

Vanessa Lacey

BEH342/442 15 October, 2005

**Molecular Structure of Biological Materials: Structure, Function & Self-assembly**

Take-home Midterm Exam (Due Thursday, Nov. 17) 30% of the total grade. Please write a short description for each of the questions (10 point each). You need to think carefully and creatively, study the textbooks and lecture notes.

1. What is the structure of water in different temperatures?

A single water molecule forms a tetrahedron, where the two hydrogen atoms and lone pairs sit at the corners, and the single oxygen molecule forms the center. These lone pairs act as Hydrogen (H) acceptors during H-bonding with neighboring water molecules, and the H atoms are of course donors. Each molecule is therefore both a donor and an acceptor, allowing clusters (and rings) to form due to the tetrahedral shape. Figure 1 depicts how a single

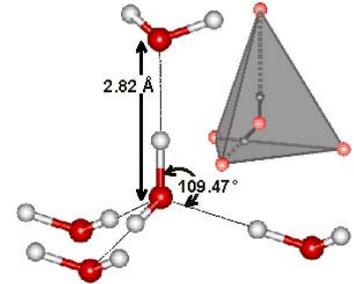


Figure 1  
Courtesy of Professor  
Martin Chaplin, London  
South Bank University.  
Used with permission.

tetrahedral water molecule can H-bond to four other molecules, setting a definitive angle between neighboring molecules, as well as bond distance (depending on temperature).

In general, liquid water is arranged into small groups of joined clusters. Since these groups are small, water is able to move and flow. Frozen water, on the other hand, has a highly structured geometric pattern. Water at high temperatures in the form of gas is highly charged with energy, causing molecules to be in constant motion, thereby reducing the likelihood of bond formation between molecules.

Water does not conform to a single structure at select temperatures due to the dynamic nature imposed by complex hydrogen bonding. The three dimensional H-bond network can undergo many configurations. Even at constant temperature of 230K (-43.1 °C), water has four average “distinct” configurations. During the initial state of cooling (quiescent state), supercooled liquid water undergoes H-bond rearrangements. Here, the network is composed of water molecules arranged in 5, 6, and/or 7-membered rings bound by H-bonds. This is the structure typical of liquid water at room temperature (Fig 2a). At room temperature (300K; 26.8C), however, the average lifetime of a H-bond is only 1ps opposed to 180ps at 230K.

As freezing begins (still at 230K), polyhedral structures with long lasting (>2ns) H-bonds begin to form (Fig 2b). A structured nucleating point moves around to exchange H-bonds with surrounding water molecules until it becomes anchored. Upon expansion of the nucleus, the network becomes mostly 6-membered rings. Once freezing is complete, water takes an ordered “stacked honeycomb” structure (Fig 2c). Ice is essentially a lattice formed by repetition of the tetrahedral hydrogen bonding pattern.

Figures removed for copyright reasons.

(2a)

(2b)

(2c)

The exact temperature at which these different conformations occur depends on pressure and volume.

Why is hydration important in keeping protein soluble? Why is the structure of water very important in the formation of protein secondary structures and protein-protein interaction?

Polar molecules on the surface of a protein keep a protein soluble. Water molecules surround each ion or polar molecule on the surface, keeping it in solution. Exposed polar residues form H-bonds with surrounding water molecules (C=O is a 2H-bond acceptor, NH<sub>2</sub> is a 2H-bond donor). Due to this H-bonding, the first hydration shell is 10-20% denser than the bulk water and probably keeps proteins sufficiently separated so that they remain in solution<sup>1</sup>. In addition, the first hydration shell is highly ordered, containing a wide range of non-random hydrogen bonding as seen by X-ray analysis. Very close to the surface of a protein (within 15Å), there are significant differences in the direction of water diffusion. The organization of water surrounding proteins creates preferred diffusive routes and favored conformational changes and interactions.<sup>2</sup>

Proteins are inactive in without water, therefore, hydration is crucial for the three-dimensional structure to have activity. Water molecules can compete with internal hydrogen bonding within a molecule, disrupting secondary structures like  $\alpha$ -helices. This exchange of internal H-bonds for water H-bonds causes molecules with internal H-bonds to dissolve readily in water. Water also acts as a lubricant, facilitating the necessary peptide amide-carbonyl hydrogen bond changes. Water can also bring further away residues into contact by forming water bridges, linking secondary structures together. The electric properties and structure of water allow it to make these critical protein connections intra- and inter-molecularly.

Protein-protein interactions are mediated by the shared H-bond network between them. In the absence of ligand (or other protein), water can occupy contact sites, thereby preserving the site in a shape accessible to the ligand, reducing the activation energy. Releasing ordered water molecules from the surface of an enzyme and the substrate is energetically favorable. While separate, two proteins are enclosed in an ordered water shell. When binding occurs, some of this ordered water is released, making the enzyme-substrate complex thermodynamically favorable<sup>3</sup>. Water bridges, again, can also reinforce the connection between two proteins, as drawn in figure 3.

Figure 3

2. Please draw the structures of 20 L-amino acids and 4 nucleotides. Write their full names, three-letter code and one letter code. Please arrange amino acids in the order of hydrophobic, hydrophilic, polar and nonpolar groups. Explain briefly their side chain's tendency to interact with water. Please draw **all possible hydrogen bonds on the side chains of amino acids** and explain how they interact with water. **Draw all possible hydrogen bonds for the nucleic acid bases** (follow the books).

**Hydrophobic and Non-polar** I've listed these in order of increasing hydrophobicity according to the "Hydropathy index" which measures the tendency of an amino acid to seek an aqueous environment. Negative values are more hydrophilic, positive values are hydrophobic. In looking at different websites, and over four textbooks, there has been some discrepancy about some of the amino acids classification, so for this question I've used this more quantitative approach.<sup>3</sup> The H-bonds drawn indicate the interaction with water. In general, a lot of H-bonding with water results in being hydrophilic.

Glycine Gly (G) -4 Not hydrophobic per se (negative index) but considered non-polar	Proline Pro (P) 1.6 A preference for turn structures means that Prolines are usually found on the protein surface where they interact with water despite the positive index.	Alanine Ala (A) 1.8 somewhat hydrophobic.	Methionine Met (M) 1.9 Has a long aliphatic side chain.
Phenylalanine Phe (F) 2.8 Aromatic and very hydrophobic. Benzene rings have been known to act as weak H-bond acceptors.	Leucine Leu (L) 3.8 Very hydrophobic.	Valine Val (V) 4.2 Very hydrophobic.	Isoleucine Ile (I) 4.5 Very hydrophobic.

**Polar** Here I list both charged and non-charged residues, in order of increasing hydrophilicity. These are grouped generally as hydrophilic according to text books, but I've written some exceptions and reasons for being somewhat hydrophobic for some of them. They all, however, do interact with water by making at least one hydrogen bond.

<p>Cysteine Cys C 2.5 Small, neutral, considered somewhat <u>hydrophobic</u> due to CH<sub>2</sub>. Very rare H-bond via SH can form with water</p>	<p>Threonine Thr (T) -7 Polar. Has aliphatic methyl with reactive hydroxyl. Either <u>hydrophobic</u> or <u>hydrophilic</u>, but either way <u>polar</u></p>	<p>Serene Ser (S) -8 Polar. Has aliphatic methyl with reactive hydroxyl. Either <u>hydrophobic</u> or <u>hydrophilic</u>, but either way <u>polar</u></p>	<p>Tryptophan Trp (W) -9 Aromatic and very <u>hydrophobic</u>, but can form a single H-bond with water so is <u>polar</u>.</p>	<p>Tyrosine Tyr (Y)-1.3 Polar. Aromatic and can participate in ring stacking so is <u>hydrophobic</u>, but has reactive OH group and is <u>polar</u>.</p>	<p>Histidine His (H) -3.2 Polar/Basic. pKa near physiological pH (easy to move protons on and off changing from neutral to positive charge). <u>ambiguous</u> whether it prefers to be buried in protein core or exposed to water.</p>
<p>Asparagine Asn N -3.5 Polar/uncharged. Side chain does not ionise, but due to polarity can donate and accept hydrogen in water H-bonds</p>	<p>Glutamine Gln Q -3.5 Polar/uncharged. Side chain does not ionise, but due to polarity can donate and accept hydrogen in water H-bonds</p>	<p>Aspartate Asp (D) -3.5 Polar/acidic.</p>	<p>Glutamate Glu (E) -3.5 Polar/acidic.</p>	<p>Lysine Lys K -3.9 Polar/basic. Ionized amino group at the end of the chain is <u>extremely polar</u>. Part of the side chain nearest to the backbone is long and hydrophobic.</p>	<p>Arginine Arg R -4.5 Polar/ Basic. Ionized amino group at the end of the chain is <u>extremely polar</u>. Part of the side chain nearest to the backbone is long and hydrophobic.</p>

**Nucleotides**. All four nucleotides undergo significant H-bonding with water. In the polymerized state (DNA) some of the hydrogen bonds with water are replaced by hydrogen bonds between bases. In addition, the bases are able to stack due to their ring like structure. In DNA, hydration is very strong around the phosphate groups running along the inner edges of the major groves.

Cytosine C (Deoxycytidylate)	Adenine A (Deoxyadenylate)
Thymine T (Deoxythymidylate)	Guanine G (Deoxyguanylate)

3. Describe in detail the basic characteristics of several secondary structural construction motifs and explain their functions in determining the structures of protein and protein-materials.

The  $\alpha$ -helix and  $\beta$ -sheet are secondary structural motifs “designed” by nature to be used mainly in the hydrophobic core of protein structures. By optimizing hydrogen bonding between N-H and C=O regions of different amino acids, these very polar regions are neutralized, making it possible for the polar backbone to exist in the hydrophobic core.

### a) Alpha-helix

Intra-chain hydrogen bonding is utilized to orient the polar chemical groups (C=O and N-H) that are part of the peptide bond. Every peptide bond (except those at the ends, which are polar and therefore almost always at the surface) participates in H-bonding, which is why the structure is so stable. Each successive turn binds to adjacent turns using three or four H-bonds. The cyclical pattern is such that the C'=O of residue  $n$  acts as the H-acceptor from N-H of residue  $n+4$  (figure 4).

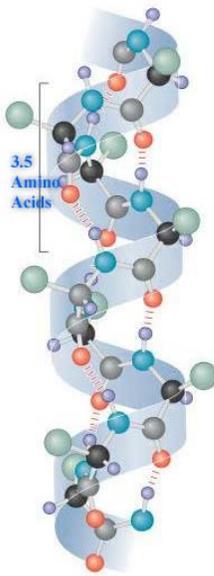


Figure 4

Courtesy of U.S. NIH.

The planes of the rigid peptide bonds lay parallel to the long axis of the helix. In a helical conformation, the relationship of one peptide unit to the next is the same for all alpha-carbons. This means that the dihedral angle pairs “phi”  $\phi$  and “psi”  $\psi$  are the same for each residue. The amino acid residues are strictly confined to conformations with  $\psi$  between  $-45$  and  $-50$  and  $\phi = -60$  for  $\alpha$ -helices, this is a shallow window (see question #4 for more information on  $\psi$  and  $\phi$ ).

Wound around an imaginary axis, the R groups protrude outward from the helical backbone at  $90^\circ$  angles for low steric hindrance, as seen for the purple “R” groups in figure 4. The twist found in all proteins is right handed due to the natural conformation of L-amino acids. One complete

turn takes a non-integral number of units, which “turns” out to be exactly 3.6 residues. Therefore, each amino acid contributes  $100 (=360/3.6 \text{ residues})$ . The rise caused by each amino acid is  $1.5 \text{ \AA}$ , and therefore there are  $5.4 \text{ \AA}$  per helical turn  $(=3.6 \times 1.5 \text{ \AA})$  along the long axis. This is why you see a bright spot every  $5.4 \text{ \AA}$  using X-ray crystallography. The average length seen in proteins is about 10 residues ( $\sim 3$  turns) per helix.

Since all the peptide units are aligned in the same orientation, H-bonds point in the same direction (figure 4). Individual dipole moments within peptide units therefore point in the same direction, resulting in a net dipole for the helix (positive at amino end, negative at carboxy end). This is one reason negatively charged groups like phosphate ions bind the amino ends of  $\alpha$ -helices.

### (Describe possible helices from different amino acids, hint: the book)

The properties of the different amino acids have a great effect on the stability of resulting helices. Here are five major stabilizing/destabilizing forces:

(1) Electrostatic repulsion between the side groups of successive amino acids destabilizes helices. For example, a long stretch of positively charged (Lys or Arg) or negatively charged (Gln or Arg) amino acids would repel each other so strongly that they overcome the stability of the H-bond structure.

(2) The bulkiness and shapes of certain amino acids (Asn, Ser, Thr, Cys, Tyr) are destabilizing.

(3) The interactions between sidegroups 3-4 residues away are particularly important. A Lys(+) residue 3-4 positions away from Gln(-) could create an ion pair. Aromatic side chains 3-4 positions away can form hydrophobic interactions, thereby stabilizing the helix.

(4) In general, Pro and Gly residues destabilize helices. Pro residues result in destabilization due to kinks from no rotation around the N-C $\alpha$  bond and no free N-H for H-bonding. Gly residues, on the other hand, have too much conformational flexibility, and long chains of Gly tend to create coiled structures instead of helices.

(5) The charge of residues at the ends of the helices can affect the net dipole. Synthetic peptides (ie DAR) where the charges of side chains at the ends oppose the net dipole (put neg charged D at + charged N terminus, and pos charged R at -charged C terminus) result in extremely stable  $\alpha$ -helices.

Although they are rare, other helices are possible by modifying the H-bond pattern as described for  $\alpha$ -helices.

Figure removed for copyright reasons.

Figure 5

**b) 3-10 helix.** This secondary structure is less common than the  $\alpha$ -helix, but can sometimes be found at the end of an  $\alpha$ -helix. Unlike the  $n, n+4$  hydrogen bonding pattern observed in  $\alpha$ -helices, 3-10 helices display a pattern where the backbone C=O residue  $n$  H-bonds to the HN of residue  $n+3$ . This results in 3 residues/turn, which is tighter than the  $\alpha$ -helix and so the resulting coil is thinner/skinnier (radius = 1.9Å vs 2.3Å  $\alpha$ -helix)(see figure 5). This tightness may result in van der Waals contact across the helical axis, and repulsion, one reason for its rarity.

The 3-10 helix has 10 atoms between the H-bond donor and acceptor. Like the  $\alpha$ -helix, all amide protons point toward the N-terminus (down) and all carbonyl oxygens point toward the C-terminus (up). Dipoles, however, are not perfectly aligned (~30 degrees off) making it less stable. Alignment does occur, however, with the side chains on the outside, creating steric interference- another cause for destabilization. The pitch of the 3-10 helix is longer than that of  $\alpha$ -helices (6Å/turn vs 5.5Å/turn), so the H-bonds must span a longer distance. For 3-10 helices,  $\phi=-74$ ,  $\psi=-4$  which are at the edge of the allowed minimum energy region of the Ramachandran map.

### **c) $\pi$ helix**

The  $\pi$ -helix is extremely rare.  $\pi$ -helices take an  $n, n+5$  repeating H-bond pattern. The resulting helix is “fatter” with a radius of 2.8Å, which is large enough to create a hole preventing van der Waals contact, but too small for water to enter. The hole can be seen in Figure 5. One  $\pi$ -helix is sometimes found at the ends of regular  $\alpha$ -helices but  $\pi$ -helices longer than a few  $n, n+5$  H-bonds are not found. The  $\phi$  and  $\psi$  angles (-57.1, -69.7) lie at the very edge of an allowed, minimum energy region of the Ramachandran map. The angle between N, C $\alpha$ , and C=O must be larger than 114.9, which is a strain from the normal tetrahedral angle of 109.5 degrees.

### **e) beta-sheet**

Beta strands are usually 5-10 residues long, with  $\phi$ ,  $\psi$  angles almost fully extended, but well within the broad structurally allowed region of the Ramachandran plot (-139, 135).  $\beta$ -strands have a rise of 3.4 Å/residue where the backbone is extended into a zigzag rather than helical structure. These zigzags, or  $\beta$ -strands, can be placed adjacent to each other to make a pleated structure, where the  $C\alpha$  atoms of adjacent strands are 350 pm (0.35 nm)

Figure removed for copyright reasons.  
See Fig. 6.10 in Garrett, R. H., and C. Grisham.  
*Biochemistry*. 2nd ed. Pacific Grove, CA:  
Brooks/Cole, 1998.

Figure 6

apart. Depending on if the orientation of adjacent strands is the same (both  $N \rightarrow C$ ) or opposite ( $N \rightarrow C$ ,  $C \rightarrow N$ ) sheets are either parallel or anti-parallel, respectively. The R groups from sequentially adjacent amino acids within a strand protrude in opposite directions, creating another zigzag if looked at in another plane (see the yellow "R" groups in figure 6). Unlike  $\alpha$ -helices that H-bond amino acids 4-5 positions away in linear sequence,  $\beta$ -strands connect to regions (linearly) far away, sometimes not even in the same chain. Hydrogen bonds are formed between the NH and C=O of residues in physically adjacent strands. In both parallel and antiparallel sheets, all possible main chain H-bonds are formed except for the strands on the ends. The cumulative effect of multiple hydrogen bonds arranged in this way contributes to the sheet's stability and structural rigidity and integrity. The sheets formed by joining strands is not planar, in fact they have a right handed twist.

There are several structural patterns that arise from the connections between separate strands. A series of antiparallel strands often form a distorted barrel within the interior of a protein. These are divided into three main groups: up-and-down barrels, Greek key barrels, and jelly roll barrels. A novel fold, the beta-helix, occurs as the result of repeating parallel beta strands.

The side chains from the amino acid residues found in a  $\beta$  sheet have important interactions. Many of the adjacent side chains on one side of the sheet may be hydrophobic, while many of those adjacent to each other on the alternate side of the sheet are polar or charged (hydrophilic), resulting in an alternating hydrophobic/hydrophilic pattern. Amino acids containing a " $\beta$ -branch" have preference for the  $\beta$ -sheet. For example, Ile, Thr, and Val have branched  $\beta$  C's, but Leu does not so Leu is preferred in  $\alpha$ -helices. Likewise, small amino acids like Gly, Ala, and Ser are also preferred.

### **d) beta-helix**

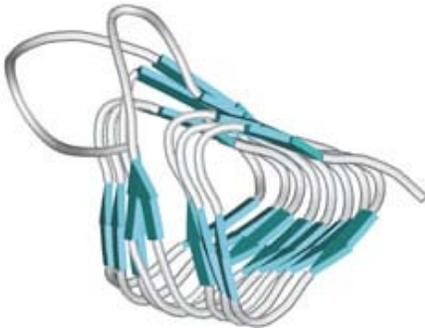


Figure 7

Courtesy of Oak Ridge National Laboratories.

$\beta$ -helices have a completely different underlying structure than the  $\alpha$ -helix and its derivatives described earlier. Instead,  $\beta$ -helices use short  $\beta$  strands (described in the  $\beta$ -sheet section) as a unit (not necessarily repeating). One turn of the  $\beta$ -helix has two or three successive  $\beta$  strands, all oriented in the same direction, and can be thought of as a rung. These strands align and form either two or three parallel beta-sheets, where the core is completely filled with side chains.

A typical helix contains from 7 to 16 triangular rungs, which twist around gradually to the right. An example of a three-stranded  $\beta$ -helix is depicted in Figure 7. Each coil of the helix represents a structural repeat that can be recognized from sequence information. Each rung contains

another landmark: a very short turn, usually only two amino acid residues long, called the T2 turn.

The  $\beta$  helix can create long, strong spikes that may be used by bacteria and allergens for attaching to or penetrating cell membranes. These structures are generally flexible and large.  $\beta$ -helix sequences often have repeats separated by Gly, Ser, Asn, Asp, or Pro rich sequences. Hydrophobic or ionic interactions between side chains within the interior probably contribute to this structure's stability. The interior is often filled with Leu residues, which separate Ca ions coordinated by interior Asp residues. So, although the overall structure is generally hydrophobic, nature has found a way to contain polar Ca ions within its core.

### **f) turns and loops**

These are found between strands of  $\beta$ - sheets, and/or  $\alpha$ -helices and necessary for protein structure because most proteins contain multiple  $\alpha$ -helices and/or  $\beta$ -sheets. Loops come in various shapes and sizes, and are always on the surface of proteins, exposed to solvent. For this reason (unlike  $\alpha$ -helices and  $\beta$ -sheets), the C=O and NH regions H-bond to water, not each other. The most common residues found in hairpin loops and reverse turns are Asn, Asp, Ser, Gly and Pro. These are all short chain amino acids and they are all hydrophilic since they are exposed to solvent. Gly and Pro are unique in reverse turns due to the conformational flexibility of Gly and Pro because of its inability to form hydrogen bonds.

The importance of loops, especially due to the fact that they are hydrophilic, is that they are an integral part in the binding site and active site of enzymes. For example, in  $\alpha/\beta$  barrel domains the active site is formed by 8 loops that connect sequential  $\beta$ -strands and  $\alpha$ -helices. Active sites are also formed by loops connecting to  $\alpha$ -helices at the edges of  $\beta$ -sheets, creating crevices ( $\alpha/\beta$  twisted open-sheet). Large loops can change conformation depending on the microenvironment, forming open/active or closed states, shielding reactive groups. Smaller loops, hairpin, and reverse turns are used to connect adjacent anti-parallel  $\beta$ -sheets. Flexible Gly residues are often found in these since such a sharp turn is made in so few amino acids.

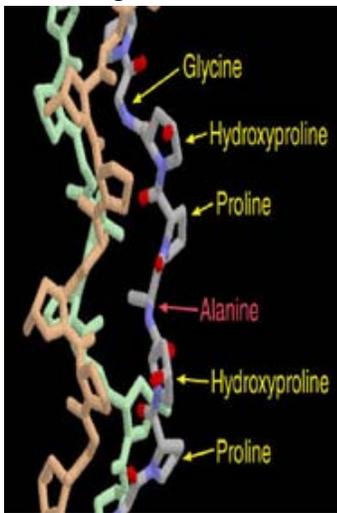


Figure 8

### **g) collagen helix.**

Collagens are generally extracellular structural proteins involved in formation of connective tissue structure. The collagen super helix is a right handed triple helix where each strand (~1400 amino acids) consists of left-handed helical strands with the repeating sequence "Glycine-Proline-Proline" as seen in figure 8. Proline is often replaced by the hydroxylated form, hydroxyproline (Pro-OH). Since Gly is so small and flexible, these residues form tiny elbows packed inside the super helix. Since Pro is by nature bent, Pro and Pro-OH efficiently bend the chain back around the helix and provide elasticity and strength due to the limited conformations it can assume. Tight wrapping of the chains in the collagen triple helix provides tensile strength that is greater than that of steel wire across an equal cross section. The three chains are H-bonded to each other where NH groups of Gly are the donors, and C=O groups on the sister chains are H-bond acceptors. The OH group of Pro-OH also

participates in H-bonding. In addition, water is involved in both inter and intra-chain H-bonding between Pro-OH residues and peptide C=O and N-H groups. Unlike the  $\alpha$ -helix

and  $\beta$ -sheet, the collagen helix does not involve every single main chain C=O and N-H group in intra-molecular hydrogen bonding. Collagen triple helices have only one such direct bond per tripeptide repeat.

The rise of the collagen helix (superhelix) is 290 pm (0.29 nm) per residue and residues have the angles  $\phi = -70$  and  $\psi = 160$ . This means residues are not quite as extended as in  $\beta$ -strands, but not as curved as the  $\alpha$ -helix (only 3 residues/turn).

See question 6 for more information on collagen structure.

4. Describe the peptide bond formation and its structure in great detail.

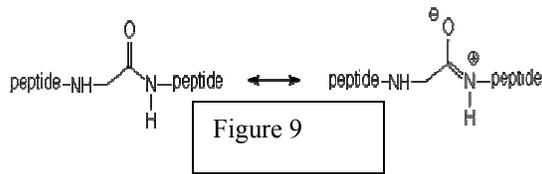


Figure removed for copyright reasons.

Amino acids are linked linearly through peptide bonds. These bonds are formed via a dehydration synthesis reaction between the carboxylate-end of the first amino acid with the amino-end of the second amino acid.

Peptide bonds are strong and fairly rigid due to a resonance structure in which there is a double bond between the amide nitrogen and the carbonyl carbon (figure 9). This planarity is an important feature to peptide bonds. Alpha carbons link these

rigid peptide units by covalent bonds. Thus, there are only two bonds within the peptide backbone that have rotational variability. The whole plane may rotate around the N-C $\alpha$  bond ( $\phi$  angle) or C-C $\alpha$  bond ( $\psi$  angle). C $\alpha$  is the carbon atom connected to the R group. Mathematically, phi ( $\phi$ ) and psi ( $\psi$ ) are the dihedral angle which is defined as the angle between the point (C $\alpha$ ) at the end of a 4-point sequence and the plane (peptide plane) occupied by the other three points. It is preferred to keep bulky R groups in a trans- and staggered formation to avoid steric hindrance.

Can you deduce other structure formations based on the peptide bond structure?

In a peptide,  $\phi$ - $\psi$  angles are restricted to certain ranges such that angle combinations are not infinite. A plot of their allowed distribution is called the Ramachandran plot. As it turns out, peptide bonds limit rotational freedom to essentially four understood structures ( $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -helix, and turns). This restricted freedom allows one to deduce possible structure formations. These possible structures made from defined bond angles are described fully in question 3.

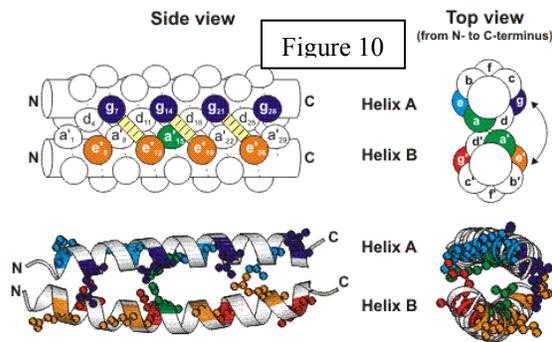
Another important property of peptide bonds is their ability to form H-bonds. These H-bonds are crucial for the formation of secondary structure in proteins, such as the  $\alpha$ -helix and the  $\beta$ -sheet. This involves the hydrogen on N (which has a significant partial positive charge due to resonance) forming a "bond" with a lone pair from a nearby carbonyl oxygen.

5. What is the characteristic of coiled-coil? Describe two, three and four strands of coiled coils. Where do you find such coiled coils in nature and can use improve the coiled-coil based materials?

A single  $\alpha$ -helix is not by itself very stable, and so  $\alpha$ -helices pack together through hydrophobic interactions between side chains to form stronger structures. Coiled coils consist of two or more right-handed  **$\alpha$ -helices** wrapped around each other in a left-handed supercoil. Induction among side chains causes compression, so that there are 5.1Å/turn instead of 5.4Å/turn in a single  $\alpha$ -helix. This reduces the number of amino acids per turn from 3.6 to 3.5. This means every seventh residue (2 turns) will be found at the same relative position along the helical axis. The types of interactions (and often specific side chains) that occur with amino acids 7 residues apart often repeat. Amino acid sequences containing coiled-coils can be identified with this characteristic repetitive “heptad sequence pattern”, where each position is labeled a-d. The top view in figure 10 shows the positions of the seven residues in each helix, and the side view demonstrates how they interact between helices.

Interhelical assembly is provided by the hydrophobic core “a” and “d” position sidechains, and hydrophilic (often charged) “e” and “g” resides. Coiled coils can form homodimers or heterodimers, form two-, three- four- (or five-) stranded complexes, and they can align in a parallel or an antiparallel orientation. (think of a “pull and peel” twizzler, and the sticky part where individual strands ( $\alpha$ -helices) meet. This sticky line can be thought of as a series of “a” and “d” residues that are hydrophobic/sticky, connected to each other.) Long, flexible dimers are often found in fibers, whereas shorter coiled-coils can be found in globular proteins (like myoglobin), and DNA/RNA binding proteins.

In a two-strand coiled-coil (like in leucine zippers of DNA binding proteins), position “d” (often Leu/Val/Ile) and position “a” are hydrophobic. Outside the



Courtesy of Katja Arndt. Used with permission.

hydrophobic core, salt bridges can form between “e” and “g” position residues with charged side chains (often Glu/Lys, Asp/Arg). These salt bridges can be seen in the yellow bonds depicted in the side view of figure 10. These ionic interactions define the relative chain alignment and orientation. Each side chain in the hydrophobic region (position d) of one strand is in direct contact with four side chains of the second strand, two at

position a, one at d, and one at e. This is called the “knobs in holes” model for fitting coiled coils together.

Three stranded coiled-coils form a hydrophobic a-d core as well, except the position “a” interacts with position “d” of one strand, and position “d” interacts with position “a” of a different strand. The core is thereby composed of 3 “d” position residues, and 3 “a” position residues per repeat.

The four-helix bundle is a complete domain structure where four  $\alpha$ -helices are arranged almost parallel along their helical axis. Water is excluded from a hydrophobic core created by packing hydrophobic side chains of all four strands. Helices often pack according to the “ridges and groves” model. Ridges are formed on the surface of  $\alpha$ -helices by helical rows of side chains separated by grooves. Ridges and groves are usually 4 residues apart in four helix bundles. There are, however, instances where dimers packed according to the “knobs in holes” model create four-helix bundles where the dimers are packed according to “ridges and groves.”

Coiled-coils are found in a variety of natural structures. DNA and RNA binding proteins often contain coiled-coils. For example, GNC4 is a dimer held together by a leucine zipper (described earlier). The N-terminal regions split, and fit into the major groove of DNA, where specificity is determined by the side chains of this region. Coiled coils are often found in membranes. Multiple  $\alpha$ -helices make rings where the outside of the structure is hydrophobic, and the interior is hydrophilic to accommodate soluble ligands. In the case of light harvesting complexes, these helices create a delicate arrangement of electron transfer reactions by providing a stable scaffold.

In mammals, almost 100% of the dry weight of hair, wool, nails, claws, quills, horns, hooves, and the outer layer of skin consist of  $\alpha$ -keratins. A-keratins are part of the family of intermediate filaments, and are formed by parallel coiled-coils. Super coiling amplifies the strength of overall structures like coiled strings can form a strong rope. Coiled-coils can form long flexible fibers, as found in hair and wool. Hair cells are elongated, holding long  $\alpha$ -keratin coiled-coils. Bonding is within one  $\alpha$ -helix, with disulfide bonding between helices. Disulfide bonds lock neighboring helices together during stretching process (intra-molecular coupling), that go back to their original position upon relaxation. Tri-helices wind into fibers assembled from 11 such strands, where 2 interior strands are surrounded by 9 on the outside (“9+2” configuration). Keratins are also found in nails and horns, but higher order structure is different. In horns, which have the strongest/toughest  $\alpha$ -keratin structure, 18% of the amino acid residues are Cys. Therefore, much cross linking accounts for this additional strength. To build a very tough coiled-coil, I would incorporate many Cys residues.

Muscle fibers (myosin) contain coiled-coils. Myosin molecules form coiled-coils of two  $\alpha$ -helices. A single coiled-coil is not that strong, but once they pack together, very strong rope-like materials can be formed. In Intermediate filaments, 2 parallel  $\alpha$ -helix monomers form stable coiled-coil dimer with 50nm tails. Two staggered antiparallel dimers form tetramer proto filaments. These tetramers take on 7+1 rope-like fiber conformation.

One could take advantage of the interactions that occur between two coils, and optimize them for specific material needs. For example, if you wanted coils that could come apart at certain pH and salt concentrations, you could take advantage of the e/g residues and their electrostatic interface. For example, if you put all negatively charged Glu (E) in one helix at position “e”, and all Lys (K) at position “g” (or “e” if you’re building a trimer), I suspect you would get a stable structure. If, however, you used Lys (K) at both positions e and g, then it would not be as stable because the same charges would repel each other at physiological pH. You could use the pI of the side groups to your advantage, however. If you used a very high pH solution, you may be able to get away with pairing Lys-Lys (above the pI=9.4, no net positive charge.) Likewise, you

could pair Glu-Glu at very low pH (<4) without repulsion. One could take advantage of these properties to design a coiled-coil that changes affinity with pH.

Within the hydrophobic core, you could use space-filling models to predict what residues would fit best and tightest within the hydrophobic region. This would be like a 3D puzzle, and you could logically choose what side chains fit best. If you wanted a more loosely-packed coil, you might use large, bulky groups within the core, such as Trp.

Another way one could manipulate coiled-coils is by adding Cys residues at specific locations. If inserted within the core, this would provide alignment between supercoils. On the other hand, by adding such “bridges” at the N or C termini of helices, one could imagine connecting two or more sets of coils together end to end.

6. Explain the significance of the high-resolution crystal structure of collagen. What have you learned from the single crystal structure? Explain why glycines and prolines are important in forming the structure of collagen?

Collagen is an abundant protein in the human body (~30% human protein content). It is a key ingredient to natural materials with functions as diverse as in tendons, skin, the organic matrix of bone, cartilage, and even the cornea of the eye. It makes sense to focus on collagen's structure as a source for useful new materials for tissue engineering, regenerative medicine, and even cosmetics. Therefore, the high resolution (1.85Å) structure of collagen was rather momentous. A fiber diffraction model was proposed in 1981, but this did not give information on the specifics of side-chain interactions, hydrogen bonding among chains, or the importance of water. Fibers are difficult to get X-ray crystal structures on. The regular packing of molecules form crystallites, but these ordered regions are separated by much less ordered regions with less regular packing. It wasn't until 1994 that a single crystal structure was determined by mutating a Gly residue to Alanine.<sup>4</sup>

The collagen super helix is a right handed triple helix where each strand (~1400 amino acids) consists of a left-handed poly-proline II helix with the general repeating sequence "Gly-Pro-Pr(or Pro-OH)" The positions occupied by Pro and Pro-OH are more variable, so this repeating triplet is often noted as Gly,X,Y. This variability is evident in the innumerable types of collagens found in nature. There are more than 30 variants of collagen, each with a slightly different sequence, and entirely different function. The amino acid composition of collagen directly relates to its unique structural constraints. Overall, collagens are 35% Gly, 11% Ala, and 21% Pro or Pro-OH.

Consisting of over one-third the amino acid content, the presence of Gly every three residues is critical to the collagen helical structure. Since Gly is so small and flexible, these residues form tiny elbows packed inside the super helix.<sup>5</sup> Such close packing requires this sterically small side chain at the central axis. In fact, a single substitution of Gly for a larger R groups (for example Ser or Cys) can result in lethal diseases due to the inability to form the triple helix where packing is so important. Osteogenesis imperfecta results in abnormal bone formation, and Ehlers-Dehlos syndrome results in loose joints. Both diseases can be lethal, and are the result of a single substitution. Even a single mutation to Ala, the next smallest amino acid, in a collagen fragment changes packing conditions, allowing the molecule to form crystals. This is consequently how the high resolution structure was finally determined. Information from this crystal structure is relevant to all collagen-like triple helices, and especially important in understanding molecularly how important Gly substitutions are.

Since Pro is by nature bent, Pro and Pro-OH efficiently bend the chain back around the helix and provide elasticity and strength due to the limited conformations it can assume. In the ER, Prolines are hydroxylated post-translationally to form the more stable hydroxyproline by prolyl hydroxylase. Prolines are also sometimes glycosylated to increase hydration properties, since it is very important to retain water in collagen.

Another important finding of the crystal structure was the formation of inter-chain H-bonds between the C=O of Pro/Pro-OH and H-N of Gly, keeping the helices together. Unlike the  $\alpha$ -helix and  $\beta$ -sheet, the collagen helix does not involve every single main

chain C=O and N-H group in intra-molecular H-bonding. Collagen triple helices have only one such direct bond per tri-peptide repeat.

The high-resolution structure has demonstrated the importance of both intra-molecular and inter-molecular bonds with water in keeping the structure stable. Triple helices are surrounded by a structured cylindrical hydration shell that becomes important in determining lateral separation in macromolecular fibril assemblies described later. Alpha water bridges form intra-molecularly within the backbone whereas  $\beta$  water bridges form inter-molecularly among backbones. Long inter-chain bridges form a network surrounding the triple helix. The OH groups on Pro-OH are also important for water-mediated H-bonding. Gamma bridges form between Pro-OH's -OH group with one backbone intra-molecularly, and delta bridges occur inter-molecularly between the three strands. The bond distances for the collagen fragment formed with a Gly  $\rightarrow$  Ala substitution were longer. In this case, four water molecules are inserted in the interior of the triple helix to mediate bonds between the three chains which are displaced from the Ala substitution. In fact, water molecules can fix inter-chain H-bonds wherever there is a point of disruption resulting in sub-optimal bonding distances, thus preserving the overall structure. This is why the positions X and Y are more variable.

The conformational flexibility/variability in collagen may be involved in recognition. Non-imino acids at position X and Y cause a change in helical pitch that could provide sites of recognition. Differences in superhelical symmetry along the collagen molecule change self-association and binding properties. Macrophage scavenger receptors have collagen triple helical structures as do extra cellular matrix proteins and collectins.

Collagen fibrils are supramolecular assemblies of the above described tri-helices. To form higher-order structures, individual collagen molecules ( $\sim 3,000\text{\AA}$  long and  $15\text{\AA}$  thick) are staggered and cross-linked for strength. Lateral spacing in these fibrils is due to the effective diameter of the hydration cylinder of the triple helices. Axial relation in staggering is from Lys, Hy-Lys, and His residues that provide cross links. Alignment and crosslinking is very tissue dependant, and results in different properties and tensile strength. (This information could be helpful in designing new materials with different crosslinking and hydration properties.) Accumulation of crosslinks in connective tissue results in rigidity and brittleness, which is often experienced by older people.

### How does collagens undergo self-assembly?

Within the ER, several modifications take place to the long procollagen peptide (hydroxylation of Hyp, glycosylation, and disulfide bond formation). The inter-chain disulfide bonds align the three procollagen chains such that they are in register to initiate formation of the triple helix. The central portions of the chain zipper from C- to N-terminus. Several forces make the triple helix a favorable conformation, including the hydrogen bonds between and among chains, and water bridges between and among chains. Helix formation probably initiates at parts where the Gly-X-Y repeat is fairly regular, and propagates down the chain forming H-bonds such that even regions with variability in X,Y positions conform to the triple helix structure.<sup>5</sup>

7. Describe the molecular structures of silkworm silk and the spider silk. Both proteins have a high content of glycine. How are these glycines arranged in both silkworm silk and spider silk structures?

Fig.11

Figure removed  
for copyright reasons.

The protein that makes up silk is called fibroin, and is produced by spiders and silkworms. Fibroin is rich in Ala and Gly residues, which forms  $\beta$ - sheets that pack closely with interlocking R groups. In Fibroin,  $\beta$ -sheets are oriented anti-parallel, and the small side chains interdigitate to allow very close packing for each layered sheet. As discussed with beta sheets, structure is stabilized by hydrogen bonding between all peptide linkages in each  $\beta$ -sheet as well as van der waals

interactions for stacked sheets due to the predominance of small side chains such as Gly, Ala, and Ser.<sup>6</sup> Figure 11 shows the stacking of  $\beta$ -sheets in silk fibroin. Notice the R-groups between sheets; there are alternating layers of Ala and Gly.

#### Silkworm silk:

Fibroin is bundled and arranged lengthwise to form a continuous fiber that can be very long made by silk worms. Fibroins overlap to form fibers as long as 2km! Fibroin from silkworm cocoons has two subunits held together covalently via disulfide bonds: Fibroin Heavy Chain (~350kDa) and Fibroin Light Chain(~25kDa). The structure of the heavy chain is composed of long fiber domains made up of repeats of the amino acid sequence (-GSGAGA-)<sub>N</sub><sup>7</sup>. These stretches are interrupted by short bulkier stretches in which the Ser is replaced by Tyr. Note, however, that both Ser and Tyr contain -OH groups and can form 3 hydrogen bonds.

More than 80% of silkworm silk forms  $\beta$ -sheets, as one would expect from alternating Ala and Gly residues. These  $\beta$ -sheets are oriented anti-parallel, and the small side chains interdigitate to allow very close packing for each layered sheet. Ala residues stagger on one side, and Gly on the other. Hydrophobic interactions occur between Ala (or Ser) residues between stacked sheets. It is not surprising to find such a high proportion of Gly residues because it is so flexible. As with spider silk described later, some silk worms also make  $\beta$ -sheets entirely from poly-alanine repeat sequences.

Looking at the primary sequence of the heavy chain from *B.mori* silk, one can see there is a pattern of hydrophobic and hydrophilic blocks. There are large hydrophilic blocks at the chain ends and six smaller hydrophilic internal blocks separating seven internal hydrophobic blocks. In water, these form micelle structures and fibroin globules upon higher concentration. Low concentrations of fibroin result in "random coils," high concentrations of fibroin without physical shear result in silk I, and physical shear or exposure to solvents like methanol result in silk II. Silk II is in the  $\beta$  form, which is composed of the described antiparallel  $\beta$  pleated sheets. In silk II, the peptide chains run parallel to the fiber axis. Sericin family proteins are utilized as glue to hold fibers together, and are hydrophilic.<sup>7</sup>

#### Spider silk:

Spider silk contains both  $\alpha$ -helices and  $\beta$ -sheets (but no sericin glue), which function as a block copolymer. The  $\beta$ -sheets form crystalline packing that provide strength, and the  $\alpha$ -

helices provide flexibility and extendibility. There are many types of spider silk (dragline, capture, frame, etc), and each contains a different proportion of  $\alpha$ -helices, 3-10 helices, and  $\beta$ -sheets, resulting in different strength, flexibility, and extendibility.<sup>8</sup>

Figure removed for copyright reasons.

Spiders undergo different gene expression for the different types of silk modules they make. Crystalline  $\beta$ -sheets (and weakly oriented sheets) exist in an amorphous glycine-rich matrix in dragline silk, which is organized as in figure 12.

Dragline silk from *N. clavipes* is characterized by polyalanine modules that form the crystalline beta sheets, providing strength.  $\beta$ - sheets are not only formed by poly-alanine, but by Gly-Ala repeats as well, just as in silkworm silk. The glycine rich matrix is formed by at least two Gly-rich motifs, which

Figure 12

occur in capture (flagelliform) silk as well. One such repeat is Gly-Pro-Gly-Gly-X, which occurs up to 63 times per ensemble repeat.<sup>8</sup> This sequence has been proposed to form type II  $\beta$ - turns. Furthermore, a long series of tandem  $\beta$ - turns likely creates a  $\beta$ -spiral, where each turn has two pentapeptide motifs. These beta-spirals act like springs, providing elasticity to flagelliform silk. Capture silk has > 200% extendibility, and a minimum of 43 contiguously linked  $\beta$ -turns. Dragline silk, on the other hand, has only 35% extendibility, and only has 9 beta-turns. So, although different types of silk share similar motifs, they occur in different amounts leading to different properties.

Another gly-rich motif is gly-gly-X, which follows the  $\beta$  spiral domains. The presence of so many glycine residues lifts restrictions from forming  $\alpha$ -helices, and 3-10 helices can be formed. Short arrays of Gly-Gly-X have been proposed to stop the current protein chain direction, causing detours as they terminate spiral regions.

## 8. Briefly explain X-ray crystallography used in determination of the molecular structure of biological materials.

Although useful for many biological applications, like visualizing a whole cell, light microscopes cannot discriminate structures at an atomic scale. In accordance with Bragg's law, in order to see an object, its size has to be at least half the wavelength of the light being used to see it. When any electromagnetic wave hits an atom, they make the electron cloud move around it. Rayleigh scattering occurs when waves re-radiate with the same frequency. The diffraction phenomenon is when these re-emitted rays interfere, giving constructive or destructive interference.<sup>9</sup> The smaller the wavelength, the smaller objects you can resolve. The wavelength of visible light (400-700nm), though small, is much bigger than an atom, making it invisible. X-rays, however, have a wavelength (.07-.15nm) short enough to resolve atoms of a protein crystal. The setup for X-ray crystallography is different than light microscopy in that there is no x-ray lens. Scattering from a single molecule is so weak that crystals must be used to amplify the signal by creating aligned arrays of molecules that scatter rays at the same angles. Atoms in the protein crystal scatter the x-rays from a narrow beam, which produce a diffraction pattern of discrete spots with different intensities when captured on x-ray film or a solid-state electronic detector. Each atom in the protein molecule makes a contribution to the intensity and location of each spot.

Since there is no lens to recombine x-rays to form an image, the diffraction pattern must be reconstructed mathematically using the Fourier transform. Since each spot is the result of interference of all x-rays with the same angle emerging from all atoms, it is difficult to extract information on individual atoms. One needs to account for the interference, positive or negative, with other emerging beams- this is the phase problem. One way to solve this problem is to compare data from native crystals with those containing heavy atoms like selenium. The number of electrons in the atom determines the intensity of its scattering of x-rays, and the amplitude of the spot is easy to determine. Heavier atoms scatter more effectively, so scientists diffuse different heavy metal complexes into the channels of pre-formed crystals, and compare data from those crystals without metal. This phase determination method is called "Multiple isomorphous replacement." Another similar method is the "Multiwavelength Anomalous Diffraction" method. Sometimes electrons absorb some of the energy from the x-rays, changing the scattering. This effect depends on the wavelength of the x-rays and type of atom. One can infuse a protein with seleno-methionine using recombinant technology, and take diffraction measurements using different wavelengths to solve the phase problem. Other methods of phase determination include "Molecular Replacement" where the structure of a known protein is used to initially phase the data, and "Direct Methods" which are computationally expensive and involve the fact that certain intensities can be generated by particular combinations of phases.

The quality of your crystal (degree of structural order) greatly determines the amount of information you can learn. Many proteins are very difficult to crystallize, and poor crystals result in errors and incorrect structures. Proteins are globular, large, spherical, ellipsoidal objects with irregular surfaces that can be difficult to pack into ordered arrays. In reporting X-ray crystallography data, one must report the "R-factor": a

measure for the deviations between the observed and calculated amplitudes of structure factors. A value of .4 is usually a starting point, .2 is acceptable, and .1 is usually required for publications.

It is important to realize that looking at a crystal structure does not necessarily tell you about the protein in its native state. Upon crystallization, a protein is frozen in time, and so the image cannot tell you about rearrangements and transformations that happen within a cell. NMR described later is more useful for this kind of information. X-ray diffraction, however, is more useful for proteins too large to be analyzed by NMR.

**Briefly explain both solution and solid state NMR in determination of the molecular structure of biological materials.**

**Solution State NMR:**

An advantage Nuclear magnetic resonance (NMR) has over X-ray crystallography is that samples do not have to be crystallized. NMR can be used to deduce information about protein dynamics such as folding, conformational changes, and interactions with other molecules. One disadvantage, however, is that molecules must be small (<200aa).

NMR is a phenomenon which occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to a second oscillating magnetic field. Some nuclei experience this phenomenon, and others do not, dependent upon whether they possess a magnetic moment, or “spin”. Most elements do have an isotope available that is detectable by NMR. An atomic nuclei’s nuclear spin causes its magnetic dipole to line up with respect to a strong external magnetic field. This equilibrium alignment can occur in parallel (low energy) or anti-parallel (high energy). To change alignment to an excited state, a short pulse (~10us) of electromagnetic energy of a suitable frequency (the Resonant Frequency, RF) is applied at right angles to the nuclei aligned in the magnetic field. The use of pulses of different shapes, frequencies, and durations, in specifically-designed patterns or “pulse sequences” allows the spectroscopist to extract many different types of information about the molecule. As the nuclei revert to their equilibrium alignment, they emit measurable RF radiation as a decaying sinusoid signal. Each atom has a different molecular environment, and emits a different radiation frequency (unless they are chemically equivalent). The electron density around each nucleus in a molecule varies according to the types of nuclei and bonds in the molecule. The opposing field and therefore the effective field at each nucleus will vary. This is called the chemical shift phenomenon. Chemical shift are taken in respect to a reference signal. Recording the chemical shifts from  $^1\text{H}$  atoms give you a one-dimensional spectrum. These can be very difficult to interpret because there are hundreds of  $^1\text{H}$  atoms even in small proteins.

Extra information can be obtained by adding more spectral dimensions. A two-dimensional NMR experiment involves a series of one-dimensional experiments. Each experiment consists of a sequence of radio frequency pulses with delay periods in between them. It is the timing, frequencies, and intensities of these pulses that distinguish different NMR experiments from one another. During some of the delays, the nuclear spins are allowed to freely rotate for a determined length of time known as the “evolution time”. The frequencies of the nuclei are detected after the final pulse. By incrementing the evolution time in successive experiments, a two-dimensional data set is generated

from a series of one-dimensional experiments. Peaks on the diagonal in the plots from such experiments correspond to normal 1D NMR. Interactions between nearby hydrogen atoms create peaks off the diagonal. One can use magnetization transferred by scalar coupling (J-coupling) in COSY (Correlation spectroscopy) experiments. Protons that are more than three chemical bonds apart give no cross signal (coupling constants are close to 0). Therefore, only signals of protons which are two or three bonds apart are visible in a COSY spectrum. Since hydrogen atoms in two adjacent residues are covalently linked through more than 3 other atoms, all COSY signals are from protons within the same amino acid. These signals can provide a fingerprint for each amino acid.

Another type of experiment that can give you more detailed information about an atom's 3D position is dipolar coupling (NOESY, Nuclear Overhauser Effect). In these experiments, signals are only observed for atoms closer than 5 Å apart. So, protons distant in amino acid sequence, but close in space due to tertiary structure can be correlated. This can provide information about interatomic distances and bond vector orientations.

#### Solid State NMR:

“Anisotropy” is the property of being directionally dependent. An anisotropic liquid is one which has the fluidity of a normal liquid, but has an average structural order among molecules relative to each other along their molecular axis. Liquid crystals are examples of anisotropic liquids.

In solution state NMR, spectra consist of a series of very sharp transitions due to averaging of anisotropic NMR interactions by rapid random tumbling. This is because Brownian motion averages out the orientation-dependent interactions between nuclear spins, allowing you to ignore them on the time scale of NMR. By contrast, solid-state NMR spectra are very broad, as the full effects of anisotropic interactions are observed. In a sample with limited mobility, such as crystal powders, molecular aggregates, and membrane vesicles, orientation-dependant interactions have a large influence on nuclear spins. Two directionally dependent interactions commonly found in solid-state NMR are the “chemical shift anisotropy” (CSA) and the “internuclear dipolar coupling”.

Sometimes the broadening effect seen in solid-state NMR is not desirable, and can therefore be limited by simulating samples tumbling in motion. High-resolution conditions can be obtained using “Magic Angle Spinning” (MAS) where one spins the sample about an axis oriented at  $54.74^\circ$  with respect to the external magnetic field. This allows one to create data similar to solution-state NMR. The residual line width of  $^{13}\text{C}$  nuclei under MAS conditions at 5–15 kHz spinning rate is typically in the order of 0.5–2 ppm, and may be comparable to solution-state NMR conditions. At faster MAS rates of 20 kHz and above, however,  $^1\text{H}$ – $^1\text{H}$  dipolar interactions can only be suppressed partially, leading to line widths of 0.5 ppm and above, which is considerably more than in optimal solution state NMR conditions.

Solid state NMR has been applied to: organic complexes, inorganic complexes, minerals, biological molecules, glasses, cements, food products, wood, ceramics, bones, semiconductors, metals and alloys, archaeological specimens, polymers, resins, and surfaces. Membrane proteins and amyloid fibrils related to Alzheimer's disease and Parkinson's disease are two applications where solid-state NMR complements solution-state NMR and X-ray crystallography methods.

## 9. Describe how nature makes various types of biominerals. How can you make one?

By creating an interface between materials derived from an organism (proteins) and inorganic materials found in the environment ( $\text{CaCO}_3$ ,  $\text{SiO}_2$ ,  $\text{Fe}_3\text{O}_4$ , hydroxyapatite), nature has effortlessly fashioned a foundation for building high-performance nanocomposite materials. I will discuss the major classes of biominerals and some of the proteins involved.

### Layers of the abalone shell:

Figure removed for copyright reasons.  
Source: Board on Army Science and Technology.  
"Opportunities in Biotechnology for Future Army Applications." 2001. Viewable online at <http://www.nap.edu/openbook/0309075556/html/42.html>

Figure 13

The structure of an abalone shell consists of layered plates of  $\text{CaCO}_3$  (~200 nm) held together by a much thinner (<10 nm) "mortar" of organic matter. Looking at figure 13, doesn't this look strikingly similar to bricks glued together with cement? There are at least three layers of an abalone shell.

The first layer, in contact with secretory cells, is an organic sheet of secreted proteins rich in Pro, Tyr, and Gly. This "nucleation protein sheet" provides a suitable surface for high-density nucleation of crystals with high order in the next layer. The second layer is the "calcite primer." This layer contains five specific polyanionic proteins (MW 14, 28, 35, 48, and 55kDa) that promote formation of only calcite polymorph.<sup>10</sup> Upon confluence of this layer, secretory cells begin to secrete a different combination of polyanionic proteins (MW 16, 20, 31 kDa) that promote formation of the "aragonite" polymorph layer in the flat pearl and shell.<sup>10</sup> This layer has highly oriented nacre tablet sheets surrounded by a specific organic protein matrix with a fracture resistance 3,000 times greater than aragonite alone.

Lustrin A is one protein within this matrix. Looking at the sequence one can recognize the presence of repeat regions separated by spacers, and a basic C-terminal domain. There are many Cys and Ser residues within the "GS" domain, implying the formation of intra- and inter-molecular disulfide bonds provided by Cys, and high solubility from H-bonding afforded by Ser. Therefore, protein combinations have the ability to change the properties of organic calcium carbonate crystals by affecting nucleation, phase, and growth dynamics. By interacting with minerals to form a highly regular laminate structure, proteins enhance the strength that aragonite crystals have by themselves, despite the fact that proteins compose only 2% of shell mass ( $\text{CaCO}_3$  98%).

### Light Harvesting Structures:

Mechanical strength is not the only property that results from the interaction between proteins and minerals. The same calcite crystals that provide skeletal construction in brittlestars also act as photoreceptors. In the arm ossicles of light-sensitive species, the calcitic skeleton forms a regular array of spherical microstructures that have the ability to focus light.<sup>11</sup> If we could understand the protein interactions that promote the architecture these domed structures, perhaps we could mimic and make synthetic optical networks.

Some sponges also have the ability to transduce light. The glass sponge *Euplectella aspergillum* has basalia spicules made of bio-silica fibers. These spicules have various layers with specific differences in the glass/organic composite (just like  $\text{CaCO}_3$  in different layers of the abalone shell). An organic axial filament (OF) runs through a smooth central cylinder (CC) embedded in a striated shell (SS) composed of a series of equally spaced layers.<sup>12</sup> The OF core consists of heavily hydrated and Na-doped silica, and there is less organic matter and/or hydration outward toward the periphery. Each layer has small silica spheres that gradually increase to four times in size from the CC region to the outer SS. Here, an organic “glue” holds the successive laminate shells together, resulting in an effective crack-arresting mechanism that would be quite useful in today’s commercial fibers that suffer from propagating cracks. This layered structural motif is highly effective at increasing the strength of composite materials, and should be considered when designing new materials.

Differences in composition between layers results in different refractive indices between the core (high-index) and cladding (low-index) just like commercial telecommunications fibers. Experimentally, it has been found that light can be transported through the core of the spicule (and a narrow region surrounding the OF), but not through the CC or SS.<sup>12</sup> Under more biologically relevant conditions with a higher refractive index difference between the spicule and the surroundings, spicules behave more like multimode light guides. Light traveling inside the fiber fills the entire cladding (not just the OF) confined by the outer surface of the fiber. Therefore, the optical properties can change depending on the surroundings.

Similar in function to the domed calcitic microlenses of brittlestars, these sponges have another interesting design element- terminal lens-like extensions located proximally and barb-like spines located along the spicule shaft. These lenses improve the efficiency of collecting light at the tips and spined regions.

#### The “Glue” and Proteins:

The interface between the organic and inorganic components of biomineral composites is not well understood for most systems. Since proteins are in effect the organic “glue” from the layered motif so often found in biomineralized materials, they are crucial to understand.

A family of proteins called “silicateins” have been identified in the silicon-rich spicules of another sponge, *Tethya aurantia*. These proteins have a periodically repeating substructure within the filament, and the most abundant one (silicatein alpha) has a cDNA sequence quite similar to the cathepsin L protease.<sup>13</sup> Thus, these proteins probably perform more than mere structural jobs; they are also involved in catalysis. Furthermore, silicateins have been found specifically to catalyze the polycondensation of silica from silicon tetraethoxide at neutral pH. Such polycondensation requires chemically harsh conditions industrially, and it is amazing that sea creatures can perform it under such mild conditions underwater.

Another class of proteins involved in silica-based biominerals are “silaffins.” These organic molecules are involved in regulating biosilicification by inducing silica to form uniform nanoparticles at neutral pH. Silaffins contain many repeats with Gly, Ser, Lys, Arg, and Tyr. Within the repeating units, there are clusters of basic residues (Lys-Lys and Arg-Arg) spaced regularly. Hydroxy amino acids (Ser and Tyr) are found in between the basic clusters. These proteins are ideal candidates for silica precipitation

because they are both cationic polymers and contain many hydrogen bonding groups (mostly Ser). Ionic and hydrogen bonding interactions are necessary with the surface of silica particles during this reaction.<sup>14</sup>

### Bone

The family of biominerals collectively called “bone” is quite diverse and complex both in structure and function. The basic building block, mineralized collagen fibrils, is shared within this wide family. All family members use collagen as the three-dimensional scaffold matrix, onto/into which minerals are formed. The mineral found in bone is dahllite, or carbonated apatite, which is a natural calcium phosphate. The proportions of water, collagen, and apatite vary between family members, and can differentiate function.

Bone structure has been described in terms of 7 hierarchical levels, four of which I will discuss, the first being the “major components”.<sup>15</sup> Dahllite is perhaps the smallest known biologically formed crystal, and has a plate shape with uniform thickness. The small size makes it difficult to analyze. Although dahllite is the only mineral found in bone, type I collagen is by no means the only protein. There are at least 200 non-collagen proteins that generally comprise less than 10% of the total protein content. Collagen fibrils (discussed in two earlier questions) are made of three polypeptide chains wound together in a triple helix. Fibril packing within bone is quite unusual. Fibrils (Level 2 structure) undertake a parallel, but staggered array pattern, creating channels where calcium and water often appear. This packing can be seen in figure 14, as well as the “hole zone,” or channels. Water, in fact, is a very important part of bone structure. The relative proportions of mineral, collagen, and water vary between bone types, where an increase in calcium generally replaces water.

The proposed arrangement of mineralized collagen bundles (level 3) is orthotropic (aligned with respect to platy crystal layers and the fibril axis). Bundles therefore form a platelet-reinforced composite. During the early stages of mineralization, crystals are kept separate from each other along the length of fibrils within separate channels (see figure 14). As the crystals grow, they compress the triple-helices and join together to form extended sheets. The process of crystal growth undoubtedly changes the mechanical properties of bone (or other materials) over time. Diffraction patterns from turkey tendon showed that in an area of 300nm, crystal layers in several neighboring fibrils are aligned in all three dimensions, and that orthotropic order may well extend to millimeter distances. This means that the mechanical properties are different in all three directions.

The diversity in fibril array organizational patterns is referred to as level 4, and at least four such patterns have been described. The first is “Arrays of Parallel Fibrils,” which is essentially an extension of the pattern described for level 3 to the millimeter size scale. The directional dependency (highly anisotropic) of parallel fibril arrays has the property that they can be aligned in different orientations to optimize for a specific mechanical function. One place that parallel alignment is found is in the tendons that attach to bone surfaces, where elasticity is desirable in one direction, and stiffness fixes the orientation.

Figure removed for copyright reasons.  
See Fig. 2 in Rho, J., et al. “Mechanical Properties and the Hierarchical Structure of Bone.” *Medical Engineering & Physics* 20 (1998).

Figure 14

Almost completely opposite to parallel fibrils is the “Woven Fiber Structure.” Here, fibril bundles have various diameters and are arranged in many different orientations. Both the matrix and the mineral are loosely organized. This structure is useful in healing and during development where quick growth is necessary.

“Plywood-Like Structures” are characterized by layers of parallel fibrils and/or fiber bundles of different orientations. This structure is the closest to the other laminated composites described earlier for shells and sponges. Sometimes the thickness of each layer can vary in a repeating nature, such as in lamellar bone. The rotation found in each layer is often regular, such that the resulting composite is rather isotropic on the macro scale, despite being anisotropic at the micro/nano scale for individual bundles. This property makes rotated plywood-like bone an ideal material to withstand compressive forces in many directions, unlike parallel-fibered bone.

The fourth pattern at level 4 is “Radial Fibril Arrays,” and this is the pattern found in dentin (inner layer of teeth). The fibril axis for intertubular dentin (ITD) runs perpendicular to the tubule direction, and all fibrils are within this plane. Figure 15a depicts the orientation of dentin tubules and odontoblasts. Fibril bundles are for the most part arranged tangentially to the tubule circumference, but oriented in all directions. Figure 15b shows how fibrils form around the tubules. Since collagen fibrils are confined to one plane, they are anisotropic. The extent of directionality is much less, however, when looking at crystal orientation.

Fig 15a

Figures removed for copyright reasons.

15b

There is a large portion of unoriented crystals, and even the crystal layers from neighboring fibril bundles are not always aligned. When designing a new material, one should take into account not only the directionality of the organic matrix, but also the alignment of the inorganic crystals that will fill it. These different orientations could have profound effects on the mechanics of new materials- for example yielding anisotropic fracture properties and isotropic elastic properties as with dentin.

By understanding the way nature constructs these strong, highly ordered materials of so many different functions, we can learn to design synthetic materials using similar strategies. For example, instead of using proteins scaffolds that attract calcite or silica ions for structural strength, one could utilize polypeptides that bind ions with different functionalities, such as semiconducting ions to build nano circuits. In addition, protein matrices can take the shapes of materials we are familiar with, like lenses in light-conducting materials to make nano optical-based devices. The multi-layered motif has been seen in many of these biomineralized structures, and could be useful in making new materials strong. One thing to keep in mind is the fact that crystals by nature grow, and the organic matrix one creates should take into consideration crystal growth, as seen in bone. Another thing to think about is the orientation of both the organic matrix, and the orientation of the crystals- they need not be the same.

If I were to design a biomineral, I would have to consider two things: what peptides I would try, and what minerals to combine them with. Since the minerals I would try are positively charged (probably Silica, Calcium, and maybe Iron or copper), I would need negatively charged amino acids to bind them. So, I would design sequences

with repeated Aspartic acid (D) and/or Glutamic acid (E) residues. I would also include hydrophobic groups to promote ordered assembly, and spacer regions as seen in some of the other matrices for biominerals. From alternating charged/hydrophobic residues, I would expect to form  $\beta$ -strands, with the charged groups on one side, and hydrophobic residues on the other. These would probably assemble into two-stranded sheets with the hydrophobic sides together, and the charged residues on the outsides available to bind minerals. I would try making sheets with different patterns on each side of the strands, and in different patterns.

Another strategy would be to make strong mineralized coiled-coils. To begin, I would need heptad repeat patterns. I could use all negatively charged residues at positions b,c,e,f,g, and hydrophobic residues at a and d to promote coiled-coil assembly, with ions attaching to the outsides. By adding Cys residues to the termini of the helices, coiled-coils could be linked together into a network. Another idea is to make  $\alpha$ helices with a single line running down the axis that would bind minerals, keeping the rest of the molecule neutral. Very long helices may make one-mineral-thick nano-wires if I used a conducting metal.

10. Use the knowledge you have learned in the class to design an interesting building structural motif (hint: use various repeating sequences). Please describe how do you know what you designed it what you have in real materials, how do you characterize them at each steps and what tools and instruments to use.

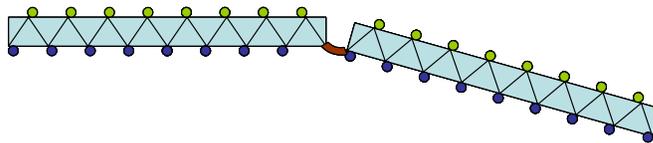
The new structural motif that I propose is the “bent  $\beta$  strand.” I think it would be interesting to make a peptide that would form a pair of stable  $\beta$ -strands separated by a bend induced by stiff Pro residues. The general scheme would be alternating hydrophilic/hydrophobic residues, a “bend” region, and another series of alternating hydrophilic/hydrophobic residues. Ultimately, I want to design the beta sheets such that the “R” groups on the top and bottom of the sheet promote stacking, similar to silk fibroins, via ionic and/or hydrophobic interactions. The resulting peptide may look something like an elongated “V”, as seen in the level I diagram in figure 16.

## Figure 16: “Bent Beta-sheet Structure”

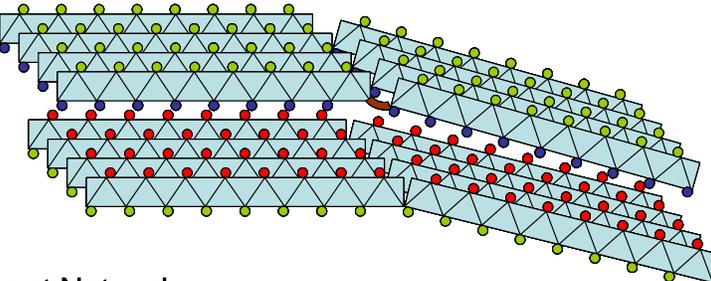
### Level I: Single two-part beta-strand

Hydrophobic R group= 

Charged R groups=  

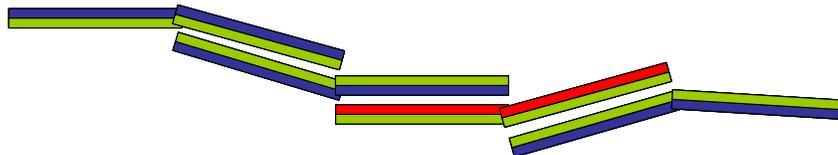


### Level II: aligned beta-sheet



### Level III: Stacked sheets

### Level IV: modular Stacked-sheet Network



## The Bend Region

Prolines would be ideal for the bend region because I've seen in the collagen helix they are ideal for twisting, and are not flexible. Prolines are far more structurally confined than the other amino acids because one of the dihedral angles ( $\phi$ ) is fixed at  $-65^\circ$ . Pro is rarely found within a periodic secondary structure such as  $\beta$ -sheets or  $\alpha$ -helices in globular proteins. Pro lacks one potential H-bond donor, and the  $\phi$  angle  $-65^\circ$  is not compatible with  $\beta$ -sheets. I would like to use this property to my advantage, and construct two segments of  $\beta$ -sheets (parallel or anti-parallel) connected by a bend induced by Pro.

B- turns are common in proteins, and sometimes contain Pro residues. Since my goal is not to make a monomeric two-stranded antiparallel sheet, I want to avoid having the Pro residue merely form a hairpin. In hairpins, Prolines are often accompanied by flexible Gly residues to make the full turn, so I will avoid adding Gly near Pro to make the "bend" as rigid as possible so that the backbone will not fold back on itself.<sup>17</sup> During analysis, I would run a PAGE gel, and if I see bands around the same molecular weight as the monomers under non-assembly conditions, then hairpin loops probably formed.

I would try various numbers of Pro repeats (1-4) to optimize the bend. Another strategy is to use modified Pro residues, or D-Proline such that hairpins don't form. One group was able to create H-bonded  $\beta$ -sheet ribbons using Z-ProNH(CH<sub>2</sub>)<sub>2</sub>NHPro-Z.<sup>18</sup> Perhaps by adding a covalent linker between two Pro's, I could create a structure where stable beta sheets could form on both sides of a Pro bend. The use of Trp (or other bulky residues) before and after the Pro region may allow better alignment or a stiffer bend. I would try small residues like Ala as well.

If the Pro residues do not create a convenient angle such that the residues on the other side of the bend line up, perhaps one side of the bend, "a" would align into a beta sheet, and side "b" would be splayed into non-aligned strands. I will use sequences that strongly promote sheet formation to avoid this.

### The $\beta$ -Sheets

I have decided to use 16 residues (8 hydrophobic, 8 hydrophilic) for each  $\beta$  strand because it has been demonstrated that 16-residue amphiphilic peptides self-assemble into stable  $\beta$  sheets.<sup>16</sup> For hydrophobic residues, I will try combinations of Ala, Ile, and Val. Ile and Val are  $\beta$  branched, and should promote  $\beta$  sheet formation. Alanine has been proven to form stacked sheets, as in silk, so I will try that as well. For charged residues, I will try Glu, Asp, Lys, and Arg. I will also try Thr as a neutral, polar molecule. The general scheme will look like this:

$\Phi$  = hydrophobic residue

$\Psi$  = hydrophilic (neutral or charged)

P = Proline

-----a----- "bend" -----b-----  
 $\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi$  P(n)  $\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi\Phi\Psi$

Since there will be two  $\beta$  sheet regions, peptides can be modular in nature. Different combinations can be used to build both sides of the bend. Possible sequences I could use for each strand are listed in the table 1 attachment. The proportion of different combinations in the peptide mixture would result in different binding properties between sheets, and therefore different macro structural properties. Figure 16 shows the structural levels I predict from the sequences in table 1.

Interactions between the “a” and “b” regions of different peptides could make interesting arrangements; I’ve drawn a few possibilities in figure 16, level IV structure. If each molecule is forced to bind to two other molecules at the hydrophobic regions, then many Y shaped arrangements may occur, resulting ultimately in a fractal system of sheets. An open network would probably form, depending on the angle between regions “a” and “b”. To make the arrangement even more interesting, instead of alternating hydrophobic with hydrophilic residues on both sides “a” and “b”, one of the sides could be replaced with hydrophobic residues alternating with positive or negative charges. This would create a case where only a complimentary side could interact with that strand. In this case, stacks of beta sheets could assemble, with either hydrophobic or ionic interaction between layers. Some of these type modules are listed in table 1, where D, E, K, and R are charged.

In general:

“a”= (-)Φ (-)Φ (-)Φ (-) could only interact with “a”= (+)Φ (+)Φ (+)Φ (+)

on the charged side of the beta sheet, or

“a”= (-)Φ (-)Φ (+)Φ (+) with “a”= (+)Φ (+)Φ (-)Φ (-),

where (-)=D or E, and (+)=K or R.

#### Peptide Analysis

The first analytical step I would take with my peptides is to make sure the peptides I ordered are pure, and of the correct molecular weight. This is relatively simple, and could be done using SDS-PAGE with a coomassie blue stain. If I see multiple bands, then my pure peptides may be contaminated with other byproducts from the synthesis process.

Next, I would need to try to promote assembly among the peptides by putting them under a variety of conditions. This is very important because apart from the primary sequence, the molecular environment is the most important factor in determining peptide conformation. Where the Prolines may completely destabilize the β sheet structure and form globules under one condition, another condition may promote stability. By simulating a membrane environment, different conditions can stabilize an α-helix containing Prolines<sup>19</sup>. Although the authors claim β-sheet structure does not change with the environment, by trying *many* different conditions I think I may be able to do something similar with β-sheets. I would try a variety of temperatures, salt concentrations, pH ranges, lipids, etc.

After assembly and during assembly, what I want to do is look at my peptide under AFM at many time points to see if my peptides created any kind of regular structure. I would put the peptides under different pH and temperature conditions to see if that changes macro assembly. I would look for the presence of globules, which would imply random coils, or a fiber structure as I would hope for.

There are many ways I envision this peptide would assemble, and a few techniques I could use to determine what actually occurs. One important method I would try would be Circular Dichroism (CD) Spectroscopy to analyze the morphology of the backbone. I would use the “far-UV” region (190-260nm), and look for a curve characteristic of a beta sheet (minimum ellipticity at 218nm, maximum at 195nm), which would look something like the red curve in figure 17. If a

Figure removed for copyright reasons.

Fig 17

portion of the peptide “a” or “b” does indeed assemble into a beta sheet, then I would expect to see something like the red line in the diagram. If, however, beta sheets do not form, I may end up seeing something like the green curve, with an ellipticity of zero due to irregularity of the peptide bond. If there is regularity in the tertiary structure as well, I may detect a signal at the near-UV region (250-300nm). I would try variations in the buffer pH, temperature, and even sonication or vortexing to see if I can detect differences at the secondary structure level as it assembles and/or disassembles.

Another technique I could use is Infrared Spectroscopy (IR). This method would probably give me similar information to CD, but is not as sensitive. Nevertheless, it would provide a good complement to CD. Since the stretching vibrations of amide N-H and C=O bonds are sensitive to H-bond geometry and secondary structure, there are peak absorbencies for different secondary structures. If my peptide makes  $\beta$ -sheets, I would look for a strong peak at  $1640-1620\text{cm}^{-1}$ . If I needed even more evidence for the  $\beta$ -sheet, I could use Raman Spectroscopy as well. Here, I would look for intensity from “Amide I” at  $1680-1655\text{cm}^{-1}$  and “Amide III” at  $1240\text{cm}^{-1}$ .

Although it would take longer, I could use NMR and X-ray diffraction to get very detailed information about the molecular assembly of my peptide, assuming CD, IR and Raman data all indicate the formation of a regular structure.

1. Rowe, A. J. Probing hydration and the stability of protein solutions--a colloid science approach. *Biophys Chem* 93, 93-101 (2001).
2. Makarov, V. A., Andrews, B. K., Smith, P. E. & Pettitt, B. M. Residence times of water molecules in the hydration sites of myoglobin. *Biophys J* 79, 2966-74 (2000).
3. Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Lehninger principles of biochemistry* (Worth Publishers, New York, 2000).
4. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* 266, 75-81 (1994).
5. Kramer, R. Z., Bella, J., Brodsky, B. & Berman, H. M. The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J Mol Biol* 311, 131-47 (2001).
6. Valluzzi, R., Winkler, S., Wilson, D. & Kaplan, D. L. Silk: molecular organization and control of assembly. *Philos Trans R Soc Lond B Biol Sci* 357, 165-7 (2002).
7. Jin, H. J. & Kaplan, D. L. Mechanism of silk processing in insects and spiders. *Nature* 424, 1057-61 (2003).
8. Hayashi, C. Y. & Lewis, R. V. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J Mol Biol* 275, 773-84 (1998).
9. Wikimedia. in Bragg's Law (October 2005).
10. Belcher, A. M., Hansma, P. K., Stucky, G. D. & Morse, D. E. First steps in harnessing the potential of biomineralization as a route to new high-performance composite materials. *Acta mater* 46, 733-736 (1998).
11. Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L. & Hendler, G. Calcitic microlenses as part of the photoreceptor system in brittlestars. *Nature* 412, 819-22 (2001).
12. Aizenberg, J. et al. Skeleton of *Euplectella* sp.: structural hierarchy from the nanoscale to the macroscale. *Science* 309, 275-8 (2005).
13. Shimizu, K., Cha, J., Stucky, G. D. & Morse, D. E. Silicatein alpha: cathepsin L-like protein in sponge biosilica. *Proc Natl Acad Sci U S A* 95, 6234-8 (1998).
14. Kroger, N., Deutzmann, R. & Sumper, M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* 286, 1129-32 (1999).
15. Weiner, S. & Traub, W. Bone structure: from angstroms to microns. *Faseb J* 6, 879-85 (1992).
16. Zhang, S., Lockshin, C., Cook, R. & Rich, A. Unusually stable beta-sheet formation in an ionic self-complementary oligopeptide. *Biopolymers* 34, 663-72 (1994).
17. Karle, I., Gopi, H. N. & Balaram, P. Infinite pleated beta -sheet formed by the beta-hairpin Boc-beta-Phe-beta-Phe-D-Pro-Gly-beta-Phe-beta-Phe-OMe. *Proc Natl Acad Sci U S A* 99, 5160-4 (2002).
18. Ranganathan, D., Makarov, V. A., Kumar, G., Kishore, R. S. K. & Karle, I. L. A unique example of a core-modified bis-proline peptide self-assembling into an infinite hydrogen-bonded -sheet ribbon: crystal structure of Z-ProNH(CH<sub>2</sub>)<sub>2</sub>NHPro-Z. *Chemical Communications* 3, 273-274 (2001).
19. Li, S. C., Goto, N. K., Williams, K. A. & Deber, C. M. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci U S A* 93, 6676-81 (1996).