

**Chemistry 5.07UE 'Biological Chemistry I
Fall Semester, 2013**

Lectures 2 and 3. Introduction to Protein Structure and Function

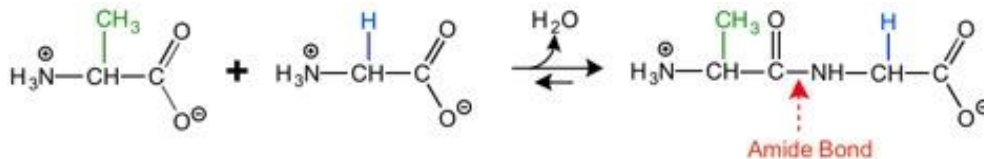
Outline:

1. Peptide bond formation
2. Hierarchy in protein structure: Primary structure; Secondary structure; Tertiary structure; Quaternary structure
3. How we get from primary to secondary structures (α helices; β strands; turns connecting secondary structures)
4. Tertiary and quaternary structures

1. Peptide bond formation via a dehydration reaction

Polymer formation from amino acid building blocks requires a dehydration reaction.

Chemists convert the HO of the carboxylate of one amino acid into a better leaving group and deprotonate the amine of the second amino acid to make a better nucleophile, that then allows formation of an amide bond concomitant with loss of H_2O . [NOTE: All polymerization reactions important in biology involve dehydration reactions.] The ribosome is the macromolecular machine inside cells that performs the same reaction. They catalyze this dehydration reaction without activation of the HO group for leaving. Note by convention, polypeptides are normally written with the N terminus on the left and the C terminus on the right.



2. Hierarchy in protein structure with hemoglobin (Hb) as an example:

I. Primary structure: The amino acid sequence of the polypeptide chain (Figure 1A).

[Anfinsen won the Nobel Prize](#) for the formulation of “Anfinsen’s Hypothesis” that the primary sequence is sufficient to determine the three dimensional structure of a protein. We now know that some small proteins can fold rapidly inside cells, but that in general, folding inside the cell involves chaperone proteins (DnaK and DnaJ), protein disulfide isomerases, and macromolecular machines that assist in the folding process (folding chambers such as GroEL-GroES in *E. coli* with similar counterparts in eukaryotic systems). You will learn about this machinery in Chemistry 5.08. Protein folding is particularly challenging given the crowded environment of the cell (Lecture 1) and hydrophobic stretches of amino acids.

II. Secondary structure: Regular features include helices (α , 3_{10} , or π ; Figure 1B), or β sheets (parallel or antiparallel; not shown in Figure 1B).

III. Tertiary structure: Packing of secondary structures to form a more complex structure (Figure 1C). This packing involves weak non-covalent interactions. Because of the structural revolution we now have solved, at atomic resolution, the structures of thousands of proteins. The structural motifs repeat themselves. Shown in Figure 1C is the structure of myoglobin (helix bundle protein).

IV. Quaternary Structure: Many proteins are composed of multiple polypeptide chains. The organization of these polypeptides is called the quaternary structure. Often, but not always, proteins with complex quaternary structures are involved in regulation. Hemoglobin (Hb) is shown as an example (Figure 1D).

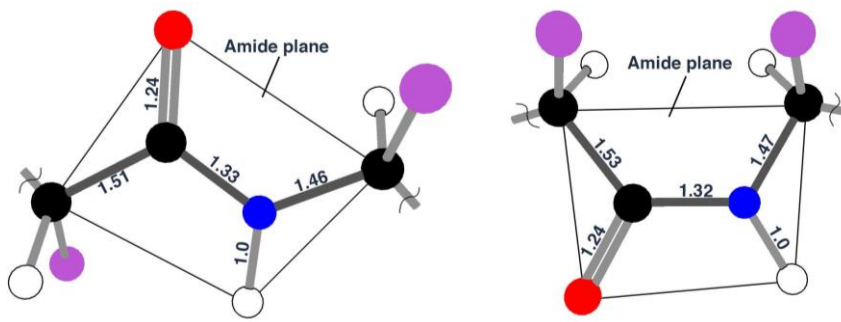
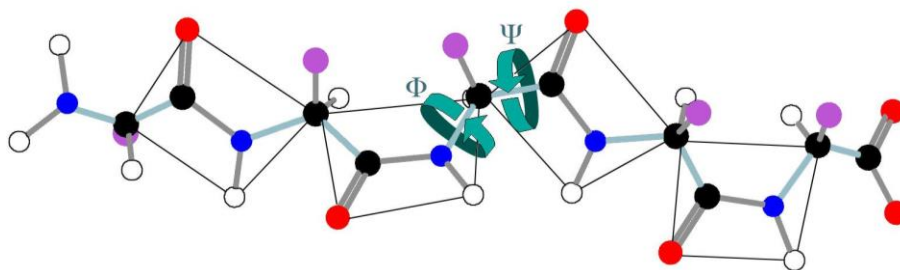


Figure by O'Reilly Science Art for MIT OpenCourseWare.

Figure 2. The planarity of the amide bond constrains the polypeptide to a limited set of secondary structures. The left panel shows a typical trans peptide bond in a polypeptide chain where the N-terminus is on the left and the C-terminus is on the right. The right panel shows a cis peptide bond.

The conformation of the polypeptide backbone is defined by the torsional angles ϕ (C_{α} -N) and Ψ (C_{α} -carbonyl) for each residue (Figure 3). Only certain values of ϕ and Ψ are allowed due to steric constraints. If the polypeptide had to sample all conformational space to fold into a protein, a simple calculation making reasonable assumptions shows that it would require the age to the universe to fold. Different secondary structural motifs have characteristic (ϕ, Ψ) angles. In the extended polypeptide conformation (ϕ, Ψ) are both 180° (Figure 3).

A.



B.

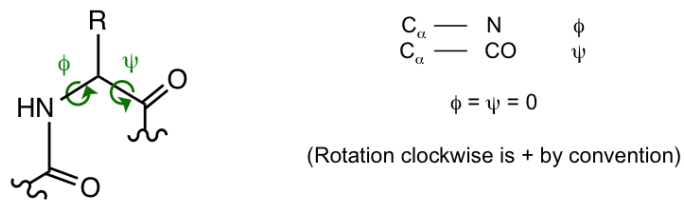
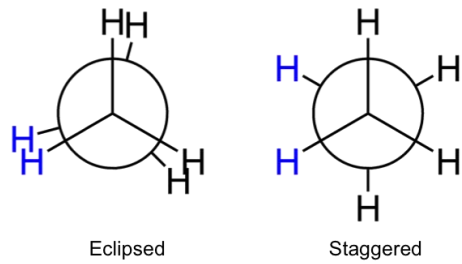


Figure 3A by O'Reilly Science Art for MIT OpenCourseWare.

Figure 3. Extended polypeptide chain and torsional angles. A. The planes of the amide bonds are shown. The chain consists of all trans peptide bonds. B. Torsional angles and rotation by definition.

This is a nomenclature convention (as you learned with R and S) and is analogous to the Newman projections that you used in organic chemistry to define the conformations of ethane or of cyclohexane rings (ethane is shown in Figure 4A).

A.



Courtesy of Leyo on Wikimedia Commons.
Newman projection images are in public domain.

B.

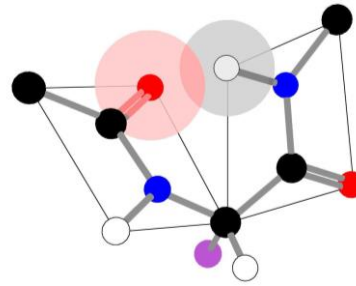


Figure 4B by O'Reilly Science Art for MIT OpenCourseWare.

Figure 4. Chemical conformations and steric hindrance. A. Newman projections of ethane in eclipsed and staggered conformations. B. Steric clash of amino acids in a polypeptide chain. $\phi = -45^\circ$, $\psi = 0^\circ$.

Look at dipeptides using models and use them to think about the (ϕ, Ψ) angles. This exercise helps you to see steric clashes which make certain conformations much less energetically favorable than other conformations. For example with ethane, the staggered conformation experiences less steric clashes than the eclipsed conformation (Figure 4A). In the dipeptide shown in Figure 4B, thinking about van der waals radii, one can see that in the conformation shown, the carbonyl sterically clashes with the H of the amide of the next amino acid (gray/pink overlap). This type of conformational analysis is used for nucleic acids and polysaccharides as well.

Seventy five percent of conformational space is inaccessible to proteins. Figure 5 is a Ramachandran plot that shows typical (ϕ, Ψ) angles found in the many proteins from the [Protein Database \(PDB\)](#) that have been examined.

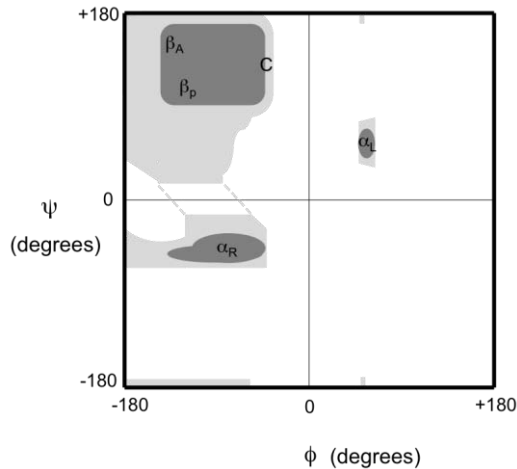


Figure 5. A Ramachandran plot of the ϕ and Ψ protein backbone dihedral angles for common secondary structures. Light grey represents allowed regions that include outliers, while dark grey represents the favored regions (not including outliers). α_R , right-handed helix (3_{10} , α , or π); α_L , left-handed helix; β_P , parallel β strand; β_A , antiparallel β strand; C, collagen triple helix. White regions represent unfavorable conformations due to steric clashes.

a. α Helices are very common 2^o structures as seen for Myoglobin (Mb) and Hemoglobin (Hb).

Generalizations about α Helices

Definitions: n = number of peptide units per helical turn; p = pitch = distance the helix rises along the axis per turn.

1. α helices have $(\phi) = -57^\circ$ and $(\Psi) = -47^\circ$.
2. There are 3.6 residues/turn (n), the pitch is 5.4 Å (p)
3. H bonding plays a key role in structure. C=O----H-N (linear, H bonds are NOT as directional as covalent bonds)
4. The α helix is a continuous region in the polypeptide chain
5. All carbonyls in the helix point in the same direction generating dipoles and these dipoles together generate a macrodipole with a (+) charge on the N-terminus of the helix. Note: negatively charged molecules often interact with the N-terminus of a helix.
6. The most common location of α helix is on the outside of the protein (Think about the location of the side chains of each amino acid.)
7. Proline is a helix breaker (Figure 6). Why?

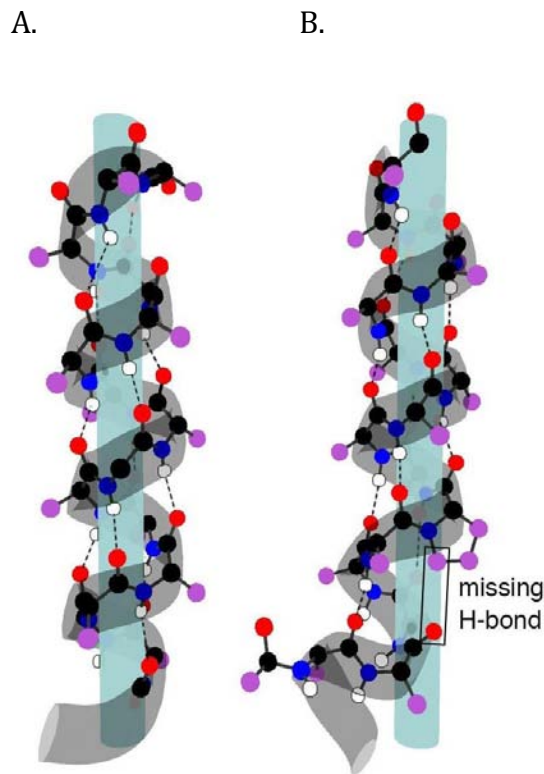


Figure by O'Reilly Science Art for MIT OpenCourseWare.

- A. PDB: [2HHB](#). Fermi, G., M. F. Perutz, B. Shaanan, and R. Fourme. "The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution." *Journal of molecular biology* 175, no. 2 (1984): 159-174.
- B. PDB: [5CTS](#). Karpusas, Mihail, Bruce Branchaud, and S. James Remington. "Proposed mechanism for the condensation reaction of citrate synthase: 1.9 Å structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A." *Biochemistry* 29, no. 9 (1990): 2213-2219.

Figure 6. A. Typical α -helical structure (left) showing residues 62-78 of Hb. B. α -helix in citrate synthase (residues 62-78) with a proline, showing a kink demarked by the loss of a H bond.

Figure 6 shows the side chains of the amino acids pointing away from the helix (purple balls). You can imagine that these side chains play an important role in tertiary structure. Note also the H bonds (black dotted lines) and that all the carbonyls are pointed in the same direction generating a dipole. Different types of helices have different ϕ , Ψ angles and H bonding. Some of the more common helices are shown below in Figure 7: the 3_{10} helix (3 peptide units/turn with a pitch of 6 Å), α helix (3.6 peptide units/turn and a pitch of 5.4 Å) and the π helix (4.4 peptides/turn with a pitch of 5.2 Å).

Three types of helices:

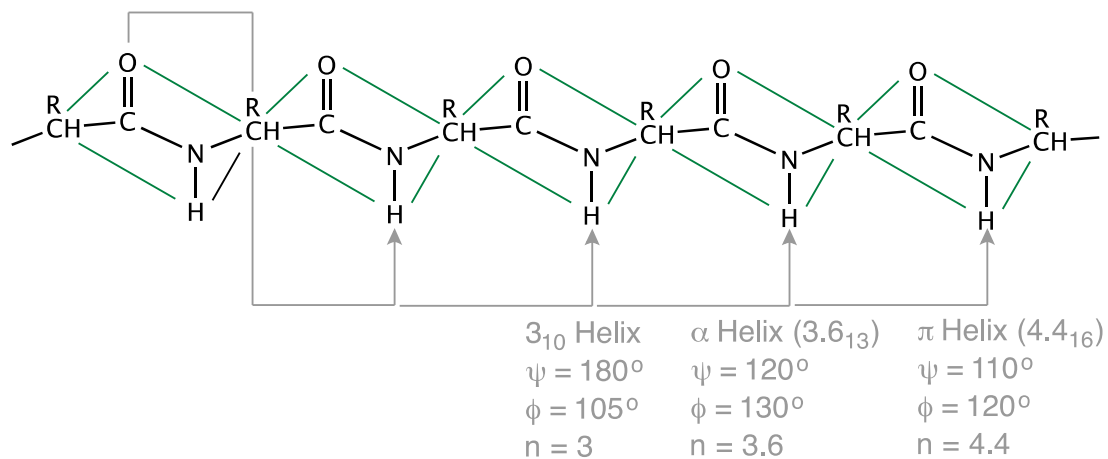
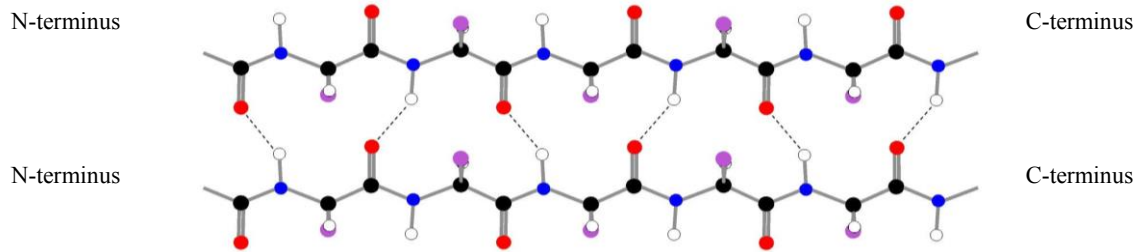


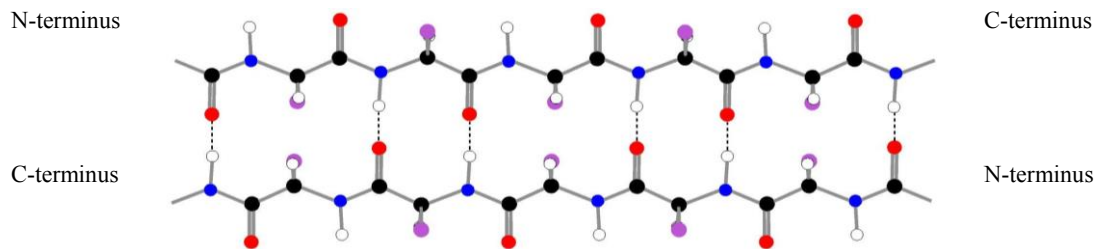
Figure 7. Types of helices. The α helix is common and contains 3.6 residues per full turn in the helix (360°). 3_{10} helices have a steeper, narrower corkscrew conformation so they only contain 3 residues per full turn. π helices have a longer, less compacted corkscrew conformation that has 4.4 residues per full turn.

b. The second type of very common **secondary structure are β strands**. These strands interact with other strands to form sheets. The sheets, as shown in Figure 8, can be formed by two strands with parallel or antiparallel predisposition.

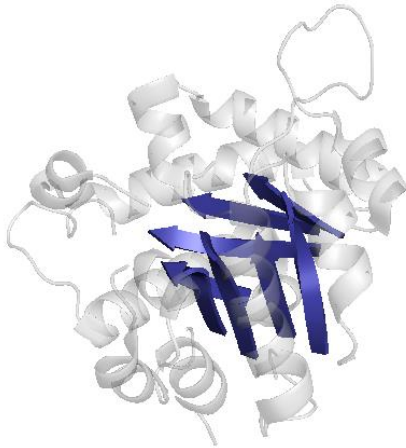
A. Parallel



B. Antiparallel



C.



D.

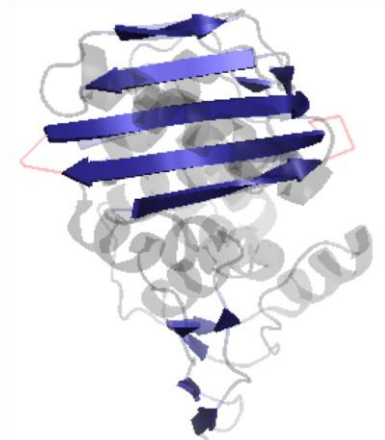


Figure 8A, B by O'Reilly Science Art for MIT OpenCourseWare.

C. PDB: [1YPI](#). Lolis, Elias, Tom Alber, Robert C. Davenport, David Rose, Fred C. Hartman, and Gregory A. Petsko.

"Structure of yeast triosephosphate isomerase at 1.9 Å resolution." *Biochemistry* 29, no. 28 (1990): 6609-6618.

D. PDB: [1ES4](#). Fonzé, Eveline, Marcel Vermeire, Martine Nguyen-Distèche, Robert Brasseur, and Paulette Charlier.

"The Crystal Structure of a Penicilloyl-serine Transferase of Intermediate Penicillin Sensitivity. The DD-Transpeptidase of *Streptomyces* K15." *Journal of Biological Chemistry* 274, no. 31 (1999): 21853-21860.

Figure 8. β strands that form sheets: A. Parallel B. Anti-parallel β strands. Hydrogen bonds are shown in dashed lines. C. Parallel β strands highlighted in blue from yeast triosephosphate isomerase. D. Antiparallel β strands highlighted in blue from *Streptomyces* K15 DD-transpeptidase.

Generalizations about β strands:

1. β sheets are composed of individual polypeptide strands and the strands are not necessarily contiguous in primary sequence space.
2. The polypeptides are fully extended and for parallel sheets (ϕ) = -119° and (Ψ) = $+113^\circ$, while for antiparallel sheets (ϕ) = -139° and (Ψ) = $+135^\circ$.
3. This motif uses full H bonding capacity. H bonds occur between neighboring strands to make sheets. The energetics of parallel and antiparallel β sheets are about the same.

Think about the location of the side chains of the sheets. The side chains as with the helices play a key role in the tertiary structure.

β turn: Turns provide connections between secondary structural motifs. Below in Figure 9 are two types of commonly observed turns in proteins (Type I and Type II β turns).

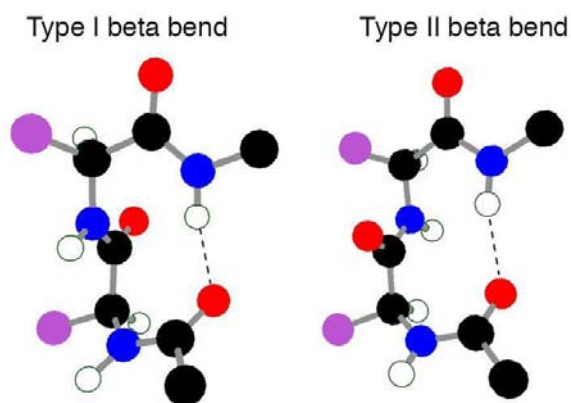


Figure by O'Reilly Science Art for MIT OpenCourseWare.

Figure 9. Helices and sheets can be connected by turns. Though there are many turns or bends, two of the most common are presented here - type I (left) and type II beta bends (right).

III. and IV. Tertiary and Quaternary Structure: Two examples of 3^o and 4^o structures will be presented in Chemistry 5.07: one a fibrous protein and one of a globular protein.

A. The extracellular matrix (ECM) is a highly organized multimolecular structure essential for life in higher organisms. It is largely composed of large fibrous polymers (polypeptides) and proteoglycans. Examples of fibrous proteins are collagen, fibrillin and fibronectin. One of the fibrous proteins best characterized structurally is collagen and this protein is the focus of our attention. Fibrous proteins, as the name implies, are elongated and the dominant motif is often repeated secondary structure that is organized into quaternary structure.

B. The second focus of our attention are globular proteins and Hb will be used as an example. Globular proteins are usually involved in catalysis (enzymes) and you will encounter many of them and their architectures over the course of the semester.

A. Collagen is the most abundant protein in all multicellular organisms (vertebrates). It is extracellular and supports the structure of most cells from the outside. In humans it is 25% of our protein. It is organized as insoluble fibers that possess amazing tensile strength. Gram for gram, type I collagen is stronger than steel. There are 30 types of collagen and the structures that they form all help tissues withstand stretching. It is a main component of cartilage, ligaments, tendons, bones and teeth. Its degradation leads to wrinkles and hence collagen treatment is frequently discussed in the press (see Figure 10)! Think about this and how collagen is synthesized in the cells, which is discussed in more detail later.



© source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/0>

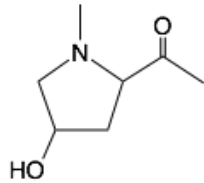
Figure 10. Collagen advertisement suggesting that daily application of a collagen-containing cream can reduce wrinkling of skin associated with aging.

1. Watson and Pauling, without success, tried to elucidate the structure of collagen. Crick and MIT's Alex Rich proposed the current structure. The inherent structure of large fibrous

polymers, flexible and often cross-linked, has impeded structural elucidation at atomic resolution.

2. Type I collagen is the best characterized and is composed of three polypeptides each of molecular weight 285 kDa. The three polypeptide chains all form left-handed helices with **NO** H-bonds within any one chain. The three polypeptides wrap around each other to form a right-handed rod that is 300 nm (3000 Å) x 15 Å and is one of the longest proteins known. The three polypeptide chains together define the quaternary structure.

3. The amino acid composition of each polypeptide is a repeat of G-X-Y where X and Y are Pro or a post-translationally modified 3-HO-Pro (HyP).



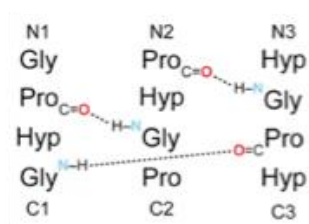
Hydroxyproline (HyP)

One third of the residues in each chain are Gs which are found in the center of the triple helix for steric reasons. The glycine amide hydrogens make H bonds with the carbonyls of prolines between chains where the interactions are **interchain** (Figure 11).

A.



B.

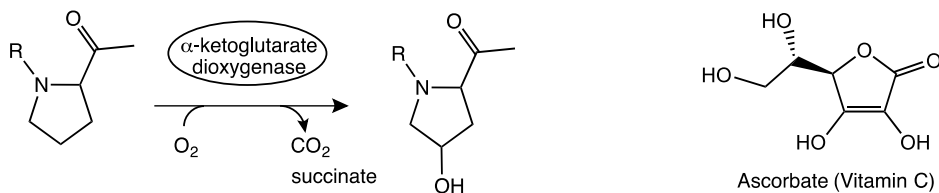


Courtesy of Annual Reviews, Inc.

Source: Figure 1BC from Shoulders, Matthew D., and Ronald T. Raines. "Collagen structure and stability." *Annual review of biochemistry* 78 (2009): 929.

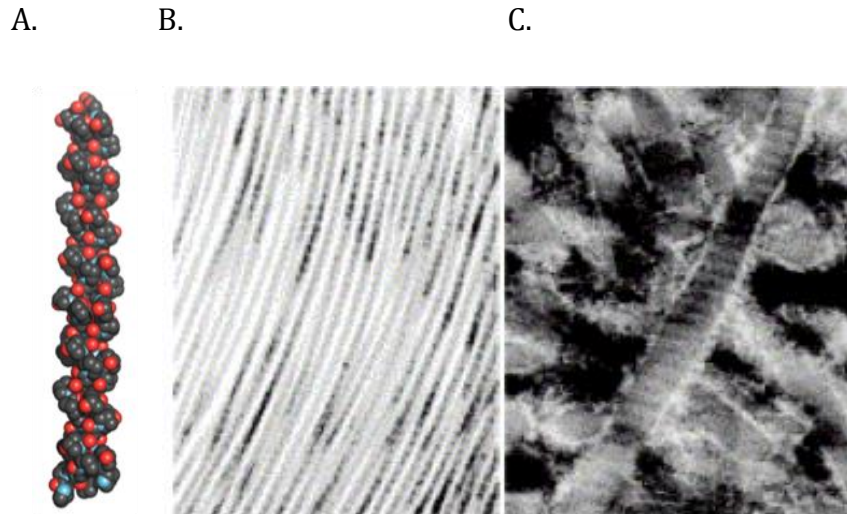
Figure 11. Collagen triple helix. A. Ball-and-stick image of a segment of collagen triple helix showing the ladder of interchain hydrogen bonds. B. Stagger of the three strands in the segment shown in A highlighting the interchain H bonds.

MEDICAL interlude: The hydroxylation of proline is an example of post-translational modification (PTM) discussed in Lecture 1. The reaction is catalyzed by an α -ketoglutarate (KG) dioxygenase and requires Fe^{2+} and ascorbate (Vitamin C). Ascorbate keeps the iron in the reduced state.



In the 16th century, Scurvy was a disease from which many sailors died. On long trips they had no access to Vitamin C, essential to making intact collagen. With no Hyp in collagen, it denatures (unfolds) into individual polypeptides at 24°C , instead of 39°C . The phenotype associated with unhydroxylated Pros is skin and blood vessel fragility. Note α -KG is an intermediate in the TCA cycle and α -KG dioxygenases are now known to play major roles in human physiology (HIF – hypoxia inducible factor) with respect to cancer.

End medical interlude.



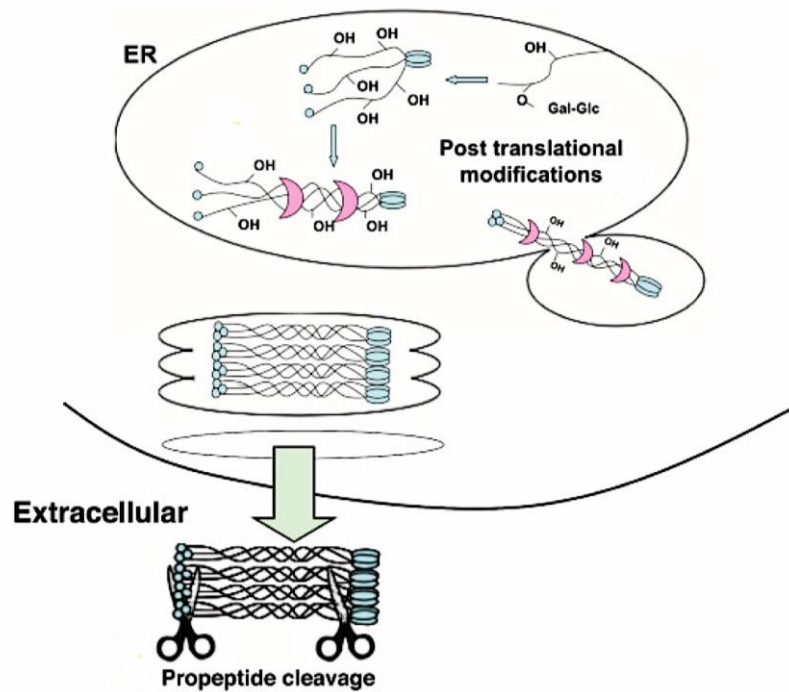
Courtesy of Annual Reviews, Inc.

Source: Figure 1A from Shoulders, Matthew D., and Ronald T. Raines. "Collagen structure and stability." *Annual review of biochemistry* 78 (2009): 929.

Elsevier Fig 2B & C from Source: Gelse, Kolja, E. Pöschl, and T. Aigner. "Collagens—structure, function, and biosynthesis." *Advanced drug delivery reviews* 55, no. 12 (2003): 1531-1546.

Figure 12. A. First high-resolution crystal structure of a collagen triple helix, formed from (ProHypGly)₄-(ProHypAla)-(ProHypGly)₅. B. Type I collagen fibrils from a tendon that are organized in a parallel manner. C. Type II collagen fibrils from articular cartilage, which is organized into a network.

The quaternary structure of collagen (Figure 12A) has inspired rope making (see [video on rope making here](#)) and the fibrils are interesting to bioengineers as scaffolds. Collagen biosynthesis is very complex (see simplified schematic shown in Figure 13). Matt Shoulders in our department is making important contributions to mapping the biosynthesis.



Courtesy of BioMed Central Ltd from Clarice Chen and Michael Raghunath. PMID: 20003476. [<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805599/pdf/1755-1536-2-7.pdf>]. CC license BY.

Figure 13. Simplified schematic of the translation and post-translational steps of biosynthesis of collagen. Single collagen strands are imported into the endoplasmic reticulum where procollagen formation and post-translational modification occur. Procollagen is exported from the cell to the extracellular matrix where it is cleaved by proteinases prior to crosslinking.

As with many biochemical systems that have been studied in detail, collagen has attracted attention due to the devastating effects of inborn errors in metabolism. As an example, a single mutation of G988C (guanine to cytosine—DNA level) at the C-terminus of type I collagen causes a genetic disease called **osteogenesis imperfecta**. The phenotype is skeletal deformation resulting in multiple fractures due to brittle bones. There is disruption of the collagen triple helix at its C-terminus that leads to extensive post-translational modification by hydroxylation and glycosylation (attachment of sugars) and the inability to form the highly ordered fibrils (Figure 12).

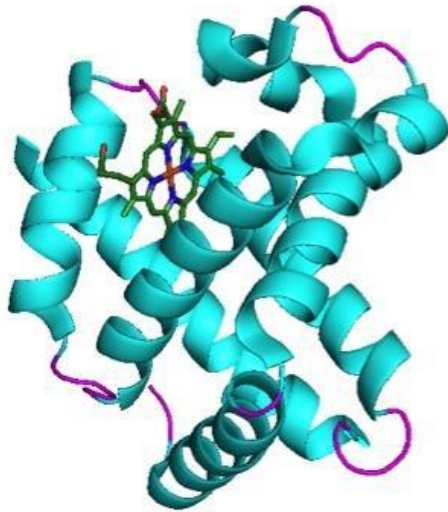
B. Globular Proteins (compact spherical molecules) that have tertiary structure and can have quaternary structure.

1. General Rules:

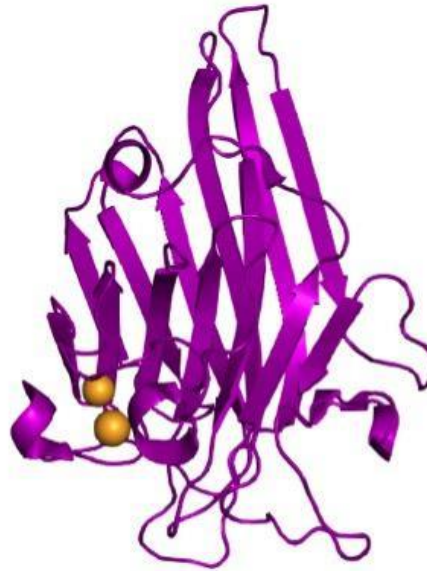
- a. Globular proteins are very compact. They have little H₂O in their interior. Their packing density is similar to that observed with crystals of small organic molecules.
- b. Side chain location: Non-polar residues are almost always in the interior of proteins [V, L, I, F, M etc]; Charged residues are almost always on the surface and have interaction with H₂O [R, K, H, E, D]; Uncharged residues [S, T, Y, N, Q, W] are on the surface or if buried have H bonding interactions with appropriate donors and acceptors.
- c. The C and N termini of globular proteins are found on their surface. Allows most proteins to be isolated by affinity chromatography using a tag ((His)_n-tag, FLAG tag) the sequence of which can be appended to the ends of the gene.

2. There are many cartoons of different types of globular proteins in [the Protein Data bank \(PDB\)](#) that you can browse. We now have solved thousands of 3-dimensional structures of proteins at atomic resolution by crystallography and by nuclear magnetic resonance (NMR) methods. There are many websites, including [SCOP](#) and [CATH](#), focused on categorizing protein folds, which are proposed to be limited in number. The four general categories of protein folds are: α structure, β structure, α / β structure and little or “unfolded” structure. An example of an α structured protein, myoglobin (Figure 14A), is highly homologous to the structure of both subunits of Hb, and binds oxygen. An example of a β structure is concanavalin A (Figure 14B), which is a lectin that binds sugars.

A.



B.



A. PDB: [3RGK](#)

Hubbard, Stevan R., Wayne A. Hendrickson, David G. Lambright, and Steven G. Boxer. "X-ray crystal structure of a recombinant human myoglobin mutant at 2· 8 Å resolution." *Journal of molecular biology* 213, no. 2 (1990): 215-218.

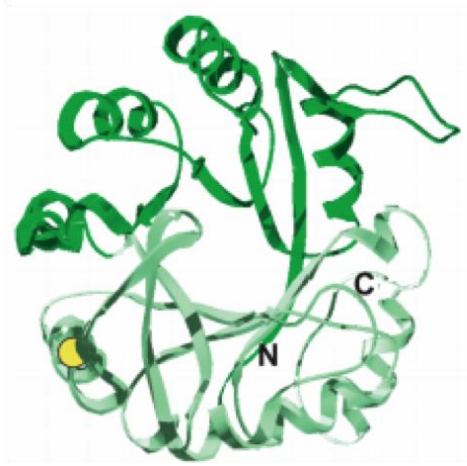
B. PDB: [2UU8](#)

Ahmed, H. U., M. P. Blakeley, M. Cianci, D. W. J. Cruickshank, J. A. Hubbard, and J. R. Helliwell. "The determination of protonation states in proteins." *Acta Crystallographica Section D: Biological Crystallography* 63, no. 8 (2007): 906-922.

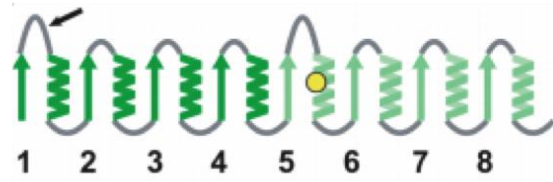
Figure 14. Crystal structures of proteins with almost exclusively either α or β secondary structure. A. Predominantly α structure of myoglobin bound to its heme group represented as balls and sticks. B. Mostly anti-parallel β structure of a single subunit with its Ni²⁺ and Ca²⁺ ions bound (orange balls) of the homotetramer concanavalin A.

α/β structure: The Enolase Superfamily will be used as an example. Triose phosphate isomerase (TIM), an enzyme you will meet again in the glycolysis pathway, is an example. Ten percent of all proteins are estimated to have this fold. A TIM barrel and the plumbing diagram that describes the secondary structure of this domain with 8 strands and 8 helices that form the barrel is shown in Figure 15. The residues involved in catalysis in the active site in all of these TIM barrels are located in the loops between the strands at their C-terminus.

A.



B.

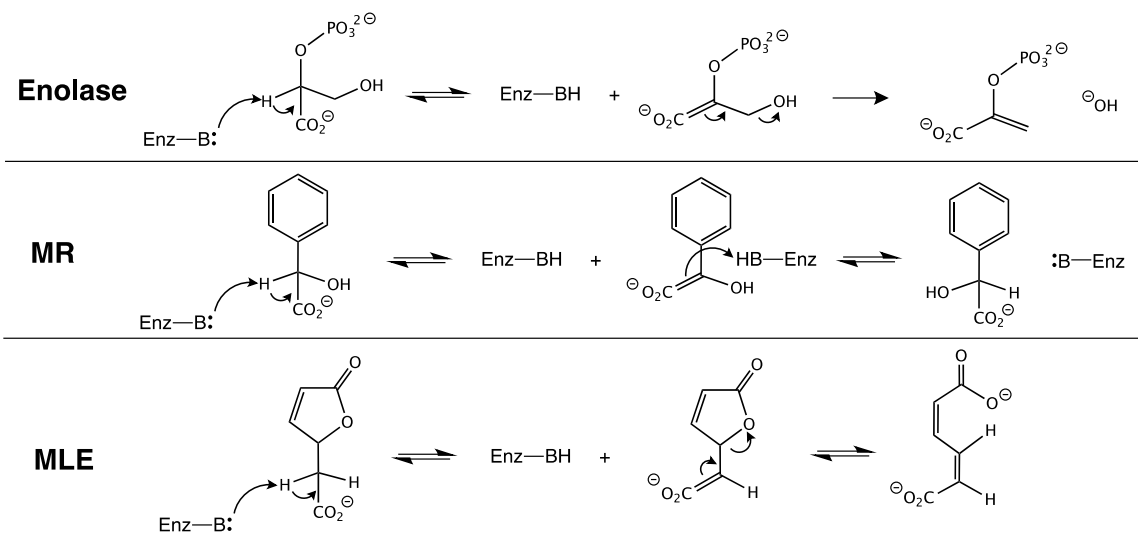


Source: Höcker, Birte, Jörg Claren, and Reinhard Sterner. "Mimicking enzyme evolution by generating new ($\beta\alpha$) 8-barrels from ($\beta\alpha$) 4-half-barrels." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 47 (2004): 16448-16453.
Copyright © 2004. National Academy of Sciences, U.S.A.

Figure 15. TIM barrel shown in the HisF protein. A. Eight beta sheets form a channel, while the eight helices surround that channel. Groups involved in catalysis reside in the loops between the β strands and α helices. B. Plumbing diagram describing secondary structure interconnectivity.

This structural motif has been conserved and used widely as a platform to evolve new catalytic activities. The three reactions shown below are catalyzed proteins that have a TIM fold (Figure 16A). Furthermore the amino acids involved in catalysis in the active site of these enzymes are also conserved (Figure 16B). Note the diversity of the substrates (small molecules) that can bind to the same fold. Also note the diversity of the chemistries (an elimination reaction, an epimerization reaction and a lactonization reaction), although all three enzymes **share the ability to remove an extremely non-acidic proton** to initiate the catalytic transformation.

A.



B.

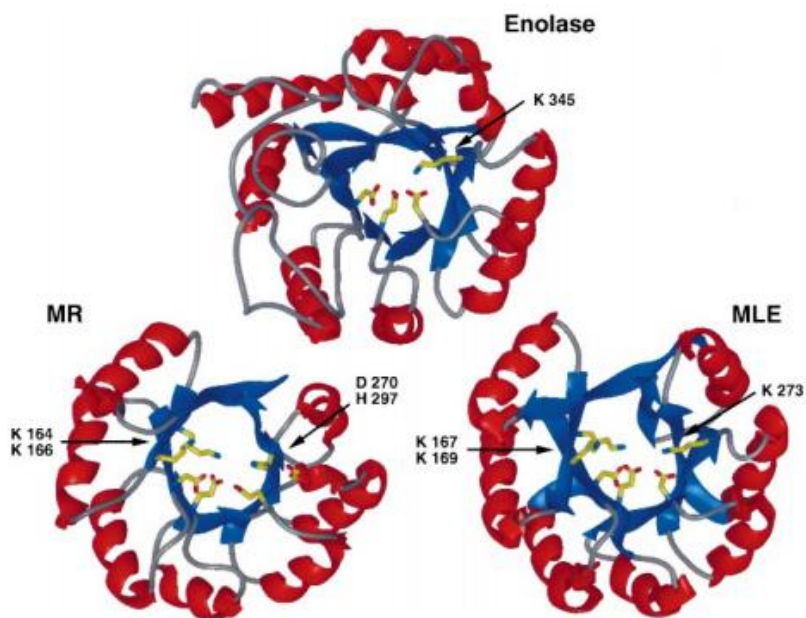
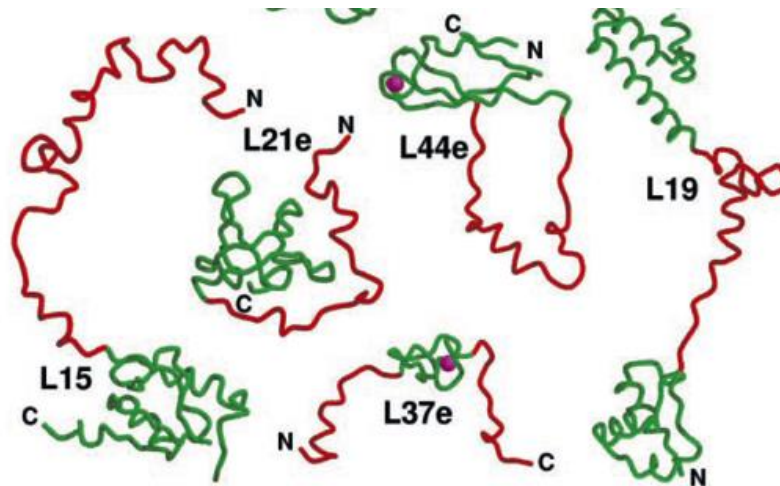


Figure 16B courtesy of Annual Reviews, Inc.

Source: Figure 2 from Gerlt, John A., and Patricia C. Babbitt. "Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies." *Annual review of biochemistry* 70, no. 1 (2001): 209-246.

Figure 16. Enolase SF members show structural similarities: enolase, mandelate racemase and muconate lactonizing enzyme. A. Each enzyme catalyzes a similar reaction, even though the substrates are quite different. B. Ligands to the two metals are all conserved.

Many proteins inside the cell are now being found to have little structure or no structure (Figure 17). A spectacular example of this was provided in 2000 (Steitz and Moore, Science 2000) by the structure of the ribosome (the conserved machine that makes polypeptides in all organisms). This machine is composed of RNA and protein; some of the proteins are shown below (L indicates proteins in the large subunit of the ribosome). The red shows the unfolded protein, while the green reveals secondary structure. Only a few of the 50 proteins within this machine are shown below.



© AAAS. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/0>
 Source: Ban, Nenad, Poul Nissen, Jeffrey Hansen, Peter B. Moore and Thomas A. Steitz. "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution." Science 289, no. 5481 (2000): 905-920.

Figure 17. Ribosomal proteins frequently contain multiple unstructured regions. Shown are the structures of the L15, L19, L21e, L37e, and L44e proteins of the large subunit of the ribosome. Unstructured regions are shown in red, while the regions in green are globular in structure. The pink dots in L37e and L44e are zinc ions associated with the globular domains (green) of these proteins.

MIT OpenCourseWare
<https://ocw.mit.edu>

5.07SC Biological Chemistry I
Fall 2013

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.