#### **Protein Structure: Three-dimensional structure**

#### Background on protein composition: Two general classes of proteins

- Fibrous long rod-shaped, insoluble proteins. These proteins are strong (high tensile strength). Examples: keratin, hair, collagen, skin nails etc...
- Globular compact spherical shaped proteins usually water-soluble. Most hydrophobic amino acids found in the interior away from the water. Nearly all enzymes are globular... an example is hemoglobin

Proteins can be simple - no added groups or modifications, just amino acids

Or proteins can be conjugated. Additional groups covalently bound to the amino acids. The naked protein is called the apoprotein and the added group is the prosthetic group. Together the protein and prosthetic group is called the holoprotein. Ex. Hemoglobin

#### **Primary Sequence: Mutations**

Mutations in our DNA occur for different reasons:

- Mistakes during DNA replication
- Environmental DNA damage

The mistakes (e.g. a Cytosine to Guanine point mutation), will be observed also at the mRNA and amino acid level.

What do you think these mutations do to the overall structure of the protein? It's function?

#### **Peptide Backbone**

#### Rotatable peptide bonds

- Due to peptide bond (review)
- Trans conformation of a carbon is more stable and primary conformation
- Consider the consequence of this arraignment
- 10% of Pro are in Cis.

#### Limited rotation around a carbon

- Dihedral/torsion angles describe the rotation around the *a carbon*
- C-N Phi and C-C Psi
- The peptide bond is creates limited possibilities. •
- Torsion angles are defined by 4 points or planes
- Notice the side groups

#### Limited Options

conformation.

helices.

possible angles.

- Peptide bond limits rotation along all bonds in backbone creating a
- 3D spatial arrangement is determined by the limited rotational options
- R groups add limitations steric
- Amide H and Carbonyl O also impact rotation options
- R groups have NOTHING to do with peptide bond issue

#### Primary Secondary/Tertiary Ramachandran Plot FUNCTION Structure Structure Protein structure is defined by the Ramachandran plot shows the peptide plane and the 2 degrees of values of the torsional angles and freedom around the peptide bond allows prediction of the Two basic predictable forms of The angle will depend on the two secondary structure amino acids. The steric hindrance Alpha helix and beta pleated sheet between the functional groups -Other forms are specialized determine the angle of stability. (collagen) There are only three real means - Beta sheets, tight turns and beta of conformation the sheets and bulge Glycine and proline alter the Amide lone $\phi = 180^{\circ}, \psi = 180$

#### Levels of Structure

Type of Structure	: Defines:	Type of Bonds:	
Primary (1°)	Order of amino acids	Covalent bonds	
Secondary (2°)	Local structure (α-helix, β-sheet, loop)	Non-covalent interactions, disulfides	
Tertiary (3°)	Overall fold, 2° elements organize and compact—low surface to volume ratio	Non-covalent interactions, disulfides	
Quaternary (4°)	Subunit organization, dimers, macromolecular assembly	Non-covalent interactions, disulfides	



#### α-**helix**

Discovered by Linus Pauling

- 2 Nobel prizes.
- Discovered folding while sick in bed!
- Treated poorly as a result of his anti-nuclear stands 2<sup>nd</sup> NB prize
- Pushed high doses of vitamin C
- The helix is a right handed twist of the backbone notice when we are looking at this the side groups are NOT considered
- Notice where the amino acids are.
- Hydrogen bonding occurs between the carbonyl and the amino group four residues away. The bonding takes place within the same chain.
- A run of proline residues lead to breaking the helix structure. Wh
- Formation of α-helices are governed by **HYDROGEN BONDS**!
- One helix turn is 3.6 amino acid residues, and involves 13 atoms from the O to the H of the H bond
- For an a-helix of n residues, there are n-4 hydrogen bonds!
- Helix capping—the last 4 amide hydrogens/carbonyl oxygens cannot H-bond. Proteins compensate by folding other parts of the protein to facilitate hydrogen bonding.

H-bonds all point in the same direction

Amide bond has a **dipole moment**—cumulatively, the helix has a large dipole moment

#### β-Pleated Sheet

Notice that there are no "turns" pictured here—the sheet here is not a contiguous amino acid sequence!

- Side chains perpendicular to plane of the sheet
- Defined by a different hydrogen-bonding network
- Note: hydrogen bonds occur interstrand
- Parallel: Adjacent chains run in the same direction
- Bent H-bonds
  - Normally large, >5 strands
  - Hydrophobic side chains distributed on both sides of sheet

Anti-Parallel: Adjacent chains run in the opposite direction

- More extended H-bond conformation
- Can consist of 2 strands
- Hydrophobic on same side—alternates hydrophilic and hydrophobic in primary sequence



(a)

(b)

Also governed by H-bonding—do you see a theme here?

- Amino acid influences/requirements for turns
- Pro, because of it's cyclic structure and fixed φ angle drives the formation of these turns!
- Gly is also commonly found in turns-why do you think this is true?

#### Two Beta Turns

Reverse turns/ bends – four amino acids each stabilized by H bonding from backbone

- Type I and II differ at peptide bond orientation at turn
- Type II angle requires Glycine on second aa position ramachandran!



residues

#### "Random coils" are not random

The segments of a protein that are *not helices or sheets* are traditionally referred to as "random coil", although this term is misleading:

- Most of these segments are neither coiled or random
- They are usually organized and stable, but don't conform to any frequently recurring pattern
- · Random coil segments are strongly influenced by side-chain interactions with the rest of the protein



#### **Intrinsic Disorder**

Coupled folding and binding is the process in which an intrinsically disordered protein folds into an ordered structure concomitant with binding to its target. For example, the phosphorylated kinase-inducible domain (pKID) of CREB is unstructured when it is free in solution but it folds on forming a complex with the KID-binding (KIX) domain

Entropy cost of increasing order in ID proteins as the protein transitions from disorder to order – paid for (thermodynamic driving force) is a change in enthalpy of many weak binding forces between proteins



#### **Intrinsic Disorder**

Some regions or domains of proteins do not become structured (predictable secondary structure) until binding to another protein or target (lipid, carbohydrate, small molecule, substrate...).

- Identified as "missing electron density in crystal structures atom positions and backbone Ramachandran angels fluctuate
- Until binding target interacts, ID appears as a random coil
- Breaks some of the structure-function paradigm
- High in Gly, Pro and Ala, low in Cys and Asn
  - Order breaking aa and order promoting aa
  - Often low hydrophobic and high net charge add to charge repulsions and less compact structure

#### Silk

Anti parallel conformations are stronger - alignment of H bonding.

- Often found in silk
- R groups can interact glycine and alanine



#### Keratin

**Fibrous protein** – makes up most of protein in hair, nails, horns and feathers >30 different keratin genes

- Two classes  $\alpha$ -mammals,  $\beta$  birds and reptiles
- Basic unit is two left handed helical wrapped around the other
- Left handed coil-coil
- Top down view shows nonpolar aa allowing hydrophobic interactions
- · Hard or soft keratin is based on Cys-Cys content
- Cys rich allows for disulfide bridges between strands and within fibers
- Hair perms use mercaptans to reduce S-S bridges
- Ammonium thioglycolate is perm salt
- Peroxide oxidizes SH back to S-S





#### Unique structure of collagen

#### A special helical protein

- biological significance fibrous, structural component
- Type I collagen is found in bone, tendon and skin, II in cartilage and III in blood vessels
- Glycine R group face inside others outside
- up to 30% are proline or hydroxyproline important for maintaining secondary structure
- hydroxyprolines involved in H bonding of three strands together
- helical structure formed by three left handed helices twisted to form a right handed superhelix (gives strength)
- hydrogen bonding between 3 helices
- (thus the glycine)
- covalent bonding of lysine between strands necessary for strength

Hydroxylations on pro are performed by an enzyme called prolyl hydroxylase, which is an enzyme that requires vitamin C as a cofactor in the reaction.

Absence of vitamin C in the diet reduces hydroxylation of pro, and collagen fibres begin to break down and new collagen not formed properly. Lack of vitamin C causes scurvy because collagen fibres are not formed properly, and this causes skin lesions, weakened gums so teeth fall out etc.

#### A special helical protein

Equally important is hydroxy-lys catalysed by lysine hydroxylase. Attached to the lys residues are three sugars gal-gal-glu, and these enable H-bonding to occur between triple helices, which is essential for stability of the greater complex that binds fibers together to form a matrix bed to binds cells to the matrix and form a tissue.



#### Collagen Related Disease

## Loss of flexibility with age is likely due to increased amount cross-linked collagen compared to younger tissue

Scurvy – problems with sea voyages, lack of food other than salted meats.

- Symptoms include, swollen gums, loose teeth, small black-and-blue spots on the skin, and bleeding from small blood vessels are among the characteristic signs of scurvy.
- Caused when vitamin C (ascorbic acid) is lost from diet
- Vit C is needed to keep Iron reduced in the active site of prolyl hydroxylase. This is the enzyme
- responsible for conversion of proline to hydroxyproline. The H bonding of hydroxyproline is vital for the connective protein's function



In 1795, the British Royal Navy provided a daily ration of lime or lemon juice to all its men. English sailors to this day are called "limeys", for lime was the term used at the time for both lemons and limes.



Several heritable diseases result from mutations in the collagen Marfan's Syndrom and Ehler's-Danlos syndromes - inherited disorder of connective tissue which affects many organ systems, including the skeleton, lungs, eyes, heart and blood vessels. All resulting from various mutation in collagen and other fibril associated proteins, ultimately affecting the structure and molecular interaction.

#### Protein Structure – Folding

#### Why study how a protein folds:

The structure that a protein adopts is vital to it's chemistry. Its structure determines which of its amino acids are exposed to carry out the protein's function. Its structure determines what substrates it can react with.

#### Levinthal's Paradox

di

Consider a 100 residue protein. If each residue can take only 3 positions, there are  $3^{100} = 5 \times 10^{47}$  possible conformations.

If it takes 10 s to convert from 1 structure to another, an exhaustive search would take  $1.6 \times 10$  years! Folding must proceed by progressive stabilization of intermediates. How can this path be found?

		Covalent: 490, O-H		
	Type of Interaction	Strength, kJ/mol	Example	Dependence of Energy on Distance
**ipole	** (a) Charge-charge Longest-range force; nondirectional	α (q⁺q⁻)/r²D	-NH₃ ))C o	1/r
	(b) Charge-dipole Depends on orientation of dipole		. <sub>NH</sub> , Ф	1/r <sup>2</sup>
	(c) Dipole-dipole Depends on mutual orientation of dipoles	0.5-4	аго <mark>н</mark> аго н н аго д н н	1/r <sup>3</sup>
	(d) Charge-induced dipole Depends on polarizability of molecule in which dipole is induced		-NH3 (g- g-)	1/r4
	(e) Dipole-induced dipole Depends on polarizability of molecule in which dipole is induced		۹ <sup>-</sup> 0 <sup>-</sup> H H	1/r <sup>5</sup>
	** (f) Dispersion Involves mutual synchronization of fluctuating charges	<1		1/r <sup>6</sup>
	(g) van der Waals repulsion Occurs when outer electron orbitals overlap			1/r <sup>12</sup>
	** (h) Hydrogen bond Charge attraction + partial covalent bond	6-20	N-H····O=C Hydrogen bond length	Length of bond fixed

# Non-covalent interactions in tertiary structure

#### Hydrophobic Effect in tertiary structure

Water adjacent to a non-polar molecule sacrifice rotational and translational freedom to maintain molecular interactions. Thus, non-polar molecules associate to reduce the surface area with aqueous solvent.

- Secondary structural elements form, then pack together to create tertiary structure
- Packing excludes water from the center of the folded protein—hydrophobic amino acids are buried in the center, no longer have to be solvated by water
- Cavities form that are complementary to any small molecules (or other large proteins)—binding sites/active sites (in the case of enzymes)----FUNCTION!

#### Levinthal's Paradox may be an energy landscape.

Protein Folding Movie: http://www.youtube.com/watch?v=swEc\_sUVz5I

**Protein Folding** - Secondary structure forms, then the protein begins to compact itself until it reaches the lowest energy state possible. Proteins fold spontaneously and on short (nanosecond!) timescales—but theoretically, a protein has a large number of degrees of freedom (all those  $\psi$  and  $\phi$  angles in the peptide bonds), and if all possibilities were sampled it would take FOREVER (well, longer than the age of the earth for a 100-mer protein)—Levinthal's Paradox



#### How Does Protein Stability Arise?



#### Thinking About Proteins as Dynamic Structures

antly in motion



Proteins can undergo conformational changes that involve the movement of whole sections of protein/sidechains/atoms—can be in response to stimuli (e.g. small molecule substrate)

### Why do proteins fold?

Figure 6.35: NMR structure-ensemble that has a variety of structures that fit the data

Protein Folding Problem-Folded state is only slightly more stable than unfolded state

 $\Delta G_{folded} = -0.1$  kcal/mol per AA For comparison: a H-bond is 1 - 4 kcal/mol

Forces that drive folding:

**Water entropy (\Delta S)** Non-polar and polar interactions in protein Disulfide bonds Forces that prevent folding: Chain entropy (△S) Polar interactions between AAs and water

Total =  $\sim 200$  kcal/mol

Total =  $\sim 200$  kcal/mol

#### Protein Structure – Domains

#### Components of tertiary structure



- Combination of Folds, Supersecondary and secondary structures.
- ② Independently folded, compact units in proteins.
- ③ Single domain may have a particular function (ie bind small molecule, catalyze rxn).
- ④ In multi-functional enzymes, each catalytic activity can be on one of several domains.





Guanidinium/Urea

Protein Structure – Structure Determination

#### X-ray crystallography – First crystal structure was myoglobin,

- Resolution: (look up the definition of resolution???!!!!\_
- 5-6 angstrom overall shape of protein
- 3 angstroms, begin to see overall folding
- about 1.5 angstroms can begin so see individual atoms
  - Most figures from textbook

#### **Protein Denaturation**

0

Urea

NH

\*disrupt H-bonding

H<sub>2</sub>N

NH<sub>2</sub>

Guanidinium

 $H_2N$ 

#### **Denaturation** - disruption of native conformation of a protein, with loss of biological activity:

- 1) Energy required is small, perhaps only equivalent to 3-4 hydrogen bonds
- 2) Proteins are denatured by heating, pH, or chemicals
- Some proteins can be refolded or renatured
- 4) Native = properly folded
- 5) Denatured = not folded
- 6) Misfolded = not properly folded7) Aggregation- the clumping together
- proteins

\*easily happens with denatured proteins

Protein Denaturation – Effect of structure & ligand binding



## Thermodynamic prediction of folding and stability – Go to van't hoff plot page for details.

Temperature

unfolded

ent

#### **Quaternary Structure**

Many proteins exist as multiple polypeptide units:

- Quaternary structure only exists when there are more than one protein subunits involved in a protein
  - Subunits are separate genes/proteins which come together with similar or different subunits to form a complete protein • Homo or hetero proteins
    - Monomer, dimer, drimer, tetramer, pentamer, hexamer...
    - Also called multimeric proteins or enzymes
    - Often involved in cooperativity or allosteric regulation

#### **Subunit Arraignment**

• Usually non-covalent interactions – held in place by non-polar aa and H bonds. Some bound by disulfide bridges

**Global Symmetry** – Refers to the symmetry of entire complex. Symmetric proteins have multiple arraignments (local symmetry)



#### Quaternary Structure : Driving Forces

#### The Bad News:

- Considerable entropy loss when subunits come together
- Loss of translational degrees of freedom
- · Residues that were able to move at the subunit interface are now restricted

#### The Great News:

- · Increased Van der Waals contacts-but nearly as many are lost with water as are made with the new oligomer
- Increased hydrophobic interactions—the money maker (roughly 100-200kJ/mol)
- Polar interactions at the interface
- · Salt bridges/disulfides

#### **Protein Stability**

#### Protein stability is labile – very little energy to denature

• U40 kJ/mol for an avg100 aa protein to denature. H bond breaking takes ~20kJ/mol

Stabilizing factors- global impact of non-covalent and covalent interactions maintaining tertiary and quaternary structure High Influence

- hydrophobic aa in center of protein keep structure stable
- Large number of van der Waals are lost if denatured and maintain overall structure

#### Lower influence

- H bonding – important for structure but the balalnce in native or denatured is same H bonding energy as H bonding will occur with water in denatured state

- salt bridges - entropy and solvation changes offset most of the ionic interactions

#### **Protein Denaturing**

#### Denaturation and Renaturation (sometimes reversible)

- **Heat** disrupts or melts the van der Waal and other forces holding protein in native form
- **pH** both basic and acidic will alter functional group charge decreasing ionic interactions within chain or at surface of protein. Also can cause loss of H bonding potential – consider Carboxyl and Amino group at high and low pH
- **Detergents** hydrophobic, non-polar amino acids will unravel and bind to hydropathic soaps/detergents
- Chatropic agents: bind water tightly away from protein.
- Reducing agents: reduce Cys- disulfides

#### **Quaternary Interactions Gone Awry: Amyloidoses**

- We've covered interactions between already folded subunits.
- Unfolded or misfolded monomers can glob together to form aggregate structures, which organized into cross-β-sheets (amyloid)
- Amyloidoses are a major health problem in the ageing population (Alzheimer's disease, Systemic amyloidoses, etc.)

#### MAD COW DISEASE

The prion protein exists in two forms. The normal, protein (PrPc) can change its shape to a harmful, disease-causing form (PrPSc). The conversion from PrPc to PrPSc then proceeds via a chain-reaction. When enough PrPSc proteins have been made they form long filamentous aggregates that gradually damage neuronal tissue. The harmful PrPSc form is very resistant to high temperatures, UV-irradiation and strong degradative enzymes.

#### Prion diseases arise in three different ways

- 1. Through horizontal transmission from e.g. a sheep to a cow (BSE).
- 2. In inherited forms, mutations in the prion gene are transmitted from parent to child.
- 3. They can arise spontaneously.

#### **Route of infection**

When cows are fed with offals prepared from infected sheep, prions are taken up

from the gut and transported along nerve fibers to the brain stem. Here prions accumulate and convert normal prion proteins to the disease-causing form, PrPSc. Years later, BSE results when a sufficient number of nerve cells have become damaged, affecting the behavior of the cows.

9



