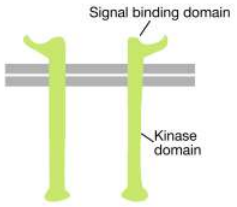

Figure 8.5.1: *Receptor Tyrosine Kinase before signal binding*

As the name suggests, a receptor tyrosine kinase is a cell surface receptor that also has a tyrosine kinase activity. The signal binding domain of the receptor tyrosine kinase is on the cell surface, while the tyrosine kinase enzymatic activity resides in the cytoplasmic part of the protein (see figure above). A transmembrane alpha helix connects these two regions of the receptor.

What happens when signal molecules bind to receptor tyrosine kinases?

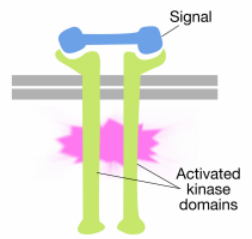

Figure 8.5.2: *Signal binding causes dimerization of receptor and activation of tyrosine kinase domains*

Binding of signal molecules to the extracellular domains of receptor tyrosine kinase molecules causes two receptor molecules to dimerize (come together and associate). This brings the cytoplasmic tails of the receptors close to each other and causes the tyrosine kinase activity of these tails to be turned on. The activated tails then phosphorylate each other on several tyrosine residues. This is called autophosphorylation.

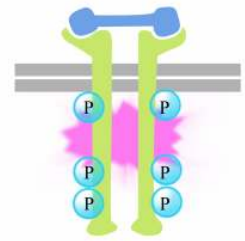

Figure 8.5.3: *Activated tyrosine kinase domains add phosphate onto each other*

The phosphorylation of tyrosines on the receptor tails triggers the assembly of an intracellular signaling complex on the tails. The newly phosphorylated tyrosines serve as binding sites for signaling proteins that then pass the message on to yet other proteins. An important protein that is subsequently activated by the signaling complexes on the receptor tyrosine kinases is called *Ras*.

The Ras protein is a monomeric guanine nucleotide binding protein that is associated with the cytosolic face of the plasma membrane (in fact, it is a lot like the alpha subunit of trimeric G-proteins). Just like the alpha subunit of a G- protein, Ras is active when GTP is bound to it and inactive when GDP is bound to it.Also, like the alpha subunit, Ras can hydrolyze the GTP to GDP.

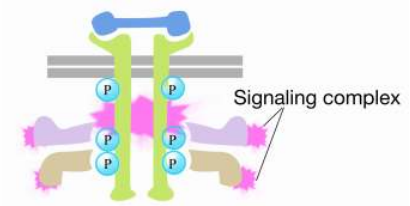

Figure 8.5.4: *Complex of signaling proteins assembles on phosphorylated RTK tails. This complex can activate Ras.*

When a signal arrives at the receptor tyrosine kinase, the receptor monomers come together and phosphorylate each others' tyrosines, triggering the assembly of a complex of proteins on the cytoplasmic tail of the receptor. One of the proteins in this complex interacts with Ras and stimulates the exchange of the GDP bound to the inactive Ras for a GTP. This activates the Ras.

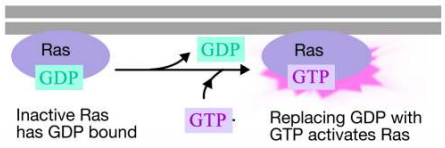

Figure 8.5.5: *Ras Activation*

Activated Ras triggers a phosphorylation cascade of three protein kinases, which relay and distribute the signal. These protein kinases are members of a group called the MAP kinases (Mitogen Activated Protein Kinases). The final kinase in this cascade phosphorylates various target proteins, including enzymes and transcriptional activators that regulate gene expression.

The phosphorylation of various enzymes can alter their activities, and set off new chemical reactions in the cell, while the phosphorylation of transcriptional activators can change which genes are expressed. The combined effect of changes in gene expression and protein activity alter the cell's physiological state.

Once again, in following the path of signal transduction mediated by RTKs, it is possible to discern the same basic pattern of events: a signal is bound by the extracellular domains of receptor tyrosine kinases, resulting in receptor dimerization and autophosphorylation of the cytosolic tails, thus conveying the message to the interior of the cell.

The message is passed on via a signalling complex to Ras which then stimulates a series of kinases. The terminal kinase in the cascade acts on target proteins and brings about in changes in protein activities and gene expression.

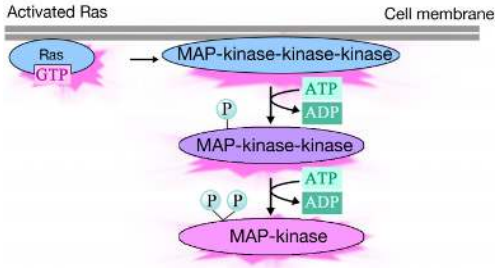

Figure 8.5.6: *Activated Ras Cascade*

The descriptions above provide a very simple sketch of some of the major classes of receptors and deal primarily with the mechanistic details of the steps by which signals received by various types of receptors bring about changes in cells. A major take-home lesson is the essential similarity of the different pathways.

Another point to keep in mind is that while we have looked at each individual pathway in isolation, a cell, at any given time receives multiple signals that set off a variety of different responses at once. The pathways described above show a considerable degree of "cross-talk" and the response to any given signal is affected by the other signals that the cell receives simultaneously. The multitude of different receptors, signals and the combinations thereof are the means by which cells are able to respond to an enormous variety of different circumstances.

CHAPTER OVERVIEW

9: Techniques

The environment of a cell is very complex, making it very diffcult, if not impossible, to study individual reactions, enzymes, or pathways within it. For this reason, biochemists prefer to isolate molecules (enzymes, DNAs, RNAs, and other molecules of interest) so they can be analyzed without interference from the millions of other processes occurring simultaneously in the cell. Many of the methods used in isolating molecules from cells involve some form of chromatography. To separate compounds from their cellular environments, one must first break open (lyse) the cells. In this section, we describe some of the methods biochemists use to do their work.

9.1: Cell Disruption

There are several ways to break open cells.

- Lysis methods include lowering the ionic strength of the medium cells are kept in. This can cause cells to swell and burst. Mild surfactants may be used to enhance the efficiency of lysis. Most bacteria, yeast, and plant tissues, which have cell walls, are resistant to such osmotic shocks, however, and stronger disruption techniques are often required.
- Enzymes may be useful in helping to degrade the cell walls. Lysozyme, for example, is very useful for breaking down bacterial walls. Other enzymes commonly employed include cellulase (plants), glycanases, proteases, mannases, and others.
- Mechanical agitation may be employed in the form of tiny beads that are shaken with a suspension of cells. As the beads bombard the cells at high speed, they break them open. *Sonication* (20-50 kHz sound waves) provides an alternative method for lysing cells. The method is noisy, however, and generates heat that can be problematic for heat-sensitive compounds.
- Another means of disrupting cells involves using a “cell bomb". In this method, cells are placed under very high pressure (up to 25,000 psi). When the pressure is released, the rapid pressure change causes dissolved gases in cells to be released as bubbles which, in turn, break open the cells.
- *Cryopulverization* is often employed for samples having a tough extracellular matrix, such as connective tissue or seeds. In this technique, tissues are .ash-frozen using liquid nitrogen and then ground to a fine powder before extraction of cell contents with a buffer.

Whatever method is employed, the crude lysates obtained contain all of the molecules in the cell, and thus, must be further processed to separate the molecules into smaller subsets, or fractions.

9.2: Fractionation

Fractionation of samples typically starts with centrifugation. Using a centrifuge, one can remove cell debris, and fractionate organelles, and cytoplasm. For example, nuclei, being relatively large, can be spun down at fairly low speeds. Once nuclei have been sedimented, the remaining solution, or supernatant, can be centrifuged at higher speeds to obtain the smaller organelles, like mitochondria. Each of these fractions will contain a subset of the molecules in the cell. Although every subset contains fewer molecules than does the crude lysate, there are still many hundreds of molecules in each. Separating the molecule of interest from the others is where chromatography comes into play. We will consider several separation techniques.

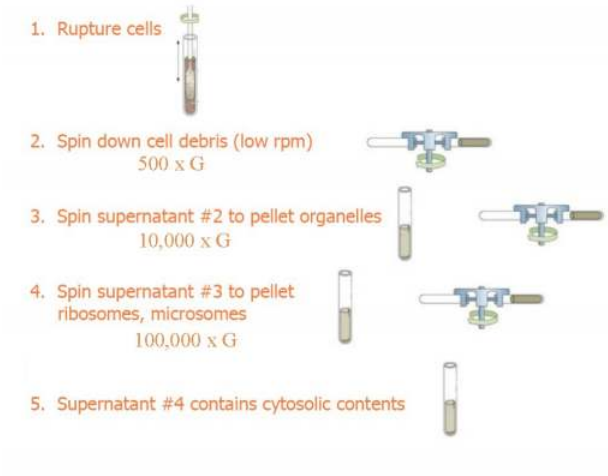

Figure 9.2.1: *Fractionation by centrifugation*

Many chromatographic techniques are performed in “columns." These are tubes containing the material (called the “support") used to perform the separation . Supports are designed to exploit the chemical, or size, differences of the many molecules in a mixture. Columns are “packed" (filled) with the support and a buffer or solvent carries the mixture of compounds to be separated through the support. Molecules in the sample interact differentially with the support and consequently, will travel through it with different speeds.

9.3: Ion Exchange Chromatography

In ion exchange chromatography, the support consists of tiny beads to which are attached chemicals possessing a charge. Each charged molecule has a counter-ion. The figure shows the beads (blue) with negatively charged groups (red) attached. In this example, the counter-ion is sodium, which is positively charged. The negatively charged groups are unable to leave the beads, due to their covalent attachment, but the counter- ions can be “exchanged" for molecules of the same charge. Thus, a cation exchange column will have positively charged counter-ions and positively charged compounds present in a mixture passed through the column will exchange with the counter-ions and “stick" to the negatively charged groups on the beads. Molecules in the sample that are neutral or negatively charged will pass quickly through the column. On the other hand, in anion exchange chromatography, the chemical groups attached to the beads are positively charged and the counter-ions are negatively charged. Molecules in the sample that are negatively charged will “stick" and other molecules will pass through quickly. To remove the molecules “stuck" to a column, one simply needs to add a high concentration of the appropriate counter-ions to displace and release them. This method allows the recovery of all components of the mixture that share the same charge.

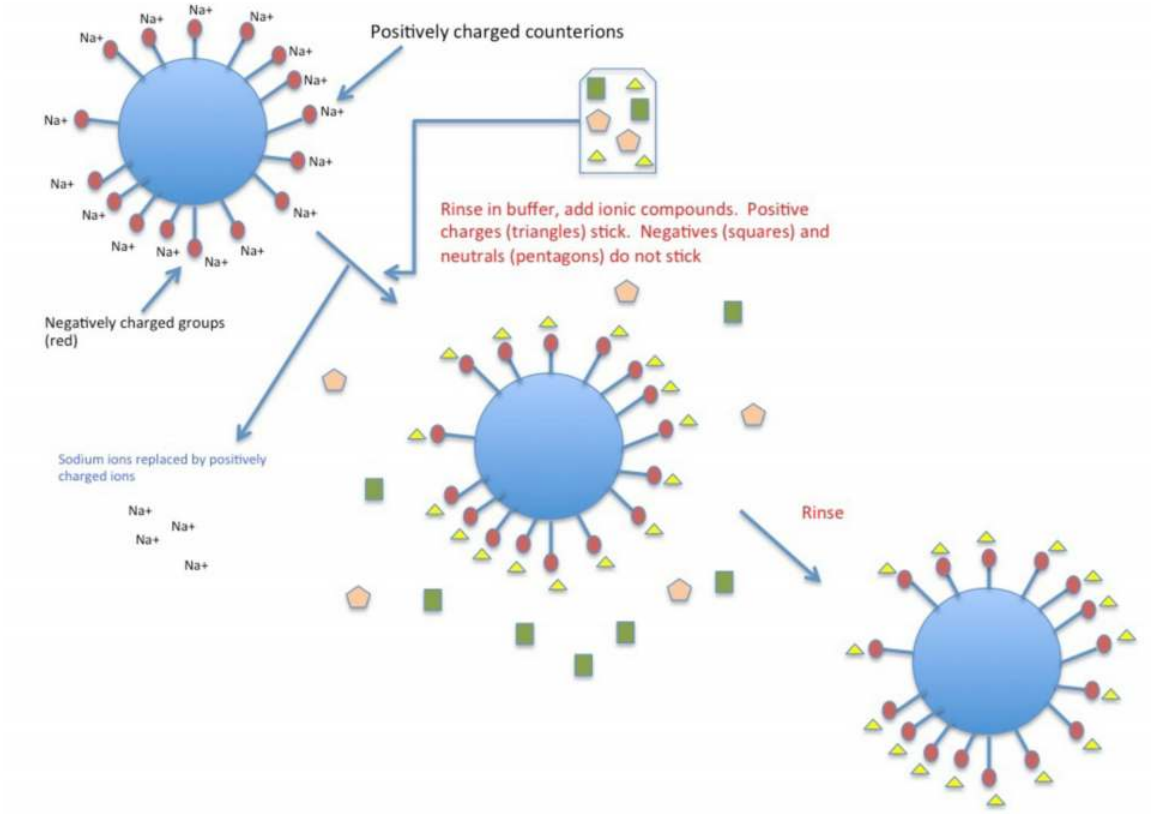

Figure 9.3.1: *Cation exchange chromatography*

9.4: Gel Exclusion Chromatography

Gel exclusion chromatography (also called molecular exclusion chromatography, size exclusion chromatography, or gel filtration chromatography) is a low resolution isolation method that employs a cool “trick." This involves the use of beads that have tiny “tunnels" in them that each have a precise size. The size is referred to as an “exclusion limit," which means that molecules above a certain molecular weight will not fit into the tunnels. Molecules with sizes larger than the exclusion limit do not enter the tunnels and pass through the column relatively quickly by making their way between the beads. Smaller molecules, which can enter the tunnels, do so, and thus, have a longer path that they take in passing through the column. Because of this, molecules larger than the exclusion limit will leave the column earlier, while those that pass through the beads will elute from the column later. This method allows separation of molecules by their size.

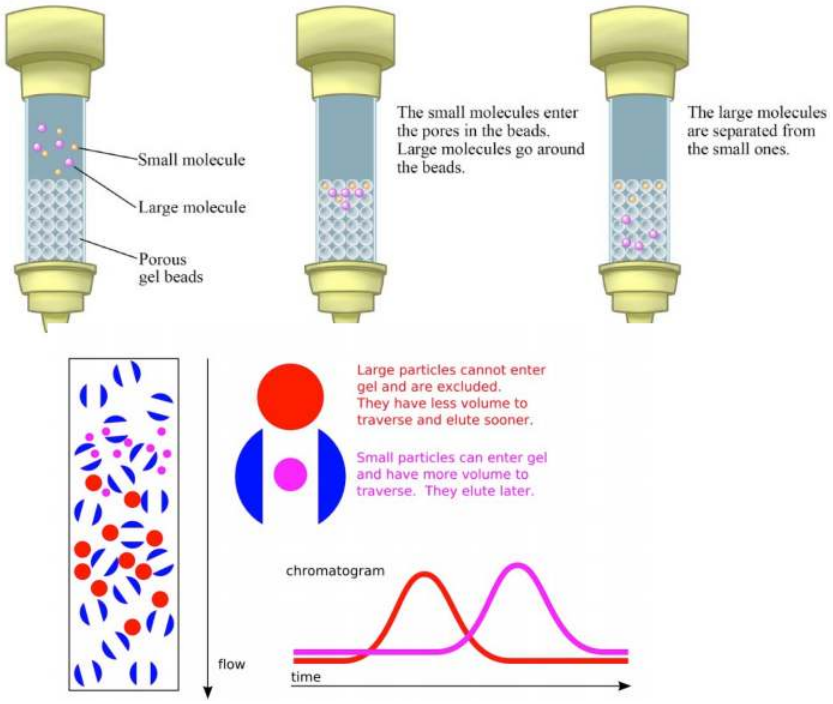

Figure 9.4.1: *Gel Exclusion Chromatography*

9.5: Affinity Chromatography

Affinity chromatography is a very powerful technique that exploits the binding affinities of target molecules (typically proteins) for substances covalently linked to beads. For example, if one wanted to separate all of the proteins in a sample that bound to ATP from proteins that do not bind ATP, one could covalently link ATP to support beads and then pass the sample through column. All proteins that bind ATP will “stick" to the column, whereas those that do not bind ATP will pass quickly through it. The proteins adhering to the column may then released from the column by adding ATP.

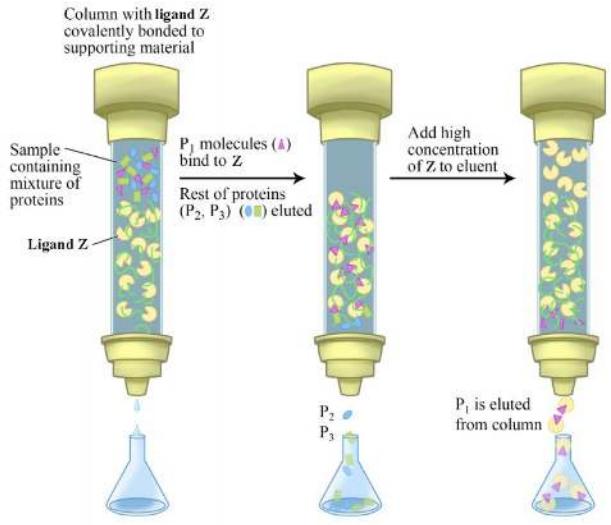

Figure 9.5.1: *Affinity Chromatography*

9.6: High Performance Liquid Chromatography (HPLC)

HPLC (also sometimes called High Pressure Liquid Chromatography) is a powerful tool for separating smaller molecules based on their differential polarities. It employs columns with supports made of very tiny beads that are so tightly packed that .ow of solvents/buffers through the columns requires the application of high pressures (hence the name). The supports used can be polar (normal phase separation) or non-polar (reverse phase separation). In normal phase separations, non-polar molecules elute first followed by the more polar compounds. This order is switched in reverse phase chromatography. Of the two, reverse phase is much more commonly employed to due more reproducible chromatographic profiles (separations) that it typically produces.

9.7: Histidine Tagging

Histidine tagging is a powerful tool for isolating a recombinant protein from a cell lysate. It relies on using recombinant DNA techniques to add codons specifying a series of histidines (usually six) to the coding sequence for a protein. The protein produced when this gene is expressed has a run of histidine residues fused at either the carboxyl or amino terminus to the amino acids in the remainder of the protein. The histidine side chains of this “tag" have an affinity for nickel or cobalt ions, making separation of histidine tagged proteins from a cell lysate is relatively easy. Simply passing the sample through a column that has immobilized nickel or cobalt ions allows the histidine- tagged proteins to “stick," while the remaining cell proteins all pass quickly through. The histidine-tagged proteins are then eluted by addition of imidazole (which is chemically identical to the histidine side chain) to the column. Histidine tags can be cleaved off using endopeptidases.

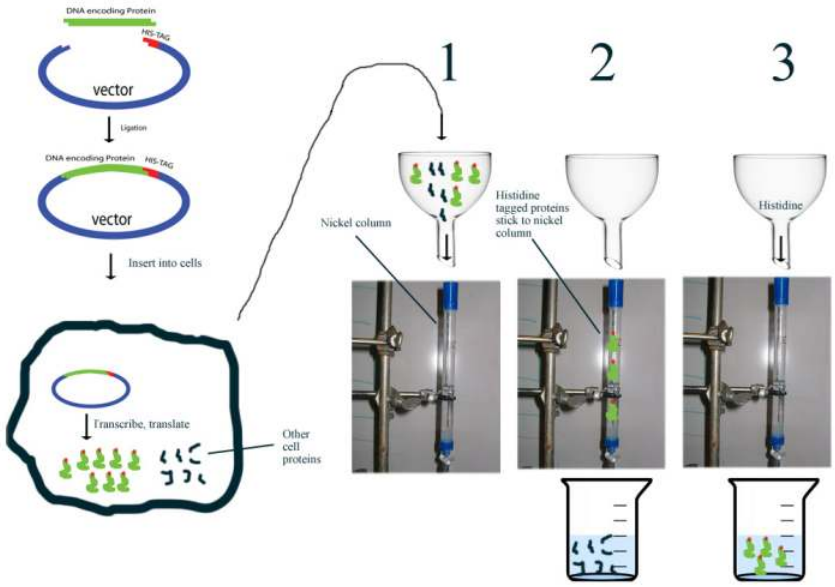

Figure 9.7.1: *Histidine Tagging*

9.8: Electrophoresis

DNA molecules are long and loaded with negative charges, thanks to their phosphate backbones. Electrophoretic methods separate large molecules, such as DNA, RNA, and proteins based on their charge and size. For DNA and RNA, the charge of the nucleic acid is proportional to its size (length). For proteins, which do not have a uniform charge, a clever trick is employed to make them mimic nucleic acids.

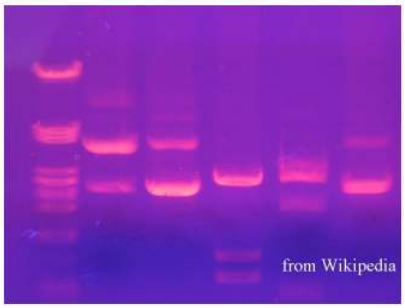

Figure 9.8.1: *DNA fragments separated by agarose gel electrophoresis*

Agarose

Agarose gel electrophoresis is a method for separating nucleic acids. It is worth noting that nucleic acids are the largest molecules found in cells, in some cases by orders of magnitude. Agarose provides a matrix which encases a buffer. The matrix provides openings for macromolecules to move through and the largest macromolecules have the most difficult time navigating, whereas the smallest macromolecules slip through it the easiest. Unlike column chromatography, electrophoresis uses an electric current as a force to drive the molecules through the matrix. Since the size to charge ratios for DNA and RNA are constant for all sizes of these nucleic acids, the size per force is also constant (since force is directly proportional to charge), so the molecules simply sort on the basis of their size - the smallest move fastest and the largest move slowest. Visualization of the DNA fragments in the gel is made possible by addition of a dye, such as ethidium bromide that fluoresces under ultraviolet light.

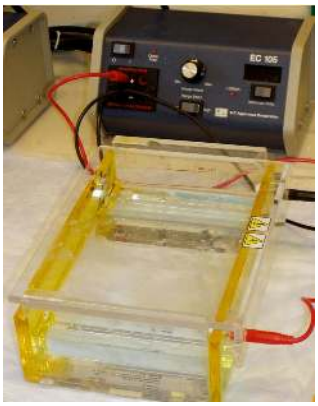

Figure 9.8.2: *Agarose gel electrophoresis*

SDS-PAGE

Like DNA and RNA, proteins are large macromolecules. Proteins, however, vary tremendously in their charge. Whereas double-stranded DNA is rod-shaped, most proteins are globular (folded up). Further, proteins are considerably smaller than nucleic acids, so the openings of the matrix of the agarose gel are simply too large to effectively provide separation. Consequently, unlike nucleic acids, proteins cannot be effectively separated by electrophoresis on agarose gels. To separate proteins by electrophoresis, one must make several modifications. First, a matrix made by polymerizing and crosslinking acrylamide units is employed. One can adjust the pore size of the matrix readily by changing the percentage of acrylamide in the gel. Higher percentages of acrylamide create smaller pores and are more effective in separating smaller molecules, whereas lower percentages of acrylamide reverse that. Second, proteins must be physically altered to “present" themselves to the matrix like the negatively charged rods of DNA. This is accomplished by treating the proteins with the detergent called SDS (sodium dodecyl sulfate). SDS denatures the proteins so they

assume a rod-like shape and the SDS molecules coat the proteins such that the exterior surface is loaded with negative charges proportional to the mass, just like the backbone of DNA. Third, a “stacking gel" may be employed at the top of the gel to provide a way of compressing the samples into a tight band before they enter the main polyacrylamide gel (called the resolving gel). Just as DNA fragments get sorted on the basis of size (largest move slowest and smallest move fastest), the proteins migrate through the gel matrix at rates inversely related to their size. Upon completion of the electrophoresis, there are several means of staining to visualize the proteins on the gel. They include reagents, such as Coomassie Brilliant Blue or silver nitrate (the latter is much more sensitive than Coomassie Blue staining and can be used when there are very small quantities of protein).

Figure 9.8.3: *SDS-Page Separation of proteins*

Isoelectric Focusing

Proteins vary considerably in their charges and, consequently, in their pI values (pH at which their charge is zero). Separating proteins by isoelectric focusing requires establishment of a pH gradient in an acrylamide gel matrix. The matrix’s pores are adjusted to be large to reduce the effect of sieving based on size. Molecules to be focused are applied to the gel with the pH gradient and an electric current is passed through it. Positively charged molecules, for example, move towards the negative electrode, but since they are traveling through a pH gradient, as they pass through it, they reach a region where their charge is zero and, at that point, they stop moving. They are at that point attracted to neither the positive nor the negative electrode and are thus “focused" at their pI. By using isoelectric focusing, it is possible to separate proteins whose pI values differ by as little as 0.01 units.

2-D Gel Electrophoresis

Both SDS-PAGE and isoelectric focusing are powerful techniques, but a clever combination of the two is a powerful tool of proteomics - the science of studying all of the proteins of a cell/tissue simultaneously. In 2D gel electrophoresis, an extract containing the proteins is first prepared. One might, for example, be studying the proteins of liver tissue. The liver cells are lysed and all of the proteins are collected into a sample. Next, the sample is subjected to isoelectric focusing as described earlier, to separate the proteins by their pI values. Next, as shown on the previous page, the isoelectric gel containing the separated proteins is rotated through 90º and placed on top of a regular polyacrylamide gel for SDS-PAGE analysis (to separate them based on size). The proteins in the isoelectric gel matrix are electrophoresed into the polyacrylamide gel and separation on the basis of size is performed. The product of this analysis is a 2D gel, in which proteins are sorted by both mass and charge.

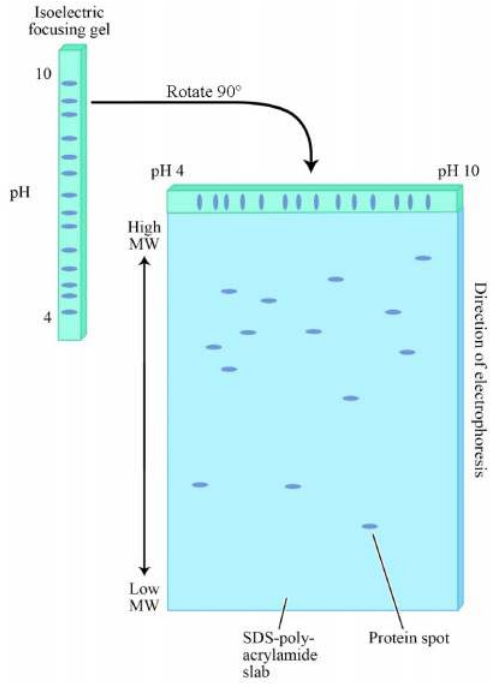

Figure 9.8.4: *2D Gel Electrophoresis*

The power of 2D gel electrophoresis is that virtually every protein in a cell can be separated and appear on the gel as a distinct spot. In the figure, spots in the upper left correspond to large positively charged proteins, whereas those in the lower right are small negatively charged ones. It is possible using high- throughput mass spectrometry analysis to identify every spot on a 2D gel. This is particularly powerful when one compares protein profiles between different tissues or between the samples of the same tissue treated or untreated with a particular drug. Comparison of a 2D separation of a non-cancerous tissue with a cancerous tissue of the same type provides a quick identification of proteins whose level of expression differs between them. Information such as this might be useful in designing treatments or in determining the mechanisms by which the cancer arose.

9.9: Protein Cleavage

Working with intact proteins in analytical techniques, such as mass spectrometry, can be problematic. Consequently, it is often desirable to break a large polypeptide down into smaller, more manageable pieces. There are two primary approaches to accomplishing this - use of chemical reagents or use of proteolytic enzymes. The table on the previous page shows the cutting specificities of various cleavage agents.

Cleavage Reagent	Specificity
1. Cyanogen bromide	COOH side of MET
2. Hydroxylamine	ASP-GLY bonds
3. 2-Nitro-5-thiocyanobenzoate	NH2 side of CYS
4. Trypsin	COOH side of LYS or ARG
5. Chymotrypsin	COOH side of TYR, TRP, PHE, LEU, MET
6. Thrombin	COOH side of arginine
7. Elastase	ALA-ALA or ALA-GLY bonds

Figure 9.9.1: *Protein cleaving agents*

9.10: Microarrays

2D gels are one way of surveying a broad spectrum of molecules simultaneously. Other approaches to doing the same thing involve what are called microarrays. DNA microarrays, for example, can be used to determine all of the genes that are being expressed in a given tissue, simultaneously. Microarrays employ a grid (or array) made of rows and columns on a glass slide, with each box of the grid containing many copies of a specific molecule, say a single-stranded DNA molecule corresponding to the sequence of a single unique gene. As an example, consider scanning the human genome for all of the known mRNA sequences and then synthesizing single stranded DNAs complementary to each mRNA. Each complementary DNA sequence would have its own spot on the matrix. The position of each unique gene sequence on the grid is known and the entire grid would represent all possible genes that are expressed. Then for a simple gene expression analysis, one could take a tissue (say liver) and extract the mRNAs from it. These mRNAs represent all the genes that are being expressed in the liver at the time the extract was made.

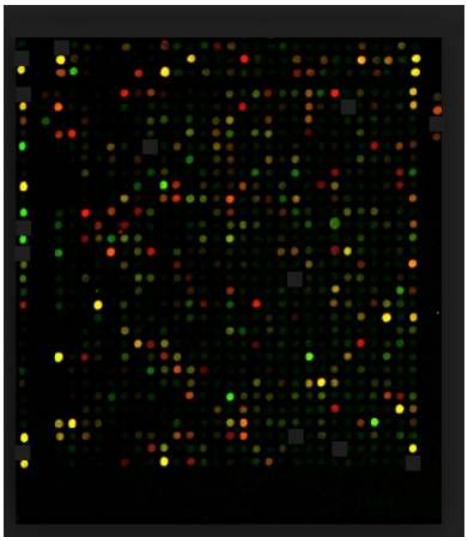

Figure 9.10.1: *Microarray*

The mRNAs can easily be tagged with a colored dye (say blue). The mixture of tagged mRNAs is then added to the array and base-pairing conditions are created to allow complementary sequences to find each other. When the process is complete, each liver mRNA should have bound to its corresponding gene on the array, creating a blue spot in that box on the grid. Since it is known which genes are in which box, a blue spot in a box indicates that the gene in that box was expressed in the liver. The presence and abundance of each mRNA may then readily determined by measuring the amount of blue dye at each box of the grid. A more powerful analysis could be performed with two sets of mRNAs, each with a different colored tag (say blue and yellow). One set of mRNAs could come from the liver of a vegetarian (tagged blue) and the other from a meat eater (tagged yellow), for example. The mRNAs are mixed and then added to the array and complementary sequences are once again allowed to form duplexes. After unhybridized mRNAs are washed away, the plate is analyzed. Blue spots in grid boxes correspond to mRNAs present in the vegetarian liver, but not in that of the meat eater. Green spots (blue plus yellow) would correspond to mRNAs present in equal abundance in the two livers. The intensity of each spot would also give information about the relative amounts of each mRNA in the tissues. Similar analyses could be done, using cDNAs instead of mRNA. Peptide microarrays have peptides bonded to the glass slide instead of DNA and can be used to study the binding of proteins or other molecules to the peptides.

9.11: Blotting

Blotting provides a means of identifying specific molecules out of a mixture. It employs three main steps. First, the mixture of molecules is separated by gel electrophoresis. The mixture could be DNA (Southern Blot), RNA (Nothern Blot), or protein (Western Blot) and the gel could be agarose (for DNA/RNA) or polyacrylamide (for protein). Second, after the gel run is complete, the proteins or nucleic acids in the gel are transferred out of the gel onto a membrane/paper that physically binds to the molecules. This "blot", as it is called, has an imprint of the bands of nucleic acid or protein that were in the gel (see figure at left). The transfer can be accomplished by diffusion or by using an electrical current to move the molecules from the gel onto the membrane. The membrane may be treated to covalently link the bands to the surface of the blot. Last, a visualizing agent specific for the molecule of interest in the mixture is added to the membrane. For DNA/RNA, that might be a complementary nucleic acid sequence that is labeled in some fashion (radioactivity or dye). For a protein, it would typically involve an antibody that specifically binds to the protein of interest. The bound antibody can then be targeted by another antibody specific for the first antibody. The secondary antibody is usually linked to an enzyme which, in the presence of the right reagent, catalyzes a reaction that produces a signal (color or light) indicating where the antibody is bound. If the molecule of interest is in the original mixture, it will "light" up and reveal itself.

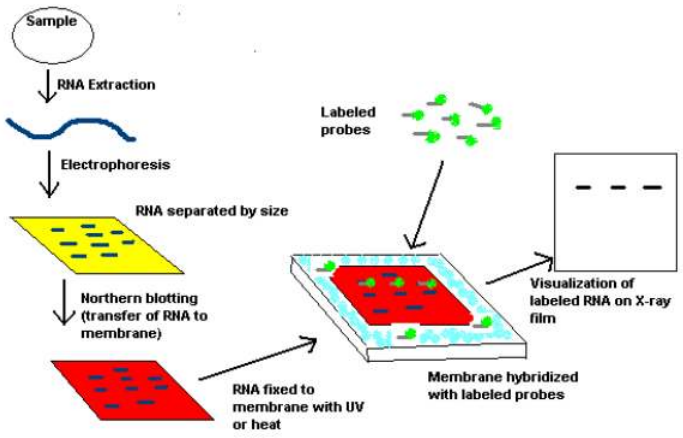

Figure 9.11.1: *Northern blotting procedure*

9.12: Making Recombinant DNAs

Molecular biologists often create recombinant DNAs by joining together DNA fragments from different sources. One reason for making recombinant DNA molecules is to enable the production of a specific protein that is of interest. For example, it is possible to engineer a recombinant DNA molecule containing the gene for human growth hormone and introduce it into an organism like a bacterium or yeast, which could make massive quantities of the human growth hormone protein very cheaply. To do this, one needs to set up the proper conditions for the protein to be made in the target cells. For bacteria, this typically involves the use of plasmids. Plasmids are circular, autonomously replicating DNAs found commonly in bacterial cells. Plasmids used in recombinant DNA methods

1. replicate in high numbers in the host cell;
2. carry markers that allow researchers to identify cells carrying them (antibiotic resistance, for example) and
3. contain sequences (such as a promoter and Shine Dalgarno sequence) necessary for expression of the desired protein in the target cell. A plasmid that has all of these features is referred to as an expression vector (see an example in the figure at left).

Plasmids may be extracted from the host, and any gene of interest may be inserted into them, before returning them to the host cell. Making such recombinant plasmids is a relatively simple process. It involves

1. cutting the gene of interest with a restriction enzyme (endonucleases which cut at specific DNA sequences);
2. cutting the expression plasmid DNA with restriction enzyme, to generate ends that are compatible with the ends of the gene of interest;
3. joining the gene of interest to the plasmid DNA using DNA ligase;
4. introducing the recombinant plasmit into a bacterial cell; and
5. growing cells that contain the plasmid. The bacterial cells bearing the recombinant plasmid may then be induced to express the inserted gene and produce large quantities of the protein encoded by it.

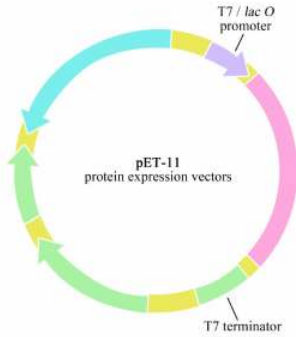

Figure 9.12.1: *An expression vector*

9.13: Polymerase Chain Reaction

PCR allows one to use the power of DNA replication to obtain large amounts of a specific DNA in a short time. As everyone knows, cell division results in doubling the number of cells with each round of division. Each time cells divide, DNA must be replicated, as well, so the amount of DNA is doubling as the cells are doubling. Kary Mullis recognized this fact and came up with the technique of PCR, which mimics DNA replication. In contrast to cellular DNA replication, which amplifies all of a cell's DNA during a replication cycle, PCR is used to replicate only a specific segment of DNA. This segment of DNA, known as the target sequence, is replicated repeatedly, to obtain millions of copies of the target. Just as in DNA replication, PCR requires a template DNA, 4 dNTPs, primers to initiate DNA synthesis on each strand, and a DNA polymerase to synthesize the new DNA copies. In PCR, the primers bind to sequences .anking the target region that is to be amplified, and are present in large excess over the template. The DNA polymerase used is chosen to be heat stable, for reasons that will be clear shortly. The first step of each PCR cycle involves separating the strands of the template DNA so that it can be replicated. This is accomplished by heating the DNA to near boiling temperatures. In the next step of the cycle, the solution is cooled to a temperature that favors complementary DNA sequences finding each other. Since the primers are present in great excess over the template, they can readily find and base-pair with the complementary sequences in the template on either side of the target sequence. In the third step in the cycle, the DNA polymerase (which has not been denatured during the heat treatment because it is thermostable) extends the primer on each strand, making copies of both DNA strands and doubling the amount of the target sequence. The cycle is then repeated, usually about 30 times. At the end of the process, there is a theoretical yield of 2^{30} more of the target DNA than there was in the beginning.

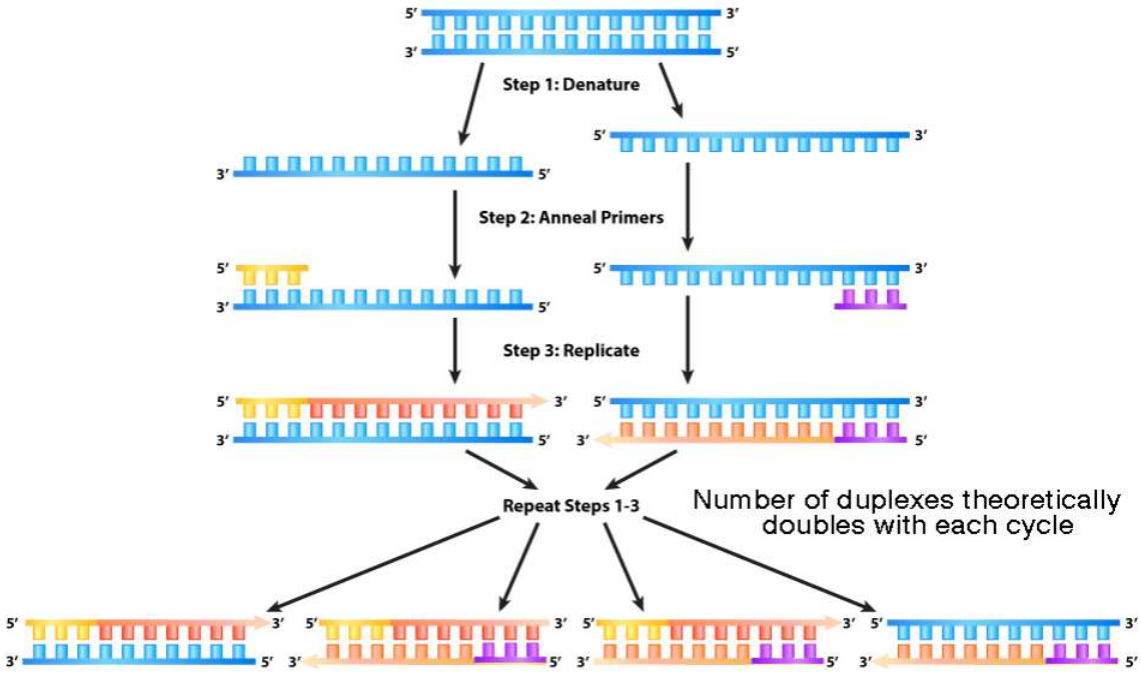

Figure 9.13.1: *PCR*

9.14: Lac Z Blue-White Screening

A powerful tool for biotechnologists is the lac Z gene. You may recall from an earlier section on the control of gene expression, that lac Z is part of the lac operon of E. coli and encodes the enzyme ß galactosidase. This enzyme catalyzes the hydrolysis of lactose into glucose and galactose, allowing the bacteria to use lactose as an energy source. ß galactosidase can also break down an artificial substrate called X-gal to produce a compound that is blue in color. X-gal can thus be used to test for the presence of active ß galactosidase. With this background, we can now look at how the lac Z gene can be of help to molecular biologists when they create recombinant plasmids. In the example described earlier, the gene for human growth hormone (hGH) was inserted into a plasmid. As we noted, the plasmid, as well as the hGH gene are cut with restriction endonucleases to create compatible DNA ends that can be ligated. While the ends of the hGH gene are, indeed, capable of being ligated to the ends of the plasmid, the two ends of the plasmid could also readily rejoin. In fact, given that the two ends of the plasmid are are on the same molecule, the chances of their finding each other are much higher than of a plasmid end finding an hGH gene. This would mean that many of the ligated molecules would not be recombinants, but simply recircularized plasmids. Five percent of the plasmids having inserts of the hGH gene would be very good. That would mean that 95% of the bacterial colonies arising from transformation would contain the original plasmid rather than the recombinant. To make the process of screening for the relatively rare recombinants simpler, plasmids have been engineered that carry the lac Z gene, modified to contain, with the coding sequence, restriction enzyme recognition sites. If one of these sites is used to cut open the plasmid and a gene of interest is inserted, this disrupts the lac Z gene. If the plasmid simply recircularizes, the lac Z gene will be intact. To find which bacterial colonies carry the recombinant plasmids, X-Gal is provided in the plates. Bacterial colonies containing plasmids with the lac z sequence disrupted by an inserted gene will not produce functional ßgalactosidase. The X-Gal will not be broken down and there will be no blue color. By contrast, bacterial cells with recircularized plasmids having no inserted hGH gene will make functional ß galactosidase, so in the presence of X-Gal and IPTG these colonies will produce a blue color. This is summarized in the figure on the previous page.

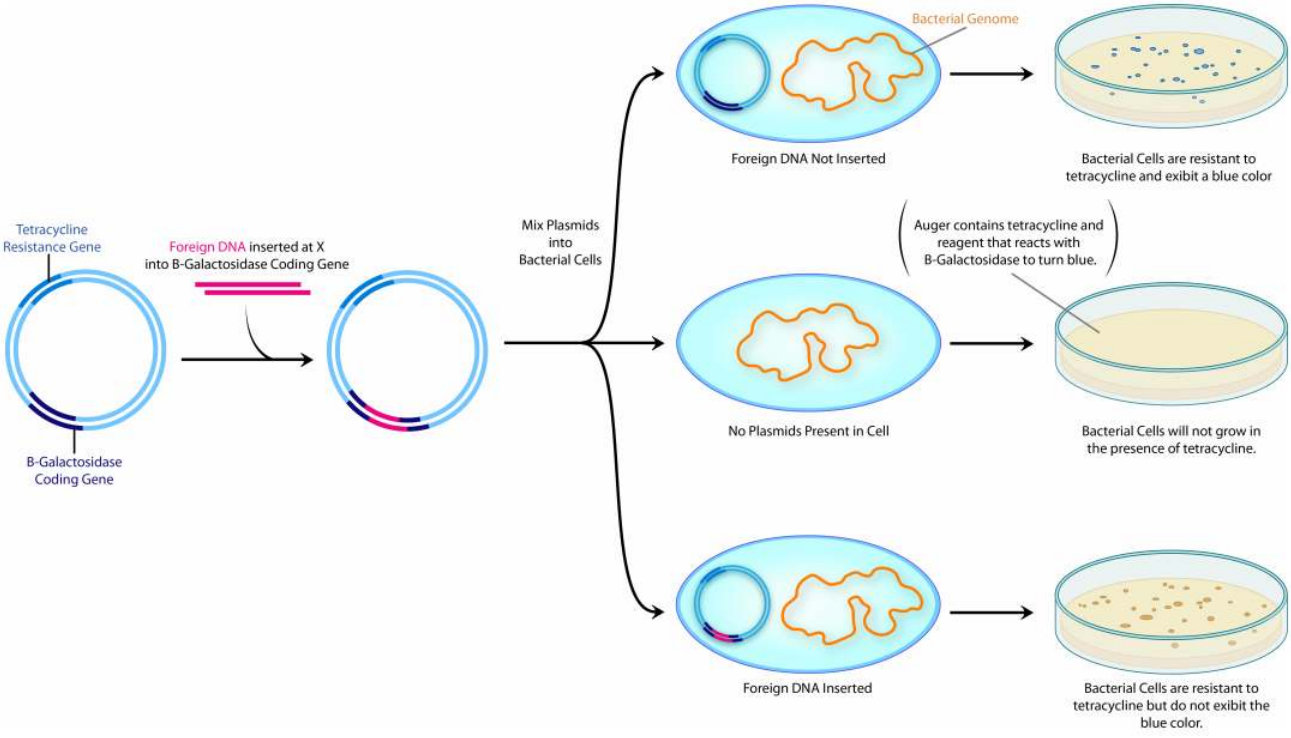

Figure 9.14.1: *Blue-White Screening Strategy*

9.15: Reverse Transcription

According to the central dogma, DNA codes for mRNA, which codes for protein. An exception to this rule is seen with the retroviruses, RNA-encoded viruses that have a phase in their replication cycle during which their genomic RNA is copied into DNA by a virally-encoded enzyme known as reverse transcriptase. The ability to convert RNA to DNA can be useful in the laboratory. For example, the power of PCR can be brought to RNA by converting RNAs of interest to DNA and then amplifying them by PCR. With reverse transcriptase, this is readily accomplished. First, one creates a DNA oligonucleotide to serve as a primer for reverse transcriptase to use on a target RNA. The primer must, of course, be complementary to a segment (near the 3' end) of the RNA to be amplified. The RNA template, reverse transcriptase, the primer, and four dNTPs are mixed. With one round of replication, the RNA is converted to a single strand of DNA, which can be separated from the RNA either by heating or by the use of an RNase to digest the RNA. The product of this process is called a complementary DNA (cDNA).

CHAPTER OVERVIEW

10: Putting It All Together

With this chapter, we tie up a bunch of loose ends and ponder what lies in the future of biochemistry.

10.1: Looking Back

The bounds of biochemistry have expanded enormously since its inception. Wöhler's demonstration, in 1828, that urea could be synthesized outside of a living cell, showed that there was no "vital force" that distinguished the chemistry of life from that of the non-living world. Chemistry is chemistry, but the term "biochemistry" was coined in 1903 by Carl Neuberg to describe the special subset of chemical reactions that happen in living cells. This specialness derives not from any exceptions to the laws of physics and chemistry, but from the way in which the chemical reactions in cells are organized and regulated, and also from the complexity and size of biological molecules.

Faced with far greater complexity than in the inorganic world, the traditional strategy of biochemists has been "divide and conquer." In this approach, individual enzymes and other biological molecules are purified from cells so that their properties can be studied in isolation. The underlying logic of this method, sometimes described as reductionist, is that we can learn about the whole by studying its individual parts. This painstaking approach, used through most of the twentieth century, teased out chemical reactions and molecular interactions that occur within cells, one by one, gradually revealing to scientists much of what we know in biochemistry today.

As increasing numbers of biochemical reactions were worked out, biochemists began to see that they were connected together in chains of reactions that we now refer to as metabolic pathways. These metabolic pathways turned out to be remarkably similar between cells across all kingdoms of life. Though there are a few pathways that are unique to certain organisms, many more are the same, or very similar, in organisms as different as bacteria and humans.

It also became clear that metabolic pathways interacted with each other via common intermediates or by regulation of one pathway by molecule(s) created by other pathway(s). The similarity of the chemical reactions in all living cells was shown to extend to the common energy currency, ATP, that cells use to power their chemical reactions, as well as the mechanism by which cells make the ATP.

Metabolic pathways trace the transformation of molecules in a cell and represent the work of enzymes, which are proteins. The discovery of the structure of DNA led to understanding of how information in genes was used to direct the synthesis of these proteins. The protein-DNA interactions that determine which genes are copied into RNA at any given time were uncovered and helped explain how cells with the same DNA came to express different proteins. The genetic code, as well as the mechanisms of transcription, translation and regulation of gene expression also turned out to be remarkably similar in cells of all kinds, leading Nobel laureate Jacob Monod to joke that what was true for E.coli was also true for E.lephant.

The "one component at a time" approach also helped biochemists understand how cells sense changes in their environment and respond to them. The ability to sense conditions outside the environs of cells extends through all groups of organisms. Even the simplest single-celled organism can follow nutrient gradients to move itself closer to food. Cells in multicellular organisms can detect chemical cues in the blood (nutrients, hormones) or impulses from nerve cells and alter their actions. These cues may trigger changes in metabolism, decisions to divide, die, or become senescent, or the performance of specialized functions (e.g., muscle contraction or enzyme secretion). Thus cells are constantly in a state of .ux, adjusting their activities in response to signals from outside themselves as well as their own changing needs.

The power of the "take things apart" analytical approach is evident from the astounding pace of discoveries in biochemistry and molecular biology. The first demonstration that an enzyme was a protein was made only in 1926, and it wasn't till twenty years later that this was sufficient well established that the Nobel Prize was awarded in 1946 for this discovery. Since that time, the methods of biochemistry have uncovered all of the information that you can find in any standard biochemistry textbook, and more.

Thousands of enzymes and their substrates have been identified, and hundreds of metabolic pathways traced. The structure of hundreds of proteins is known down to the position of every atom. Following the elucidation of the structure of DNA in 1953, scientists have discovered a dizzying number of facts about how information is stored, used and inherited in cells. Cloned and transgenic animals and gene therapy were a reality in less than 50 years. And the discoveries still keep coming.

10.2: Looking Forward

Toward the end of the twentieth century, new methods began to change the face of biochemistry. The launching of the Human Genome Project and the development of faster and cheaper sequencing technologies provided biochemists with entire genome sequences, not only of humans, but of numerous other organisms. Huge databases were set up to deal with the volume of sequence information generated by the various genome projects. Computer programs cataloged and analyzed these sequences, making sense of the enormous quantities of data.

Protein coding regions of genomes could be identified and translated “in silico" to deduce the amino acid sequence of the encoded polypeptides. Comparisons could be made between the gene sequences of different organisms. In parallel with the growth of sequence information, more and more protein structures were determined, by using X-ray crystallography and NMR spectroscopy. These structures, too, were deposited in databases to be accessible to all scientists.

The accumulation of vast amounts of sequence and structure information went hand in hand with new and ambitious goals for biochemistry. Modern biotechnology techniques have provided tools for studying biochemistry in entirely new ways. The old ways of dividing and conquering to study individual reactions are now being supplemented by approaches that permit researchers to study cellular biochemistry in its entirety.

These fields of research, which collectively are often referred to as the ‘-omics’ include genomics (study of all the DNA of a cell), proteomics (study of all the proteins of a cell), transcriptomics (study of all the transciption products of a cell), and metabolomics (study of all the metabolic reactions of a cell), among others. As an example, let us consider proteomics. The field of proteomics is concerned with all of the proteins of a cell. Since proteins are the ‘workhorses’ of cells, knowing which ones are being made at any given time provides us with an overview of everything that is happening in the cells under specific conditions.

How is such an analysis performed? First, one extracts all of the proteins from a given cell type (liver, for example). Next, the proteins are separated in a two-step gel method, where the first step resolves proteins based on their charge and the second separates them by mass. The product of this analysis is a single gel (called a 2-D gel) on which all of the proteins have been separated. In the left-right orientation, they differ in their original charge and in the up/down orientation, they differ in their size.

By using such a technique, as many as 6000 cellular proteins can be separated and visualized as spots on a single gel. Robotic techniques allow excision of individual spots and analysis on mass spectrometers to identify every protein present in the original extract.

Why is this useful? There are several ways in which this information can be illuminating. For example, by comparing the proteins in a normal liver cell with those in a cancerous liver cell, one can quickly determine if there are any proteins that are expressed or missing only in the cancer cells. These differences between normal and cancerous cells may provide clues to the mechanisms by which the cancer arose or suggest ways to treat the cancer. Or, the same sort of analysis could be done on cells to find out about the effects of a hormone or drug treatment. Comparison of the proteins found in untreated and treated cells would give a global view of the protein changes resulting from the treatment.

Similar analyses can be performed on the mRNA of cells, employing devices called microarrays. In this case, all the RNAs that are being made at the time that the cell extract is made can be identified by the signals generated when the RNAs hybridize with oligonucleotides complementary to their sequence, that are immobilized in ordered arrays on the surface of a plate. The position and strength of these signals indicates which RNAs are made and in what amounts.

The techniques of proteomics and transcriptomics, together with other “global view" approaches of molecules like lipids, carbohydrates, etc., are allowing biochemists to have, for the first time, a “big picture" view of the activities of cells. While these techniques have already provided valuable new insights, they are still incomplete, as a description of what goes on in cells. This is because they provide us with a snapshot that captures what is happening in cells at the moment that they were disrupted to make the extract. But cells are not static entities. At every moment, they are adapting their activities in response to changing combinations of internal and external conditions. Changes in response to any one signal are modified and in.uenced by the every other condition, within and outside the cell, and understand these complex systems as an integrated whole is the new holy grail of biochemistry.

The aim, then, is to develop models that depict these dynamic interactions within cells, and to understand how such interactions give rise to the properties and behavior that we observe. This is the goal of the emerging field of systems biology that constructs mathematical models and simulations, based on the large data sets generated by transcriptomic, proteomic and other broad-range techniques. Systems biology is truly an interdisciplinary venture, drawing as it does on mathematics and computer science as much as traditional “bench biochemistry". While the original laboratory techniques of biochemistry are by no means obsolete, they will no longer be the sole tools used to understand what goes on inside of cells.

These newer approaches are already leading to applications that are of tremendous value. Understanding the system level differences between normal and diseased cells can lead to major changes in the way diseases are detected, treated or altogether prevented.

One recent triumph of systems biology has been in an intriguing discovery about how antibiotic drugs work. System level studies of many classes of antibiotics revealed that, regardless of how we think they work to kill bacteria, all of the drugs appear to have a common effect – that of increasing the level of oxidative damage, leading to cell death. This observation suggested that the potency of antibiotics could be enhanced by blocking bacterial responses that protect against oxidation damage. This idea was tested by screening large numbers of compounds for the ability to inhibit a pathway that bacteria use to repair their oxidation-damaged DNA. This screen yielded several compounds, the best of which was able to increase the effectiveness of the drug gentamicin by about a thousand-fold. Such compounds will be of increasing value in a world where antibiotic resistance is on the rise.

Another application of systems biology is in the development of more effective vaccines. Till recently, most vaccines have been developed with little understanding of how exactly they stimulate the immune response. As systems biology approaches give us a better understanding of the changes that vaccines bring about to mediate immunity, it will be possible to identify the patterns that characterize stronger immune responses or adverse reactions to vaccines and even to predict how well particular vaccines may work in specific populations or individuals. Similarly, system level studies can help identify which drugs might be most effective, with the fewest side-effects, for a given patient, leading to a new era of personalized medicine.

Related to systems biology, and heavily dependent on it, is synthetic biology, which aims to use the knowledge gained from the former to engineer novel biological systems and pathways. Because the technology now exists to synthesize extremely long pieces of DNA, entire genomes can be made synthetically and used to program cells that they are inserted into. It also allows for the possibility of custom-designing an organism to create particular chemical compounds through artificially assembled pathways.

These methods have already been used to produce the drug artemisinin, which is used to treat malaria. The pathway for making a precursor of artemisinin was created by combining a metabolic pathway from yeast with part of another derived from the plant Artemisia annua, the natural source of artemisinin. Similar efforts are underway for anticancer drugs, novel drugs, favoring compounds, etc. One major goal is to create organisms programmed to make biofuels that could potentially replace petroleum.

The successes of systems and synthetic biology, even in their infancy, promise great advances both in our understanding of living systems and in the applications that arise out of that knowledge. The next fifty years in biological research may well eclipse even the amazing accomplishments of the last. The practice of medicine will be transformed. Regenerative medicine will improve, as a better knowledge of stem cells allows us to use them more effectively to replace cardiac muscle lost in a heart attack, neurons damaged in Parkinson’s or Alzheimer’s, or even to regrow limbs lost in accidents or war. Treatments for our illnesses can be tailored to be optimal for each individual. Biofuels may bail us out when oil supplies run out and engineered organisms may help clean up a polluted planet. And research on longevity may give us the best gift of all- lives extended long enough to witness these advances and to participate in the creation of a new and better world.

Index

R

Ramachandran plot
3.3: Proteins

Receptor Tyrosine Kinases (RTKs)
8.5: Receptor Tyrosine Kinases (RTKs)

replication
5.1: DNA Replication

ribonucleotide reductase (RNR)
7.12: Deoxyribonucleotide de novo Biosynthesis

RNA splicing
5.5: RNA Processing

rubisco
7.3: Calvin Cycle

S

secondary structure
3.3: Proteins

Sonication
9.1: Cell Disruption

Southern blot
9.11: Blotting

stop codons
5.6: Translation

Superhelicity
3.4: Nucleic Acids

11.2: Appendix 2 - Key Points

Basics

Buffers

Protein Structure

Macromolecular Characterization

Hemoglobin

Enzymes

Catalytic Strategies

Allostery and Regulation

Carbohydrate Nomenclature, Structure, and Function

Signaling

Metabolic Control

Glycolysis & Gluconeogenesis

Glycogen Metabolism

Citric Acid Cycle

Lipids and Membranes

Membrane Transport

Mitochondria and Energy

Photosynthesis

Lipid and Steroid Metabolism

Fatty Acid Metabolism

Nucleotide Metabolism

Nitrogen Metabolism

DNA Replication, Repair, Recombination

Transcription

Translation

Gene Expression

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Biochemistry Free & Easy (Ahern and Rajagopal)

Webpages: 109

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 91.7% (100 pages)
- Undeclared: 8.3% (9 pages)

By Page