# Week 3

# Proteins

## 3.1 Amino Acids, Peptides, and Protein Synthesis

10/11:

- Initial impressions of the homework: More difficult than expected.
  - Tang did not raise the difficulty of the course, but was told to in course evals two years ago.
  - Literature problems: People believe reading the papers did help with the questions. We should be able to do these problems without the papers, though — some of these problems are past exam problems and were expected to be answered in a closed-book setting. We can expect similar questions on exams this year. Purpose: Show us how concepts from the class are used in research.
- There will be a practice exam posted.
- We can bring a one-page (single-sided A4) review sheet to the exam.
- OH Monday via Zoom.
- For every Thursday midterm, the content from the preceding Tuesday will not be covered.
- **Lesion**: Something bad that your cell will recognize and repair.
- Types of lesions:
  - Double-stranded breaks, mismatches, pyrimidine dimers, and damaged bases.
  - Are mutations lesions? It depends.
    - If the mutation is a mismatch, there will be a repair.
    - If it shows up as matched, your cell will not know to repair it.
  - DNA modifications.
    - Damaged backbones (e.g., pyrimidine dimers or a methyl group on the $O^6$ of guanine). Things your cells know shouldn't be there. Will be repaired.
    - However, there can also be intentionally placed modifications on DNA to regulate it. These will not be repaired.
  - Bulges don't usually occur during synthesis, but they can occur during recombination. These will definitely be repaired.
  - This should clarify some points on the homework.
- Natural base modification in mRNA.
  - mRNA is less diversely modified than tRNA and rRNA, but mRNA modifications do still happen.
  - Most abundant internal modifications in mammalian mRNA: $N^6$-methyladenosine ($m^6A$).

        - There's on average 1-3 of these per mRNA. However, there can still be 0.
        - How it's detected: Highly related to ChIP-Seq. You fragment your DNA, introduce antibodies that will bind to $m^6A$, do immunoprecipitation for a specific DNA binding protein to enrich the target DNA sequence, and sequence both the input and the enriched pool. The sequences that got enriched are the ones that carry the modification.
    - 5-methylcytosine ($m^5C$) and pseudouridine ($\psi$) are also present in mRNA, but their functions are less well studied.
        - Pseudouridine is a flipped uracil base with a carbon connected to ribose instead of a nitrogen connected to ribose.
        - The W-C interaction surface is basically unchanged, though, so it will still be detected as U. However, it has alternate regulatory functions, such as helping ribosomes read through premature stop codons.
        - The $\psi$ detection method is messy (noisy): Introduce a chemical that selectively reacts with pseudouridine and gives a stop-signal during transcription. Not testable.
        - 5mC detection for RNA is identical to for DNA (bisulfite chemistry — see the discussion associated with Figure 2.7). Note, however, that since RNA is less stable, more will decompose upon heating; thus, you need a larger initial sample size.
    - In addition to $m^6A$, $m^5C$, and $\psi$, other base modifications can occur (we are not responsible for these, though).

- Summary of what we've learned so far:
    - DNA synthesis and transcription (the DNA $\rightarrow$ RNA part of the central dogma).
    - DNA methylation and epigenetics.
    - mRNA methylation and epitranscriptomics.
    - These three things function as a network (many feedback mechanisms). Moving forward, we will add proteins and metabolites to this network.

- Not testable: Arms race between bacteria and bacteriophages.
    - Answers how weird DNA modifications develop.
    - Bacteriophages are the most abundant life organism on this earth.
    - Round 1: Bacteria evolve restriction enzymes and base modification X; purpose: cleave phage DNA while avoiding suicide.
    - Round 2: Phages evolve X or Y modification in DNA; purpose: escape cleavage.
    - Round 3: Bacteria evolve X/Y-dependent restriction enzymes and additional self base modification Z; purpose: cleave phage DNA while avoiding suicide.
    - And on and on.

- Diverse base modifications in bacteriophages.
    - Guanine converts $N^7$ to a carbon and adds a functional group; you need multiple modifications to get to this result (called deoxyarchaeosine).
    - Cytosine attaches to glucose instead of deoxyribose.
    - Some bacteriophage DNA/RNA base modifications overlap with those in higher organisms, who evolve these modifications for completely different reasons.
    - And more.

- We are now done with last lecture's content; we are moving onto amino acids, peptides, and proteins.
    - Note that many of the mechanisms of RNA are more complicated than those of proteins, so if you have trouble with the latter, review the former.

  - Primarily amino acids this lecture; peptides, proteins, and higher-order structures next lecture.
  - Hopefully, these first six lectures will be foundational for the week 5-7 lectures on organelles and cell biology.

- A chemical look at proteins.
  - Made of proteinogenic amino acids (natural L-amino acids save glycine).
  - Can be post-translationally modified.
  - Post-translational rearrangement (lecture on this later).
  - We will look at amino acid properties.
  - Next lecture: Determinants of protein structure and...
  - Secondary and higher order structures.

- **Protein**: A polymer composed of amino acids.
  - Grows from the N-terminus to the C-terminus.

- Chirality is key.
  - Except for the achiral glycine, (almost) every amino acid is in its L-form.
    - Amino acids in their D-form are used as monomers, not for protein synthesis.
  - Steve Kent synthesized the D-form of HIV protease; he's a giant in the field. Taught here.
    - His big contribution is the development of **native chemical ligation**, while he was at Scripps.
    - We will talk about this more when we cover bioorthogonal chemistry.
  - No ribosomal D-protein synthesis because we would need an entire mirror image biological system.
    - People are trying to build a mirror ribosome, which Tang thinks is crazy, but they are making progress.
  - Total protein synthesis hasn't gone beyond 300 amino acids.
    - Solid state protein synthesis: Add one amino acid at a time. Highly efficient. 99.5% efficiency is great, but we have an exponential decrease of yield. Thus, we can't synthesize more than 50-100 amino acid peptides at a time.
    - Strategy: Fragments of 50 amino acids ligated together with natural chemical ligation.
    - But since proteins are folded as they're built in real life, we natural chemical ligation doesn't necessarily result in an accurately folded protein.

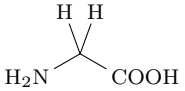
- **Native chemical ligation**: Connecting two peptides with an amide bond.
  - A very hard chemical problem; requires activating the amine of an amino acid.

- Taking advantage of D-proteins.
  - Why we want to do this: To challenge nature. Tang thinks this is stupid, though.
  - Favorable features of D peptides/proteins.
    - Similar to L proteins: bind to DNA/RNA/proteins, can catalyze reactions.
    - Cannot be degraded by natural protease (much more stable than natural peptides/proteins).
  - Challenges in identifying D peptides/proteins that bind specifically to a natural protein.
    - Rational design? — Hard to do with so many variables.
    - Screening? — Synthesize a D peptide library that can be amplified between selection rounds.
  - You can't synthesize a D library to look for hits on an L target (too hard; no mirror ribosomes). So instead, synthesize an L library, look for a hit on a D target, and then synthesize the D version of your L hit, which (flipping both chiralities) will react with your L target.

    - A brilliant idea but didn't turn out that well, though.
  - D-proteins aren't as big as they might be because there are many ways proteins can be degraded *in vivo* (not just natural proteases).

- Protein basics.
  - Classification of amino acids is somewhat arbitrary, but they are loosely categorized into hydrophobic, charged, polar, and glycine (in a class by itself since it's achiral).
    - For example, tryptophan could conceivably be hydrophobic or polar.
    - Histidine can frequently be charged.
    - Knowing properties is more important than knowing classes.
  - Knowing the amino acids is essential for predicting things like how amino acids interact with each other, what their role is in a reaction, how they catalyze a reaction, etc.
  - Memorize amino acids!
    - The 3-letter and 1-letter shorthand is often (but not always) the first 3 (resp. 1) letter(s) of the name.

- Achiral amino acid.

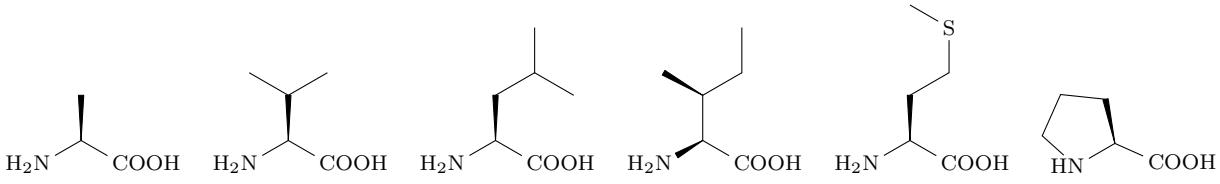



(a) Glycine.

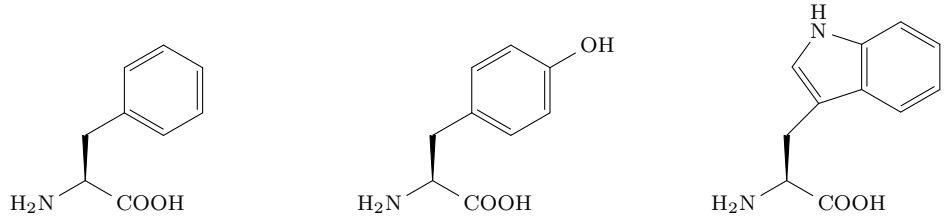
Figure 3.1: Achiral amino acid.

  - Glycine.
    - Flexible since it's unsubstituted on its $\alpha$-carbon; can sample multiple conformations.
    - Whenever you have a glycine in your protein, you can assume the protein is flexible in that region.
    - If you want to fuse two proteins together but you're worried about sterics, you typically use a GGS (glycine, glycine, serine) linker.
    - Name: Glycine, Gly, G.

- Hydrophobic amino acids.




(a) Alanine. (b) Valine. (c) Leucine. (d) Isoleucine. (e) Methionine. (f) Proline.




(g) Phenylalanine. (h) Tyrosine. (i) Tryptophan.

Figure 3.2: Hydrophobic amino acids.

- We start with the *aliphatic* hydrophobic amino acids.
- Alanine.
  - Simplest chiral amino acid. That's what makes it important. No other important features.
  - If you think an amino acid is important, mutate it to alanine. If the protein is nonfunctional, then you know that it was important. You use alanine over glycine because it's less flexible.
  - Name: Alanine, Ala, A.
- Valine.
  - The simplest branched amino acid.
  - Name: Valine, Val, V.
- Leucine.
  - Name: Leucine, Leu, L.
- Isoleucine.
  - The second chiral center is not required (it is S though).
  - Name: Isoleucine, Ile, I.
- Note on valine, leucine, and isoleucine:
  - All are considered bulky, aliphatic amino acids.
  - Example (possible test question): Suppose you have an enzyme that fits ATP perfectly. If you want the active site to kick ATP out, you can mutate some of the amino acids to these three to make the pocket smaller.
  - Takeaway: Used to change the size of pockets.
  - Phenylalanine is another possibility, but it comes with other features as an aromatic system.
- Methionine.
  - One of the two amino acids containing sulfur; the other one (cysteine) forms disulfide bridges.
  - Frequently seen as a start codon (ATG), though it can appear in the middle of proteins, too.
    - ➢ Their are only two proteins that are encoded by a single codon; the other is tryptophan.
  - When we see a methyl modification, that methyl group is coming from a methionine derivative (specifically **SAM**).
  - Name: Methionine, Met, M.
- Proline.
  - Proline has a strained structure.
  - Whenever you have proline, the chain naturally has less flexibility.
  - You can only have two conformations: *cis*- and *trans*-proline (with respect to the nitrogen). *trans* is more common.
  - Proline is not in $\alpha$-helices or $\beta$-pleated sheets because it typically induces a turn.
  - Name: Proline, Pro, P.
- We now move on to *aromatic* hydrophobic amino acids.
- Phenylalanine.
  - Name: Phenylalanine, Phe, F.
- Tyrosine.
  - Some people categorize tyrosine as polar.
  - The hydroxyl group is often phosphorylated; this derivative is called a **tyrosine kinase**.
  - Tyrosine kinases have been the most successful cancer drug target: You can somehow develop things that fit into the active site of one tyrosine kinase without affecting the rest of them.
  - Tyrosine kinases are less diverse then serine kinases and threonine kinases, aiding selectivity.
  - Name: Tyrosine, Tyr, Y.

  - Tryptophan.
    - Contains an indol moiety.
    - Only has one codon corresponding to it.
    - The heaviest amino acid.
    - The biosynthesis of tryptophan tends to be important, but we will not discuss it in this class. Proceeds through chromic acid.
    - Name: Tryptophan, Trp, W.

- (S)-adenosylmethionine (SAM) is a very important cofactor in our bodies.

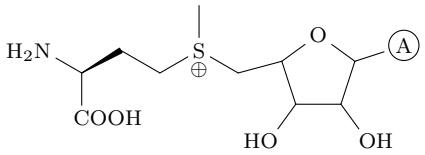
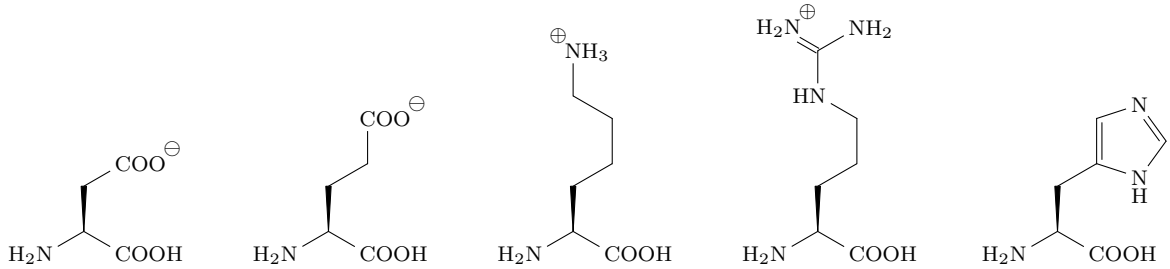



Figure 3.3: (S)-adenosylmethionine.

  - It donates the methyl group in DNA, RNA, and protein modification.
  - When the constituent moieties combine, S takes on a positive charge. This makes the lone methyl group on the sulfur a particularly good donor.

- Charged amino acids.



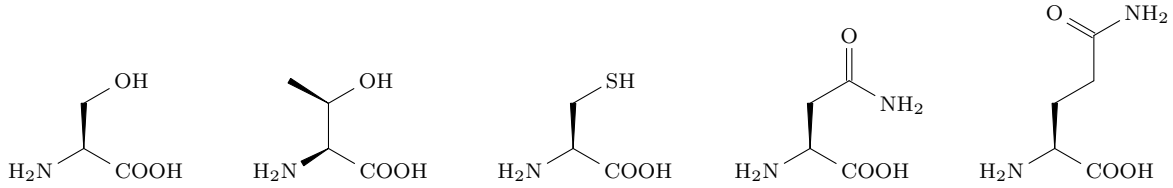

(a) Aspartate. (b) Glutamate. (c) Lysine. (d) Arginine. (e) Histidine.

Figure 3.4: Charged amino acids.

  - We start with the *negatively* charged ones.
  - Aspartate.
    - Under physiological pH, we draw the top "COOH" deprotonated; the other will be reacted.
    - Name: Aspartate (aspartic acid, if protonated), Asp, D. Alanine plus a carboxylic acid.
  - Glutamate.
    - Name: Glutamate (glutamic acid, if protonated), Glu, E.
  - We now move onto the *positively* charged ones.
  - Lysine.
    - An amine with $pK_a \approx 9 - 10$.
    - Name: Lysine, Lys, K.
  - Arginine.
    - Positively charged, but even more so under physiological pH. $pK_a \approx 12$.
    - Name: Arginine, Arg, R.

- Histadine is somewhat unique.
- Histidine.
  - Sometimes recognized as polar, but Tang prefers charged because it so frequently serves as the general base and acid in enzyme catalysis.
  - Contains an imidazole moiety.
  - The top nitrogen has $pK_a \approx 6$, so it can easily be protonated or deprotonated at physiological pH. Thus, it functions as a good **proton shuffle** to help catalyze acid/base reactions.
  - Some acid/base reactions can be catalyzed by lysine or aspartic acid.
    - For the nucleic acid polymerization reaction, the side chain is made of aspartic acid, which coordinates a metal ion to promote the reaction.
  - The other nitrogen does not easily lose its hydrogen.
  - Name: Histidine, His, H.

- **Proton shuffle**: A group that receives a proton from one group and donates it to another.
- Polar, uncharged amino acids.




(a) Serine. (b) Threonine. (c) Cysteine. (d) Asparagine. (e) Glutamine.

Figure 3.5: Polar amino acids.

- Serine.
  - Can be phosphorylated.
  - The hydroxyl group frequently serves as a nucleophile in the active site of enzymes.
    - Example (next time): **Serine protease**.
  - Name: Serine, Ser, S.
- Threonine.
  - Chirality: Same drawing style as isoleucine; R, though. Again, this chirality is not required.
  - Name: Threonine, Thr, T.
- Cysteine.
  - Very similar to serine.
  - Forms disulfide bonds to bring distal ends or subunits of a protein together.
  - Name: Cysteine, Cys, C.
- Asparagine.
  - Related to aspartate; we just change the carboxylic acid to an amide.
  - Frequently found as a metal coordinate.
  - Can also form H-bonds with other amino acids.
  - Name: Asparagine, Asn, N.
- Glutamine.
  - Related to glutamate; we just change the carboxylic acid to an amide.
  - Same metal-coordinating and H-bonding properties as asparagine.
  - Name: Glutamine, Gln, Q.

- A colleague asked Tang what AA he should substitute for alanine to prove that it's absolutely conserved.
  - She suggested fellow small amino acids (valine or leucine) as well as achiral glycine to determine if either size or chirality is important in that position.
  - Overall, this is a very hard to answer question.
  - You often find that alanine is needed because it doesn't disrupt anything; it's an inert filler and doesn't play a role. Other things will typically play a role.
- Amino acids: Hydrophobic side chains, acidic side chains, basic side chains, and special residues.
  - On the acidic side-chain amino acids: Sometimes the active site can be so well organized that replacing an D with an E will disrupt it.
  - In addition to cysteine, we sometimes have selenocysteine (it does occur in our bodies, but it's not considered one of the 20 natural amino acis).
    - Has selenium instead of sulfur.
    - Name: Selenocysteine, Sec, U.

## 3.2 Protein Structure and Function

10/13:

- Callie: PSet should take about 3 hours.
- Office hours Monday evening on Zoom.
- Review of amino acids.
  - Tang did her grad work on cysteine.
  - Proline the amino acid is the chiral asymmetric organocatalyst from OChem III!
- Proline biosynthesis:
  - Uses the precursor of glutamate and cyclization happens in the final step.
- Amino acid properties given.
  - Residue mass (minus $H_2O$ because when we add peptides to the chain, we lose water [dehydration synthesis/condensation reaction]).
  - Van der Waals volume (related to mass).
    - Anything else we need to know about this??
  - Frequency in proteins.
    - There are three proteins that have six codons corresponding to them: Leu, Arg, and Ser. L does occur the most frequently, but R and S are farther down the list.
    - Ala has four associated codons and occurs relatively frequently.
    - Met and Trp have one associated codon, each, and are pretty far down the list (W is the least frequent, but His occurs less frequently than M for example).
  - Not testable: W has a relatively unique UV-Vis absorbance at 280 nm (Phe and Tyr can contribute a bit but not much).
    - Another way to quantify the concentration of protein in the lab is to use a **blackfield assay**.
  - Cysteine is rare because it can form disulfide bonds, so we don't want it everywhere.
    - 2 corresponding codons.
- **Blackfield assay**: An assay in which you use a dye that changes color when it interacts with the protein; you then quantitatively measure the change in color.

- p$K_a$'s of side-chain groups.
  - The $\alpha$-amino and $\alpha$-carboxyl groups are not super important (they are only present at the ends of the proteins).
  - The negatively charged amino acids have p$K_a \approx 3.9 - 4.0$ (D) and p$K_a \approx 4.3 - 4.5$ (E), making them properly acidic.
  - Arg's guanidinium moiety has p$K_a \approx 12$.
  - Lys's amino moiety has p$K_a \approx 10$ (or a bit larger).
  - Thiols, imidazoles, and phenolic hydroxyls are all in the viscinity of 6-10 as well.
  - p$K_a$'s are highly context dependent.
    - Depends on electrostatic effects, H-bonding effects, and inside a protein (more difficult to ionize a residue here).
    - Recall the context-dependent pH of A2486 in the ribosomal mechanism (see the discussion associated with Figure 2.5).
- Main chain ionization.
  - If pH $< 2$, everything will be protonated ($NH_3^+$, COOH).
  - If pH $> 9$, everything will be deprotonated ($NH_2$, $COO^-$).
  - If $2 <$ pH $< 9$, we will have the zwitterionic form ($NH_3^+$, $COO^-$).
- Table of p$K_a$.
  - p$K_{a_1}$ corresponds to the carboxylic acid, p$K_{a_2}$ corresponds to the amine, and p$K_{a_3}$ corresponds to the side chain (if applicable).
- Aspartic acid pH analysis.
  - As the pH increases, $\alpha$-carboxylic acid is deprotonated, then the side chain carboxylic acid, then the $NH_3$.
  - We might have a test question like this.
- Chemical environment affects p$K_a$ values.
  - The $\alpha$-carboxy group in amino acids is much more acidic than in carboxylic acids because the conjugate base is stabilized significantly by the zwitterionic form.
  - The $\alpha$-amino group in amino acids is slightly less basic than in amines because the electronegative oxygen atoms of the $\alpha$-carboxy group act as EWGs.
- Formulation of peptides.
  - Peptides are small condensation products of amino acids.
  - They are "small" compared to proteins ($M_w < 10\,\text{kDa}$).
  - Review of peptide bond formation (see the discussion associated with Figure 2.5).
  - Overlap between "polypeptides" and "proteins."
    - Arbitrarily, people use 10 kDa as a cutoff.
- Peptide ends are not the same.
  - Synthesis occurs from the N terminus to the C terminus (remember, the new amine is attacking the carboxylic acid of the previous).
  - Solid phase peptide synthesis occurs in the reverse direction.
  - You number the amino acids $\text{AA}_1, \ldots, \text{AA}_n$ starting from the N terminus.

- How do such simple building blocks result in diverse functions?
  - Enzymes play the majority of the structure and catalysis functions in our body, e.g., enzymes, receptors, antibodies, hormones, regulatory roles, structural, etc.
  - A protein of 100 amino acids has $20^{100} \approx 10^{130}$ possible sequences.
  - An average 100 amino acid protein has a mass of 14 000 Da; thus, $10^{130}$ such molecules would have a mass of $1.4 \times 10^{134}$ Da. For reference, the mass of the universe is about $10^{80}$ Da.
- Note on directed evolution.
  - Nature has not sampled every AA sequence, but has optimized over some.
  - A reasonable (but still tough) sample size library to achieve is $10^8$. Thus, the sampling space for any directed evolution experiment is necessarily limited.
  - Your directed evolution experiment only works because there are multiple answers in the solution space.
- Favorable interactions in proteins.
- **Hydrophobic effect**: Release of water molecules from the structured solvation layer around the molecule as protein folds increase the net entropy.
- **Hydrogen bonds**: Interaction of N–H and C=O of the peptide bond leads to local regular structures such as $\alpha$-helices and $\beta$-pleated sheets.
- **London disperson effect**: Medium-range weak attraction between all atoms contributes significantly to the stability of the interior of the protein.
- **Electrostatic interactions**: Long-range strong interactions between permanently charged groups.
  - Example: Salt bridges, especially those buried in the hydrophobic environment which strongly stabilize the protein.
- 4 levels of protein structure.
  - Primary: Amino acid sequence.
  - Secondary: $\alpha$-helix or $\beta$-pleated sheet.
  - Tertiary: Larger protein chunks, made of a single polypeptide chain.
  - Quaternary: Most proteins are made of multiple such moieties/are assembled subunits.
- Natural protein function beyond proteinogenic amino acids.
  - Proteinogenic amino acids have limited chemical functionality.
  - Natural proteins are especially bad at redox reactions.
  - This functionality is expanded in nature with...
    1. Small molecule cofactors.
       - Enable redox reactions and can serve as an electron sink.
       - Example: NADH, NADPH.
       - Some of these reactions may have come up in OChem III. We can also learn more about them if we take some of Tang's other courses.
    2. Post-translational modification.
       - Example: Phosphorylation.
    3. Post-translational rearrangement.
       - In GFP, for instance, we have a post-translational rearrangement between serine, glycine, and tyrosine that leads to the formation of a chromophore.

- The peptide bond.
  - Carbonyl oxygen and amine hydrogen are *trans*.
  - This is because, drawing resonance structures, we see that the peptide C–N bond has some $\pi$ character. $p$-orbital overlap. Thus, we will say it does not rotate meaningfully.
- The rigid peptide plane and the partially free rotations.
  - Rotation around the peptide bond is not permitted.
  - Rotation around bonds connected to the $\alpha$ carbon is permitted.
  - $\phi$: Angle around the $\alpha$-carbon – amide nitrogen bond.
  - $\psi$: Angle around the $\alpha$-carbon – carbonyl carbon bond.
  - In a fully extended polypeptide, both $\psi, \phi = 180°$, but this is not common.
  - Even without the side chain, some angles are not permitted due to sterics. Bulkier side chains narrow the allowable range still further.
- The polypeptide is made up of a series of planes linked at the $\alpha$ carbon.
- Distribution of $\phi$ and $\psi$ dihedral angles.
  - Some $\phi$ and $\psi$ combinations are very unfavorable due to sterics.
  - Some $\phi$ and $\psi$ combinations are more favorable because of the chance to form favorable H-bonding interactions along the backbone.
  - We won't be asked to memorize any good or bad $\phi, \psi$ angles.
- **Ramachandran plot**: A plot showing the distribution of $\phi$ and $\psi$ dihedral angles that are found in a protein.
  - Gives the distribution of secondary strucures in the $\phi\psi$-plane.
  - Shows the common secondary structure elements and reveals regions with unusual backbone structure.
  - There are characteristic regions for $\alpha$-helices (lower left) and $\beta$-pleated sheets (upper left), but loops are harder to detect (though they may appear to some extent in the upper right).
  - Glycine has density in all four quadrants (it is the most flexible, after all).
- Protein conformational space.
  - $\phi, \psi$ describe torsional angles: $-180° < (\phi \text{ or } \psi) < 180° = -180°$.
- Secondary structure: $\alpha$ helix.
  - Not testable, but $\phi = -57°$ and $\psi = -47°$.
  - Right-handed helix (just like double stranded DNA).
  - 3.6 residues per turn.
  - 5.4 Å rise per turn.
  - H bonds between $i, i + 4$.
  - Proline is a good helix starter. A, R, K, L, and M are good in an $\alpha$-helix; P, G, T, and S are poor.
    - P is too rigid, G is too flexible, and T,S have additional H-bonding donors and acceptors that might disrupt the $i, i + 4$ pattern.
- mRNA is more complex and weighs more than a protein.

  - B-DNA is about 28 Å per turn, larger than the $\alpha$-helix. This makes sense since there are so many moieties involved in DNA (sugar, phosphate, base pair).
  - It's not like it's a small template we're using to synthesize something much larger.
  - Tang explores this theme in the arena of mRNA vs. protein vaccines.
    - For COVID vaccines, we can deliver either the COVID spike protein or the mRNA. The former is a smaller thing to deliver.

- Secondary Structure: Antiparallel $\beta$-sheet.
  - Adjacent strands run in opposite directions.
  - Hydrogen bonds are neatly stacked one carbons apart, always.

- Secondary Structure: Parallel $\beta$-sheet.
  - Adjacent strands run in the same direction.
  - Hydrogen bonds are not neatly stacked. They are bent and evenly spaced, though.

- Loops connect secondary structure.
  - Usually rich in polar residues.
  - Loops are irregular structures.
  - H-bond with solvent.
  - Gly = common start or end.
  - Often contain binding sites or enzyme active sites.
    - Loops are more flexible, so they can test out more conformations.
    - The lock and key model is misleading — it's not a rigid interaction, but rather the protein adjusts when the substrate binds.

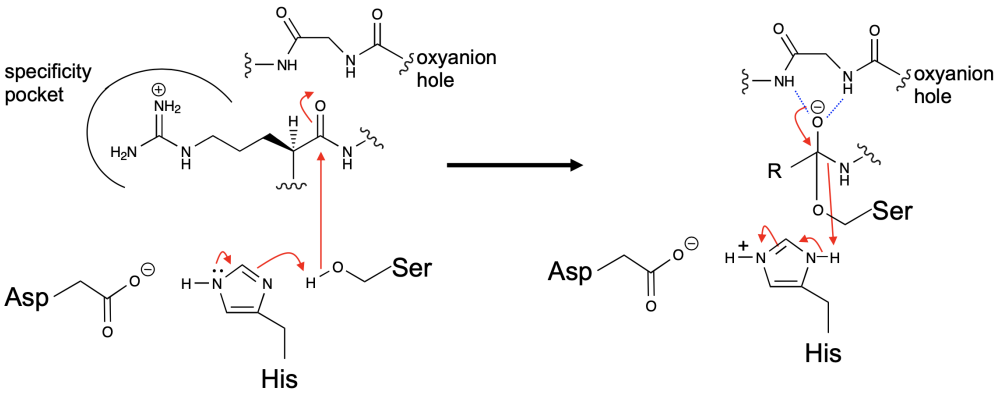
- Higher order protein structures.
  - $\alpha$-helix, loop, $\beta$-strand $\rightarrow$ motif $\rightarrow$ domain $\rightarrow$ protein.

- **Tertiary structure**: Overall 3D structure of the protein; describes how the peptide chains fold and pack.
  - Covalent structures (peptide, disulfide bonds).
  - Hydrogen bonds, hydrophobic interactions, electrostatic interactions, and van der Waals interactions.
  - Recall that disulfide bonds bring distal ends of a protein together.

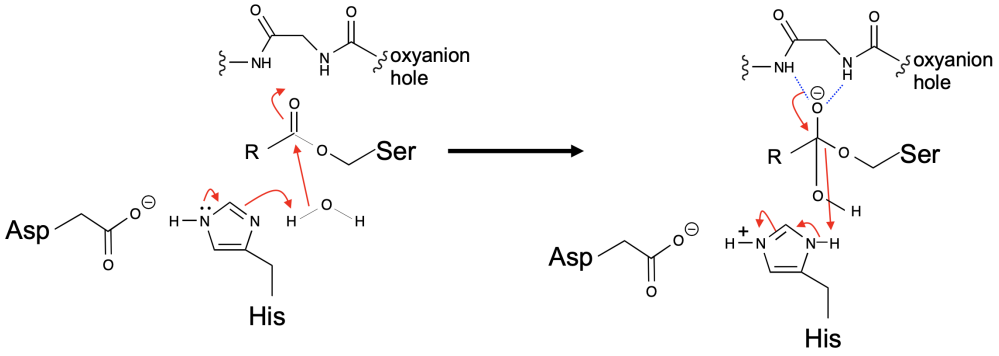
- Quaternary structures are formed of multiple tertiary motifs.

- **Enzyme catalysis**: Enzymes are protein catalysts of biologically relevant chemical reactions and typically display a high efficieency and selectivity.
  - Part of **mechanistic enzymology** (which Tang used to cover).
  - Enzymes can accelerate some reactions by 17 orders of magnitude.
  - Enzymes are typically highly selective.
  - Enzymes function either by stabilizing the transition state of a reaction more than they stabilize the ground state of the substrate or by providing an alternative reaction pathway (mechanism) that involves a lower activation barrier.

- One example of enzymatic catalysis: Protease (also called a peptidase or proteinase) is an enzyme that catalyzes the hydrolysis of a protein amide bond.




(a) Step 1.




(b) Step 2.

Figure 3.6: Serine protease.

- There is either 1-step or 2-step protease catalysis.
  - In 1-step, an acidic side chain (e.g., D, E, or a metal ion [perhaps bound to N or Q]) activates a water molecule to attack the peptide bond.
  - In 2-step, a nucleophilic side chain (e.g., C, S, or T) splits the peptide bond and then water comes in to break off the side chain from the carbonyl derivative.
- Serine protease is a classic example of a protein participating in catalysis covalently; a good example in protein engineering, where it can be developed into a ligase.
- Serine protease has a catalytic triad within the active site.
  - Histidine (good for its proton shuffle ability) deprotonates serine protease.
  - The now positively charged histidine is stabilized by the negatively charged aspartate (though protonation/deprotonation does not occur).
  - Now that serine has been deprotonated, it is an excellent nucleophile and can attack the peptide amide via nucleophilic acyl substitution. The departing amine grabs an extra proton back from histadine.
  - In step 2, we have another histadine-involved deprotonation to start, but this time of water.
  - The lone hydroxyl group then attacks the serine-bonded amino acid's carbonyl group again, doing a transesterification nucleophilic acyl substitution, and serine regenerates itself from the histadine proton.
- Note that most serine proteases come with a specificity pocket that recognizes one specific type of amino acid at which we want to cleave the peptide bond.

  - Serine proteases are quite important.
  - Penicillin (the antibiotic) works by inhibiting a kind of serine protease in bacteria (a transpeptidase used to build the cell wall) that's not found in humans. This makes it so that bacteria cannot construct their cell walls.

- Lecture summary.
  - 20 natural amino acids offer different features.
  - Post-translational modifications and rearrangements expand the functional repertoire of proteins.
  - Protein structure is determined by the nature of the amide bond and by at least four noncovalent interactions (and disulfide bonds).
  - Natural protein structures are highly modular (helix, sheet, loop; motif; domain; protein) and typically satisfy several constraints.
  - Side chain structure determine accessible conformational space at each peptide bond.
  - A subset of proteins serve as catalysts for chemical conversions by lowering the activation barrier.