Biochemistry study guide/outline #1, Fall 2025

Adapted from one I give my students to help them focus on core concepts so they can apply them

Functional groups

- You should be able to identify these functional groups for any structure you see:
  - Hydroxyl
  - Amine
  - Amide
  - Thiol/sulfhydryl
  - Ester
  - Ether
  - Carbonyl
  - Carboxylate and carboxylic acid (and tell them apart)

Intermolecular forces (IMFs) & the hydrophobic effect

- IMFs can be over-simplified to be thought of in terms of electrostatics (charge-charge based attractions)
  - Their strength thus depends on:
    - Strength of the charges (more (opposite) charge, more attraction)
    - Distance between them (shorter distance, more attraction)
    - Polarity of the intervening medium/solvent (lower polarity of solvent (as reflected in lower dielectric/relative permittivity, ε), more attraction thanks to less distractions from the solvent)
      - This is why salt bridges are stronger in the nonpolar interior of a protein than on the surface, where there are more charge distractions

- The **main forms of IMFs** are (in order of decreasing strength): ionic bonds (salt bridges) > H-bonds > other permanent dipole - permanent dipole > other transient dipole-based attractions (London dispersion forces & other VDW)

- You should be able to predict the types of IMFs various molecules can form with one another (including the solvent)

  - If a molecule is non-polar, it can only participate in VDW. By definition, it does not have a permanent dipole and can therefore only be temporarily polarized due to random chance (instantaneous dipole) or being near a charge or partial charge (induced dipole)

  - You need both a full + and a full – charge (not just a partial charge, but an actual difference in the number of protons and electrons) to form a salt bridge/ionic bond

  - You need both a donor and an acceptor to form an H-bond

- Water molecules are highly polar, so they only want to give up water-water interactions if they're given another polar and/or ionic interaction to swap for
  - More polar and/or ionic molecules tend to be more water-soluble
  - Signs of polarity are nitrogen and oxygen, which are electronegative, thus they pull shared electrons toward themselves, creating a dipole
  - Molecules that are mostly just hydrocarbons are typically hydrophobic (water-avoided/excluded) & thus not very water-soluble (but can be fat-soluble)

    - Molecules with lots of oxygens and/or nitrogens are typically hydrophilic (water-loved) & thus very water-soluble

- Binding affinity ultimately comes down to thermodynamics - you want a negative ΔG, which can come from:
        - A decrease in enthalpy (forming more favorable bonds and/or interactions)
        - And/or an increase in entropy (giving molecules more freedom to move about randomly)
    - When considering these factors you need to consider interactions with and between the solvent (typically water) molecules as well

- Hydrophobic molecules are ***not*** intrinsically attracted to other hydrophobic molecules - instead, the energy gain from the **hydrophobic effec**t comes from the increase in water **entropy** as water molecules are freed from solvation shells when hydrophobic molecules are pushed together as water maximizes its water-water bonds, minimizing the surface area that must be covered
    - And then dispersion forces and other transient attractions can occur between the hydrophobic molecules once forced together as water tries to maximize its water-water bonds)

- Hydrogen bonds (H-bonds) are a special case of permanent dipole-permanent dipole interactions
    - They require:
        - A **Donor (D)**: A H on something electronegative (O, N, or sometimes S)
        - An **Acceptor (A)**: An electronegative atom with a lone pair of electrons (O, N, or sometimes S)
    - They're extra-strong (compared to other dipole-dipole but still individually weak) because of their short distance (since H is so small), their partial covalent character, and their typically large dipole
    - To be maximally strong, they need a particular orientation/alignment (which, in addition to steric hindrance, helps explain why protein secondary structures take distinct forms)
    - You should be able to identify potential donors and acceptors and draw dotted lines for potential H-bonds

Acids, bases, pH, pKa, pI, & buffers

- **pH is at the solution level**

    - The **pH** is a measure of free proton availability, which comes from the contributions of all acids/bases in a solution.

- The more protons there are (more acidic), the lower the pH -> the more likely things are to be protonated, but how likely something is to be protonated will depend on how much it wants a proton (reflected in the pKa)

- **pKa is at the acidic group level**

    - The **pK_a** tells you about the strength of a single acid site (any acid, not only acids in an amino acid side chain)
        - The pK_a is the pH at which half the copies of an acidic site is deprotonated (and thus in conjugate base state) and half is protonated (and this in conjugate acid form)
            - At a pH < pK_a, there's more protons available to take, so more conjugate acid
            - At a pH > pK_a, there are fewer protons available, so more conjugate base

- **pI is at the multiprotic molecule level**
    - The **pI (isoelectric point)** is the pH at which a molecule is net neutral (has no net charge)
        - At a pH < pI, a molecule is positively-charged
        - At a pH > pI, a molecule is negatively-charged
        - The pI comes from the combination of acid/base groups in a molecule

- Putting things together:
    - The protonation state of any acidic group depends on the pH.
        - The lower the pH, the higher the proton concentration, thus the more likely something is to be protonated (in conjugate acid state).
        - The higher the pH, the lower the proton concentration, thus the more likely something is to be deprotonated (in conjugate base state).
    - A conjugate acid can be + or neutral, but is always protonated – look to and/or draw the structure to see which it is
    - A conjugate base can be – or neutral, but is always protonated – look to and/or draw the structure to see which it is
    - Whether an acidic group is protonated at a given pH depends on the strength of the acid, as reflected in the pKa:
        - The pKa is the pH where a single acid site is half protonated, half deprotonated.
        - A lower pKa corresponds to a stronger acid/weaker base
        - A higher pKa corresponds to a weaker acid/stronger base
    - The pI is the pH at which a multi-protic molecule is electrically neutral.
        - The charge on such a molecule depends on the protonation statuses of its acidic sites, each of which has its own pKa.

- A **buffer** is a solution of a weak acid/base containing relatively equal concentrations of conjugate acid and conjugate base -> this allows it to counteract added acid or base to maintain a steady pH
    - Since the pKa is the 50:50 mark, a buffer works best close to its pKa (typically +/- 1 pH unit)

Amino acids:

- Amino acids are almost always in the zwitterionic form where the N-terminal amino group (pKa ~9-10) is protonated & positive, & the C-terminal carboxyl group (pKa ~2) is deprotonated, so the backbone is net neutral

- The pK_R is the pK_a for an amino acid side chain (aka R-group). Not all amino acids have them.

- I do not need you to memorize pK_R's, but do want you to know the approximate pK_as of the termini, as well as the predominant form of each amino acids at physiological pH (~7.4)

- The **"basic" amino acids** are Arginine (Arg, R), Lysine (Lys, K), and Histidine (His, H)
    - We call them basic because in their neutral form they are conjugate bases
        - And remember this refers to neutral as in uncharged sidechain, *not* neutral as in neutral pH
    - In their protonated (conjugate acid) form, they are positively-charged
        - <u>Lys & Arg</u> have high pK_R's (~10.5 and ~12.5, respectively) and are thus <u>predominantly (and really almost always) positively-charged at physiological pH</u>
        - <u>His</u> has a pK_R (~6.0) closer to physiological pH, so it's <u>predominantly neutral</u> at physiological pH, but can more easily go back-and-forth
    - They have amine groups
    - They are often colored blue (think of the N, or think of basic-blue)

- The **"acidic" amino acids** are Aspartate (Asp, D) & Glutamate (Glu, E)
    - We call them acidic because in their neutral form they act as acids
    - In their deprotonated (conjugate base) form, they are negatively-charged
    - They have carboxylic acid groups
    - They have very low pK_R's (~4) and are thus predominantly (and basically almost always) negatively-charged at physiological pH
    - They are often colored red (think of the O)

- The polar amino acids have hydroxyl, amine, amide, or thiol groups

- You should know just by seeing the 1-letter abbreviation, its 3-letter abbreviation & full name; how to classify it; and what (if any) functional groups it contains

- When amino acids link together through peptide bonds, they do so by combining their amino & carboxylic acid groups), so we now call them ***residues*** (because they're no longer amino acids)

- When thinking about what interactions 2 amino acid side chains can form:

    - If both are nonpolar, then only VDW

    - If one is nonpolar, the other polar, then only VDW

    - If both are polar, but uncharged, then VDW & permanent dipole-permanent dipole, potentially H-bond

    - If both are polar, one charged, then VDW & permanent dipole-permanent dipole

    - If both are charged with the opposite charge, VDW & ionic

Protein structure:

- There are 4 main levels of protein structure:
  - o **<u>Primary (1°)</u>**: the sequence of amino acids (not an actual "structure")
  - o **<u>Secondary (2°):</u>** Involves H-bonding between peptide backbone (the amine H & carbonyl O)
    - ▪ Main forms are: α-helix & β-strands -> connected (often with β-turns in between) to form β-sheets
    - ▪ Gives protein local structure
  - o **<u>Tertiary (3°)</u>**: Involves side-chain (R-group) interactions (with other side chains and/or backbone) -> gives protein chains their overall shape
  - o **<u>Quaternary (4°)</u>**: Involves interactions between chains/subunits
    - ▪ Not all proteins have
      - Ones that do are called multimers
        - o Homomers if all chains the same
        - o Heteromers if some are different
- The hydrophobic effect plays a large role in protein folding
  - o Non-membrane proteins fold so that polar regions are on the surface and nonpolar regions are in the center, hanging out with nonpolar regions, so water doesn't have to get stuck in a shell around them

- **Primary (1°)**:
  - o This is "just" the sequence of amino acids (not an actual "structure")
  - o But it <u>influences the entire protein structure!</u> Because the R group…
    - ▪ Restricts what bond angles can be taken and what 2° structures are preferred
    - ▪ Influences where that part of the structure wants to hang out (e.g. near aqueous solvent or hydrophobic interior)
    - ▪ Influences what type of tertiary & quaternary interactions are possible

- **2° structure**
  - o Involves **<u>H-bonding</u>** between peptide amine H's & carbonyl O's, which are evenly spaced out along the **<u>backbone</u>**
  - o Comes mainly because of the restricted nature of the peptide backbone
    - ▪ Rotation is only allowed at specific places (phi & psi angles - on either side of the Cα) because <u>the peptide bond is resonance stabilized & thus needs to be planar</u>
      - Phi & psi angles are then restricted by steric hindrance
        - o The angles are further fine-tuned to optimize H-bonding interactions
          - ▪ Leads to characteristic 2° structures like α-helices & β-strands/sheets
  - o Different amino acids prefer to be in different secondary structures, largely due to their size & the relative amount of space available in the different structures
    - ▪ In an **α-helix**, side chains all face out and have to compete for molecular elbow room -> **long, skinny** amino acids (or alanine since it's small) are favored
    - ▪ In a **β-strand**, side chains alternate up & down, so side chains have more space -> **big, bulky amino acids** are okay here
      - Every other amino acid in a side chain is on the same face, so you could have something like polar-nonpolar-polar-nonpolar-polar-nonpolar which would give you a hydrophobic & a hydrophilic face (but you don't always)

        - Glycine & proline are unlikely to be found in α-helices or β-strands
            - Glycine: too flexible
            - Proline: too inflexible, too awkward of angles, no H-bonding with backbone
    - Ramachandran plots show phi-psi angles, with characteristic angles corresponding to α-helices, β-strands, and other secondary structures. You do not need to interpret Ramachandran plots

- **3° structure**
    - Involves side-chain (R-group) interactions (with other side chains and/or backbone)
    - Gives protein chains their overall shape
    - The bonds & forces holding them together are (from strongest to weakest): disulfide bonds (cysteine crosslinks) > ionic bonds (salt bridges) > H-bonds > other permanent dipole - permanent dipole > other transient dipole-based attractions (London dispersion forces & other VDW)
        - Disulfide bonds are special because they're the only of these forces that are covalent
            - But they're still reversible -> can be broken by reducing agents
                - The intracellular environment is relatively reducing, so they're mostly found in secreted proteins

- **4° structure**
    - Involves interactions between chains/subunits
        - Not all proteins have
            - Ones that do are called multimers
                - Homomers if all chains the same
                - Heteromers if some are different
    - Same bonds & forces as 3° structure

- "**Domain**" is a really vague term we use in all sorts of ways to refer to a "part" of a protein - often it's a largish part that is functionally and/or structurally distinct
    - Domains are like rooms in a house & may contain motifs
    - Domains are often evolutionary conserved
    - Domains can often be studied independently

- A **motif** is a small arrangement of secondary structures that is seen over and over in different proteins & sometimes has some known function
    - This is like furniture in your room