Amino Acids

Basic Acidity Concepts (did you like the pun?)
Remember the Henderson-Hasselbach Equation? Well, if you don’t, let’s re-derive it here!
As you know, the acid constant \( K_{a} \) is defined as:

\[
K_a = \frac{[A^-][H^+]}{[HA]} \Rightarrow -\log(K_a) = -\log\left(\frac{[A^-][H^+]}{[HA]}\right) = -\log([H^+]) - \log\left(\frac{[A^-]}{[H-A]}\right)
\]

As you know, pH is defined as pH = –log [H⁺] and likewise pKₐ = –log [Kₐ]. Therefore, we have:

\[
pK_a = pH - \log\left(\frac{[A^-]}{[H-A]}\right) \quad \text{OR} \quad pH = pK_a + \log\left(\frac{[A^-]}{[H-A]}\right) \quad \text{The Henderson-Hasselbach Equation}
\]

\[
pH - pK_a = \log\left(\frac{[A^-]}{[H-A]}\right) \quad \text{OR} \quad 10^{pH-pK_a} = \frac{[A^-]}{[H-A]}
\]

We see that the quantity (pH-pKₐ) directly correlates to acid strength. At a given pH, the higher the pKₐ, the smaller the ratio of [A⁻]/[HA], the weaker the acid. This makes sense because the smaller the concentration of the conjugate base, the weaker the acid. Hence, we can correlate pKₐ to acid strength. The more positive the pKₐ, the weaker the acid. The more negative the pKₐ, the stronger the acid. Remember that species with pKₐ < 0 are very strong acids! On the other hand, species with pKₐ > 5 are weak acids. Carboxylic acids are considered relatively weak acids.

Factors Affecting Proton Acidity
1) Electronegativity – The electronegativity of the atom or group to which a proton is bonded significantly affects its acidity. For example, a hydrogen bonded to an sp³ oxygen (EN ≈ 3.5) is much more acidic (pKₐ ≈ 15–16) as compared to a hydrogen bonded to an sp³ carbon (EN ≈ 2.5) (pKₐ ≈ 45–50). Likewise a hydrogen bonded to an sp² carbon (EN ≈ 3.0) is much more acidic than a hydrogen bonded to an sp³ carbon.

2) Molecular Orbital Overlap – An atom’s bond to hydrogen significantly decreases in strength as we move down a given column (or family). For example, although sulfur (EN ≈ 2.5) is less electronegative than oxygen, thiols (pKₐ ≈ 10–12) are much more acidic than alcohols. This results from the poor molecular orbital overlap between the H1s orbital and the S3sp³ orbital due to size and energy constraints.

3) Inductive Effects – electronegative substituents in the vicinity of the hydrogen can influence a hydrogens acidity. This is because electronegative substituents stabilize the conjugate base by withdrawing electron density from the negatively charged deprotonated atom. Thus, electronegative substituents (F, O, N, etc.) in the vicinity of a proton will reduce its pKₐ.

4) Resonance Stabilization – Once again, this is an effect on the stability of the conjugate base. If the conjugate base is resonance stabilized, as in enolates (pKₐ ≈ 9–13) or phenols (pKₐ ≈ 10) like Tyrosine, the proton’s acidity will also greatly increase!

5) Solvent Effects – These are very important. Notice that all the cited pKₐ on the next page are in water. This is convenient because proteins are synthesized in aqueous solutions. However, the hydrophobic core of a protein is a very different environment. The
conjugate base of any acid is highly destabilized, because charges are not dissipated in non-polar environments. For this reason, the $pK_a$ of any species increases in the hydrophobic core of a protein as compared to the cytosol.

**The Amino Acids and $pK_a$**
There are 20 naturally occurring amino acids. I would memorize their structures (it’s not so hard and it will save you time on the exam!) Know their $pK_a$s and functionalities (functional groups)!

| Carboxy Terminus | 2.0-2.6 (predominantly deprotonated under physiological conditions) |
| Amino Terminus | 6.8-8.0 (variable protonation states at physiological pH) |
| Aspartic Acid | 3.9 (predominantly deprotonated under physiological conditions) |
| Glutamic Acid | 4.1 (predominantly deprotonated under physiological conditions) |
| Histidine | 6.1 (variable protonation states at physiological pH) – USUALLY BASIC |
| Cysteine | 8.0 (variable protonation states at physiological pH) – NUCLEOPHILIC |
| Tyrosine | 10.0 (predominantly protonated under physiological conditions) |
| Lysine | 10.5 (predominantly protonated under physiological conditions) |
| Arginine | 12.5 (predominantly protonated under physiological conditions) |

**Amino Acid Structure**

All natural amino acids (except Glycine which is achiral) are **L-amino acids** (they rotate plane-polarized light counterclockwise). They all have the same relative stereochemical configuration (as shown to the left), and all, except Glycine (which is achiral) and Cysteine, have **S absolute stereochemistry** at the $\alpha$-carbon. Because of the particularities of the Cahn-Ingold-Prelog rules, Cysteine’s thiol group takes priority over the carboxylate, thereby giving it **R absolute stereochemistry** at the $\alpha$-carbon. Remember that the amino acid exists as a **zwitterion** (a doubly charged neutral species) at physiological pH. This is because the carboxylic acid has a $pK_a$ less than the physiological pH and the **protonated amine** has a $pK_a$ greater than the physiological pH.

**Enantiotopic Hydrogens – A Look at Glycine**

Glycine, shown to the left, is the simplest of amino acids. Its R group is a hydrogen, and for this reason, it is achiral. However, the two hydrogens are **enantiotopic**, meaning that they behave differently in chiral environments (such as the active sites of enzymes!). There is a nomenclature associated with this **prochirality**, with each hydrogen labeled as **pro-R** or **pro-S**. If you replace the pro-R hydrogen with a deuterium, you get an R stereocenter. If you replace the pro-S hydrogen with a deuterium, you get an S stereocenter. Verify this using the diagram above for glycine.

**Additional Reading**

**Further Suggested Reading:** Bugg (Chapter 2) Focus on pages 8 – 12.
Conformational Analysis

The Ethane Barrier

Staggered ethane is 3.0 kcal/mol more stable than eclipsed ethane. As you can see from the space filling models above, this is most definitely not the result of steric clash between the hydrogens in the eclipsing state because the Van Der Waals Radii of two eclipsing hydrogens do not interact. Rather, it is a purely electronic reason: there is better hyperconjugative overlap between the $\sigma$ C-H bond and the $\sigma^*$ C-H antibond in the staggered conformation, as you can see in the diagram above. Since there are three such interactions, we conclude that each eclipsing interaction costs 1 kcal/mol energetically.

Propane: The Me-H Eclipsing Interaction

As we notice, the propane barrier is slightly larger, at 3.4 kcal/mol. Since each hydrogen eclipsing interaction costs +1.0 kcal/mol, we conclude that the H-Me eclipsing interaction costs +1.4 kcal/mol (out of which 1.0 kcal/mol is an electronic cost and 0.4 kcal/mol is a steric clash).

Gauche Butane

With butane, we have the new gauche butane interaction, shown below which costs 0.9 kcal/mol. The gauche butane conformer is +0.9 kcal/mol less stable than the anti conformer.

n-Butane Torsional Energy Profile
Syn-Pentane

Pentane has several conformers, as shown to the left. Draw each of these in Newman Projection for your own gratification to understand the steric and electronics that dictate the relative energetics. The most unstable conformers are the SYN-PENTANE (g+-g- and g--g+ on the diagram). You can see how the Syn-Pentane results more clearly in this diagram to the left. If you start with a cyclohexane and remove the sixth carbon, replacing it with two hydrogens, you get a synpentane. Recognize that both hydrogens are in the same place – the location of the removed carbon! The two hydrogens are occupying the same physical space – this is extremely destabilizing!

Cyclohexane – Axial and Equitorial Conformations

The chair is the most stable conformation and is the conformation we will deal with for the most part. The boat will pop up occasionally in various contexts. With cyclohexanes, analyzing the most stable conformation becomes tricky, especially with multiply substituted rings. Let us first examine mono-substituted cyclohexanes, where there are two conformations: the substituent-axial chair and the substituent-equatorial chair. As shown to the left, a chair can have the substituent (R) in the axial or equatorial position. The difference in energy between the two conformations differs based on the substituent and is called the A-value for that substituent. There is a common chair-flipping convention, and the carbon mapping is shown in full-color-coded fashion above. Use this convention when interconverting chairs, as I have shown above with the R-cyclohexane. Once you move past mono-substituted cyclohexanes, it becomes trickier to estimate the energy. Generally, maximizing the equatorial substituents is going to be favorable. For a 1,3 substituted system,
when both substituents are cis, you get a SYN-PENTANE interaction in the diaxial conformation. This is called **1,3 diaxial interaction**, but it is nothing more than a syn-pentane (convince yourself of this). Vicinal substituents (1,2) and geminal substituents (1,1) can also be tricky. Watch out for the acyclic conformational parameters that I have previously discussed (gauche butane, syn pentane) to determine the relative stabilities!

**Alkenes – A-1,3 Strain**

With alkenes, we have a new interaction called the **allylic strain**, in which the double bond preferentially takes an eclipsing interaction with the least sterically taxing substituent on the **allylic carbon**. First examining the simplest system, propene, as shown to the left (X = CH₂). There are two interactions that we must analyse in this system. The first, and more obvious interaction, is the eclipsing interaction between the two protons in the **staggered conformation**, which becomes favourably anti-periplanar in the **eclipsed conformation**.

Take note of the terminology: **staggered** and **eclipsed** specifically refer to the alkene configuration! This is worth the usual +1.0 kcal/mol energetically, as we see with the ethane barrier. The other interaction is a new interaction between the **π system of the alkene and the two σ C-H bonds** (or two σ C-R bonds in the case of a substituted molecule) on the allylic carbon. In the staggered conformation, there is a destabilizing interaction between the **filled π-orbital (bonding)** and the **two σ C-H (bonding)**, as shown above, whereas in the eclipsed conformation, there is a stabilizing hyperconjugation between the **empty π*-orbital (antibonding)** and the **two σ C-H (bonding)**. These are both shown above. Each interaction is worth +0.5 kcal/mol (adding up to +1.0 kcal/mol for both interactions). Think carefully about why this hyperconjugation is only worth 0.5 kcal/mol whereas the hyperconjugation between a σ and a σ* orbital (as in the two hydrogens) is worth 1 kcal/mol. An alkene will always take on a conformation which places the least sterically demanding substituent on the allylic carbon parallel to the alkene, as shown above to the right. When we have a *cis*-1,2-disubstituted alkene, as shown above, there is an especially strong energetic preference for the least sterically demanding substituent to be in-plane with the alkene. This is minimization of **A 1,3 strain**; the right two structures are essentially isoenergetic and very unfavorable (compare this interaction to the syn-pentane interaction discussed earlier). This is a very powerful means of **acyclic stereochemical control on conformation** and is extremely important in **peptide and protein conformational analysis**.

**The Diamond Lattice**

The diamond lattice is a convenient way to draw conformations for tetrahedral carbon. The diamond lattice is nothing but several cyclohexane chairs all connected to ensure tetrahedral
The Peptide Backbone – A1,3 Strain Minimization

The peptide is essentially a substituted polyamide. The amide, as you know, has significant C=N double bond character with a large rotation barrier of 20 kcal/mol (due to the amide conjugation), and thus the C=N bond must be treated as a double bond, which must follow all of the allylic strain minimization parameters. Verify that the structures shown below are indeed those in which allylic strain has been minimized (there are two different allylic interactions – find both of them in the peptide backbone)!

Energetics Summary

1) H-H eclipsing interaction – 1.0 kcal/mol (σ-σ* MO overlap, NO sterics)
2) Me-Me gauche interaction – 0.9 kcal/mol (pure sterics)
3) H-Me eclipsing interaction – 1.4 kcal/mol (ethane barrier + sterics)
4) Me-Me eclipsing interaction – 3.1 kcal/mol (ethane barrier + significant sterics)
5) Synpentane interaction – 3.7 kcal/mol (pure sterics)
6) A-1,3 Strain – 3.5 – 3.9 kcal/mol (sterics for large groups; favorable σ-π* MO overlap in eclipsed conformation, unfavorable σ-π MO overlap in staggered conformation; H-H eclipsing contributes 1 kcal/mol)
7) Chair Cyclohexane Me A-value – 1.8 kcal/mol (pure sterics – 2 gauche interactions)

The Amino Acid Side-chains

You must keep in mind all of the aforementioned rules when constructing amino acid side chains. Start with the peptide backbone as drawn above. Remember that methionine is more flexible because C-S bonds are long and therefore less sterically demanding (the electron density is much more diffuse). Otherwise, you are all set to go!
1) Valine

The low energy conformation of valine contains two gauche and two anti interactions. The high energy conformations of valine contain three gauche and one anti interaction. Thus the energy difference between the two is about +0.9 kcal/mol.

2) Leucine

The low energy conformations of leucine avoid syn-pentane interactions. Rotations of \( \chi_1 \) and \( \chi_2 \) lead to the creation of syn-pentane interactions. Two of these rotations are shown, although there are more. The two low energy conformers are isoenergetic and thus equally populated.

3) Isoleucine

Here notice that isoleucine has a second stereocenter which restricts its conformations! Be careful to avoid inverting this stereocenter! Isoleucine is a very rigid amino acid because rotation of \( \chi_1 \) generates two gauche interactions while rotation of \( \chi_2 \) generates a syn-pentane interaction. Therefore, isoleucine is 95% populated by this single low energy conformer.

4) Methionine

Methionine is a conformationally flexible amino acid, a result of the increased length of the C-S bond relative to the C-C bond due to the larger and more diffuse S orbitals. This greater length and associated more-diffuse electron density greatly diminishes steric effects that lead to conformational bias. Thus we see a large variety of conformers. For this reason, many general enzymes that must accept a wide variety of substrates (such as chaperonins, signal recognition particles, and others) incorporate this flexible, hydrophobic, amino acid into their active site.

**Further Reading:** Chem 206 Lecture 4-6:
http://www.courses.fas.harvard.edu/~chem206/Fall_2005/Lectures_and_Handouts/
Protein Folding

The Theoretical Basis

1) Gibbs Free Energy ($\Delta G$) – determines the spontaneity of a given physical or chemical process

$$\Delta G = \Delta H - T\Delta S$$

1) A process is spontaneous if $\Delta G < 0$ (exergonic)
2) A process is non-spontaneous if $\Delta G > 0$ (endergonic)

2) Entropy ($\Delta S$) – Usually defined as ordering. Increases as systems become more disordered – the greater the degrees of freedom in a system, the larger the entropy! For example, hydrogen bonding requires proper positioning of the donor and acceptor atoms, reducing their entropy. However, going from one hydrogen bonded state to another causes no difference in entropy because similar special constraints are placed both before and after the change. Notice that as temperature increases, the $\Delta S$ becomes a greater and greater part of the free energy determination!

3) Enthalpy ($\Delta H$) – Changes only when new bonds are formed or broken.

Notice that a POSITIVE $\Delta H$ makes a process LESS SPONTANEOUS but a POSITIVE $\Delta S$ makes a process MORE SPONTANEOUS. We must minimize $\Delta H$ and maximize $\Delta S$ to make a process more spontaneous.

Hydrogen Bonding

Does not affect $\Delta S$ because amino acid residues hydrogen bond to water before folding and hydrogen bond to other amino acid residues after folding. Since there is no net change in hydrogen bonds, there is no significant change in entropy. Similarly, because no new bonds are formed (the same number of hydrogen bonds are present before and after folding), there is no change in $\Delta H$. Thus there is no change in $\Delta G$. However, hydrogen bonding is essential to ensuring the proper folding, which maintains the most efficient hydrogen bonding for each amino acid. There are two important generalizable and regular modes of hydrogen bonding: $\beta$-sheet formation and $\alpha$-helix formation.

Verify for yourselves that A1,3 strain is minimized in both of these depictions!
**Electrostatic Interactions**

Salt Bridges between oppositely charged residues (for example between aspartate or glutamate and arginine or lysine) also **contributes to ensuring the overall fidelity of the folding of a protein**, but because charges are well solvated in water, these interactions do **not contribute significantly to the spontaneity of the process**. The charge would be equally stable solvated by water or electrostatically interacting. **Π - Cation** interactions between aromatic residues and positively charged amino acid side chains are also important to fidelity but not to spontaneity.

**Disulfide Bonds**

These are strong S-S bonds that are made by oxidizing two cysteine residues. The bond is much stronger than the two original S-H bonds, so there is a negative $\Delta H$. However $\Delta S$ is also negative because disulfide bonds significantly constrain a protein’s degrees of freedom by linking two distant parts of the polypeptide. Thus there is **little change in $\Delta G$**. However, they **contribute again to fidelity of folding**. Remember the **90° Conformation**. In the gauche conformation, there is maximum overlap between the S Lone Pair (a good donor orbital) and C-S $\sigma^*$ orbital (a good acceptor orbital). However, this is sterically demanding, as you know from analysis of gauche-butane. The anti-conformation, on the other hand, minimizes the steric clash, but does not have any stabilizing hyperconjugative effects. The **90° Conformation**, shown in the middle, is the most stable conformer that minimizes the energy based on these two opposing forces.

**The Hydrophobic Effect**

This is the major process contributing to the spontaneity of folding. Water molecules form highly ordered ice-like lattice cage called **clathrate water** around non-polar residues (valine, phenylalanine, etc.) to diffuse the repulsion. These are especially strong hydrogen bonds which hold water rigidly in place around the non-polar residue. When folding occurs, these stable bonds are broken, increasing $\Delta H$. However, breaking the bonds makes the system much less ordered, so $\Delta S$ is also increased significantly. Examining the equation for Gibbs Free Energy, $\Delta G$, we see that at low temperatures, $\Delta H$ dominates, and that at high temperature $\Delta S$ dominates. Looking back to the first section of this handout, we see that minimizing $\Delta H$ and maximizing $\Delta S$ will maximize the spontaneity. **Thus, at lower temperatures, there is less spontaneity and at higher temperatures there is more spontaneity.** Protein folding is more spontaneous at high temperatures because of hydrophobic interactions.
Protein Denaturation

1) **Chemical Denaturation** – Urea disrupts both hydrogen bonding and electrostatic interactions. It does so because of its high polarity and tendency to form polydentate hydrogen bonding interactions with substrates. Additionally, thiols like glutathione or β-mercaptoethanol can oxidize disulfide bonds and thus further disrupt the protein folding.

2) **Temperature Dependent Cold-Denaturation** – This is an effect which concerns the hydrophobic effect. Recall that the hydrophobic effect is the primary driving force behind protein folding thermodynamically, and that it occurs with both $\Delta H$ and $\Delta S$ are positive. Thus, by the Gibbs Free Energy relation, we see that at sufficiently low temperatures, $\Delta H$ can outweigh the $T \Delta S$ term, making the overall $\Delta G$ positive. This disfavors folding, allowing for denaturation.

3) **Temperature Dependent Heat-Denaturation** – This is due to the overall increase in the thermal energy of the environment. Recall that both hydrogen bonds and electrostatic interactions are relatively weak interactions on the order of only a few kilocalories per mol. These weak interactions are subject to regulation by the thermal energy of the environment. If the thermal energy of the environment ($RT$) reaches a few kilocalories per mol, then these bonds become extremely labile, and the protein has a tendency to denature at its surface, leading to loss of enzymatic activity.

**Further Suggested Reading:** Bugg (Chapter 2) – Focus especially on pages 13 – 24.
Edman Degradation – An Analysis of the Reaction Mechanism

1) Step 1: Attack of the Isothiocyanate

Thiocyanates (cyanide analogs of thioesters):

\[
\begin{array}{c}
\text{N} \\
\text{C} \\
\text{S} \\
\text{R}
\end{array}
\]

Isothiocyanates:

\[
\begin{array}{c}
\text{R} \\
\text{N} \\
\text{C} \\
\text{S} \\
\text{R}
\end{array}
\]

The most nucleophilic site on the peptide chain is the DEPROTONATED amino-terminus. It must be deprotonated to attack! This means that the reaction must be done at pH somewhat above the pKa of the amino terminus (pKa = 6.8 – 8.0). The isothiocyanate is highly electrophilic. This is because of the poor \( \pi \) bonding between carbon and sulfur which results from 1) poor molecular orbital overlap between the C2p and the S3p orbitals due to size and energy constraints and 2) the bond length, which is too long to promote strong \( \pi \) bonding. Thus, the middle resonance structure above, which is formally uncharged, is actually not a dominant resonance structure. The other resonance structures, especially the first, which gives the central carbon cationic character, is significant and results in the electrophilicity of the species.

2) Step 2: Formation of the Thiourea – The nucleophilic addition in the previous step has left behind what is essentially a deprotonated thiourea. The negative charge on the sulfur is resolved by protonation of the nitrogen, creating a species which has significant zwitterionic character. NOTE THAT IT IS THE NITROGEN LONE PAIR AND NOT THE \( \pi \) BOND THAT PICKS UP A PROTON! THE \( \pi \) BOND IS NOT REACTIVE BECAUSE OF THE RESONANCE!

3) Notice that the first resonance structure shown above for the thiourea is dominant for the electronic reasons presented earlier (the poor \( \pi \) bonding in C=S). Therefore, the sulfur has a slight negative charge. Moreover, this charge is not significantly stabilized by the proticity of the aqueous environment because S-H bonds are extremely weak, again for reasons of poor molecular orbital overlap between the S-3sp\(^3\) and the H-1s orbitals. Thus, this is a naked charge that is highly nucleophilic. Indeed we see sulfur is an excellent nucleophile in many reaction conditions. Here, when we lower the pH, the amide carbonyl is activated by protonation. Carbonyls are up to \( 10^6 \) (one-million) times more reactive upon protonation. Remember that the Sulfur is NOT DEACTIVATED by the acidic conditions because it does not form H-bonds! Hence we have increased the electrophilicity of the amide significantly.
without affecting the nucleophilicity of the sulfur. Under these conditions, we observe the nucleophilic sulfur attacking the activated amide carbonyl, forming a tetrahedral intermediate.

4) Upon formation of the tetrahedral intermediate, the amide nitrogen loses conjugation. Its lone pair, previously involved in the amide conjugation, now becomes highly basic, and under the acidic reaction conditions, it gets protonated. This results in an activated peptide leaving group, and collapse of the tetrahedral intermediate expels the peptide chain leaving a phenylthiohydantoin intermediate behind.

5) Once again, we see an isothiourea moiety within the thiazolinone (thia = sulfur, azole = nitrogen, one = ketone). This intermediate has a thioester moiety, which is highly susceptible to hydrolysis because of the lack of typical ester resonance:

\[
\begin{align*}
\text{R' & S_2} \\
\text{O & R} \\
\end{align*}
\]

The middle resonance structure is not very significant, lending greater importance to the third than is usually observed in oxyesters and amides. Hence we see the greater electrophilicity of thioesters compared to oxyesters and amides. In fact, in terms of electrophilicity, thioesters are about as electrophilic as ketones and aldehydes. The carbonyl is once again further activated by protonation under the acidic conditions, and water in the aqueous medium attacks the carbonyl center. Remember, activation is nothing but a shift in the relative contributions of the resonance structures shown above. Protonation causes the molecule to adopt greater cationic character at the carbonyl carbon (resonance structure 3 above). The sulfur is expelled during collapse of the tetrahedral intermediate. At this point, the amine from the resulting thiourea attacks the ester, causing cyclization. The driving force here is not only the entropic driving force behind all cyclizations but also the formation of the more stable amide. We are thus left with our stable phenylthiohydantoin product for analysis. A rudimentary reaction coordinate diagram is provided to the left, with 1 representing the thiazolinone, 2 representing the acyclic ester, and 3 representing the final phenylthiohydantoin product. The tetrahedral intermediates are represented as intervening minima between high energy transition states.

6) Monodegradation is made possible by the sequential nature of the reaction conditions. Only the terminal amino acid is released during each cycle. This is the beauty of the dual-condition system.
Protein Sequencing by Mass Spectrometry

The Theoretical Basis
Mass Spectrometers separate particles based on their charge-to-mass (z:m) ratio. Particles must be charged in order to be analyzed by the mass spectrometers. Conveniently, peptides readily develop a charge (or multiple charges) by protonation at Nitrogen or Oxygen. The process can be neatly divided into three phases: 1) Ionization, or the creation of charged vaporized peptide molecules, 2) Mass Analysis, or the separation of the particles by their z:m ratio, and 3) Detection of the particles after separation.

Ionization
1) Matrix-Assisted Laser Desorption Ionization (MALDI) – The protein is solubilized in a volatile solvent and mixed with a solution of acidic matrix containing synapinic acid or α±-cyano-4-hydroxycinnamic acid (simple proton sources for ionization). This solution is coated in a thin film onto a metallic surface and allowed to crystallize. The surface is shot with a high energy laser beam which vaporizes the matrix and protein, creating charged peptide ions which are separated by the analyzer.

2) Electrospray (ESI) – The protein is solubilized in a volatile solvent and pumped into a steel capillary tube. Very high voltage is applied to the tip of the capillary, and this causes the sample to ionize and disperse as a fine mist. The mist is dried in a chamber of warm nitrogen, and the dried particles are passed through a small aperture into the analyzer chamber.
Particle Separation

1) Time-of-Flight (TOF) – Particles pass through an electric field (E) that accelerates the particles based on their charge z. The acceleration is dependent on the charge of the particle and its mass: F = ma = Ez gives us a = E (z/m). Since \( \Delta v = at \), the velocity is directly dependent upon the acceleration, and thus upon the z/m ratio. The greater the charge, the greater the acceleration, and thus velocity; the smaller the charge, the less the acceleration, and thus velocity. After this initial acceleration, particles are allowed to enter a drift chamber, where they migrate at their constant velocity, further separating based on their differential velocities. Since velocity is difficult to detect, we will use one more of Newton’s Equations: \( \Delta x = \frac{1}{2}at^2 + vt \), and detect time instead. Because there is no field in the drift chamber, there is no acceleration, so the equation simplifies to \( \Delta x = vt \). If we fix the distance x, then we get \( t = \frac{x}{v} \). Thus, we can measure the time it takes a particle to come out of the analyzer of a fixed distance and thereby determine its z/m ratio. The slower a particle, the longer it takes (makes sense!) and the smaller its z/m ratio!

Here is the derivation using Newton’s Second Law of Motion:

\[
F = ma = zE \Rightarrow a = E \left( \frac{z}{m} \right)
\]

\[
\Delta v_a = aT_a \Rightarrow \Delta v_a = ET_a \left( \frac{z}{m} \right)
\]

\[
\Delta x_a = \frac{1}{2}aT_a^2 \Rightarrow T_a = \sqrt{\frac{2\Delta x_a}{E(z/m)}}
\]

We must express \( T_a \) in terms of constant parameters

\[
\Delta v_a = ET_a \left( \frac{z}{m} \right) \Rightarrow \Delta v_a = \sqrt{2E\Delta x_a \frac{z}{m}}
\]

This gives us an expression for the velocity

\[
\Delta x_d = v_a T_d \Rightarrow T_d = \Delta x_d / v_a \quad v_a = \Delta v_a \text{ by setting } v_o = 0 \text{ for simplicity.}
\]

\[
T_d = \frac{\Delta x_d}{v_a} = \frac{\Delta x_d}{\sqrt{2E\Delta x_a \frac{z}{m}}} \Rightarrow (m/z) = \frac{2E\Delta x_a T_d^2}{\Delta x_d^2}
\]

\[
(m/z) = \frac{2E\Delta x_a}{x_d^2} T_d^2
\]

From this equation, we see that we set constant the electric field (E), the accelerator distance (x_a), and the drift chamber distance (x_d). This gives us a direct relationship between the drift time and the (m/z) ratio.

This could just as well be derived from Conservation of Energy:

\[
\Delta KE = \Delta U \quad \text{Conservation of Energy}
\]

\[
\Delta U = \int \vec{F} \cdot d\vec{x}_a = F_x \cos \theta = F \vec{x}_a \quad \text{The electric field is parallel to the direction of motion}
\]

\[
KE = \frac{1}{2}mv^2 = F \vec{x}_a = zE \vec{x}_a \quad \text{Using the Electric Force}
\]

\[
\Delta v_a^2 = 2Ex_a (z/m) \Rightarrow \Delta v_a = \sqrt{2E\Delta x_a \frac{z}{m}}
\]

\[
\Delta x_d = v_a T_d \Rightarrow T_d = \Delta x_d / v_a \quad v_a = \Delta v_a \text{ by setting } v_o = 0 \text{ for simplicity.}
\]

\[
T_d = \frac{\Delta x_d}{v_a} = \frac{\Delta x_d}{\sqrt{2E\Delta x_a \frac{z}{m}}} \Rightarrow (m/z) = \frac{2E\Delta x_a T_d^2}{\Delta x_d^2}
\]
2) Quadrupole (QMS) – The device consists of four parallel rods, two of which are positively charged and two of which are negatively charged, both at a set DC current. Further, there is an applied AC current which causes a fluctuation in the voltage and thus a resonance of the particles. Only particles with particular resonant frequencies (which is related to the m/z ratio) will pass through the quadrupole to the detector. All others will hit the rods and decompose.

The math is more complex for this one, so I won’t bother showing it here. I will derive it for anyone in person who so desires. Suffice it to say for the purposes of the course that a particle’s m/z ratio is known because only particles of a singular m/z ratio are allowed to pass through the analyzer to the detector, which measures the concentration by counting the number of particles in a given time period for each m/z setting.

Detection
Particles are detected by a counter which counts their relative abundance. Plots are constructed with m/z on the x-axis and the relative abundance on the y-axis. This will allow you to sequence a peptide very accurately. One conundrum: Leucine and Isoleucine have the same mass, and thus, they will have the same m/z ratio!!! How do you differentiate between them?

Peptide Fragmentation
In order to sequence a peptide, it must be fragmented into sequential fragments. This will allow you to sequence a peptide accurately. One conundrum: Leucine and Isoleucine have the same mass, and thus, they will have the same m/z ratio!!! How do you differentiate between them?

Fragmentation is accomplished in a collision chamber, which fragments the peptide into B-ions and Y-ions. **You must know how to fragment these peptides and reconstruct a peptide from its ions!** Then, these ions are subjected to one of the analyzer methods above, and their m/z ratios are recorded as usual! **REMEMBER THAT THE FIRST N-TERMINAL AMINO ACID IS NEVER FRAGMENTED – THERE IS NO B1!**
Don’t forget that the mass of b-ions is the mass of the contained amino acid residues + the mass of a proton (m = 1 amu) and that the mass of y-ions is the mass contained in the amino acid residues + the mass of water (m = 18 amu) + the mass of a proton (m = 1 amu).

**Interpreting a Spectrum**

In the problem sets (and perhaps on an exam) you may be given a spectrum from a mass spectrometer and asked to determine the sequence.

Look for the B-ions and read them straight across. The difference in the masses between any two B-ions is the mass of an amino acid residue. An amino acid residue is the mass of an amino acid – the mass of water (the water from the condensation of the carboxylic acid and amine during peptide bond formation).

**Additional Reading On-Line (Quick Overviews)**

http://www-methods.ch.cam.ac.uk/meth/ms/theory/tofms.html
http://www.chem.vt.edu/chem-ed/ms/ms-intro.html
http://www.astbury.leeds.ac.uk/Facil/MStut/mstutorial.htm
Polypeptide Synthesis – Amino Acid Coupling Reactions

**Amide Protection**

Amide Protection (BOC)

Amide Deprotection – Acidic Conditions

Deprotection Mechanism

**Carboxyl Activation**

**Sidechain Protection**

The various amino acid side chains are variably reactive, and this is what gives them their versatility in enzymatic catalysis and structural fidelity. However, during synthesis, these side chains may participate in numerous side reactions. For this reason, the side chains must be protected before polypeptide synthesis is pursued. For the list of amino acid protecting groups, see the Chem 27 Sourcebook Page 27.

**Laboratory Polypeptide Synthesis – ALWAYS C → N**

Cellular Synthesis is always N → C. However, Laboratory Synthesis is always C → N to prevent epimerization (racemization of individual residues) of the forming polypeptide chain! See the diagram to the left – how can the C-terminal amino acid epimerize – verify the mechanism.

The C-terminal residue of the desired polypeptide is amino-protected and sidechain protected and then coupled to a solid-phase bead by S$_2$N$_2$ attack of a primary chloride. Subsequently, the amine is deprotected with acid treatment. The amine is then neutralized with base (triethyl amine) to make it nucleophilic. Separately, the next amino-protected residue is activated by DCC. The activated amino acid is then passed through the solid phase and attacked by the amine.
of the attached amino acid. Remaining free amino acids are washed away. The resulting dipeptide is then deprotected by acid treatment. The amine is then neutralized with base (triethyl amine) to make it nucleophilic. Separately, the next amino-protected residue is activated by DCC. The activated amino acid is then passed through the solid phase and attacked by the dipeptide amino terminus. This process is repeated several times until the desired polypeptide is produced. Treatment with hydrofluoric acid and thiophenol removes all sidechain protecting groups and releases the polypeptide from the solid phase beads.

**Kent’s Convergent Synthesis Chemistry**

Kent developed a method to synthesize large peptides efficiently by synthesizing parts and then coupling these parts. This requires slight modifications, such as the use of a thioester instead of an amide to allow for coupling, as well as using an $\alpha$-Bromo ketone instead of an acid for easy coupling. Be sure to read this carefully!

**Native Peptide Ligation**

Using a thioester once again, we can allow for the attack of a cysteine thiol followed by the formation of the more stable amide bond (amides are more thermodynamically stable than thioesters). **This ligation must occur on the N-terminal side of a cysteine residue!**

**Cellular Synthesis**

Cellular synthesis is always N $\rightarrow$ C. Amino acids are coupled to a tRNA molecule, which delivers the amino acid to the ribosome. The ribosome, a complex protein and ribonucleic acid macroenzyme catalyzes the peptide bond forming reaction in a sequential fashion. The amino acid forms an ester linkage with the 3’-hydroxyl on the 3-adenine on the tRNA. This is the activated bond which is attacked by the amine of the next amino acid in the sequence. This is a highly efficient process which progresses at a rate of 1000s of amino acids per second!
Proteases and Proteolysis – Cleavage of the Amide Bond

**Serine Proteases**
Serine Proteases possess a **Catalytic Triad**: Serine, Histidine, and Aspartate, which coordinate the efficient cleavage of a peptide bond.

1) Serine is the nucleophile which attacks the amide carbonyl, forming a covalent ester linkage with the peptide after expulsion of the peptidic amine. This ester is subsequently cleaved by the nucleophilic attack of a water molecule to result in the hydrolysis.

2) Histidine acts to deprotonate the serine in the first nucleophilic attack and water in the second nucleophilic attack.

3) Aspartate hydrogen bonds to Histidine, increasing its basicity, and hence its catalytic activity.

4) Additionally, serine proteases possess an **oxy-anion hole** which activates the carbonyl and stabilized the negative charge after nucleophilic attack.

Trypsin cleaves on the C-terminal side of Lys and Arg (positively charged long-chain residues). Chymotrypsin cleaves on the C-terminal side of aromatic and large hydrophobic residues.

**Aspartyl Proteases**
Aspartyl proteases possess two aspartate residues which function in coordination to directly facilitate the nucleophilic attack of water on the amide carbonyl and subsequent hydrolysis through deprotonations and hydrogen bonding interactions. Unlike serine proteases, there is no acyl-enzyme intermediate. The enzyme simply catalyzes the direct attack of water on the peptide bond by protonating the carbonyl and deprotonating the water. Collapse of the tetrahedral intermediate results directly in the two peptide fragments and regenerates the enzyme active site in its original state.

**N-Terminal Nucleophile Hydrolases (Very Similar to Serine Proteases)**
N-Terminal Nucleophile Hydrolases have an **N-Terminal Serine** which is the catalytic residue. The N-terminus deprotonates a water molecule, which subsequently deprotonates the serine, allowing it to attack the peptide carbonyl, forming a tetrahedral intermediate. The amide nitrogen gains a proton from the protonated N-terminus. This tetrahedral intermediate then collapses, expelling the C-terminal peptide fragment and leaving behind an **acyl-enzyme intermediate**. Subsequently, the N-terminus deprotonates a water molecule which attacks the acyl-enzyme intermediate. The tetrahedral intermediate then collapses, expelling serine and the N-terminal...
peptide fragment. This mechanism is very similar to the serine protease mechanism, except that the histidine and aspartate are replaced by the N-terminus in the same capacity.

**Protease Inhibitors**

**Competitive Inhibitors** bind in the active site of the enzyme, preventing the binding of the natural substrate. Inhibitors must first possess proper targetability. They must bind to the active site of the enzyme by mimicking the enzyme’s natural target, especially the transition state of the enzyme catalyzed reaction. There are two types of competitive inhibitors: Non-covalent Transition State Analogs and Covalent Modifiers.

**Non-competitive inhibitors** do not bind in the active site bind elsewhere on the enzyme, usually causing allosteric changes in the active site. They bind usually in a separate site called the allosteric site, and their binding causes a change in the conformation of the active site, preventing effective catalysis of the enzymatic reaction.

Most of the focus has been on **competitive inhibitors** which bind in the active site. These competitive inhibitors must resemble the natural substrate sufficiently to bind the active site. For example, cyclotheonamide contains the arginine residue which binds to the active site of trypsin (remember that trypsin cleaves C-terminal to arginine residues). In addition to these recognition moieties, they must possess some means of inhibition, some of which are discussed below.

**Serine Protease Inhibitors**

1) **Suicide Substrates**

   ![Suicide Substrate](image)

   Aldehydes are very electrophilic and form hydrates, acetics, and hemiacetals (like this serine hemiacetal) readily. This is a stable species, and thus it is a suicide substrate which forms a stable enzyme-inhibitor linkage. **This is a reversible inhibition.**

2) **Covalent Modifiers**

   ![Covalent Modifier](image)

   The electrophilic ketone is attacked by the serine, and the expelled serine is properly positioned for direct $S_N^2$ attack on the chloride. **This is an irreversible inhibition.**

3) **Transition State Analogs**

   ![Transition State Analog](image)

   Boron esters readily accept tetrahedral geometry. The serine attacks the empty boron p-orbital, forming a tetrahedral intermediate which resembles the transition state. B-O bonds are longer than C-O bonds and therefore resemble the transition state bonding lengths more closely. **This is a reversible inhibition.**

**Aspartyl Protease Inhibitors**

Because there are no acyl-enzyme intermediates, transition state analogs have been the focus of research in protease inhibitors. The transition state is an orthoamide (a carbon with two bonds to oxygens and one bond to a nitrogen). This is highly unstable and will collapse immediately. In order to mimic this intermediate, a secondary alcohol has proven to be reasonably effective. Phosphates have also been effective, because the P-O bond is longer than C-O bonds and
therefore resemble the transition state bonding lengths more closely. Additionally, it has two oxygens, thus more closely resembling the transition state than the secondary alcohol.

Remember that HIV Protease is symmetric because it a homodimer (a dimer consisting of 2 identical subunits). Because of this, symmetric or nearly symmetric inhibitors were designed that would bind the symmetric active site. These are C2 symmetric.

**Autoproteolysis – Peptide Splicing**

Autoproteolysis is the process by which a peptide catalyzes its own cleavage. This is also known as peptide splicing or protein splicing. There are two extein fragments (that are joined together after splicing) which are separated covalently by an intein fragment (which is expelled after splicing). Inteins are always flanked by a nucleophilic alcohol (Serine or Threonine) or thiol (Cysteine) on the N-terminal side which act to displace the peptide bond in the first step. The C-terminal extein also is flanked on its N-terminal side by a nucleophilic alcohol (Serine or Threonine) or thiol (Cysteine), and this residue then replaces the nucleophilic residue from the intein. This is shown in the first diagram above.

Inteins are also always flanked on the C-terminal side by an asparagines residue, which nucleophilically attacks the peptide bond between the intein and the C-terminal extein. This cleaves the bond, releasing the intein and leaving behind a free amine on the extein. This N-terminus attacks the serine ester of the N-terminal extein, forming the final peptide bond, as shown in the diagram to the right.

**Additional Reading**
Bugg Chapter 5 (81 – 89, 95 – 98)