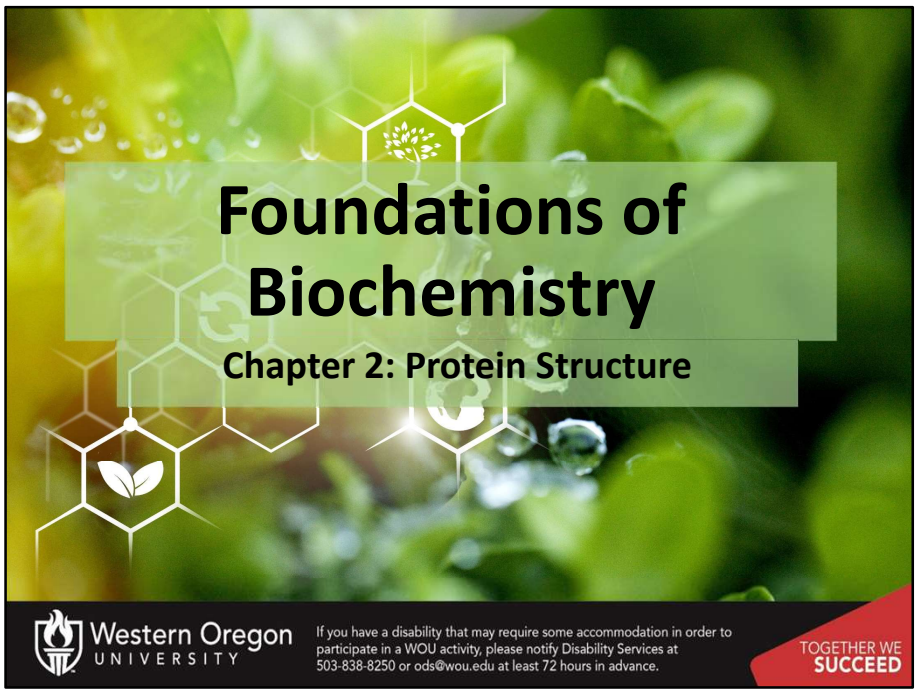
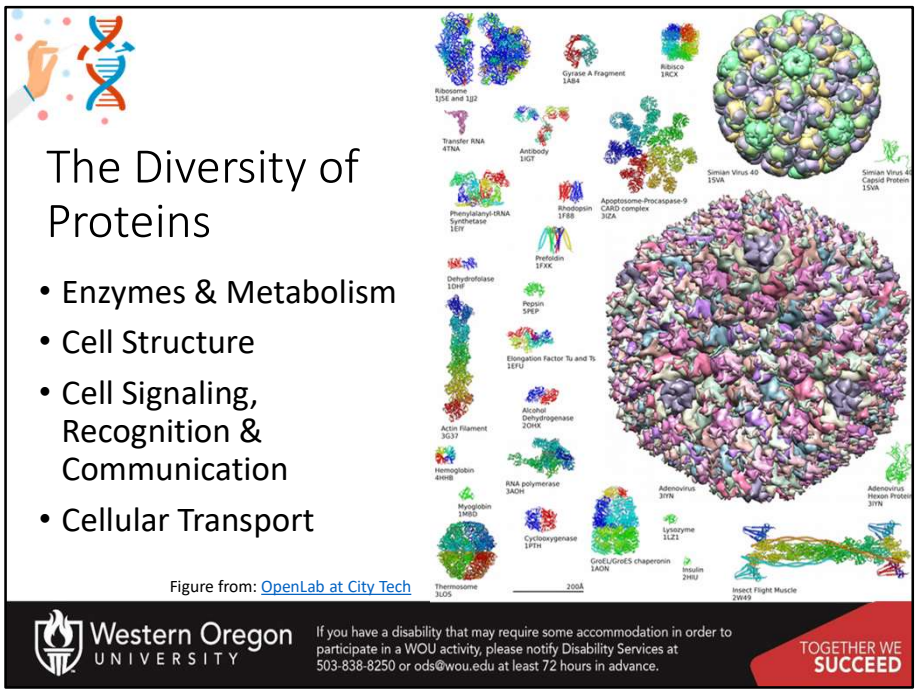
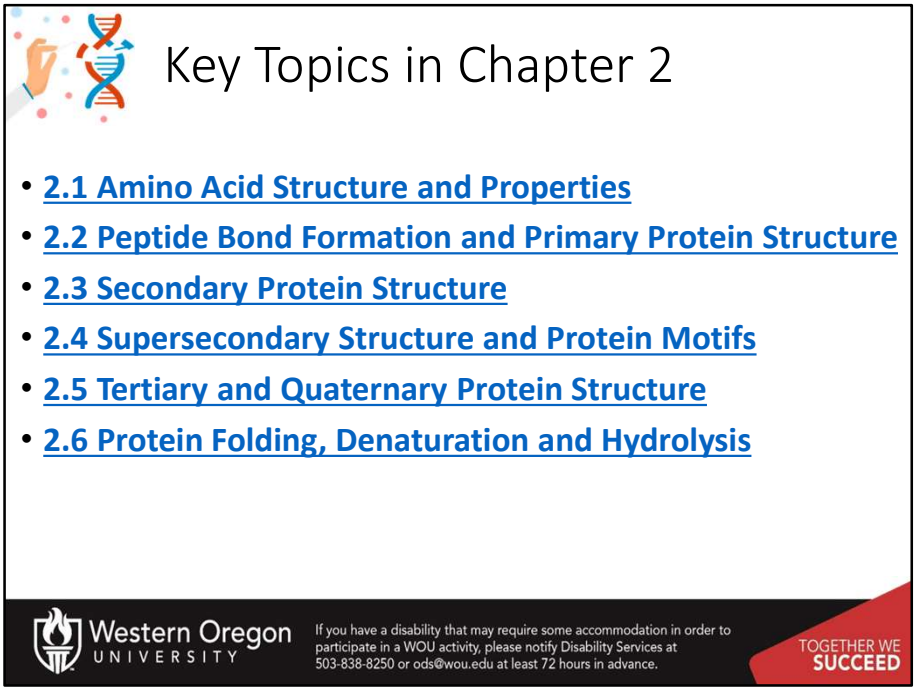


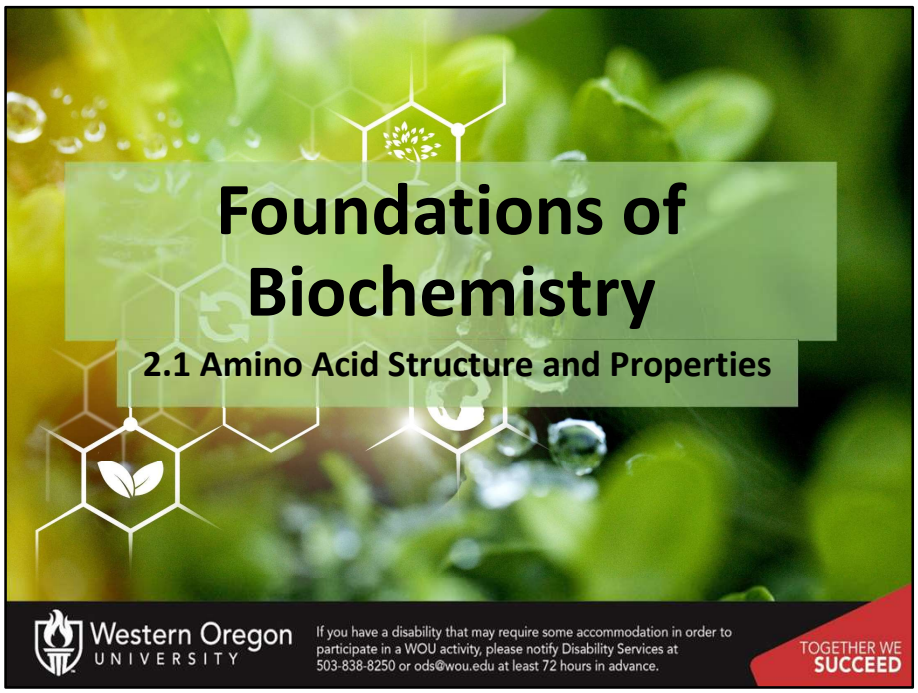
The focus of Chapter 2 is on protein structure. Proteins are



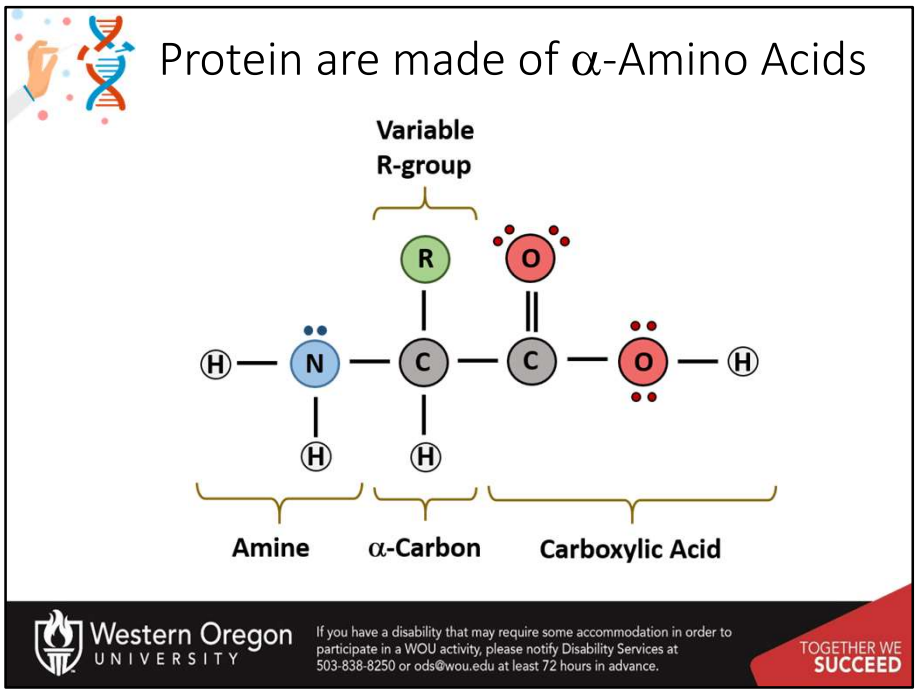
key components in regulating cellular processes through their abilities to function as enzymes, play structural support roles, mediate communication and cell signaling, and transport molecules to different locations. They come in many shapes and sizes, yet are constructed from the same set of 20 building block amino acids. Understanding this vast diversity is important for understanding disease, medicine, and treatment strategies.



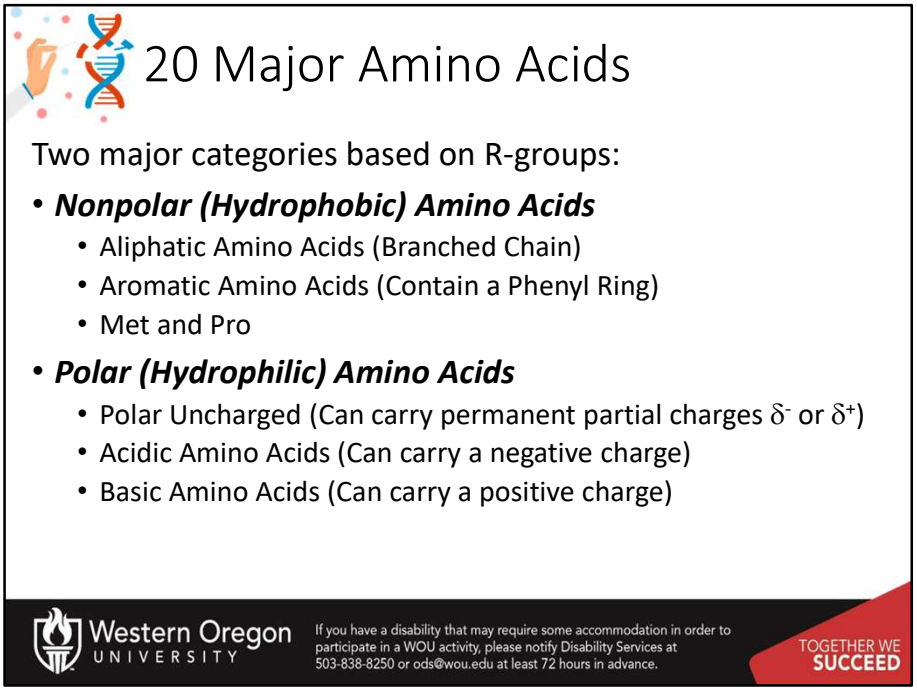
Here we will focus on: amino acid structure and properties, peptide bond formation and primary protein structure, secondary protein structure, supersecondary structure and protein motifs, tertiary and quaternary structure, and the processes of protein folding, denaturation and hydrolysis.



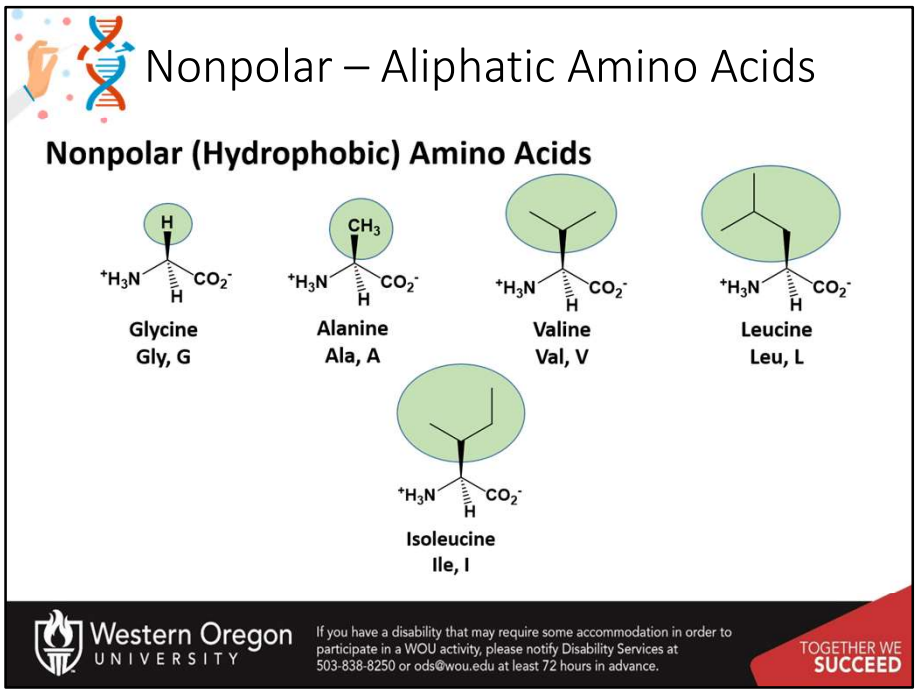
Let’s start with a close look at amino acid structure and properties.



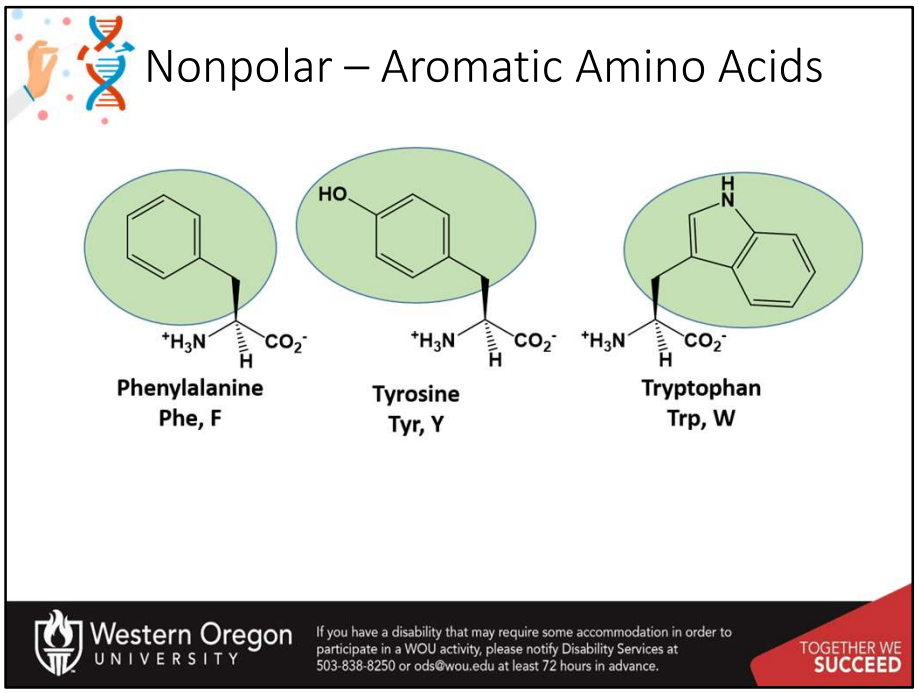
The major building block of proteins are called ***alpha (α) amino acids***. As their name implies they contain a carboxylic acid functional group and an amine functional group. The alpha designation is used to indicate that these two functional groups are separated from one another by one carbon group. In addition to the amine and the carboxylic acid, the alpha carbon is also attached to a hydrogen and one additional group that can vary in size and length. In the diagram, this group is designated as an R-group. Within living organisms there are 20 amino acids used as protein building blocks. They differ from one another only at the R-group position.



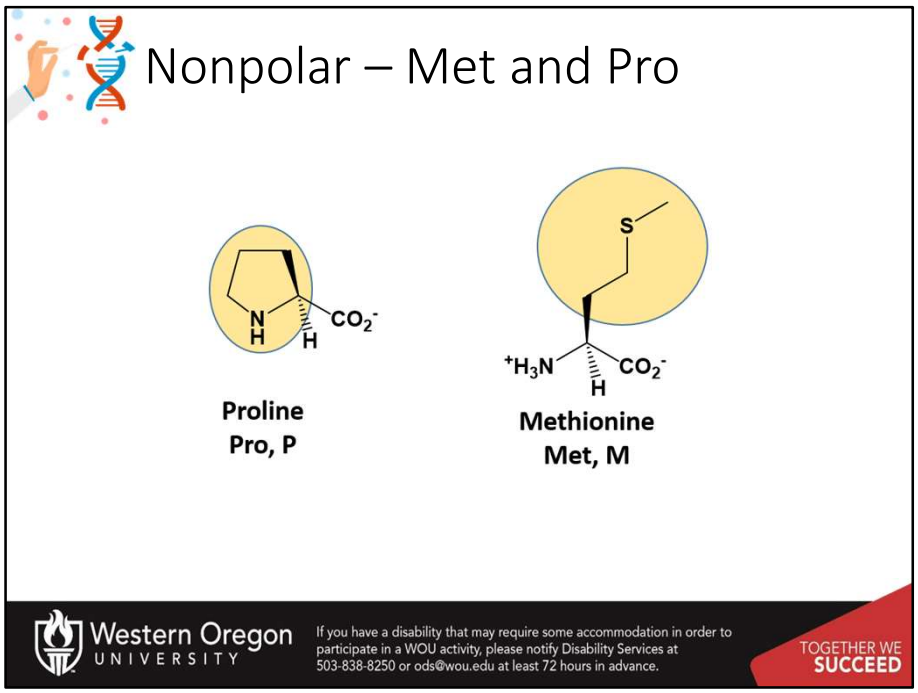
The 20 amino acids used in protein synthesis can be subdivided into two major types, based on the structure of the R-group. They can be divided into Nonpolar or hydrophobic amino acids or Polar (hydrophilic) amino acids. Note that this classification scheme is really a sliding scale between these two states. Amino acids with the highest polarity potential will be able to form full charges, others will carry strong, permanent dipoles, some will have weaker permanent dipoles and others will be very nonpolar and only have London dispersion interactions with other molecules. Each of these broad categories contains subcategories that help us evaluate their reactive potential. For the nonpolar amino acids, this includes the aliphatic and aromatic amino acids as well as Methionine (met) and proline (pro) which don't fit well into the other two categories. The hydrophilic amino acids can be subdivided as polar uncharged, acidic, or basic.



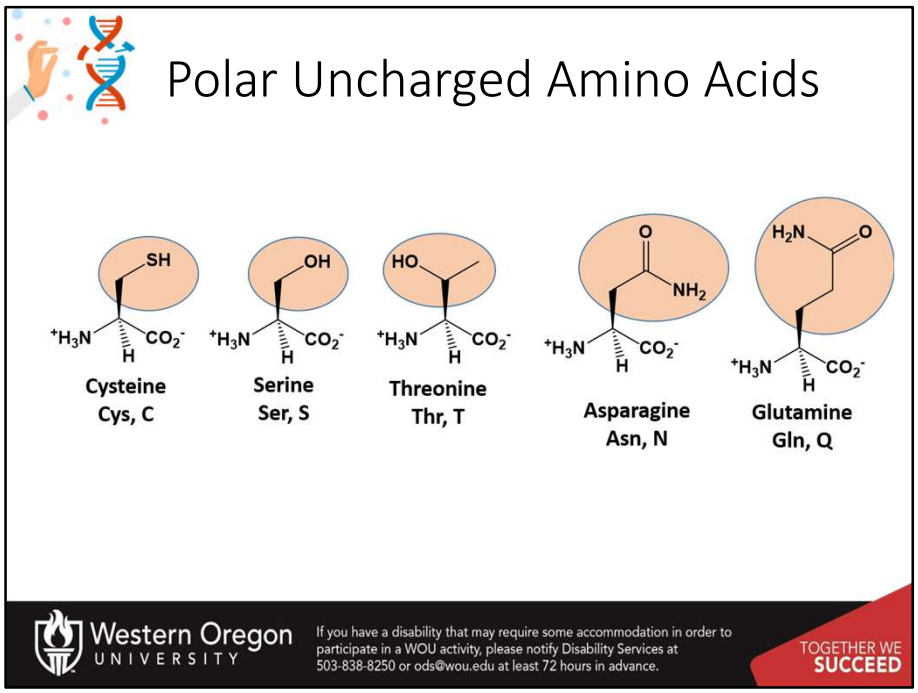
Let's take a closer look at each of these categories, starting with the nonpolar aliphatic amino acids. The term aliphatic is used to refer to organic compound in which carbon atoms form open chains, rather than aromatic rings. Thus, the aliphatic amino acids are Gly, Ala, Val, Leu, and Ile. You should start learning their three letter and one letter codes, and be able to recognize their side chains.



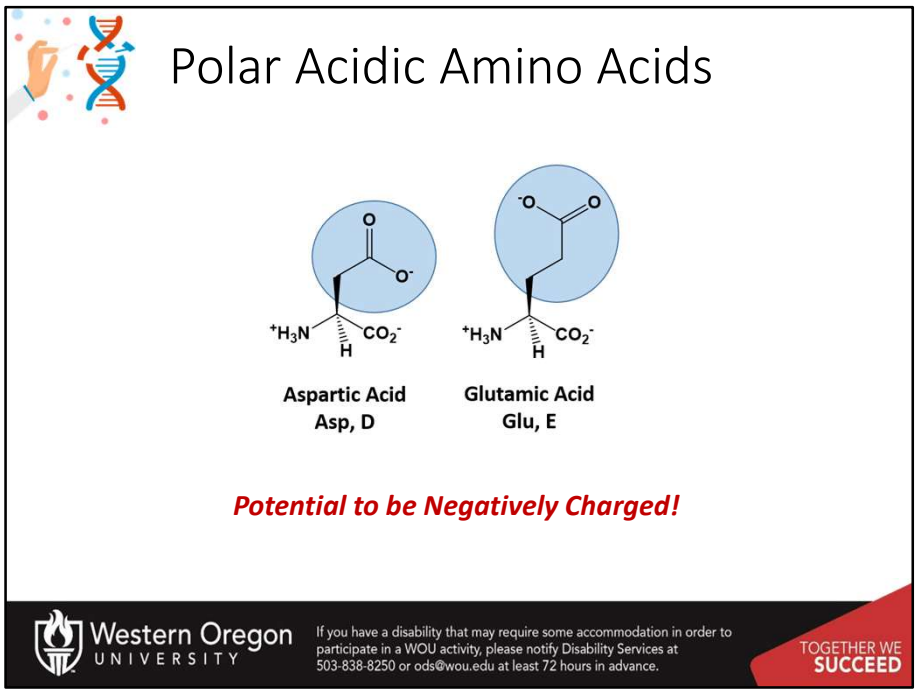
The next category are the aromatic amino acids, that contain the phenyl ring structure. These include Phe, Tyr, and Trp.



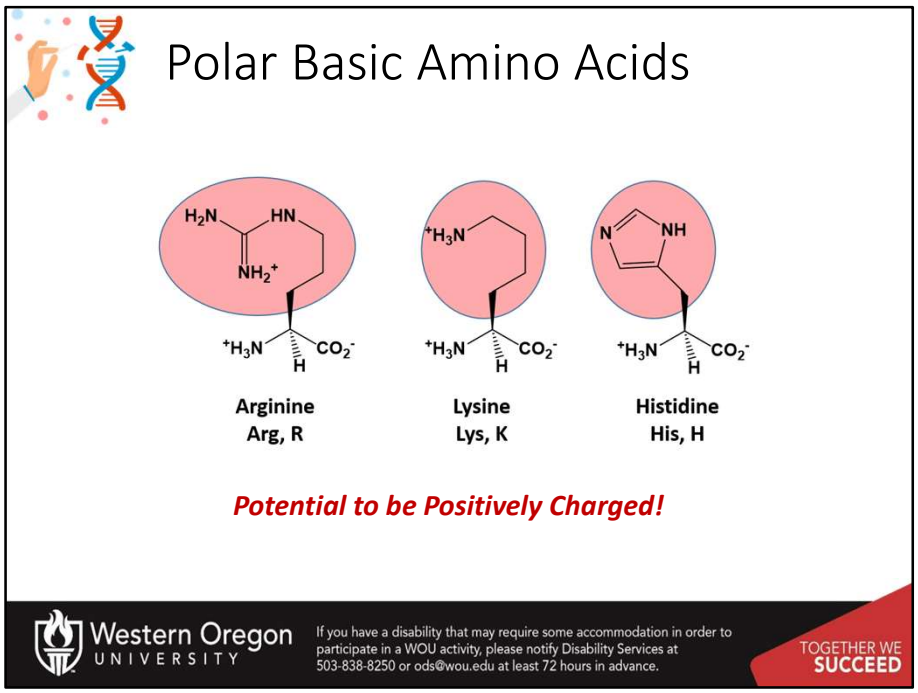
The other two amino acids that have high nonpolar character are Pro, and Met. Notice that the R-group of pro circles back and forms a closed ring structure with the amine functional group in the backbone. This causes proline to adopt an alternate conformation (the cis conformation) within the protein backbone compared with the other amino acids that adopt a trans conformation. Thus, prolines often cause direction changes within the 3-dimensional structure of the protein and can destabilize secondary structure such as alpha helices or beta pleated sheets. Met is also an important amino acid, as it is amino acid that most often begins the translation of a protein.



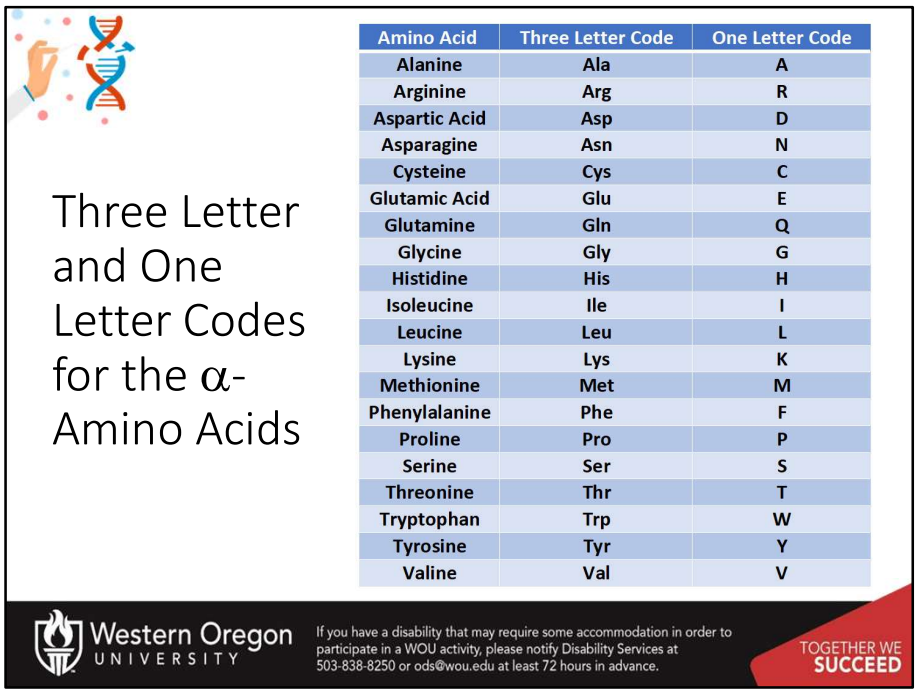
The polar uncharged amino acids, carry permanent dipoles that allow them to form hydrogen bonds with water molecules and other dipole-dipole interactions. These include amino acids that contain alcohols, thiols and amides within their R-groups.



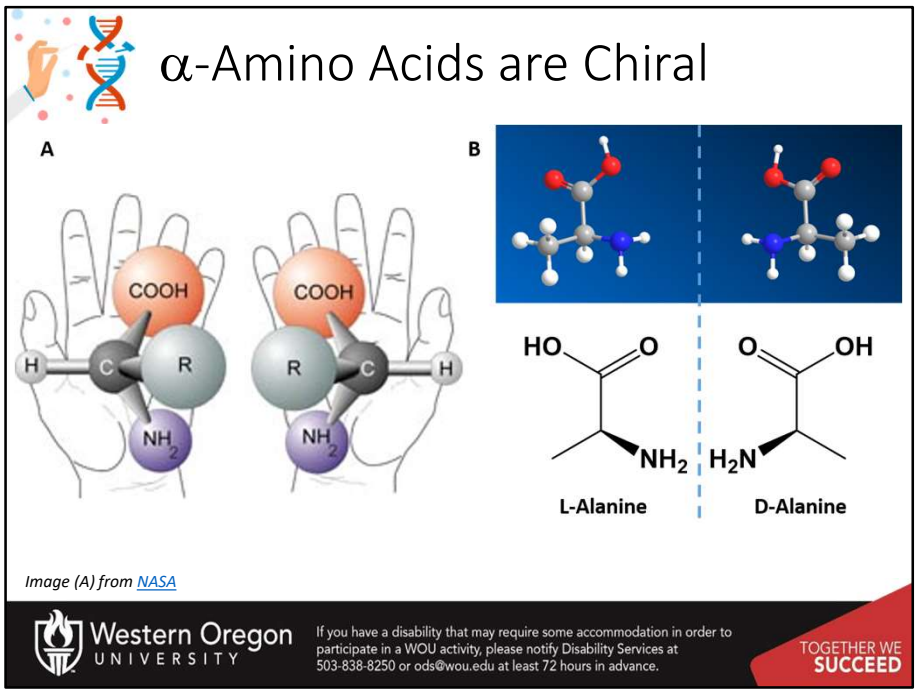
Some polar amino acids may also fully ionize and carry charges. There are two acidic amino acids (glutamic acid and aspartic acid) that can lose a proton from a carboxylic acid functional group and carry a full negative charge on their R-groups.



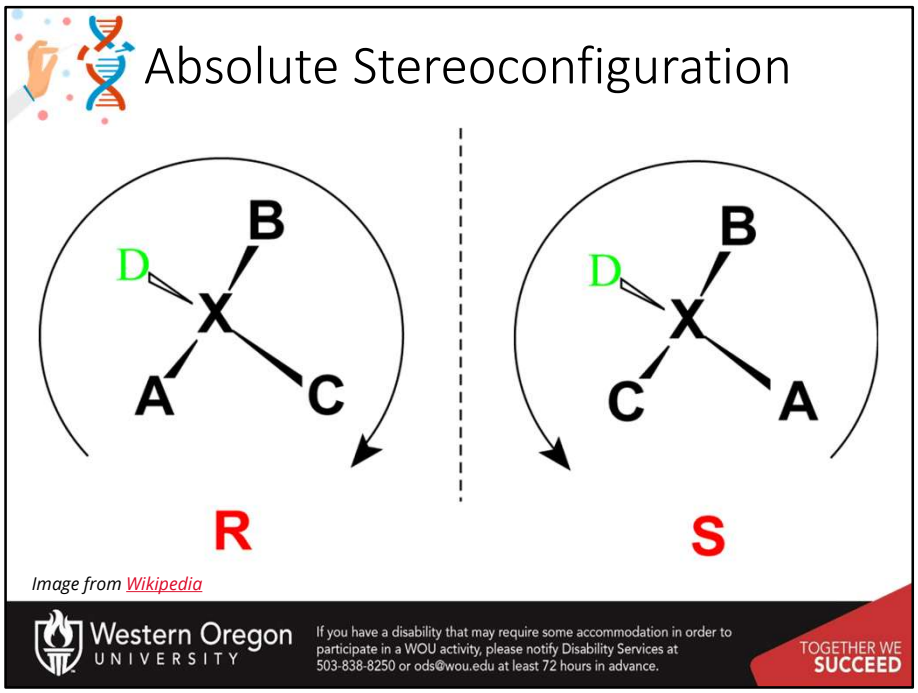
There are also three basic amino acids, Arg, Lys, and His that have amine functional groups that can serve as a proton acceptor and act as a Lewis Base. When they serve as a proton acceptor, they will carry a positive charge.



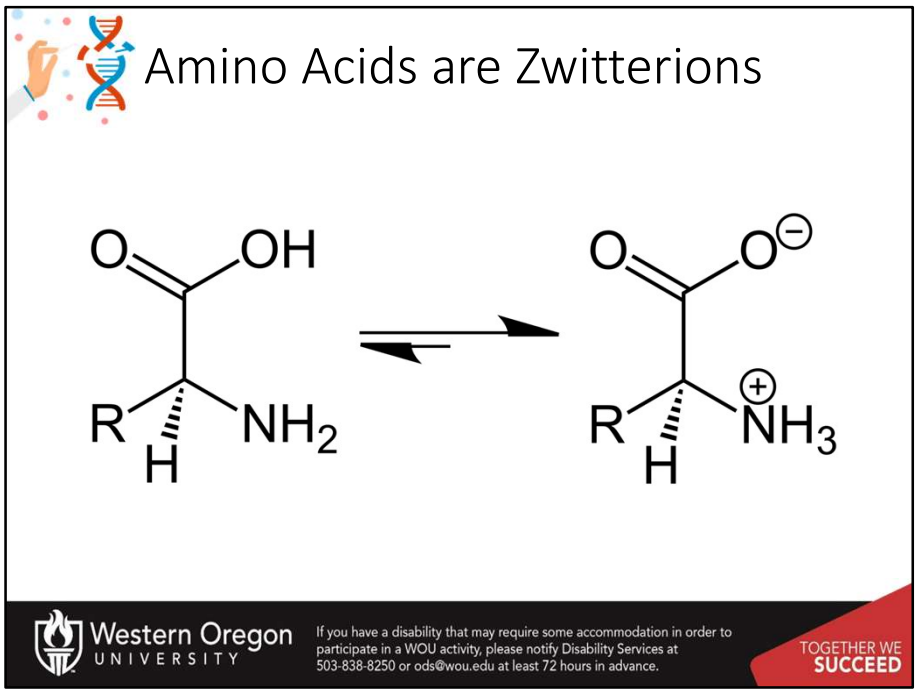
Here is a chart with all of the three letter and one letter amino acid codes. You will need to know these. We will be using these designations all term as we continue to study the structures of proteins. Knowing their abbreviations and being able to recognize their side groups and understand their charge states will enable you to better understand enzyme reaction mechanisms and features of protein folding.



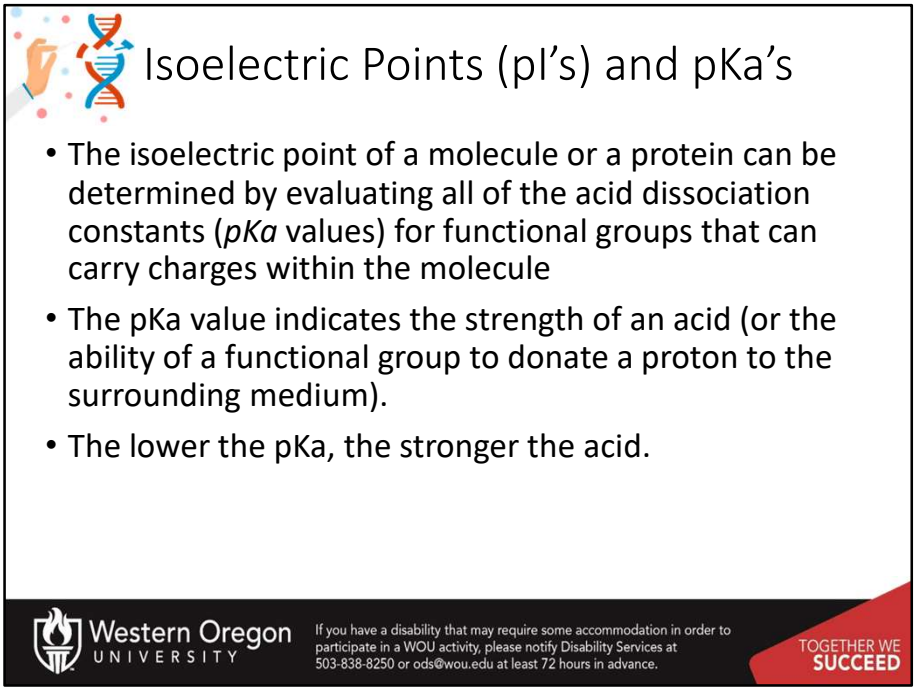
All but one of the alpha amino acids used for protein synthesis are chiral in nature. Glycine is the exception, which has –H as its R-group. Molecules that are ***chiral*** are not superimposable with their mirror image. Because of this nature, they are given a special stereoisomer name called ***enantiomers*** and in fact, the compounds themselves are given the same name! These molecules do differ in the way that they rotate plain polarized light and the way that they react with and interact with biological molecules. Molecules that rotate the light in the right-handed direction are called dextrorotary and are given a D- letter designation. Molecules that rotate light in the left-handed direction are called levorotary and are give an L- letter designation to distinguish one enantiomer from the other. The D- and L- forms of alanine are shown here. The L-forms of the alpha amino acids are common in nature and are the ones that are incorporated into proteins.



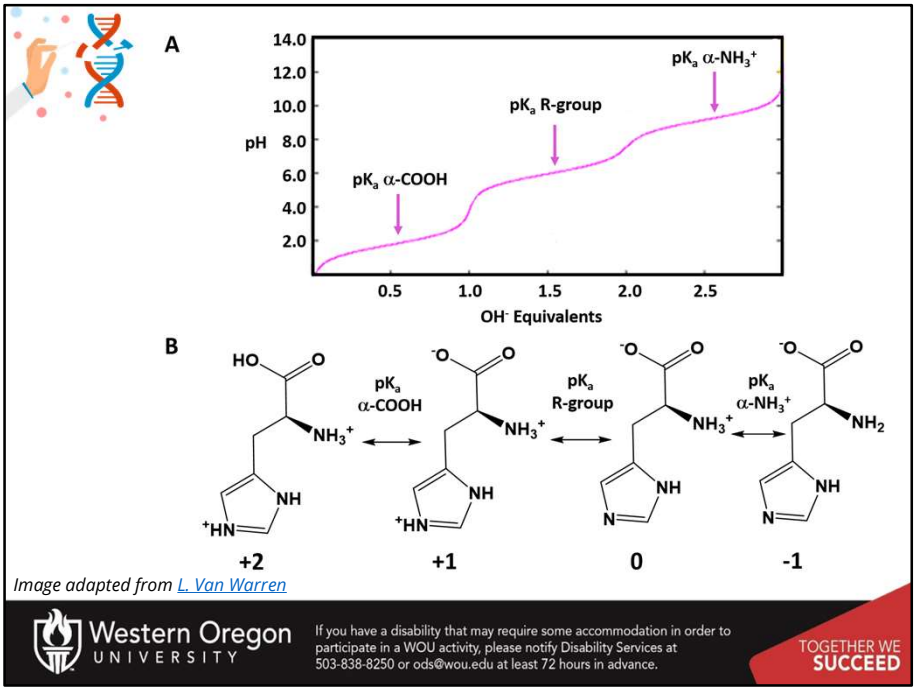
An **absolute configuration** refers to the spatial arrangement of the atoms of a chiral molecular entity (or group) and its stereochemical description e.g. **R** or **S**, referring to **Rectus**, or **Sinister**, respectively. Absolute configurations for a chiral molecule (in pure form) are most often obtained by X-ray crystallography. Alternative techniques are optical rotatory dispersion, vibrational circular dichroism, use of chiral shift reagents in proton NMR and Coulomb explosion imaging. When the absolute configuration is obtained the assignment of *R* or *S* is based on the **Cahn–Ingold–Prelog priority rules,**



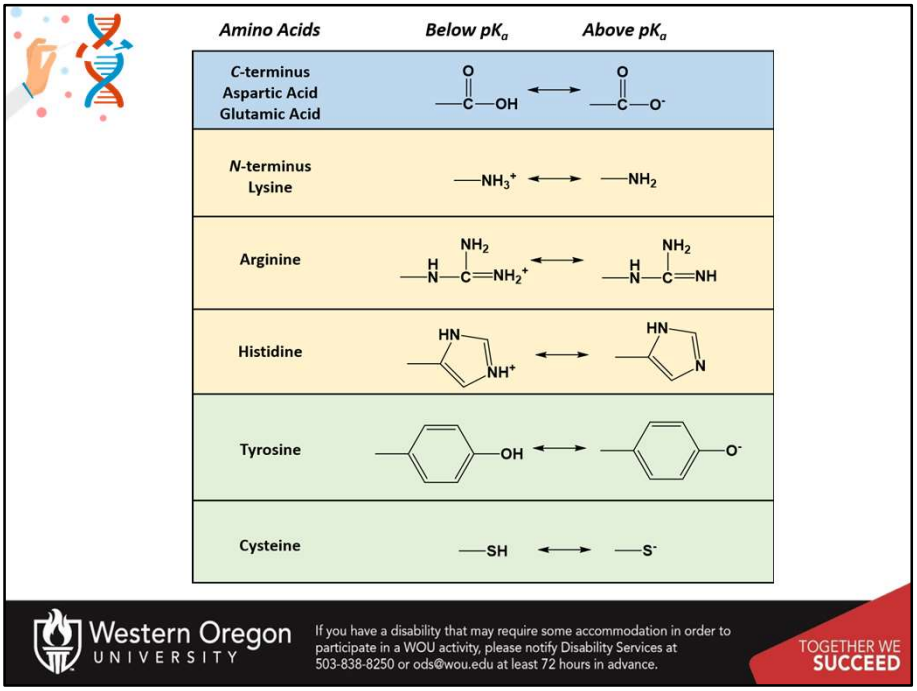
Amino acids are also known as zwitterions. A ***zwitterion*** is a molecule with two or more functional groups, of which at least one has a positive and one has a negative electrical charge and the net charge of the entire molecule is zero at a specific pH. Because they contain at least one positive and one negative charge, zwitterions are also sometimes called ***inner salts***. The charges on the different functional groups balance each other out, and the molecule as a whole can be electrically neutral at a specific pH. The pH where this happens is known as the ***isoelectric point (pI).***



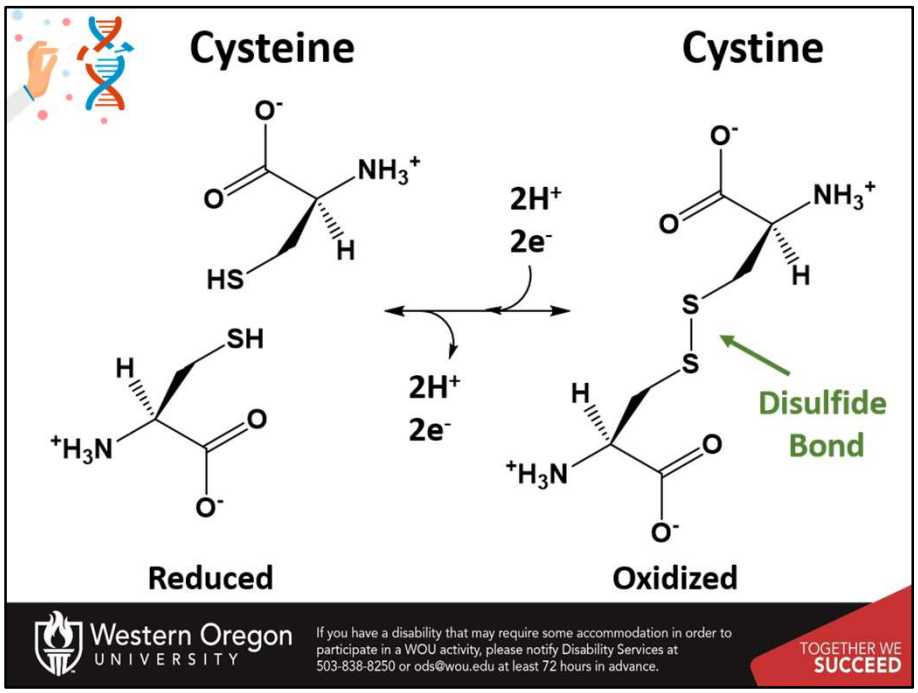
In chapter 1, we reviewed the concept of pKa which indicates the strength of an acid (or for our purposes, indicates the ability of a functional group to donate a proton) Thus, acidic functional groups (carboxylic acids) have low pKa values and amine (or basic) functional groups tend to have high pKa values. Each ionizable group found within the molecule will have a distinct pKa value. The pH at which the overall molecule carries no net electrical charge is known as the isoelectric point (pI). Amino acids all have at least two pKa values for the main chain carboxylic acid and amine functional groups. Some have additional ionizable groups within their R-groups.



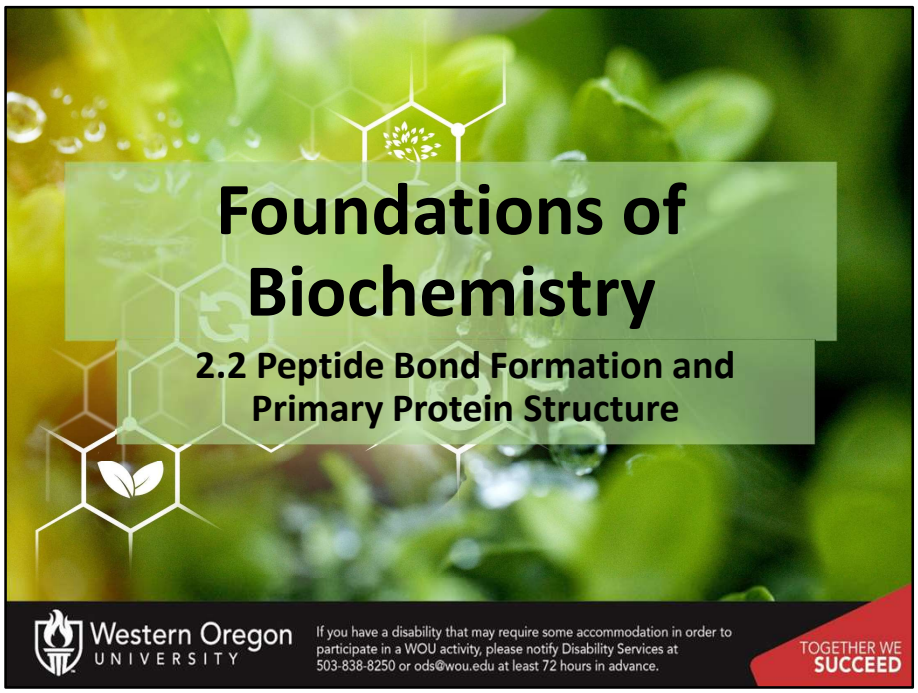
Due to their variable pKa values, the charge state of amino acids will vary at different pH conditions. When a functional group is below it's pKa value, it will predominantly be in its protonated form. Above the pKa value it will be predominantly in the deprotonated form. The graph shown here is the titration curve for histidine. His has three pKa values: the alpha carboxylic acid at 1.86, the R-group amine at 6.04, and the alpha amine group at 9.17. Thus, at different pHs, you can predict the charge state of each ionizable group. For example, at pH values below 1.86, all of the ionizable groups will be protonated. Protonation of the carboxylic acid will neutralize it, while protonation of the amine functional groups means that they will each carry a +1 charge. Thus, the overall molecule is at a +2 state. In between pH 1.86 and 6.04, the carboxylic acid functional group will be deprotonated, but both amine groups will be protonated. This yields a net charge of +1 on the amino acid. Between a pH of 6.04 and 9.17, both the alpha carboxylic acid and the R-group amine will be in the deprotonated state, while the alpha amine will still be protonated. At this stage, the net charge on histidine is zero, as the charges on the alpha carboxylic acid and the alpha amine functional group cancel each other out. In fact, the isoelectric point for histidine will be at this inflection point (pH = 7.64). Above a pH of 9.17, all of the functional groups will be deprotonated and the histidine will carry an overall -1 charge, as both of the amine groups are neutral in their deprotonated states.



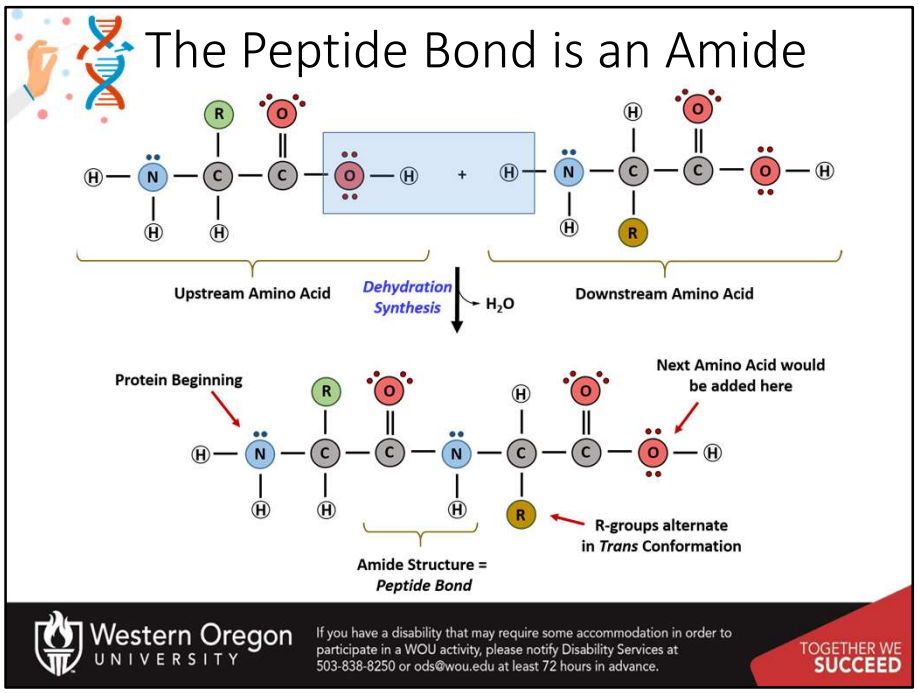
The following diagram shows the amino acids that have ionizable side groups. These include the acidic amino acids (Asp and Glu) and basic amino acids (Arg, Lys, and His). It also includes two polar (typically uncharged amino acids) Tyr and Cys. The R-group of Cys has a pKa of 8.33 and the alcohol group of Tyr has a pKa of 10.07. Thus, for these amino acids to behave as an acid, the pH of the solution has to be quite basic, beyond that which could be supported in biological organisms. However, as we will see, the active site of many enzymes is buried deep inside the folded regions of the protein and often excludes water. Under these conditions, the pKa values can shift causing unexpected behavior. Cys and Tyr are often active site residues involved in the reaction mechanism of enzymes.



The thiol functional group of Cys is also highly susceptible to oxidation and will readily form disulfide bond linkages with other Cys residues. The formation of disulfide bonds within or between protein sequences have a profound impact on the 3-dimensional structure, and hence, the biological activity of many proteins and peptides. For example, the biologically active form of insulin requires intact disulfide bond formation. In the next section, we will see how the amino acid building blocks are linked together to form the primary structure of peptides and proteins.

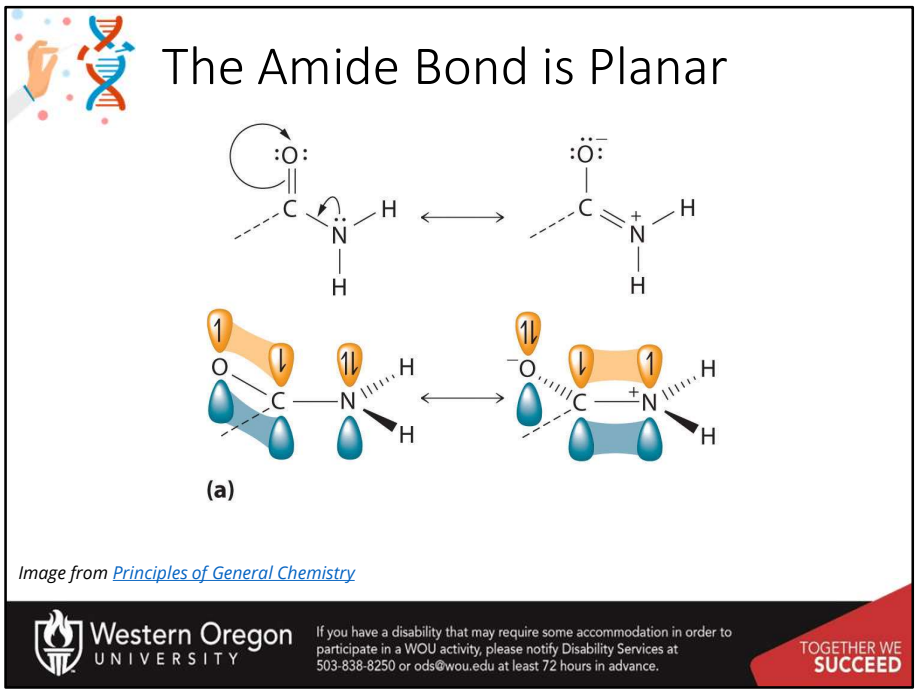


In this section, we will learn how amino acids are joined together forming peptide bonds, as well as learn key characteristics about a protein's primary structure.

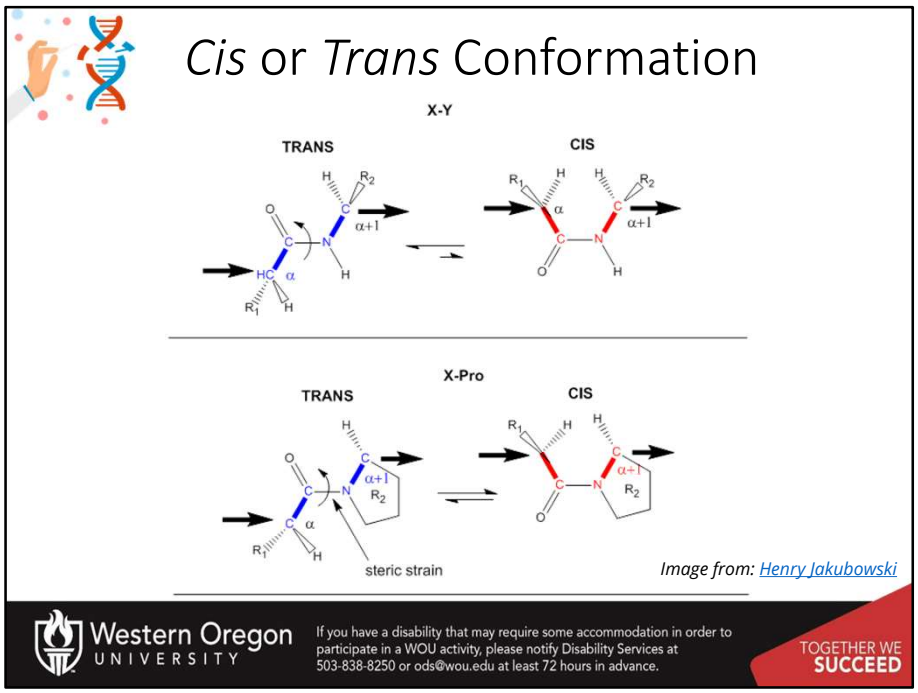


Amino acids are linked together through dehydration synthesis. A key feature of the amino acids joining together is that the R-groups of the amino acids are NOT involved in this process. Only the alpha carboxylic acid and alpha amine functional groups are involved in the peptide bond formation. During this reaction the amine nitrogen from the downstream amino acid, mediates attack on the upstream carbonyl carbon of the carboxylic acid functional group. This leads to the loss of the hydroxyl group from the carboxylic acid and a proton from the amine function group, resulting in water formation. (in reality, the oxygen comes the carboxylic acid and the two hydrogens from the amine functional group as amino acids typically exist in their zwitter ionic state at cellular pH levels. But it is helpful to see the water removal in the deionized state). Some key features of the peptide bond, is that it is an amide functional group. The nascent peptide is also, always put together in the N to C direction (ie the amine of the starting amino acid on the lefthand side and the carboxylic acid of the last amino acid added on the right hand side). This directionality of protein synthesis is created by the action of the ribosome. If we are doing protein synthesis in a test tube, both ends of each amino acid would be reactive and create great difficulty in being able to create a peptide with a desired sequence order. Another key feature that we notice is that the R-groups tend to alternate on either side of the peptide in the *trans* configuration. This is because the R-groups are usually quite large and bulky and if they were on the same side of the growing peptide, they would create too much steric hindrance to be stable. Thus, they are assembled in the more favorable trans conformation.

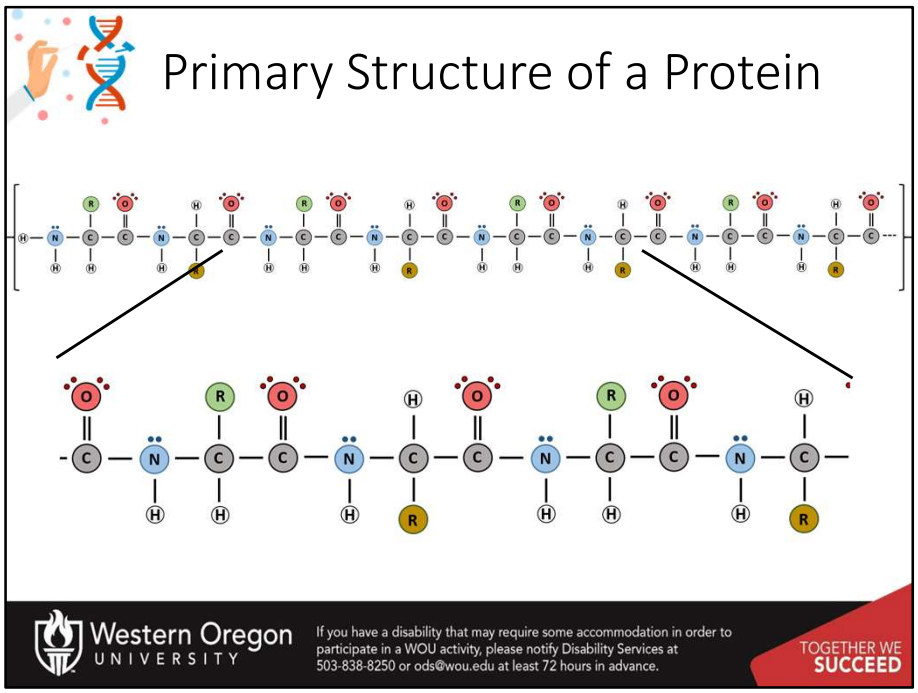
(as we have noted in the previous lecture, proline is the one exception to this rule).



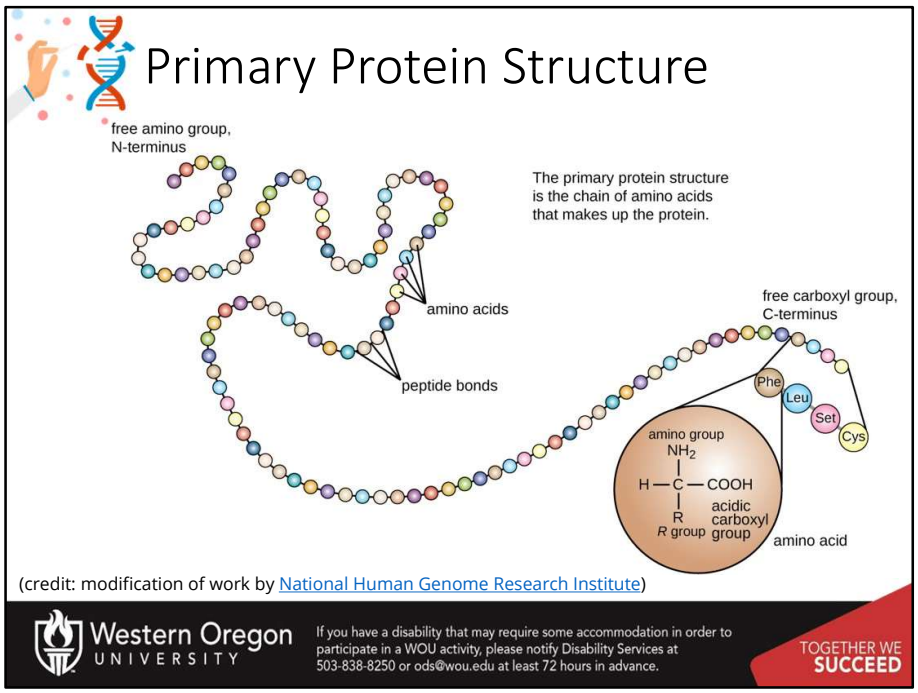
The cis and trans configuration within the protein structure is created because the amide bond formed during peptide synthesis is planar in nature. This is due to the resonance structure of the amide bond. Because the nitrogen atom as a lone pair of electrons, that lone pair can shift in and create double bond character between the nitrogen and the carbonyl carbon atom. This would shift the electrons up to the oxygen momentarily creating a charged state of the oxygen. The electronegativity potential of oxygen allows this state to be possible, even though it is not the favored conformation. Recall that double bonds are created using pi orbitals and that the molecule cannot rotate freely when they are present. This creates the fixed/planar nature of the peptide bond. Note that the resonance structure of the amide bond is also the reason that the amide cannot act as a Lewis Base, like free amines can. The lone pair electrons on the nitrogen are not free to act as a proton acceptor. They are committed to the resonance structure and double bond formation with the carbonyl carbon.



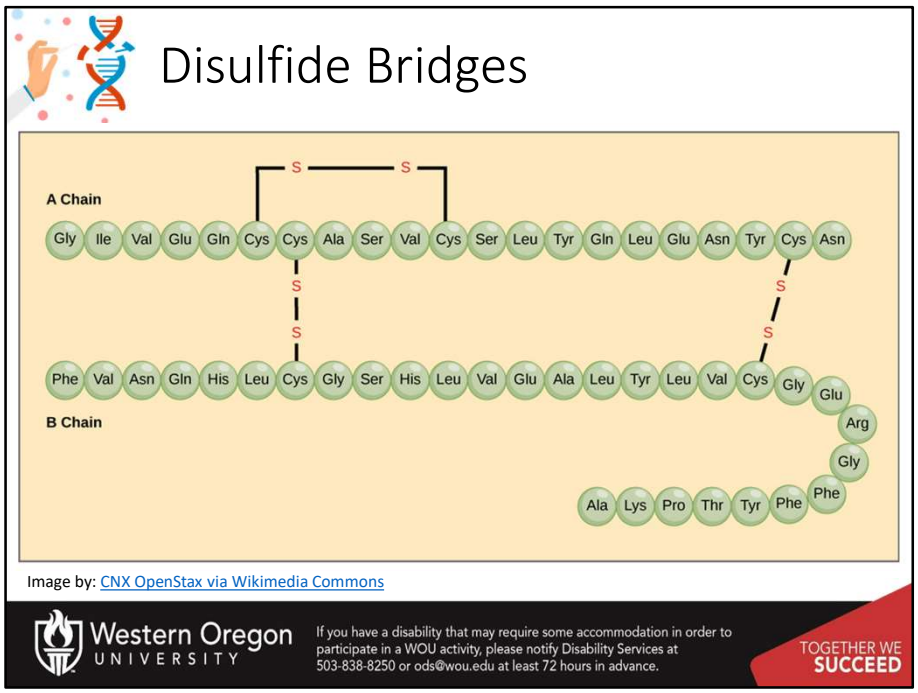
As you can see from this diagram, that most amino acids will adopt the trans conformation to avoid steric hindrance that is created when the R-groups are on the same side of the peptide bond. Proline is the notable exception to this rule. Because proline forms a cyclic ring structure with the alpha amine group, the trans conformation of this amino acid creates more steric strain than the cis conformation. Thus, prolines will adopt the cis conformation and commonly result in bends or kinks within the protein structure.



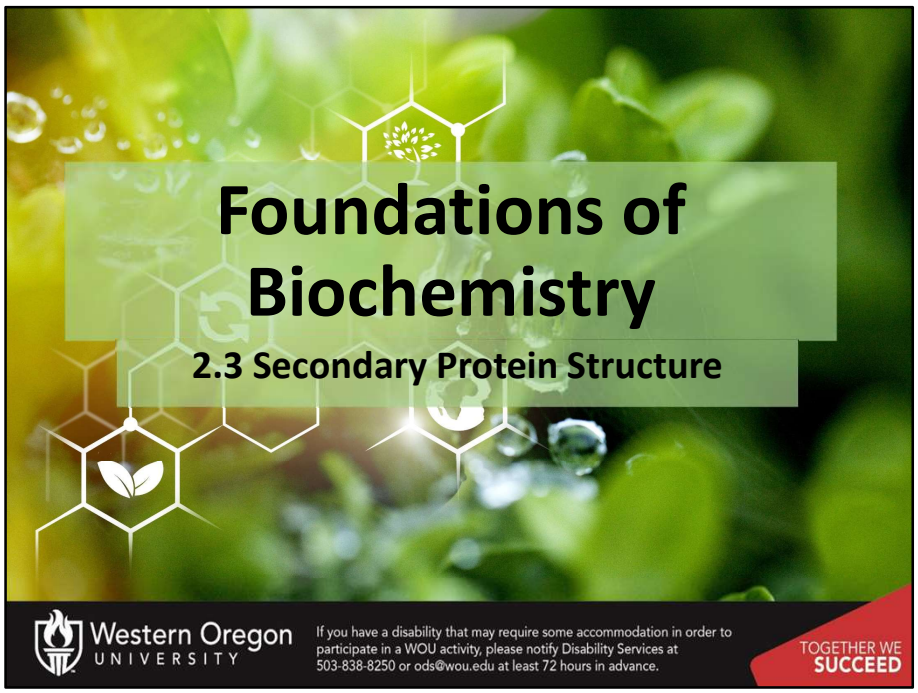
Here is a longer peptide sequence. You should study this sequence such that you can identify where one amino acid starts and stops. Where are the amide linkages? Can you identify the N-terminal of the peptide? This is the starting location of the protein. How about the C-terminal end of the protein? The peptide will always be put together in the N- to C-orientation, and in that respect can be read like a book. Also notice that each R-group is alternating in the trans conformation in this diagram. (only proline would adopt the cis conformation).



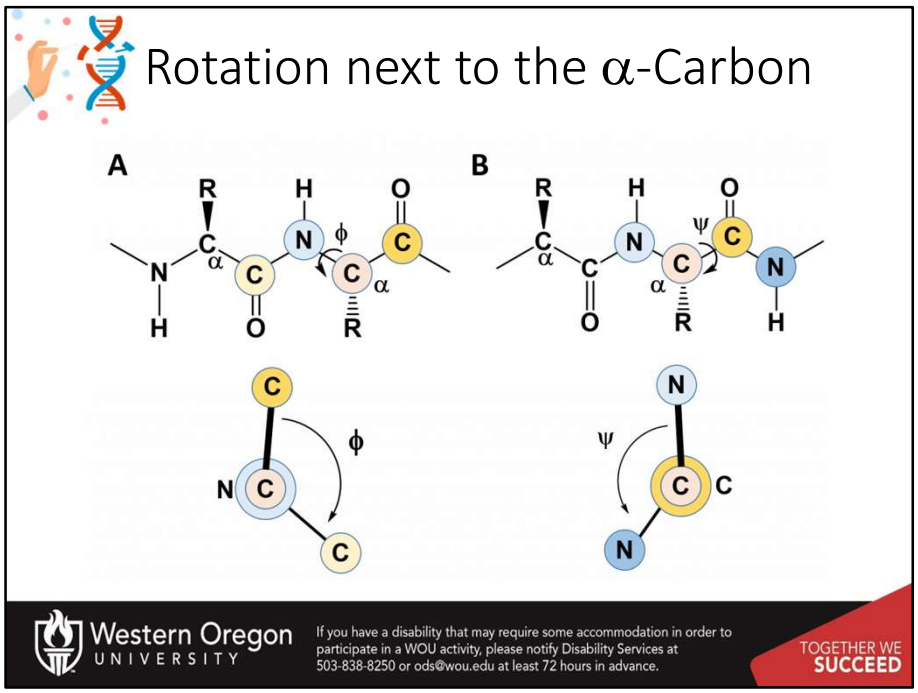
The overall primary structure of the protein is defined by the linear sequence of amino acids put together in the N- to C- orientation. Each protein will have a unique ordering of amino acids creating hundreds of thousands of possibilities. The C-terminal sequence of this protein is shown …Cys-Ser-Leu- and the last amino acid, Phe.



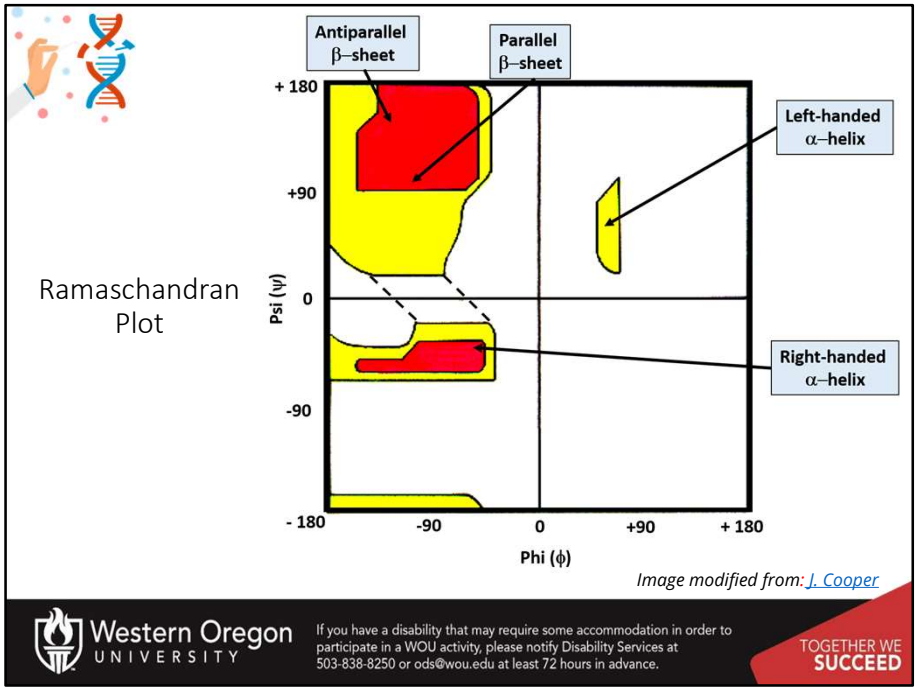
As noted in the previous talk, Cys residues are capable of being oxidized to form disulfide bridges within the protein structure. Often times these disulfide bridges are required for the correct 3-dimensional structure of the protein to be created. The insulin peptide hormone is shown in the diagram above. Within this structure the A and B peptide chains are held together by disulfide bridges. Insulin would lose its activity without these critical linkages. Note that disulfide bond formation is NOT a spontaneous process. It requires the action of an enzyme to mediate formation. Thus, not all Cys residues within a protein structure will be involved in disulfide bond formation. In the next section, we will discuss how repeating structural features arise within the protein structure to create regional secondary structural features.



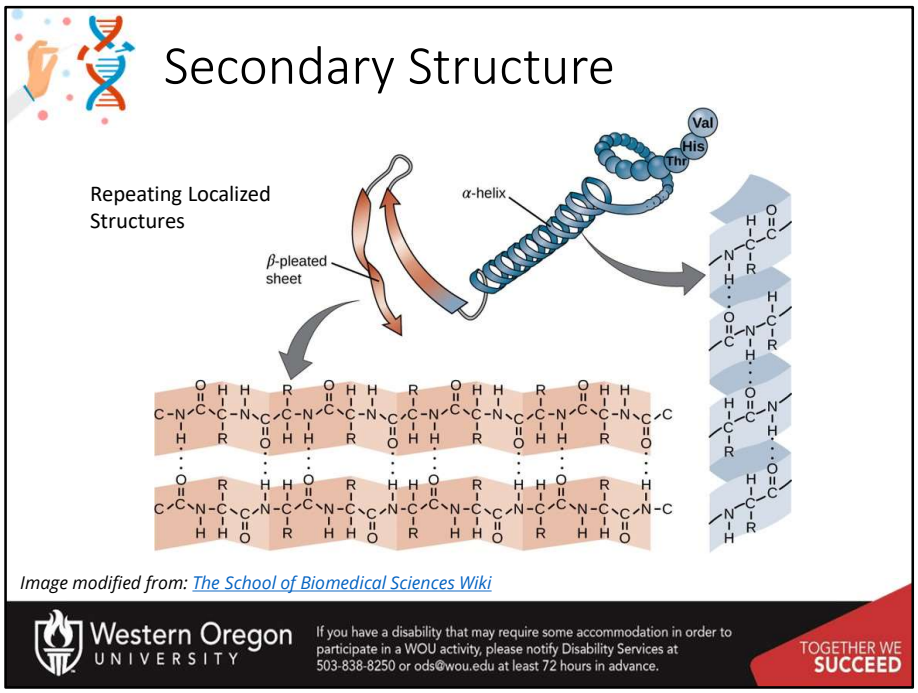
In this section, we will focus our discussions on the most common secondary structural features in proteins.



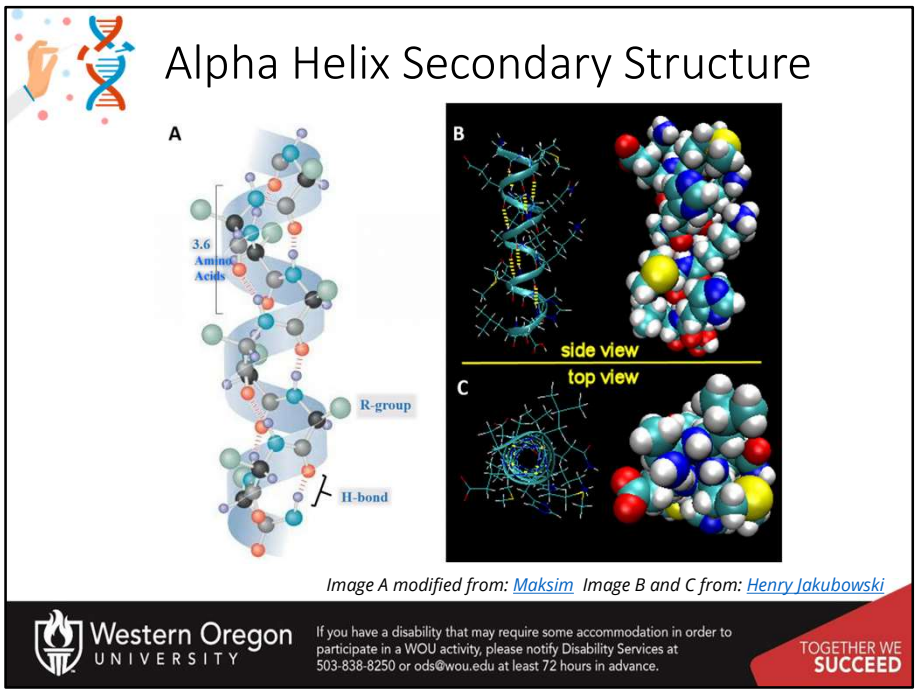
In the previous lecture, we learned that the C-N bond in the amide linkage is planar and rigid, and that the R-groups favor the *trans* confromation (except for proline which favors the *cis* conformation). This rigidity with the protein backbone limits the folding potential and patterns of the resulting protein. However, the bonds attached to the α-carbon can freely rotate and contribute to the flexibility and unique folding patterns seen within proteins. To evaluate the possible rotation patterns that can arise around the α-carbon, the torsion angles Phi (Φ) and Psi (ψ) are commonly measured. The torsion angle Phi (Φ) measures the rotation around the α-carbon – nitrogen bond by evaluating the angle between the two neighboring carbonyl carbons when you are looking directly down the α-carbon – nitrogen bond into the plane of the paper. Conversely, the torsion angle Psi (ψ) measures the rotation around the α-carbon – carbonyl carbon bond by evaluating the angle between the two neighboring nitrogen atoms when you are looking directly down the α-carbon – carbonyl carbon bond.



The Ramaschandran Plot graphs the Phi (Φ) and Psi (ψ) torsion angles together and shows the favorable bond angles shown in yellow, and the highly favorable angles shown in red. The bond angles for some of the most common secondary protein structures are indicated, and include the beta-pleated sheet and the alpha helix. We will talk about the details of these structural motifs in more detail over the next few slides.



The most favorable Phi and Psi angles give rise to repeating localized structures that form recognizable features in the protein. These include beta-pleated sheet and alpha helix. These structures are held together by hydrogen bonding within the backbone of the protein structure. The R-groups are not involved in the hydrogen bonding within the structure, but they can have an affect on structure due to steric hindrance or hydrophobic/hydrophilic characteristics within that protein region.



So let's look at the alpha helix structure in a little bit more detail. A total of 3.6 amino acids are required to form one turn of an α-helix. Hydrogen bonding between the carbonyl oxygen and the nitrogen of the 4th amino acid stabilize the helical structure. In figure (A), the black atoms are the alpha carbon, grey are carbonyl carbons, red are oxygen, blue are nitrogen, green are R-groups, and light purple are hydrogen atoms. Note that the R-groups are on the outside of the helix structure. Figure (B) shows the Expanded Side View Linear Structure and Space-Filling Model and (C) shows the Expanded Top View Linear Structure and Space-Filling Model.

Key Points of Alpha Helices

- The alpha helix is more compact than the fully extended polypeptide chain with phi/psi angles of 180°
- In proteins, the average number of amino acids in a helix is 11, which gives 3 turns.

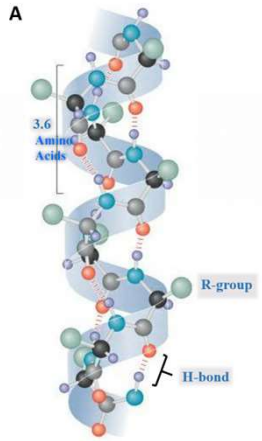


Image A modified from: Maksim

In most proteins, the average number of amino acids involved in an alpha helix is 11, which will give a total of 3 turns, although this is just the average.

Key Points of Alpha Helices

- The Left-handed Alpha Helix is Rare!

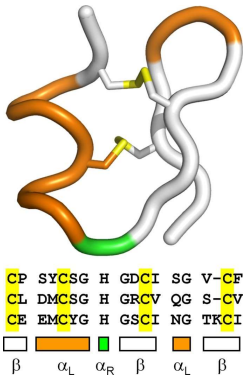


Figure from: Annavarapu, S., Nanda, V. (2009) *BMC Struct Biol* **9**, 61

The left-handed alpha helix, although allowed from inspections of a Ramachandran plot, is rarely observed, since the amino acids used to build protein structure are L-amino acids and are biased towards forming the right-handed helix. When left-handed helices do form, they are often critical for the correct protein folding, protein stability, or are directly involved in the formation of the active site. The left-handed helix shown in orange is stabilized by two disulfide bridges, shown in yellow.

Key Points of Alpha Helices

- The core of the helix is tightly packed.

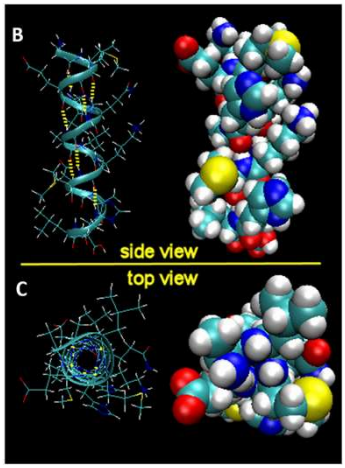
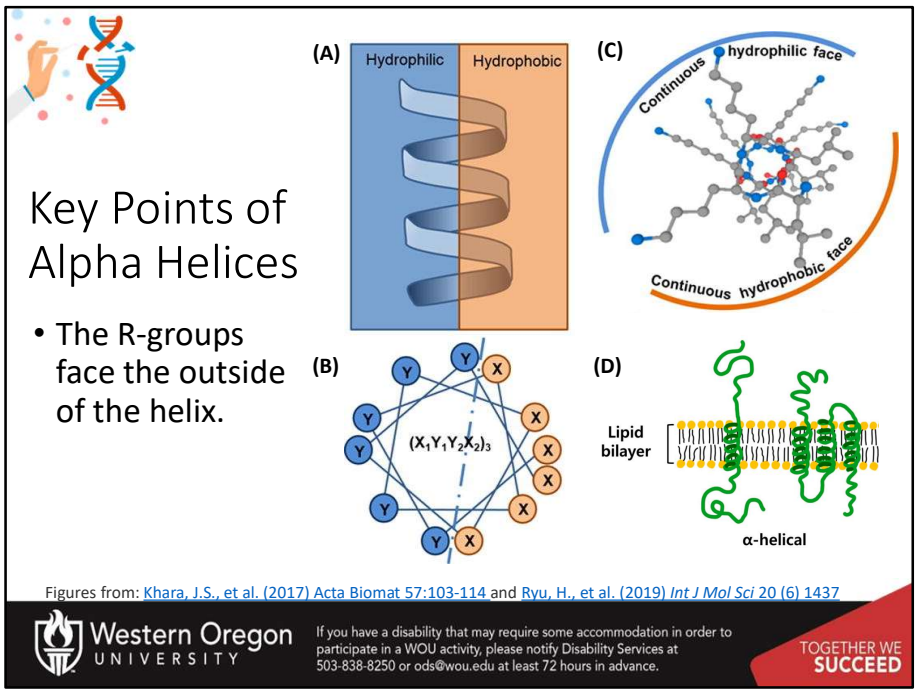
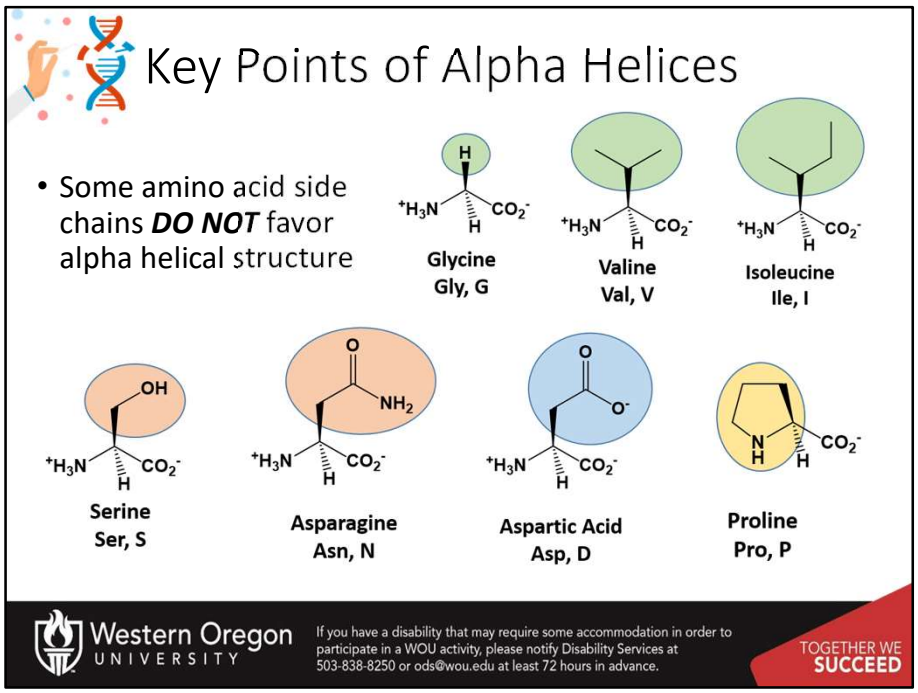


Image A modified from: Henry Jakubowski

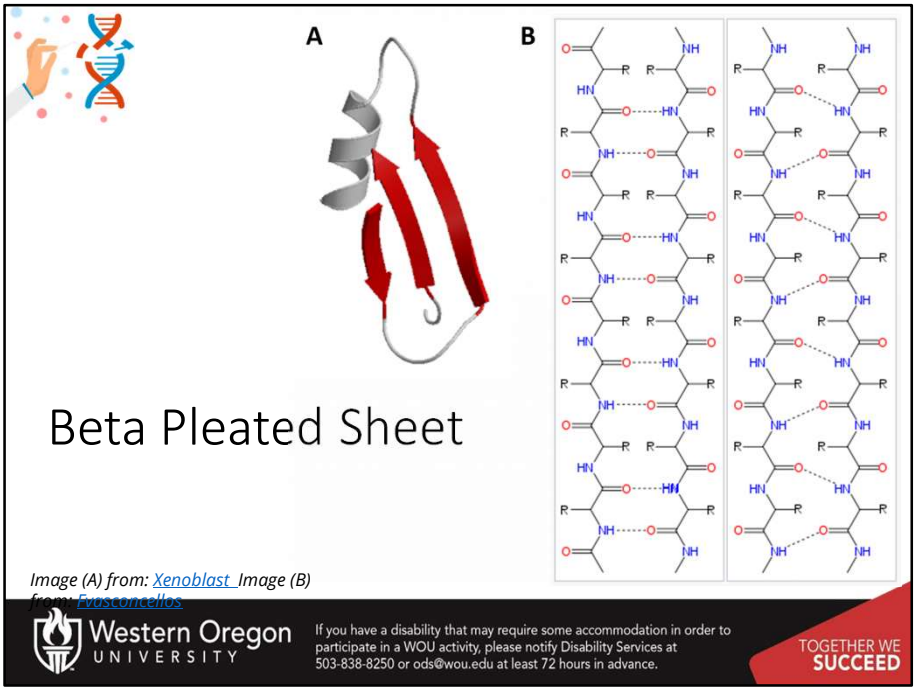
As you can see from the space-filling model, the core of the alpha helix is packed tightly. There are not holes or pores in the helix.



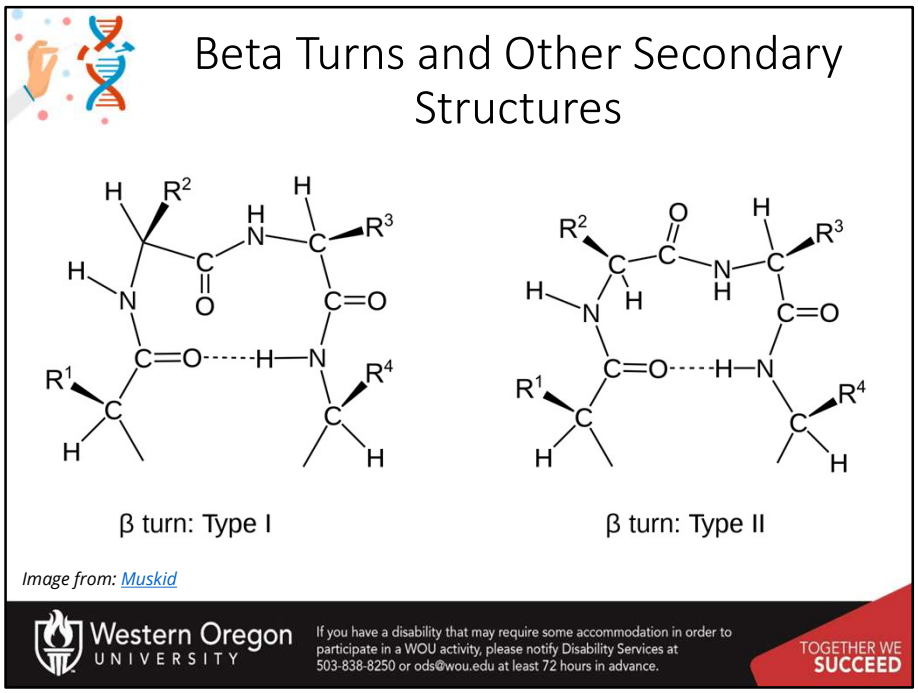
All the R-groups extend outward and away from the helix axis. The R-groups can influence the folding of proteins based on their characteristics and help to determine localized properties of the protein within that region. Hydrophilic portions of alpha helix can interact with the water environments while hydrophobic regions may fold to the inside of the protein to be shielded away from water, or may be found in the helices that span the plasma membrane. Some helices, such as the one shown, has one hydrophilic face and one hydrophobic face mediating these differential interacting surfaces. Others may be fully hydrophobic or hydrophilic or show hydrophilic and hydrophobic regions in the other orientation depending on their localization and function.



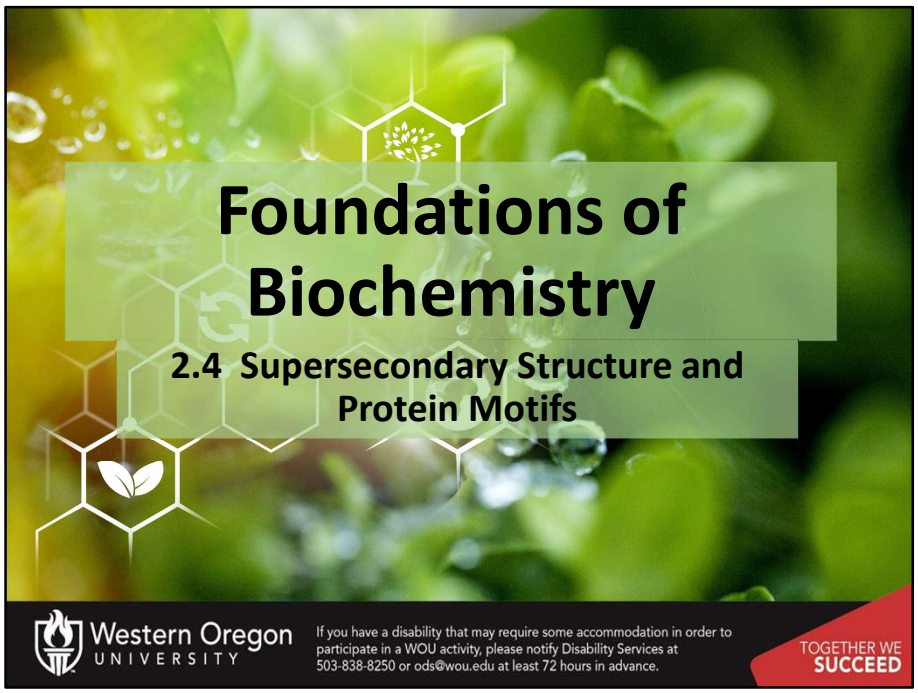
Some amino acids are more commonly found in alpha helices than others. Here are the amino acids that are typically ***NOT*** found in alpha helical structures: ***Gly*** is too small and conformationally flexible to be found with high frequency in alpha helices, while ***Pro*** is too rigid and in the *cis*-conformation. ***Pro*** often disrupts helical structure by causing bends in the protein. Some amino acids with side chains that can H-bond (***Ser, Asp,*** and ***Asn***) and aren't too long appear to act as competitors of main chain H bond donor and acceptors, and destabilize alpha helices. Early branching R-groups, such as ***Val*** and ***Ile,*** destabilize the alpha helix due to steric interactions of the bulky side chains with the helix backbone.



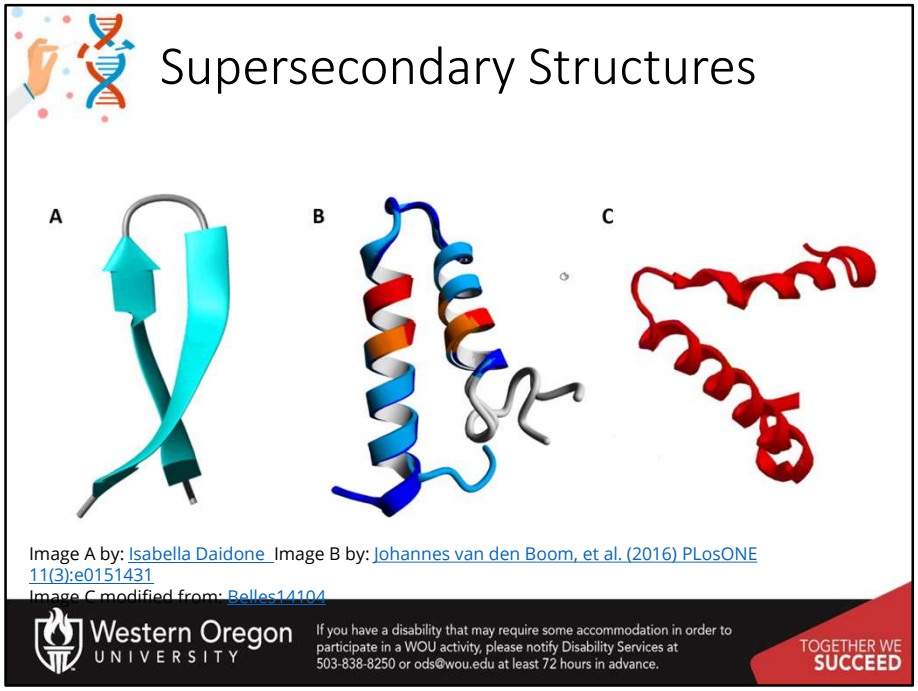
In the β-pleated sheet, the "pleats" are formed by hydrogen bonding between atoms on the backbone of the polypeptide chain. The R groups are attached to the carbons and extend above and below the folds of the pleat in the *trans* conformation. The β-pleated sheet can be oriented in the parallel or antiparallel orientation, shown in (A) above with the β-pleated sheet represented by the red ribbon arrows. The direction of the arrow indicated the orientation of the protein with the arrow running in the N- to C- direction. In parallel pleated sheets, both strands of the protein are in the N to C orientation. In antiparallel orientation one strand is in the N to C orientation and the other is flipped over heading in the opposite direction. Hydrogen bonding between the backbone carbonyl and the backbone amine functional groups stabilized both the antiparallel (B left) and the parallel (B right) β-pleated sheet structures.



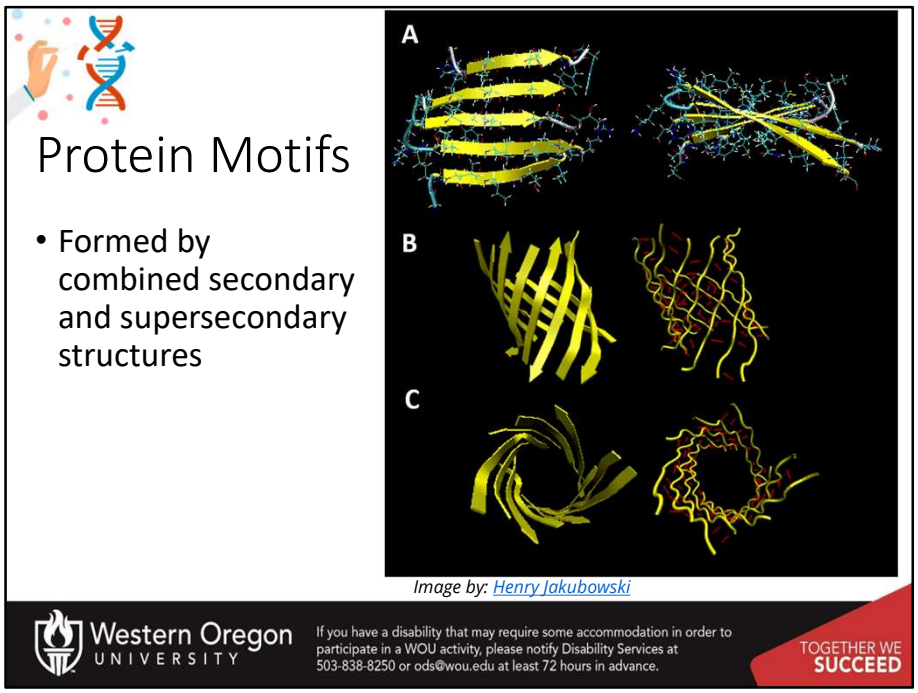
Other important secondary structures include ***turns, loops, hairpins*** and ***flexible linkers***. There are many different classifications of ***turns*** within protein structure, including ***α-turns, β-turns, γ-turns, δ-turns*** and ***π-turns***. ***β-turns*** (the most common form) typically contain four amino acid residues. Proline and Glycine are commonly found in turn motifs, as the cis conformation of Proline favors sharper conformational bends, while the minimal Glycine side chain allows for tighter packing of the amino acids to favor the turn structure. In the next section, we will see how secondary structural elements combine to form supersecondary structures and motifs.



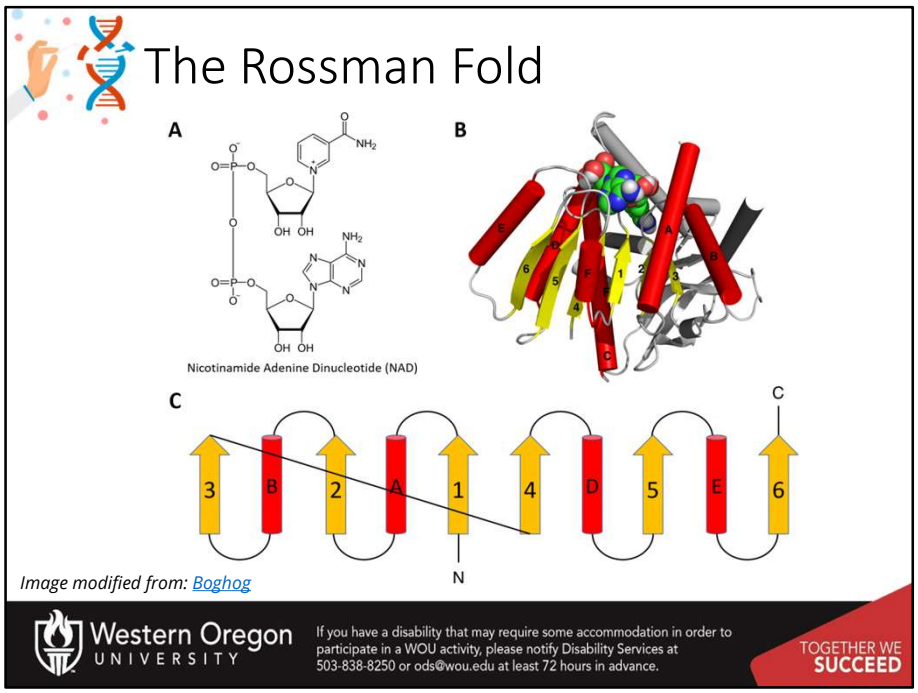
In this section, we will explore protein supersecondary structures and motifs.



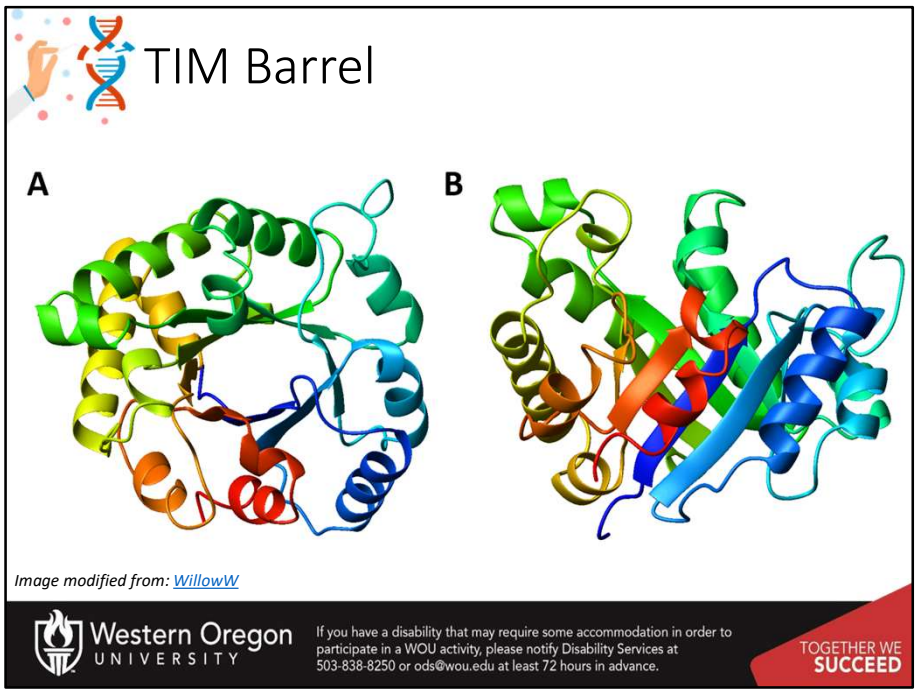
Supersecondary structure is usually composed of two secondary structures linked together by a turn. Simple supersecondary structures include (A) β-hairpin-β structures are characterized by a sharp hairpin turn that does not disrupt the hydrogen bonding of the two β-pleated sheet structures. (B) the helix-turn-helix structure and (C) the α-α corner structure present in the Myoglobin protein.



Beta strands have a tendency to twist in the right hand direction to help minimize conformational energy. This leads to the formation of interesting structural motifs found in many types of proteins. Two of these structures include ***twisted sheets*** or ***saddles*** as well as ***beta barrels***

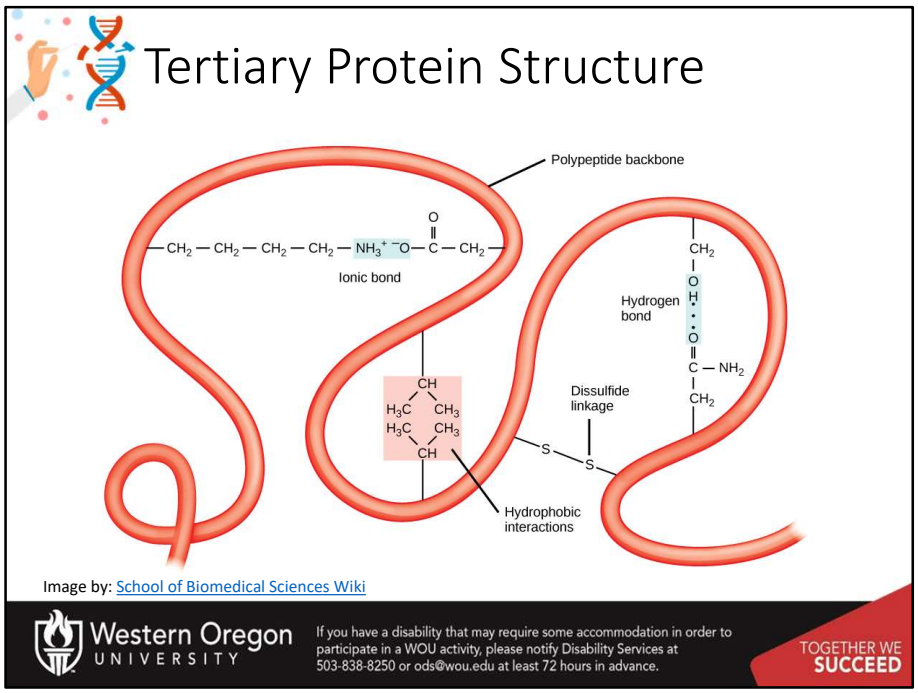


Structural motifs can serve particular functions within proteins such as enabling the binding of substrates or cofactors. For example, the ***Rossmann fold*** is responsible for binding to nucleotide cofactors such as nicotinamide adenine dinucleotide (NAD^+). The yellow arrows indicate beta pleated sheet structures and the red tubes are alpha helices.

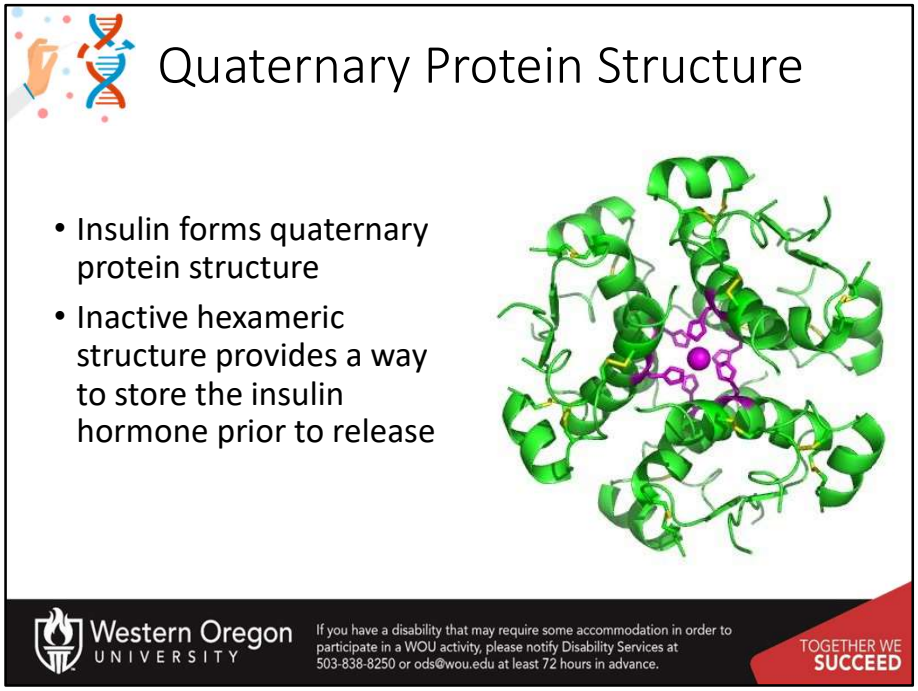


TIM barrels are considered α/β protein folds because they include an alternating pattern of α-helices and β-strands in a single domain. In a TIM barrel the helices and strands (usually 8 of each) form a solenoid that curves around to close on itself in a doughnut shape, topologically known as a toroid. TIM barrels are one of the most common protein folds. One of the most intriguing features among members of this class of proteins is although they all exhibit the same tertiary fold there is very little sequence similarity between them. At least 15 distinct enzyme families use this framework to generate the appropriate active site geometry. This is an example of ***convergent evolution***, where unrelated sequences converge on the same structural motif. In the next section, we will see how secondary and supersecondary structures come together to form tertiary and potentially quaternary protein structure.

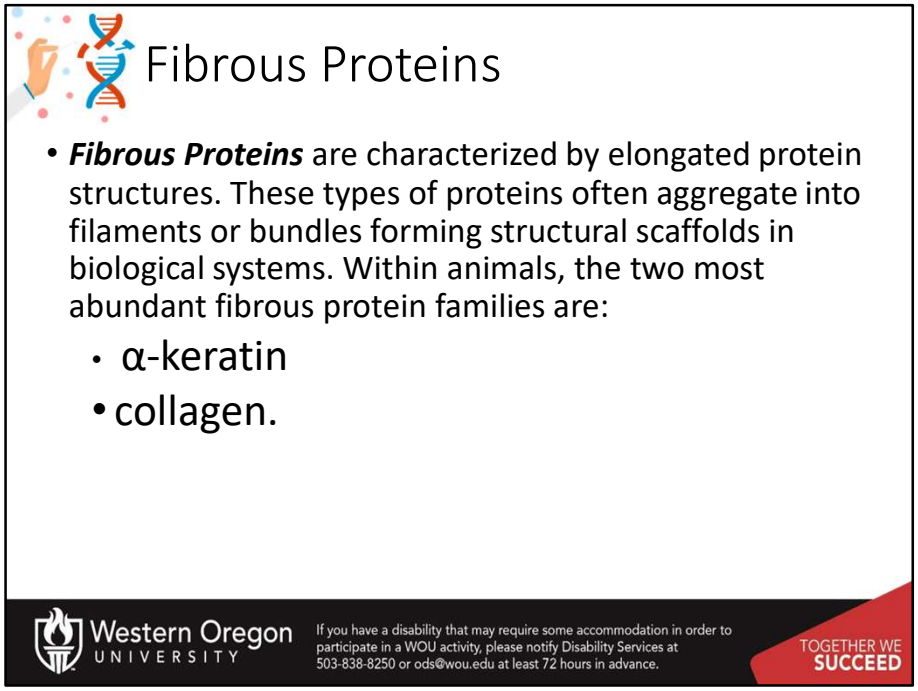
In this section, we will focus on tertiary and quaternary protein structures. In this section, we will discuss some examples of proteins that have different functions within the body, including cell communication, structural support, and cellular transport.



The complete 3-dimensional shape of the entire protein (or sum of all the secondary structural motifs) is known as the ***tertiary structure*** of the protein and is a unique and defining feature for that protein. The ***tertiary structure*** of proteins is determined by a variety of chemical interactions. These include a combination of all the different intermolecular forces from hydrophobic interactions, ionic bonding, hydrogen bonding, and dipole-dipole interactions. It also includes the covalent disulfide linkages that form between two Cys residues.

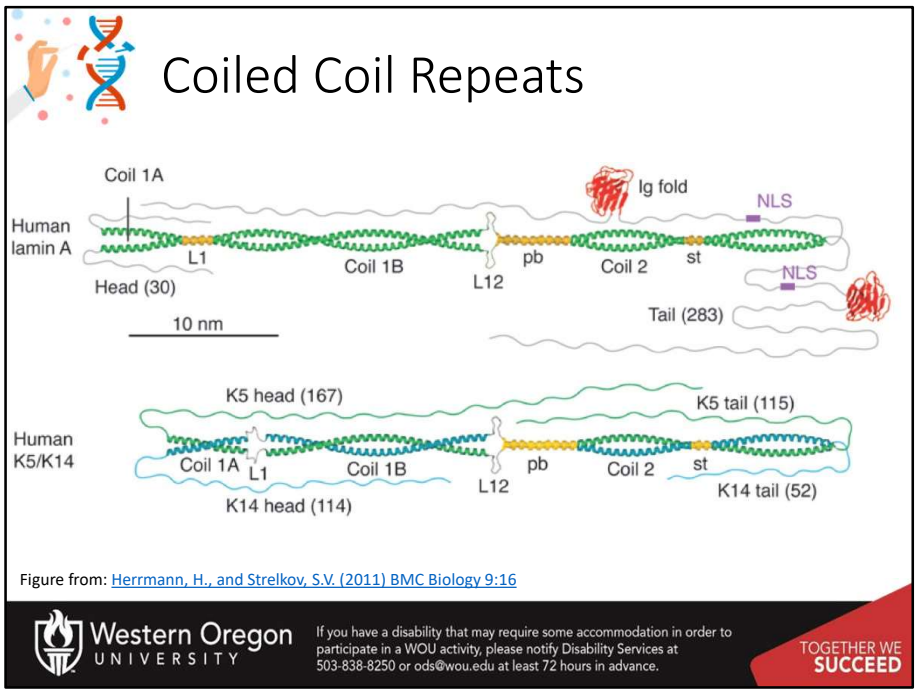


In nature, some proteins are formed from several polypeptides, also known as subunits, and the interaction of these subunits forms the ***quaternary structure.*** Weak interactions between the subunits help to stabilize the overall structure. The ***insulin peptide,*** shown here, has quaternary structure. Insulin starts out as a single polypeptide and loses some internal sequences during cellular processing that form two chains held together by disulfide linkages as shown. Three of these structures are then grouped further forming an inactive hexamer. The hexamer form of insulin is a way for the body to store insulin in a stable and inactive conformation so that it is available for release and reactivation in the dimer form (A and B chains)



Proteins can be classified into different groups based on their major structural features. The first class that we will look at are the fibrous proteins. They are characterized by elongated structures that often are linked together to form durable filament structures that provide support and scaffolds within the body. The two most abundant proteins in animals are the fibrous proteins known as the keratins and the collagens. We will take a deeper look at both of these protein families.

α-keratin is the key structural element making up hair, nails, horns, claws, hooves, and the outer layer of skin. Due to its tightly wound structure, it can function as one of the strongest biological materials and has various uses in mammals, from predatory claws to hair for warmth. α-keratin is synthesized through regular protein biosynthesis, utilizing transcription and translation, but as the cell matures and becomes full of α-keratin, it dies, creating a strong non-vascular unit of keratinized tissue.



Keratin proteins are long and fibrous in nature. There are two distinct but homologous keratin families which are named as Type I keratin and Type II keratins. There are 54 keratin genes in humans, 28 of which code for type I, and 26 for type II. Type I proteins are acidic, meaning they contain more acidic amino acids, such as aspartic acid, while type II proteins are basic, meaning they contain more basic amino acids, such as lysine. This differentiation is especially important in α-keratins because in the synthesis of its sub-unit dimer, the ***coiled coil***, one protein coil must be type I, while the other must be type II. They form strong ionic interactions to create the ***coiled coil structure.***

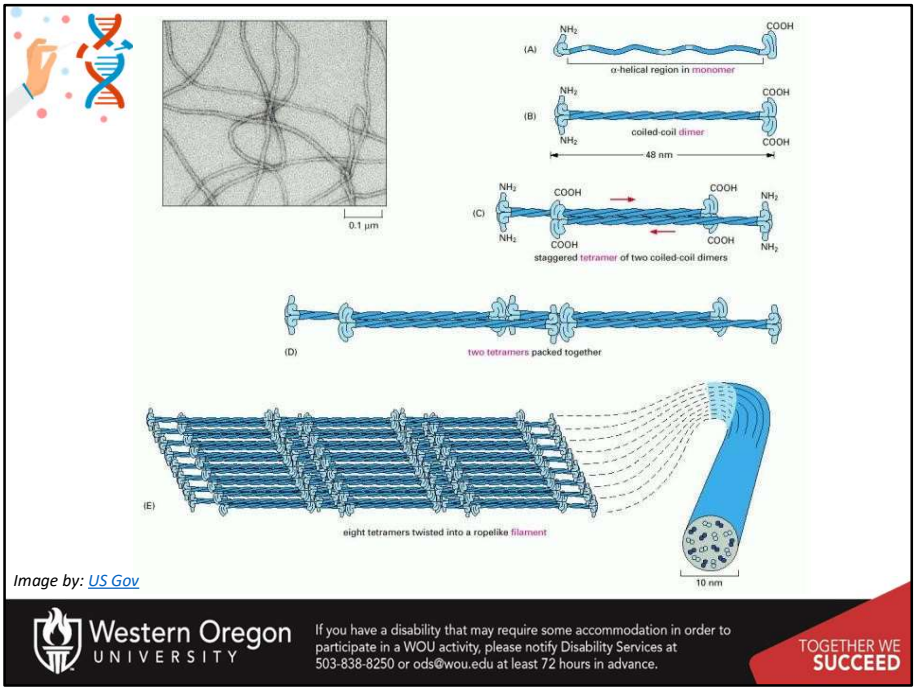
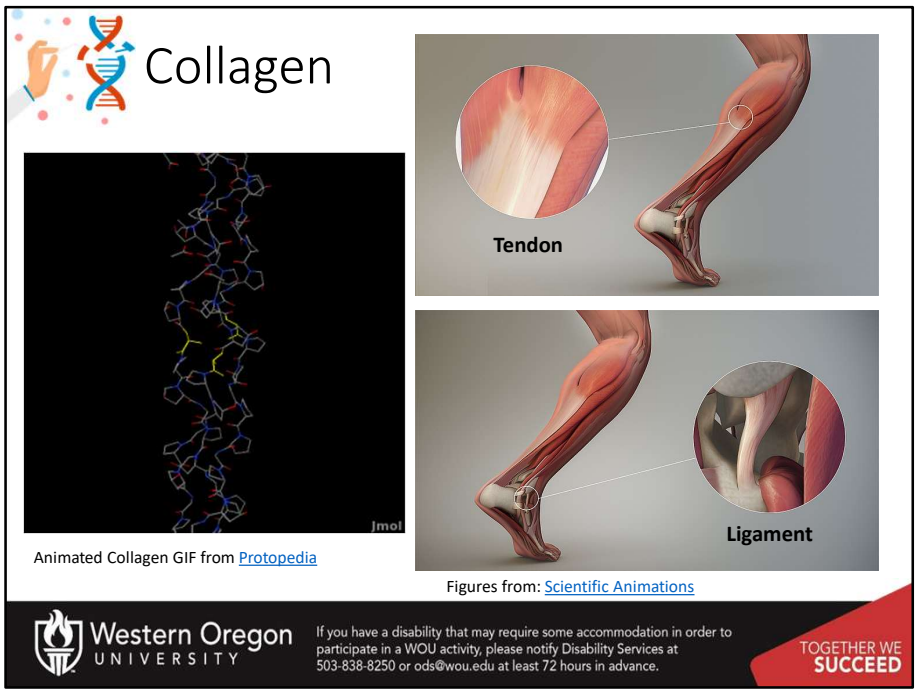
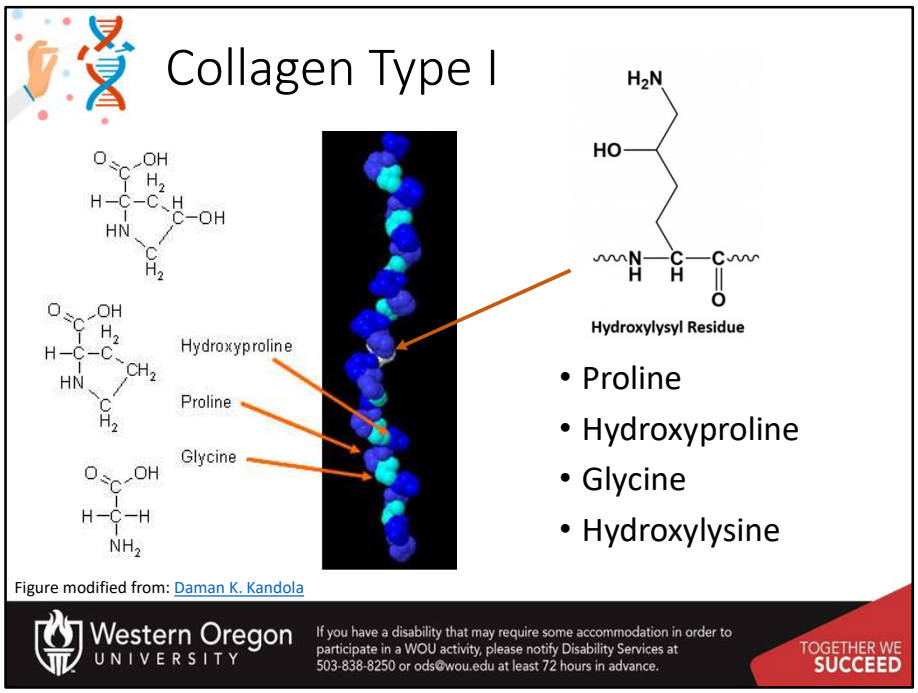


Image by: US Gov

Coiled-coil dimers then assemble into overlapping protofilaments or ***intermediate filaments,*** in a very stable, left-handed superhelical motif which further multimerises, forming filaments consisting of multiple copies of the keratin monomers. The major force that keeps the multiple coiled-coil structures associated with one another in the intermediate filament are hydrophobic interactions between nonpolar residues along the keratins helical segments.



The fibrous protein, **Collagen** is the most abundant protein in mammals, making 25% to 35% of the whole-body protein content. It is found predominantly in the extracellular space within various connective tissues in the body. Collagen contains a unique quaternary structure of three protein strands wound together to form a ***triple helix.*** It is mostly found in fibrous tissues such as tendons, ligaments, and skin.

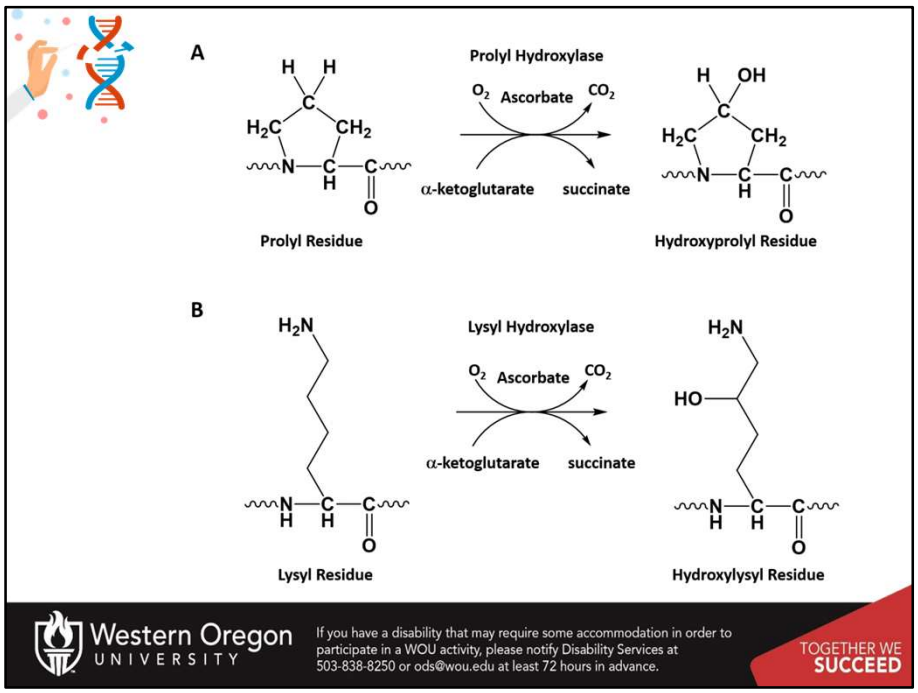


Collagen Type I has an unusual amino acid composition and sequence:

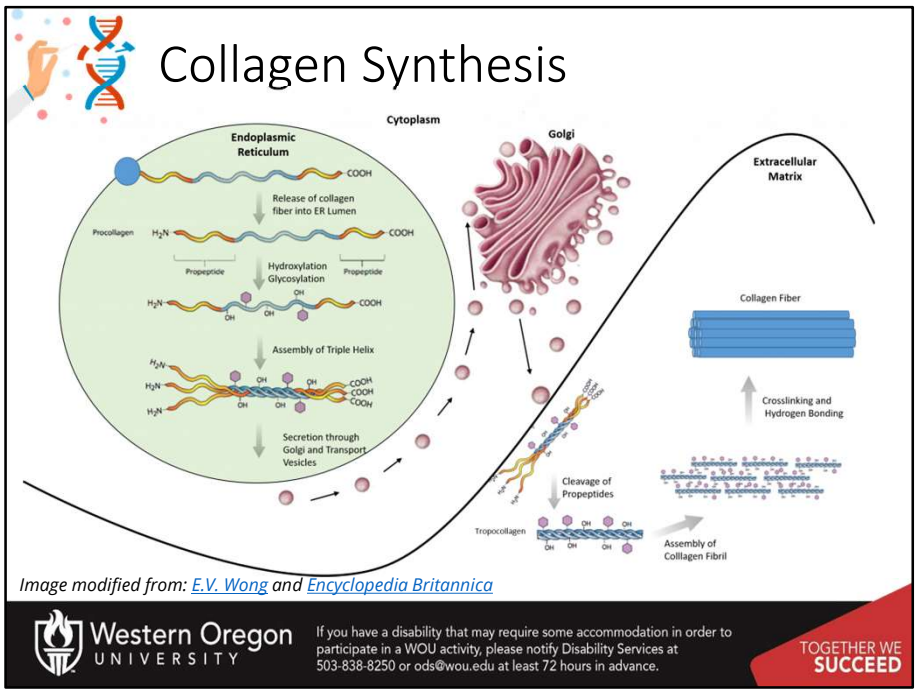
- Glycine is found at almost every third residue.
- Proline makes up about 17% of collagen.
- Collagen contains two uncommon derivative amino acids not directly inserted during translation. These amino acids are found at specific locations relative to glycine and are modified post-translationally by different enzymes, both of which require vitamin C as a cofactor.

    Hydroxyproline derived from proline

    Hydroxylysine derived from lysine – depending on the type of collagen, varying numbers of hydroxylysines are glycosylated (mostly having disaccharides attached).



The enzymes ***prolyl hydroxylase*** and ***lysyl hydroxylase*** are required for the hydroxylation of proline (A) and lysine (B) residues, respectively. (Note: While position 3 is shown above, prolyl residues may alternatively be hydroxylated at the 4-position). The hydroxylase enzymes modify amino acid residues after they have been incorporated into the protein as a post-translational modification and require vitamin C (ascorbate) as a cofactor. (C) Further modification of the hydroxylysine residues by glycosylation can lead to the incorporation of the disaccharide (galactose-glucose) at the hydroxy oxygen.



This slide provides an overview of collagen synthesis, which takes place predominantly in the rough endoplasmic reticulum, shown in light green here.

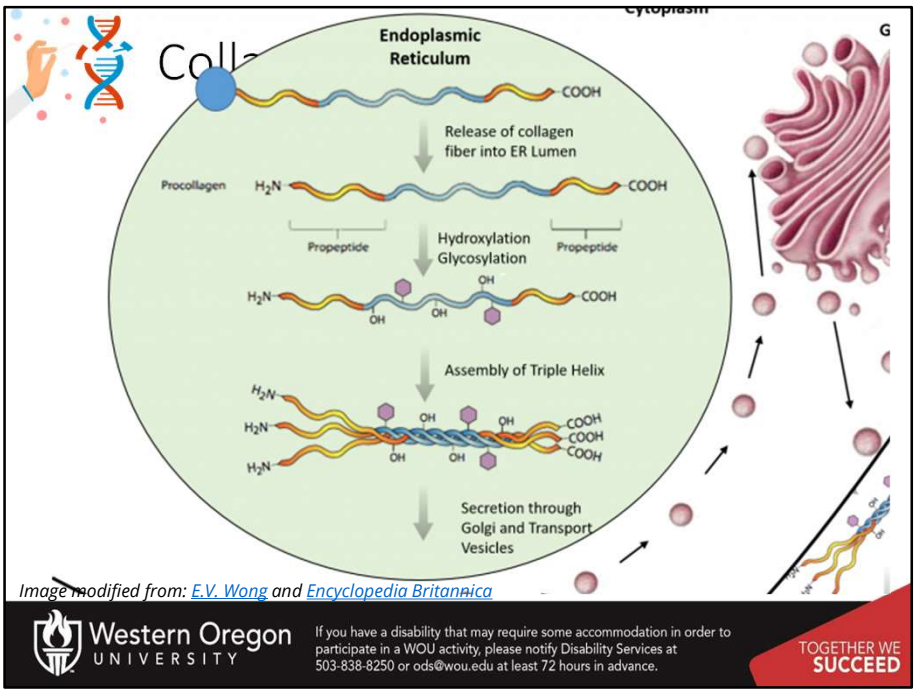
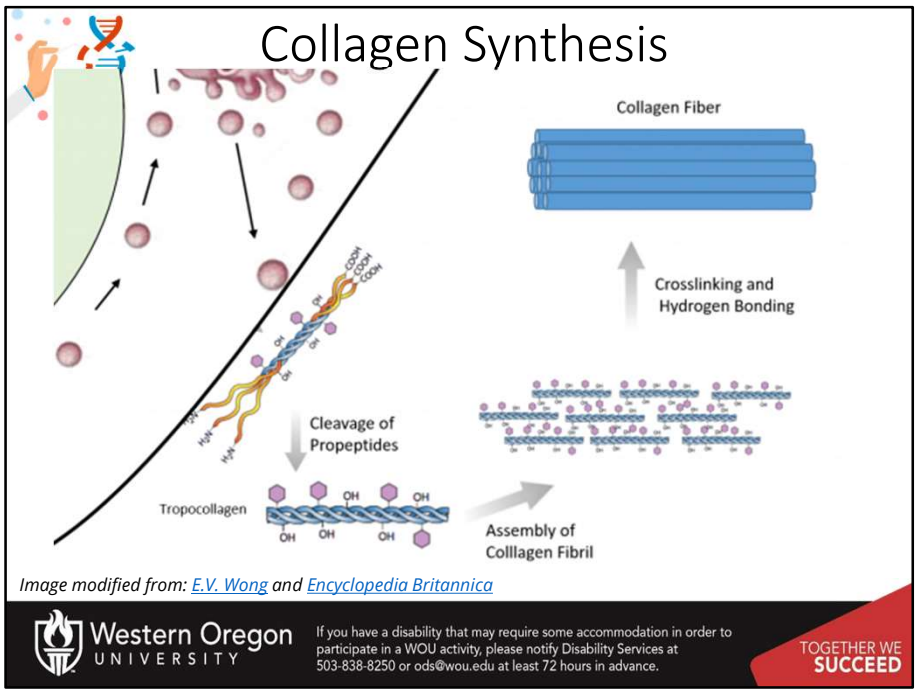
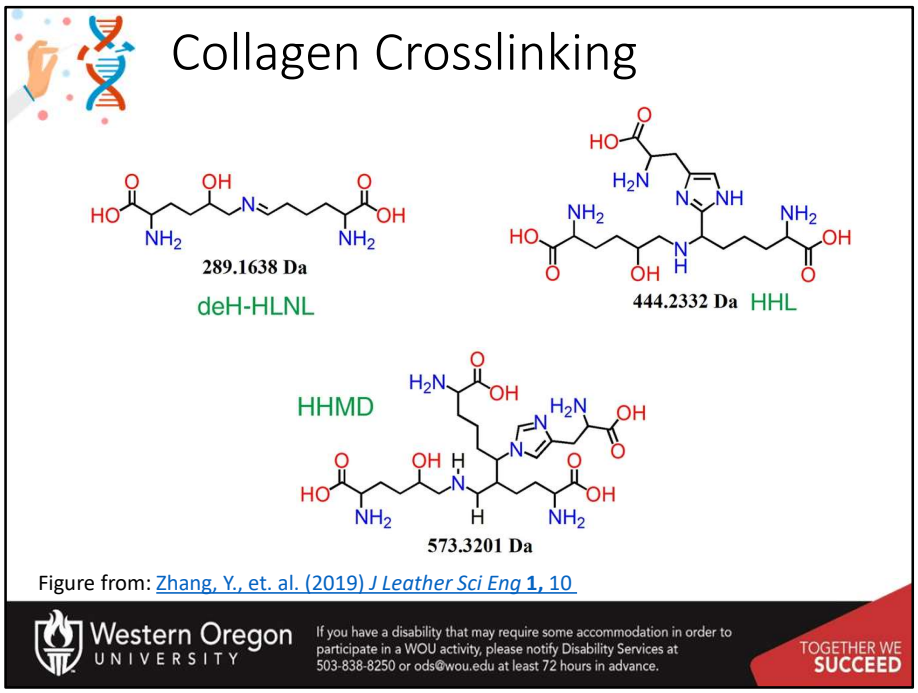


Image modified from: E.V. Wong and Encyclopedia Britannica

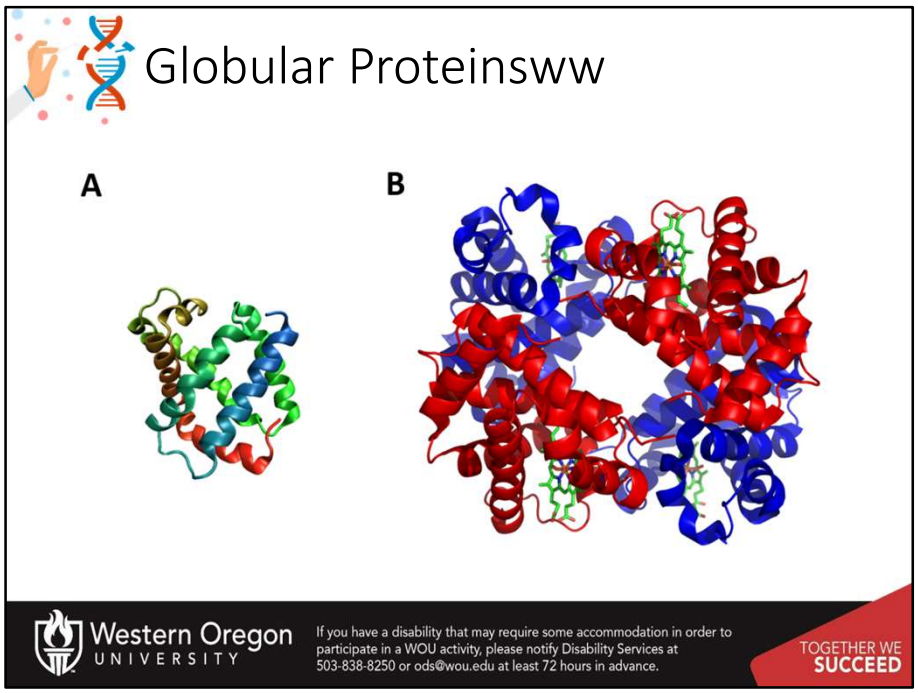
Taking a closer look, we can see that the collagen mRNA is translated on ribosomes associated with the ER and released as procollagen into the ER lumen. Hydroxylation of Pro and Lys residues occurs, followed by glycosylation (or the addition of sugar residues to the structure). The collagen monomers are then assembled into the triple helix.



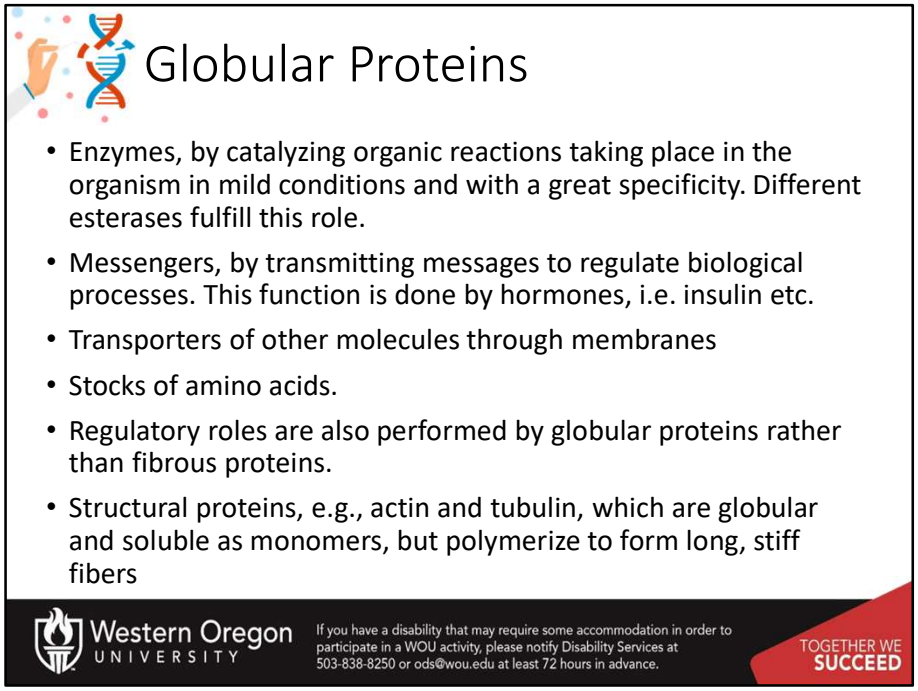
The procollagen is then packaged into secretory vesicles, transported through the golgi apparatus (where more post-translational modifications can occur) and then secreted outside of the cell into the extracellular matrix (ECM). Once in the ECM, procollagen is cleaved to form tropocollagen and then assembled into a collagen fibril. You can see that these are overlapping, strong fibrils that are held together predominantly by covalent crosslinking and hydrogen bonding.



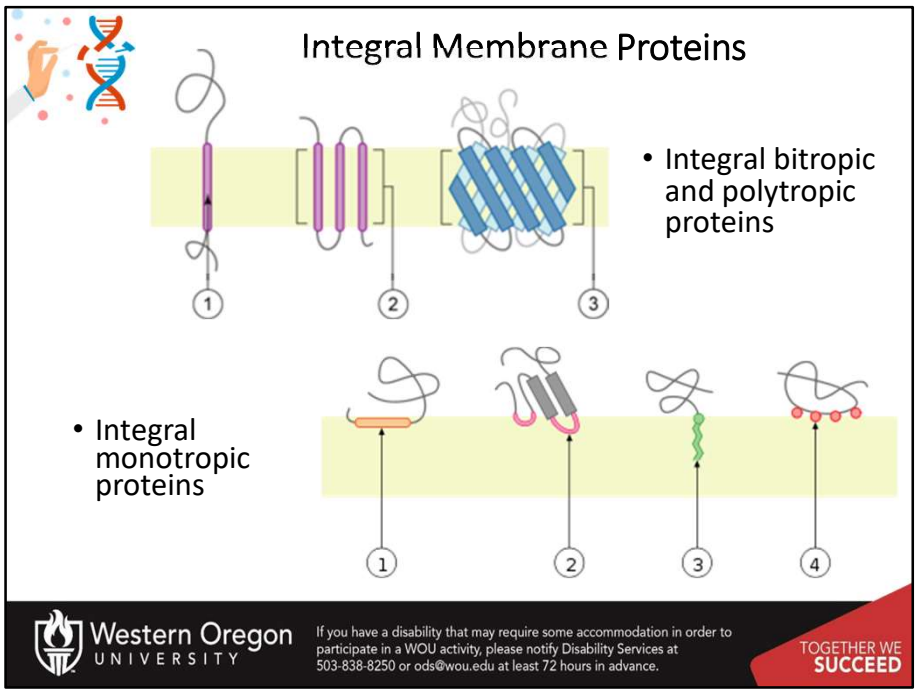
Several types of naturally-occurring covalent crosslinks have been identified in different collagenous materials, and include the examples shown here: dehydrohydroxylysinonorleucine (deH-HLNL), histidinohydroxylysinonorleucine (HHL), and histidinohydroxymerodesmosine (HHMD). The process is quite complicated and can take an extended period of time for a collagen fibril to completely mature.



Thus far, we have been focusing primarily on fibrous protein structures. ***Globular proteins*** or ***spheroproteins*** are spherical ("globe-like") proteins and are one of the common protein types. Globular proteins are somewhat water-soluble (forming colloids in water), unlike the fibrous or membrane proteins. There are multiple fold classes of globular proteins, since there are many different architectures that can fold into a roughly spherical shape. The term ***globin*** can refer more specifically to proteins including the globin fold. The ***globin fold*** is a common three-dimensional fold in proteins and defines the globin-like protein superfamily An example of the globin fold, the oxygen-carrying protein myoglobin (PBD ID 1MBA) from the mollusc *Aplysia limacina.* (B) Structure of the tetrameric hemoglobin protein containing a total of four globin folds. We will talk about the hemoglobin protein in more detail in a later lecture.

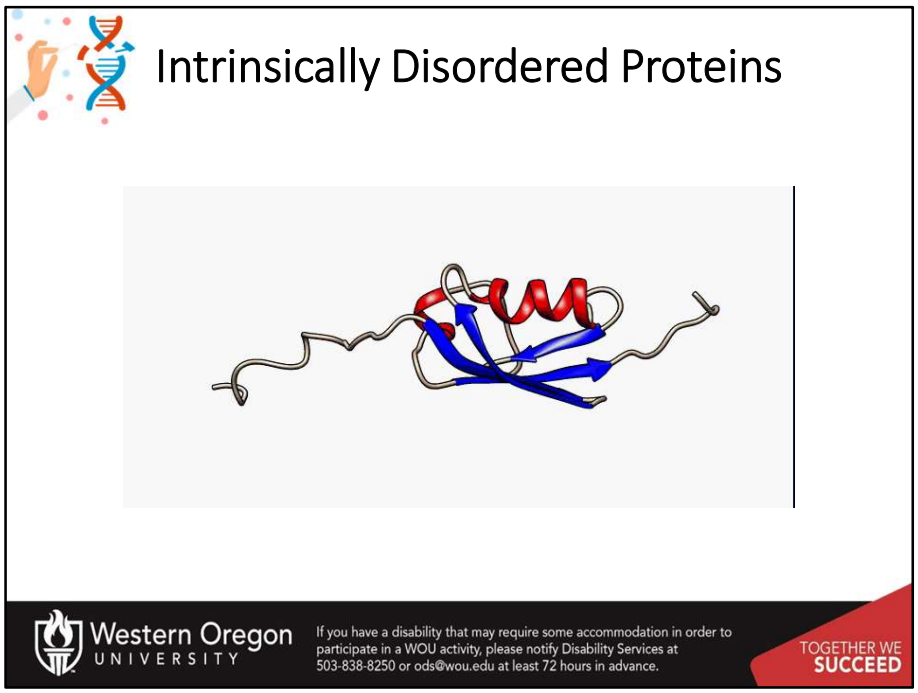


Overall, globular proteins are involved in many cellular processes and can serve as enzymes, cellular messengers, transporters, regulatory proteins, and structural support proteins. Throughout the term, we will become more familiar globular protein structure and activity in a multitude of biological processes.

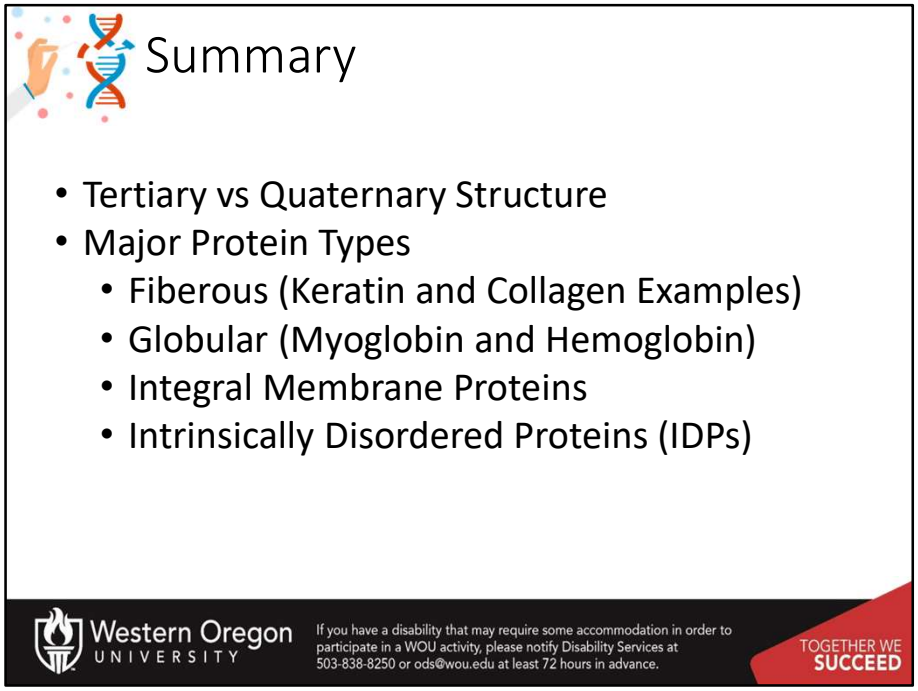


Another major type of proteins are ***Integral membrane proteins. Integral membrane proteins*** are permanently attached to the membrane. Such proteins can be separated from the biological membranes only using detergents, nonpolar solvents, or sometimes denaturing agents. They can be classified according to their relationship with the bilayer. ***Integral polytopic proteins*** are transmembrane proteins that span across the membrane more than once. These proteins may have different transmembrane topology. These proteins have one of two structural architectures: ***helix bundle proteins (2 in top diagram)***, which are present in all types of biological membranes; ***beta barrel proteins (3 in top diagram)***, which are found only in outer membranes of Gram-negative bacteria, and outer membranes of mitochondria and chloroplasts. ***Bitopic proteins (1 in top diagram)*** are transmembrane proteins that span across the membrane only once. Transmembrane helices from these proteins have significantly different amino acid distributions to transmembrane helices from polytopic proteins. ***Integral monotopic proteins (lower diagram)*** are integral membrane proteins that are attached to only one side of the membrane and do not span the whole way across. 1 in lower diagram shows the interaction by an amphipathic α-helix parallel to the membrane plane (in-plane membrane helix) 2 in the lower diagram shows the interaction by a hydrophobic loop 3 in the lower diagram shows the interaction by a covalently bound membrane lipid (*lipidation*) 4 shows the electrostatic or ionic interactions of the protein with membrane lipids. Peripheral membrane proteins (which are not shown here) are temporarily attached either to the lipid

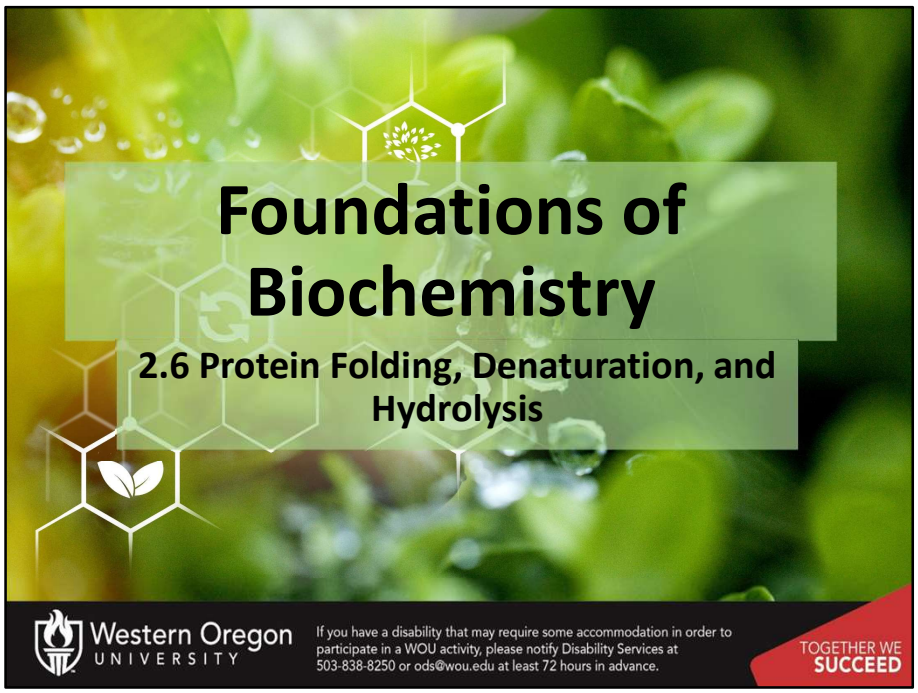
bilayer or to integral proteins by a combination of hydrophobic, electrostatic, and other non-covalent interactions. Peripheral proteins dissociate following treatment with a polar reagent, such as a solution with an elevated pH or high salt concentrations.



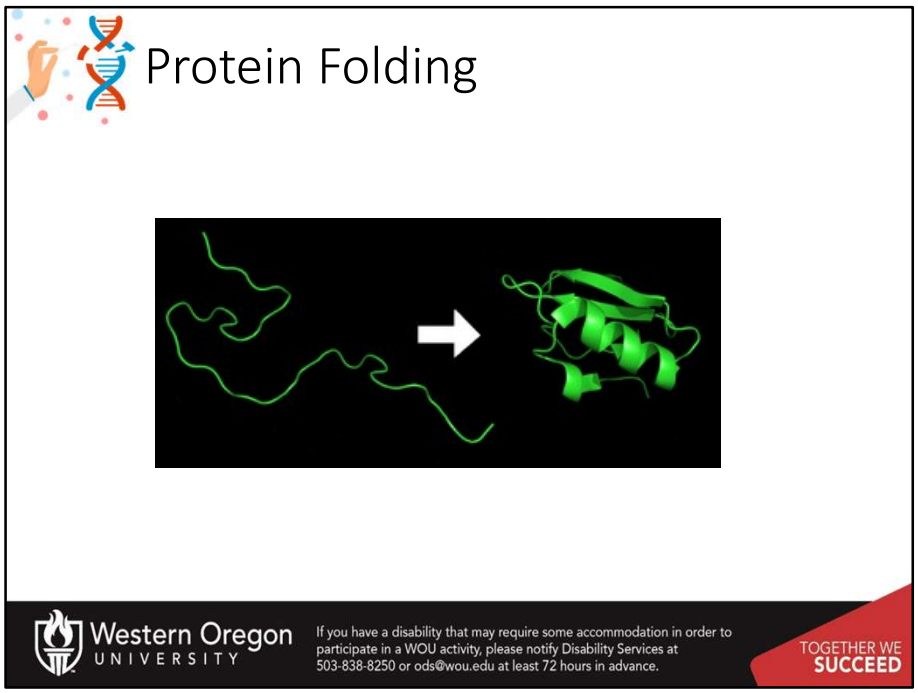
More recently, researchers have been characterizing ***intrinsically disordered proteins or IDPs***. An ***intrinsically disordered protein*** ***(IDP)*** is a protein that lacks a fixed or ordered three-dimensional structure. IDPs cover a spectrum of states from fully unstructured to partially structured and include random coils, (pre-)molten globules, and large multi-domain proteins connected by flexible linkers. They constitute one of the main types of protein (alongside globular, fibrous and membrane proteins). Many disordered proteins have the binding affinity with their receptors regulated by post-translational modification, thus it has been proposed that the flexibility of disordered proteins facilitates the different conformational requirements for binding the modifying enzymes as well as their receptors. Intrinsic disorder is particularly enriched in proteins implicated in cell signaling, transcription and chromatin remodeling functions.



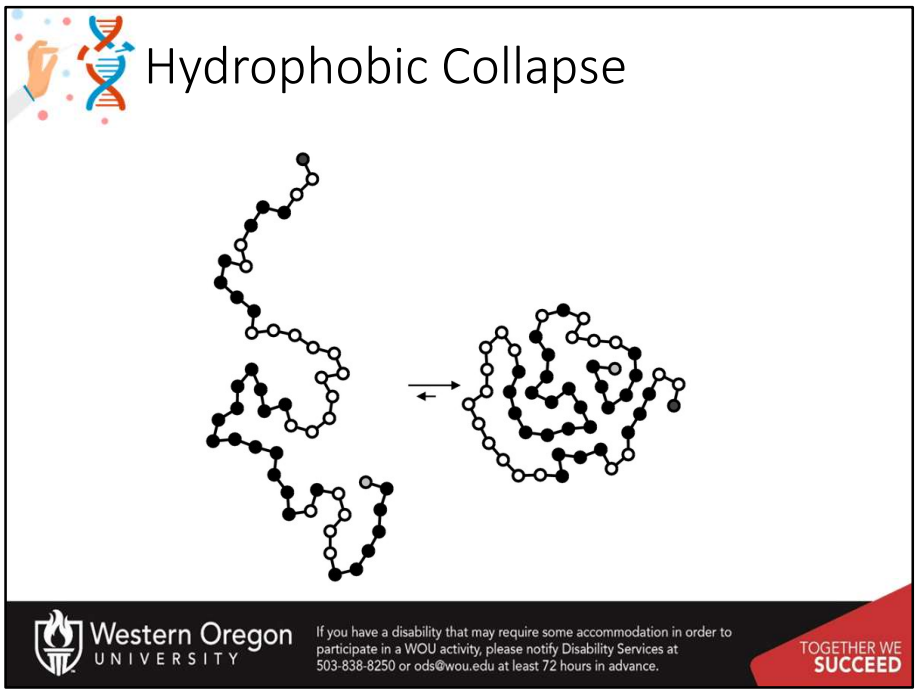
In this section you have learned about differences and complexities between tertiary and quaternary protein structures and the major protein structural types. We spent a lot of time introducing two of the major fibrous proteins in the body, keratin and collagen, as we will not really revisit this protein type through the rest of the term. We only briefly mentioned globular proteins and IDPs, as most of the proteins that we will focus on during the rest of the term will fall into these two classifications. We also introduced the different types of integral membrane proteins, which will play a bigger role during next terms lectures CH451.



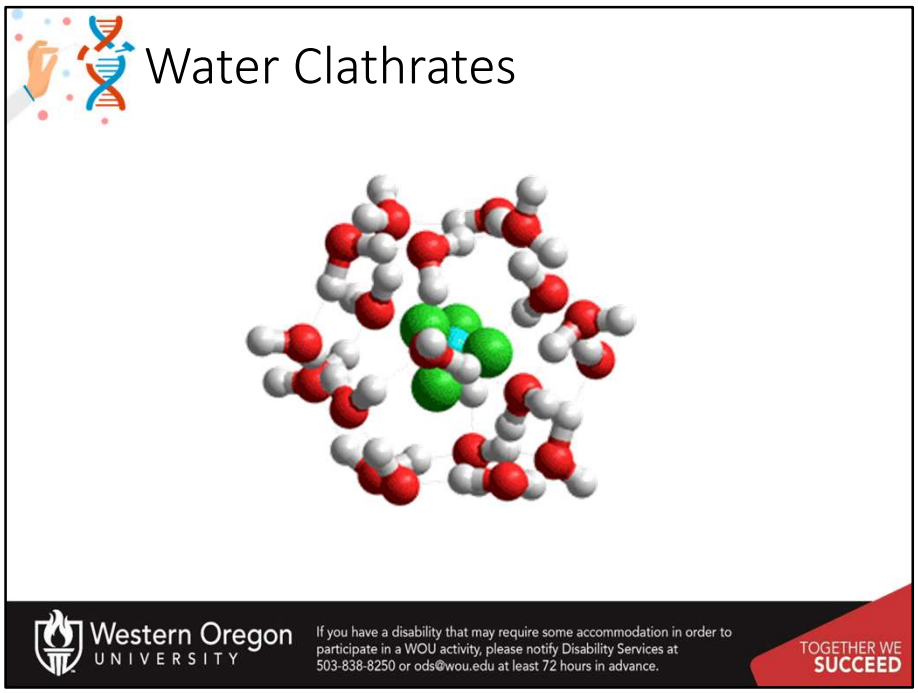
Our final section in chapter 2 deals with protein folding, denaturation and hydrolysis.



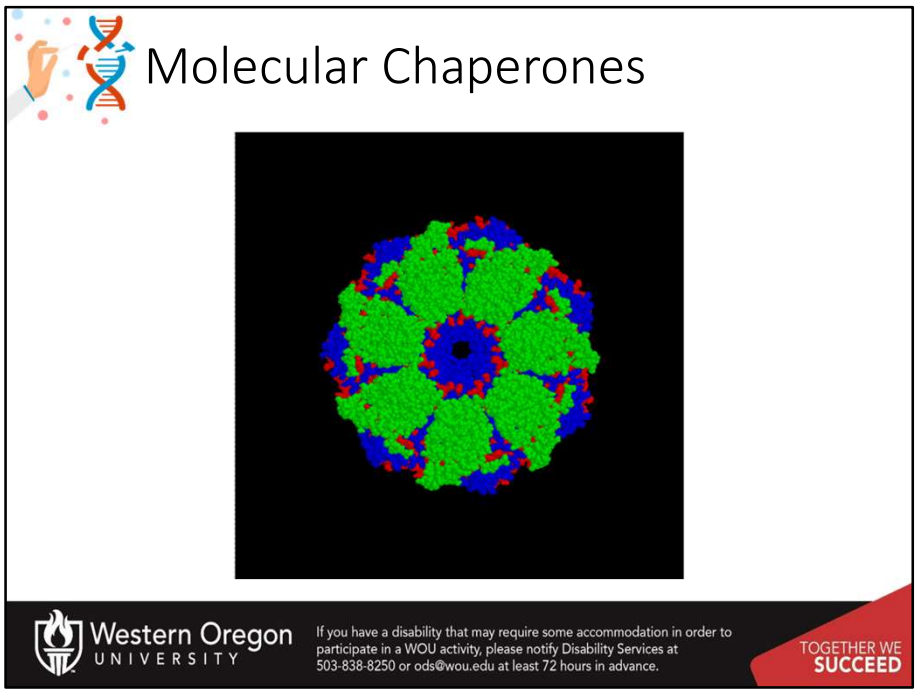
Correct protein folding is critical for the correct function of the protein, and is very difficult to predict! We will discuss some of the models thought to induce protein folding in vivo.



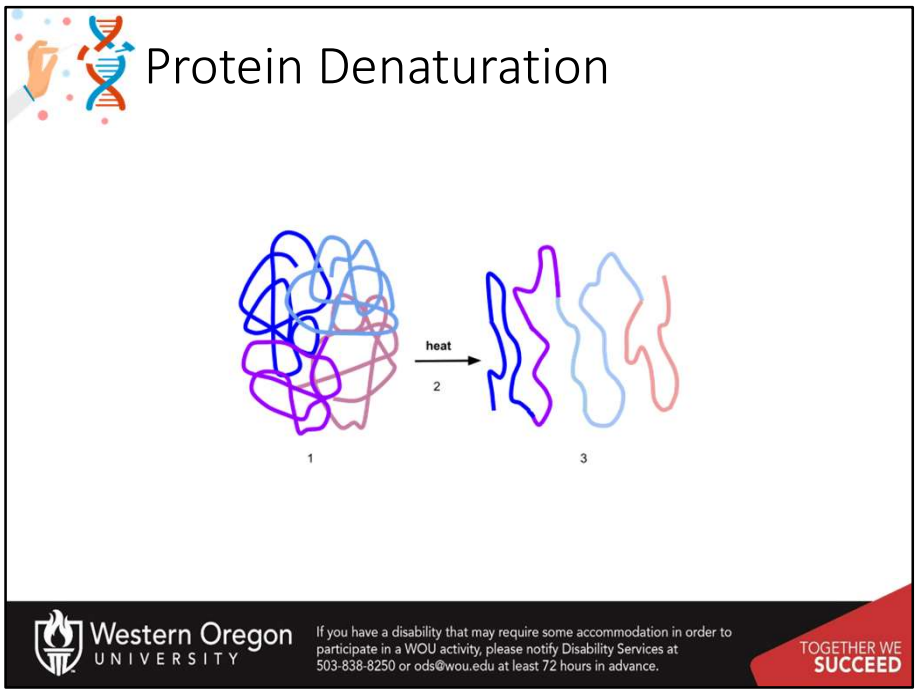
Hydrophobic collapse is thought to help minimize the number of hydrophobic side-chains that are exposed to water and is an important driving force behind the folding process.



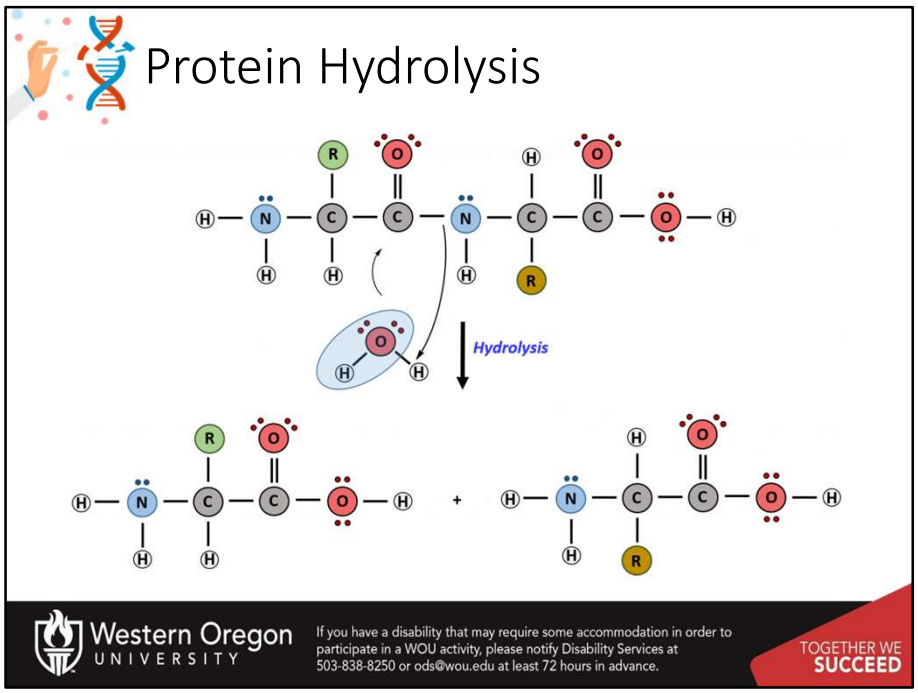
In an aqueous environment, the water molecules tend to aggregate around the hydrophobic regions or side chains of the protein, creating water shells of ordered water molecules. An ordering of water molecules around a hydrophobic region increases order in a system and therefore contributes a negative change in entropy (less entropy in the system). The water molecules are fixed in these water cages which drives the hydrophobic collapse, or the inward folding of the hydrophobic groups. The example above shows chloroform, a hydrophobic compound, with a surrounding water clathrate. Essentially when choroform is dissolved in water forming a hydrate, the hydrophobic hydration is accompanied by a negative entropy change due to the increased order in the surrounding water and the positive heat capacity change, often causing a positive ΔG. Similar water cages can associate around hydrophobic protein residues prior to correct folding. The folding inward of hydrophobic groups releases the water clathrate structure and increases the free movement of water, leading to a negative ΔG.



Molecular chaperones are a class of proteins that aid in the correct folding of other proteins *in vivo*. Chaperones exist in all cellular compartments and interact with the polypeptide chain in order to allow the native three-dimensional conformation of the protein to form. They operate by binding to and stabilizing an otherwise unstable structure of a protein during its folding pathway. Chaperones do not contain the necessary information to know the correct native structure of the protein they are aiding; rather, chaperones work by preventing incorrect folding conformations. Along with their role in aiding native structure formation, chaperones are shown to be involved in various roles such as protein transport, degradation, and even allow denatured proteins exposed to certain external denaturant factors an opportunity to refold into their correct native structures.



So what exactly is protein denaturation. Fully denatured proteins lack both tertiary and secondary structure, however, their primary protein sequence remains intact and the protein exists as a random coil. In some cases the protein may be able to refold, but often times denaturation is irreversible. The external factors involved in protein denaturation or disruption of the native state include temperature, external fields (electric, magnetic), molecular crowding, and even the limitation of space, which can have a big influence on the folding of proteins. High concentrations of solutes, extremes of pH, mechanical forces, and the presence of chemical denaturants can contribute to protein denaturation, as well. These individual factors are categorized together as stresses. Chaperones are shown to exist in increasing concentrations during times of cellular stress and help the proper folding of emerging proteins as well as denatured or misfolded ones. We will explore some of the chemical denaturants used in analyzing protein structures in Chapter 3.



Hydrolysis is the breakdown of the primary protein sequence by the addition of water to reform the individual amino acids monomer units. In the hydrolysis reaction, water is added across the amide bond incorporating the -OH group with the carbonyl carbon and reforming the carboxylic acid. The hydrogen from the water reforms the amine. Under standard biological pH, the hydrolized amino acids will take their zwitter ionic form, with the carboxylic acid in the deprotonated state and the amine in the protonated state. This concludes chapter 2. In the next chapter, we will explore laboratory methodologies used to characterize and study protein structure and function.