Part I
Chemical and Physical Foundations

1 Thermodynamics and Kinetics

1.1 Laws of Thermodynamics

The **First Law of Thermodynamics** is a restatement of the Law of Conservation of Energy. From the perspective of thermal systems, it can be expressed as:

$$\Delta E = q + w \tag{1}$$

where E represents the change in energy of a system, q represents the heat flow into or out of a system, and w represents work done by or on the system. The numbers here are, by convention, positive if the system gains energy (so if heat moves into the system, or if work is done on the system) and vice versa. Under constant pressure (which is where most biological systems operate), q is also ΔH, or **enthalpy**. ΔH and ΔE are **state functions**, in the sense that they are properties of a system at a specific state and are independent of how one arrives at that state.

For a process to be **reversible**, it must occur such that the system is constantly at **equilibrium** through the course of the process.

The **Second Law of Thermodynamics** states that the **entropy** of the universe must increase or stay constant. Change in entropy is oftentimes referred to by ΔS (also a state function) and refers to the change in the "disorder" of a system, or, more mathematically precise, in the multiplicities of the system. This law simplifies into the understanding of the concept of **free energy**, or ΔG (referring to Gibb's Free Energy, the free energy at constant pressure and temperature).

$$\Delta G = \Delta H - T\Delta S \tag{2}$$

The formula above gives a state function, the Free Energy, which serves as an indicator of the overall entropy change of the universe, whereby a positive ΔG indicates an overall reduction in universal entropy and vice versa.

$$\Delta G = \Delta G_0 + RT \ln Q \tag{3}$$

Refers to how to calculate ΔG at non-standard conditions where R is the ideal gas constant, ΔG_0 is the free energy under certain standard conditions, T is the temperature, and Q is a mathematical calculation representing the deviation from standard condition (as Blincoe would say, "products over reactants").

These state functions can be used to calculate properties of reactions. This can be done by taking the total ΔG, ΔH, or ΔS of formation of the products (which can be found in a table and represent the change in a given state function of producing the substance from its original elements) and subtracting the total values of formation for the reactants ("products minus reactants").

1.2 Kinetics

A reaction is first order if the rate of the reaction is proportional to the first power of a reactant ($\frac{d[A]}{dt} = -k[A]$). A reaction is then nth order for a specific reactant if the rate is proportional to the nth power of that reactant ($\frac{d[A]}{dt} = -k[A]^n$). For 1st order reactions:

$$\begin{aligned} \frac{d[A]}{dt} &= -k[A] \\ A(t) &= [A]_0 e^{-kt} \end{aligned} \tag{4}$$

For second order reactions:

$$\begin{aligned} \frac{d[A]}{dt} &= -k[A]^2 \\ \frac{1}{[A](t)} &= kt + \frac{1}{[A]_0} \end{aligned} \tag{5}$$

It should be noted that the rate of the reaction depends on the **activation energy** of a reaction (the hump in the reaction-energy diagram), and not on the free energy difference (ΔG) or any other state function difference between products and reactants.

2 Redox States

There are two usages of "redox state." The most formal refers to the charge on an atom. In ionic compounds, this is simply the charge of the atom in the ion (ie in NaCl, Na's redox state is +1 and Cl's is -1). In covalent structures, this is simply the formal charge on the atom in question.

There is another usage which is tied to the first, but refers more loosely to how electron-deficient an atom is. A carbon bonded to another carbon, for instance, would not be particularly electron-deficient. But, a carbon bonded to a nitrogen or an oxygen would be. Therefore a carboxylic acid is oxidized relative to a ketone is oxidized relative to an alcohol ... It would be even more so if it were doubly bonded to an oxygen. This state of electron deficiency is considered to be oxidized relative to the original carbon-carbon bond status.

Because of the change in the charges involved, reduction and oxidation involve electron transfers.

3 Water, pH, Acid-Base Reactions

The BrØnsted Lowry definition of an **acid** is a molecule which donates H^+ and the definition of a **base** is a molecule which reduces H^+ levels. The acidity of a substance is given by its K_a which is evaluated at equilibrium to be:

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{6}$$

where $[A^-]$ represents the concentration of the **conjugate base**, or the acid after it has donated a proton, and $[HA]$ represents the concentration of undissociated acid. The basicity of a substance is given by its K_b which is ultimately $\frac{K_w}{K_a}$ where K_a refers to the acidity of the base's conjugate acid (the base upon gaining a proton) and $K_w = [H^+][OH^-] = 10^{-14}$.

The equation of importance here is the **Henderson-Hasselbach Equation** which defines the pH of a solution under specific circumstances of acid/conjugate base levels:

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \tag{7}$$

This relationship suggests that solutions made using the salts of acids/bases or their conjugates can have specific pH's.

Buffers are solutions made from acids/bases and the salts of their conjugates which are able to resist pH change. Mathematically, this is accomplished by having a sufficiently large volume of buffer such that the ratio $\frac{[A^-]}{[HA]}$ does not change substantially.

3.1 Solutions and Equilibria

At equilibria ($\Delta G = 0$), all reactions have a K property (products over reactants) which is dependent not on the rate of the reaction but on the overall free energy difference between the products and the reactants, according to:

$$\Delta G = -RT \ln K \tag{8}$$

This K is related to kinetics indirectly in the sense that it is a quotient of the forward and reverse rates of reaction (depending on how one defines K). A similar K dependence occurs also for dissolving substances.

4 Solute-Solvent Interactions and Chemical Interactions and Bonds

Solute particles are able to affect some of the solvent's properties (eg: partial vapor pressure, boiling point, freezing point, etc). These occur because dissolving requires solute-solvent interactions to

overcome the solute-solute and solvent-solvent interactions.

The four types of non-covalent chemical interactions are **polar dipole-dipole interactions**, **hydrogen bonding**, **Van Der Waals dispersion forces**, and **ionic interactions**. Polar dipole-dipole interactions refers to the affinity that molecular dipoles, molecules which have an uneven distribution of charge, have for one another. Hydrogen bonding refers to especially strong dipole-dipole interactions between hydrogens connected to strongly electronegative atoms and other strongly electronegative atoms which typically have lone pairs on them. The hydrogen bond is usually depicted as an interaction between the lone pair (hydrogen bond acceptor) and the hydrogen (hydrogen bond donor). Van Der Waals dispersion forces refer to attraction with nonpolar molecules due to charge imbalances arising by chance in **nonpolar** molecules. These explain octane's higher boiling point relative to methane and iodine's nongaseous state - the molecules are larger and more likely to become dipole-like. Ionic interactions, oftentimes called **salt bridges** in biological environments, are the affinity between cations and anions.

A general rule is that "like dissolves like", meaning that polar solvents will dissolve polar solutes and nonpolar solvents will dissolve nonpolar solutes, but the two do not mix.

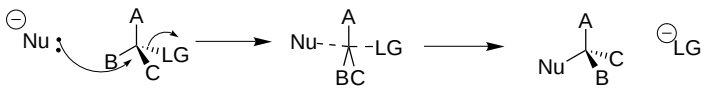
The **hydrophobic effect** is used to describe how many biological macromolecules fold in *aqueous* environments. Because polar water molecules do not associate with hydrophobic groups, to dissolve a hydrophobic group in water requires that the water molecules arrange themselves in a lattice, held together by hydrogen bonding, surrounding the hydrophobic group. Thus, dissolving nonpolar compounds in water has a weakly negative ΔH (releases energy) but a negative ΔS (decrease in entropy), making ΔG temperature-dependent. The decrease in entropy is due to the water molecules forming the tight, rigid structure. The hydrophobic effect refers to the release in free energy from allowing nonpolar groups to face the inside of the macromolecule (ie a protein folding) which thus eliminates the need for the crystalline water structure (and removes the ΔS cost) and hence is a strong driving force for folding.

These interactions also are the basis for not only macromolecular folding, but also for interactions between enzyme and substrate and between receptors and ligands. In each case, there must be a driving force as well as complementarity between the surfaces. The driving force for most interactions is typically the hydrophobic effect. Complementarity refers to the "match" between substrate and enzyme or ligand and receptor with regards to positioning and type of interactions (i.e. two hydrogen bond donors must be paired with two hydrogen bond acceptors). It should be noted that hydrogen bonding, in an aqueous environment, is NOT a good driving force for macromolecule folding or DNA base-pairing in general because hydrogen bond acceptors and donors are all satisfied by hydrogen bonding to water molecules in the unfolded state. Hydrogen bonding and salt bridges are, however, important for specificity and complementarity.

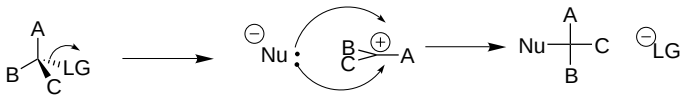
5 Chemical Reaction Mechanisms

(consult an organic chemistry textbook if more is needed)

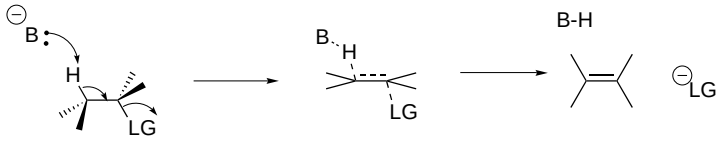
S_N2 reaction is a direct nucleophilic substitution where a **nucleophile** (an atom with a lone pair that wishes to donate it) attacks at a position with a good **leaving group** (an atom who can take on a negative electric charge well). Mechanistically, this causes the leaving group to leave and replaces the leaving group with the nucleophile as shown. The attack and leaving is simultaneous and occurs through the shown transition state causing the final conformation to be an exact reversion of the original stereochemistry. S_N2 reactions work best with good leaving groups and good nucleophiles, but also when the nucleophile is not sterically hindered from attack (ie if A and B and C are small and oftentimes hydrogen).



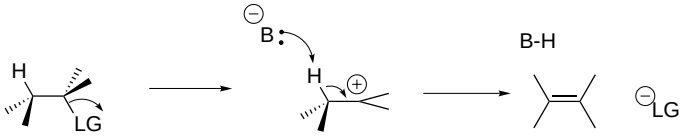
S_N1 reactions occur when the leaving group first leaves, creating a flat **carbocation** intermediate which can then be attacked by the nucleophile from either direction, creating a final product that is indeterminate in terms of stereochemistry. S_N1 reactions work best with good leaving groups (do not need a particularly good nucleophile) and when the carbocation is stabilized, preferably when A, B, and C are carbon chains or larger.



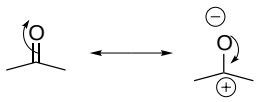
E$_2$ eliminations occur when a base deprotonates at a position exactly *anti* to a good leaving group such that a flat transition state resembling the final double bond is present. E$_2$'s work best when the proper stereochemistry is present and a strong enough base is near.



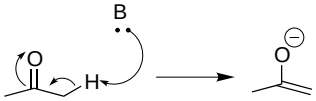
E$_1$ elimination occurs when the leaving group first leaves, leaving a flat carbocation intermediate. Deprotonation by a base pushes electrons into the carbocation, completing the elimination. E$_1$'s thus work best when there are stabilizing groups around the carbocation and when the leaving group is very good (the base does not need to be very strong).



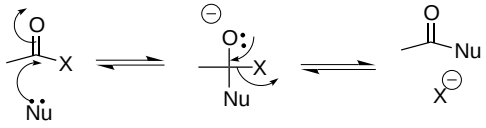
Much of biological chemistry has to do with carbonyl chemistry. There are three principles that one needs to be kept in mind when dealing with carbonyls. First, carbonyls are **electrophiles** at the carbon group which is very oxidized because of its two bonds to oxygen.



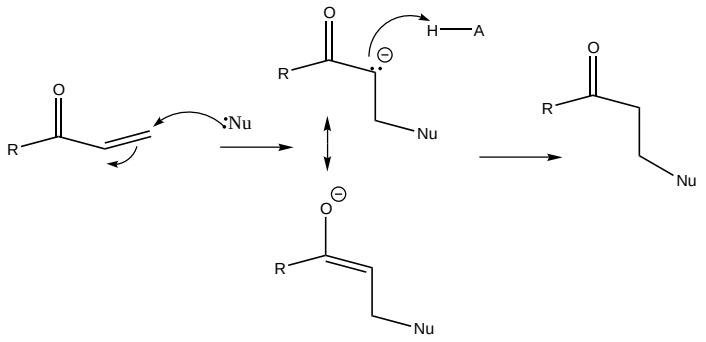
Secondly, the protons alpha to a carbonyl are fairly acidic because the electrons can flow into the electrophilic carbonyl.



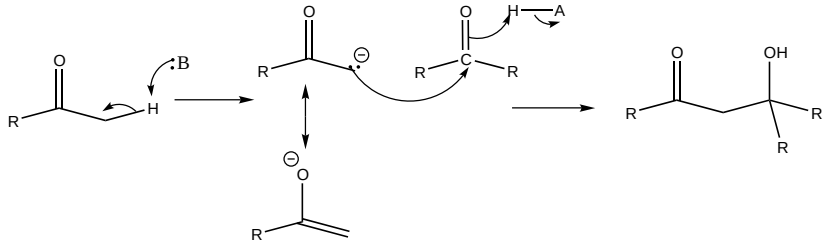
Thirdly, nucleophilic attacks on carbonyls go through a tetrahedral intermediate (the stereochemistry depends on where the nucleophile attacks from) which collapses to recreate the carbonyl structure and kick off a leaving group.



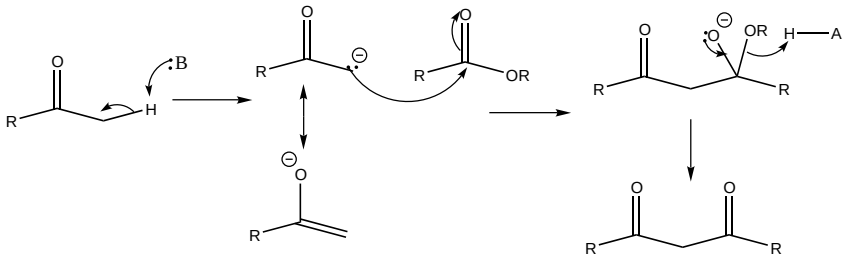
Carbonyls undergo a variety of condensation reactions. The acidity of the α-proton allows for three particular condensation reactions which occur repeatedly in biology. One is the **Michael addition** where a carbonyl conjugated to a double bond can convey the carbonyl's electrophilicity to the position β to the carbonyl:



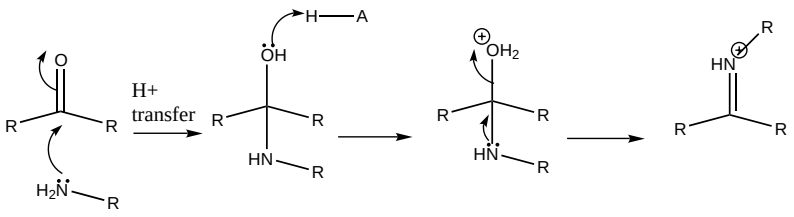
The **aldol condensation** employs the acidity of the α-proton to convert the α position into a nucleophile which can then attack other carbonyl's, generating carbonyl's with alcohol groups at the β position. This is a key step in the gluconeogenesis pathway and its reverse, the **retro-aldol** is the key step in glycolysis:



The **Claisen condensation** employs a similar mechanism to convert the position α to a carbonyl into a nucleophile and displaces the alcohol group of an ester, giving a β-ketoester. This is a key step in fatty acid synthesis, and the reverse **retro-Claisen** is the key step in fatty acid metabolism:



Another key condensation is the replacement of a carbonyl with an imine group, oftentimes used by enzymes to covalently attach a carbonyl to a lysine residue. The aldolase responsible for catalyzing the key aldol and retro-aldol condensations in glycolysis and gluconeogensis operate in this fashion. Effectively, a series of proton transfers converts the carbonyl oxygen into a water molecule which can act as a leaving group, allowing the nitrogen to create the imine linkage, which is oftentimes called a **Schiff base**, and is notably more electrophilic than a carbonyl because of the positive charge:



Part II
Catalysis and Binding

6 Enzyme Reaction Mechanisms and Kinetics

6.1 How Enzymes Work

Every different enzyme employs a different reaction mechanism. However, they all work by lowering the **activation energy** (free energy difference between the transition state and substrate which for quantum mechanical reasons allows a reaction to progress faster):

1. Enzymes position reactive amino acids to "activate" the substrate (raising the free energy of the substrate, making it easier to proceed through the reaction)
2. Enzymes bind the transition state more tightly than the product or reactant states, lowering the transition state energy

Alternative approaches (not all are mutually exclusive with each other or with the transition state stabilization theory described here, the bold terms are important to know though) to understanding how enzymes work are:

1. **Induced Fit Model** which claims that the binding of the correct substrate triggers a change in the structure of the enzyme that brings catalytic groups into the right position to catalyze the reaction
2. **Nonproductive Binding** claims that only the correct substrates bind in one, specific, and productive orientation to an enzyme for catalysis, while incorrect substrates bind in multiple orientations and thus cannot react.
3. **Strain and Distortion Model** argues that the enzyme active site strains the bonds and atoms of the substrate molecule such that it assumes the transition state

6.2 Enzyme Limitations

It should be noted that enzymes have certain limitations. They do not affect the final equilibrium between products and reactants (which is dependent on ΔG which is dependent only on the initial and final states rather than the process), are thus oftentimes reversible, are oftentimes energy-dependent, oftentimes require cofactors, and cannot process reactions which are not energetically favorable.

6.3 Enzyme Kinetics

Enzyme reactions can be summarized as follows:

$$S + E \rightleftharpoons ES \rightarrow E + P$$

The velocity of the reaction, defined as $\frac{d(P)}{dt}$, thus shows first-order kinetics with respect to the enzyme-substrate complex with the constant k_{cat} (also called the **turnover rate**:

$$v = \frac{d(P)}{dt} = k_{cat}(ES)$$

The rate of appearance of the enzyme-substrate complex, then, is given by the rate of the forward reaction (with kinetic constant k_1) subtracted by the rate of the reverse reaction (with kinetic constant k_{-1}) and the rate of enzymatic catalysis (as it also consumes the enzyme-substrate complex). Thus:

$$\frac{d(ES)}{dt} = k_1(E)(S) - k_{-1}(ES) - k_{cat}(ES) \tag{9}$$

At steady-state equilibrium, this rate is 0. One can thus define an equilibrium value K_m:

$$K_m = \frac{k_{cat} + k_{-1}}{k_1} = \frac{(E)(S)}{(ES)} \tag{10}$$

As (E) and (ES) are constantly varying, we seek a quantity which is constant which will allow us to calculate the reaction rate under the presence of enzyme. Because the total enzyme is constant, an expression for (ES) can be found:

$$E_{tot} = (E) + (ES) = (ES)\left(\frac{K_m}{(S)} + 1\right) \Rightarrow (ES) = \frac{(E_{tot})}{1 + \frac{K_m}{(S)}}$$

Substituting into the original equation gives the **Michaelis-Menten equation** for enzyme kinetics:

$$\frac{d(P)}{dt} = \frac{k_{cat}(E_{tot})(S)}{(S) + K_m} \tag{11}$$

As the maximum possible enzyme velocity is when all the enzyme is complexed (hence when $(ES) = (E_{tot})$) equation [11] can also be written:

$$\frac{d(P)}{dt} = \frac{v_{max}(S)}{(S) + K_m}$$

Hence, enzyme kinetics are determined primarily by two parameters: K_m and k_{cat}.

6.4 Analysis of Enzyme Kinetics

There are three plots which are always associated with Michaelis-Menten Kinetics. The first are called **Lineweaver-Burk** plots which show the relationship between the reciprocal of the velocity and then reciprocal of the substrate concentration:

$$\begin{aligned} v &= \frac{v_{max}(S)}{(S) + K_m} \\ \frac{1}{v} &= \frac{1}{v_{max}} + \frac{K_m}{v_{max}}\left(\frac{1}{(S)}\right) \end{aligned} \tag{12}$$

which is in other words a line with vertical intercept of $1/v_{max}$ and a slope of K_m/v_{max}.

The second are called **Dixon** plots which show $\frac{(S)}{v}$ versus (S) (which is essentially the Lineweaver-Burk relationship multiplied by (S) on both sides):

$$\frac{(S)}{v} = \frac{1}{v_{max}}(S) + \frac{K_m}{v_{max}} \tag{13}$$

which is a line with vertical intercept of $\frac{K_m}{v_{max}}$ and a slope of $\frac{1}{v_{max}}$.

The last is the **Eadie-Hofstee** plot which plots v versus $v/(S)$ (which is essentially the Lineweaver-Burk relationship multiplied by $v_{max}v$):

$$v = -K_m\left(\frac{v}{(S)}\right) + v_{max} \tag{14}$$

which is a line with vertical intercept v_{max} and a slope of $-K_m$

6.5 Competition and Inhibition

If an enzyme has two substrates, it will react with both, thus slowing the rate of the reaction with one specific substrate because of partial saturation with the other substrate. The relative reaction rate, then, is determined by the specificity of the enzyme as well as the substrate concentrations. Recall that at equilibrium:

$$v(t) = \frac{k_{cat}(E)_t(S)}{K_m} \tag{15}$$

Thus, if $[A]$ and $[B]$ are two substrate for the same enzyme, then because the free enzyme $(E)_t$ is the same for both enzymes, the relative rate of reaction is:

$$\frac{v_A}{v_B} = \frac{\left(\frac{k_{cat}}{K_m}\right)_A [A]}{\left(\frac{k_{cat}}{K_m}\right)_B [B]} \tag{16}$$

and thus $\frac{k_{cat}}{K_m}$ is the bspecificity constant of an enzyme for a particular substrate.

Competitive inhibitors (ie most small-molecule drugs), on the other hand, do not get reacted with and simply bind the enzyme, preventing the enzyme from binding the substrate. This binding shows a binding parameter K_I such that:

$$K_I = \frac{[E][I]}{[EI]} \tag{17}$$

Therefore the total enzyme concentration is actually the sum of the free enzyme, enzyme bound to substrate, and enzyme bound to inhibitor:

$$E_{tot} = (E) + (ES) + (EI) = (E)\left(1 + \frac{(I)}{K_I}\right) + (ES) \tag{18}$$

Using [18] and re-deriving the Michaelis-Menten relationship gives the formula for the rate of an enzyme-catalyzed reaction subject to a competitive inhibitor:

$$v = \frac{k_{cat}(E_{tot})(S)}{(S) + K_m\left(1 + \frac{(I)}{K_I}\right)} \tag{19}$$

This new relationship has a modified K_m of:

$$K'_m = K_m\left(1 + \frac{(I)}{K_I}\right) \tag{20}$$

It should be noted here that the inhibition can be made irrelevant by sufficiently large levels of substrate. On a Lineweaver-Burke plot, this inhibition is represented by a rotation of the curve, with increasing inhibitors showing an increased rotational effect. Because a competitive inhibitor does not affect the maximum velocity, the vertical intercept ($\frac{1}{v_{max}}$) stays at the same position while the horizontal intercept representing K_m shifts.

Inhibitors that are not competitive cannot be overcome by large levels of substrate and can happen in three ways:

1. An irreversible modification to the enzyme active site (e.g. 5-FU binding to thymidylate synthase)
2. Binding of an inhibitor to the enzyme at a site other than the enzyme active site (e.g. NNRTIs [as opposed to NRTIs])
3. Binding of the inhibitor to the enzyme-substrate complex

This is where the nomenclature becomes difficult and confusing. There are two types of nomenclature employed. The first and most used refers to three broad types of inhibition: competitive, noncompetitive, and uncompetitive. In this system, the term **uncompetitive inhibitor** refers to inhibitors that are not competitive which are able bind to the enzyme-substrate complex and are thus only able to exert influence when substrate concentrations are high (when the enzyme-substrate complex is in abundance), as compared to competitive inhibitors which cannot exert influence when substrate levels are high and can only do so when the concentrations are low. In a Lineweaver-Burke plot, these curves are shown as lines which have been displaced vertically in the presence of inhibitor. In this system, the term **noncompetitive inhibitor** refers to inhibitors that are able to bind both the free enzyme and the enzyme-substrate complex and thus exert multiple effects on the Lineweaver-Burke plot.

In the second system of nomenclature, inhibitors are designated according to which constant they affect. Competitive inhibitors affect K_m (refer to [20]), while noncompetitive inhibitors affect k_{cat}, and **mixed inhibitors** affect both. While this is not as oftentimes used, its important to remember that enzyme kinetics can be almost exclusively understood from their ability to impact the parameters of the Michaelis-Menten relationship, which is why this is included. Therefore, in dealing with enzyme kinetics calculations regarding inhibitors that act in a mode different from simple competitive inhibition, one only has to consider the effects on the Michaelis-Menten equation (e.g. if an irreversible inhibitor knocks out 10% of an enzyme, then $(E)_{tot}$ is reduced by 10%).

7 Ligand-Protein Interactions

7.1 Mathematics of Complex Formation

The value which determines complex formation equilibria is K_D:

$$K_D = \frac{(R)(L)}{(RL)} \tag{21}$$

where (R) is the equilibrium level of free receptor, (L) of free ligand, and (RL) of receptor-ligand complex. From [8], we know, therefore, that a smaller K_D means a tighter and more energetically favorable interaction. Using the fact that the total receptor $((R) + (RL))$ level is constant, then one can deduce:

$$\begin{aligned} R_{tot} &= (R) + (RL) \\ &= K_D \frac{(RL)}{(L)} + (RL) \\ (RL) &= \frac{R_{tot} L}{K_D + L} \\ \frac{(RL)}{R_{tot}} &= \frac{L}{K_D + L} \end{aligned} \tag{22}$$

If, as is often the case, that the total ligand concentration is much greater than the total receptor concentration, then the concentration of free ligand L can be approximated by the total ligand concentration L_{tot}.

7.2 Scatchard Equation

[22] is oftentimes plotted as a **Scatchard diagram** which plots the fractional saturation divided by the ligand concentration against the fractional saturation:

$$\begin{aligned} \frac{(RL)}{R_{tot}}(K_D + L) &= L \\ K_D \frac{(RL)}{R_{tot}} &= L\left(1 - \frac{(RL)}{R_{tot}}\right) \\ \frac{\frac{(RL)}{R_{tot}}}{L} &= \frac{1}{K_D} - \frac{1}{K_D}\left(\frac{(RL)}{R_{tot}}\right) \end{aligned} \tag{23}$$

which gives a linear relationship where the vertical intercept is the reciprocal of the dissociation constant and the slope is the negative of the reciprocal of the dissociation constant.

[23] was derived for the case of a single receptor. It turns out, though, if there is a protein with multiple identical receptors for the same ligand, and if each of these receptors acts independently of each other (binding at one site does not affect the affinity at another site), then the equation can still be used if one understands that $\frac{(RL)}{R_{tot}}$ is actually ranges from 0 to N, where N is the total number of receptor sites on the molecule. In a Scatchard diagram, then, the vertical intercept would thus be $\frac{N}{K_D}$, the horizontal intercept would be N and the slope would be $-\frac{1}{K_D}$ where K_D is the dissociation constant for a single receptor:

$$\frac{\frac{(RL)}{R_{tot}}}{L} = \frac{1}{K_D}\left(N - \frac{(RL)}{R_{tot}}\right) \tag{24}$$

It should be noted that many texts write [24] as:

$$\frac{y}{L} = K_A(N - y) \tag{25}$$

where y is the fractional saturation (measured from 0 to N) and K_A is the association constant, which is simply the reciprocal of K_D.

In some cases, the Scatchard Equation, is written as a relationship between r, the number of bound ligand molecules per monomer, and n, the number of binding sites per monomer:

$$\frac{r}{L} = \frac{1}{K_D}(n - r) \tag{26}$$

7.3 Hemoglobin and Myoglobin and the Hill Equation

A comparison between hemoglobin and myoglobin binding of oxygen is the standard example given when dealing with complex receptor-ligand interaction systems that deal with **cooperativity** or **allostery**. Both terms refer to binding relationships where the receptors are not independent. Cooperativity refers to what happens when binding of one receptor with ligand alters the affinity of the other receptors, while allostery refers to what happens when the binding of the receptors with a non-ligand at an allosteric position affects their affinity for substrate.

Cooperativity can be either positive or negative with the former referring to cases where binding of ligand increases the affinity of the other receptors and the latter refers to a decrease. Cooperativity can be observed through a sigmoidal relationship between fraction of receptors bound and concentration of ligand and by a non-linear relationship when a Scatchard diagram is made. If one takes [25], one can rearrange it:

$$\begin{aligned} \frac{y/N}{L} &= K_A(1 - y/N) \\ \frac{y/N}{1 - y/N} = \frac{f}{1 - f} &= K_A L \end{aligned} \tag{27}$$

[27] shows a linear relationship between the ratio of the fractional saturation f (recall that y ranges from 0 to N and that the true fractional saturation is thus y/N) to $1 - f$ goes linearly with L. Cooperativity relationships do not show this linearity and the degree of linearity can be quantified using the **Hill Coefficient** n such that:

$$\frac{f}{1 - f} = K_A L^n \tag{28}$$

where n ranges from 1 (in the case of complete non-cooperativity as shown in the Scatchard relationship in [27]) to N (in the case of all-or-nothing cooperativity, where if one receptor is bound all the other receptors are forced to be bound as well). **Hill plots** are often made to assess n at different ligand concentrations. This is best visualized by taking the logarithm of both sides:

$$\log \frac{f}{1 - f} = n \log L + \log K_A$$

which results in a graph where the slope of the curve at any point is the Hill coefficient, or degree of cooperativity, at that ligand concentration/degree of saturation.

In the comparison between Hemoglobin and Myoglobin one finds that Myoglobin (which is in muscle cells) binds oxygen with a Hill coefficient of 1 (no cooperativity, as there is only one binding site) and Hemoglobin, which has a lower affinity for oxygen than myoglobin, has several Hill coefficients with the physiologically relevant one being around 2.8. This is evolutionarily advantageous as it means that Hemoglobin molecules will quickly and cooperatively release oxygen when oxygen concentrations are low (i.e. in the tissues) and quickly and cooperatively bind oxygen when oxygen concentrations are high (i.e. in the lungs). This sigmoidal relationship allows oxygen to be quickly and efficiently delivered from the lungs to the body.

There are several mechanisms by which this is postulated to occur. The most common one is the **Monod-Wyman-Changeux** model (or MWC model) which postulates that hemoglobin's four

oxygen binding sites exist in two forms, a relaxed form, and a tense form with the former binding oxygen more tightly than the latter. What happens is that while the tense form predominates in the absence of ligand, one receptor's binding of ligand shifts the equilibrium of the other binding sites towards the relaxed form, making binding a cooperative event.

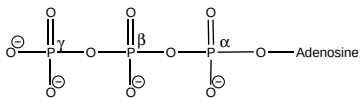
Part III
Major Metabolic Pathways and Bioenergetics

8 Organic Chemistry of Metabolic Biochemistry

8.1 Metal Cations

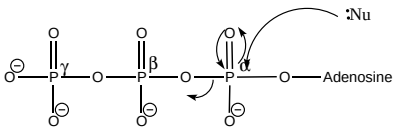
The purpose of Magnesium and Zinc and Calcium ions in most metabolic pathways is to stabilize the buildup of negative charges (ie in kinases, metal cations interact with the negative phophate groups) or to increase the acidity (lower the pK_a) of a proton (ie in alcohol dehydrogenases).

8.2 XTP

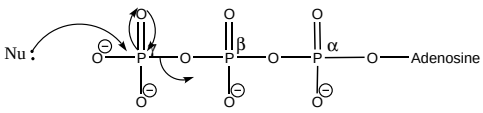


ATP

ATP is the most commonly used, although GTP, CTP, and TTP are also used occasionally. These cofactors function as sources of good leaving groups which thus lower the free energy ΔG of a reaction, making them sources of energy, or as sources of negatively charged phosphates. The chemistry for reaction with an XTP molecule (although ATP is shown, all the chemistry is essentially the same) occurs in two ways. The first means is by attacking the α-phosphate (the one closest to the nucleotide) to displace a **pyrophosphate** ion, which is a good leaving group that spontaneously hydrolyzes into two phosphates (hence releasing more free energy):



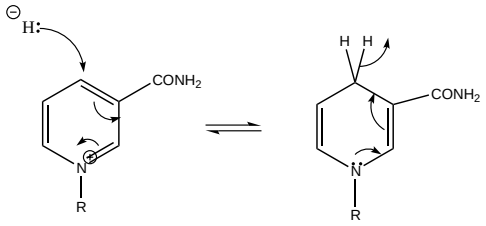
The second means is by attacking the γ-phosphate (the one furthest from the nucleotide) to displace ADP (with only two phosphates rather than one) as a leaving group, leaving the nucleophile phosphorylated:



8.3 NAD(P)H

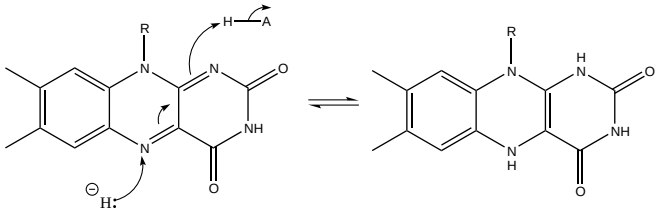
Nicotinamide adenine dinucleotide (NAD) is a cofactor involved in redox reactions which is derived from the vitamin niacin. It has a nicotinamide ring which can act as a hydride acceptor or donor depending on the charge and double bond structure on the ring. When the amine nitrogen is positively charged, the cofactor is called NAD(P)$^+$ and acts as a hydride acceptor (and hence an oxidizing agent). When the amine nitrogen is neutral, the cofactor is called NAD(P)H and acts as a hydride donor (and hence a reducing agent). The only difference between NADPH and NADH is in an extra phosphate group in the R group. The chirality of the ring (the amide substituent gives the

ring a handed-ness) and the fact that all enzyme reactions take place within a complicated three-dimensional structure means that enzyme-catalyzed reactions involving NAD(P)H or NAD(P)+ abstract a specific proton (with exclusively pro-R or pro-S stereochemistry on the substrate) and add it to one specific side of the nicotinamide ring.



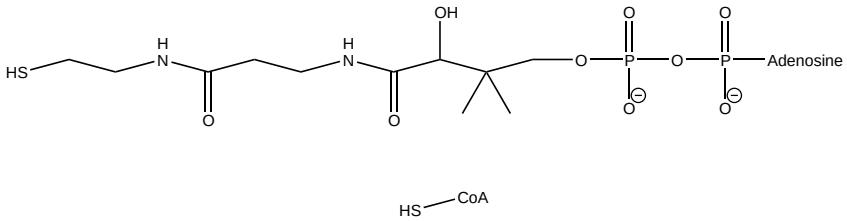
8.4 FAD

Flavin Adenine Dinucleotide (FAD) is a cofactor also involved in redox reactions which is derived from the vitamin riboflavin. It is primarily used as an oxidizer. Its specific mechanism of action depends a great deal on the substrate, as FAD can participate in hydride transfers and single-electron transfers/radical chemistry (ie it can be used to reduce oxygen to hydrogen peroxide).

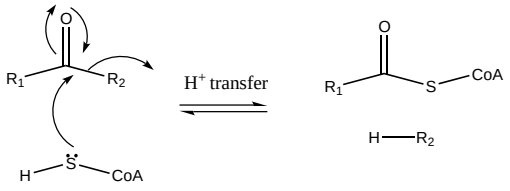


8.5 Coenzyme A

Coenzyme A (CoA) is used to activate acyl substrates to transfer them. It effectively holds acyl groups through the use of a a phosphopantetheine arm which ends with a thiol group which is capable of acting as a nucleophile to attack carbonyl's. The arm is oftentimes abbreviated "CoA" as shown below (where the two structures represent the same thing):

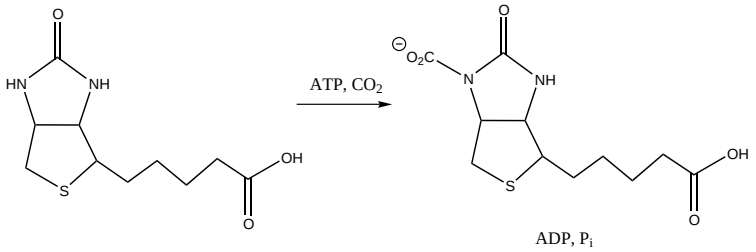


Enzymes oftentimes attack the α-phosphate (closest to the nucleotide) with a nucelophilic group (ie serine), displacing AMP as a leaving group and conjugating the phosphopantetheine arm to the enzyme. CoA helps catalyze reactions which replace substituents on carbonyl groups:



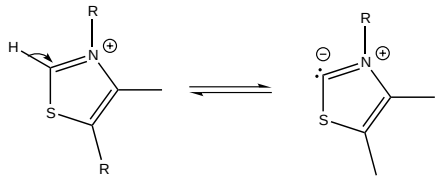
8.6 Biotin

Biotin, a B vitamin derivative, is employed in carboxylation reactions. Although its precise mechanism is not well understood, it is used to hold carbon dioxide such that it can then be transferred onto another molecule.

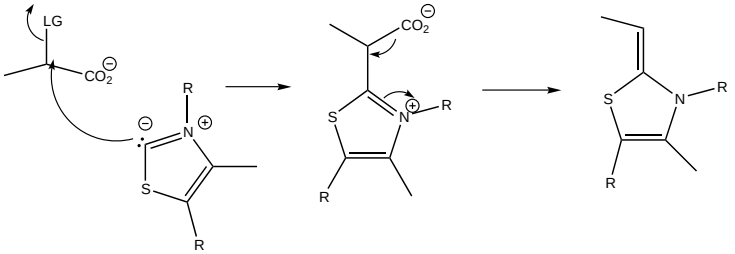


8.7 TPP

Thiamin Diphosphate (TPP) is a cofactor involved in a number of reactions. It is derived from the B-vitamin thiamine and is most useful because of its **ylid** functional group. The positioning of a carbon between two heteroatoms (nitrogen and sulfur) lowers the pK_a of the carbon's proton.

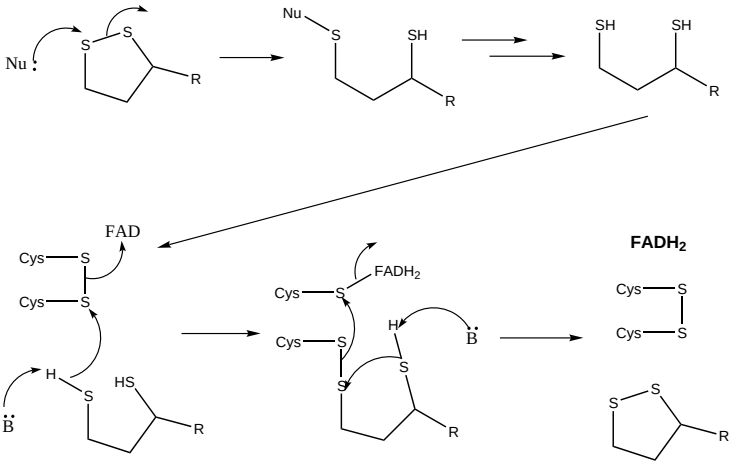


This enables it to act as a nucleophile. which is connected to an iminium ion which can thus function as an electron sink.



8.8 Lipoamide

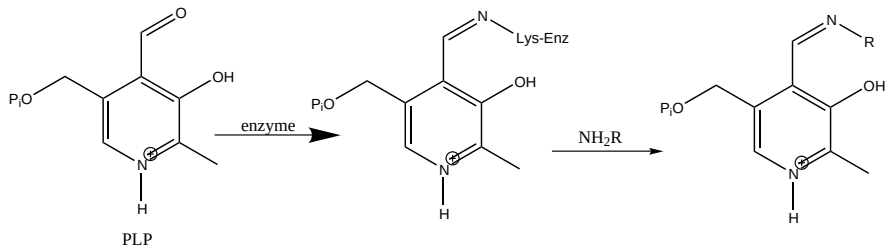
Lipoamide, a derivative of lipoic acid, has an electrophilic disulfide bridge. Upon breaking of the ring, FADH2 is required to recreate the original lipoamide structure.



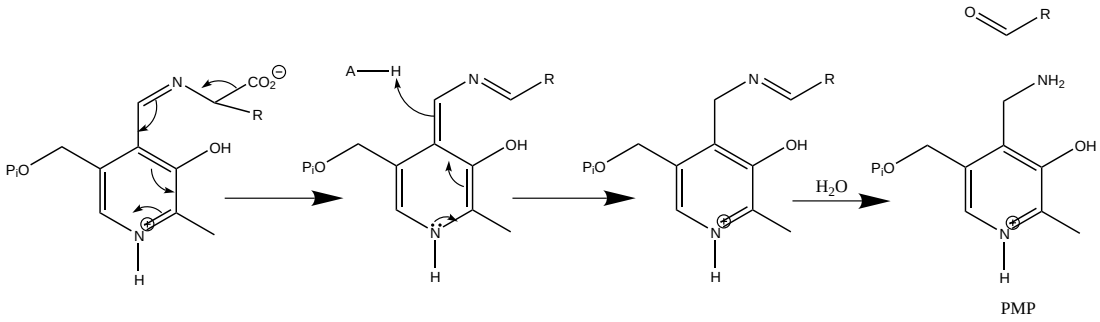
8.9 PLP

PLP, or **pyridoxal phosphate**, a cofactor derived from pyridoxine (vitamin B_6) is used widely in amino acid metabolism. Its aldehyde functionality serves as an attachment point for various

lysine-containing enzymes, creating a Schiff base link. PLP-mediated steps always begin by a transamination reaction in which a substrate amine group becomes covalently linked to the cofactor.



As PLP possesses an electrophilic nitrogen connected to a conjugated pi system, PLP acts as an electron sink, allowing various reactions to occur adjacent to or on the amine group, sometimes giving pyridoxamine phosphate (PMP) as a by-product (which can then be converted back to PLP by transamination with an α-keto acid, and disruption of the Schiff base linkage by hydrolysis or transamination).



8.10 Miscellaneous

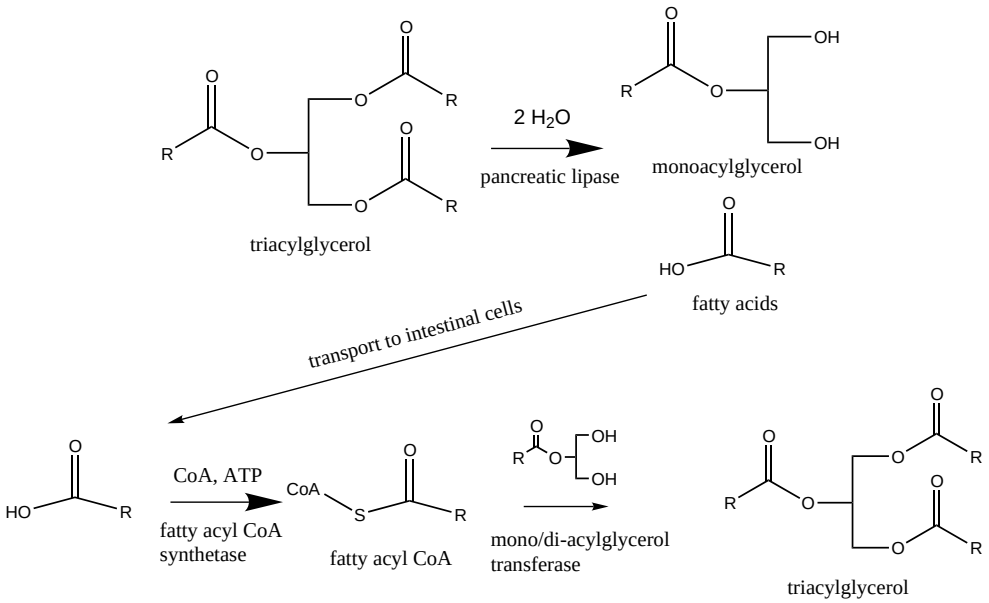
There are a number of cofactors which chelate a metal ion to help mediate radical chemistry. **Coenzyme $\mathbf{B}_{12}$** is a cobalt containing cofactor which helps to mediate radical rearrangement reactions. It mediates this by having a readily homolytically cleaved bond to a methyl-deoxyadenosine moiety. **Heme** is an iron containing cofactor which helps bind oxygen species to mediate oxygenation/oxidation reactions.

Tetrahydrofolate (THF) and its derivative 5,10-methylenetetrahydrofolate acts as a means to transfer 1-carbon groups.

9 Lipid Metabolism

9.1 Digestion

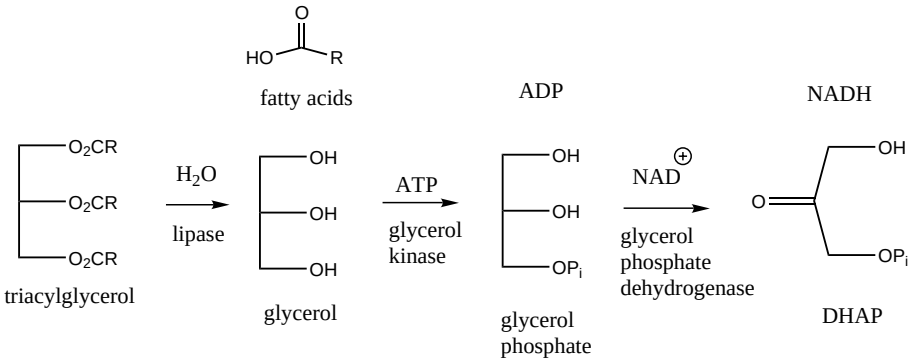
Because dietary fats are insoluble in water, they are emulsified by bile acid salts (such as taurocholate and glycocholate) which act as detergents. Once emulsified, they are hydrolyzed by pancreatic lipase at the C1 and C3 positions to yield a mono-acylglycerol and two fatty acids. These are then transported to the cells lining the intestinal wall and recombined to produce triacylglycerols. This step is catalyzed by the use of ATP and CoA. The use of ATP would make this enzyme a **synthetase** as opposed to a **synthase**.



Once reprocessed into triacylglycerols, they combine with lipoprotein assemblies called chlyomicrons which transport them through the bloodstream to target tissues.

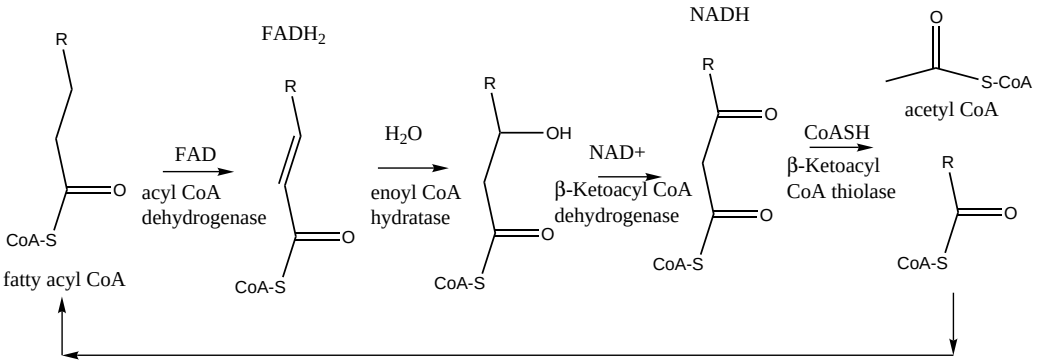
9.2 Glycerol Catabolism

Once at the target tissues, the triacylglycerols are again broken down by a lipase into three fatty acids and glycerol. The fatty acids are processed in the mitochondria while the glycerol is sent to the liver for metabolism. Glycerol is then phosphorylated at the pro-R alcohol by glycerol kinase, which uses ATP (source of phosphate) and Mg ions as cofactors. The glycerol phosphate is then oxidized by glycerol phosphate dehydrogenase, which uses NAD+ as a cofactor, to **dihydroxyacetone phosphate (DHAP)** which can enter the carbohydrate metabolic pathway.



9.3 Fatty Acid Catabolism

The fatty acids in the mitochondria are catabolized by a repetitive sequence called the **β-oxidation pathway** which produces **Acetyl CoA** molecules which can enter the carbohydrate metabolic pathway. Fatty acids are first converted to fatty acyl CoA's. These are then loaded onto **carnitine** which shuttles the acyl CoA's into the mitochondria for oxidation (the **carnitine shuttle**). In the mitochondria, the acyl CoA's have a double bond introduced beta to the acyl carbonyl by an acyl-CoA dehydrogenase which employs FAD to introduce a trans double-bond. Water is then syn-added by enoyl-CoA hydratase to the double bond. This introduced alcohol is then oxidized by a hydroxyacyl-CoA dehydrogenase using NAD+ as a cofactor to produce β-Ketoacyl CoA. β-Ketoacyl CoA is the product of a Claisen condensation (two carbonyls beta to each other). β-Ketoacyl CoA thiolase which performs a retro-Claisen, removing the terminal CoA-linked carbons to produce an acetyl CoA molecule and also catalyzes the loading of the remainder of the fatty acid onto another Co-enzyme A which then goes through the same cycle again.

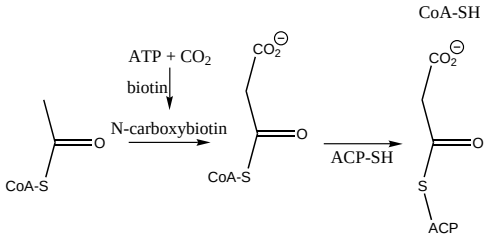


9.4 Fatty Acid Biosynthesis

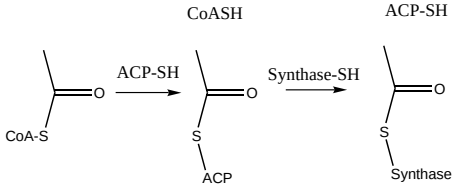
Fatty acids typically have an even number of carbon atoms because they are synthesized from two carbon-acetyl CoA precursors. In order for both processes to be energetically favorable, fatty acid biosynthesis occurs in the cytoplasm and is slightly different from β-oxidation which occurs in the mitochondria. Fatty acid biosynthesis employs NADPH to reduce the double bond while β-oxidation employs FAD to introduce one. In bacteria, fatty acid biosynthesis is catalyzed by multiple enzymes which handle each step, while in eukaryotes, a large, multienzyme complex catalyzes all the steps in the pathway.

The acetyl CoA molecules for fatty acid synthesis are produced in the mitochondria. The transport of acetyl CoA to the cytosol where fatty acid synthesis occurs is mediated by oxaloacetate (OAA). Acetyl CoA undergoes a condensation with OAA to form citrate which can then leave the mitochondria. In the cytosol, an ATP-dependent citrate lyase liberates OAA and acetyl CoA. The OAA then undergoes a series of transformations to become pyruvate (where it is reduced by NADH and oxidized by NADPH, thus converting an NADH to NADPH) and then re-enters the mitochondria where it is converted in another ATP-dependent step into OAA (thus moving one equivalent of acetyl CoA requires two equivalents of ATP).

Acetyl CoA in the cytosol has two fates. The first involves carboxylation by acetyl-CoA carboxylase which uses a bicarbonate ion, ATP, and the cofactor biotin to add a carboxylate group to acetyl CoA, producing malonyl CoA. Malonyl CoA is then loaded onto an acyl-carrier protein side chain. This protein then "holds" the growing fatty acid chain.

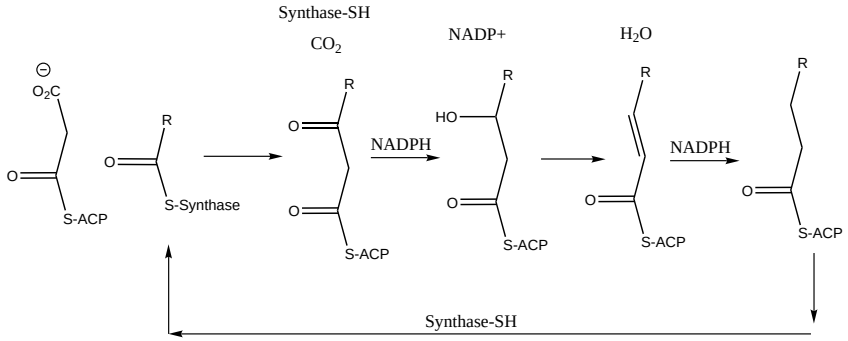


The second fate of acetyl CoA is to be added onto the growing fatty acid chain. The first step is the loading of acetyl CoA onto an **acyl carrier protein (ACP)**. This is then transferred onto another protein, acetyl synthase.



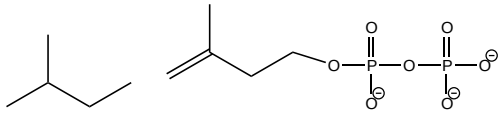
The enzyme complex then catalyzes a Claisen condensation between the malonyl CoA and the acetyl CoA by decarboxylating the fatty acid chain on the acyl-carrier protein. The resulting ketone is then reduced by NADPH to create an alcohol which is dehydrated to remove the hydroxyl group and introduce a double bond. This double bond is then reduced by NADPH.

The resultant fatty acid chain, now longer by two carbons and conjugated to the acyl-carrier protein is transferred onto acetyl synthase which allows it to undergo another Claisen condensation with a newly produced malonyl CoA, thus repeating the cycle and lengthening the fatty acid chain to whatever length is desired. It should be noted that the enzyme complex here can "choose" to leave out a reduction step or a dehydration step in order to produce modified fatty acid chains (ie unsaturated fatty acid).



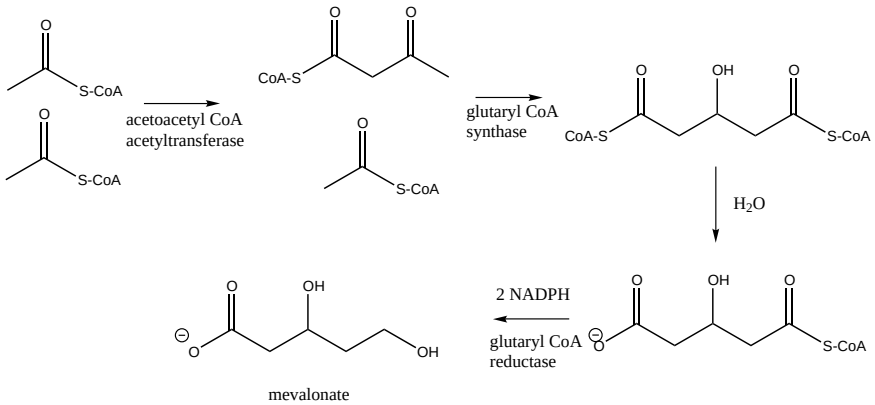
9.5 Terpenoid Biosynthesis

Many of the non-glycerol lipids are **terpenoids**, carbon structures built from **isoprene units**, 5-carbon structures which can be combined head-to-tail. Isopentenyl diphosphate is the functional form of the isoprene unit consisting of a double bond at the head and a diphosphate group towards the tail which allows for conjugation.

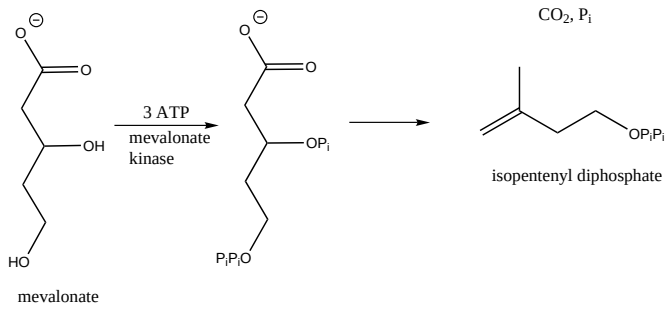


isoprene isopentenyl diphosphate

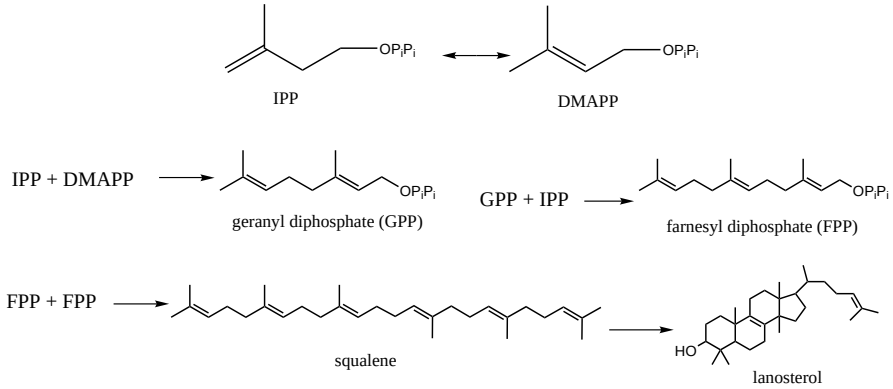
Isopentenyl diphosphate is synthesized primarily by the mevalonate pathway. Two molecules of acetyl CoA undergo a Claisen condensation catalyzed by acetoacetyl CoA acetyltransferase to form acetoacetyl CoA. This is then loaded onto an enzyme (displacing the CoA linkage) and then undergoes an aldol condensation with another molecule of acetyl CoA catalyzed by a glutaryl-CoA synthase. The enzyme linkage is then hydrolyzed and the resulting 3-hydroxy-3-methylglutaryl CoA is then reduced with two equivalents of NADPH to remove the final CoA linkage to yield mevalonate.



Mevalonate is then converted into **isopentenyl diphosphate (IPP)**. Three equivalents of ATP are employed to doubly phosphorylate the tail and to singly phosphorylate the middle alcohol. Decarboxylation then produces the double bond, displacing one of the phosphates.



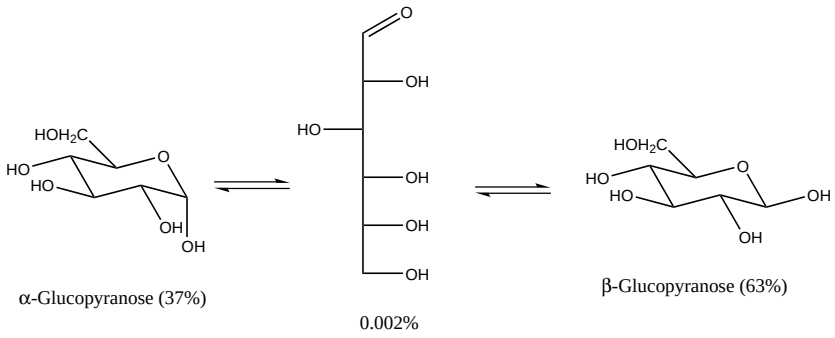
To couple multiple isoprene units, an isomerization reaction takes place converting IPP into **dimethylallyl diphosphate (DMAPP)** as catalyzed by IPP isomerase, requiring magnesium and zinc cations as cofactors. DMAPP then couples with IPP to give geranyl diphosphate which can then be combined with another molecule of IPP to form farnesyl diphosphate. Dimerization of farnesyl diphosphate yields squalene which is the precursor to a number of steroid molecules.



10 Carbohydrate Metabolism

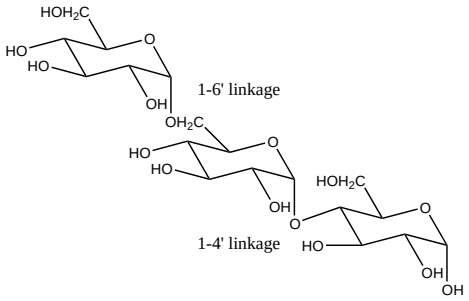
10.1 Structure

Carbohydrates exist in multiple forms. Glucose, for instance, exists in an α-form, a β-form (with α and β forms differing in the positioning of the hydroxyl group at position 1, α being axial and β being equatorial), and in the polyhydroxyaldehyde form with the last form being the most rare.



10.2 Digestion

Starches and other polysaccharides are linked between individual carbohydrate monomers by glycosidic linkages. Glucose polymers are typically linked by α-1-4' linkages (between the 1 position of an α form and the 4 position of the next) and α-1-6' linkage between the 1 position of an α form and the 6 position of the next.

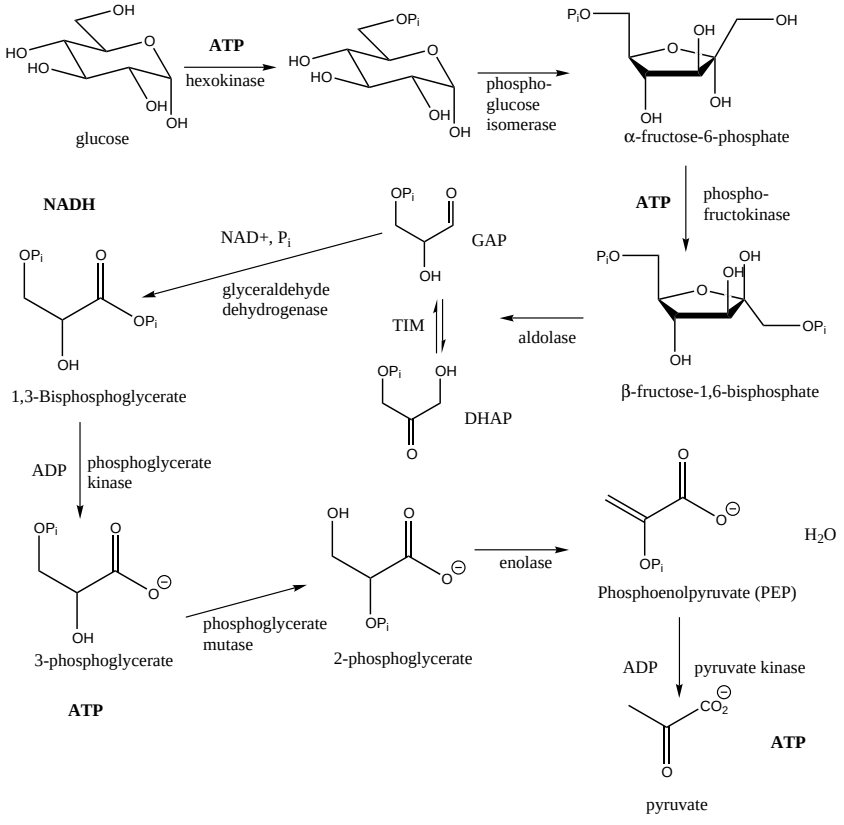


A salivary enzyme called amylase hydrolyzes the 1-4' linkages. Processing in the intestine with glycosidases yields glucose which enters the bloodstream.

10.3 Glycolysis

Glycolysis, also called the **Embden-Meyerhoff Pathway** is the process by which glucose is catabolized into **pyruvate**. The hormone **insulin** stimulates it, while the opposing hormone **glucagon** inhibits it. It is turned on in muscle cells but turned off in the liver by epinepherine. This is most likely due to regulation of phosphofructokinase.

First, glucose is phosphorylated by hexokinase. The resultant glucose-6-phosphate then undergoes a number of keto-enol tautomerizations catalyzed by glucose-6-phosphate isomerase to give α-fructose-6-phosphate. This is in turn phosphorylated and isomerized to give β-fructose-1,6-bisphosphate. This ring sugar is then opened and cleaved with a retro-aldol reaction into **glyceraldehyde 3-phosphate (GAP)** and **dihydroxyacetone phosphate (DHAP)**. The enzyme makes a Schiff base linkage with the substrate in order to increase the carbonyl's electrophilicity. GAP and DHAP rapidly interconvert via **triose phosphate isomerase (TIM)**. GAP is then oxidized and phosphorylated by GAP dehydrogenase to 1,3-bisphosphoglycerate. This is readily dephosphorylated by phosphoglycerate kinase, leaving behind 3-phosphoglycerate. This is then isomerized by phosphoglycerate mutase to 2-phosphoglycerate. A dehydration step catalyzed by an enolase creates **phosphoenolpyruvate (PEP)** which is dephosphorylated and tautomerized from enolpyruvate to pyruvate by pyruvate kinase.

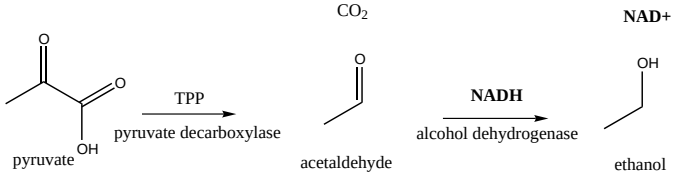


It should be noted here that glycolysis takes one molecule of glucose, consumes two molecules of ATP (to phosphorylate the glucose molecule), and produces two molecules of NADH and four molecules of ATP (two per pyruvate) and two molecules of pyruvate.

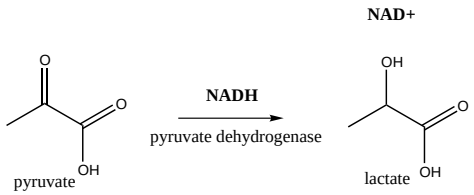
10.4 Transformations of Pyruvate

Pyruvate can have three fates.

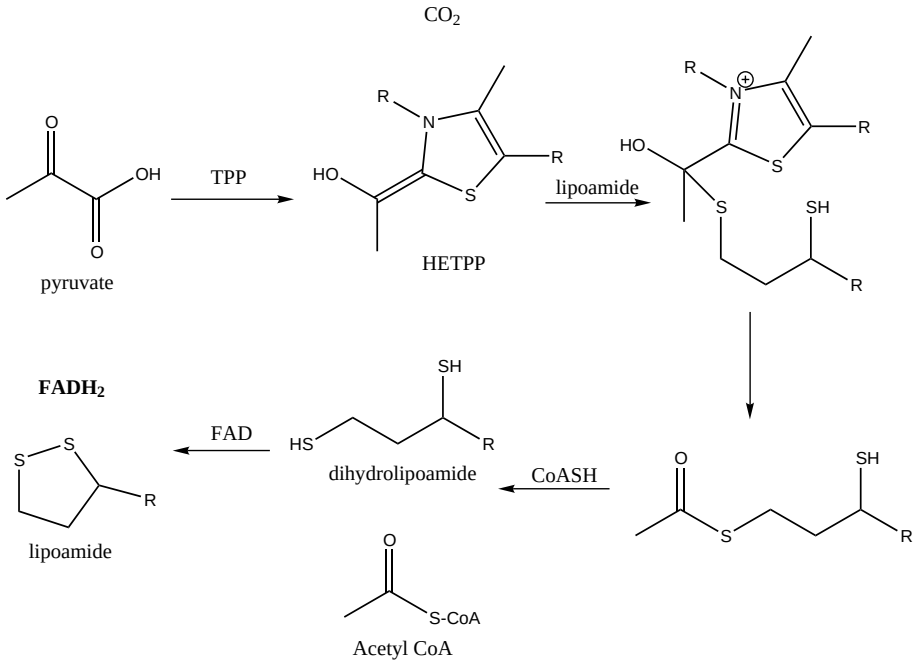
In yeast under anaerobic conditions, pyruvate is converted to ethanol and carbon dioxide. The process is a Thiamin diphosphate-dependent process catalyzed by pyruvate decarboxylase whereby pyruvate is decarboxylated and converted into acetaldehyde. Acetaldehyde is then reduced by an alcohol dehydrogenase into ethanol, thus recycling NADH back into NAD+ and allowing glycolysis to continue.



The second fate is to lactic acid which is produced under anaerobic conditions (ie during muscle activity), again the purpose of which is to regenerate NAD+ for glycolysis. This is carried out by lactate dehydrogenase which uses NADH to reduce pyruvate into lactate.



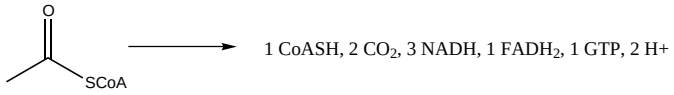
Under normal situations, pyruvate is converted to acetyl CoA by a three-enzyme-complex called the **pyruvate dehydrogenase complex**. The first step is the TPP-dependent decarboxylation of pyruvate and conversion into HETPP. The TPP ylid is then removed in a lipoamide dependent fashion. The lipoamide is finally replaced by coenzyme A (producing acetyl CoA) and the dihydrolipoamide is oxidized by FAD back into lipoamide (and hence producing FADH2 which is then used to reduce NAD+ to NADH).



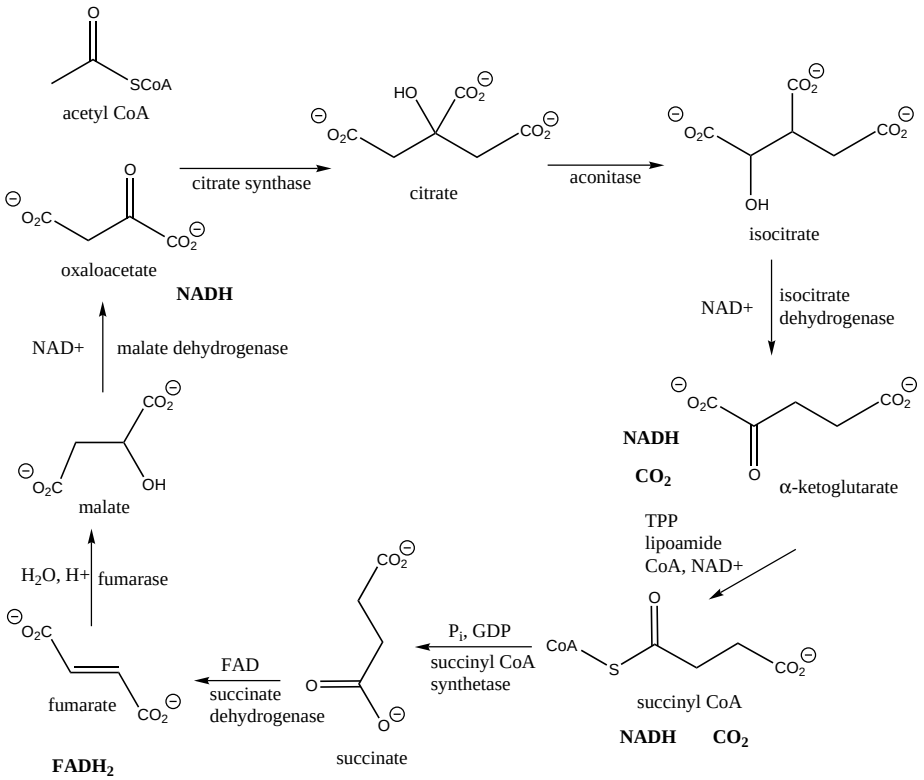
This conversion takes one molecule of pyruvate and converts it to one molecule of acetyl CoA, one molecule of carbon dioxide, and in the process reduces one equivalent of NAD+ to NADH and uses one molecule of Coenzyme A.

10.5 Citric Acid Cycle

Also called the Krebs cycle and the tricarboxylic acid (TCA) cycle, the Citric Acid cycle aerobically combines acetyl CoA with oxaloacetate and in a sequence of reactions produces two molecules of carbon dioxide and a number of reduced coenzymes (NADH and $FADH_2$).



The first step is a citrate synthase-catalyzed aldol condensation between acetyl CoA and **oxaloacetate (OAA)**, producing citryl CoA which is hydrolyzed to citrate. Aconitase then catalyzes an isomerization, moving the hydroxyl group from the 2 position to the 3 position in a stereospecific manner, producing isocitrate. Isocitrate is then oxidized by isocitrate dehydrogenase to give oxalosuccinate which undergoes decarboxylation to become α-ketoglutarate. α-ketoglutarate then undergoes another round of decarboxylation to yield succinyl CoA, a process which also requires TPP and lipoamide and Coenzyme A. Succinyl CoA is converted to succinate by a nucleophilic acyl substitution with a phosphate ion as a nucleophile and then removal of the phosphate to produce GTP and succinate. The succinate is dehydrogenated by succinate dehydrogenase using FAD as a cofactor and producing fumarate which is then hydrolyzed by the enzyme fumarase into malate. Finally, malate is oxidized by malate dehydrogenase to regenerate OAA and re-initiate the cycle.



10.6 Oxidative Phosphorylation

The bulk of the energy output from glucose catabolism comes from oxidative phosphorylation in the mitochondria, specifically the reduction of oxygen gas to water. The electrons to reduce oxygen come from the hydride carriers NADH and $FADH_2$. In a series of redox reactions, the electrons are "passed" in the **electron transport chain** by a sequence of carriers culminating at the end with the

reduction of oxygen to water. The energy released by each transfer is used to transport proton's from the inside compartment (**the matrix**) to the **intermembrane space**. This concentration gradient powers the F_0F_1 ATPase which converts the energy of protons moving down their concentration gradient into the synthesis of ATP in a process called **oxidative phosphorylation** (the oxidation of NADH and $FADH_2$ coupled to the phosphorylation of ADP to ATP). Generally, one molecule of NADH supplies sufficient energy to produce three equivalents of ATP, and one molecule of $FADH_2$ supplies sufficient energy to produce two equivalents of ATP.

It should be noted here that the NADH produced in glycolysis was primarily produced in the cytoplasm and hence needs to be trafficked into the mitochondria. There are two paths involved:

1. The **α-glycerol phosphate shuttle** operates by using NADH in the cytoplasm to reduce DHAP to α-glycerol phosphate which is then transported into the mitochondria whereby it is oxidized back to DHAP, and reduces FAD to $FADH_2$. DHAP is then exported back to the cytoplasm. Note, that in this shuttle, NADH is converted to $FADH_2$ which adds a bit of uncertainty to the final ATP output.

2. The **malate-aspartate shuttle** is used in muscle, heart, and brain tissues. NADH is used to reduce oxaloacetate to malate which is then transported into the mitochondria where it is re-oxidized to form oxaloacetate and to regenerate NADH. Oxaloacetate is then converted to aspartate which is then exported back to the cytoplasm where it is re-converted to oxaloacetate to re-initiate the cycle.

10.7 Glucose Catabolism: In Summary

1. Glycolysis converts one molecule of glucose into two molecules of pyruvate. In the process, it consumes two molecules of ATP and produces four molecules of ATP and two molecules of NADH, thus giving a net production of **2 ATP + 2 NADH**

2. Pyruvate is converted to acetyl CoA in a process which produces one equivalent of carbon dioxide and one equivalent of NADH, thus considering one molecule of glucose as the starting point, the net production is **2 CO_2 + 2 NADH**

3. Every cycle of the Kreb's cycle converts Acetyl-CoA into two molecules of carbon dioxide, three equivalents of NADH, one equivalent of $FADH_2$, and one equivalent of GTP (which can be used to produce ATP). Thus, giving a net production of **2 ATP + 4 CO_2 + 6 NADH + 2 $FADH_2$**

4. The electron transport chain produces approximately 2 ATP for every $FADH_2$ and 3 ATP for every NADH. As there are a total of 10 NADH molecules and 2 $FADH_2$ molecules, the maximum energy output of the electron transport chain is **34 ATP**

The maximum total, then, would be approximately 38 ATP, although because the electron transport chain is used to transport things other than protons and because of the two different types of NADH shuttles, the number can vary.

10.8 Poisons Which Affect Glucose Metabolism

A number of poisons affect steps in glucose metabolism. **Arsenate** ($HAsO_4^{2-}$), which is an analogue of phosphate, impacts glycolysis. It impacts the step catalyzed by GAP dehydrogenase in where the arsenate ion is incorporated by the enzyme rather than a phosphate ion. While glycolysis can still continue, the instability of the arsenate ester causes it to be readily hydrolyzed, producing 3-phosphoglycerate without generating ATP, thus eliminating the net ATP gain of glycolysis.

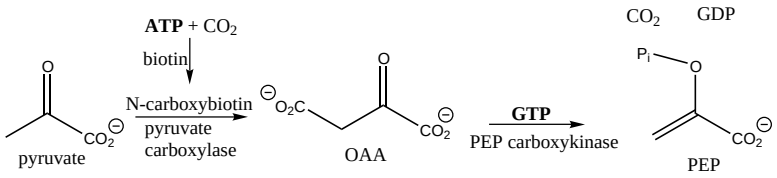
Other poisons are called **uncouplers** as they uncouple the connection between ATP synthesis and electron transport chain function. **Dinitrophenol (DNP)** and **valinomycin** are classic examples of this – they effectively punch holes in the mitochondrial membrane, allowing protons to diffuse down their electrochemical gradient, and decoupling it from ATP synthesis.

Cyanide and rotenone and antimycin inhibit specific components of the electron transport chain and hence prevent ATP synthesis in that fashion. Atractyloside and bongkrekate block the entry of ADP into the mitochondria and block the export of ATP, thus preventing oxidative phosphorylation from continuing once all the ADP has been phosphorylated.

10.9 Gluconeogenesis

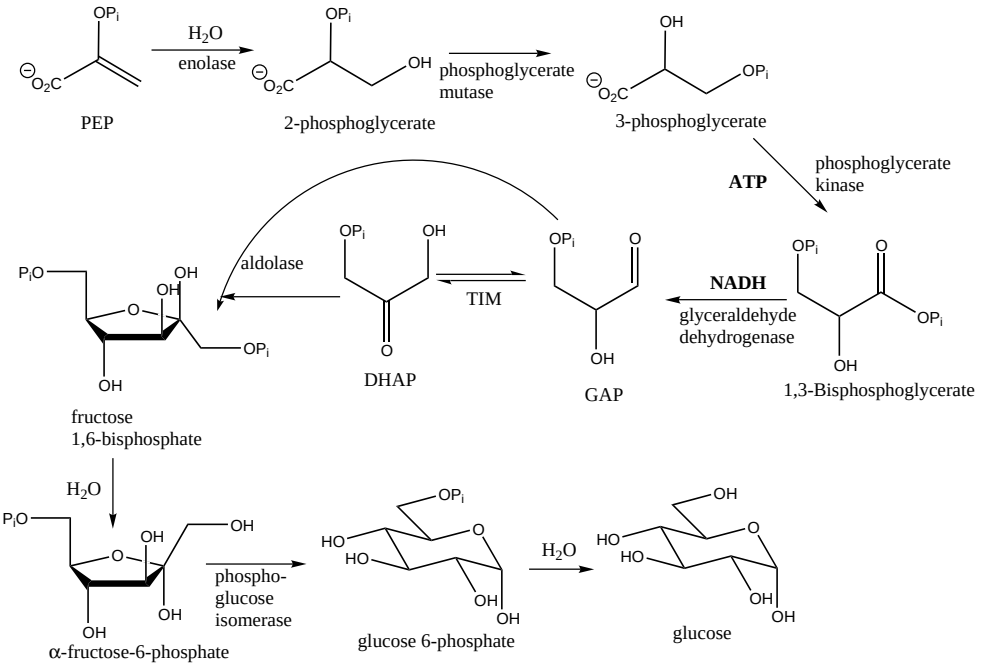
Glucose can be biosynthesized from pyruvate in an 11-step pathway called **gluconeogenesis**. As with fatty acid anabolism and catabolism, gluconeogenesis and the breakdown of glucose are not the exact reverse.

The first step is a biotin-mediated and ATP-dependent carboxylation of pyruvate catalyzed by pyruvate carboxylase to yield OAA. OAA is then decarboxylated and converted into PEP by phosphoenolpyruvate carboxykinase which transfers a phosphate from GTP (converting it to GDP) to the molecule. This may seem redundant (attaching a carboxyl and then removing it) but it is not as the attachment of the carboxyl is coupled to the hydrolysis of a molecule of ATP, thus releasing energy, making the subsequent reaction of OAA with GTP more favorable.



Water is then added across the double-bond of phosphoenolpyruvate to give 2-phosphoglycerate which is then isomerized to 3-phosphoglycerate, which are essentially exact reverses of the dehydration and isomerization reactions in glycolysis. 3-phosphoglycerate is phosphorylated, consuming a molecule of ATP, at the 1-position. Glyceraldehyde 3-phosphate dehydrogenase then reduces the molecule to GAP with NADH and displaces the phosphate group. As before, the enzyme TIM catalyzes the reversible conversion of GAP to DHAP. A molecule of GAP and a molecule of DHAP then undergo an aldol condensation (as opposed to the retro-aldol reaction in glycolysis) to give fructose 1,6-biphosphate.

The 1-phosphate is then hydrolyzed to give fructose 6-phosphate. Recall that in glycolysis, a molecule of ATP was consumed to ADP in order to phosphorylate the 1 position. This is not the case here as the production of ATP from ADP is not energetically favorable (the reverse of a favorable reaction cannot also be favorable). Fructose 6-phosphate is then converted to glucose 6-phosphate by the same series of keto-enol tautomerizations, but in reverse, which converted glucose 6-phosphate to fructose 6-phosphate in glycolysis. The final step is another hydrolysis reaction to hydrolyze the phosphate and produce glucose (rather than pushing the phosphate onto ADP to make ATP which is, as mentioned above, energetically unfavorable).



It should be noted that all the steps of gluconeogenesis are essentially the reverse of those used in glycolysis *EXCEPT* for:

1. The conversion of PEP to pyruvate in glycolysis is mediated by ATP hydrolysis, while the conversion of pyruvate to PEP in gluconeogenesis occurs through an OAA intermediate and requires a biotin-dependent-carboxylation, the hydrolysis of ATP, a decarboxylation, and the phosphorylation of OAA by GTP.

2. In glycolysis, fructose 6-phosphate is converted to fructose 1,6-bisphosphate by ATP hydrolysis. In gluconeogenesis, the reverse step is done simply by a phosphatase-mediated hydrolysis of the phosphate linkage, rather than the production of ATP.

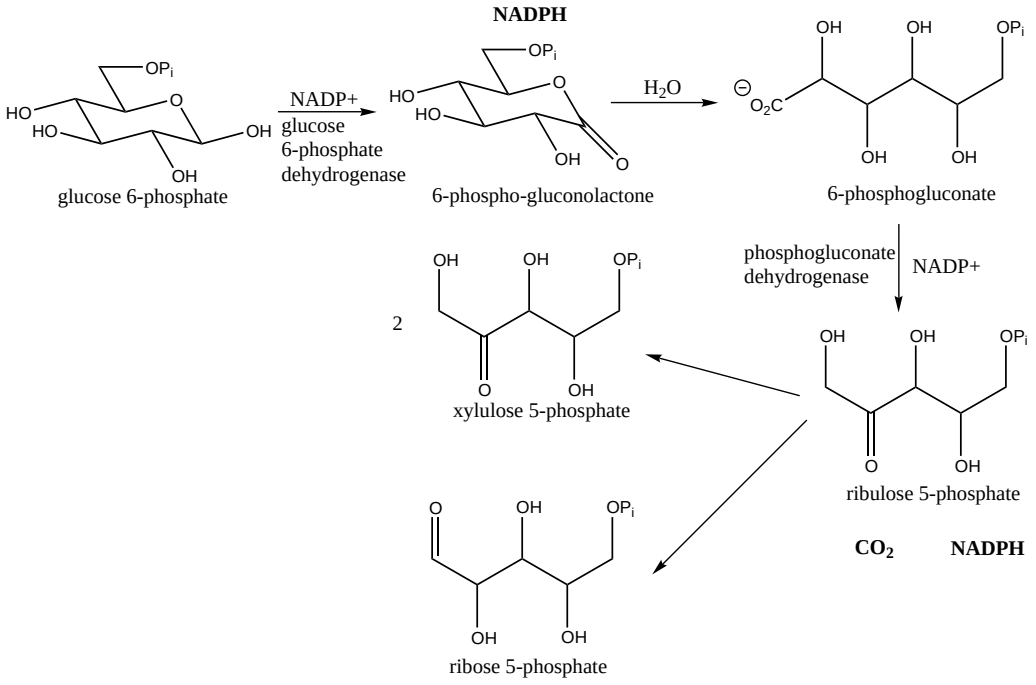
3. In glycolysis, glucose 6-phosphate is converted to fructose 6-phosphate by ATP hydrolysis. In gluconeogenesis, the reverse step is done simply by a phosphatase-mediated hydrolysis of the phosphate linkage, rather than the production of ATP.

Energetically, forming one molecule of glucose from two molecules of pyruvate requires 4 molecules of ATP, 2 molecules of GTP, and 2 molecules of NADH.

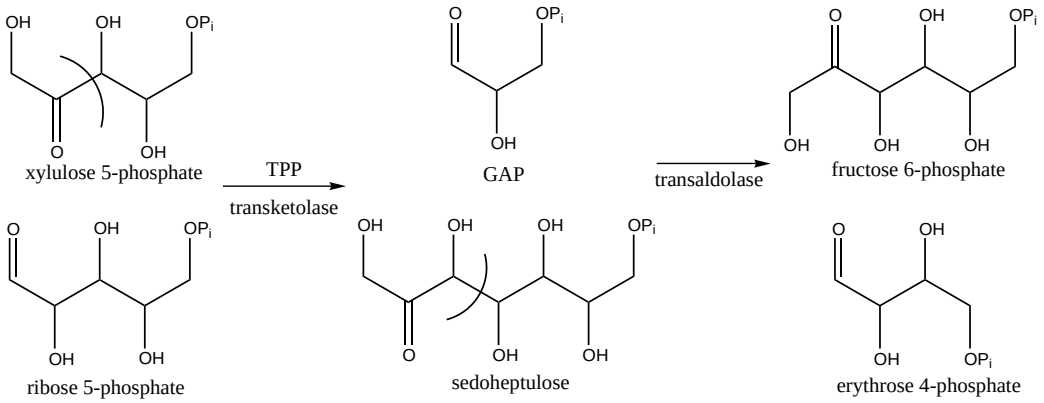
10.10 Pentose Phosphate Pathway

The **pentose phosphate pathway** is an alternative means of processing glucose. It is used to metabolize 5-carbon sugars, produces NADPH, and produces ribose 5-phosphate which is needed for nucleotide biosynthesis.

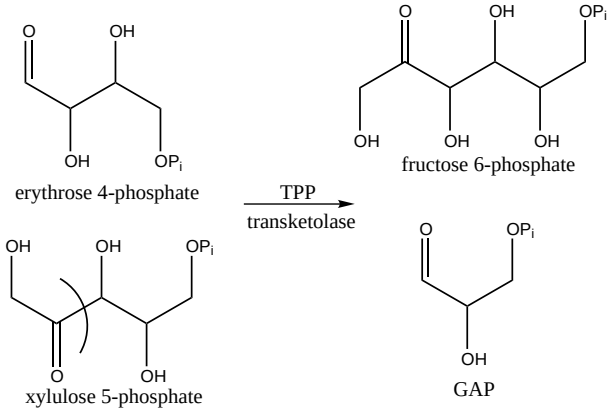
Glucose 6-phosphate is first oxidized by NADP+ as catalyzed by glucose 6-phosphate dehydrogenase to produce 6-phospho-gluconolactone and NADPH. This is then hydrolyzed to 6-phosphogluconate which undergoes another oxidation at the 3-position. This β-ketoester is then decarboxylated to give ribulose 5-phosphate, a two-step process which is catalyzed by one enzyme, phosphogluconate dehydrogenase, which creates another molecule of NADPH. Ribulose 5-phosphate is isomerized by a series of keto-enol tautomerizations catalyzed by ribulose 5-phosphate isomerase to produce ribose 5-phosphate. Simultaneously, ribulose 5-phosphate epimerase alters the stereochemistry at one carbon of ribulose 5-phosphate to yield xylulose 5-phosphate (**epimers** are diastereomers that differ at only one chirality center, and an epimerase thus converts between epimers). The relative rates of reactions typically give a two-fold excess of xylulose 5-phosphate. When ribose 5-phosphate is needed (for ribonucleotide synthesis), the epimerase can simply catalyze the reverse reaction, enabling more ribose 5-phosphate to be produced.



If ribonucleotide biosynthesis is not occurring, then the pentose phosphate pathway continues. One equivalent of xylulose 5-phosphate then reacts with ribose 5-phosphate in a TPP-dependent reaction catalyzed by transketolase which transfers two carbons from xylulose 5-phosphate to the ribose 5-phosphate, producing GAP and sedoheptulose 7-phosphate (transfers two carbons from the ketose [xylulose] to the aldose [ribose]). This is then followed by a transaldolase reaction which reacts sedoheptulose 7-phosphate with GAP, moving three carbons from the sedoheptulose (the ketose) to GAP (aldose). Note how despite the names, they both transfer carbons to the aldose sugar. This step creates fructose 6-phosphate (which can undergo glycolysis) and erythrose 4-phosphate (which is metabolized by some organisms).



Erythrose 4-phosphate can also react with xylulose 5-phosphate (the extra equivalent that did not react with ribose 5-phosphate). This is catalyzed by a TPP-dependent transketolase which transfers a two-carbon unit from xylulose 5-phosphate to erythrose 4-phosphate to give GAP and a second molecule of fructose 6-phosphate.



10.11 Photosynthesis Light Reactions

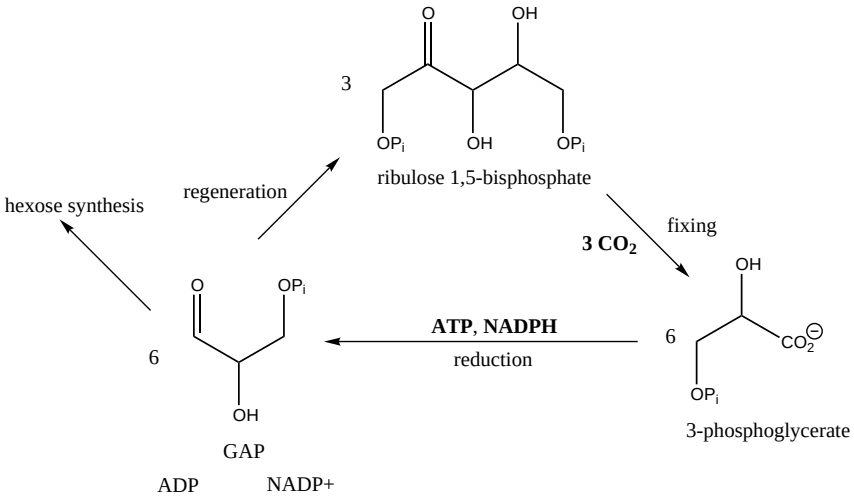
The light reactions are so called because of their dependence on sunlight. The sun mediates a process whereby chlorophyll captures solar energy and employs it to split one equivalent of water into two equivalents of protons, half an equivalent of oxygen gas, and 4 equivalents of high-energy electrons. These electrons are funnelled through an electron transport chain which are used to reduce NADPH and power another electrochemical gradient by pumping protons into the interior of the thylakoid discs, driving the production of ATP upon the diffusion of the protons down their gradient. Some of these electrons do not reduced NADP+ and undergo **cyclic phosphorylation** where they repeatedly pass through the electron transport chain, constantly pumping protons. The produced ATP and NADPH are then used in the light-independent "dark reactions".

10.12 Reductive Pentose Phosphate Cycle

Also called the **Calvin Cycle**, or the **reductive pentose phosphate (RPP) cycle**, these are the "dark reactions" of photosynthesis which employ the NADPH produced in the light reactions

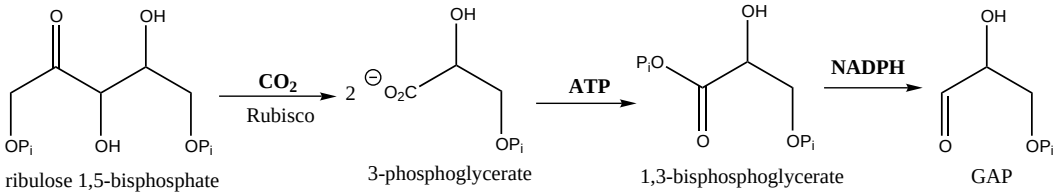
to reduce carbon dioxide to give carbohydrates. The process has three stages:

1. **Fixation** which reacts three molecules of carbon dioxide with 3 molecules of ribulose 1,5-bisphosphate to give 6 molecules of 3-phosphoglycerate
2. **Reduction** which reduces the 6 molecules of 3-phosphoglycerate to 6 molecules of GAP
3. **Regeneration** which takes 5 of the GAP molecules and regenerates 3 molecules of ribulose 1,5-bisphosphate. The last molecule of GAP enters the gluconeogenesis pathway to become glucose



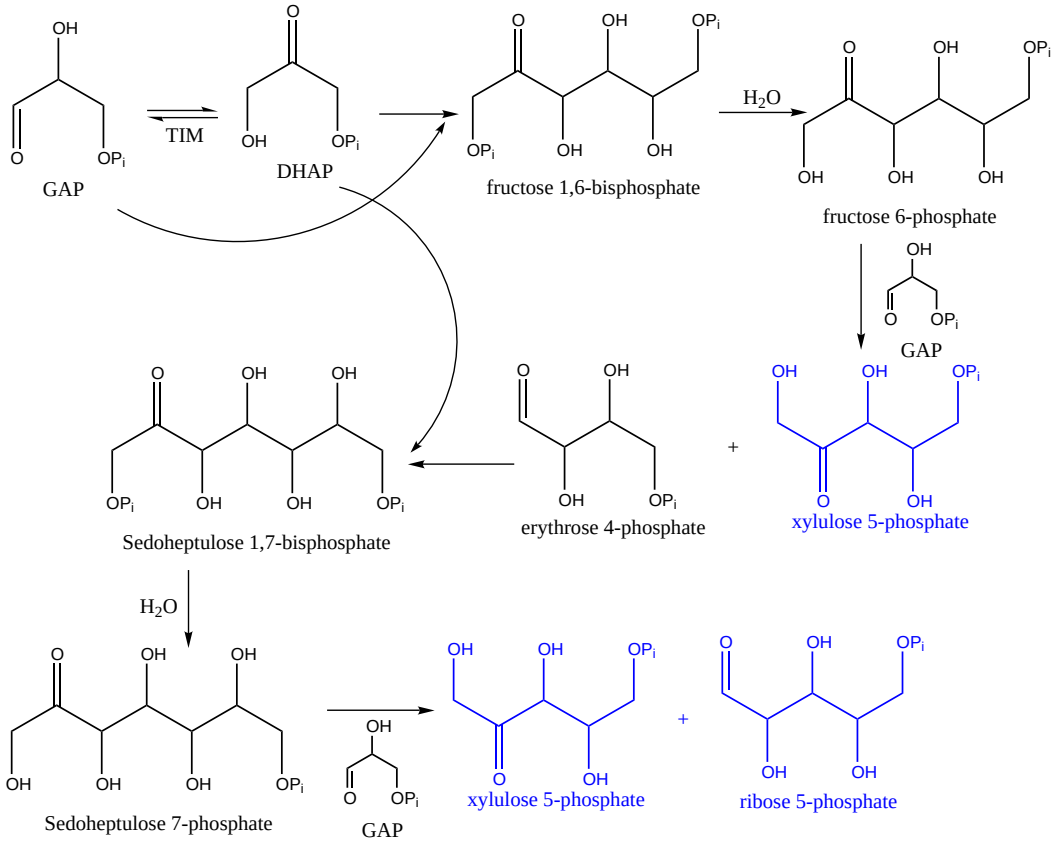
The RPP cycle is initiated by the carobxylation of ribulose 1,5-bisphosphate to give two molecules of 3-phosphoglycerate as catalyzed by the enzyme **rubisco**. The mechanism of the catalysis occurs via the production of a cis enediolate which is stabilized by a magnesium ion. This enediolate is able to attack carbon dioxide and hydrolysis at the carbonyl of the resulting β-keto acid generates two molecules of 3-phosphoglycerate, thus completing the fixation stage. Rubisco, it should be noted, oftentimes nonproductively fixes oxygen as well as carbon dioxide. Plants that have a danger of operating in low-carbon dioxide environments have adapted to this problem that un-adapted **C3 plants** do not deal with. These **C4 plants** form OAA (a four carbon compound) by combining carbon dioxide with PEP in cells specifically designed to collect carbon dioxide. OAA is then converted into malic acid and shunted to cells which engage in the Calvin Cycle where the acid is broken down back into PEP and carbon dioxide. These cells are removed from high oxygen levels and hence avoid the problem of rubisco catalyzing the wrong reaction. **CAM plants**, rather than spatially separating the site of carbon dioxide gathering and fixing, temporally separate the two.

3-phosphoglycerate is then phosphorylated at the carboxyl location by a molecule of ATP to generate 1,3-bisphosphoglycerate. This is then reduced to an aldehyde by NADPH at the site of carboxylation, also displacing an equivalent of inorganic phosphate to generate GAP, thus completing the reduction stage.

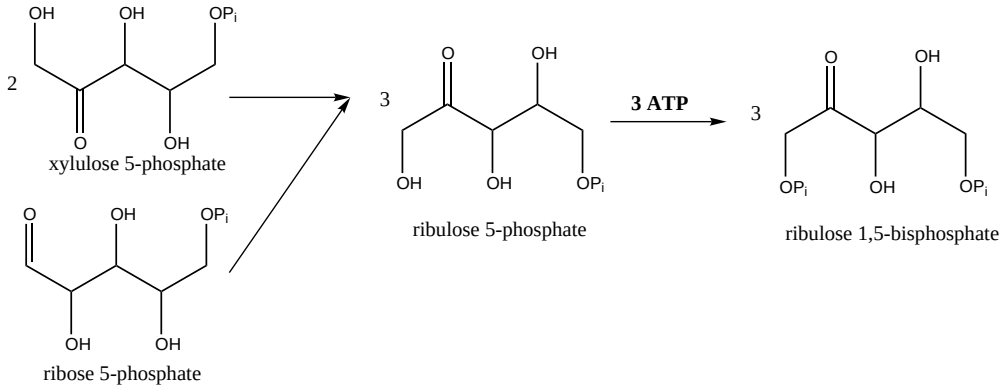


While one molecule of GAP enters gluconeogenesis, the remainder enter the regeneration stage. Two equivalents of GAP are isomerized to DHAP by the enzyme TIM. One DHAP molecule then undergoes an aldol condensation with another GAP molecule, producing fructose 1,6-bisphosphate which is hydrolyzed to fructose 6-phosphate. Fructose 6-phosphate then reacts with a molecule of GAP to give xylulose 5-phosphate and erythrose 4-phosphate as catalyzed by transketolase.

Erythrose 4-phosphate itself undergoes an aldol condensation with the second equivalent of DHAP to give sedoheptulose 1,7-bisphosphate. This is then hydrolyzed to sedoheptulose 7-phosphate which undergoes a reaction with another molecule of GAP which is catalyzed by a transketolase to give xylulose 5-phosphate and ribose 5-phosphate.



At this point, all 5 molecules of GAP have been used and have been utilized to produce 2 equivalents of xylulose 5-phosphate and 1 equivalent of ribose 5-phosphate. The xylulose 5-phosphates and the ribose 5-phosphate are all isomerized to ribulose 5-phosphate. These are then phosphorylated with ATP to give ribulose 1,5-bisphosphate to complete regeneration.

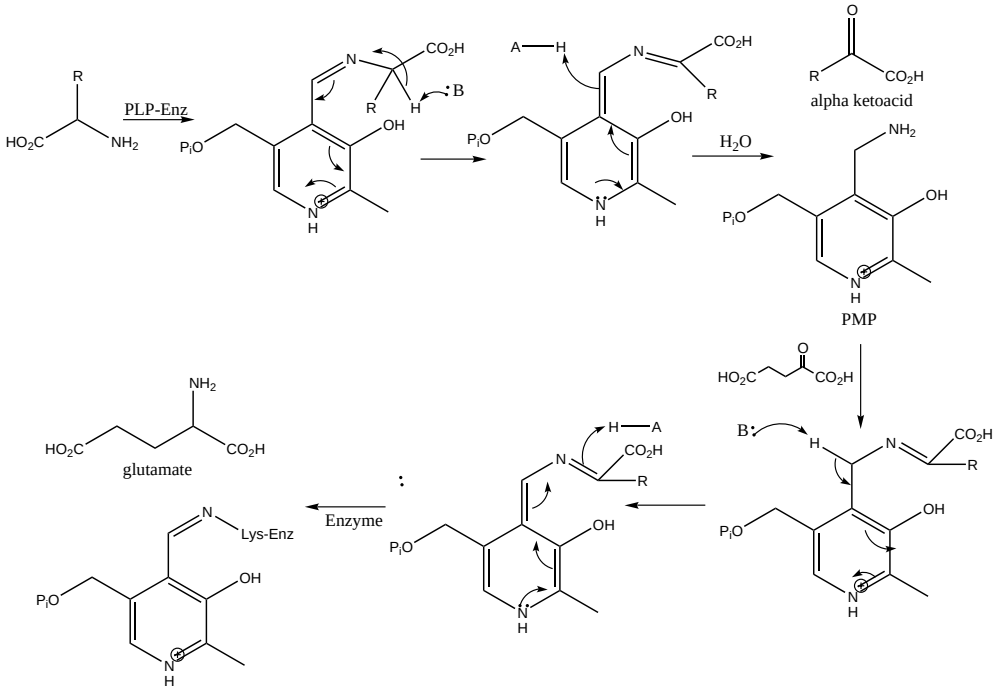


11 Amino Acid Metabolism

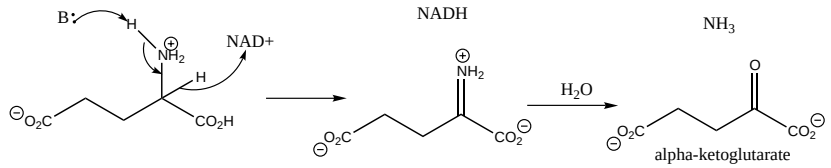
11.1 Urea Cycle

The first stage in most amino acid catabolism is **deamination**. This is handled by the cofactor PLP which mediates a transamination between the amino acid and α-ketoglutarate to produce the

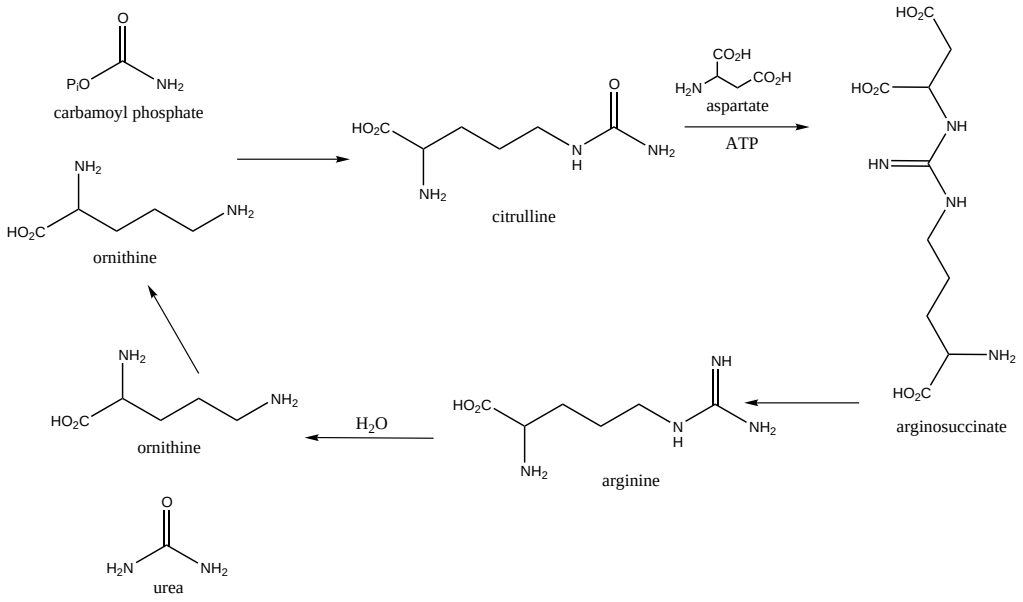
de-aminated amino acid and glutamate. This PLP-reaction is handled by deprotonation of the α proton upon transamination, and concludes by the formation of PMP upon hydrolysis. PMP then aminates α-ketoglutarate to glutamate, regenerating PLP.



Glutamate is then oxidatively deaminated to generate an imine (with the alpha proton removed). This imine is then hydrolyzed back to α-ketoglutarate and ammonia.



This ammonia enters the **urea cycle** where it is converted to urea. The first step is the conversion of ammonia to carbamoyl phosphate requiring two molecules of ATP and one of bicarbonate. This then undergoes the four-step transformation whereby it combines with ornithine to produce citrulline. Citrulline then undergoes an ATP-coupled condensation with aspartate to give arginosuccinate. Arginosuccinate then decomposes into arginine and fumarate. Arginine is then hydrolyzed to produce urea and regenerate ornithine, thus completing the cycle:



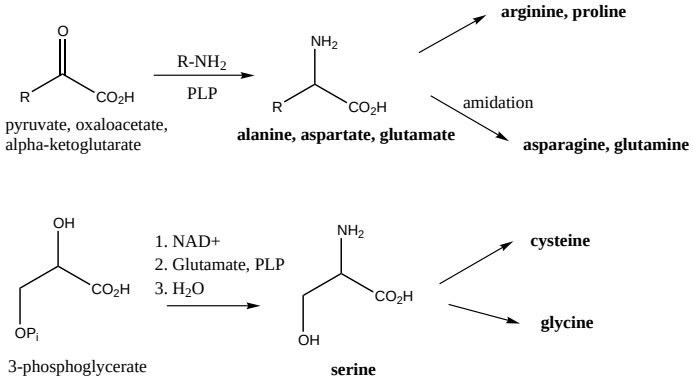
11.2 Amino Acid Side Chain Catabolism

Each amino acid has a specific degradation path. Many of them converge (ie glutamine being converted to glutamate and thus shares glutamate's catabolic pathway). These degradation paths allow one to characterize amino acids as **glucogenic** or **ketogenic**, as they are all ultimately degraded into components which can enter the gluconeogenesis pathway or the fatty acid biosynthetic pathway. Some can be both. Tryptophan, for instance, is both, as it produces acetyl CoA (which enters fatty acid biosynthesis) and pyruvate (which enters gluconeogenesis). Catabolism of all amino acids is completed by entry of each product into the Citric Acid Cycle.

11.3 Amino Acid Biosynthesis

Humans are only able to synthesize the 11 **non-essential amino acids**. The other 9 are **essential amino acids** which must be obtained from the diet.

7 of the non-essential amino acids are biosynthesized from pyruvate or other citric acid cycle intermediates via PLP-dependent aminations (and subsequent transformations in some cases). 3 of them are synthesized from the oxidation of 3-phosphoglycerate, amination by glutamate, and hydrolysis (and subsequent transformations).

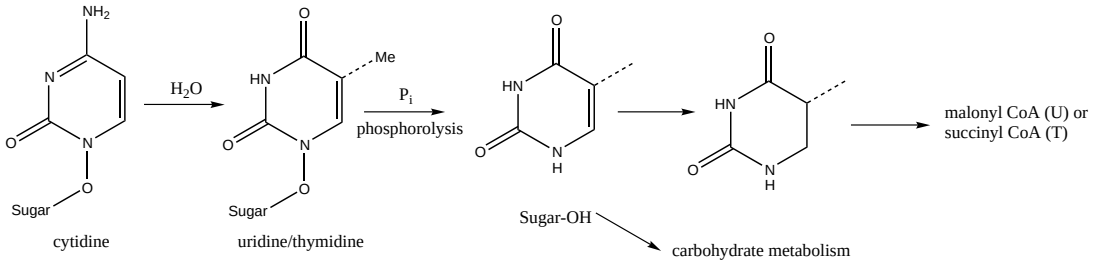


12 Nucleotide Metabolism

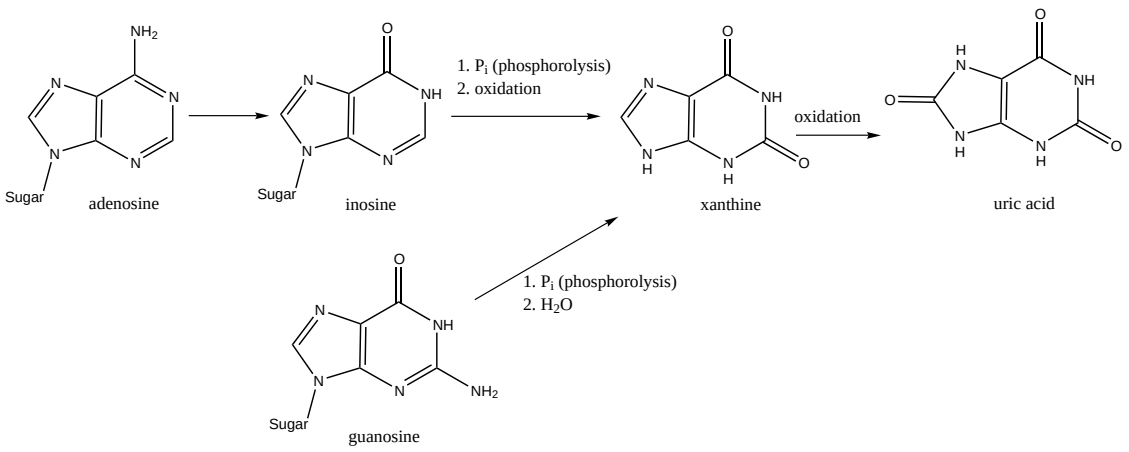
12.1 Nucleotide Catabolism

Dietary nucleic acids are hydrolyzed to the individual nucleotides and some are transported to be used for nucleotide/DNA/RNA synthesis while the rest are catabolized.

The pyrimidines are metabolized in similar ways. Cytidine is converted to uridine before processing. Uridine and thymidine undergo phosphorolysis to cleave the base from the sugar. The free base is then reduced by a coupled NADPH/$FADH_2$ system and then metabolized to malonyl CoA or succinyl CoA.



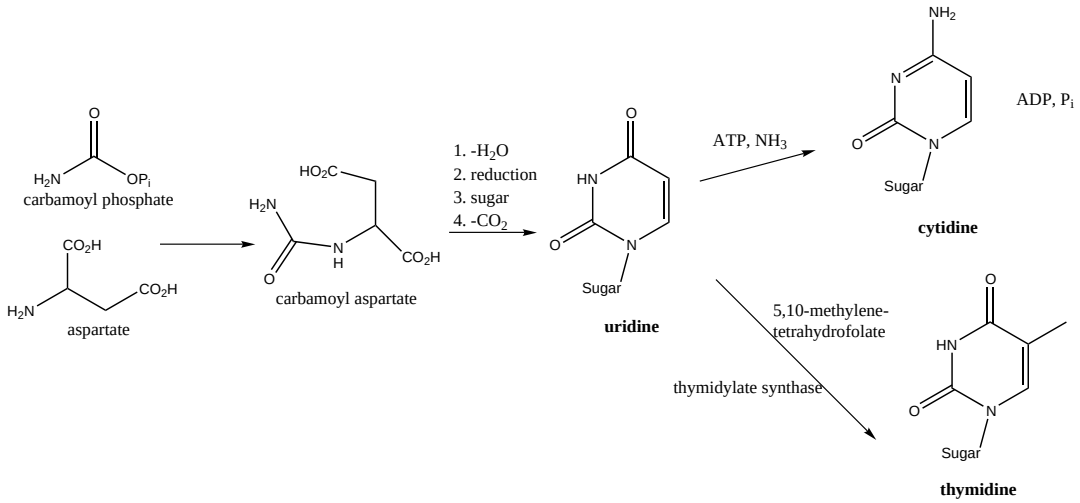
The purines are metabolized by conversion to free xanthine which is then oxidized to uric acid. Adenosine is converted to inosine which is then phosphorolytically cleaved to produce free hypoxanthine (which is then oxidized to produce xanthine). Guanosine is phosphorolytically cleaved and then de-aminated to xanthine.



12.2 Nucleotide Biosynthesis

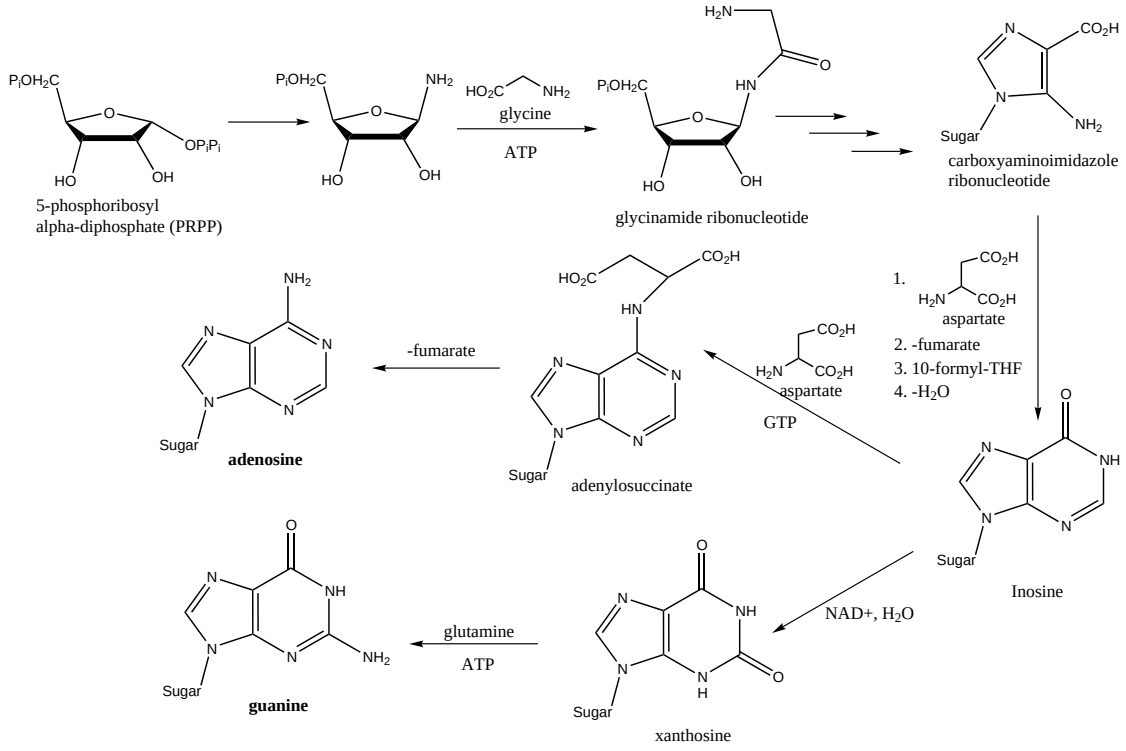
Uridine is produced from the condensation of carbamoyl phosphate and aspartate. This carbamoyl phosphate is then cyclized, reduced, coupled to the appropriate sugar, and then decarboxylated. Cytidine is then made from this by an amination (the opposite of the de-amination step in cytidine catabolism).

Thymidine is produced by the methylation of uridine, a process catalyzed by **thymidylate synthase**, an enzyme which is inhibited by the anti-cancer agent 5-FU. Thymidylate synthase uses 5,10-methylene-tetrahydrofolate as the methyl donor and forms a substrate-enzyme covalent linkage which activates the correct position of uridine to pick up the methyl group.



Unlike the pyrimidine biosynthesis which couples sugar to a completed base, purine biosynthesis assembles the base on the ribose sugar, producing inosine monophosphate which is the precursor to guanosine and adenosine. The process of inosine formation involves attaching an amine group to the 1' position of the sugar. This group then attacks an activated glycine residue, creating the backbone of the five-membered ring component of inosine. Aspartate and a formaldehyde group (donated by 10-formyl-tetrahydrofolate) then form the backbone of the six-membered ring part of inosine.

Inosine is then converted to adenosine by GTP-coupled condensation with aspartate to form adenylosuccinate and then loss of fumarate to give the final nucleotide. Inosine is converted to guanosine by an NAD+ oxidation and the ATP-coupled amination of the newly formed ketone.



13 Integration of Energy Metabolism

13.1 Universals

All metabolic pathways are regulated to serve a basic purpose: to supply a ready amount of energy to all cells. This energy is ultimately in the form of ATP. Indirectly, this ATP comes from glucose which is necessary, first because glycolysis and the citric acid cycle are the dominant sources of energy, and secondly because pyruvate, the by-product of glycolysis, can undergo **anaplerotic reactions** to form OAA and Malate which are needed to keep the citric acid cycle running. Fat, protein, and glycogen act as sources/supplies of energy.

13.2 Hormonal Regulation

Three hormones are the master regulators of metabolism: **insulin**, **glucagon**, and **epinepherine**. Insulin is produced in response to high levels of glucose, and induces cells to take up glucose from the bloodstream and store excess energy as fat, glycogen, and protein. It activates a Receptor Tyrosine Kinase system. Glucagon responds to and exerts the exact opposite effects. It activates a G-Protein Coupled Receptor system. Epinepherine induces quick retrieval of glucose and fat for quick bursts of energy. Thus, this defines four basic metabolic states:

1. **Feeding** is eating. During feeding, the precursors to energy storage molecules are in high amounts, and the pancreas responds by secreting insulin.
2. **Fasting** is not eating. During fasting, the body withdraws from the energy supplies that were stored during feeding, and the pancreas responds to this and induces this by secreting glucagon.
3. **Starvation** is not eating for long periods of time. At this point, glycogen stores have been depleted and metabolic signals are used to induce starvation behavior in the brain and the body (ie use of protein for energy). Note, diabetics experience constant starvation despite high blood glucose levels.
4. **Excitement** refers to the immediate need for energy. It is triggered by the adrenal glands releasing epinepherine into the circulation.

Glucagon's binding to GPCR's induces **adenylyl cyclase** which produces **cAMP**. cAMP activates various protein kinases which phosphorylate and hence regulate a variety of pathways which exert glucagon's effects. Regulation of metabolism, then, can be mediated by regulation of cAMP production/degradation (and thus overall phosphorylation), by the levels of products of glycolysis, NADH, and Acetyl-CoA.

13.3 Metabolic Rules

1. ATP and Glucose levels must be relatively constant
2. The utilization of fat for energy requires carbohydrates (anaplerotic reactions are needed)
3. Glucose cannot be made from fat. β-oxidation results in formation of acetyl-CoA molecules. Glucose production requires OAA which can only be made from acetyl-CoA through the citric acid cycle. But, because the citric acid cycle begins by combining OAA with acetyl CoA, there can be no net new production of OAA from acetyl CoA.
4. Synthesis and degradation are tightly regulated such that **futile cycles** (where energy is consumed going nowhere) are minimized.
5. Low energy turns on lipolysis and glycolysis.
6. Low glucose levels turn on protein degradation and gluconeogenesis and glycogen breakdown. While muscles and liver both store glycogen, only the liver and kidneys are able to engage in significant glycogen breakdown for other tissues to use (due to muscle cells lacking a particular enzyme).

7. Protein phosphorylation due to cAMP activates pathways that maintain glucose levels and retrieve energy and inactivates enzymes that store glucose, fat, and protein.

13.4 Tissue Cooperation

Different tissues may cooperate to help maintain energy levels. One example of this is the relationship between muscle and liver. During starvation, muscle and brain adapt by activating systems to use ketone bodies and protein as a source of energy, while the liver activates pathways to make ketone bodies and process protein metabolic byproducts. During excitement, the liver is actively engaged in producing glucose while the muscle is engaged in burning it.

Another example of this are the **Interorgan Cycles**. The **Cori Cycle** refers to the muscles consuming glucose and producing lactate in response to activity, while the liver will take up lactate to produce glucose which it then releases back into the bloodstream.

The **alanine cycle** refers to the interorgan cycle where starving muscles catabolize protein and as a result some of the nitrogen wastes are used to transaminate pyruvate (from glycolysis) into alanine. This alanine is then sent to the liver where it is converted back to pyruvate and then converted back to glucose, thus processing the nitrogen wastes and recycling carbon products back to glucose.

The formation of **ketone bodies** is a consequence of starvation and the accompanied metabolism of fat. The liver's rapid processing of fatty acids into acetyl-CoA can consume the stores of free CoA. The liver can respond by condensing three acetyl-CoA molecules to form one molecule of HMG-CoA which is then split by HMG-CoA lyase to produce acetoacetate (or hydroxybutyrate, the reduced form), acetyl CoA, and two free CoA molecules. Acetoacetate and hydroxybutyrate are then transported to muscle and brain tissues which can adapt to utilize them by converting them into acetyl-CoA molecules.

Part IV
Gene Expression

14 Basic Genetics

14.1 Mendel

Gregor Mendel's experiments were conducted on **true-breeding** pea plants, plants that self-pollinating would only produce offspring with a specific **trait** for a given **character** (purple flowers is the trait, flower color is the character). Mendel's experiments revolved around his willingness to apply numerical calculation to analyzing **P generation**, **F_1 generation**, and **F_2 generation** traits. Using it, he found that the traditional understanding of traits, **blending**, was false and that inheritance was governed by **particulate factors**, or genes.

Mendel found that individuals had two alternative versions of genes, or **alleles**, the combination of which determined the **genotype** and the observed trait (**phenotype**). One allele came from the mother, one from the father. Typically, one of these alleles was **dominant**, in the sense that it was the dominant allele which determined the phenotype despite the **recessive allele** (one bearing two dominant or two recessive alleles is said to be **homozygous**, one bearing one of each is **heterozygous**). These alleles would then segregate during gamete production, such that half of the produced gametes would receive one allele and the other half would receive the other, making the F_1 generation's genotypes to be half from the father, half from the mother – which was **Mendel's Law of Segregation**. This can be shown in traditional **Punnett square** depictions (which I will not show because if you don't know Punnett square, you shouldn't even think about taking this GRE) and is studied in humans using **pedigrees**.

Mendel also found that if you looked at two separate traits (ie pea coat and pea color) that the F_2 generation yielded trait quantifications which showed that the alleles segregate into gametes independently (ie the allele for color and the allele for shape "choosing" gametes independently of one another). This is referred to as **Mendel's Law of Independent Assortment**.

14.2 Beyond Mendel

While Mendel's laws are for the most part true, it is an oversimplified view for several reasons:

1. Mendel's pea crosses used traits that showed **complete dominance** in that a dominant allele was the only allele that mattered. In cases of **incomplete dominance**, both traits show. For instance, snapdragon color is an example of incomplete dominance where the F_1 generation of a cross between homozygous red flowers and homozygous white flowers (no functioning pigment protein) is a pink flower (half as much red). This is not a case for blending, however, as the F_2 generation shows red, pink, and white flowers in a classic 1:2:1 ratio confirming particulate inheritance and demonstrating that dominance need not be complete.

2. **Codominance** can be thought of as incomplete dominance when, instead of one gene being "null" (like the white snapdragon flower) in phenotype, both genes are expressed equally. One example of this is blood group genes, where a person with the alleles coding for type A blood and type B blood both express themselves equally on the surface of red blood cells.

3. Mendel's experiments and conclusions suggested that traits only have two types of alleles. Many genes have **multiple alleles**, such as blood group (A, B, O).

4. Mendel's experiments and conclusions also suggested that alleles only control one trait. **Pleiotropy** refers to genes which control multiple traits (ie a prevalent transcription factor).

5. Mendel's experiments argue that assortment of alleles occurs independently of one another. Even if this is so, one can still see some sort of "linkage", or **epistasis**, between two genes if the genes themselves affect each other's expressions. The classic example of this is mice coat color as one gene controls whether or not the color is black or brown and one gene controls whether or not pigment gets to the skin.

6. **Polygenic inheritance** refers to how many traits, especially quantitative ones, are determined by many genes. Characters that are **multifactorial** are determined not only by many genes but my environmental factors as well.

14.3 Chromosomal Inheritance

Mendelian inheritance has its physical basis in chromosomes. This was discovered by Thomas Hunt Morgan who discovered that some traits are linked with sex (**sex-linked genes**), demonstrating that the gene was on the sex-determining chromosome. This explained why some genes were **linked** (did not assort independently of one another).

Morgan also found that even genes on the same chromosome were not completely linked. This is due to meiotic **recombination**, or a **crossing-over** of genes in neighboring chromosomes. Using this fact, one can then create a map of the chromosome (**linkage map**), as one expects genes closer together to be less likely to cross over relative to one another. This degree however cannot be fully used as a means of measuring physical distance, especially as genes that are spread about also have a higher chance of crossing-back to their original chromosome. A 1% chance of crossing over relative to another gene is thought of as a "distance" of 1 **map unit** or 1 **centiMorgan**.

Sex-linked traits express themselves dominantly in the gender that lacks the two sex-determining chromosomes (males in most mammals, females in birds, some fishes, and some insects). Male mammals, for instance, also pass on these sex-linked traits to all their daughters.

Chromosomal inheritance also sheds light on the issue of **X-Inactivation**. Female cells will randomly inactivate one of their X chromosomes by methylation and compacting of the chromosome into a compact **Barr body**. For females heterozygous for a sex-linked gene, this can lead to some cells expressing one and some cells expressing the other.

If in meiosis, the chromosomes are not distributed equally (a **nondisjunction**), the offspring will have an abnormal chromosome number, or **aneuploidy**. This often leads to various defects in humans, but surprisingly, plants with **polyploidy** appear to be healthier.

It should be noted that there are forms of non-chromosomal inheritance. Specifically, genders often **imprint** their gamete genetic material with **epigenetic modifications** (modifications not in

the genetic sequence itself). Additionally, **cytoplasmic inheritance** also occurs in the sense that some traits are passed on solely by what is occurring in the cytoplasm (plastids in plants, mitochondrial DNA), and due to the inability of sperm to provide these, these come almost exclusively from the mother.

14.4 What are Genes?

Bacteria are known to be able to change their traits. The acquisition of such traits can occur in three ways:

1. **Conjugation** refers to two bacteria exchanging genetic material directly through physical contact.
2. **Transformation** refers to bacteria acquiring genetic material from the environment (ie from dead bacteria). Avery, Macleod, and McCarty found that the acquired material was DNA using experiments with transformation.
3. **Transduction** refers to acquiring genetic material from other bacteria by using bacteriophages as vectors.

Using simple experiments, scientists were able to find that genes which seemed to give the same overall phenotype (ie inability to metabolize a certain sugar) could be disrupted in such a way that gave a hierarchy of the genes. As biochemically it was known that metabolism occurred in steps, this led to the idea that genes were responsible for those individual steps (**one-gene-one-enzyme hypothesis**).

This was confirmed by experiments dealing with **recombination** and **complementation**. Plates of bacteria were infected with two bacteriophage strains which had two different mutations which rendered them unable to infect the particular strain. They found instances of recombination where bacteriophage chromosomes were able to cross-over, such that the phage progeny of the infections were able to infect the strain, and the frequency of crossing over enabled mapping of the viral chromosome. They were also able to find evidence of complementation, where no crossing over occurs, but infection is still possible. In these experiments, progeny phage are unable to infect the bacterial strains, yet the two individual strains together are able to cause an infection. Complementation showed that individual genes come in units, such that if one strain lacks gene A, but has gene B, and one strain lacks gene B, but has gene A, then the two viruses can complement one another and together infect the cell. But, two phage lacking gene A cannot complement one another because the function of gene A has not been restored – thus confirming the one-gene-one-enzyme hypothesis.

15 The Genetic Code

15.1 Degeneracy

The Central Dogma claims that the information found in the sequence of nucleotides in DNA is transferred from DNA to RNA to protein. This genetic code is read in series of three nucleotides (**codons**). As there are four nucleotides, there are $4^3 = 64$ possible codons. However, ribosomal translation typically only employs 20 amino acids, signifying substantial **degeneracy** in the genetic code such that several codons are **synonyms**, or they code for the same amino acid. This degeneracy is an evolutionary necessity as it makes allowances for inevitable mutations which occur in the genetic code. For example:

- Mutations in the first position often give a similar amino acid
- Because **transition** mutations are more common (where purine becomes purine and pyrimidine becomes pyrimidine), transition mutations in the second position often give similar amino acids (a second position with a pyrimidine is typically hydrophobic and with a purine is typically polar)
- Mutations in the third position often give the same amino acid

- If the first two positions of the codon are G or C, the strength of the bond makes the third position redundant so mutating the third position in these cases gives the same amino acid

15.2 Wobble

It was at first believed that every codon has one tRNA, but characterization of tRNA species showed that this was not the case. Additionally, they found that the 5' end of the tRNA anticodon (which corresponds to the 3' end of the codon, hence the last position) used a non-Watson-Crick base, **Inosine**, which is a de-aminated form of adenine. Crick, in 1966, devised the **Wobble Concept** to explain the smaller numbers of tRNA then expected, which argues that the 5' position of the anti-codon is not as spatially confined and is thus able to participate in a wider variety of contacts, thus allowing the third position to be a bit more flexible. This has been confirmed by structural analysis of tRNAs.

15.3 Stop Codons

Stop codons will be discussed in the protein translation section but I thought it would be good to put in here that stop codons are NOT recognized by tRNAs.

15.4 Historical

The genetic code was deciphered by using cell-free translation systems with ribosomes and high Mg concentration (which remove the need for the typical starting factors) and synthetic mRNAs. By using different sequences, scientists were able to deduce the entire genetic code (sounds a lot easier than it is, because they didn't use solid phase synthesis back in the day, they could only make random orderings of polynucleotides).

15.5 Point Mutations

Several types of point mutations can occur:

1. a **missense mutation** is a mutation which alters a single amino acid in a polypeptide
2. a **nonsense mutation** is a mutation which prematurely introduces a stop codon in the sequence creating a nonfunctional/terminated protein
3. a **frameshift mutation** is an insertion or deletion of a base pair into a genetic code which knocks all subsequent bases out of frame with their intended reading frame

15.6 Suppression

If a mutation induces a specific phenotype, subsequent mutations which are able to neutralize the first one are called **suppressor mutations**. There are three types. The first and most obvious are **reverse (back) mutations** which simply undo the first mutation. The second are **intragenic mutations** which are mutations in another part of the mutated gene which have the ultimate effect of undoing the first mutation. The third are **intergenic mutations** which are mutations in another gene locus altogether which still suppresses the first mutation.

Another form of suppression separate of mutations exists. In many species of bacteria, mutant tRNAs exist which are capable of suppressing nonsense mutations by binding the stop codon and simply introducing an amino acid. (In case you're curious as to how the bacteria continues to read properly, it turns out that E. Coli does not use certain stop codons very often, so it makes sense to suppress those as nonsense mutations)

15.7 The Code is Universal

Except in some cases (ie mitochondrial DNA, bacterial DNA), the genetic code is interpreted the same way universally.

16 Transcription

16.1 RNA Polymerase

The enzyme that drives transcription is **RNA Polymerase**. In bacteria, there is only one RNA polymerase while Eukaryotic cells possess three (RNA Polymerase 1, 2, and 3). RNAP1 and RNAP3 transcribe ribosomal RNA and transfer RNA, respectively. RNAP2 transcribes mRNA. Not surprisingly, RNAP2 and the bacterial core enzyme are very homologous at the sequence level and are even more similar on a structural level (closer towards the active site). RNA Polymerases look like "crab claws". The space between the claws is the beginning of a channel and the "palm" of the claw is the active site of the enzyme which holds the two Mg^{2+} ions necessary for polymerase (both RNA and DNA) activity.

16.2 Chemistry

Transcription resembles DNA replication in that the nucleotides are linked together with the 3' hydroxyl group of the last nucleotide acting as a nucleophile to attack the α phosphate of the next nucleotide and displace pyrophosphate. Such a reaction requires metal cations to stabilize the negative charges on the phosphates.

16.3 Initiation

Transcription itself is divided into three steps: **initiation**, **elongation**, and **termination**. In initiation, the RNA polymerase begins by binding to promoter sequence and then melting the DNA before undergoing **abortive initiation**, whereby RNA polymerase constructs short stretches of RNA without advancing, regularly stopping and going back to the transcription initiation site. This abortive initiation is then followed by a transition to the processive transcription of the elongation phase.

In prokaryotes, the σ factor determines RNA Polymerase's promoter specificity. The RNA Polymerase-σ factor complex is called the bacterial **RNA Polymerase holoenzyme**. The **consensus sequence** promoter recognized by σ are a set of two 6-base sequences upstream of the transcription initiation site - a TTGACA (at -35) followed by 17-19 bases and then a TATAAT (at around -10). The consensus sequence is derived statistically by noting which elements of the promoter are most commonly shared amongst all bacterial genes. The closer to the consensus sequence a promoter is, the stronger it is said to be, in that it promotes more transcripts in a given period of time.

The initial binding event between the holoenzyme and the promoter leaves the DNA in a double-stranded conformation (termed the **closed complex**). This is followed by an **isomerization** (structural change) of the holoenzyme into a form which can actually initiate transcription which is called the **open complex**. This structural change causes DNA melting to occur somewhere between -11 and +3.

Because RNA Polymerase does not require a primer (unlike DNA polymerases), the protein structure must make a number of contacts with the DNA backbone (one reason that most transcription initiation sites start with A) in order to create a stable complex from which to initiate transcription. This initiation is then abortive, where short mRNA strands are created but terminated and the process repeats itself several times. Once the RNA polymerase gets past abortive initiation, it is said to **"escape" the promoter** and begin the elongation phase.

Eukaryotic initiation is similar, but somewhat more complex. As with prokaryotes, RNA polymerase II in eukrayotes recognizes specific promoter sequences, although the promoters are themselves more complicated and, instead of being recognized by a σ factor, are recognized by a wide range of **General Transcription Factors** (GTFs). The complex of RNAP2, GTFs, and the promoter is called the **Pre-Initiation Complex**.

The promoters are also very diverse, although four common elements (not all of which are found in every promoter) are the **BRE** (TFIIB Recognition Element) and **TATA Box** which are upstream of the transcription initiation site, the **Inr** (Initiator Element) which happens just at the transcription initiation site, and the **DPE** (Downstream Promoter Element) which occurs about 28 base pairs downstream of the initiator site. The TATA Box is recognized by the TATA Binding

Protein (TBP) which is itself a component of the TFIID complex which also has components which recognize the Inr and DPE sites. In the formation of the Pre-Initiation Complex, TFIID first binds the TATA box and bends the DNA. This structure can then recruit TFIIB and TFIIA which can then recruit RNA Polymerase and **TFIIH** which is critical for the melting of the DNA complex and for proceeding to elongation after several rounds of abortive initiation. The move to elongation is accompanied by TFIIH's multiple-phosphorylation of RNAP2's **C Terminal Domain** (CTD) tail which consists of repeats of phosphorylatable amino acids. Phosphorylation of the CTD assists the RNAP2 in "shedding" off the GTFs and mediator complex and in recruiting the RNA processing machinery which acts on the RNA as it is transcribed. It should be noted that there are other factors that have not been described here and each play a role in creating the Pre-Initiation Complex and in recruiting and stabilizing RNA Polymerase II.

Eukaryotes also make use of an extensive **Mediator Complex** which consists of further elements which binds DNA elements, other mediators, and/or RNA Polymerase. These components are needed due to the inherent complexity of the eukaryotic genome which is packaged into chromatin and other difficult-to-access superstructures. As large structures consisting of the mediator complex, many GTFs, and RNA Polymerase II have been assembled without DNA, some have termed this complex the Eukaryotic RNA Polymerase II Holoenzyme, although it is not certain how relevant *in vivo* this is. Interestingly, many components of this "holoenzyme" do not actually interact directly with the polymerase (ie TFIID).

16.4 Elongation

Elongation refers to the **processive** (meaning that it is not stop-release-bind again-start again, but continually adds nucleotides) addition of ribonucleotides inside the enzyme active site. The DNA enters the enzyme at one place, splits into the template and non-template strands, and then reforms behind the polymerase, while RNA nucleotides enter from one channel, and the transcribed RNA leaves through yet another channel. At any given time, only 8 or 9 RNA nucleotides are actually paired with DNA.

In prokaryotes, elongation is accompanied by two forms of proofreading. The first is **pyrophosphorolytic editing** which is simply the reverse reaction of transcription which reincorporates pyrophosphate, as opposed to displacing it. This process can remove correct base pairs as well as incorrect base pairs, but because of the enzyme active site structure, the enzyme tends to hover over incorrectly made pairs longer, and is thus more likely to remove an incorrect nucleotide (even if it removes a correct nucleotide there is a reasonably high chance that a correct nucleotide will add back on).

The second form of proofreading is **hydrolytic editing** which is where the enzyme hydrolyzes parts of the RNA which are coded improperly. This process is stimulated by **Gre factors** which not only enhance proofreading but also increase transcription speed by decreasing the time the enzyme spends at hard-to-transcribe areas (RNA transcription does not occur at a constant speed throughout the transcript).

In eukaryotes, elongation is started and enhanced by a component of the initial complex called **P-TEFb** which phosphorylates a residue on the CTD tail and is responsible for phosphorylating and recruiting TAT-SF1 and hSPT5 which, along with P-TEFb, are **elongation factors**. Similar to the Gre factors in prokaryotic elongation is the eukaryotic **TFIIS** which in addition to reducing the time that transcription stalls at specific sequences, also enhances the pyrophoshorolytic proofreading as well as an internal RNase function of the RNA Polymerase which is comparable to the hydrolytic editing in prokaryotes. Eukaryotic elongation is also accompanied by the recruitment of RNA processing enzymes and factors needed for termination.

16.5 Termination

In prokaryotes, sequences in the RNA called **terminators** signal for termination of transcription. These terminators are either **Rho dependent** or **Rho independent**.

Rho independent terminators, also called **intrinsic terminators**, usually have a short inverted repeat of about 20 nucleotides followed by a stretch of several A's. The inverted repeat is prone to forming a hairpin loop structure which, when coupled with the more weakly bound A-U pairs

leads to dissociation of the RNA from the polymerase (although the specific mechanism is not quite clear).

Rho dependent terminators are less well characterized but recruit the protein **Rho** which is an ATPase which uses ATP hydrolysis to trigger termination. Rho is unable to terminate transcription of RNA that is being translated.

Eukaryotic termination is achieved by the recognition of termination sequences. These sequences cause termination factors such as CPSF and CstF (which have already been recruited to the RNAP2 CTD tail during elongation) to bind to the RNA and recruit **poly-A-polymerase** which attaches a polyA tail to the RNA and also triggers dissociation of the RNA from the RNA polymerase.

17 RNA Processing

17.1 mRNA Capping

Eukaryotic mRNAs are capped at the 5' end with an "inverted" methyl guanine. This process involves three steps. The first involves removing a phosphate from the 5' end of the mRNA transcript with **RNA triphosphatase**. The β-phosphate then acts as a nucleophile to attack the α-phosphate of a GTP nucleotide catalyzed by **guanylyl transferase**. This displaces pyrophosphate and creates a 5'-5' linkage (the "inverted" guanine). Lastly, **methyl transferase** methylates the 7 position on guanine. This machinery is recruited by the phosphorylated CTD tail of the elongating eukaryotic RNA Polymerase II.

17.2 RNA Splicing

The Eukaryotic RNA that is first transcribed is called **pre-mRNA**, and still contains non-expressed **introns** in addition to the expressed **exons**. pre-mRNA undergoes **RNA splicing** to excise the introns and produce the final mRNA strand. Many pre-mRNAs, in fact, have multiple splice products (**alternative splicing**).

The process of splicing is fairly universal. The 5' end (**5' splice site**) of the intron and the 3' end (**3' splice site**) of the intron and an internal site (**branch site**) in the intron contain sequences which signal for splicing activity. A **transesterification** reaction occurs where the 2'-hydroxyl group of the branch site acts as a nucleophile, attacking the phosphate linkage at the 5' splice site, thus creating a **lariat**, or a loop in the intron, and liberating the exon 5' to the intron. The freed 3'-hydroxyl of the exon then attacks the phosphodiester linkage at the 3' splice site, thus linking the 5' splice site to the 3' splice site, and excising the intron, now a lariat structure. The increase in entropy from the production of the freed lariat is the driving force of the reaction and the reverse is also prevented by rapid degradation of free RNA. In alternative splicing, some exons are left out, and the 5' exon may attack a later 3' splice site (**trans-splicing**).

There are three categories of splicing. The first two involve the reaction as described above, and the third involves a different mechanism. The introns involving the splicing reaction described above are **Group 2 Introns**. The first type of Group 2 Intron uses a mostly-RNA machinery called the **spliceosome** which catalytically binds and induces the structural changes which lead to lariat formation and splicing. The second type of Group 2 intron splicing is self-splicing (the RNA of the intron adopts the appropriate structures needed), and needs no spliceosome.

The third category are the **Group 1 Introns** which are introns which self-splice and release linear introns rather than lariats. The mechanism by which this occurs is that the introns form a RNA structure capable of binding a guanine nucleotide. The intron then positions the guanine nucleotide to attack the 5' splice site, bypassing the lariat intermediate. This is followed by the usual 5' splice site attacking the 3' splice site.

Splicing errors are rare for two reasons. First, the splicing machinery is recruited as transcription occurs, making it less likely that the spliceosome will miss an exon. Secondly, exons themselves are oftentimes bound by **exonic splicing enhancers (ESEs)** which help to recruit the spliceosome to the correct position.

Alternative splicing is regulated by ESEs and also **exonic splicing suppressors (ESSs)** as well as intronic splicing enhancers and suppressors (ISE's and ISS's) which can be induced in response to different stimuli and thus favor different splice products under different circumstances.

Splicing has several evolutionary implications/concepts:

1. Self-splicing introns suggest that these elements were introduced externally as they are, like viruses, able to replicate themselves on their own and are able to incorporate into the genome
2. The spliceosome's dependence on RNA for catalytic activity lends support to the RNA world hypothesis and the idea that ribozymes are prevalent in nature
3. The presence of splicing allows for **exon shuffling** which allows for domains to be mixed and matched allowing for greater evolutionary diversity in protein structure and makeup
4. Splicing is widely believed to be the reason why the genomes of higher organisms (ie mammals) are not much longer than the genomes of lower organisms.

18 Translation

18.1 RNA

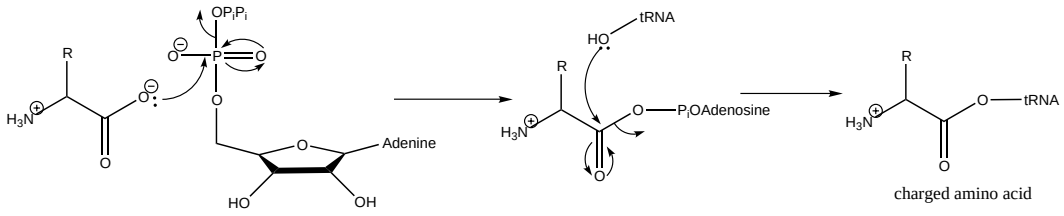
Messenger RNA is read by the ribosomal translation machinery. As codons are groups of three nucleotides, there are three possible ways to translate any given sequence. Only specified **open reading frames** with a specific start location and a stop location are translated. Most mRNA strands to be translated only have one open reading frame (**monocistronic**), while those with multiple open reading frames are **polycistronic**.

In prokaryotes, the mRNA oftentimes bears a **Shine-Dalgarno Sequence** also known as a **Ribosome Binding Sequence (RBS)** which is complementary to a sequence in the ribosomal RNA, directing ribosomes to translate the strand.

Eukaryotic mRNA recruits the translational machinery in three ways. First, the inverted methyl-guanine cap on the RNA recruits ribosomal machinery. Secondly, many eukaryotic mRNAs bear a **Kozak Sequence** (which consists of a purine three bases upstream and a guanine immediately downstream of the start codon) which increases translation efficiency. Lastly, the poly-adenylated tail of eukaryotic mRNAs promotes translation efficiency.

In addition to mRNA, translation requires an adaptor molecule to couple codon recognition with peptide chemistry. These are the **transfer RNAs** (tRNAs). These structures are mostly single-stranded, employ non-conventional bases (ie pseudouridine and dihydrouridine) and adopt a cloverleaf-like secondary structure and an L-like three-dimensional structure. At one end of the tRNA is a site which recognizes the codon (the **anticodon**). Although the sequences of tRNAs vary greatly, the 3' end of the tRNA always ends with 5'-CCA-3'. The terminal adenine is the position where **charged** tRNAs have an amino acid attached.

The process of charging an amino acid is two-fold but both steps are catalyzed by **tRNA synthetases**. In the first step, the carboxy terminus of an amino acid is used as a nucleophile to attack the α-phosphate of an ATP, displacing pyrophosphate, and creating an adenylated amino acid. This complex is then attacked by the 2' or 3' hydroxyl (it can rapidly equilibrate between both) of the tRNA terminal adenine, charging the tRNA in the process. One tRNA synthetase is responsible for one amino acid, so many tRNA synthetases recognize multiple tRNAs. This charging process is made accurate by multiple contacts between the enzyme and the tRNA and multi-step proofreading and is necessary because ribosomes cannot discern correct versus incorrectly charged tRNAs.



18.2 Ribosome

The ribosome is a very large complex of RNA and protein. It consists of two components — a large and small subunit. The small subunit is where codon recognition takes place and the large subunit is where the catalytic machinery is housed. The **ribosome cycle** refers to the formation of the ribosome-mRNA complex and the dissociation once the polypeptide chain has been formed:

1. The small subunit associates with the mRNA
2. The small subunit recruits the large subunit
3. After translation occurs, the complex dissociates

Multiple ribosomes can translate a given mRNA strand at any time, allowing for greater efficiency of translation and explains the relatively little mRNA relative to the number of ribosomes in cells.

18.3 Chemistry

Peptide bond formation chemistry is fairly simple. As with DNA and RNA production, peptide bond formation is directional. New amino acids are added at the C terminus of a growing polypeptide chain. The catalysis is handled by the RNA in the ribosome which positions the tRNA bearing the polypeptide chain next to the tRNA charged with the next amino acid. The amino group of this amino acid then attacks the acyl linkage between the polypeptide chain and the tRNA, transferring the new polypeptide to the tRNA formerly only bearing an amino acid (the **peptidyl transferase reaction**).

18.4 Initiation of Translation

The initiation of translation relies on the small subunit of the ribosome binding in the correct reading frame. This is mediated in prokaryotes by binding to the RBS. This binding places the start codon in the **P site** of the ribosome. The P site is responsible for holding the tRNA which holds the growing polypeptide chain. The adjacent **A site** is the amino site, where new charged tRNAs enter the ribosome and perform peptide chemistry, and the **E site** next to that is the exit site, where used tRNAs exit the channel. Immediately following ribosome binding to the mRNA, an **initiator tRNA** bearing **N-formyl methionine**, a modified amino acid, enters the P site to become the first amino acid in the polypeptide chain. The initiation factors **IF1**, **IF2**, and **IF3** are then responsible for initiating translation. The three factors prevent association of the large subunit (which does not associate with the small subunit until after initiation is complete), prevents other tRNAs from associating with the mRNA, and facilitates the association of the initiator tRNA, the small subunit, and the mRNA. When the initiator tRNA and the mRNA successfully pair, a change in conformation displaces IF3 which then allows the large subunit to associate with the small subunit. The resultant ribosome complex causes IF2, a GTPase bound to GTP during initiation, to hydrolyze GTP to GDP. IF2-GDP has a much weaker affinity for the ribosome complex and also becomes displaced which releases IF2 and IF1 from the ribosome, thus completing prokaryotic initiation.

In eukaryotes, the small unit of the ribosome is recruited by the mRNA cap. The small unit, already pre-bound to an initiator tRNA prior to association with the mRNA, then searches for the start site and then initiates translation there. The proteins involved are much greater in number and in complexity, but lead to a similar sequence of events except that the first amino acid is methionine (not fMet) and there are two GTP-bound GTPases which are associated with the complex rather than the one in prokaryotes. The scanning for the start site is driven by ATP hydrolysis by one of the many initiation factors. Binding of the initiation factor then induces the GTPase activity which displaces the initiation factors and allows the large subunit to associate and end initiation.

Eukaryotes increase translation efficiency through the use of the poly A tail on mRNAs. **Poly A Binding Protein** bends the mRNA into a circle allowing initiation factors associated with the poly A tail to interact with the initiation factors which associate with the guanine cap of the mRNA. This stabilized circle allows ribosomes to quickly "jump back on" to the mRNA after finishing one round of translation, thus increasing translation efficiency.

18.5 Elongation

The elongation of the polypeptide chain requires the translocation of the ribosome such that the tRNAs move over one step: from A to P, P to E. This translocation is assisted by elongation factors.

EF-Tu is a GTPase which, while bound to GTP has high affinity for tRNA and loses this affinity when bound to GDP. EF-Tu is one of the critical mechanisms by which accuracy is maintained during translation, as a codon-anticodon mispairing decreases the GTPase's efficiency and a bound EF-Tu blocks peptide coupling, thus preventing the wrong peptide bond from forming.

The ribosome itself also contributes to increasing accuracy by positioning ribonucleotides in the minor groove of the codon-anticodon pairing which can recognize when the pairing is incomplete. Additionally, the tRNAs upon binding, in order to properly form a peptide bond must be able to rotate in a process called **accomodation**. Incorrect tRNA base pairing leads to the tRNA being displaced from the ribosome during accomodation.

EF-G is another GTPase which, when bound to GTP, is able to bind to ribosomes following peptide bond formation. Upon binding, it hydrolyzes GTP to assist in the translocation process, pushing both tRNAs and also moving the mRNA. **EF-Ts** then replaces the GDP bound to EF-G with GTP so that the cycle can continue (and also on EF-Tu).

EF-Tu, EF-G, and EF-Ts have very similar counterparts in eukaryotes.

18.6 Termination

Translation termination occurs at stop codons. These codons are not recognized by tRNA but by proteins called **release factors**. There are two classes of release factors. **Class I release factors** recognize the stop codons themselves and trigger hydrolysis of the peptidylacyl-linkage of the P-site tRNA to free the polypeptide. They resemble tRNAs in structure in order to fill out this role. **Class II release factors** induce the dissociation of the Class I release factor from the ribosome. They are also GTP regulated although they have a stronger affinity for GDP so are found in the GDP bound form. The GDP bound form has a high affinity for the ribosome-Class I release factor complex. This binding causes the Class II release factor to become bound to GTP instead of GDP. This transition displaces the Class I factor. Loss of the Class I factor stimulates the hydrolysis of the bound GTP which leads to the GDP bound form of the protein which dissociates from the ribosome.

A **ribosome recycling factor** (RRF) recruits EF-G which uses GTP hydrolysis to dissociate the ribosome, tRNA, and mRNA from each other and to allow the ribosome to translate another transcript and the tRNAs to become acylated again.

18.7 Energy

One common strand here is the use of ATP hydrolysis and GTP hydrolysis to energetically push the process of translation forward. Every extra amino acid adds a cost of two GTP (one for EF-Tu, one for EF-G) and one ATP (for the tRNA synthetase to charge the tRNA). The initiation itself uses GTP and ATP hydrolysis to set up the complex and to scan for the start codon and to properly place the appropriate tRNAs there. Termination also requires GTP hydrolysis as the RRF step involves a GTP and the Class II factor uses one as well.

Part V
Gene Regulation

19 Gene Regulation in Prokaryotes

19.1 Transcriptional Control

Much of expression regulation in living organisms occurs at the level of transcription, oftentimes in the initiation phase. Proteins which become activated given specific signals are used to enhance (**activators**) or repress (**repressors**) transcription by affecting the steps in initiation.

- **Simple Activation/Repression:** The most elementary of transcriptional control are DNA binding proteins that recognize DNA sequence motifs near the promoter. These proteins then either recruit or prevent the binding of RNA polymerase (and hence modulate transcriptional levels relative to the **basal** levels by affecting the time the RNA polymerase is at a given promoter)
- **Allostery:** Certain proteins can also modulate transcription by affecting RNA polymerase conversion to the open configuration or the ability to undergo promoter escape
- **Action-at-a-Distance:** Many proteins which engage in simple recruitment/repression or in allostery bind to DNA sequences removed from the promoter. These interactions are mediated by the bending or looping of DNA and various proteins can facilitate the proper DNA bending to facilitate transcription.
- **Cooperativity:** This is not a regulatory method per se, but the effects of various regulatory proteins can be enhanced through cooperativity, whereby the effects of multiple proteins which interact with RNA Polymerase can be enhanced if they interact with one another. Cooperativity leads to switch-like behavior for transcription as genes will be transcribed reasonably poorly in the absence of full cooperativity and will quickly turn on as the binding of some factors recruits others. Cooperativity also allows the integration of multiple control signals to control one gene's transcription.
- **Combinatorial Control:** Also not a regulatory method per se, this refers to the fact that a single protein can act as a regulator for multiple genes
- **Alternative σ factors:** As mentioned before, the σ factor is responsible for promoter selectivity. The most prominent σ factor in *E. coli* is σ^{70}. Under heat-shock conditions, *E. coli* will switch to σ^{32}. Other alternative σ factors can direct RNA polymerase towards different promoter elements.
- **Altering Conformation of DNA:** A few transcriptional regulators are known to affect RNA polymerase by altering the structure of DNA in such a way to encourage binding or activity of RNA polymerase. **MerR**, for example, activates transcription of *merT*, a gene responsible for a cell coping with high levels of mercury. In the absence of mercury, MerR stabilizes the promoter which is structured so that the -35 and -10 elements of the promoter are not optimally placed to lead to RNA Polymerase binding. In the presence of mercury ions, however, MerR twists the promoter DNA forcing it into a more optimal configuration.
- **Preventing RNA Polymerase from Escaping the Promoter:** Some repressors act by preventing RNA polymerase from escaping the promoter by binding very strongly to RNA polymerase. These bind, much as one would expect activators to, upstream of the promoter. However, these proteins are able to bind tightly enough to prevent RNA polymerase from leaving the promoter.
- **Co-Repressors and Co-Activators:** These terms refer to substances which are needed for repression or activation but which are not DNA-binding proteins and do not interact with the transcriptional machinery themselves.
- **Dual Activator-Repressor:** Some proteins, like MerR alluded to above, act as both activators and repressors which can recruit/activate RNA polymerase under some conditions, but block transcription in others.
- **Attenuation:** Many bacteria use attenuation to regulate genes clusters which encode the enzymes needed to synthesize certain amino acids. This is done by preventing or enabling the formation of an early termination hairpin structure given different concentrations of the amino acid in question. There are other systems of control that are similar to attenuation which use mRNA structure responsive to different conditions to block transcription or translation.
- **Antitermination:** Antitermination is not well-understood, but involves the use of proteins which recognize stretches of mRNA or DNA during transcription and become loaded onto RNA polymerase as a result. The RNA polymerase-protein combination becomes resistant to certain termination sequences, thus acting as another level of transcriptional control.

- **Retroregulation:** Retroregulation occurs when the mRNA transcribed downstream of the gene being regulated is capable of recruiting nuclease activity. Thus, the mRNA is degraded before it can be translated in a step that occurs after transcription of the gene (and moves backwards).

19.2 Classical Example: *lac* Operon

The ***lac* operon system** of *E. coli* refers to the regulatory system which controls the transcription of the polycistronic mRNA which encodes β-galactosidase (the specific gene is lacZ), an enzyme which metabolizes lactose. The other two genes on the mRNA are lacY and lacA which encode transporters that import lactose. The lac operon is designed so that these three genes are only active in the absence of glucose and in the presence of lactose. Jacob and Monod did a classical series of experiments where they found that control of the system required two sequences, one which could only act in cis and one which could act in trans relative to the gene in question.

It turns out that the sequence capable of acting in trans coded for a regulatory protein and the sequence only capable of acting in cis was the sequence the regulatory protein recognized. In the lac system, the **lac repressor** is the protein which binds to the DNA sequence, called the **lac operator**, which is close enough to the promoter that the binding activity prevents RNA polymerase from binding. The lac repressor, upon binding to allolactose, a lactose metabolite, becomes unable to bind the operator and hence allows for transcription. The cis and trans conclusions then make sense, as the operator can only control the expression of one gene (cis) while a diffusible protein can act on any gene having the correct protein binding sequence.

The protein responsible for glucose detection in *E. coli* is called **CAP**. CAP binds cAMP, a sign of low glucose levels, and this binding enables it to bind to a site upstream of the promoter. CAP then interacts with the C-terminal tail of RNA polymerase (the region which normally interacts with the UP element which the lac promoter lacks) to recruit RNA polymerase.

19.3 Classical Example: Bacteriophage SPO1 Lifecycle Regulation

The **bacteriophage SPO1** infects *B. subtilis*. In order to produce viral progeny, a specific subset of genes must be expressed at different times. SPO1 regulates this via alternative σ factors. The early-phase SPO1 genes bear a promoter which is recognized by *B. subtilis* σ^{70} . One of these genes, gene 28, is an alternative σ factor which later displaces σ^{70} and directs transcription of the middle-phase genes. One of these genes, in turn, encodes yet another alternative σ factor which directs RNA polymerase towards the viral late phase genes.

19.4 Classical Example: NtrC

NtrC is a protein which regulates transcription of *glnA*, a gene responsible for nitrogen metabolism under low nitrogen conditions. NtrC showcases multiple forms of regulation. Under low nitrogen conditions, it is phosphorylated, exposing its DNA binding domain. NtrC binds four sites upstream of the promoter, thus binding is highly cooperative and interaction with RNA polymerase requires the looping of DNA. NtrC does not recruit RNA polymerase per se, but induces an allosteric change in RNA polymerase, causing it to associate with σ^{54} which directs transcription to the *glnA* promoter. Finally, NtrC also has an ATPase activity which, upon hydrolysis of ATP, induces RNA polymerase to switch from the closed conformation to the open conformation and activating transcription.

19.5 Classical Example: *araBAD* System

The ***araBAD* system** controls the genes responsible for arabinose metabolism. **araC**, the regulatory protein, binds as a dimer in the presence of arabinose to two sites, $araI_1$ and $araI_2$, upstream of the promoter to recruit RNA polymerase. In the absence of arabinose, araC then binds to $araI_1$ and a site many bases upstream of the promoter $araO_2$, causing the DNA to loop, and also preventing RNA polymerase recruitment. The degree of induction is very large which is why this system is used in many expression vectors.

19.6 Classical Example: *trp* Attenuation

The *trp* cluster of genes encode the enzymes needed for tryptophan biosynthesis and is regulated by a tryptophan-dependent repressor-operator system. The system also uses attenuation. Recall that in bacteria, translation occurs during mRNA transcription. The polycistronic mRNA which has the *trp* genes encodes a dummy **leader peptide** before the genes. This leader peptide is unique in that it has two tryptophan residues one after another. The mRNA encoding these two residues are part of a stretch of mRNA 5' of a stretch of mRNA that it can form a hairpin loop with. This latter stretch can also form a hairpin structure with another stretch of mRNA 3' to it which is itself a part of the termination hairpin loop of the leader mRNA. In a tryptophan-poor environment, the ribosome stalls at the double-tryptophan-encoding portion of the mRNA. This stalling causes the ribosome to block the 5' stretch of mRNA complementary to the middle stretch, enabling the middle stretch to form a hairpin loop with the 3' stretch. This hairpin loops prevents the formation of the terminator hairpin (as the 3' stretch is not available) and allows RNA polymerase to continue transcription. In a high-tryptophan environment, the ribosome does not stall at that first stretch, and proceeds. This allows the 5' stretch to form a hairpin loop with the middle stretch, allowing the 3' stretch to form the termination hairpin loop which then blocks transcription of the remaining *trp* genes. A similar system is used by various bacteria for other amino-acid biosynthesis systems.

This method of control is similar to the **riboswitch** method of transcriptional control where a leader mRNA can adopt a complex tertiary structure which can bind various molecules (ie methionine). The binding then controls the formation or prevention of a termination hairpin or the formation of RNA structures which are difficult to transcribe through or terminate translation.

This is also similar to the system which regulates ribosomal protein levels. Ribosomal proteins bind to the mRNA's which encode them in such a way as to prevent ribosomes from translating them.

19.7 Classical Example: Phage λ

Bacteriophage λ upon infection can either destroy the cell lytically or enter a dormant lysogenic phase. This decision is determined by a variety of factors including the nutrient density in the surrounding media (high nutrient density, lytic, low nutrient density, no point in being lytic). The lysogenic phase can also be exited given certain stimuli, such as DNA damage (which threatens the existence of the phage).

This complex behavior is regulated by a relatively simple system which, at its core, involves three promoters: P_L (left), P_R (right), and P_{RM} (repressor maintenance). P_L and P_R are strong promoters which are capable of inducing strong transcription and flank the weak promoter P_{RM} and its target gene, ***cI*** which encodes the λ **repressor**. P_R also directs the transcription of **Cro**. Lytic behavior occurs when P_L and P_R are active but P_{RM} is not, while lysogenic behavior occurs when P_L and P_R are inactive but P_{RM} is active.

The key to this system is that there are two operators (to the left and to the right) which regulate all three promoters. Each operator has three protein binding sites, for a total of six binding sites. O_{R1}, O_{R2}, O_{R3}, O_{L1}, O_{L2}, and O_{L3} are each capable of binding Cro and λ repressor. While Cro dimers can only act as repressors, λ repressor can bind as a dimer, a tetramer, and even an octamer and depending on its position and the units involved can act as either an activator or a repressor.

The operator of most interest is the one on the right as it simultaneously regulates P_{RM} and P_R in such a way that a bound O_{R3} blocks P_{RM} and bound O_{R2} and O_{R1} blocks P_R (and a similar pattern blocks P_L), but recruits RNA polymerase to P_{RM}. λ repressor binds O_{R1} readily but binds O_{R2} and O_{R3} weakly, although due to cooperativity, a λ repressor dimer bound at O_{R1} can recruit another dimer to O_{R2}, while Cro binds O_{R3} readily but binds O_{R1} and O_{R2} weakly. Therefore, in the lytic phase, Cro repressor binds O_{R3}, but not O_{R1} and O_{R2}, thus blocking P_{RM} and enabling P_R and P_L. In the lysogenic phase, λ repressor binds O_{R1} and O_{R2}, but not O_{R3}, thus blocking P_R (and also P_L) and enabling P_{RM}.

This system, then, is capable of a very fine-tuned system of control of the levels of λ repressor. λ repressor acts as its own inducer, as its presence at O_{R1} and O_{R2} stimulates transcription (**positive autoregulation**). If repressor levels are too high, then it will bind O_{R3} and shut down its own transcription levels (**negative autoregulation**).

Thus, if a virus is already in the lysogenic phase, **induction** (to the lytic phase) merely involves

changing this balance. In conditions of DNA damage, for example, cellular processes lead to the degradation of λ repressor's self-association domains which prevent it from being able to bind O_{R1} and O_{R2}, thus freeing P_R (and by freeing the left positions, P_L) and stopping transcription at P_{RM}.

The original decision to enter the lysogenic phase or the lytic phase is governed by two additional genes **cII** and **cIII** (which are genes under control of P_R and P_L respectively). cII acts as a transcriptional activator, activating the P_{RE} (repressor establishment) promoter which, like P_{RM}, is capable of inducing cI/λ repressor (thus λ repressor autoregulation can only be *maintained* by P_{RM} after it is *established* by P_{RE} and cII). As both cII and Cro are under the control of the same promoter and exert opposite effects, the lysogenic/lytic decision is governed by the balance between Cro and cII. cII is degraded by cellular proteases in times of high nutrients (and cIII prevents degradation of cII), thus biasing the lytic phase, and is kept stable in times of low nutrients, thus biasing the lysogenic phase.

20 Gene Regulation in Eukaryotes

20.1 Different but the Same

For the most part, transcriptional regulation in eukaryotes resembles that of prokaryotes except for in degree of complexity. There are typically more regulatory protein binding sites (these sites are called **enhancers** or **regulatory sequences**) and they are spread out farther and even downstream of transcription initiation and the regulatory proteins themselves are more diverse (structure-wise), but they are governed by the same principles (as shown by famous experiments with the *gal4* system and yeast-two-hybrid system), except, however, that instead of directly recruiting or interacting with RNA Polymerase, most regulatory proteins only interact or recruit elements of the transcriptional machinery.

20.2 Notable Differences in Regulation of Transcriptional Initiation

- **Histone Modification:** Eukaryotes possess histones which wind up nuclear DNA for packaging, and also for regulation of access to the nucleic material. Many regulatory sequences serve to bind proteins which promote histone remodelling. Histone acetyl transferases acetylate DNA and encourage transcription, while histone deacetylases do the opposite.

- **Insulators:** As many regulatory sequences are found very far from the gene in question, eukaryotes have evolved insulators, proteins which bind regulatory sequences and act to block regulatory interactions. Therefore, an activator will only activate the gene it is intended to target and not random genes nearby.

- **Locus Control Regions (LCR):** The expression of some gene clusters are regulated by a large group of regulatory sequences called Locus Control Regions (LCR) (ie globin cluster). The LCR for a gene cluster contains numerous enhancer sites, insulator sites, and histone remodelling protein binding sites which enable complex control of the gene clusters in question.

- **Architectural Proteins:** As the regulatory sites are oftentimes very far from the promoter in question, it is necessary to form large loops for enhancers to exert their effects. Architectural proteins such as *HMG-1* in the human interferon-β system are needed to facilitate the formation of such loops to facilitate cooperativity or other protein-protein interactions.

- **Transcriptional Repression:** Many prokaryotic repressors block transcription by physically blocking RNA polymerase from binding the promoter. Eukaryotic repressors rarely do this. Eukaryotic repression typically involves histone modification (de-acetylation or methylation), interacting with the transcription machinery to prevent proper transcription initiation, direct competition for enhancer binding site, and/or blocking the interaction between the enhancer and the transcriptional machinery.

- **Signal Transduction Pathways:** Stimuli is oftentimes relayed to the transcriptional level of control by signal transduction pathways which convert a raw signal (ie growth factor levels) into a biological one. In prokaryotes, this biological signal was oftentimes the regulator's

ability or inability to bind DNA. In eukaryotes, what is more common is enabling/preventing the regulator from acting by modifying the protein's structure (ie phosphorylation) or altering the nuclear import/export of the regulator.

- **Gene Silencing:** Gene silencing refers to the transcriptional shutdown of genes because of their location, not in response to a particular stimulus. It is mediated by histone modification, usually by histone deacetylation or methylation, or by methylating DNA directly. The former packs the chromatin together more tightly into **heterochromatin**, rendering the DNA inaccessible to transcriptional machinery, and the latter blocks the binding of proteins needed to initiate transcription and by recruiting histone remodelling complexes to do the same. This direct DNA methylation is the basis of **imprinting** where despite the presence of both a maternal and paternal allele, only one is active (the unmethylated one). These changes involving gene silencing are termed **epigenetic** because they are real and can be inherited through cell division but are not in the genetic code itself. These changes are oftentimes maintained by so-called maintenance methylases which methylate hemi-methylated DNA (from DNA replication).

20.3 Regulation of Post-Initiation Transcription

It should be noted first that all prokaryotic mechanisms of regulation which were dependent on co-transcriptional translation (ie attenuation) do not apply as translation and transcription in eukaryotes are separated.

- **Control of Transcriptional Elongation:** Like the bacterial anti-terminators, some eukaryotic genes require the binding or recruitment of factors to assist in transcriptional elongation. This includes **P-TEF**, a kinase which is recruited to some positions to phosphorylate the RNA polymerase CTD and allow it to pass sequences that would otherwise stall it.
- **Regulation of Splicing: Splicing enhancers** and **splicing repressors** can be activated given different cellular stimuli and can thus direct mRNA to be alternatively spliced at different levels under different conditions. This is the basis of *Drosophila* sex-determination. The expression of a master regulator gene is controlled by the ability of X-linked splicing enhancers to properly splice a specific mRNA transcript.
- **Translational Control:** Self-explanatory, but several genes are regulated on the level of translation. *Gcn4*, a yeast gene responsible for amino acid biosynthesis, is only translated in low amino acid conditions. This is regulated because the mRNA of Gcn4 contains 4 short open reading frames upstream of the Gcn4 gene itself. Ribosomes are unable to translate all 4 in low amino acid conditions and can thus reach the Gcn4 open reading frame, while in high amino acid conditions, the ribosome dissociates from the mRNA after translation is complete.

20.4 Classic Example: Yeast Cell Type Control

S. cerevisiae can be three types of cells: a, α, and diploid. The three types of cells have different expressions of genes. a type cells express a1, a a-type specific regulator of transcription, and Mcm1, a general regulator. α type cells express α1, α2, and Mcm1. Diploid cells express a1, Mcm1, and only α2.

In a cells, Mcm1 acts as an activator for a-specific genes and the absence of an activator prevents the activation of α-specific genes. In α cells, Mcm1 binds cooperatively with α1 at the enhancer for α-specific genes, thus turning them on. And α2 binds as a repressor at the enhancer for a-specific genes and blocks the capability of Mcm1 to activate a-specific genes, thus turning them off.

In diploid cells, the lack of α1 prevents the activation of α specific genes. The presence of α2 then inhibits a-specific genes (as in α cells) and binds with a1 to block haploid specific genes.

20.5 Classic Example: Glucocorticoid Receptor

The **glucorticoid receptor (GR)** is normally held in the cytoplasm. Upon binding its ligand, however, it is then transported to the nucleus where it can bind at various **glucocorticoid response**

elements (GREs) in the DNA. At some GREs, the GR adopts a conformation which allows it to recruit **CBP**, a histone acetylase and recruiter of transcriptional machinery used by many other enhancers thus activating transcription, and at other GREs, the GR adopts a conformation which allows it to recruit histone deacetylases and thus shutting down transcription.

20.6 Biological Diversity and Gene Regulation

It's commonly cited that the human genome is not much larger or more complicated than much simpler organisms. And even between species as different as humans to worms, most of the genes are clear homologues of one another. The primary explanation for the increase in complexity without great difference in genome itself is gene expression regulation.

Many genes appear to have been the products of **gene duplication**. This gene duplication is then followed by mutations in the primary structure, giving it slightly different function, or mutations in the regulatory sequences controlling the gene's expression. With the exception of bacteria, it seems that much of evolution has been driven by the latter. In eukaryotes this is further assisted by alternative splicing which can be used to "mix-and-match" gene components and by the fact that it is very commonplace for regulatory elements to be many kilobases away from the genes in question.

21 Viruses

21.1 Basics

Viruses are small parasites that employ the machinery of the cells they infect to reproduce, and in fact cannot reproduce without infecting cells. They are essentially bundles, called **nucleocapsids**, of genetic information (DNA or RNA) packaged in protein shells called **capsids**. These nucleocapsids may or may not have a lipid bilayer, called an **envelope**, which while derived from the infected cell if present contain viral rather than cellular glycoproteins and cell surface proteins, associated with them. These capsids and their envelopes often bear proteins and structures which assist in infection and in identifying cell types (ie by binding to cell surface markers of certain cell types) The study of viruses has allowed for the exploration of host immune responses as well as the molecular biology of all living things.

Capsids are usually assemblies of relatively few proteins into a three-dimensional structure. There are two main classes of capsid structure. The first involves simply wrapping the nucleic acids within a protein helical structure. The second involves assembling proteins into icosahedral structures (20 identical equilateral triangular faces in a quasi-spherical shape).

Simply mixing the nucleic information with the capsid proteins of many simple viruses promotes spontaneous self-assembly. More complicated viruses require specific assembly stages before mature virus are produced. Viruses have shown a remarkable ability to not only use cellular machinery, but to completely hijack the machinery to end the production of cellular components and only produce ones needed for viral assembly. This dependence on cellular machinery for infection and replication limits the **viral range**, the range of host cell types that a virus can infect, of most viruses. Nomenclature-wise, viruses that only infect bacteria are called **bacteriophage**, and viruses that only infect animal cells and plant cells are called (wait for it) **animal viruses** and **plant viruses**. Some plant viruses are able to infect and use insects as **vectors** for spread of infection.

Viruses demonstrate one of two general lifecycle types. The **lytic cycle** describes how viruses infect cells, use their nucleic information to synthesize viral proteins, replicate their nucleic information, assemble within the cell, and then lyse the cell to release new infectious viral particles. The alternative **lysogenic cycle** involves viruses infecting cells, but instead of culminating in lysis, these cycles culminate in the insertion of their own nucleic material into the genome of the host, as a **prophage**, where the DNA lies dormant and is replicated when the cell replicates. Under certain conditions, the prophage will then exit the nucleic material and then resume the lytic cycle. This control allows viruses to prevent themselves from exterminating a population and encourages faster growth when it is sustainable. Lysogenic viruses are oftentimes called **temperate viruses**.

21.2 Animal Virus

Bacteriophages have ready access to the nucleic information (there is no nucleus in a prokaryote) and much simpler regulatory systems to contend with. There are six classes of viruses divided mostly by type of nucleic information and process by which it is used:

1. **Class 1** - Class 1 viruses are viruses that use dsDNA as a source of nucleic information. The dsDNA of **Class 1a** viruses enters the nucleus and cellular machinery is employed to transcribe it to mRNA. The dsDNA of **Class 1b** viruses carry their own transcription enzymes and all of replication occurs in the cytoplasm.

2. **Class 2** - Class 2 viruses have ssDNA (selection of plus or minus strand is up to the virus and sometimes the virus doesn't care). This is then copied into dsDNA and then used to make mRNA.

3. **Class 3** - Class 3 viruses use dsRNA. The minus strand is then used to direct the synthesis of plus strand mRNA. To date, all Class 3 viruses have been shown to have **segmented genomes**, multiple dsRNA molecules each encoding only one or two polypeptides. They have their own enzymes to produce mRNA.

4. **Class 4** - Class 4 viruses use a single plus strand of RNA. This is used to create a minus strand which is then used to create more plus strand RNA which is then used as mRNA. **Class 4a** viruses synthesizes a single polypeptide form a long mRNA strand and then cleaves the polypeptide into the individual proteins. **Class 4b** viruses synthesize at least two strands of mRNA (at least one of which is the full-genome-length) all of which are translated into polyproteins.

5. **Class 5** - Class 5 viruses contain a single minus strand of RNA. The viruses each use a RNA-dependent RNA polymerase to synthesize mRNA for use. **Class 5a** viruses encode multiple genes on the one strand of mRNA, of which individual mRNA's encoding each of the genes are made. **Class 5b** viruses employ segmented genomes, although some of the RNA strands can encode for multiple peptides by having the ribosome shift reading frames.

6. **Class 6** - Class 6 viruses are more commonly referred to as **retroviruses**. Their capsids encapsulate two identical plus strands of RNA. These are then used to synthesize minus strand DNA via **reverse transcriptase**. This is then used to make plus strand DNA and the dsDNA is then integrated into the genome to enter a lysogenic-like phase as a provirus.