The content review is organized into two chapters is organized into two chapters:

1. Amino acids/peptides

2. Proteins (Including separation techniques)

Amino acids, peptides, and proteins are listed in the AAMC Content Guide in BOTH the biological and physical sciences tests. They are commonly emphasized on the MCAT. The following topics will be discussed here. Further topics on protein structure and function can be found throughout the various assessment tests available on our website.

Chapter 1: Amino Acids/Peptides

- ☐ Structure of canonical 20 amino acids
- ☐ Absolute configuration at the α position and dipolar ions
- ☐ Acid/Base chemistry and Titration curves
- ☐ Amino acids as precursors
- ☐ Focus on specific Amino acids, including cysteine, glycine, proline, aromatic, glutamine and glutamate, etc.
- ☐ Alanine-glucose cycle
- ☐ Hydrophilic and hydrophobic
- ☐ Amino acid metabolism: glucogenic vs ketogenic
- ☐ Amino acid transamination
- ☐ Nitrogen balance and the urea cycle
- ☐ Essential vs. Non essential amino acids
- ☐ Amino acid synthesis: Strecker and Gabriel

Chapter 2

Protein structure and function

Proteins

- ☐ **1°, 2° , 3° , 4° structure of proteins including Hemoglobin**
- ☐ **Denaturing and folding**
- ☐ **Hydrophobic interactions**

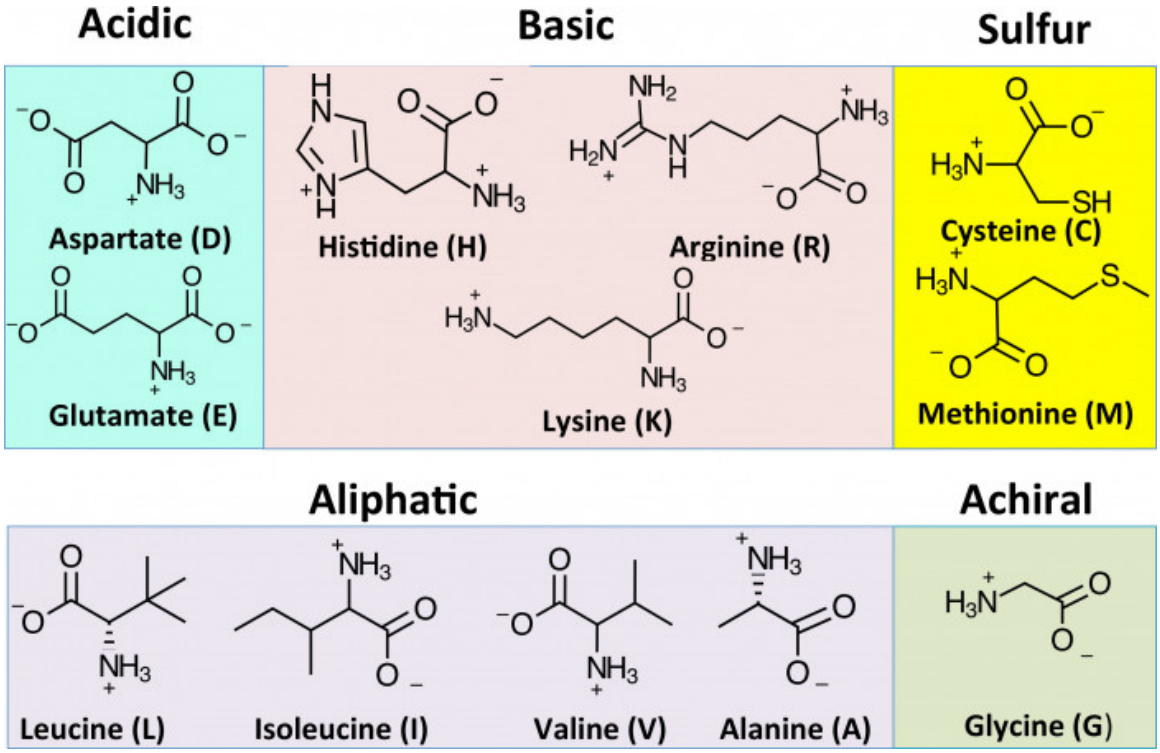
- ☐ **Hill coefficient and ligand binding (Hemoglobin)**
- ☐ **Amino acid separation & Protein Purification**

Introduction

Structure and function: The key paradigm.

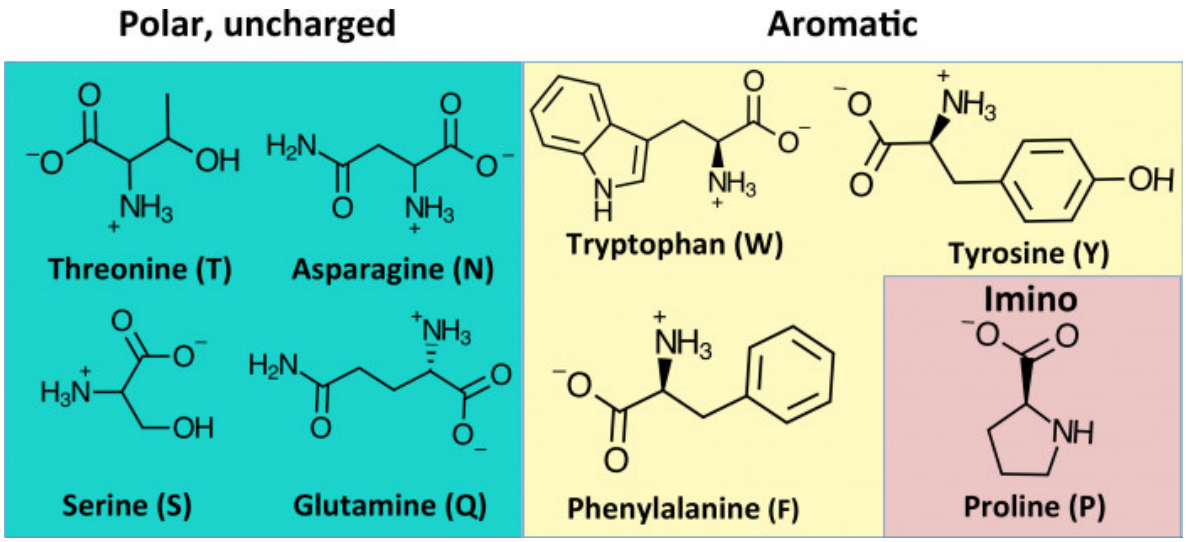
Enzymes and other proteins fold into their respective three dimensional shapes that dictate consequent cellular function(s). Numerous pharmaceutical reagents exploit structure/function relationships. In particular, development of specific drugs that bond as agonists (i.e. activators) and antagonists (i.e. inhibitors) has been used in clinical medicine for decades.

The vast majority of pharmaceutical drugs target cellular proteins, ranging from enzyme inhibition to targeting receptors as agonists and antagonists. As proteins are composed of linear chains of amino acids that fold into a functional three dimensional conformation, understanding amino acid structure, particularly in the context of protein function, is a central objective in biochemistry and we know the MCAT loves this topic. Further, amino acids and their derivatives play important roles outside of protein structure and function in diverse processes. This is because many amino acids are turned into important biomolecules ranging from serotonin (neurotransmission) to histamine (allergic response, vasodilation). It is therefore clear to see why amino acid/protein structure and function formulates a very widely covered



topic area on the MCAT.

The primary amino acid sequence of proteins is determined via the mRNA sequence and assembled during ribosomal translation. This is a major theme of Content Category 1B. The chemical diversity of amino acids and their numerous sequence combinations allows for the generation of tertiary folding structures that adopt a plethora of conformations and active sites specifying any number of functions. The canonical 20 amino acids are presented below.

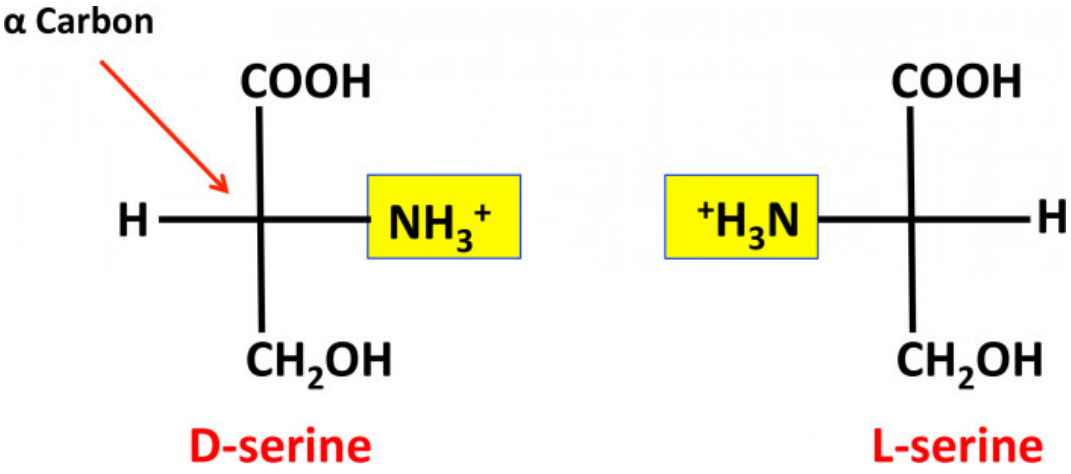


Absolute configuration at the α carbon

Each of twenty canonical amino acids, with the exception of glycine, is chiral. The structures are shown. Chirality, or handedness, is an important aspect of biology and chemistry. It is the subject of the "Principles of Stereochemistry" learning/testing module. Stereochemistry is presented in the Chemical and Physical Foundations of Biological Systems in Section 5B of the AAMC MCAT outline.

With the exception of glycine, each of the amino acids incorporated into proteins displays optical activity. This is based upon the presence of an asymmetric carbon center (α carbon). For amino acids, there are two possible stereogenic states: L and D that are related to each other as non-superimposable mirror images. Only the L-isomer is incorporated into proteins in humans, but bacteria use D amino acids in their peptidoglycan walls.

Amino acids are often represented as Fischer projections as shown. In a Fischer projection, the D isomer has the amino function on the right and its L enantiomer has the amino group on the left. This is shown with serine.



Amino acids as dipolar ions

By definition each amino acid minimally contains a single carboxyl and a single primary amine group. The structures of the common twenty amino acids are shown and are grouped into those possessing similar chemical features. Note that proline is a secondary amine and it therefore technically not an amino acid, but rather an "imino" acid.

Do you have to memorize the structures of the amino acids? Perhaps, but at the minimum you should be able to recognize a given amino acid and immediately associate its structure with function, particularly with respect to its general classification as acidic/basic, hydrophobic/hydrophilic, sulfur containing, posttranslational modifications etc. That is, you must understand the chemistry of the amino acid side chains.

Acid/Base chemistry

Amino acids are diprotic Bronsted-Lowry acids because they dissociate protons. After dissociation, the protonated species is in equilibrium with its conjugate base. You should readily be able to distinguish Bronsted-Lowry acids from Lewis acids. In contrast to Bronsted-Lowry acids, the dissociation of protons is not important. Rather, Lewis acids accept electron pairs from donor compounds.

Bronsted-Lowry acids dissociate protons as shown in the equilibrium dissociation with a generic acid designated as HA. The amino and carboxyl protons are considered weak acids because the protons do not completely dissociate under equilibrium conditions. This is in contrast to strong acids that completely dissociate protons. Strong acids include HNO_3 (Nitric acid), H_2SO_4 (Sulfuric acid), HCl (Hydrochloric acid), and $HClO_4$ (Perchloric acid).

$$HA \longleftrightarrow H^+ + A^- \qquad K_a = \frac{[A^-][H^+]}{[HA]}$$

$$pK_a = -\log[K_a]$$

The extent of proton (H^+) ionization can be measured by determining the equilibrium constant K_a, which is defined in this case as the ratio of $[A^-][H^+]/[HA]$. The equilibrium constant of a reaction depends on various factors including the ionic strength of the buffer and the temperature. High K_a values signify the formation of more product(s) than reactant(s) at equilibrium. As K_a values can be very large or small, a more convenient and useful method of describing equilibrium is the "pK_a". By definition:

$pK_a = -\log[K_a]$

The pK_a of a typical carboxyl group is approximately 2.4 and that of a primary amino group is about 9.5. You might see other values such as the whole numbers 3.0 and 9.0. However, keep in mind that the true pK_a value is governed by the local environment and can change as a function of environment.

The pK_a value quantifies the strength of a Bronsted-Lowry acid in solution and is therefore related to the acid dissociation constant: $pK_a = -[\log]K_a$. You should be very familiar with the term pK_a and how it applies to the Henderson Hasselbach equation:

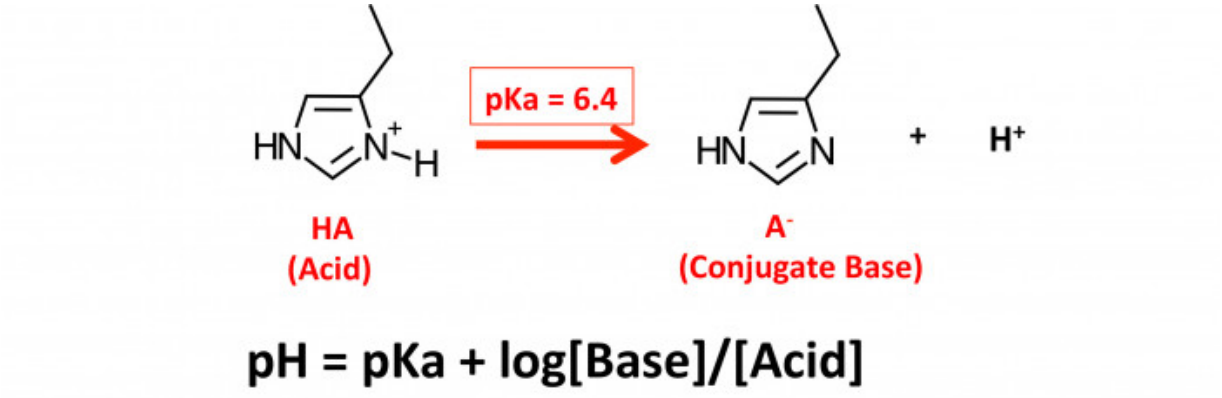
$pH = pK_a + \log[A^-]/[HA]$ or

$pH = pK_a + \log[BASE]/[ACID]$

This equation is an algebraic manipulation of the generic acid dissociation relationship: HA = H^+ + A-. Always appreciate that large pK_a values indicate that a higher pH is required to titrate (dissociate in this case) the proton. That is, a high pK_a value represents a weak acid.

Using the Henderson Hasselbach equation

The pK_a value of the imidazole side chain of histidine is ~ 6.4. This means that at physiological pH = 7.4, only 10 % (1/10) of the molecules in solution will exist in the charged, protonated form. This calculation is shown below in the image. You should be quite familiar with these acid-base calculations and Med-Pathway has multiple examples and applications of this in this section as well as the Passage Workbook.



Q: What is the fraction of histidine residues that are positively charged at pH = 7.4?

Solve for [Base]/[Acid] knowing that pKa = 6.4 and pH = 7.4.

pH – pKa = Log (X) X= [A^-][HA]

A: Antilog (pH-pKa) = antilog[log(X)] Antilog 1 = X ; X = 10
Therefore, as X= [Base][Acid], the ratio of [Base]/[Acid] = 10.

There are 10X as many His residues in conjugate base than His residues than in acidic, positively charged form.

Because you appreciate that amino acids are zwitterions at physiological pH, the overall charge will change as a function of pH. This is commonly observed with titration curves. For some amino acids, the side chains possess ionizable groups that contribute to the overall charge of the amino acid at a given pH.

We just saw this for histidine. In the context of protein primary structure (i.e. the sequence of amino acids as determined through the genetic code), only the side chain will be charged. The only exceptions are the N and C terminal amino acids.

Side chain charges are important for protein structure and function. The ionization states of amino acid side chains often make the difference between a functional enzyme and a catalytically inactive one. This important point will be seen over and over.

Once again, the ionization status of the carboxyl and amino groups depends on their respective pK_a values and remember that:

$pK_a = -\log [K_a]$

The pK_a is dependent on several factors including the pH, micro-environment, ionic strength, temperature, and the dielectric constant.

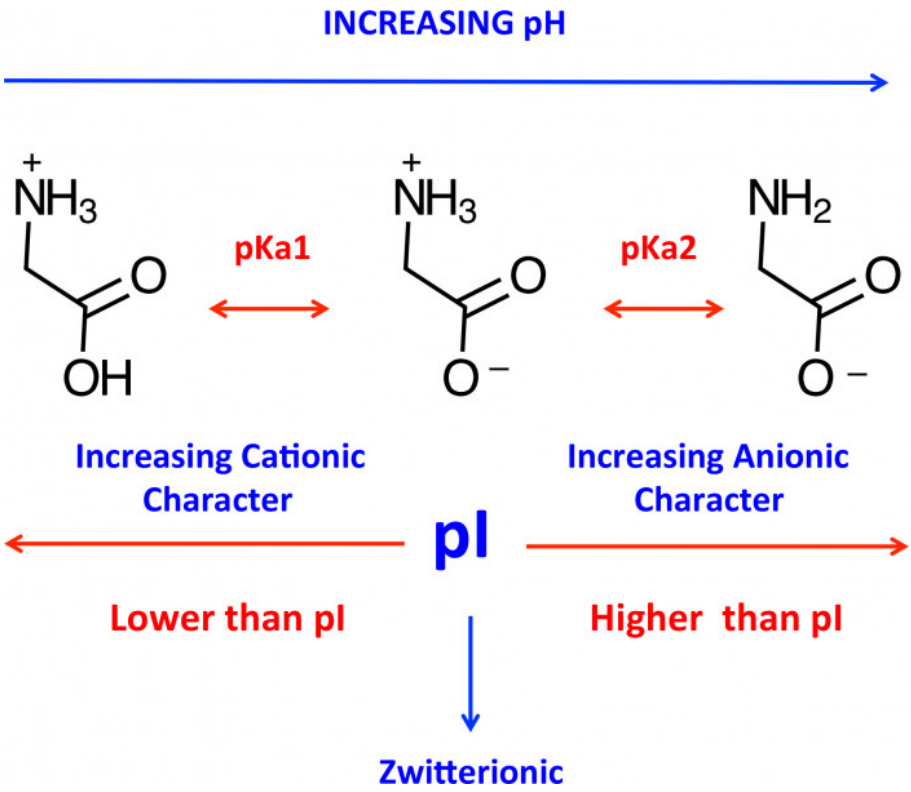
Titration of Alanine

As a free amino acid, alanine exists in three ionization states (A, B, and C), each with a distinct charge. At low pH, both the carboxyl and amino groups of alanine are both protonated (Form A, +1 net charge). As more hydroxide ion is added to the alanine solution, the pH rises until it reaches a value equal to pK_a1. At this point, ½ of the carboxyl groups will be protonated and the other half will be ionized. This is the ½ equivalence point and can be calculated from the Henderson Hasselbach equation. That is, when the pH = pK_a, the concentration of COOH = the concentration of the conjugate base [COO-], or think of it this way: the concentration of protonated acid (HA) is equivalent to the concentration of conjugated base [COO-]. Keep this in mind:

if pH = pK_a, then [HA] = [A-]

As the pH is raised through the addition of more equivalents of HO^-, the carboxyl group becomes completely titrated to COO^-, the conjugate base. This is form B. This pH represents the isoelectric point (pI), the point at which the net charge on the molecule is zero. Through the addition of more hydroxide, the pH will become equal to pK_a2. This is the pH at which ½ of the amino protons are titrated (i.e. form C).

The following graph shows how the overall charge of an amino acid changes as a function of the pH and the pI of the amino acid. In this case, the diprotic amino acid glycine is used as an example. Note that when the pI > pH, the overall charge is positive, but when the pH > pI, the overall charge is negative.



Amino acid classes

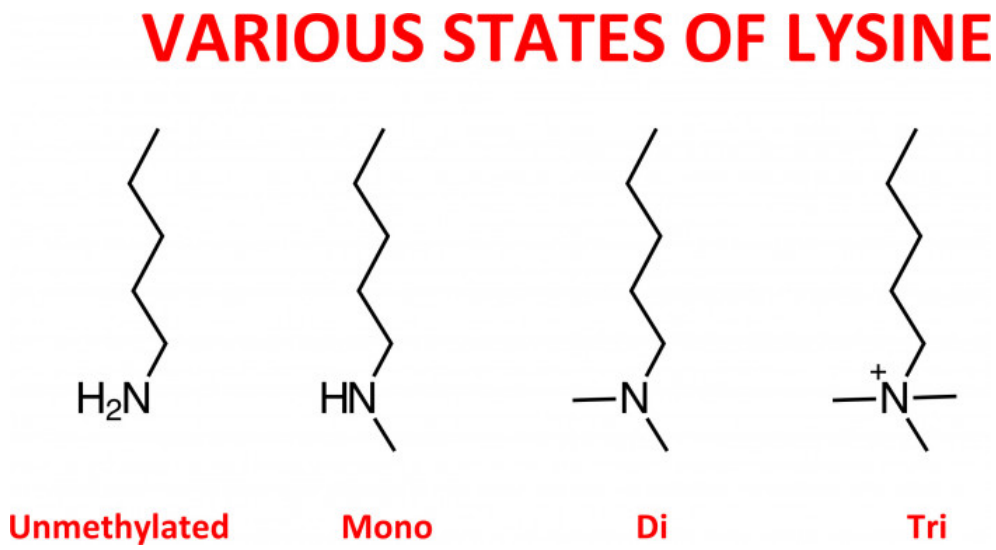
Acidic/Basic. Aspartic acid and glutamic acid possess carboxylic acid side chains with similar pK_a values usually designated as being anywhere between 2.0-3.0. The amino acid structure shows the carboxylate side chains as the ionized conjugate bases, the expected chemical state at physiological pH.

Lysine (ε-amino side chain) and arginine (guanadinium side chain) have basic side chains with pK_a values of approximately 11.0 and 13.0, respectively. The positively charged forms are shown in the table at the beginning of this section, as these are the states expected at the physiological pH of 7.4.

The pK_a of the side chain of histidine (imidazole) is commonly listed between 6.4 and 6.8. This pK_a value is the closest of all amino acids to physiological pH (7.4). Thus, at physiological pH a substantial fraction of the histidine side

chains are deprotonated and uncharged. We will see the important role that histidine plays in enzyme catalysis with the classic example of chymotrypsin.

In many cases, lysine methylation plays important roles in various biological properties, especially with respect to gene expression. Lysine can exist in four distinct states. This includes the familiar unmodified form as well as the mono-, di-, and tri-methylated forms. Each distinct methylation state can bind to various proteins. In the case of chromatin, enzymes that deposit methyl groups on various lysine moieties have been identified and shown to either facilitate or repress gene expression.



Roles of Arginine in Nitrogen regulation.

Arginine plays important roles in regulating nitrogen metabolism by multiple manners that have significant clinical impact and should be studied for the MCAT and beyond.

1. Precursor to Urea in the Urea Cycle

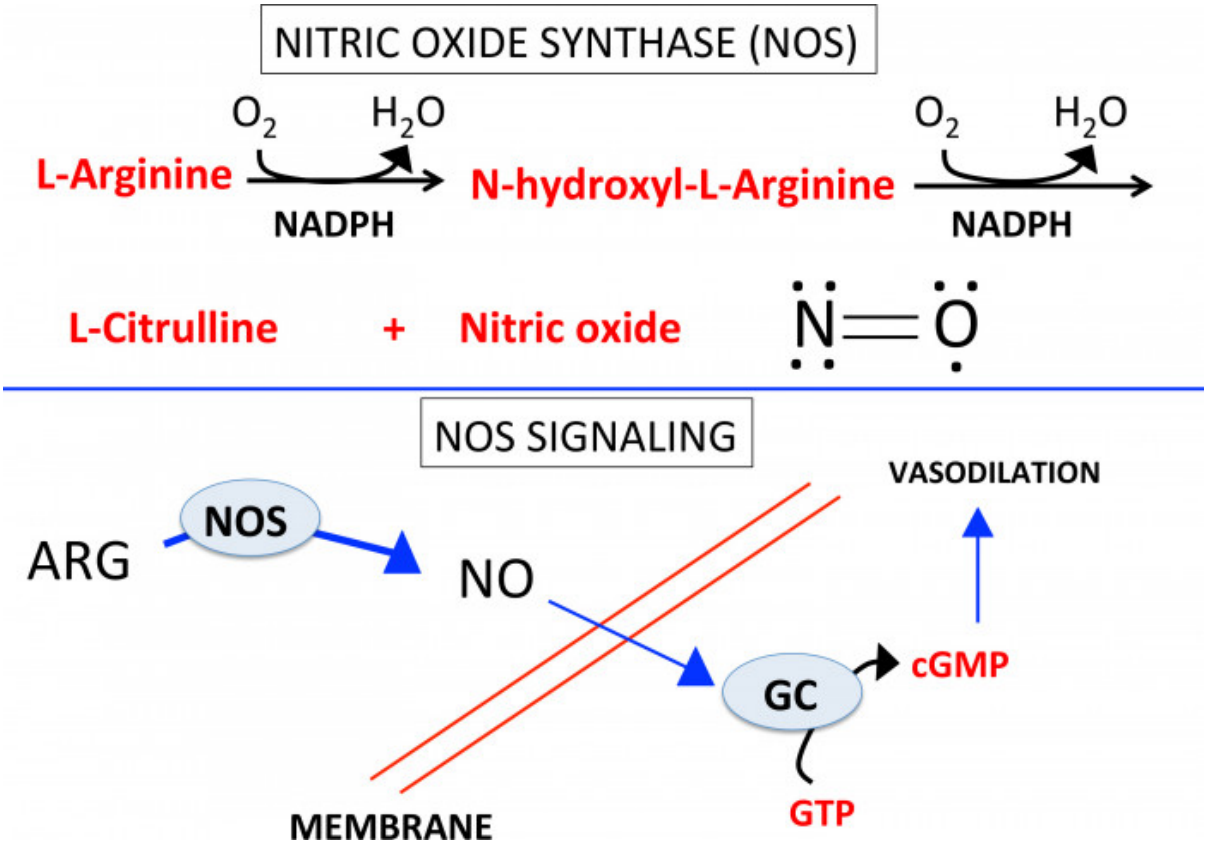
2. Required for the synthesis of nitric oxide

3. Required for synthesis of creatine and creatinine

The urea cycle will be discussed below.

Nitric oxide synthesis. Arginine is converted into L-citrulline and nitric oxide, a paramagnetic compound with an unpaired electron through the action of nitric oxide synthase (NOS). NO diffuses through cell membranes where it activates

soluble guanylyl cyclase (GC). This enzyme converts GTP into cGMP, a signaling molecule that is involved in multiple cellular responses including vasodilation.



Arginine and Creatinine.

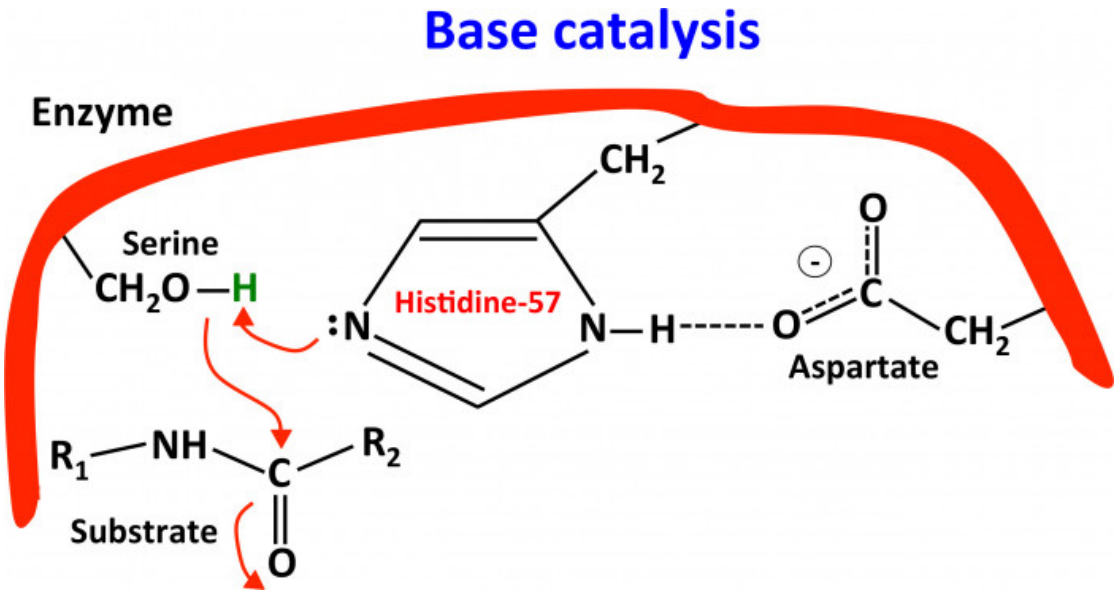
Glycine and arginine combine in the kidney to form guanidine acetate. This compound is shipped to the liver and methylated by SAM to form creatine. Creatine circulates to the brain, heart, and skeletal muscle. The fate of creatine is shown below:

Creatine-phosphate (P) is used an energy source because it phosphorylates ADP to generate ATP. Creatine-P is spontaneously converted into creatinine in

the muscle and brain. Creatinine production occurs at a fairly uniform rate as a function of muscle mass. This feature is often exploited by clinicians for assessing renal function. As very little creatinine is reabsorbed in the proximal tubule, an accurate assessment of the glomerular filtration rate can be derived from the creatinine clearance rate.

Acid/base side chain chemistry in catalysis.

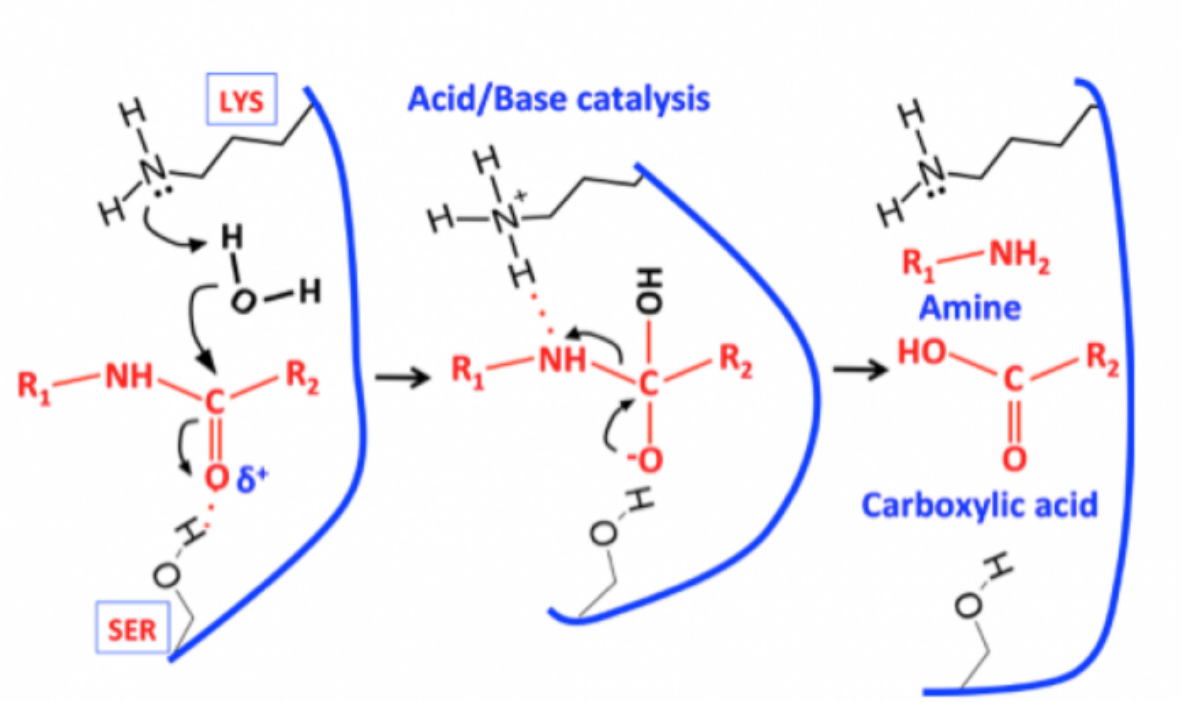
The ability of various amino acids to act as either acids or bases makes them key players at the active sites of enzymes. This is commonly seen on the MCAT. Despite the fact that there are numerous enzymes for the MCAT to test you on, the themes are always very similar. Classic examples include lysozyme and chymotrypsin. One such scenario of base catalysis is seen in the protease chymotrypsin, a pancreatic enzyme secreted into the duodenum during digestion.



As shown above, chymotrypsin exemplifies a "catalytic triad" where the amino acids serine, histidine, and aspartate collaborate in promoting hydrolysis of a peptide bond. In this example, we focus on the role of base catalysis. Note that histidine-57, which is held in place through a hydrogen bond between an imidazole nitrogen and aspartate, acts as a base because the second imidazole nitrogen accepts a proton from an active site serine. The OH group is converted into a stronger, negatively charged nucleophilic oxygen anionic species that attacks the electrophilic carbonyl carbon of the peptide bond. This is the first step in hydrolyzing the bond. Recall that the stronger the base, the stronger the nucleophile. As O- is a stronger base than OH, deprotonation of

the serine side chain promotes catalysis through the generation of a stronger base.

In the second scenario of acid/base catalysis, a hypothetical scheme for hydrolysis of a peptide bond is shown with two key active site residues: lysine (Lys) and serine (Ser). In this scenario, an uncharged lysine side chain acts as a base by accepting a proton from water, generating the stronger, negatively charged hydroxyl nucleophile.



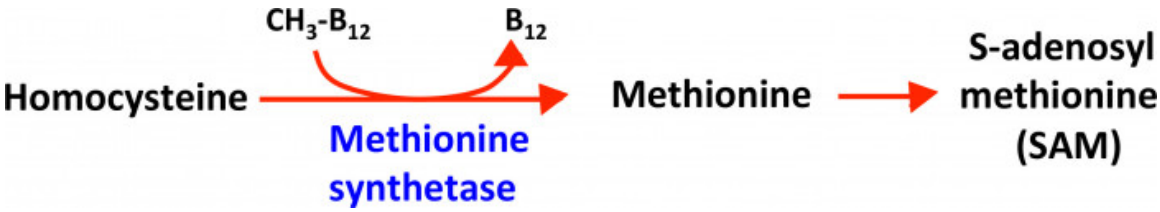
As shown, the HO- nucleophile attacks the electrophilic carbon in the carbonyl group of the peptide bond, forming a tetrahedral intermediate. Note that the serine residue forms a hydrogen bond with the carbonyl oxygen that helps keep the peptide bond situated in the active site of the enzyme. During the collapse of the tetrahedral intermediate, the newly protonated lysine residue behaves as a Bronsted-Lowry acid through donating its proton to the nitrogen atom in the peptide bond. This generates a stronger leaving group (NH_2 vs NH-). Recall that the best leaving group is the weakest base. (Weak bases do not share electrons very well, making the bond easier to break.) Without receiving the proton from lysine, the leaving group would be $R1NH^-$, but with the addition of the proton, the leaving group becomes $R1NH_2$, a neutral species and weaker base than $R1\ NH^-$.

Sulfur containing. Methionine contains a sulfide group ($–CH_2CH_2$-S-CH_3) and is notorious for starting polypeptide chains through its AUG initiation codon during ribosomal translation in both prokaryotes and eukaryotes.
Methionine is a non-essential amino acid because it is synthesized in the body through a reaction that uses Vitamin B_{12} as a methyl donor (CH_3-B_{12}). The pathway is shown below. Addition of a methyl group to homocysteine creates methionine. Afterwards, appreciate that methionine can be used in two major ways:

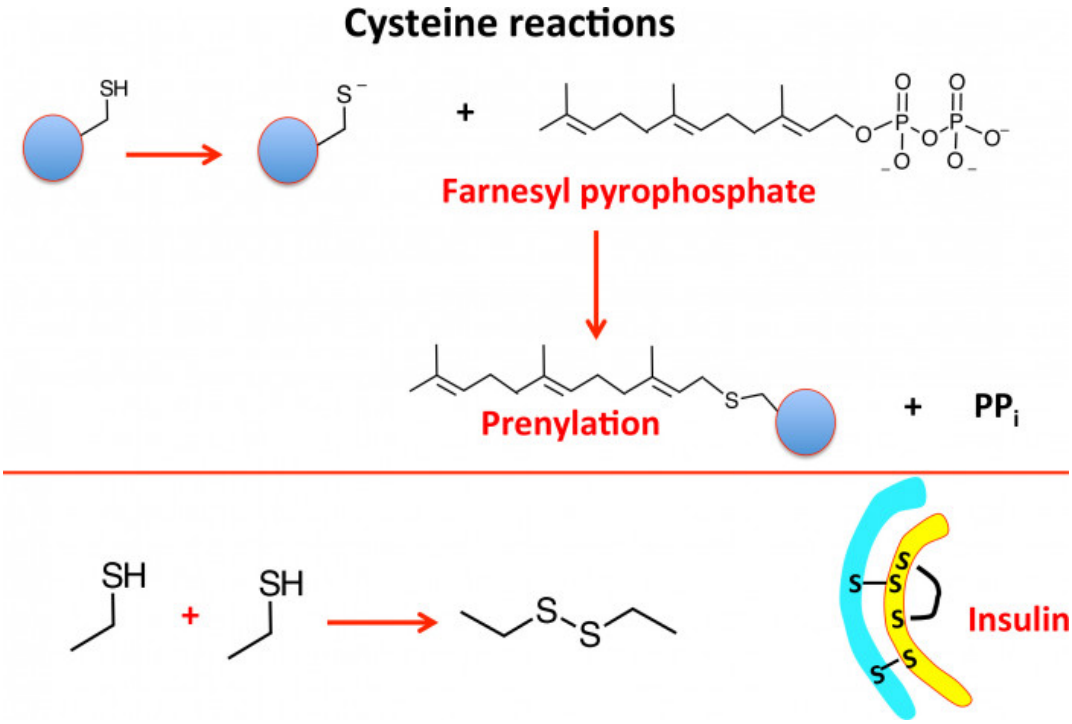
1) as an amino acid component of proteins.

2) conversion into the methyl donor, S-adenosyl methionine (SAM).

SAM is used to add methyl groups (one carbon) to numerous substrates including nucleic acids and proteins.



Cysteine
Cysteine contains a sulfhydryl side chain (-SH) as part of its side group and this has important ramifications for multiple aspects of its biology and chemistry. Reactive cysteine residues in active sites of enzymes often have pK_a values that are considered low for generic sulfhydryl groups. Indeed, once ionized, the negatively charged anion serves as a more potent nucleophile relative to its protonated form. This is shown below through the modification of cysteine with farnesylpyrophosphate, an isoprenoid molecule derived from the cholesterol biosynthetic pathway. Isoprene molecules are also called terpenes and are listed in the AAMC Content Outline in Foundation 3A.



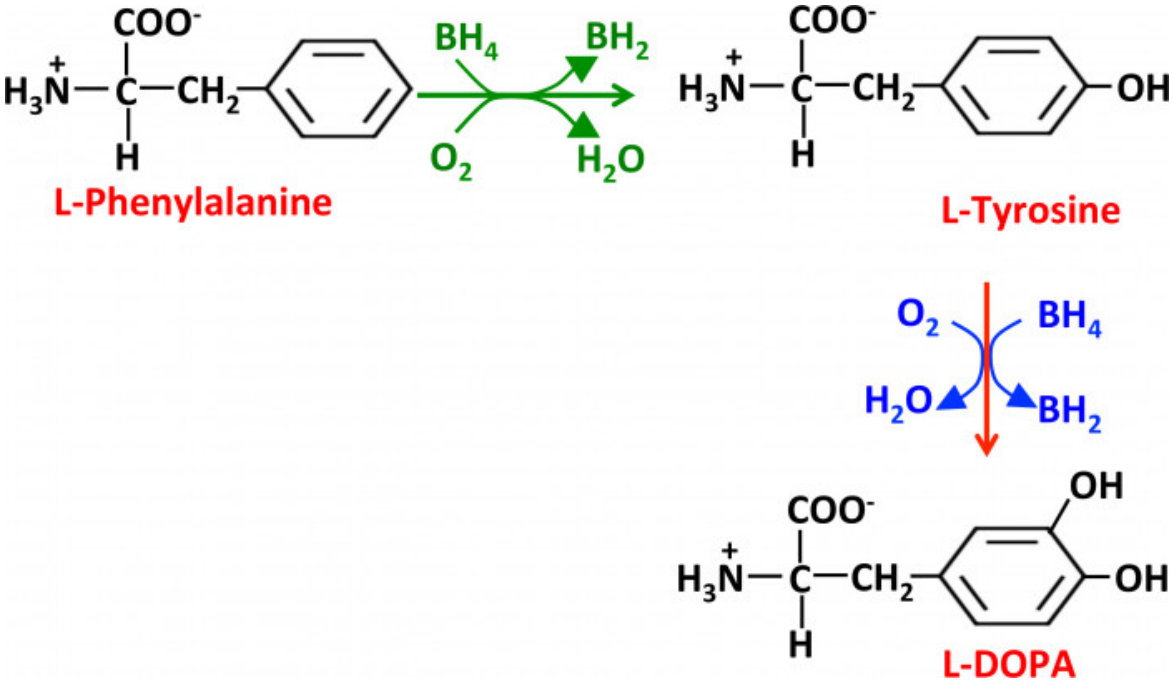
The addition of isoprenoid molecules to cysteine residues is often called "prenylation", a reference to the addition of isoprenoid groups. Think of prenylation as the addition of a large alky chain to a substrate. As such, the addition of these hydrophobic groups often serves to anchor the protein into the cell membrane. The enzyme mechanism of prenylation is further examined in the Med-Pathway Passage Workbook.

Disulfide Linkage

Cysteine most notoriously participates in disulfide bonds that are major contributors to protein structure and function. Disulfide bonds are formed through the oxidation of two cysteine residues, forming a cystine linkage (R-S-S-R) as shown above in the figure. These disulfide linkages are introduced into various proteins (i.e. insulin) in the endoplasmic reticulum and are commonly found in secreted proteins. Either intrachain or interchain cysteine side chains may form covalent linkages as is the case for insulin. Further, all cysteine residues are NOT obligated to participate in disulfide bonding.

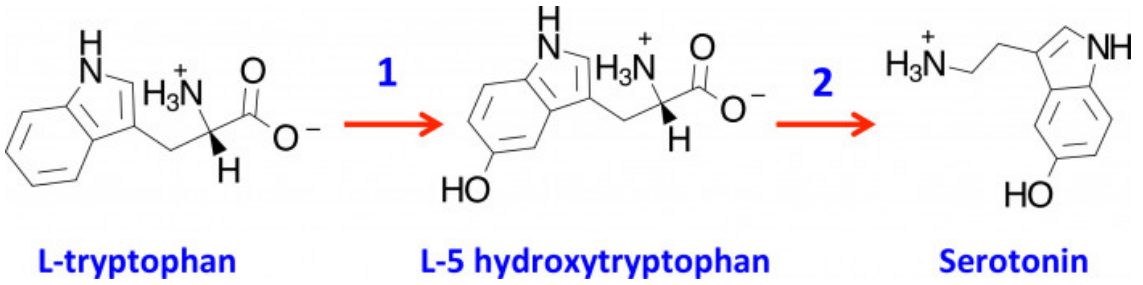
Aromatic amino acids. Phenylalanine (Phe) is converted into tyrosine (Tyr) in an important reaction that forms the basis for the disease phenylketonuria. Both amino acids are also precursors for important molecules such as L-DOPA and epinephrine.

Both phenylalanine (Phe) and tyrosine (Tyr) contain benzene rings as part of their side chains. The difference between the two residues is the presence of the hydroxyl group in tyrosine. Right off the bat, interpret this to mean that the side chain of tyrosine can participate in hydrogen bonding, but the side chain of Phe cannot. Tyrosine has a large side chain, and despite the presence of the alcohol group, it is a relatively hydrophobic.



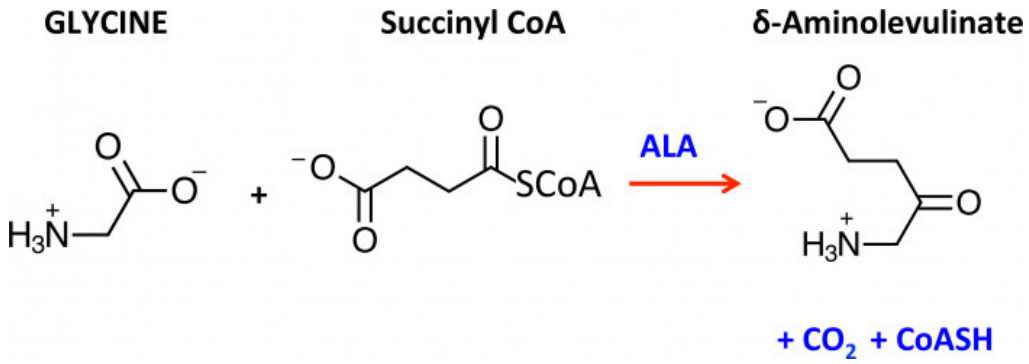
L-phenylalanine hydroxylase (PAH) converts Phe into Tyr. PAH is a monooxygenase enzyme that uses tetrahydrobiopterin (BH_4) as a redox co-factor to add the hydroxyl to Phe, generating tyrosine. This is seen in the green step in the figure. Tyrosine hydroxylase conducts the rate limiting synthesis of a class of molecules called catecholamines (i.e. L-DOPA) and also uses BH_4 as a co-factor. Other important catecholamines include epinephrine, norepinephrine, and dopamine. Therefore, both Phe and Tyr are precursors to important neurotransmitters and hormones.

Tryptophan. L-tryptophan is incorporated into proteins and contains the indole side chain. L-tryptophan is a precursor in the synthesis of melatonin, the essential human nutrient niacin (nicotinic acid), and serotonin, an important neurotransmitter. The synthesis of serontonin is shown below. Hydroxylation of L-tryptophan via tryptophan hydroxylase. Like the enzymes that catalyze the synthesis of L-tyrosine and L-DOPA, tryptophan hydroxylase is a biopterin-dependent enzyme.

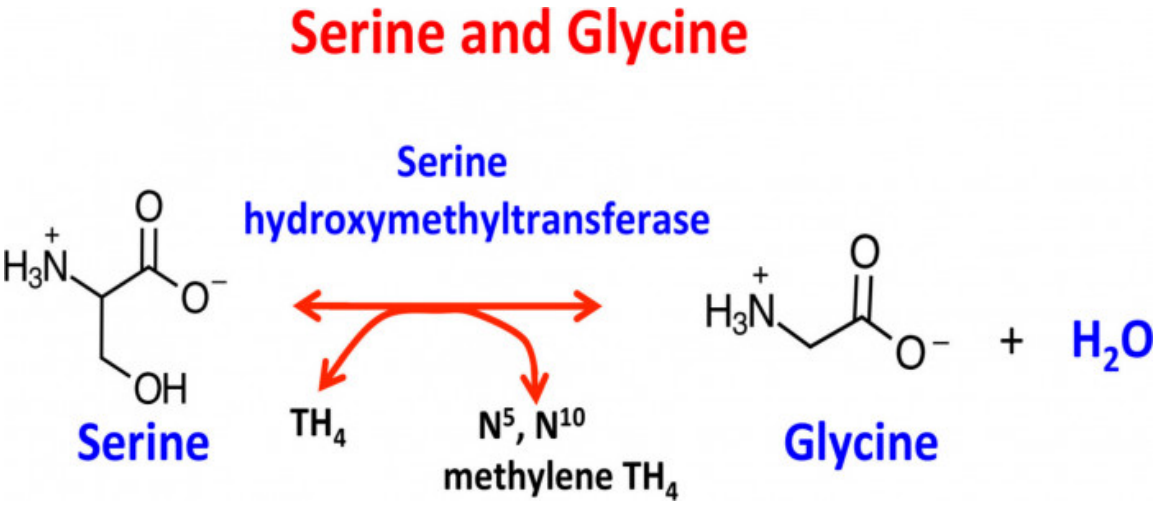


Glycine and Proline

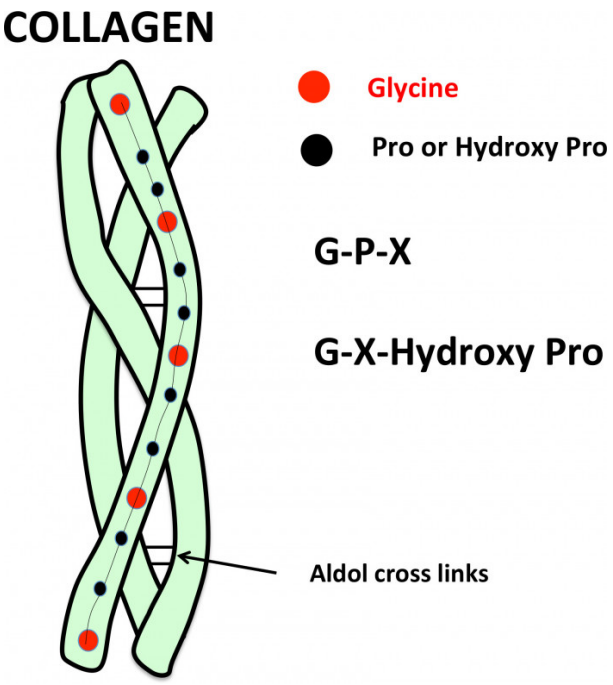
Glycine is the smallest amino acid and is achiral by virtue of having two hydrogen atoms bonded to the same α carbon atom. Besides its important role in protein structure and function, glycine is used in the biosynthesis of both purines and heme. Importantly, in the first step of heme biosynthesis, the enzyme aminolevulinic acid (ALA) synthase generates aminolevulinic acid (ALA) from glycine and succinyl CoA. This enzyme is a key regulator of heme biosynthesis. As the amino group is on the delta position, the product is often referred to as δ-aminolevulinate.



Glycine is a nonessential amino acid and is generated from serine as shown below. The reaction is reversible. The enzyme serine hydroxymethyltransferse uses a carbon moiety from tetrahydrofolate (TH_4), a one-carbon methyl donor. Remember that in the conversion of uracil into thymine, TH_4 is also used as a critical methyl donor. For the conversion of serine into glycine, serine donates its carbon to increase the cellular pools of N5, N10 methylene TH_4, a form of folic acid poised to donate a methyl group in additional biochemical reactions. Thus, serine and glycine differ by one carbon.

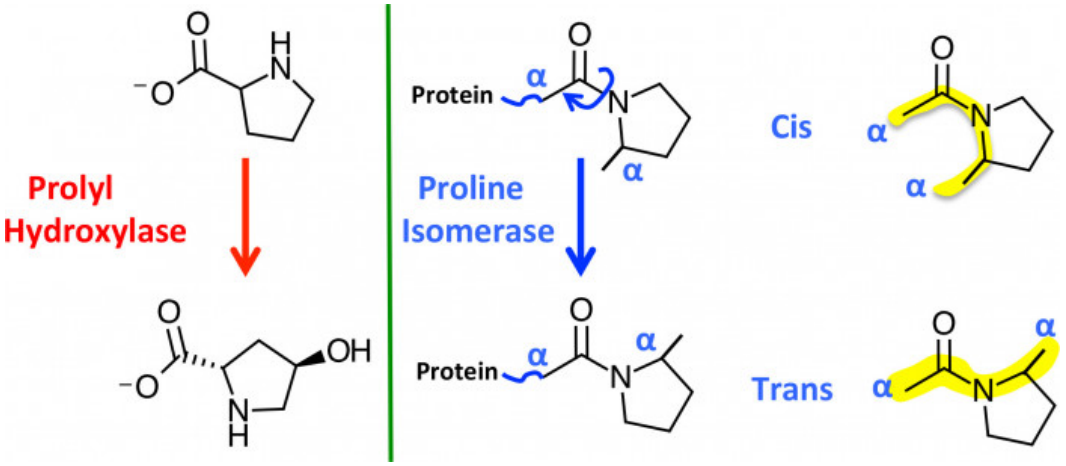


Proline is technically an imino acid as the nitrogen is a secondary amine. Proylyl hydroxylases generate hydroxyproline, a highly enriched posttranslational modification in the protein collagen as well as in other important cellular regulators. Collagen is an important structural protein and proline residues play a key role in its function. Collagen is an abundant structural protein found in various types of connective tissues in animals. This includes tendons, skin, cartilage, and ligaments.



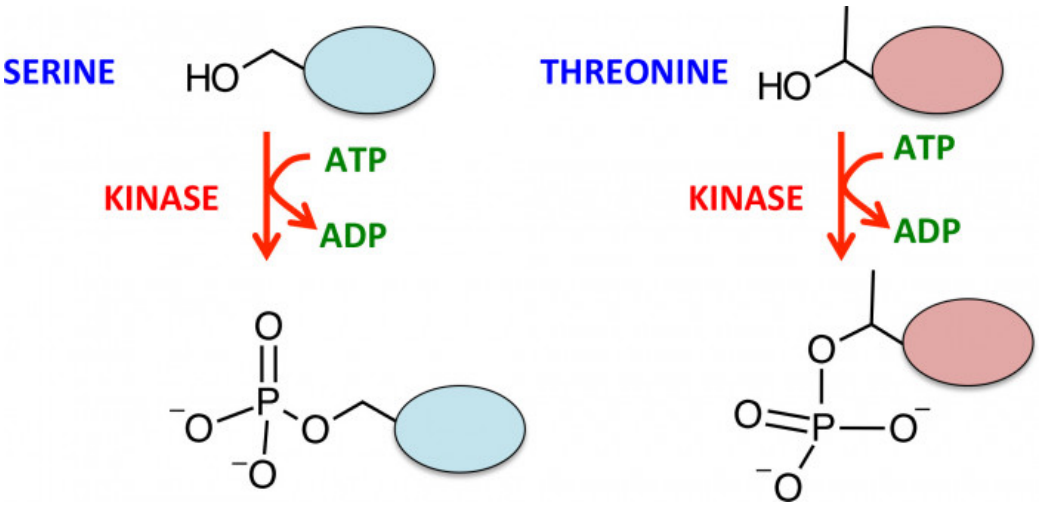
Collagen adopts a triple helical structure that largely consists of repeating amino acid motifs composed of either G-P-X or G-X-hydroxyproline. Proline hydroxylation is added in the lumen of the ER and this modification contributes to the high tensile strength characteristic of a structural protein. Aldol cross links between lysine residues generate interstrand cross links that further strengthen the collagen triple helix and collagen fibers. Defects in collagen, particularly through missense mutations in glycine and proline (or hydroxyproline) residues are causal for many diseases including osteogenesis imperfecta, commonly known as brittle bone disease.

The unique ring structure of the imino acid proline grants it several interesting properties important in regulating protein folding, structure, and function. Under normal conditions, the trans rotamer is highly favored over the cis form in the peptide bonds of proteins. This is because there is a significant activation energy barrier (20 kcal/mol) between converting the trans to the cis form that arises due to steric hindrance. However, in X-Pro peptide bonds (X = any amino acid), the energy barrier between cis and trans is greatly reduced. Thus, the cis peptide bond is more frequently seen in X-Pro sequences. Despite this, the cis proline bond is not formed spontaneously, but rather is catalyzed by proline isomerase enzymes as shown below.

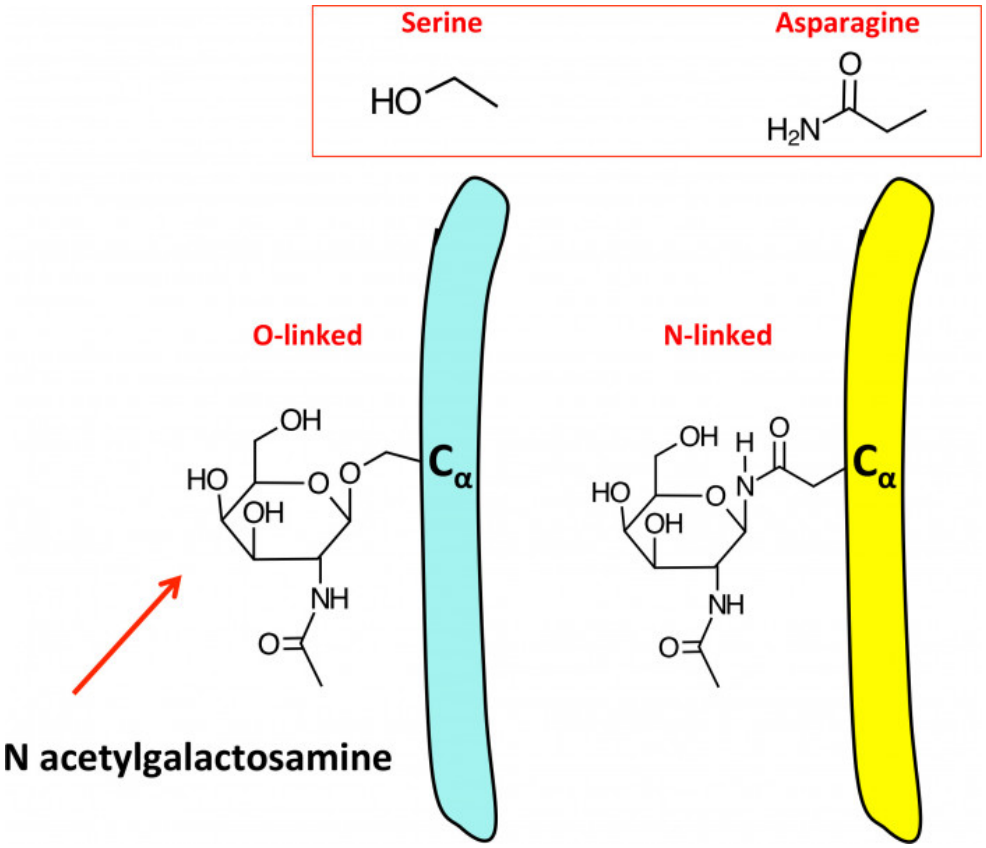


Altering between the cis and trans states has important functional ramifications for proteins. Think of proline isomerization as a molecular switch where cis/trans represent on/off or off/on switches. Roles for proline isomerization have been reported in numerous cellular processes including the cell cycle and neurotransmission. For example, one study published in Nature by Lummis

and colleagues showed that cis and trans isomerization regulated the function of a neurotransmitter-dependent ion channel.



Polar, non charged amino acids. We have included serine, threonine, asparagine, and glutamine in this category. At various target points within polypeptide chains, serine and threonine, two uncharged polar amino acids with hydroxyl side chains, serve as substrates for kinases that transfer the phosphate from ATP to the targeted side chain in the protein. This generates an overall negative charge due to the ionization state of phosphate. Phosphorylation of proteins can activate or downregulate the activity of proteins.



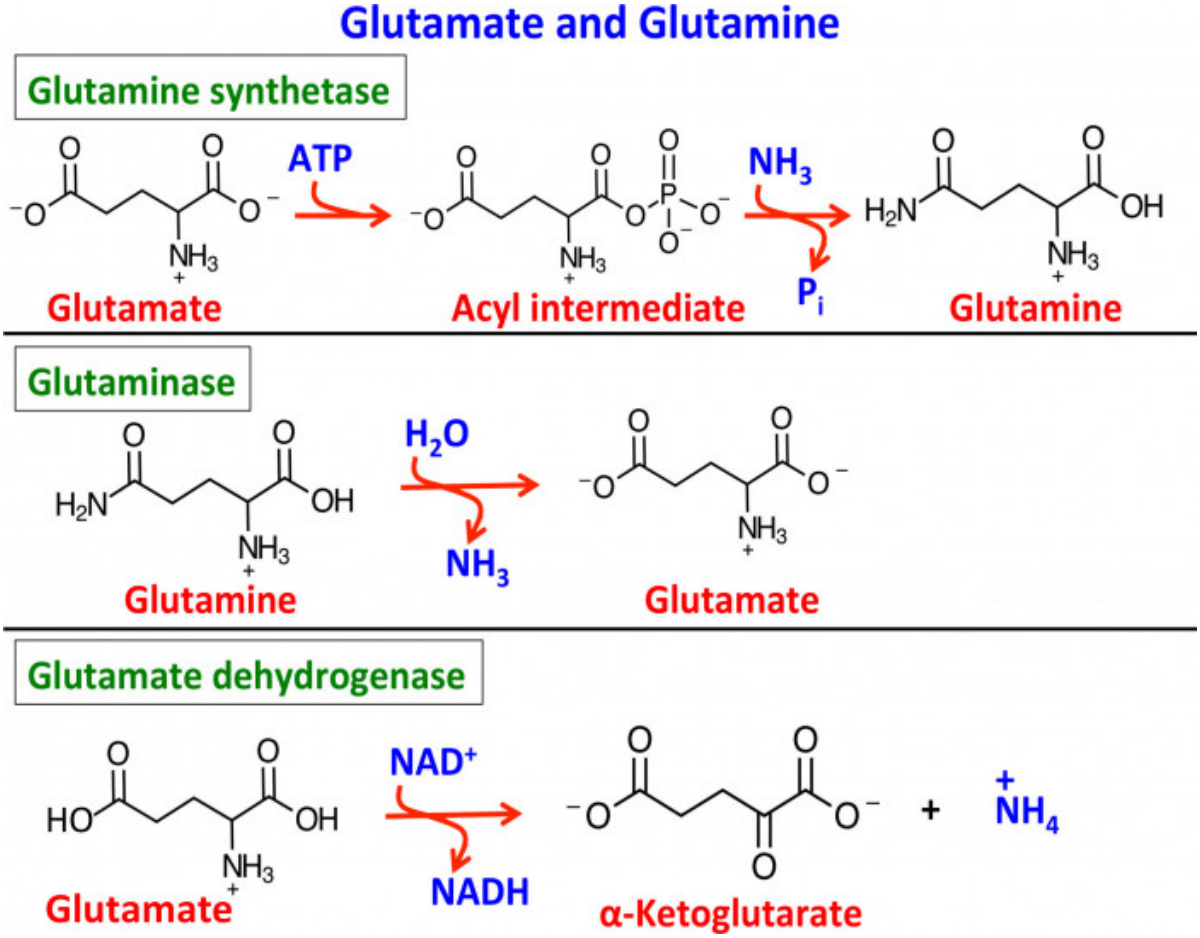
The aliphatic alcohols serine and threonine each have pK_a values expected for this functional group (~ 13.0). Serine is synthesized from the glycolytic intermediate 3-phosphoglycerate and serves as a substrate for the synthesis of glycine.

Serine and threonine (O-glycosylation) are glycosylated in many membrane and secreted proteins. Further, asparagine can also be linked through its side chain where it forms N-linked sugars such as N-acetylgalactosamine. These sugars are added posttranslationally in the Golgi apparatus by specific enzymes.

Glutamine. Glutamine is a very important amino acid because it is the major nitrogen carrier in the body. Through this, glutamine is intimately related to glutamate. Think of glutamine as carrying two amino groups and glutamate as carrying one. Because it can be synthesized from glutamate via glutamine synthetase, glutamine is not an essential amino acid.

Three major enzymes in the metabolism of glutamine are shown below. The majority of glutamine is synthesized in the muscle and liver, but glutamine synthetase is also active in the intestinal tract, the residential site of the microbiota. Here, bacterial metabolism often generates free ammonia, a

potentially toxic agent. To rid the system of this, glutamine synthetase couples free ammonia to glutamate, generating glutamine. In contrast to this, kidney glutaminase hydrolyzes glutamine, releasing free ammonia and glutamate. The resulting ammonia assists in urine buffering.

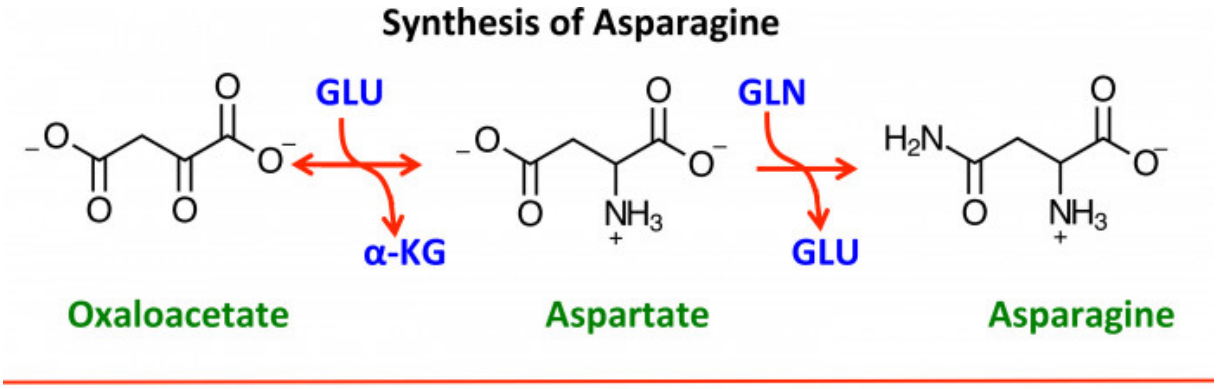


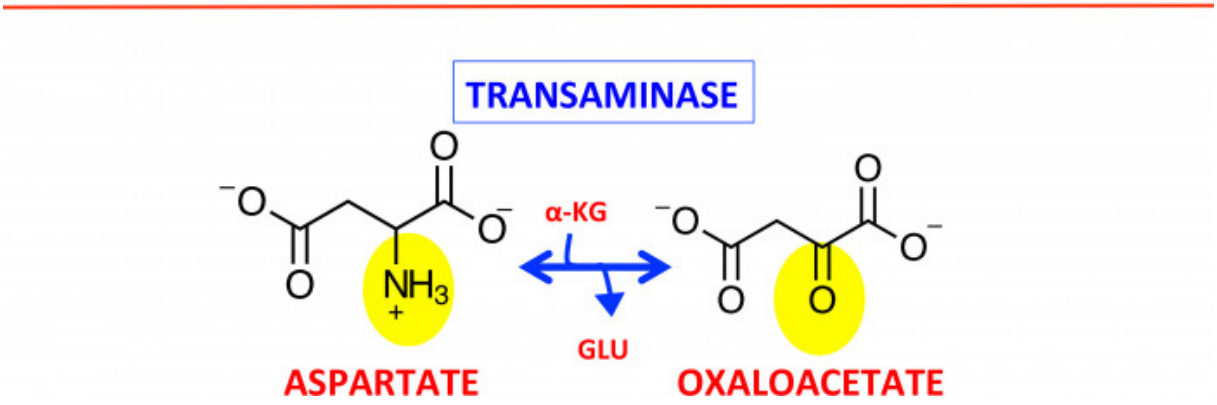
In the liver the amino groups from excess glutamine and glutamate are released via glutamate dehydrogenase and glutaminase. The free ammonia is rapidly coupled with CO_2 and ATP to generate carbamoyl phosphate. This nitrogen carrier enters the urea cycle, a hepatic specific process that converts ammonia into urea, a nitrogenous waste product. The urea cycle is discussed in more detail below.

Asparagine. Asparagine is synthesized in two steps starting from oxaloacetate (OAA). OAA and aspartate are intraconvertible through a reversible transamination reaction as shown. Transamination plays a central role in regulating nitrogen balance, a process directed by transaminase enzymes. These pyridoxamine-dependent enzymes catalyze the exchange of keto groups for

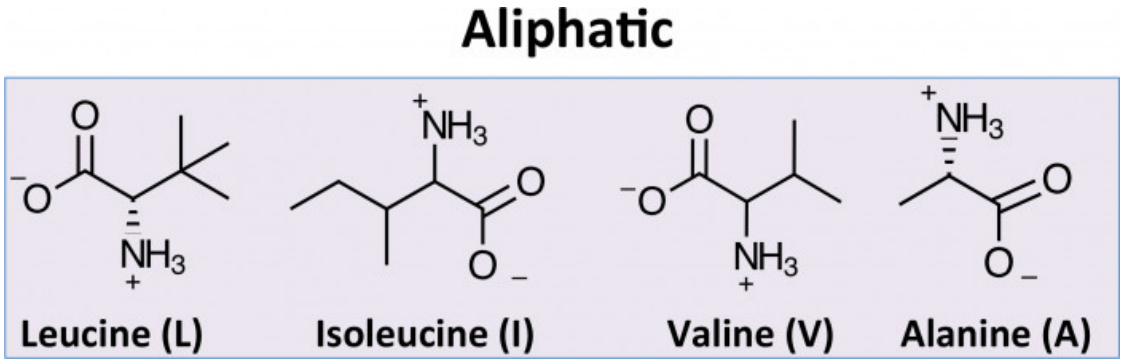
amino groups. As often seen, nitrogen is donated to alpha-ketoglutarate to form glutamate as a product. Transamination is discussed in more detail below.

Next, asparagine synthetase uses aspartate to generate asparagine as shown. Note again the presence of glutamate and glutamine as nitrogen donors. Asparagine synthetase deficiency is a rare autosomal recessive disease and is considered a neurometabolite disorder because this is accompanied by microencephaly, seizures as well as spasitc quadriplegia. Genes that are deficient in amino acid metabolism are known as "aminoacidopathies".

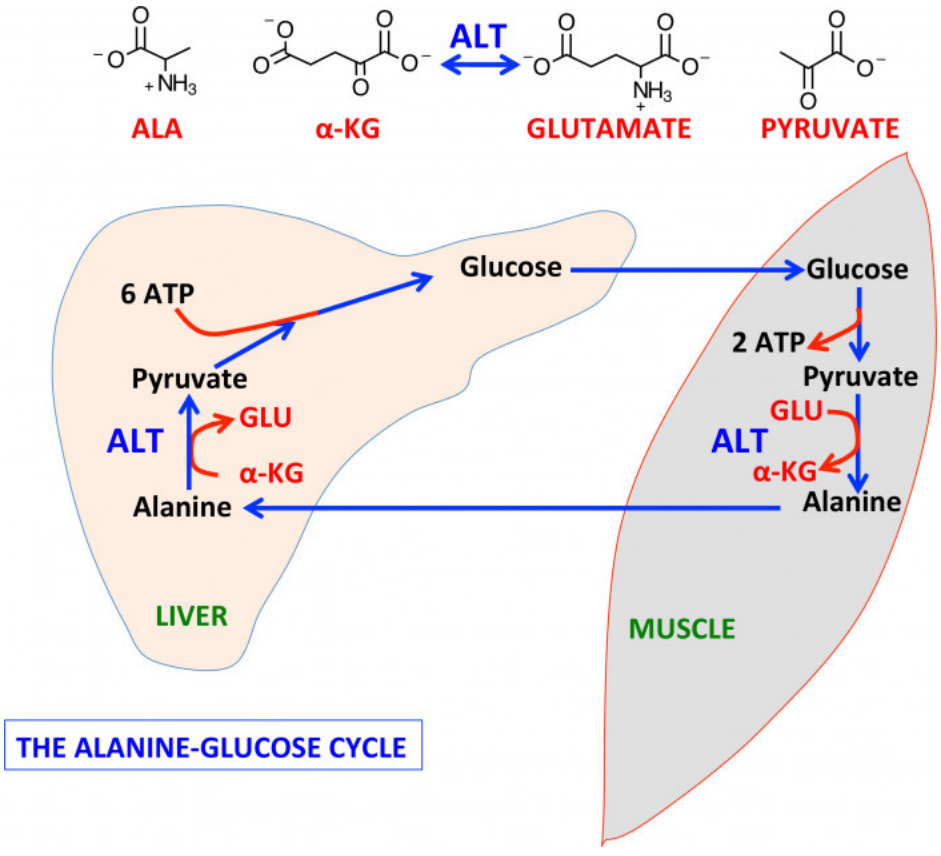




Aliphatic amino acids. We listed four amino acids as aliphatic. Three of the four (leucine, isoleucine, and valine) are commonly referred to as the "branched chain amino acids" or BCAA based upon the structures of their side chains. Each of the BCAAs is essential and is abundant in muscle as they comprise up to 30% of muscle mass. This is why they are often referred to as the "building blocks" of muscle and are used as supplements by body builders. The fourth aliphatic amino acid, alanine, does not contain a branched side chain and is nonessential as it is synthesized from pyruvate.



Alanine is critical in glucose homeostasis because upon transamination it is converted into pyruvate, a precursor to glucose in hepatic gluconeogenesis. This important role for alanine in the "alanine-glucose" cycle is shown in the figure and will be re-visited as we highlight tissue specific roles for alanine in muscle and liver metabolism. Note from the image that the conversion between alanine and pyruvate requires the action of alanine transaminase, a pyradoxamine-dependent enzyme present in both skeletal muscle and liver. However, due to mass action, the reaction proceeds in opposite direction in each tissue (see image below).

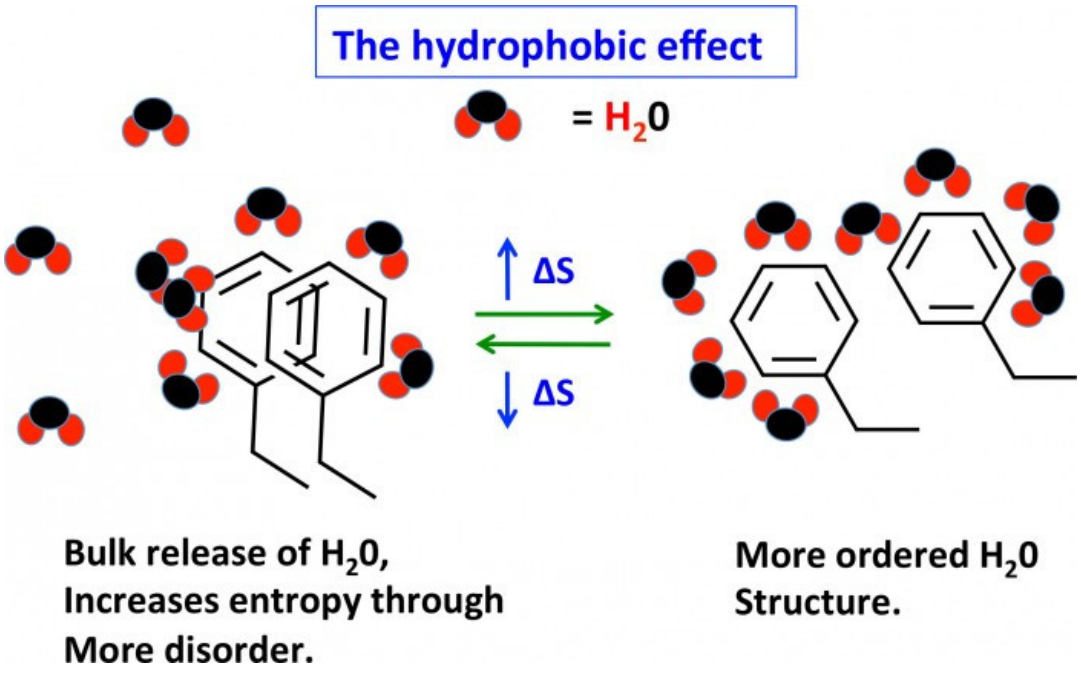


The aliphatic amino acids have nonpolar, hydrophobic side chains and this is discussed below. In addition, catabolic pathways for branched chain amino acids are important as deficiencies generate metabolic diseases known as organic acidemias. These rare metabolic disorders are highlighted by maple syrup urine disease (MSUD) and are further characterized below in the "Catabolism of amino acids" section.

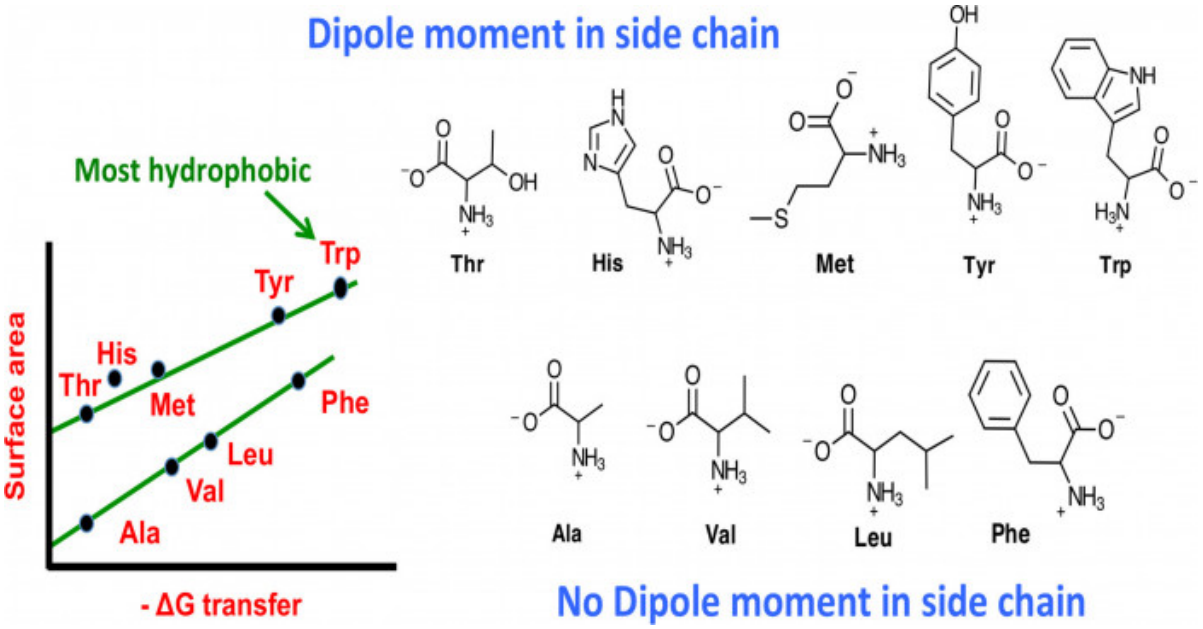
Hydrophilic amino acid side chains are soluble in water because they are comprised of polar side groups. They are further classified into acidic (aspartate and glutamate), basic (lysine and arginine and histidine at low pH values), polar amides (asparagine and glutamine), and polar alcohols (serine and threonine). Tyrosine is also an alcohol group, but its aromatic character classifies this amino acid as hydrophobic.

Hydrophobic amino acids.

Hydrophobic (or non-polar) amino acids contain side chains with a varying number of carbon and hydrogen atoms that formulate methylene groups or sp^2 carbons (i.e. phenylalanine). The amino acids commonly associated with this group including alanine, valine, phenylalanine, isoleucine, and leucine. Formulation of the hydrophobic core of globular proteins is an important driving force in the protein folding process, the process of adopting the three dimensional active structure. Therefore, due to their aversion to water ("like dissolves like"), amino acids with hydrophobic side chains are biased for residing in the interior of proteins. The physical basis for the hydrophobic effect intimately involves the thermodynamic concept of entropy. Recall that all spontaneous chemical reactions increase the overall entropy of the universe. As shown in the image, hydrophobic interactions as depicted by aromatic phenylalanine stacking, have more disordered water structures associated with them.



The "hydrophobicity" of an amino acid can be experimentally determined. This is shown by measuring the free energy of transfer ($\Delta G_{transfer}$) of an amino acid from water to an organic solvent as a function of its water accessible surface. From these experiments, the relative degrees of amino acid hydrophobicity can be determined. Ideally, a solution that mimics the environment believed to resemble protein interiors should be used in the experiment. These relative $\Delta G_{transfer}$ values are shown. The results indicate that the strength of hydrophobicity is dependent on the surface area of the side chain and its dipole moment.



Immediately recognize that because negative ΔG values are associated with spontaneous events, the more negative ΔG values signify more solubility (or capacity to transfer like into like) in non-aqueous solvents. Using your understanding of the structures of the side chains of amino acids, immediately appreciate that the amino acids on each line represent amino acids with either zero (Ala, Val, Leu, and Phe) or one dipole moment (Thr, His, Met, Tyr, and Trp).

For those amino acids with no dipole moment a linear relationship is achieved between hydrophobicity as a function of size. Also observe that hydrophobicity depends on the presence of a dipole moment and the water accessible surface area.

Additional trends can be discerned from the data, leading to some conceptual formulations regarding the amino acids. Tryptophan is the most non-polar amino acid, despite the fact that it has a dipole moment. Of the amino acids presented, threonine and alanine are the most polar, despite their differences in dipole moment and accessible surface area. Further, observe that although tyrosine has a hydroxyl group capable of hydrogen bonding with water, it is still a largely hydrophobic amino acid.

Note that the graph presented above shows the values obtained for the entire amino acid. What is the contribution of the side chain? How could one experimentally determine the ΔG values for the side chain as opposed to the entire amino acid? One way to do this is to standardize the measurements by

determining the $\Delta G_{transfer}$ value for glycine, the smallest amino acid with only a single H as its side chain. By subtracting this value, the ΔG contributions from an individual side chain are approximated in a relative fashion. In this scenario, one would say that the experimental ΔG values presented have been normalized to glycine. The second, more difficult way to do this experiment would be to chemically synthesize each of the side chains and then measure these values in parallel to each other.

Amino acid metabolism

In addition to their roles in protein structure and function, amino acids have multiple metabolic fates:

1. Increase blood glucose levels (glucogenic).

2. Increase ketone body levels (ketogenic).

2. Conversion into storage products (fatty acids, glycogen) in excess.

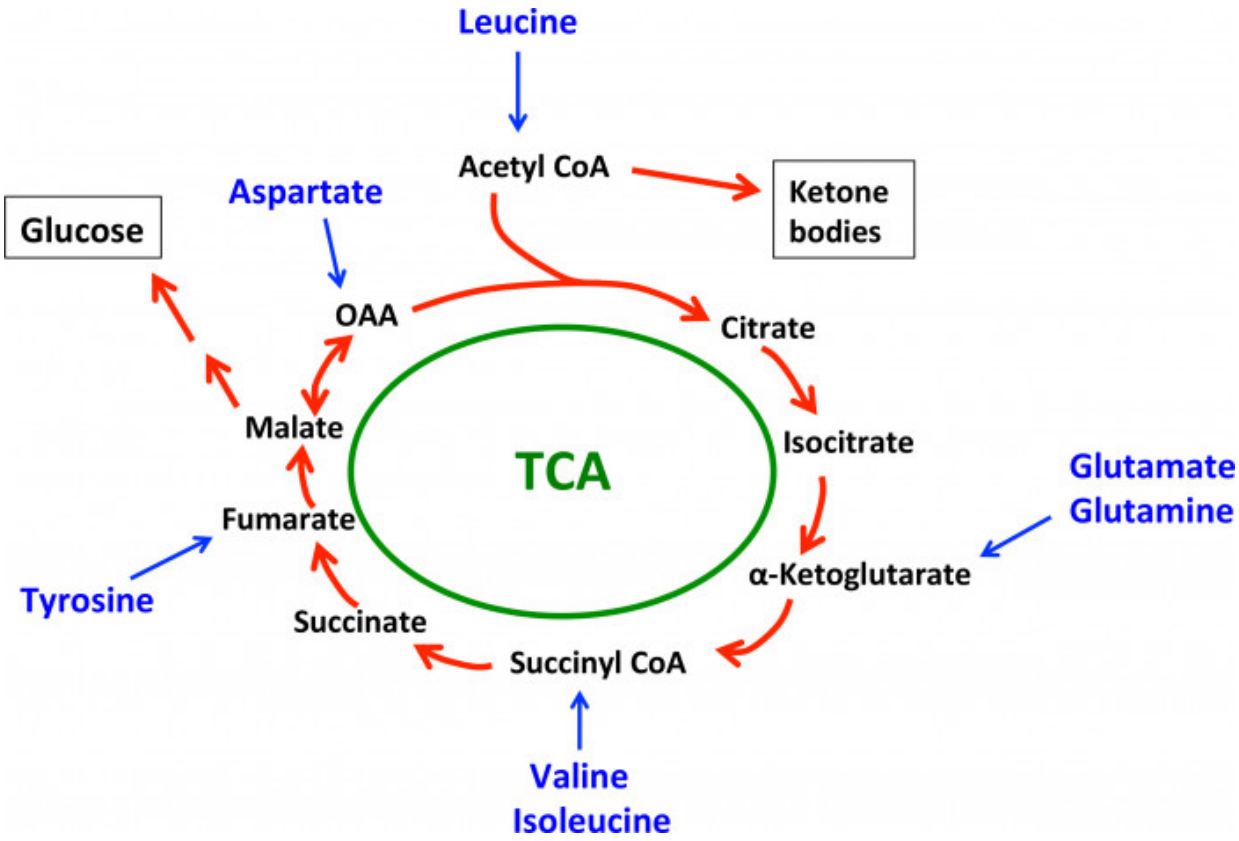
3. Generation of bioactive molecules (serotonin and histamine).

Ketogenic and glucogenic amino acids.

After removal of the α amino group, AAs are converted into carbon skeletons used in energy production (i.e. oxidation) or storage (reduction). The fate of the amino acids depends on the nature of the metabolic state. During the low energy state (i.e. low blood glucose or fasted state) amino acids will be broken down for energy, but in the high energy state (high blood sugar, fed state), excess amino acid carbon skeletons are used to generate storage molecules such as triglycerides. The carbon atoms of amino acids are processed into storage molecules (i.e. fat) or oxidized to generate energy. Those AAs metabolized into substrates used to increase de novo synthesis of net glucose via hepatic gluconeogenesis are termed gluconeogenic. Gluconeogenesis occurs under high glucagon and low insulin conditions. This subject is covered in Biological Foundations 1D and in various metabolism passages seen in the Med-Pathway workbook and in a specialized testing module on the website.

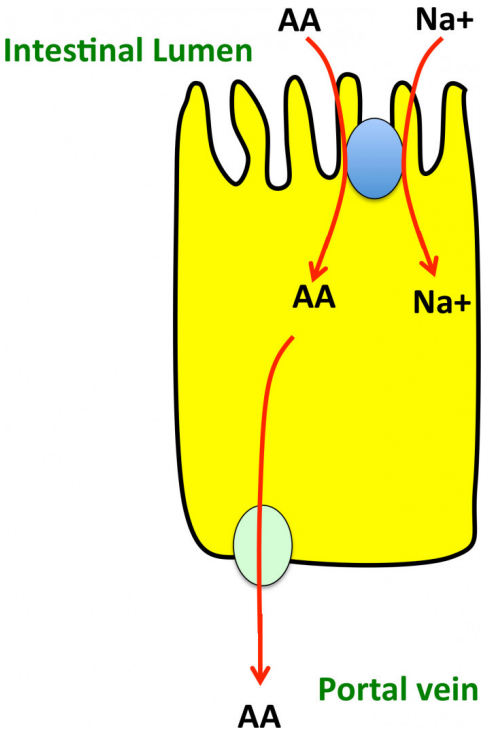
To visualize which amino acids are classified as gluconeogenic or ketogenic, refer to the figure below. Because gluconeogenesis uses the TCA intermediate malate as a substrate to generate glucose, those AAs that are metabolized into TCA intermediates convertible to malate are considered glucogenic. A partial list is shown below. Note how the AAs are linked to Krebs intermediates.

Those AAs degraded into metabolites that can form ketone bodies (i.e. acetyl CoA), are termed ketogenic.



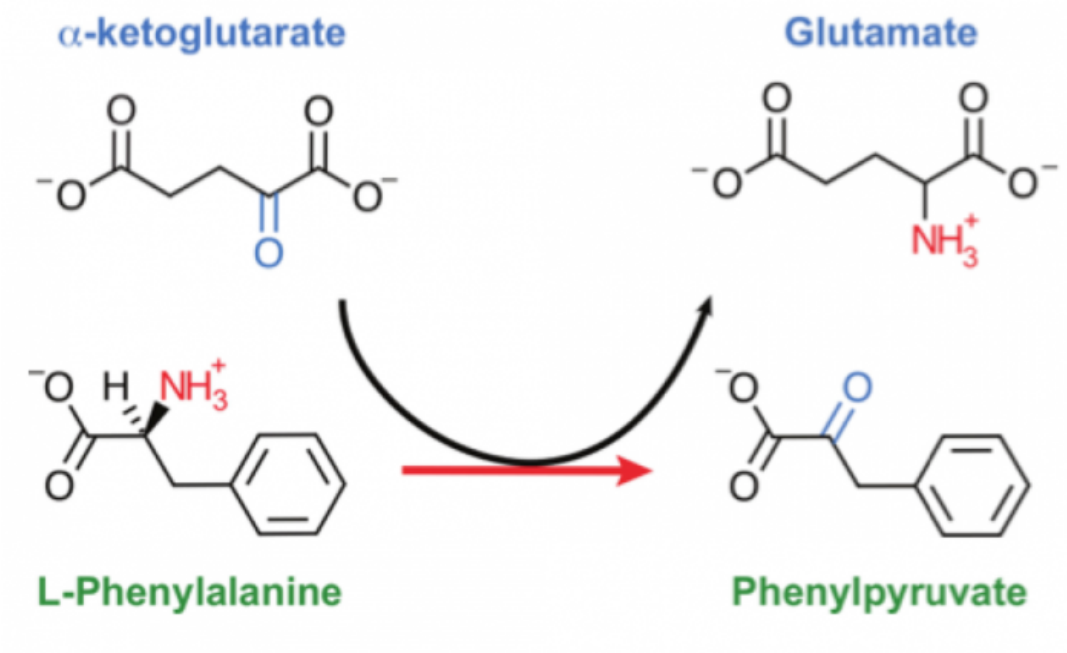
Excess free amino acids (AAs), arise from food in an energy rich state, or are derived from the breakdown of protein during low energy states. Transaminase enzymes initially metabolize these AAs. Recall that these important enzymes remove the α-amino group from amino acids and transfer them to substrates.

In the case of excessive AAs from dietary sources, transport commences in the duodenum via the action of Na+-dependent transporters as shown below. The increased levels of intracellular AAs result in facilitated transport into the portal vein as shown. The liver uses the excess amino acids for energy storage in the form of fats, glycogen, and protein and demonstrates the link between AA chemistry and biology.

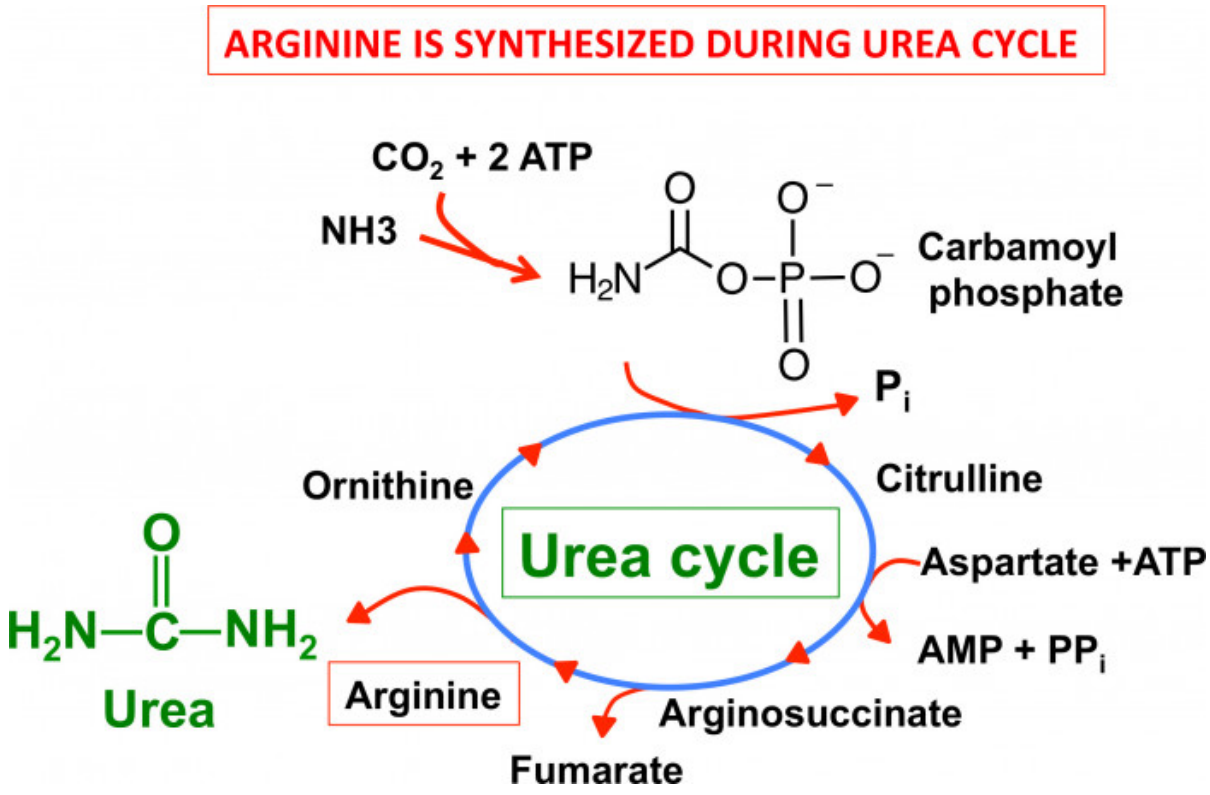


Aminotransferase enzymes use pyridoxal-phosphate as a co-factor to transfer amino groups to α –ketoglutarate. This creates the corresponding α-ketoacid, (i.e. the transaminated product) and glutamate, a major carrier of nitrogen. The transamination of L-phenylalanine into phenylpyruvate is shown again.

By removing the nitrogen group and transferring it to α-ketoglutarate, the carbon skeleton of the amino acid is "free" to be used in metabolism. AA degradation generates available carbon skeletons. As carbon represents energy, the products of AA catabolism can be used to generate net glucose (glucogenic AAs) or net acetyl CoA and ketone bodies (ketogenic AAs). The first step in AA catabolism is to remove the nitrogenous ammonium group and to transfer it to a substrate, eliminating free ammonia, a potential neurotoxin.



Transaminases and nitrogen balance. Transaminases are critical enzymes and as such are also important in disease. A classic example is the biochemical disorder phenylketonuria (PKU), a defect in metabolism of phenylalanine (Phe or PHE). The Med-Pathway Passage Workbook examines this in more detail in the physical sciences section. Phe is essential and when ingested via food, has three biological fates as shown below. First, Phe (or any other AA) can be incorporated into polypeptides during ribosomal protein synthesis. Second, excess Phe is transaminated into phenylpyruvate, a metabolite that is potentially toxic when levels build up in the metabolic disorder phenylketonuria (PKU). Lastly, Phe is substrate for conversion into the AA tyrosine (TYR). Tyrosine synthesis is catalyzed by phenylalanine hydroxylase (PAH), a monooxygenase enzyme that converts the essential amino acid phenylalanine (Phe) into TYR. Monooxygenase enzymes are important in multiple crucial biological reactions. These enzymes hydroxylate two substrates by "splitting" diatomic oxygen (O_2) where one oxygen atom is added as a hydroxyl to the substrate and the second oxygen atom is used to generate water.



Observe the flow of nitrogen from the amino acid carriers glutamine and glutamate to urea, a process that regenerates ornithine as shown in the figure. Note that the α-amino group of the transaminated amino acid is introduced into the urea cycle as carbamoyl phosphate, an activated nitrogen carrier composed from the reaction: CO_2 + NH_3 + 2 ATP = carbamoyl phosphate + P_i. This nitrogen and one from aspartate formulate the nitrogen atoms in the waste product urea. Glutamate and glutamine arrive in the liver and is a substrate of mitochondrial glutamine dehydrogenase. This enzyme releases free ammonia. As this compound is toxic, ammonia is immediately fixed into carbamoyl phosphate in a reaction that requires ATP.

Carbamoyl phosphate is a phosphoester that combines with ornithine to generate citrulline as shown. In the process of regenerating ornithine, aspartate and ATP is consumed. In contrast both arginine and urea are generated in the cycle. Thus, arginine is a nonessential amino acid. Urea is excreted through the kidneys and the blood urea nitrogen (BUN) value is an important indicator of hepatic health.

Alanine and pyruvate

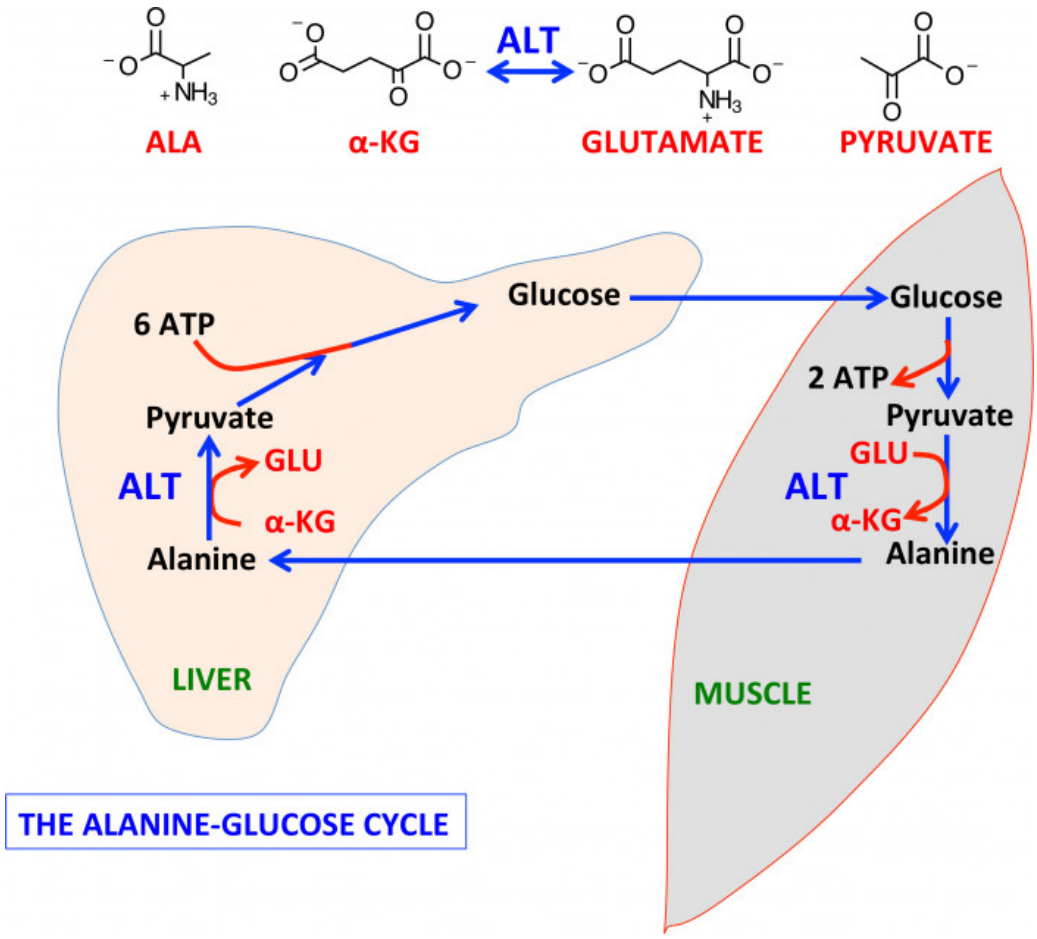
Transaminase enzymes in the liver and muscle catalyze the interconversion between alanine and pyruvate. This is diagrammed below. Alanine is an important amino acid in energy metabolism as seen through its behavior in the glucose-alanine cycle (Cahill cycle). One key enzyme in this process is alanine aminotransaminase (ALT), a pyridoxal phosphate-dependent enzyme. The reaction is presented below.

The conversion between alanine and pyruvate is an example of amino acid catabolism that occurs during the physiological low energy state (i.e. fasting). Under these conditions, hormones such as glucagon, epinephrine, and cortisol activate pathways that promote degradation of proteins into amino acids (i.e. ubiquitin proteosomal pathway). Those glucogenic amino acids serve as important substrate for generating glucose in order to maintain blood sugar levels (glycemia). This amino acid catabolism is coupled to Beta oxidation of fats as well as gluconeogenesis and ketone bodies.

In the muscle, pyruvate derived from glycolysis is converted into alanine. This, in conjunction with alanine derived from muscle protein degradation in the fasted, low energy state is shipped to the liver. Therefore, the carbon skeletons of alanine are used to generate glucose in the liver. By shipping off alanine to the liver, the carbon skeletons from the amino acid alanine are turned into glucose via the gluconeogeneic pathway. In this sense, the muscle can be seen as "storing" proteins and using their carbon atoms for conversion into glucose.

In the liver, ALT removes the amino nitrogen group from alanine, re-generating hepatic pyruvate, a key precursor to glucose in hepatic gluconeogenesis. The amine is transferred onto alpha-ketoglutarate, creating glutamate. Always think of glutamate (and glutamine) as nitrogen carriers. As such, the ammonia group is released via glutamate dehydrogenase where it is eventually incorporated into the excreted product urea through the urea cycle. Note the ALT enzymes catalyze the reciprocal reactions in the liver and muscle.

ALT is an important enzyme that is often measured in blood tests on patients. People with liver disease (i.e. cancer, cirrhosis, hepatitis) have increased levels of ALT in the serum due to release from damaged hepatocytes. Thus, high ALT levels positively correlate with liver dysfunction.



Catabolism of BCAAs. Catabolism of branched chain amino acids such as leucine generates carbon skeletons for energy formation. The initial step involves the removal of the alpha-amino group through transamination. The transaminated product is decarboxylated through a dehydrogenase reaction requiring coenzyme A (CoASH). The carbon atoms in this high energy intermediate are further metabolized through oxidation pathways that generate energy in the form of reducing power NADH and $FADH_2$.

Genetic defects in leucine decarboxylation are causal for "maple syrup urine disease (MSUD), a metabolic deficiency accompanied by mental retardation, seizures, and ataxia amongst others clinical presentations. The failure to catabolize leucine builds up the metabolic levels of the α-ketoacid as well as leucine itself. This gives off a distinctive odor in the urine, hence MSUD.

Essential/non essential amino acids

Amino acids are often classified as being essential or nonessential. Essential amino acids are those that cannot be synthesized in the body. For example Phenylalanine (Phe or F) is used as a substrate by the Phenylalanine hydroxylase (PAH) enzyme to generate tyrosine (Tyr or Y). Therefore, phenylalanine is an essential amino acid and cannot be derived from any source in the body, but tyrosine is not essential as it is made from phenylalanine. People lacking PAH function have phenylketonuria (PKU) and cannot metabolize Phe into Tyr. Thus, tyrosine is synthesized in the human body, so it is non-essential. However, tyrosine becomes an essential amino acid in patients with PKU. Thus, TYR is conditionally essential.

Synthesis of Amino Acids

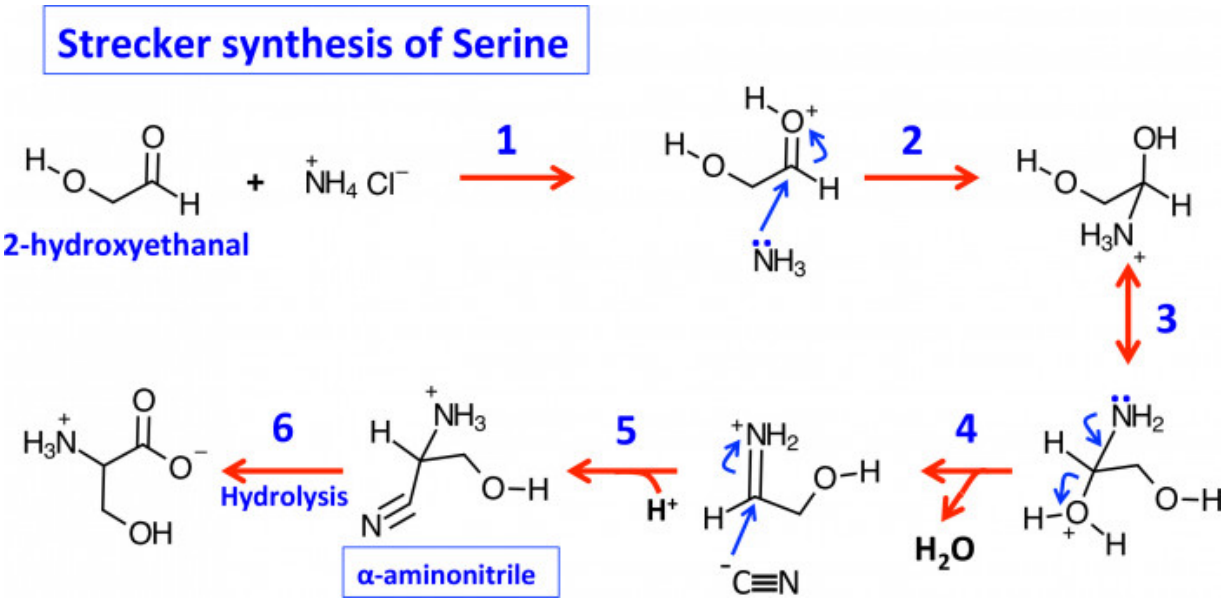
The AAMC MCAT Content Outline lists two synthetic reactions for synthesizing amino acids:

1. Strecker synthesis

2. Gabriel synthesis

Strecker synthesis of amino acids

Strecker described the first synthesis of amino acids over 150 years ago. The typical reaction scheme uses an aldehyde in the presence of ammonium and cyanide to generate an α-amino nitrile product that, after several steps, is hydrolyzed into a racemic mixture of the amino acid product.



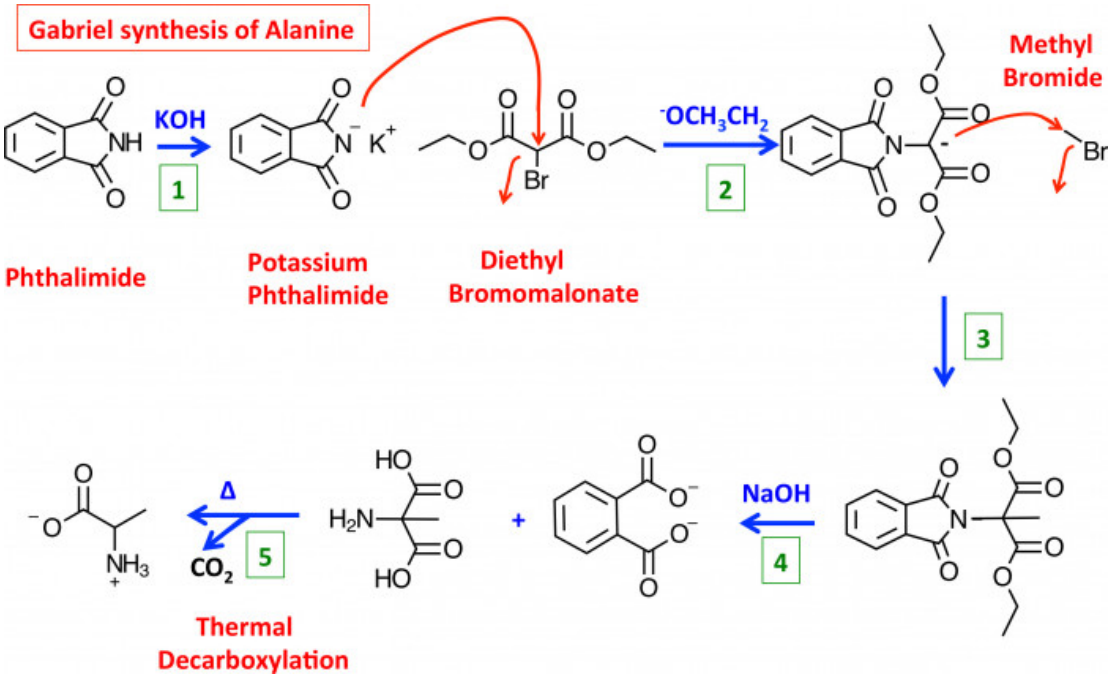
The reaction is composed of multiple steps and the synthesis of serine is shown in the image. Each step consists of chemical principles that you should be familiar with as they are not only applicable to the Strecker synthesis, but many other types of chemical reactions you might encounter on the MCAT.

The synthesis is summarized as follows:

STEP 1: Ammonium chloride acts as an acid and protonates the aldehyde substrate, 2-hydroxyethanal. As a result the conjugate base NH3 is formed.
STEP 2: The ammonia nucleophile attacks the electrophilic carbon atom in the carbonyl generating a tetrahedral intermediate.
STEP 3: The tetrahedral structure is in equilibrium between two protonated forms.
STEP 4: As the tetrahedral intermediate collapses, a water molecule is released, forming an imine that is attacked by the cyanide nucleophile.
STEP 5: Addition of a proton generates the α-amino nitrile.
STEP 6: Although represented by one arrow, several steps are required for complete hydrolysis of the α-amino nitrile into the final amino acid product (serine in this case).

Gabriel synthesis of amino acids

The Gabriel synthesis of amino acids reaction sequence is shown for alanine in the below schematic. Several important steps, each with different organic chemistry principles is highlighted below.



Step 1: The reaction starts with the addition of the strong base KOH to phthalimide. Note that the heterocyclic, 2° nitrogen in phthalimide is resonance stabilized in the ionized form, meaning that in the presence of the strong base, an acidic proton is released. The resulting negatively charged nitrogen is a strong, resonance-stabilized nucleophile.

Step 2: Potassium phthalimide reacts with diethyl bromomalonate. This S_N2 reaction releases bromide as the leaving group in the formation of the product.

Step 3: In the presence of the strong ethoxide base, a newly created nucleophile performs a second S_N2 reaction with methyl bromide formed as the product of reaction 3. Methyl bromide is required to generate alanine. This alkylation reaction adds on the methyl group that will form eventually form the side chain of alanine. Note the presence of the two ethyl ester groups in the product of reaction 3.

Step 4: Hydrolysis of the esters with hydroxide in reaction 4 generates a carboxylic acid and an amine as shown.

Step 5: In the presence of heat, decarboxylation generates alanine.

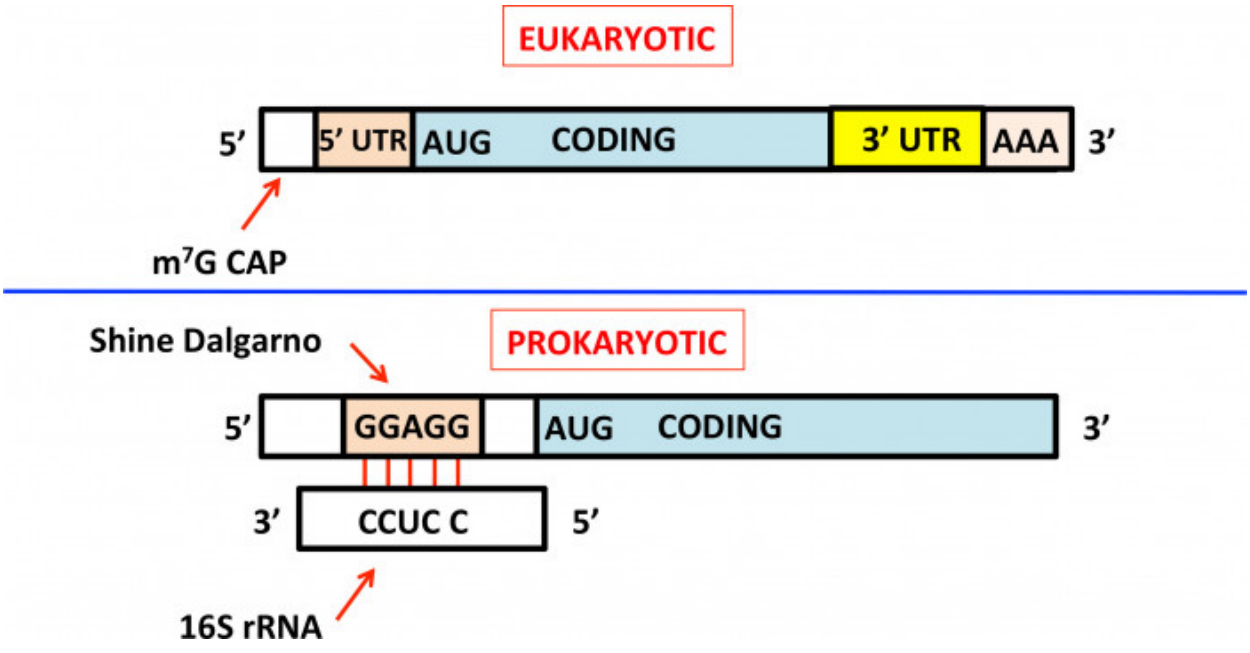
Chapter 2

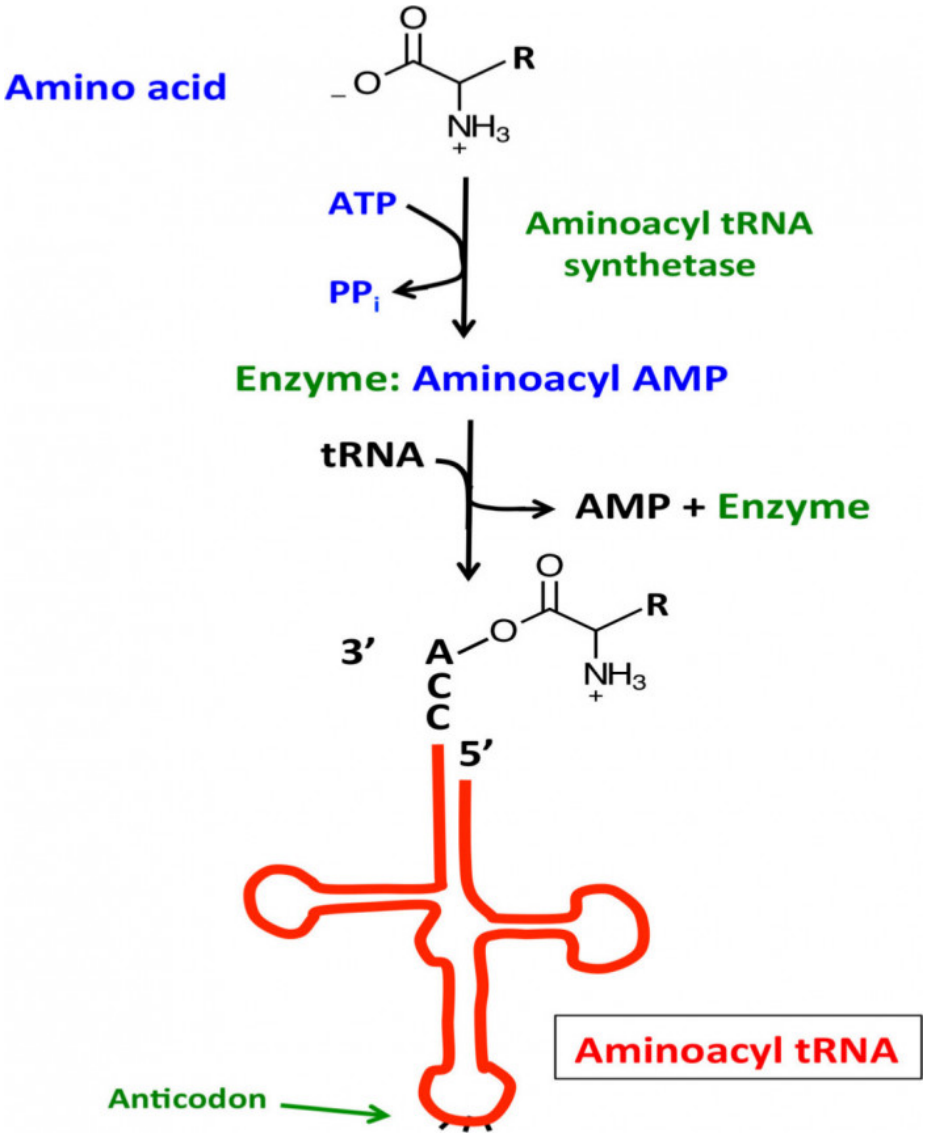
Protein structure and function;
Protein purification

- ☐ 1°, 2° , 3° , 4° structure of proteins including Hemoglobin
- ☐ Denaturing and folding
- ☐ Hydrophobic interactions
- ☐ Hill coefficient and ligand binding (Hemoglobin)
- ☐ Amino acid separation & Protein Purification

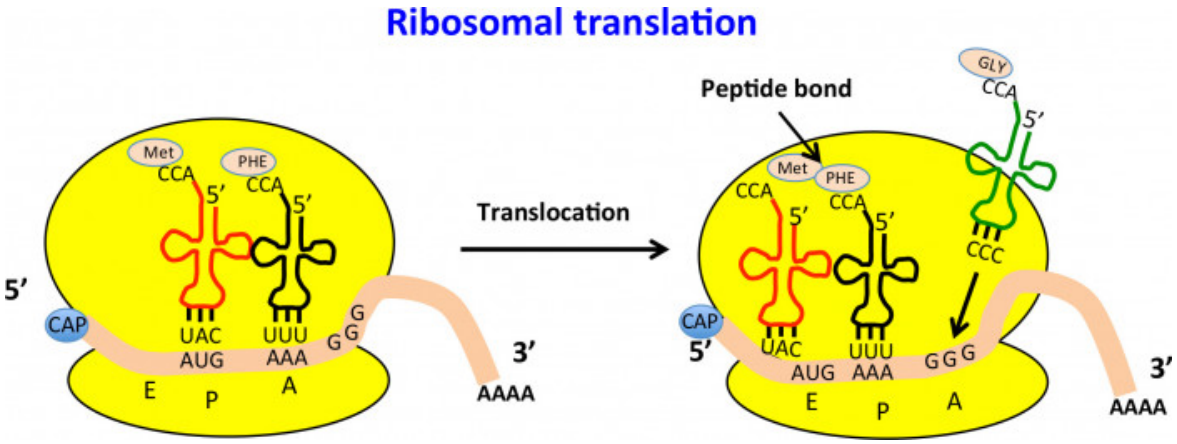
Formation of protein 1° structure

The primary (1°) structure of proteins is determined by the sequence of their amino acids arranged in consecutive peptide bonds from the N-terminus to the C-terminus. The sequence as envisioned by Francis Crick's "Central Dogma" is determined through the process of ribosomal translation. The 5' end of eukaryotic mRNAs are composed of a CAP structure that aids in ribosomal recognition and binding. In prokaryotes, a Shine Delgarno sequence rich in AT residues is recognized by the ribosome. This sequence is approximately 10-12 base pairs upstream of the initiating AUG methionine codon.





mRNA sequence triplet codons are recognized by complementary anti-codon sequences located in the stem loop structure of transfer RNA molecules (tRNAs). Importantly, amino acyl tRNA synthetases covalently links an amino acid to its cognate tRNA in a two-step mechanism as shown in the schematic above. Note that energy derived from ATP hydrolysis is required to "charge" a tRNA with its matched amino acid. The initial reaction of the enzyme creates an aminoacyl AMP intermediate that is displaced by a tRNA molecule possessing an anticodon sequence complementary to the mRNA codon that specifies the amino acid to be incorporated into the protein chain.

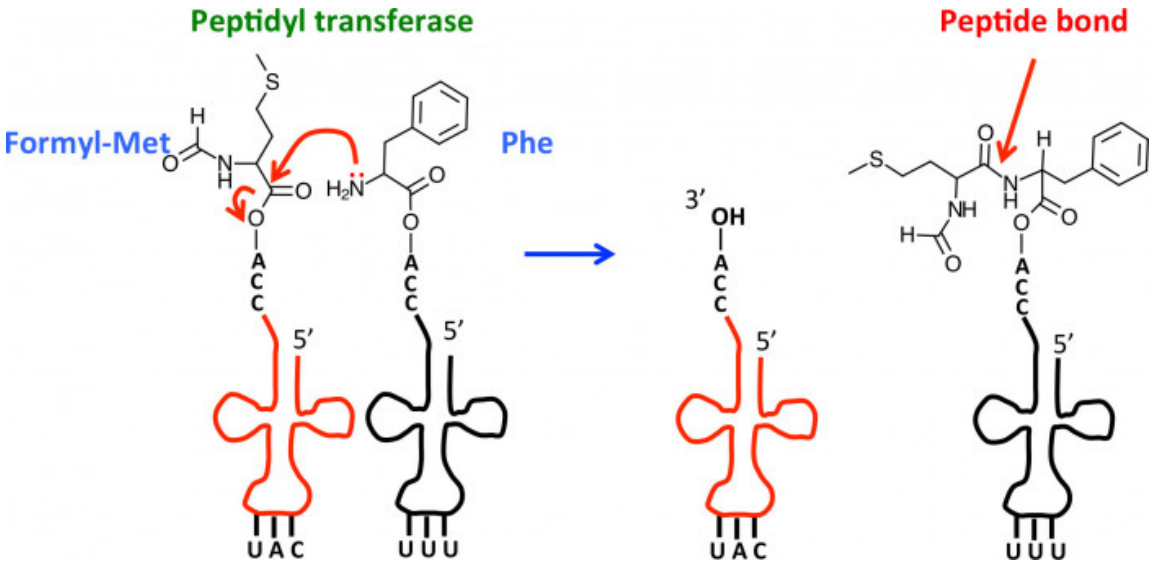


Charged tRNA molecules are recruited to ribosomal-mRNA complexes through various initiation factors. Shown in the image is the translation of an mRNA sequence: 5'-AUG AAA GGG-3' into the tri-peptide MET-PHE-GLY. As virtually all protein sequences begin with Methionine (or Formylmethionine), a charged methionine tRNA with an anticodon complementary to the AUG initiation codon is recruited to the ribosomal P site as shown. Next, the second codon (AAA) recruits a phenylalanine-charged tRNA with the complementary sequence (UUU) to the A site. Various elongation factors use the energy of GTP hydrolysis to translocate along the length of the mRNA transcript.

Eukaryotic translational factors are known targets of bacterial toxins. In particular, diptheria toxin inactivates elongation factor-2 (EF-2). Although many of the principles of ribosomal translation are similar between lower organisms and humans, prokaryotic and eukarotyic ribosomes are different enough that they are recognized by different antibiotics. For example, chloramphenicol, a broad spectrum antibiotic declared an essential medicine by The World Health Organization, specifically inhibits bacterial peptidyl transferase.

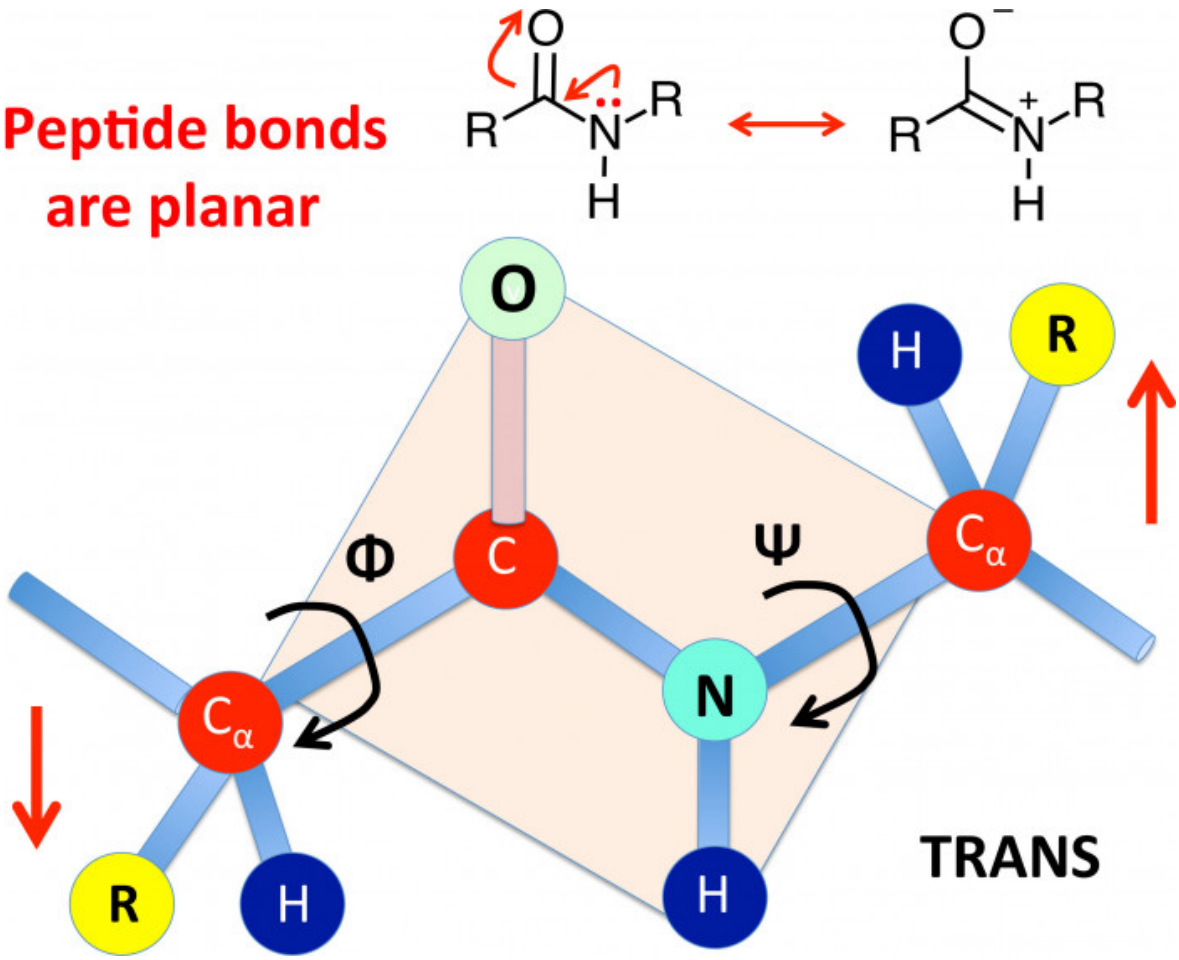
The occupancy of charged tRNA molecules at the P and A sites of the ribosome juxtaposes the first two amino acids of the primary protein sequence for the catalysis of the first peptide bond by peptidyl transferase, a ribozyme whose entire catalytic function is conducted by RNA sequences within the ribosome. The reaction scheme is outlined in the figure. Note that a nucleophilic substitution-elimination reaction converts the ester linkage in the aminoacylated tRNA into a peptide (amide) bond. The reaction occurs through

the formation and collapse of a tetrahedral intermediate, but this is not shown for simplicity.



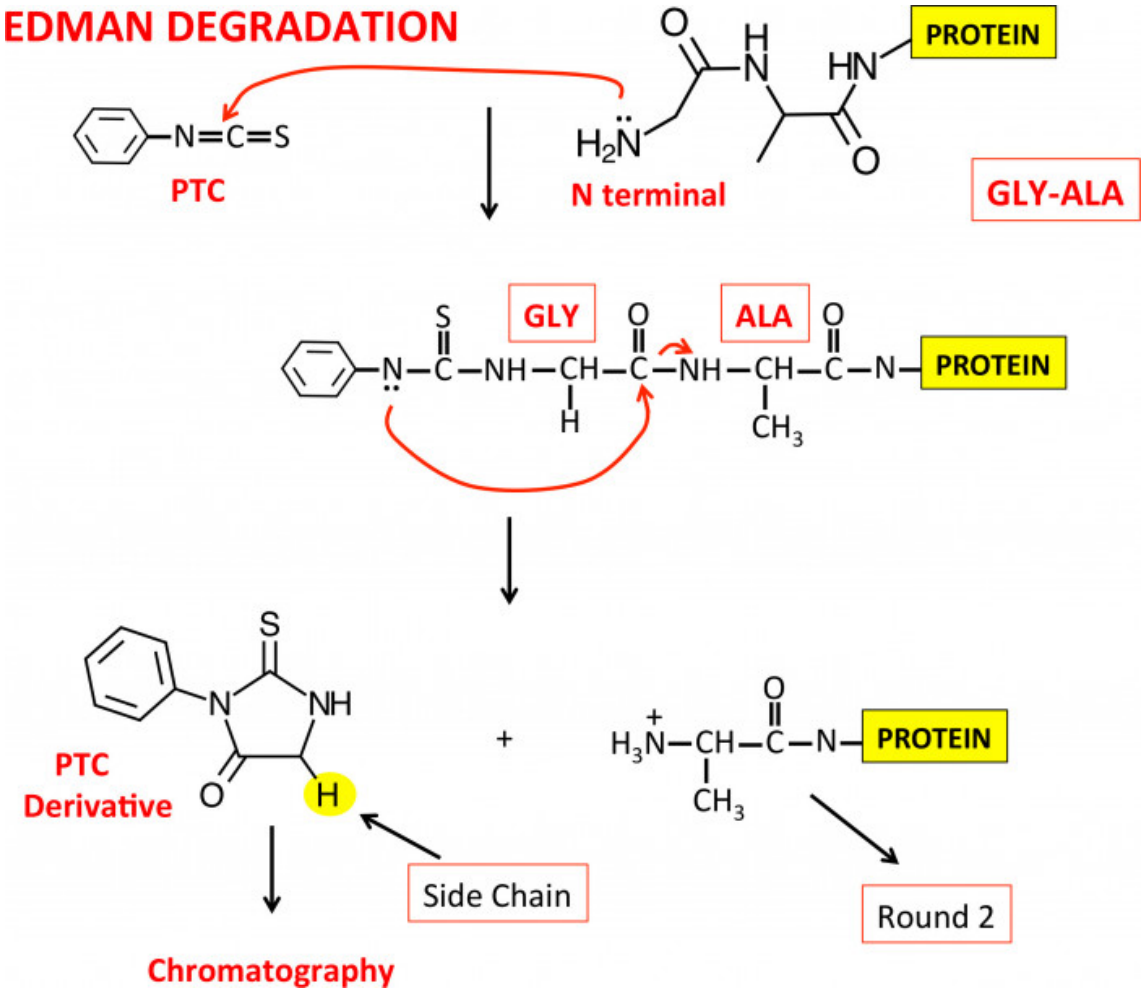
After the synthesis of the peptide bond, ribosomal translocation occurs, opening up a new A site for the third charged tRNA. This places the initiating codon in the E (exit) site.

Peptide bonds are synthesized during ribosomal translation. Due to the formation of resonance structures (see image below), peptide bonds have partial double bond character. As rotation is limited around double bonds, peptide bonds adopt a planar structure that favors the trans position as shown. However, as previously discussed above for proline, the unusual side chain structure of proline stabilizes the cis conformation as well. The proylyl isomerase enzymes catalyze the conversion of trans to cis.



Determining the 1° structure of Proteins

The Edman degradation procedure has been widely used to determine the sequence of proteins up to approximately 40 amino acids. The procedure is outlined below for the protein sequence that begins with GLY-ALA.



Step 1: Phenylthioisocynate (PTC) is reacted with the protein or peptide of interest under basic conditions that promote an uncharged, nucleophilic amine group. The carbon atom in PTC is the electrophile.

Step 2: The PTC intermediate forms an intramolecular cyclization reaction. This amide product is hydrolyzed in the presence of acid in order to generate the PTC derivative and the N-1 protein or peptide that is ready for a second round of reaction with PTC. The PTC derivative and its bound amino acid side chain can be determined through various chromatography procedures including TLC.

Proteins as polyprotic acids

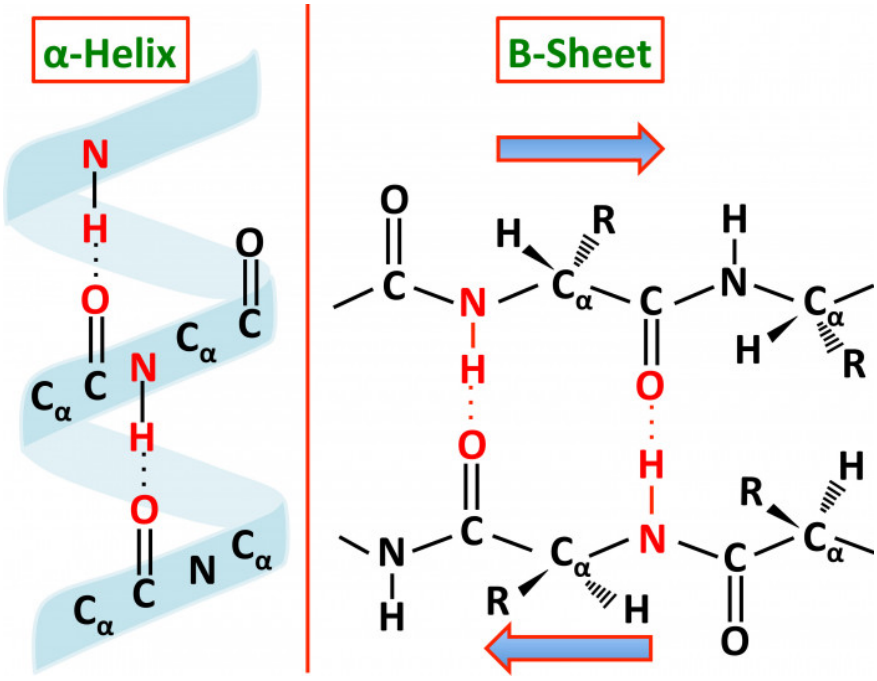
Many of the amino acid side chains of proteins are charged due to ionization of side chains. As we saw with amino acids, the isoelectric point (pI) is the pH where the net charge is zero. As with amino acids, proteins also have isoelectric points. As a general rule, you should know that if the pI > pH, then the

protein will have a net positive charge when all of the charges from the side groups are considered. Likewise, if pI < pH, the protein will have a total net negative charge. Both amino acids and proteins can be separated by exploiting their charges as a function of pH.

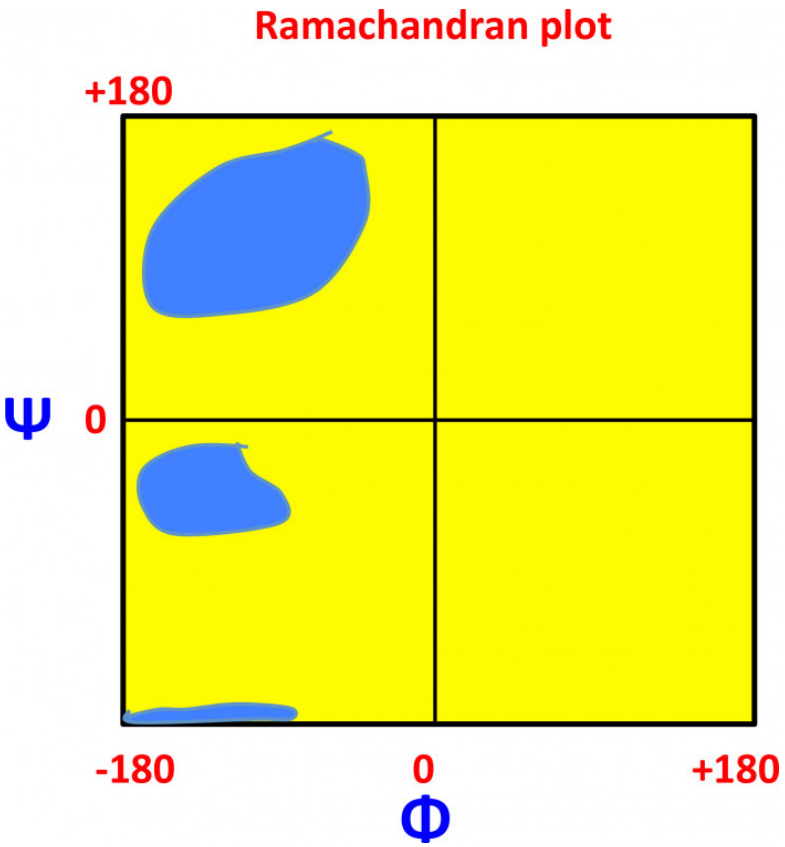
Secondary structure of proteins

The primary (1°) structure of proteins is comprised from their sequence of amino acids. Some sequence arrangements are conducive for forming localized structural features described as secondary (2°) structure. The most commonly discussed 2° structural features are α helices and β sheets, or β turns. These are shown below. The structural protein collagen adopts a 2° structural feature known as a triple helix consisting of repeating units of the amino acid X-Y-G. In this setting, every third amino acid is glycine. Although X and Y can be any amino acid, proline and its derivative hydroxyproline are most commonly observed in this motif.

The α helix is a fundamental structural feature of many proteins. Linus Pauling conceived of the notion of the classical α helix while sick in bed! He made a model from a sheet of paper and described the α helix using his knowledge of bond angles. This included his hypothesis that peptide bonds would exist in a planar configuration due to resonance as has been shown. In most cases, the trans configuration is favored. Note that rotation is permitted around the Phi (Φ) and Psi (Ψ) dihedral angles due to their single bond character.



L-amino acids most commonly form right handed helices and the α helix has 3.6 amino acids residues/turn. Both types of 2° structures utilize hydrogen bonding between the amide nitrogen hydrogen and the carbonyl groups of the main chain. This is shown above. Notice how the side chains do not participate in hydrogen bonding in these particular structures. Recall that hydrogen bonding occurs between a donor hydrogen (H), an electronegative atom (O or N in this case), and a second hydrogen (an acceptor) as shown in the figure with the red-colored atoms. As discerned from the images, the α helix and β sheet structures can be envisioned to exhibit regular geometric patterning. Multiple, consecutive secondary structures found within a polypeptide chain are known as motifs.



Secondary structures can be described through their phi (Φ) and psi (Ψ) angles in a Ramachandran plot of a given protein or peptide. In this model, atoms are envisioned as hard spheres with their corresponding van der Waals radii. Recall

from above that Φ and Ψ angles represent those bonds on either N or C side of the peptide bond.

Through rotating each Φ and Ψ angle within a given protein, permissible angles for bonds in a protein (i.e. conformation as found in α helices and β sheets) are shown in blue. These angles do not sterically interfere with each other in response to rotation as the van der Waal radii do not overlap. In contrast, those angles that have overlapping radii are considered unstable conformations as shown in the yellow regions of the Ramachandran plot.

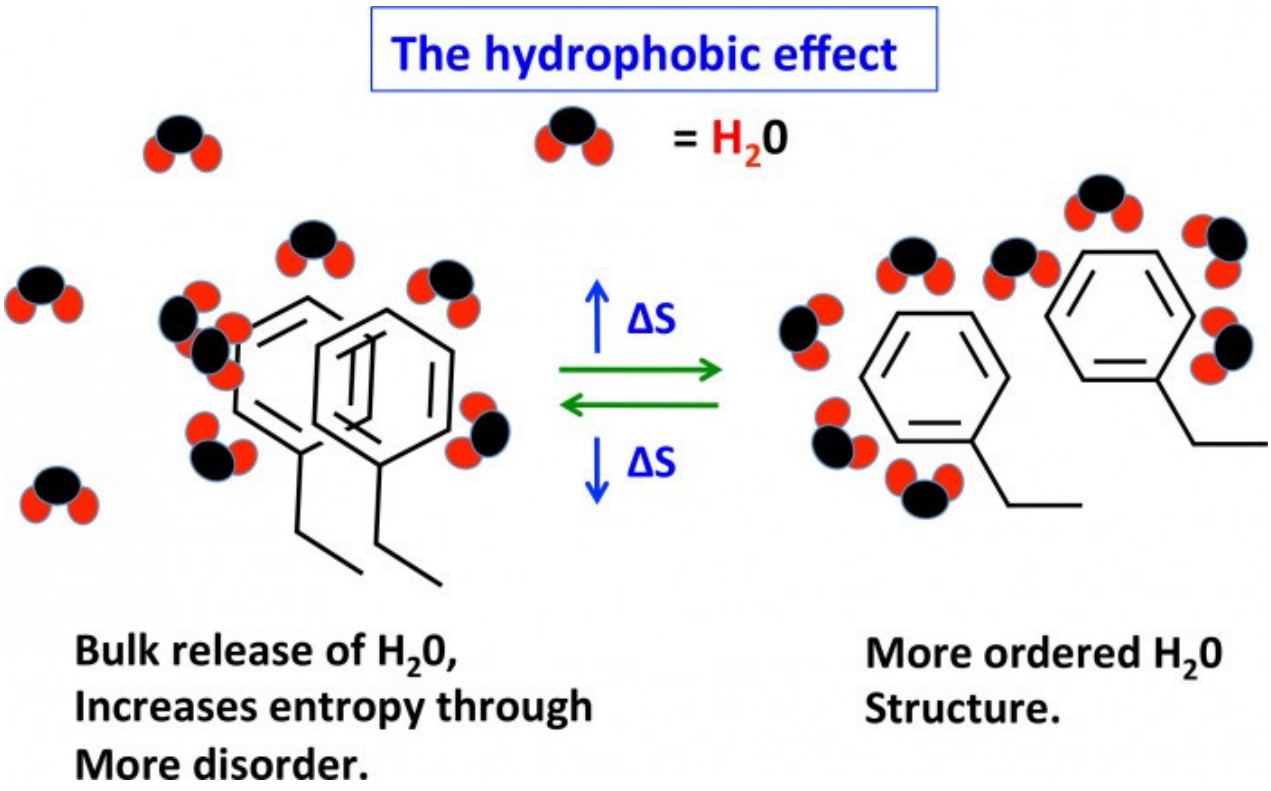
Tertiary structure. The tertiary structure (3°) of a protein represents the adopted three-dimensional configuration arrived at through protein folding. In some cases, co-factors can help facilitate protein folding. The 3° structure represents the lowest energetic state and is stabilized by various interactions including covalent (i.e. disulfide bridges), electrostatic (salt bridges), hydrogen bonding, and weaker van der Walls interactions.

How a protein adopts its 3D conformation is the subject of the field of protein folding. The process is spontaneous and assisted by chaperones (i.e. heat shock proteins) in many cases. Protein folding begins as the nascent amino acid chain is released from the core of the ribosome. Levninthal's paradox notes that if a protein were to arrive at its intrinsic 3° structure through sequential sampling of every possible Φ and Ψ angle, then it would take an amount of time greater than the age of the universe to properly fold. Thus, folding is driven by certain processes, including the "hydrophobic effect" as described below.

In the folded state, many proteins adopt a globular, compact structure composed of an inner core of hydrophobic residues and a surface occupied by hydrophilic, polar amino acids that interact with the aqueous solvent. Thus, hydrophobic residues limit their association with water, serving as the basis for the "hydrophobic effect". This results in a structure where water orderly surrounds hydrophobic structures. Intrinsic to the concept of order (and disorder for that matter) is entropy. Think of molecules as energy sources (i.e. kinetic energy of rotation and potential energy in chemical bonds). Entropy will therefore increase as they become more dispersed in space. As quoted from entropysite.ocy.edu, "Entropy change is the measure of how more widely a specific quantity of energy is dispersed in a process."

Apply this concept to two interacting phenylalanine (Phe) residues in a polypeptide chain in the primary sequence (see image below). The hydrophobic effect contributes to protein folding involving these two Phe residues. By virtue

of its hydrophobic benzene ring, Phe prefers to occupy the inner core of the folded protein in the final 3D configuration. Appreciate that the transfer of Phe from the hydrophobic interior of a protein to the hydrophilic exterior (or vice versa) will cause a change in the arrangement of the solvent (water) around the amino acids (i.e. dispersion). This hydrophobic effect is shown in the diagram.



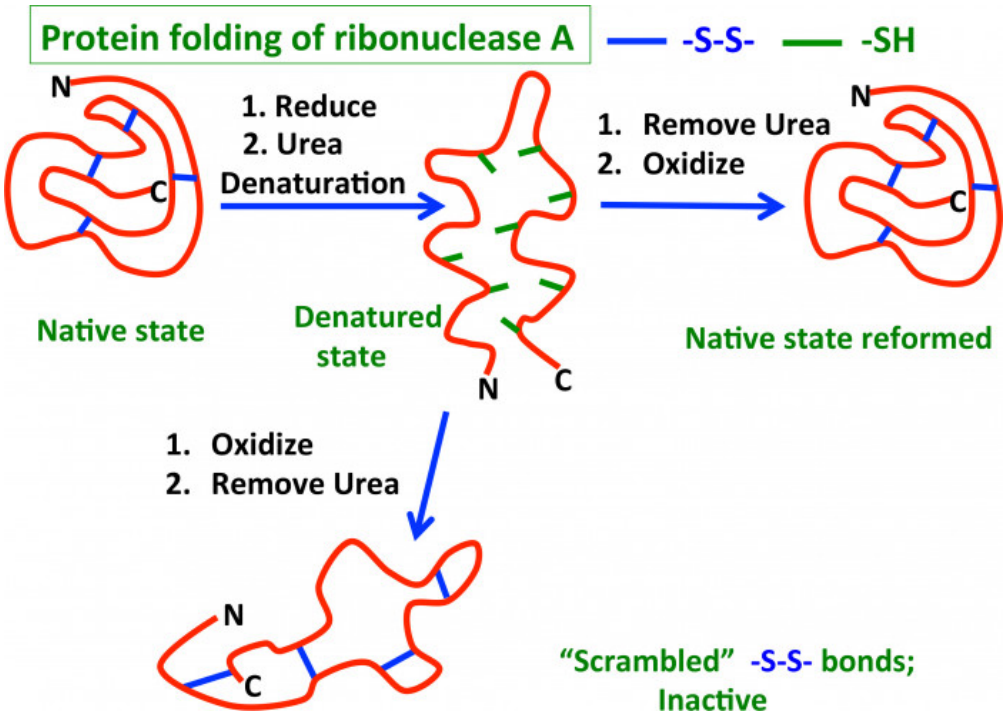
Think of the re-ordering of solvent molecules like water as an entropic effect. Because the hydrophobic Phe residues are being exposed to the aqueous solvent by virtue of translation, there would be an initial decrease in total entropy due ($\Delta S < 0$) to the ordered solvation of the Phe side chain with water. Note that through the aromatic rings in Phe --arranged in a base-stacking configuration through π-π interactions as shown-- that water molecules are released into the bulk solvent. This makes the interaction more energetically favorable as it increases entropy, or the number of disordered states.

1° structure determines the 3° structure.

Anfinsen showed that the primary amino acid sequence is sufficient for directing the three dimensional folding pattern (3° structure), the active form of a protein. This was demonstrated in an elegant experiment with ribonuclease A

(RNAse A), for which he was awarded the Nobel Prize. This important experiment is an MCAT staple and is outlined below.

If the native (active) structure of the RNAse enzyme is treated with a reducing agent (i.e. β-mercaptoethanol) in in the presence of urea, then the RNAse A enzyme becomes denatured and inactive. However, if the denatured enzyme solution is dialyzed to remove urea and then oxidized through exposure to air, then the native structure will reform as shown below in the image. This demonstrates that the 1° sequence information is sufficient for folding into the active, tertiary (3°) structure. The folding pattern of RNAse A is driven through the formation of disulfide bonds through the oxidation of cysteine residues. As a control experiment, when the denatured RNAse A enzyme is oxidized in the presence of urea, inactive, "scrambled" structures with inappropriately formed disulfide linkages are formed upon removal of urea.



Proteins with undefined structure

Not all proteins have ordered three-dimensional structures. Such intrinsically disordered proteins (IDPs) range from completely denatured states to random coils separated by partially folded domains. Random coils have no definitive structure and represent a distribution of shapes in a population. The only consistent structural feature is the polypeptide backbone composed of peptide bonds. IDPs are important in disease.

The tau protein is an IDP enriched in neurons where it stabilizes axonal microtubules. In the hyperphosphorylated state, tau adopts a helical structure that can "tangle" in neurons, abrogating their normal function. Such neuronal aggregates of tau are causal in multiple diseases (i.e tauopathies) including chronic traumatic encephalopathy (CTE) and Alzheimer's. A postmortem neuropathological examination of 111 brains from former American football players showed that 110 (nearly 100%) had CTE. In addition to tau, Alzheimer's disease is also driven by aberrant protein folding events derived from the amyloid precursor protein (APP). Under normal conditions, APP is a membrane protein enriched in some neurons. However, abnormal proteolytic processing of APP creates the amyloid β (Aβ) peptide that, in a fashion analogous to hyperphosphorylated tau, can aggregate at high concentration. At high concentrations the normally soluble Aβ peptide forms amyloid fibers that precipitate. This forms neuronal plaques, a causative feature in Alzheimer's disease.

Quaternary structure (4°)

Proteins exhibiting quaternary structure (4°) have multiple subunits (i.e. dimers, trimers, tetramers, and multimers) that often cooperate with each other to regulate function. Two classic examples are the enzyme aspartate transcarbamoylase (ATCase) and the oxygen carrier hemoglobin (Hb). Both of these proteins exhibit cooperativity, a phenomenon often seen in multisubunit protein complexes. Cooperativity is the condition observed when the binding site of a ligand (substrate) changes the affinity of a second ligand binding event. Both positive and negative cooperativity exist, and not all ligand binding events exhibit cooperativity. Cooperative binding can be quantitated by the Hill equation. This topic is frequently tested on the MCAT and is discussed in detail below.

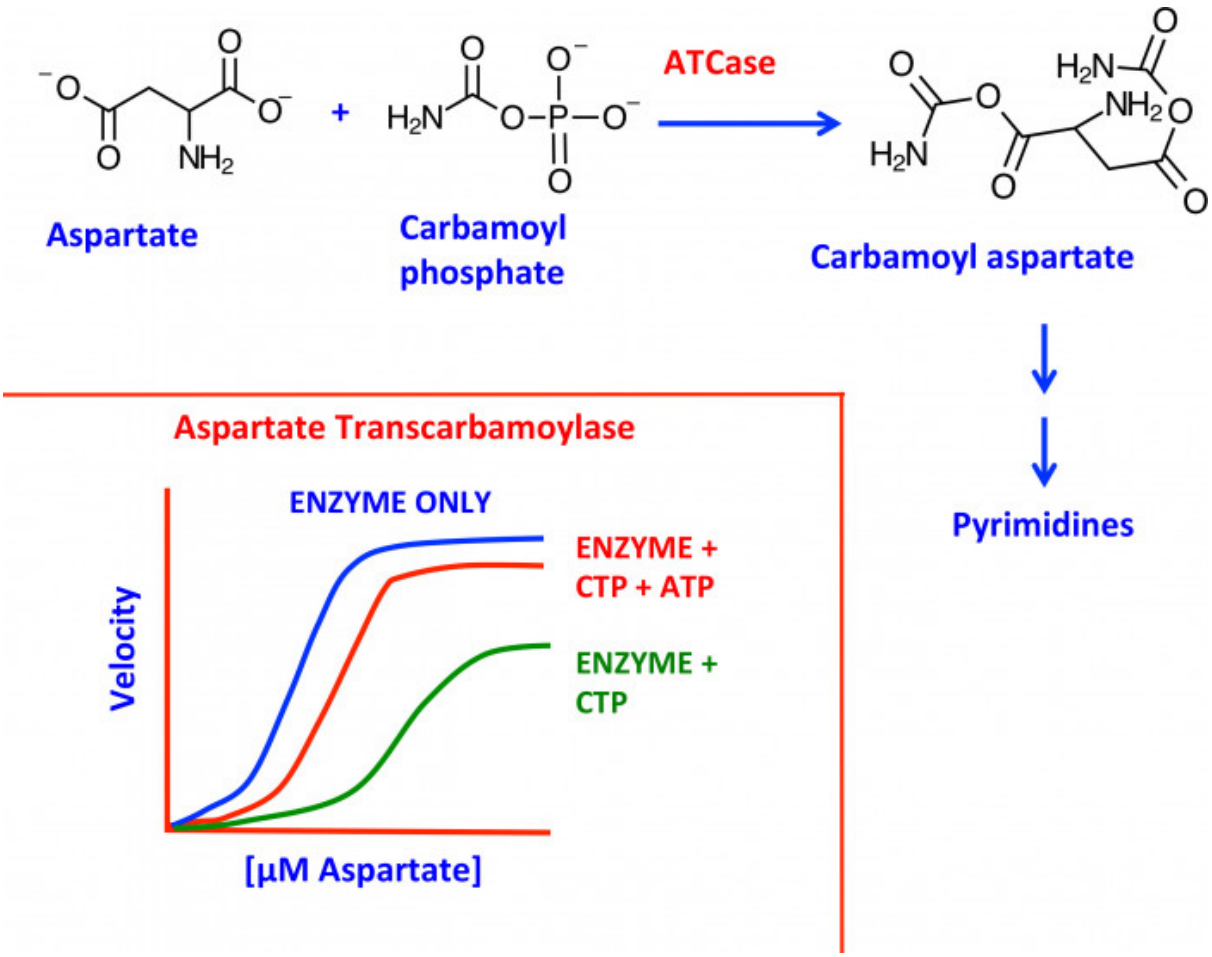
First, we will examine the regulation of two important proteins (aspartate transcarbamoylase [ATcase] and hemoglobin [Hb]). Both of these proteins have quaternary structure and are regulated through allosterism and cooperativity.

Aspartate transcarbamoylase (ATcase). This enzyme catalyzes the committed step in pyrimidine biosynthesis. ATCase is composed of 12 total subunits: six catalytic and six regulatory subunits (C_6R_6). ATcase catalyzes the formation of carbamoyl aspartate from aspartate and carbamoyl phosphate (see image). The structure and activity of ATcase is regulated by allosterism, the phenomenon where small molecule regulators alter protein (i.e. enzyme) function through

binding to a site distinct from the active site. Allosteric regulation can be either positive or negative and commonly occurs in proteins that exhibit quaternary structure.

A kinetic analysis of ATcase is shown and reveals that enzyme activity is altered in the presence of ATP and CTP. The reaction is conducted with increasing amounts of the substrate aspartate in the presence of saturating levels of carbamoyl phosphate, a second substrate. The effects of ATP and CTP on enzyme activity can be seen through measuring the K_M, the concentration of aspartate at maximal velocity (V_{max}). Such regulatory modulation of enzyme activity via nucleotides ensures balance in the production of nucleotides.

ATcase exists in two major forms: R (relaxed) and T (taught). The R form is an open, active form with high affinity for substrate (aspartate). In contrast, the T form is a low activity form with low affinity for substrate. Conversion from the T to the R form is accompanied by structural changes (i.e. van der Waals forces, hydrogen bonds, formation of helices). Note that the binding curve for ATcase and aspartate follows a sigmoidal curve. Such a curve is indicative of "cooperativity", the phenomenon where the binding of one substrate molecule influences the binding of subsequent molecules. In the case of ATcase, positive cooperativity is observed. Most instances of cooperativity involve multi-subunit proteins exhibiting quaternary structure.



Hemoglobin and myoglobin.

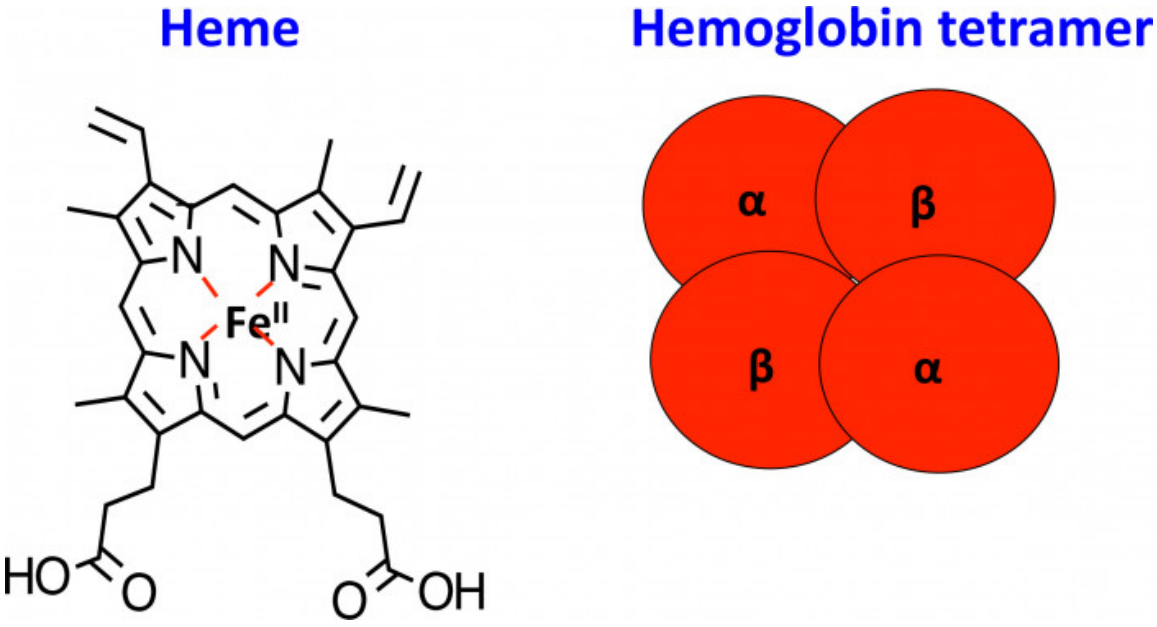
Hemoglobin (Hb) and myoglobin (Mb) both bind to heme and are classical examples of protein structure and function. Given their physiological importance and clinical relevance, you can expect to see them on the MCAT in any of a variety of ways.

As we know, the charge of side chains on amino acids is of fundamental importance to protein function and this principle is readily observed in Hb. The most famous example of this is the case of missense mutations in hemoglobin. Mutant substitution of the hydrophobic side chain valine for the acidic amino acid glutamate in hemoglobin underlies sickle cell anemia. The mutation changes the overall charge of Hb, effectively making the molecule more hydrophobic. The increased level of hydrophobicity causes the mutant hemoglobin subunit to polymerize, giving the red cell the classical sickled shape observed in sickle cell disease. Given the importance of hemoglobin and the

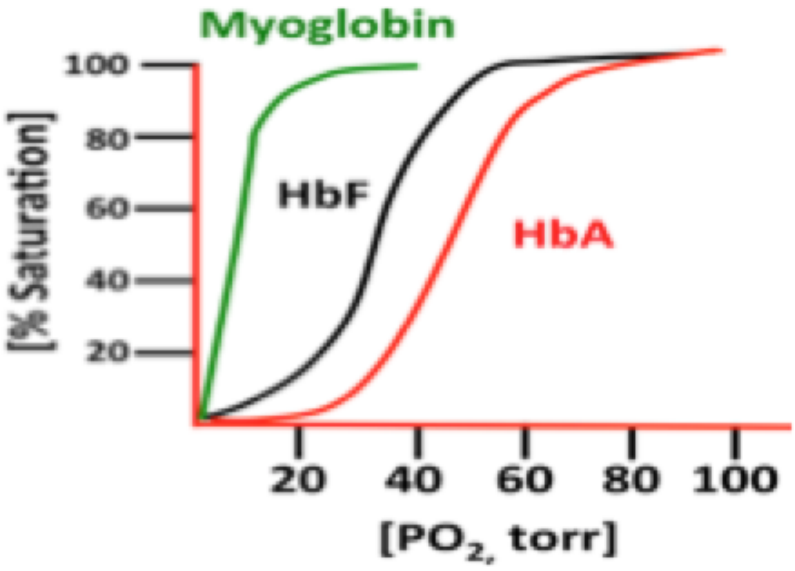
fact that sickle cell anemia is the first disease to be described with a molecular basis, it is quite likely that you will see some application of this, particularly with respect to structure and function on the MCAT.

Hemoglobin is a heterotetramer ($\alpha_2\beta_2$). It's quaternary structure and modes of regulation via allosterism and cooperativity make it the perfect protein for oxygen transport. Hb binds heme and is located in red blood cells (RBCs) where it functions as the principle O_2 carrying molecule in the body. Through a unique structure-function relationship, Hb is equipped to pick up O_2 in the lungs and deliver it to the peripheral tissues. During this cycle, Hb also picks up CO_2 and sends this waste to the lungs for expiration.

The ferrous II ion (Fe^{+2}) that occupies the central position in the heme structure forms coordination bonds with six distinct ligands, four of which are derived from the pyrrole nitrogen atoms in the backbone of the heme structure (see image below). The other two ligands are environmental oxygen and a histidine residue from Hb. Carbon monoxide competes for oxygen binding to heme iron. The bonding event between a diatomic O_2 and the Fe^{+2} ion in the heme moiety causes a reduction in the ionic radius of the metal that induces a conformational change involving a nearby histidine. This alters the angle between the subunit interfaces and breaks ionic bonds, making it easier for additional subunits to bind O_2 molecules. Thus, each O_2 binding site on Hb functions in concert with the other O_2 binding sites. This is the essence of "cooperativity".



The cooperative properties of HbA and HbF relative to myoglobin can be seen in their respective O_2-binding curves (see figure below). The sigmoidal nature of Hb is indicative of its cooperative behavior with respect to oxygen binding. Because the binding of ligands to proteins can alter their structure and activity through allosterism. Think of the R and T forms of Hb as allosteric forms of the protein. Note that the oxygen saturation curve for Mb fails to display cooperative behavior as it is not sigmoidal. The curves show the % of oxygen binding sites saturated with O_2 ligand as a function of the partial pressure (PO_2) of oxygen expressed in torr units (i.e. mm Hg). Two important PO_2 values to keep in mind are those for the lungs (~100 torr) and the peripheral tissues (~30 torr).



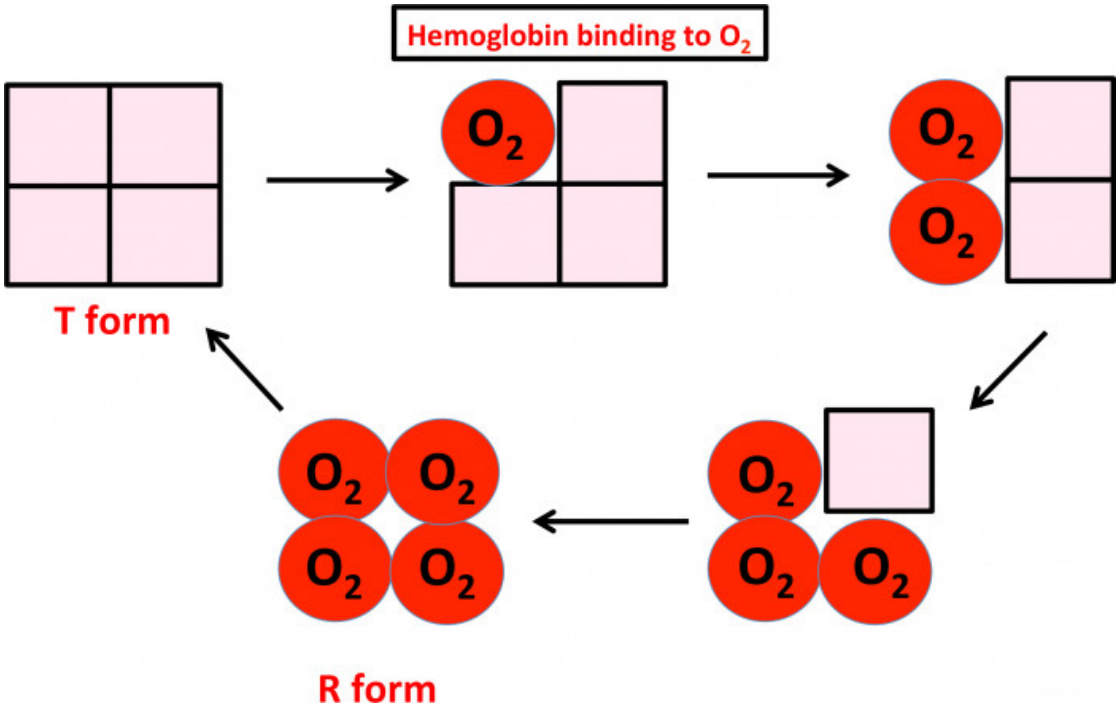
Myoglobin (Mb), also binds heme and has similar protein structural elements to Hb and is found in muscle where it is believed to serve as an oxygen storage resorvoir in this tissue. Myoglobin is monomeric while hemoglobin exhibits a quaternary structure which creates a large difference in the biochemical and physiological elements of O_2 transport.

Unlike Mb which is normally absent from blood, Hb levels are routinely measured in complete blood counts. Lab tests ususally measure the "mean corpuscular hemoglobin" or MCH value, a number that represents the amount of Hb per volume of red blood cells. Because > 90%of the mass of a red blood cell is composed of Hb, changes in this value have significant clinical implications in diseases such as anemia.

Similar to Hb, myoglobin also binds to a heme group with a central Fe^{+2} ion that binds O_2. However, unlike hemoglobin, Mb is a monomer and is therefore devoid of subunit-subunit interactions that cooperatively regulate the binding interaction between oxygen and Hb.

In contrast, adult Hb (HbA) is a heterotetramer ($\alpha_2\beta_2$) composed largely of α helices. Fetal Hb (HbF) is also a heterotetramer, but is composed of two subunits (γ) that are expressed during fetal development ($\alpha_2\gamma_2$). As shown in the image, the binding curve shows that HbF has a higher affinity for oxygen than HbA ("left shifted"). This can be readily observed by examining the pressure of O_2 required to achieve 50% saturation for each of the three proteins presented. Note that higher affinity binding proteins will achieve 50% saturation at lower pressures of oxygen.

Perutz determined the structure the HbA tetramer. This immediately revealed the link between structure, function, and oxygen delivery to peripheral tissues. As HbA picks up oxygen in the lungs and delivers it to peripheral tissue, the tetramer exists in two major forms: R (relaxed) and T (tense). Such allosterism is reminiscent for what we previously saw with aspartate transcarbamoylase. In the case of Hb, the R form has a higher affinity for oxygen than the T form. In the low affinity T form, Hb is in a structure resistant to O_2 binding (see image). As discussed below, the T form is stabilized through the binding of protons, carbon dioxide, and BPG.



The reversible binding of O_2 to myoglobin and hemoglobin can be represented as:

1) Myoglobin (Mb): $Mb + O_2 = MbO_2$

2) Hemoglobin (Hb): $Hb\text{-}T + 4(O_2) = Hb\text{-}R(O_2)_4$

where R and T represent the Tense (low affinity) and Relaxed (high affinity) states of hemoglobin. Notice that this nomenclature is not used for myoglobin as it is monomeric and does not exhibit cooperativity.

In the case of myoglobin, an oxygen equilibrium binding relationship can be described as:

$K_A = MbO_2/[Mb][O_2]$

Where K_A represents the equilibrium association constant. The reverse reaction is often represented in the form of the equilibrium dissociation constant K_D where:

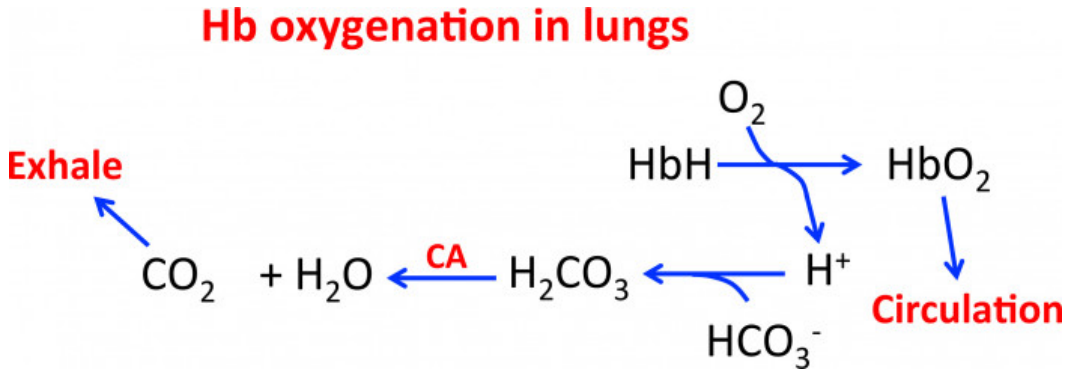
$K_D = [Mb][O_2]/MbO_2$

Notice that K_A and K_D have a reciprocal relationship. Low K_D values represent higher affinity ("tighter binding") between substrate and ligand, but low K_A values represent a low affinity for substrate. Be able to mentally flip back and forth between K_A and K_D; always appreciating the magnitude of the number and how it relates to the affinity between substrate and protein (or enzyme).

Hb binding to oxygen in the lungs.

As the partial pressure of O_2 increases in the lungs (from approximately 20 to 100 torr), O_2 begins to bind Hb. The binding of oxygen at one subunit facilitates the binding of additional subunits, changing the T form to the relaxed R form. Hb has four Fe+2 iron atoms bound in heme per tetramer, but the binding affinities for each O_2 are not equivalent. The binding of O_2 to one subunit of Hb makes it easier for additional molecules of O_2 to bind to the other subunits. This is an example of cooperativity. In this case, the cooperativity is considered positive as O_2 binding affinity increases the affinity for Hb with additional oxygen molecules.

Hb is oxygenated in the lungs and this is shown below. Here, the binding of Hb to O_2 releases a proton as shown. In this context, think of deoxygenated Hb as an acid (HbH). This free proton combines with red cell bicarbonate to generate carbonic acid (H_2CO_3). This acid dissociates through the action of the enzyme carbonic anhydrase (CA) and releases CO_2 for exhaling.



Iron-dependent binding of O_2 to Hb.

Each subunit of Hb binds O_2 via an iron-centered heme group, a prosthetic group that binds to each of the four subunits in the Hb tetramer. Thus, one molecule of Hb can bind four O_2 molecules. Oxygen binding to Hb induces changes in the conformation of the heterotetramer. This includes moving the position of the Fe^{+2} towards the center of the plane of the heme.

Heme is synthesized in the bone marrow and liver in several steps from a reaction that starts with the amino acid glycine and the TCA metabolite succinyl CoA. Accumulation of heme precursors is causal for a class of disorders known as "porphyrias", a condition that leads to severe gastrointestinal discomfort and mood disorders.

Oxygen dissociation from Hb in peripheral tissue.

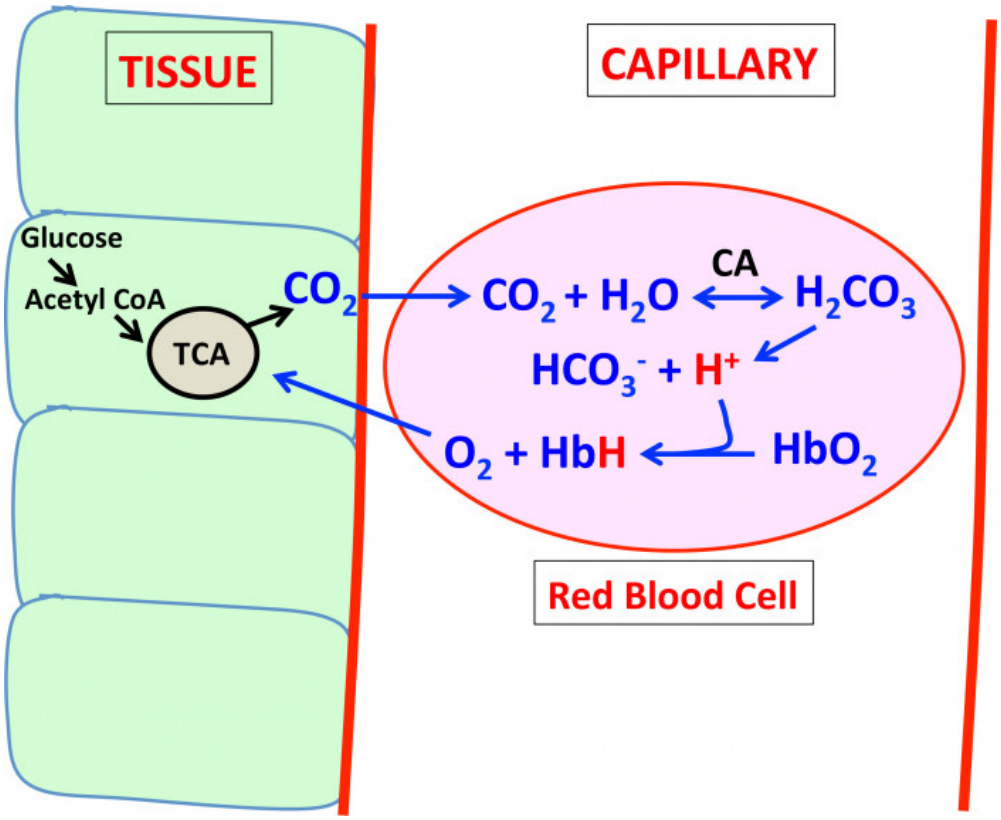
Various allosteric effectors, including acidic protons (i.e. Bohr effect) and the RBC metabolite "BPG", an acronym for 2,3 biphosphoglycerate, bind to hemoglobin and promote O_2 dissociation for delivery to tissue. These allosteric

modulators stabilize the T form. In addition, covalent modification of Hb by CO_2 also stabilizes the T form.

The Bohr effect

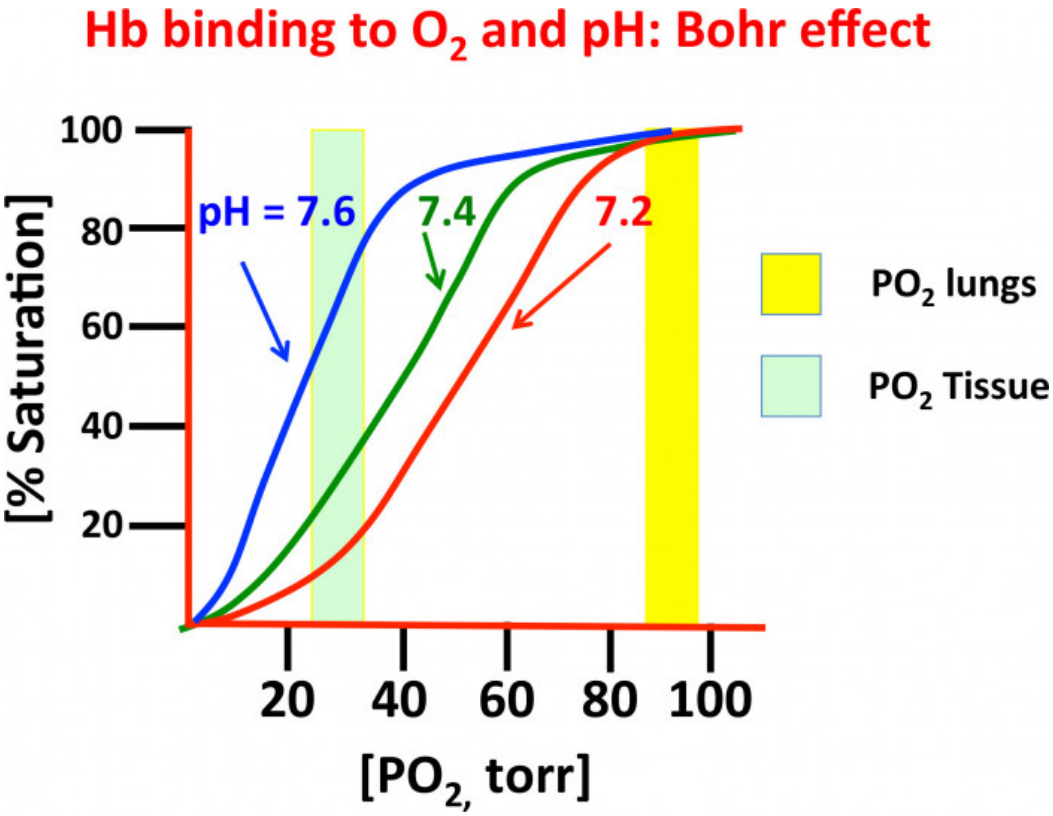
The ability of H^+ to induce dissociation of O_2 from Hb is known as the Bohr effect. Actively respiring tissues generate CO_2, largely in the mitochondria through the oxidation of pyruvate (pyruvate dehydrogenase [PDH] complex) and the carbon equivalents of acetyl CoA (TCA cycle). As shown in the figure, CO_2 diffuses into the capillaries and enters red blood cells.

Carbonic anhydrase (CA) catalyzes the formation of carbonic acid (H_2CO_3) from CO_2 and water. Carbonic acid dissociates into bicarbonate (HCO_3^-) and H^+. Thus, carbonic anhydrase is used in both the binding and release of oxygen.



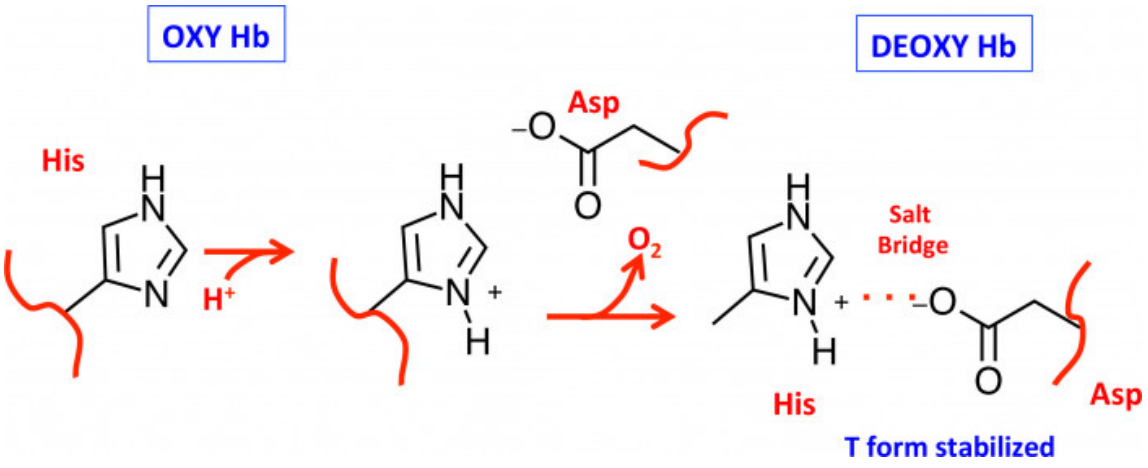
As shown below in the image, the pH influences the affinity for Hb and O_2. Appreciate that the oxygen saturation curve for Hb shifts to the right as the concentration of H^+ increases (pH decreases). As active tissues like respiring muscle generate acids through glucose oxidation (i.e. pyruvate and succinate), any Hb traveling to areas of lower pH will tend to release protons.

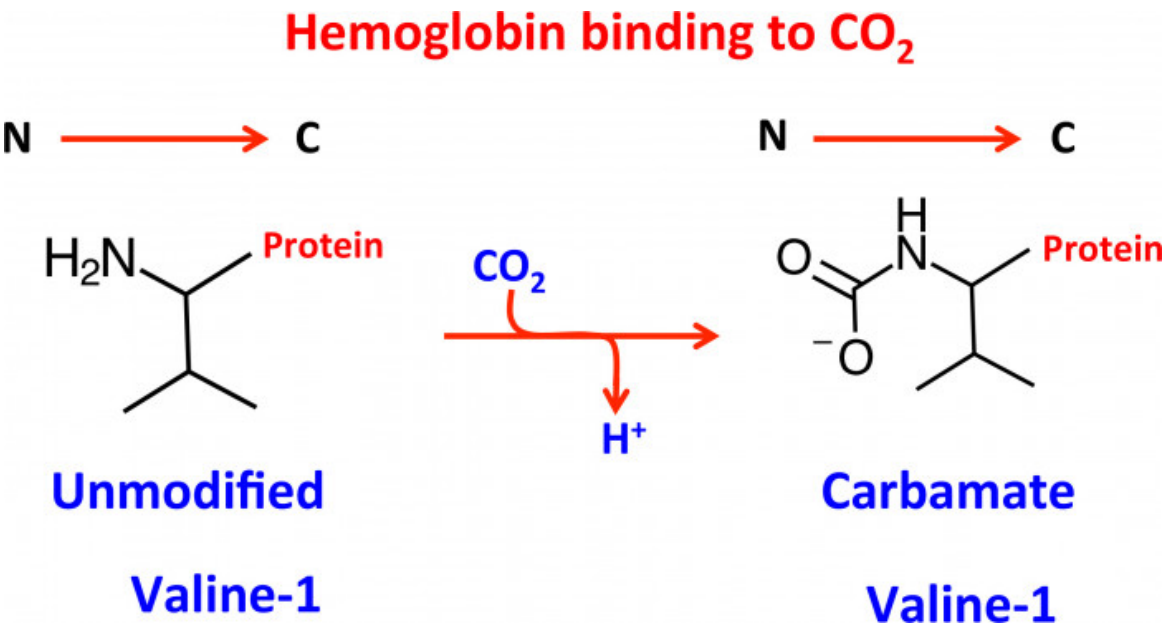
Understand how to interpret Hb binding curves. Note that in the interpretation of the pH binding curve, lower affinity is achieved by shifting to the right. For example, compare the amount of oxygen (P_{O2}) required to saturate Hb for each of the three shown pH values. As it takes more oxygen to achieve the same level of saturation as the pH decreases, it appears that the affinity between O_2 and Hb decreases as the $[H^+]$ increases.



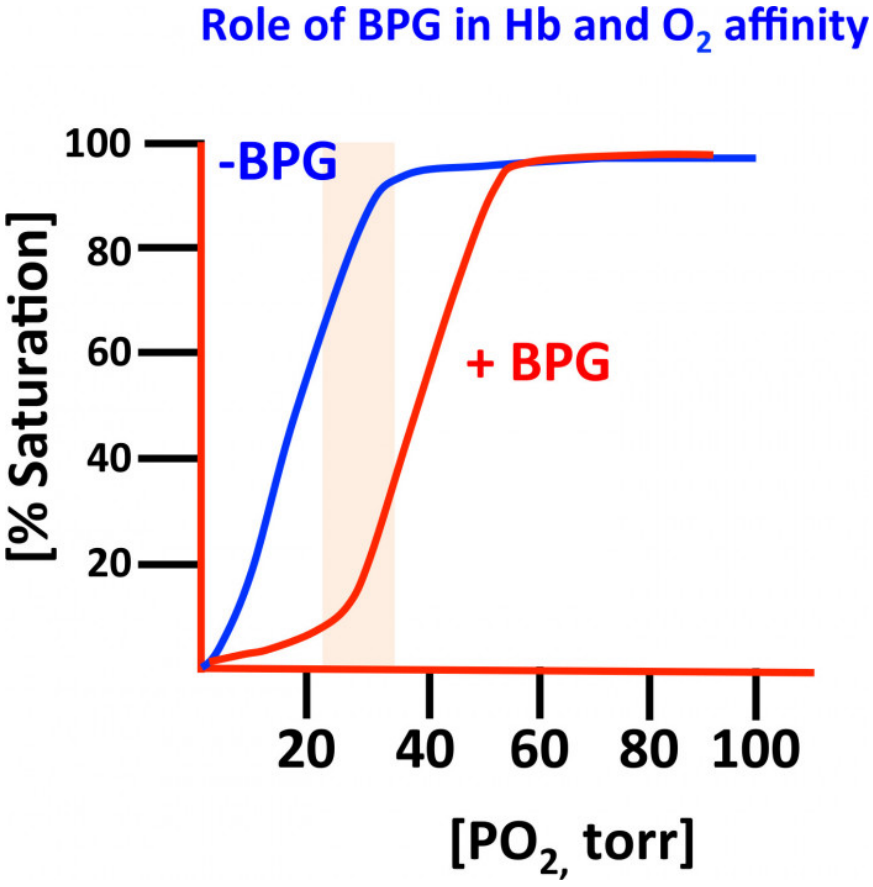
Protons released from carbonic acid can bind to a histidine side chain in Hb (generating HbH), a key factor that contributes to the release of O_2. Just like binding of O_2 to Hb, the release of O_2 is a cooperative process too. The oxygen diffuses back to the tissues, resupplying the mitochondria for further rounds of

electron transport and the generation of reducing power ($FADH_2$, NADH) and ATP from oxidative phosphorylation. As shown, proton binding to Hb creates a positive charge on the histidine side chain. The newly acquired positive charge on histidine residue 146 allows for the formation of a salt bridge with a neighboring aspartate residue 94, moving the equilibrium towards the low affinity "T" state (see image below). Further stabilization of the T form occurs through a covalent modification between CO_2 and an N terminal valine residue in Hb as shown below. This covalent modification is known as carbamate and is known to further stabilize the T form, promoting oxygen release.





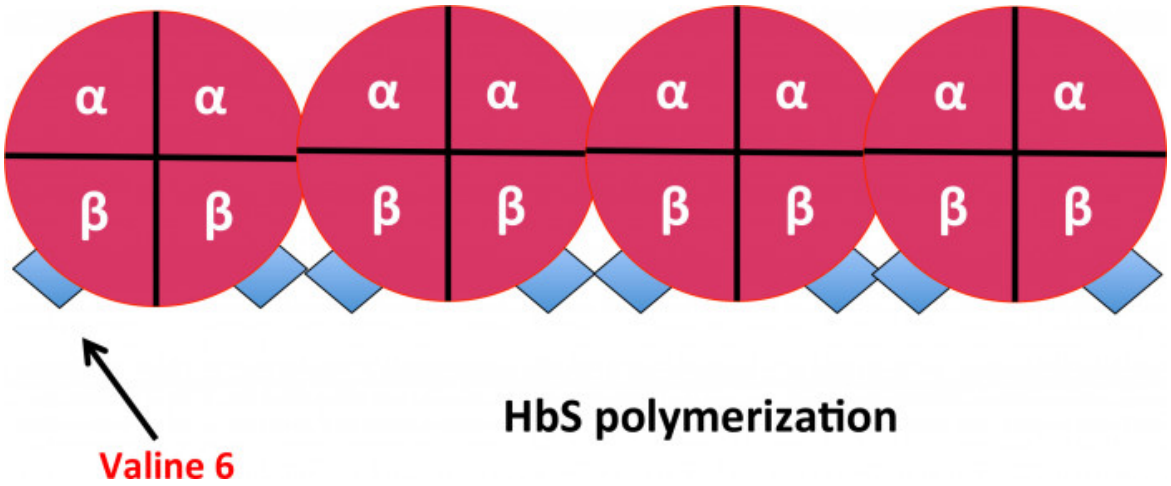
2, 3-BPG. 2, 3-BPG (or BPG as abbreviated here) is synthesized in red blood cells and stabilizes the T form of HbA. HbF, the fetal form of hemoglobin, is refractory to BPG binding. As a consequence, HbF has a higher affinity for O_2 than adult HbA. Consistently, Hb purified in the absence of BPG has a higher affinity for oxygen (i.e. left shift in the presented graph) than Hb purified in the presence of BPG (i.e. from red blood cells). The conclusion from these two observations is that BPG promotes the release of oxygen. BPG is synthesized from the glycolytic pathway in red blood cells. The enzyme biphosphoglycerate mutase (mutase) phosphorylates 1, 3 biphosphoglycerate, generating BPG. The same mutase enzyme also performs a de-phosphorylation reaction that feeds the carbon skeleton back into the glycolytic pathway.



Hemoglobinopathies

Various pathologies result from defects in hemoglobin function. Aberrant expression of Hb subunits gives rise to thalassemias and abnormal accumulation of heme precursors, which is causal for porphyrias. Further, a Glu to Val mutation at position 6 near the N terminus in the β subunit of HbA

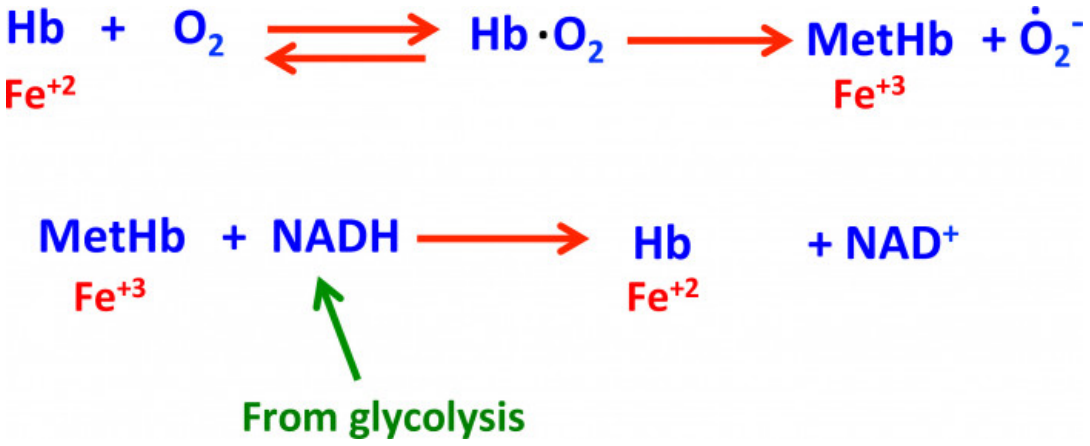
gives rise to sickle cell anemia. The introduction of the valine residue changes the charge on HbA, yielding a more hydrophobic tetrameric molecule called HbS. In the deoxy form (but not the oxy form), HbS forms polymers that precipitate in red blood cells, changing their shape to their characteristic sickling phenotype. Sickled RBCs occlude capillary flow and cause reduced oxygen delivery to tissues. Such hypoxic conditions trigger server pain and are referred to as a vaso-occlusive crisis that can lead to organ damage. One treatment for sickle cell anemia is to administer hydroxyurea, a drug that increases the expression of the fetal hemoglobin γ subunit. Incorporation of HbF subunits into the tetramer dilutes the levels of incorporated, diseased β subunits in the hemoglobin tetramer.



Methemoglobin

In order for Hb to bind oxygen, the heme iron must be in the reduced Fe^{+2} form (ferrous). However, oxygen is a highly reactive molecule that participates in numerous oxidation reactions. In the red blood cell, the oxidative formation of Fe^{+3} (ferric ion) generates methemoglobin (MetHb). This form of Hb gives blood a dark bluish, brown color. Methemoglobin reductase reduces Fe^{+3} to Fe^{+2} as shown.

The levels of MetHb increase under certain circumstances including pyruvate kinase deficiency. Recall that pyruvate kinase (PK) catalyzes the last step in glycolysis where both ATP and pyruvate are synthesized. Defects in the PK enzyme cause a build up of glycolytic intermediates, reducing the levels of NADH, the co-factor required for converting MetHb back into normal, oxygen-binding Hb.



The Hill coefficient

Many proteins and enzymes with quaternary structure exhibit cooperatively. The Hill coefficient (n) quantitates the extent of cooperativity in macromolecules. For hemoglobin:

$Hb + nO_2 = Hb(O_2)n$

In the case of hemoglobin, the maximal value of n can equal four as each tetramer binds to oxygen. Conversely, for myoglobin the Hill coefficient equals one as it displays no cooperativity. The Hill coefficient for HbA, the adult form, is measured to be approximately 3.0, but this number can change based upon mutation and experimental conditions (i.e. buffer and pH).

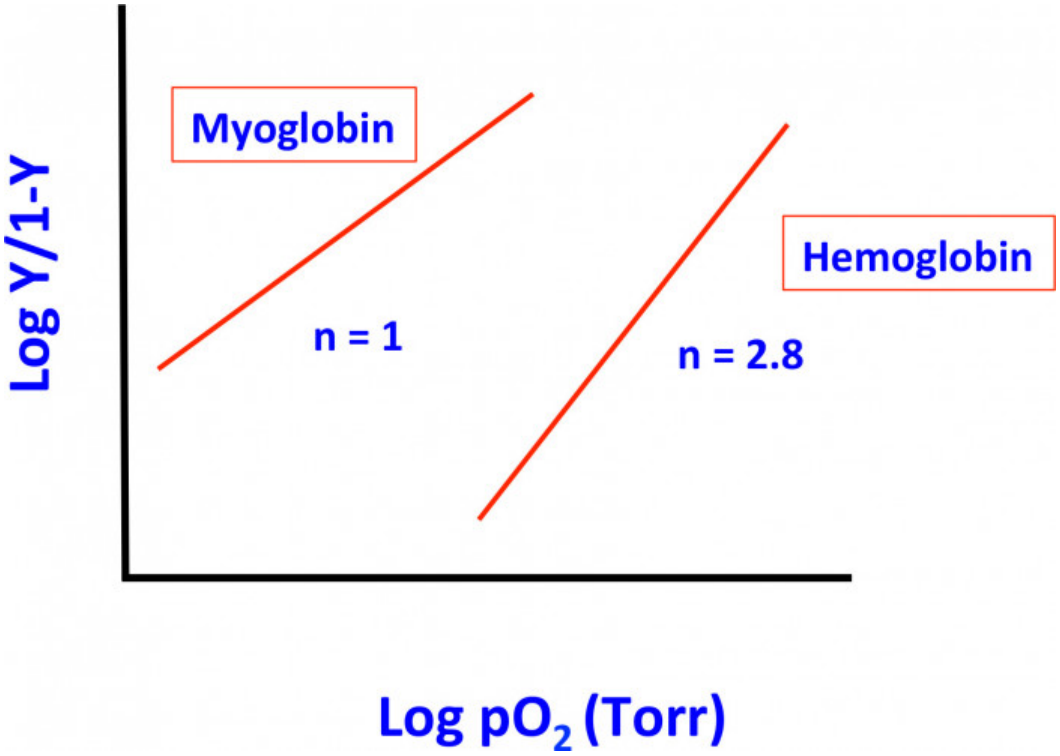
The Hill coefficient is experimentally measured through a Hill plot. Through manipulating the hemoglobin binding equation (not shown), the Hill equation can be expressed as:

$\Theta = [L]n = Kd + [L]n$

Θ = fraction of receptor protein bound to ligand
[L] = unbound (free) ligand concentration
K_D = dissociation constant

There are various ways to express and apply the equation. By graphing the log of the fraction of bound oxygen ligand (often represented as Y/Y-1) vs the log of the partial pressure of oxygen (pO_2), the slope will equal n, the Hill

coefficient. By measuring the rise/run, the graph below shows that the Hill coefficients for myoglobin and hemoglobin are 1.0 and 2.8, respectively.



Ligand binding to proteins: The Scatchard equation

Another well described example of a ligand binding to a protein is the interaction between a hormone and a receptor. In an ideal setting at equilibrium, the interaction between a hormone (H) and its receptor (R) can be designated as:

H + R = HR and K = [HR]/[H][R]

As the receptor is either unbound or bound to ligand, the total amount of receptor (HRT) can be expressed as:

HRT = [R] + [HR] where [R] and [HR] equal free and bound hormone, respectively.

From this, we can determine that the ratio of hormone bound to receptor [HR] to free hormone is:

[HR]/[H] = K([HRT] – [HR])

where K = equilibrium constant, or affinity constant.

The equation assumes several things including:

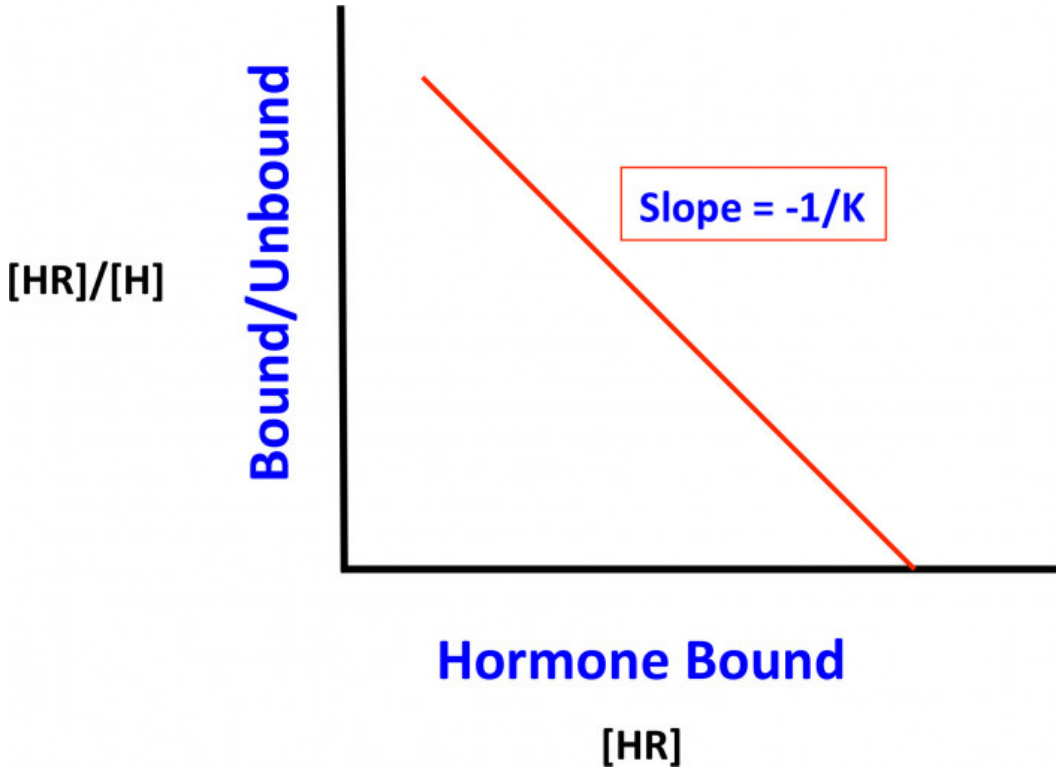
A. The reaction occurs at equilibrium

B. One hormone binds to one cognate receptor

C. Binding is reversible.

D. There is no non-specific binding. For example, the free hormone does not significantly interact with the cell membrane, fails to cross it, and is not taken up by pinocytosis etc.

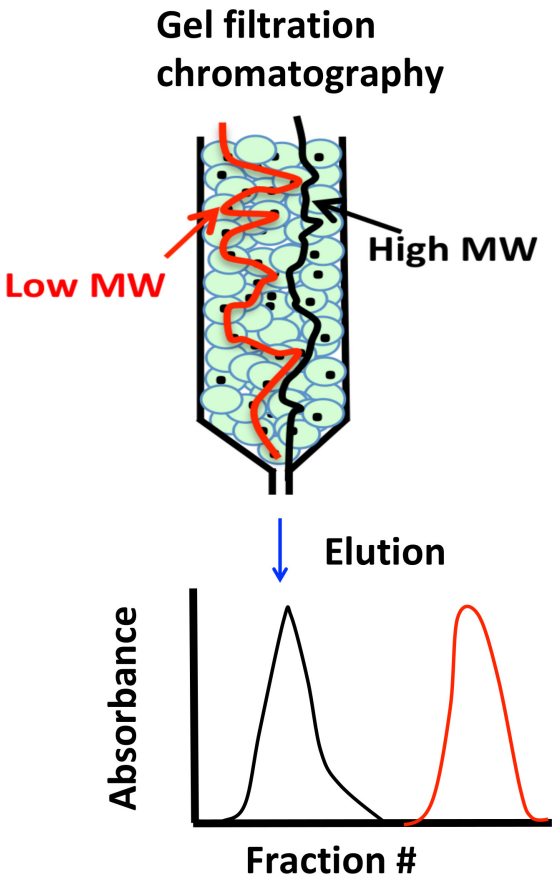
Think of this equation in terms of a linear relationship: y = mx +b. By plotting [HR]/[H] vs [HR], the slope is equal to -1/K.



Protein Purification

Proteins and peptides are purified by various means that exploit differential physical properties, including size, charge, and polarity. Three forms of chromatography exploiting these properties are described below. For the two examples of column chromatography, both high and low (i.e. gravity) pressure can be applied to move a sample through an absorption matrix.

Gel filtration (Size exclusion chromatography).



Agarose beads with defined pore sizes are used to separate molecules based upon differences in their molecular weights. This technique is most accurate with globular proteins. As the proteins flow through the column, lower molecular weight proteins are "trapped" within the pores of the agarose matrix. They are said to be "included". However, the higher molecular weight forms cannot fit inside the pore. The are considered "excluded". Therefore, smaller proteins elute last and the higher molecular weight proteins elute first.

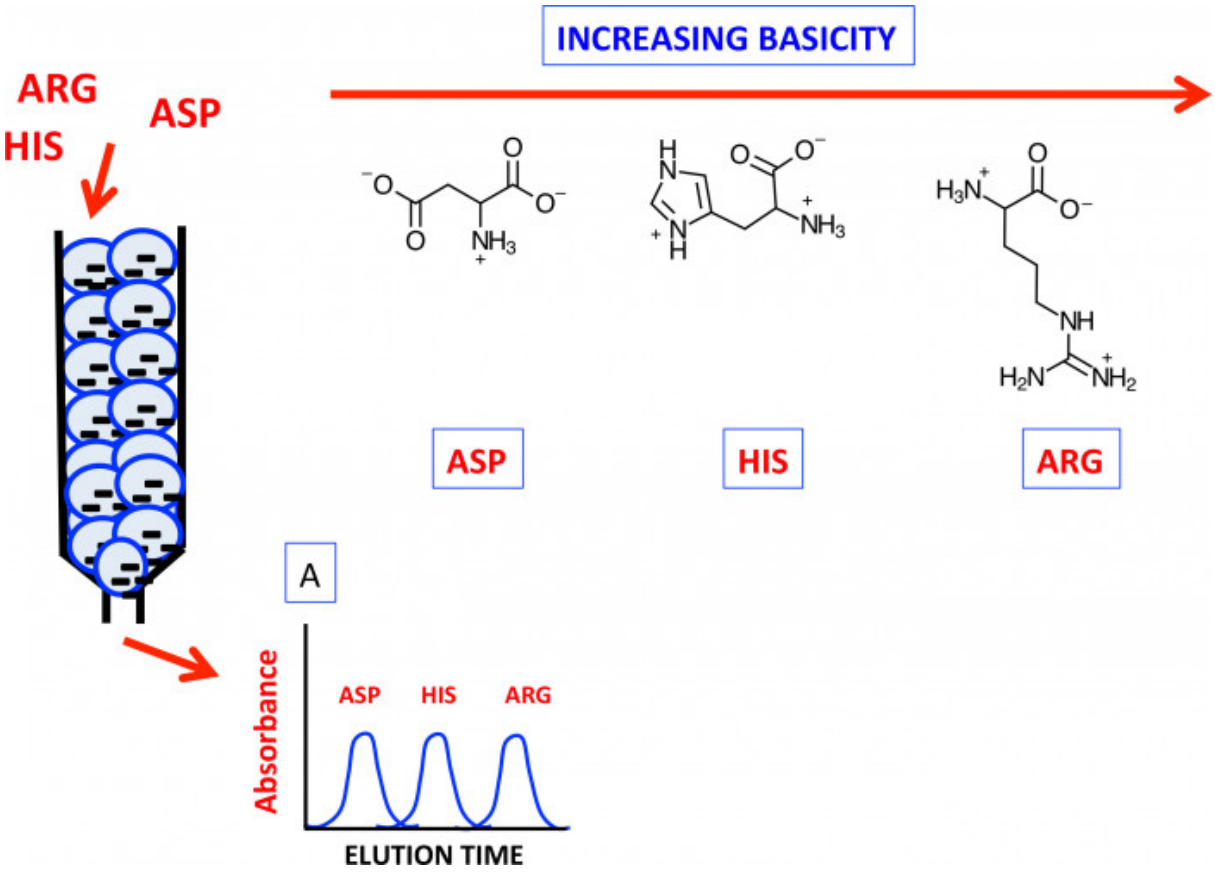
Ion exchange chromatography

Ion exchange columns are composed of a solid matrix coated with a charged species. As amino acids, peptides and proteins have multiple charges (polyacids), they will exhibit an overall charge at any given pH. A positively charged matrix exchanges anions and a negatively charged matrix exchanges cations.

For this example, two species with differing, yet overall negative, charges are loaded onto a positively charged matrix. This anion exchange column will bind the negatively charged species. Any net positively charged species present in the sample would not be expected to bind. It would therefore be found in the flow through fractions. After binding, the column is exposed to increasing levels of salt (NaCl). This agent serves to disrupt the interaction between the negatively charged protein species and the column matrix. As the two column-bound species vary in the magnitude of their negative charge, the least negatively charged species would be expected to elute first from the column. This is independent of the size of the molecule.

Take the case of a solution of three amino acids (ASP, HIS, and LYS) at pH = 6.0 that was subjected to an cation exchange column (negatively charged matrix). After binding of the amino acids, a pH gradient was use for elution? What is the order of elution of these amino acids

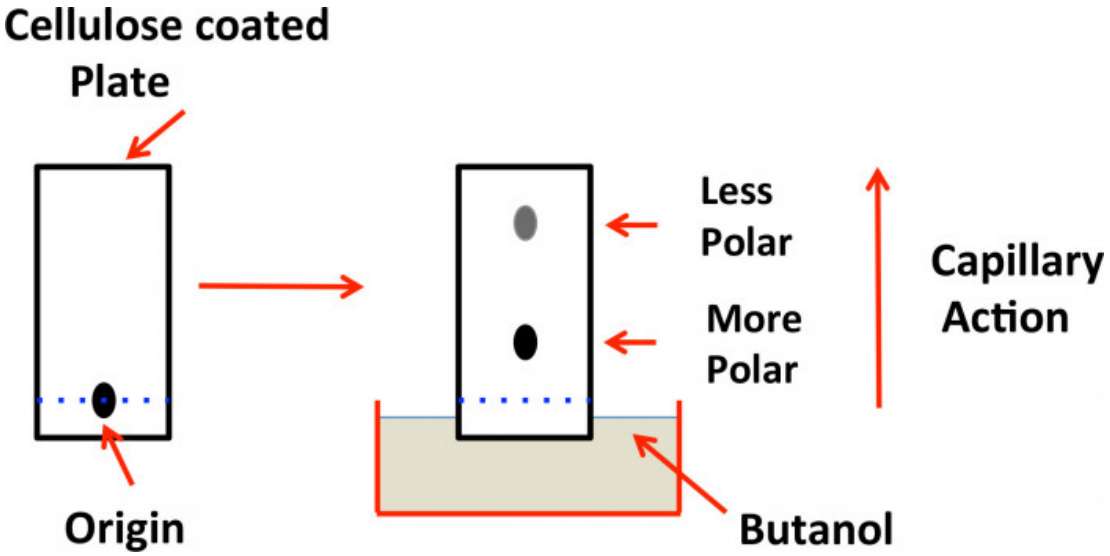
Appreciate that a cation exchange column is composed of a matrix consisting of negatively charged species (see image below). This means that those molecules with the least positive charge at pH = 6.0 (the pH of the buffer) will elute first. In other words, the least basic species (or most acidic) will elute first. At pH = 6.0, aspartate has a net charge of -1 and since the pKa of the histidine side chain is approximately 6.0, then ½ of the histidine molecules will be charged and the remaining 1/2 will be uncharged (determined from using the Henderson Hasselbach equation). Therefore, the net charge of histidine at pH = 6.0 is ~ +0.5. Both nitrogen groups in arginine would be positively charged as the pKa for each group is approximately 12. The net charge on arginine is therefore +1 (-1 + 1 + 1 = +1). Therefore, the amino acids in order of increasing positive charge (basicity) are ASP, HIS, ARG.



The binding step occurs at pH = 6.00. At this pH the negatively charged ASP would bind to the column with the least strength as it has only one positively charged nitrogen and two negatively charged carboxyl groups, ASP elutes first followed by HIS and then ARG.

Thin Layer Chromatography (TLC)

TLC is useful for separating amino acids and small peptides based upon differences in polarity. Detection is determined through radiolabeling or staining techniques such as ninhydrin for amino acids. In the example shown, a sample of two molecules varying in polarity are spotted into a cellulose coated TLC plate. The origin is marked as shown. Afterwards, the samples are immersed into a trough of butanol (a more nonpolar medium). Through capillary action, the molecules move up the plate. However, as polar molecules interact more favorably with the polar, cellulose surface, their mobility is hindered relative to more nonpolar molecules that disfavorbaly interact with the matrix. This is how separation is achieved. By determining the distance traveled from the origin as a function of time, the ratio to front (RF) value can be determined. A relatively lower RF value indicts a more polar molecule.


Cellulose coated
Plate
Origin
Less
Polar
More
Polar
Capillary
Action
Butanol