

3. TERTIARY & QUATERNARY STRUCTURE

As noted at the end of the last section, it is difficult to obtain a peptide that stably adopts a given secondary structure. Entropy is too big a barrier to overcome, and the enthalpic benefits of adopting a given conformation are too skimpy to make it happen. However, peptide chains do adopt α -helix and β -strand conformations. They just don't do it in isolation. Here we'll examine the additional influences that can stabilize a peptide in a secondary structure conformation.

Fibrous Proteins

Heptad Repeats in Coiled Coils

The earliest investigations into protein structure were performed on **fibrous** proteins. Examples include elastin and collagen (found in connective tissue) and fibroin (found in silk). Another example is **keratin**, the protein that comprises rhino horns, finger nails and wool (and hair too, but so what). The bulk of each polypeptide chain in keratin, running about 300 residues, is α -helical, and each keratin chain partners with a second to form a **coiled coil** structure, in which the two helical polypeptides snake around each other. This is a stable helical protein thanks to the interactions taking place *between* the chains rather than within chains. Interestingly, those interactions are virtually all between non-polar residues.

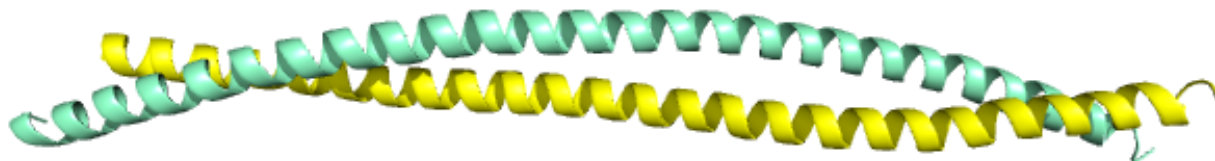


Figure 3.1 Cartoon diagram of the coiled coil structure of keratin. One strand is colored yellow and the other pale green. Helices will commonly be displayed in this fashion. (Model adapted from PDB ID 3TNU using PyMOL.)

Keratin is a relatively complex protein, but a simple motif, known as the **heptad repeat**, has been extracted from a study of its structure. The interface of the two strands is composed of non-polar residues that appear at the 1st and 4th position in a 7-residue pattern that repeats (Sch. 3.1).



Each position, labeled “A” or “D” above will be a non-polar residue. Moving away from keratin to a simpler, soluble model peptide, the **leucine zipper** motif found in the yeast protein GCN4 places leucine residues reliably at each A/D position over a 30-residue span. Because of the 3.7 residue/turn structure of the α -helix, the heptad repeat places the leucine residues reliably on one face of the helix (Figure 3.2). This creates what is known as an **amphiphilic** or **amphipathic** helix, possessing one face that is polar and one that is non polar. As a result of this structure, the non-

polar leucine residues are buried from solution upon formation of the dimer, creating a **hydrophobic core** that assists in stabilizing the protein structure. This is one of the key features of all stable, folded proteins – an interior composed of non-polar side chains that are segregated from contact with the surrounding aqueous medium.

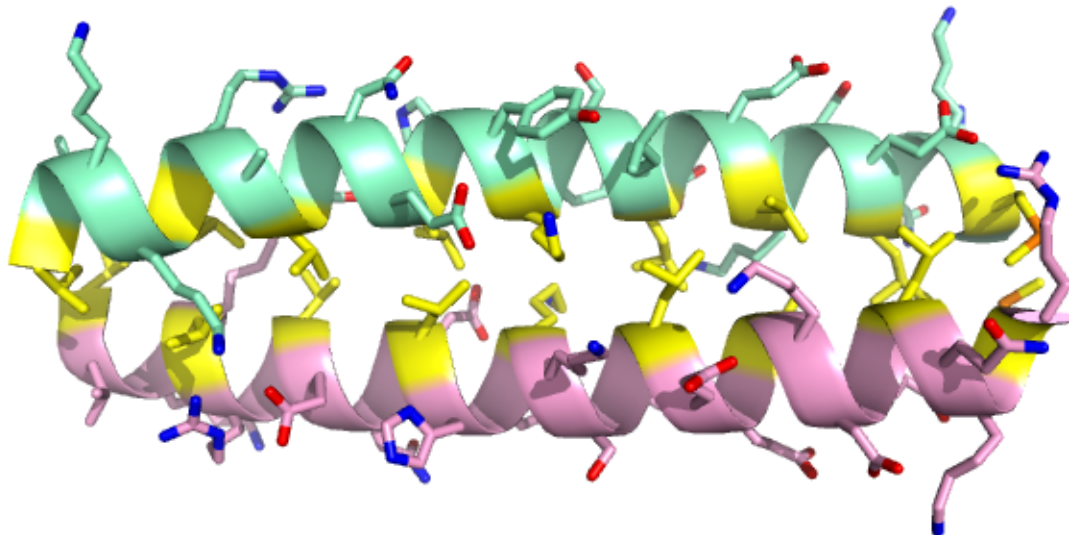


Figure 3.2 Leucine zipper motif in the GCN4 protein. The two peptides (colored pink and pale green) share a face of non-polar residues. Those residues that fit in the heptad pattern AbcDefgA' etc., are colored yellow. The helices are amphiphilic, in that the non-polar residues are segregated to one face of the helix, while polar residues are concentrated on the opposing face and point to solution.

A Pathological Fibrous Protein with β -Strands – Amyloid

While fibroin and even stretched keratin (β -keratin) are naturally occurring fibrous proteins with extensive β -strand content, the most well-studied all- β fibrous proteins are related to human disease. In the 1980's and 1990's Stanley Prusiner began investigating the causative agent of kuru (laughing sickness); a disease in Papua New Guinea that appeared linked to ritualistic cannibalism. Interestingly, at that time another disease arose in the UK linked to cannibalism in cattle. Cows were obtaining protein in their diet from rendered bovine offal taken from slaughterhouses. Mad Cow Disease (bovine spongiform encephalopathy) was typified by stumbling among cattle, loss of cognition, such as it is in cows, and death. Soon the disease was found to have crossed into humans. It is related to a genetic disease with similar symptoms, Creutzfeldt-Jacob.

The story towards understanding is a long one, but the upshot is that these diseases and a variety of other neurological diseases (including Parkinson's, ALS and Alzheimer's) all appear to be linked to the build up of **amyloid fibrils** in the brain. "Amyloid" because they looked like starch fibers in electron micrographs, but in fact they are made of protein, and they are amazingly resistant to degradation. Fibrils remain infectious after heavy processing of neurological tissue and are able to

act as seeds for fibril formation in a new host, even if transmitted through ingestion (not a protein-friendly environment).

Fibril formation occurs when two native proteins unfold slightly, and the unfolded portions become linked in interprotein β -strand interactions. The two-strand “sheet” then is a seed for adding new strands until a long fiber is obtained running perpendicular to the β -strands. As with α -helices, β -strands are not stable in isolation, but the packing that occurs during fibril growth gives increased stability, much as the coiled coil does for helices. In a solid-state NMR study, Tycko and colleagues have provided the first experimental model for the fibrils that appear in Alzheimer’s disease. A naturally occurring 40-residue peptide, **amyloid β** , forms two strands that place non-polar residues to the center (Figure 3.3). These strands don’t interact with each other but rather recruit other peptides to H-bond between strands, forming two layers of β -sheet protecting a hydrophobic set of residues. This structure depends on every other residue along the peptide to be hydrophobic, since β -strands have alternating side chain orientations along the chain. Indeed, the sequence of amyloid β is such that nine non-polar residues are buried in each peptide contributing to the fibril. As with keratin, amyloid fibrils take advantage of the hydrophobic effect to stabilize otherwise entropically costly conformations.

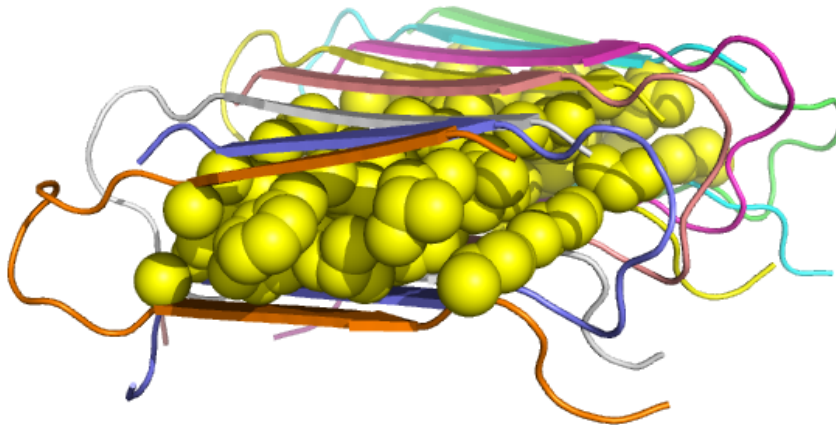


Figure 3.3 Structure of an amyloid fibril derived from solid-state NMR data. This model contains 8 peptide strands (each with a different ribbon color) aligned parallel to each other but perpendicular to the fibril axis. The sequence **QKLVFFA**ENVS**NKKAI**GL**MVGG**VV allows the β -strand conformation to hide seven **non-polar** residues from solution.

The Tertiary Structure of Globular Proteins

The proteins that we will focus upon this semester are **globular** not fibrous. They are typified by a collapse of the long polypeptide chain (often over 500 residues) into compact structures that, if not

spherical, at least adopt shapes that significantly reduce the surface area exposed to solvent. In the process of folding from the **denatured** to **native** state, a hydrophobic core forms.

One simple structure is the four-helix bundle, which is actually a variant on the coiled coil, but with the chain folding into four distinct helices that collapse on top of each other. The driving “force” for that collapse is, in part, the hydrophobic effect. Non-polar residues end up in the protein interior while polar residues are on the exterior. Each of the four helices is amphiphilic (Figure 3.3). Essentially one is generating a twinkie-like structure. Hydrophobic “creme” on the inside and sponge cake on the outside.

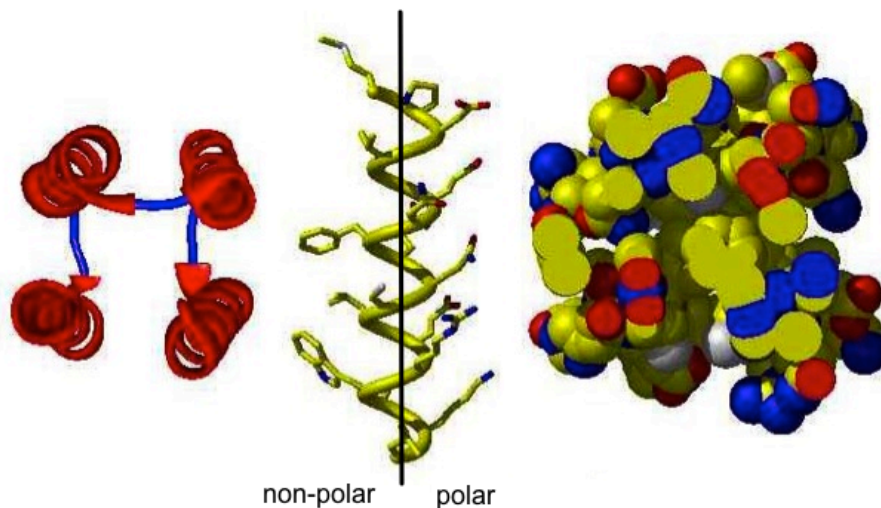


Figure 3.4. Structure of a 4-helix bundle protein. At right is a cross-section of the protein, looking down the same axis as the cartoon diagram at left. Note the non-polar (yellow) residues in the protein interior.

Similar behavior can be seen with β -sheets. The simplest all β -strand structure is composed of two sheets lying atop each other and burying non-polar residues to the center. The so-called β -sandwich structure is in some respects reminiscent of a grilled cheese sandwich. Again, grease to the middle and more polar residues to the surface. These structures are made of amphiphilic strands, not helices. As you may recall, the β -strand alternates the orientation of side chains 180° apart along the strand, so that an amphiphilic strand will have an alternating pattern of hydrophobic/hydrophilic side chains (Figure 3.5).

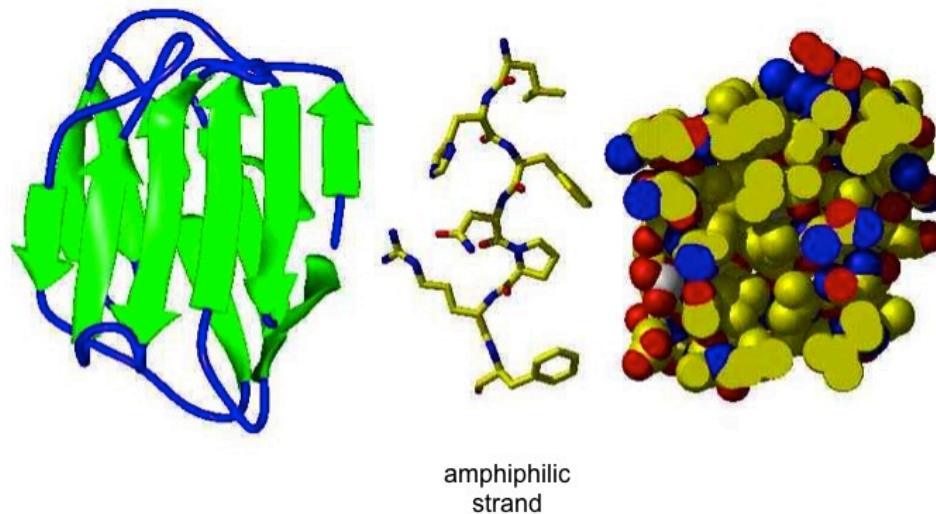


Figure 3.5. A β -sandwich protein. Notice the pattern of alternating polar and non-polar residues along the amphiphilic strand. The non-polar interior of the protein contains the non-polar face of the many individual strands.

The above examples of globular proteins introduce two out of the many, many “folds” that one can find among the structural databases. They are an introduction to the **tertiary** structure of proteins. Where primary structure is sequence and secondary structure is local conformation, tertiary structure can be viewed as the global conformation of a single polypeptide chain.

Additional Structural Elements Leading to Stability

Tertiary folds may be stabilized by bonding interactions in addition to intermolecular forces and the hydrophobic effect. The two most common bonding interactions to stabilize proteins are disulfide bridges and metal-ligand interactions

Cysteine is a redox active protein (see Ch. 2). Two cysteine residues can be oxidized by two electrons to form a **cystine** residue (note the absence of the “e” after oxidation. cute.) This cross-links two portions of the polypeptide strand and stabilizes a compact fold. Disulfides are commonly found in small, monomeric, extracellular proteins. Because the interior of the cell is reducing, structural disulfide are rarely found in intracellular proteins. The classic protein to be stabilized by disulfides is ribonuclease A (RNase), a mammalian digestive enzyme. It contains 120 residues and four disulfide bonds (Figure 3.6B). The protein can be heated to 90°C and lose all secondary structure, but the disulfide bonds remain intact and allow rapid refolding of the protein.¹

¹ Christian Anfinsen won the Nobel Prize in part for showing that even if you reduced the disulfides, RNase could refold to its native state along with reoxidation. That observation led to the argument that the tertiary structure of proteins is determined entirely by the primary sequence of the protein. As the amyloid fibril story shows – it’s a little more complicated than that. A protein may adopt different structures depending on environmental conditions, and in fact some proteins are unable to fold independently under any circumstances and require helper proteins known as chaperones.

Metal-ligand bonds can vary from nearly ionic to virtually covalent in quality. Across the spectrum, they provide additional enthalpic stabilization to the tertiary fold. The most common structural metal in proteins is zinc. Zn^{2+} ions are electron acceptors and form bonds with lone pair donors such as carboxylate (Asp and Glu), imidazole (His) and thiolate² (Cys). Zinc fingers are an archetypal tertiary fold stabilized by zinc. Consisting of 30 or so residues, the “classic” zinc finger folds into two β -strands and an α -helix, with donation of lone pairs from two cysteine and two histidine residues, forming four metal-ligand bonds (Figure 3.6B). In other zinc fingers, either one or both of the histidine electron donors may be replaced with cysteines.

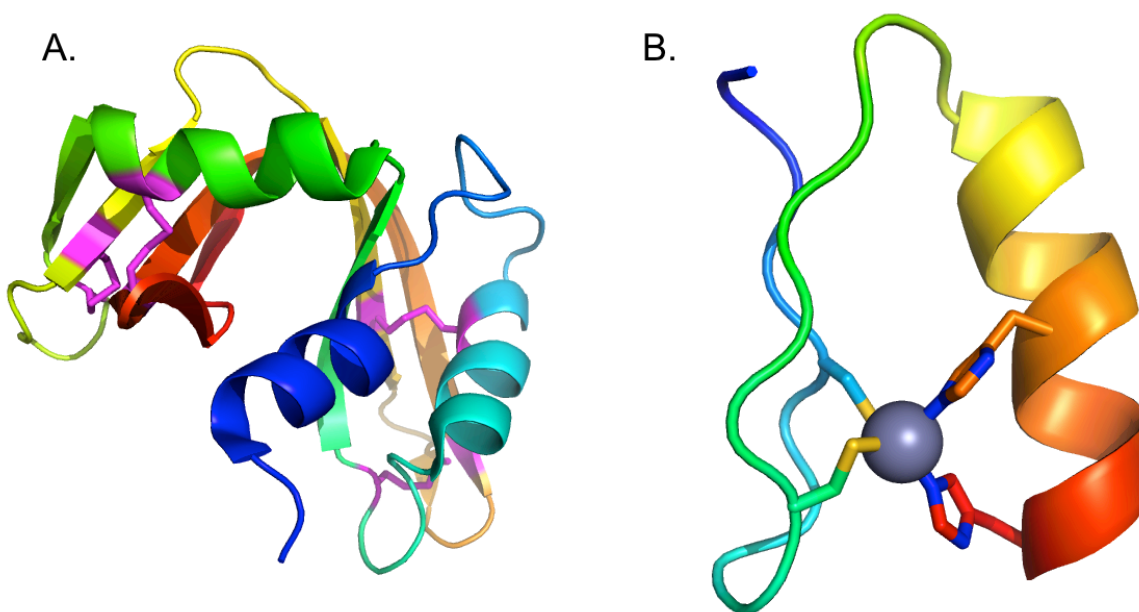


Figure 3.6. (A) Structure of ribonuclease A. Note the four disulfide bonds colored in magenta. (B) Structure of a zinc finger protein. The zinc ion is shown as a gray sphere bonded to two histidines and two cysteines. The backbone is colored to start with a blue N-terminus and progress through the spectrum to a red C-terminus.

Domain Structure in Proteins

Tertiary structure can be quite complex for a large multi-functional protein. Often, the overall tertiary structure is dissected into discrete **domains**, that are independently folded portions of the polypeptide chain. Those domains may then pack together to create a dense structure, or they may remain connected by a poorly structured “linker” that allows significant motion between domains (for example, see Figure 3.7). It is possible for domains to include non-contiguous stretches of the protein chain, but more commonly the distinct folds lie like discrete units along the chain.

² The Cys-SH thiol has a pK_a of about 8.0. However, when interacting with a Zn^{2+} (or other metal cation), the conjugate base form, Cys-S⁻ (a thiolate) is stabilized in comparison to the thiolate in solution. By selectively stabilizing the conjugate base in an acid/base pair, the Zn^{2+} ion lowers the pK_a of cysteine significantly, so that a zinc-bound cysteine is completely deprotonated at pH 7.

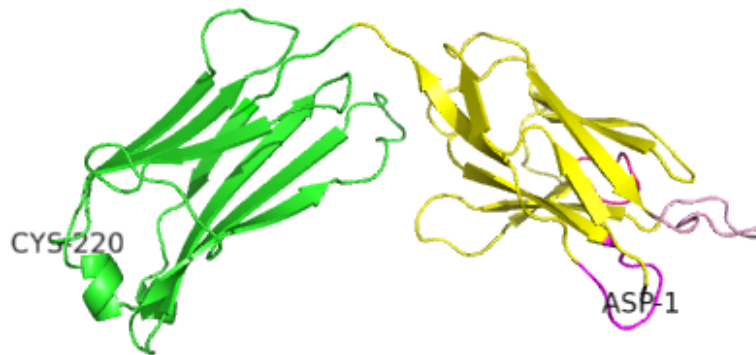


Figure 3.7. A two domain immunoglobulin chain. The 220-residue protein folds into two discrete domains, one comprising the first 110 residues of the protein (yellow) and the second comprising the second 110 residues (green).

Protein domains are often correlated to the function of a protein. They move independently of each other, allowing the protein to adopt alternate three-dimensional structures with alternative functional qualities (Figure 3.8), or it could be that each domain possesses a unique function, and tied together along the peptide strand it brings multiple functions together on one polypeptide strand.

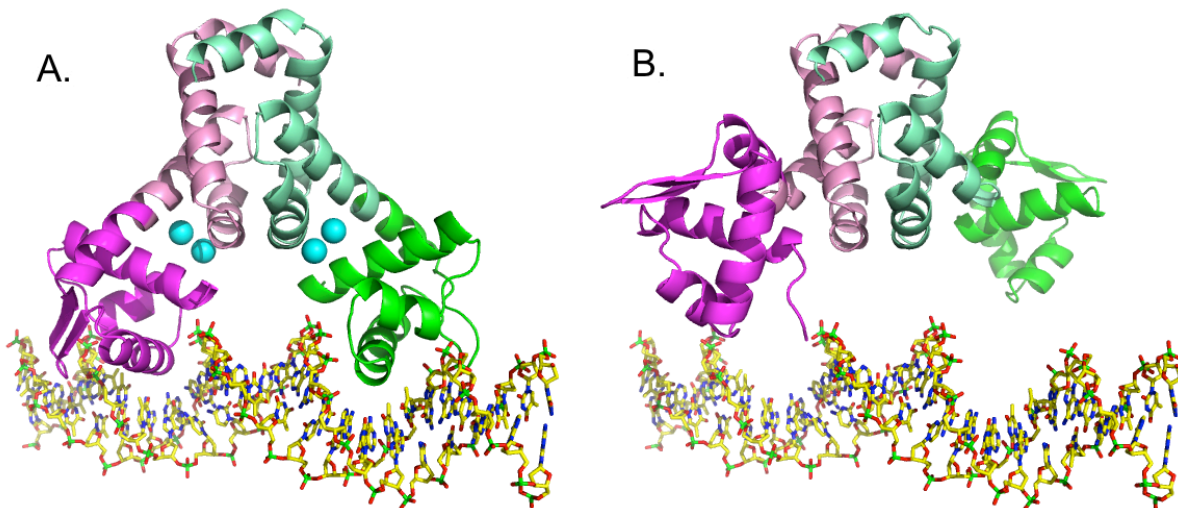


Figure 3.8. Two conformations of the DNA-binding protein, MntR, determined by movement of domains. (A) Mn^{2+} ions, shown as blue spheres, hold the protein in a conformation to bind DNA. (B) In the absence of Mn^{2+} , two domains (in dark magenta and green) move with respect to the rest of the protein, creating a new shape that no longer interacts with DNA.

Aggregation of Protein Chains: Quaternary Structure

In addition to common phenomenon that polypeptide chains “collapse” to form compact globular structures, it is also common for multiple globular proteins to aggregate to form **multimers**: dimers, trimers, tetramers and so on. To the degree that the formation of a tertiary fold can stabilize individual secondary structure elements by the power of the hydrophobic effect, it is also true that tertiary structure can be further stabilized by burying hydrophobic surface area with other proteins in defined structural complexes. In a survey of proteins found in *E. coli*, 19% are monomeric, 38% are dimeric, 5% are trimeric, 21% are tetrameric while the rest are of higher orders.³ The aggregation state, and interactions creating it, are referred to as quaternary structure.

Proteins may exist as homooligomers or heterooligomers, depending on whether the aggregate is formed from only one chain or several chains. The identity of chains is often denoted by using the Greek alphabet. For example, an α_2 dimer is a dimeric protein complex composed of two identical polypeptide chains (homodimer). An α_4 tetramer, a homotetramer, has four of the same chain. Hemoglobin is an $\alpha_2\beta_2$ tetramer, indicating that the heterotetramer is constructed from two of each of two kinds of polypeptide chains. There are heterotrimeric $\alpha\beta\gamma$ proteins with three different chains. It can go on and on.

On my list for an update... Maybe by next week.

³ Goodsell and Olsen (2000) *Ann. Rev. Biophys. Biochem.* **26**, 105.