Outline of the presentation: what is covered here?

- Relationship between structural elements (secondary-tertiary and quaternary) and folding
- Difference between proteins and random polymers
- Forces and energetic that constrain protein to particular structure

How are the proteins built up? The hierarchy of protein structures. Does this tell us anything about how they probably fold?

1. **Primary** structure
   1.1. AA-sequence, read from the DNA sequence.

2. **Secondary** structure
   2.1.
   2.2. α-helices and β-strands as structural modules
   2.3. These are repeating structures – short often (5-7 residues).
   2.4. There are strong tendencies for certain residues (within certain sequences of consecutive AAs on the primary backbone chain) to form definite types of secondary structures. But no real rigid "code" of folding is known.

3. **Tertiary** structure – this is where the "oil drop" idea enters.
   3.1. Interacting secondary modules with loops in between – these sections are called turns.
   3.2. Also interactions between the residues and backbones of the loops, where there is not a well defined repetitive secondary structure.

3.3. Sometimes the secondary structures are closely related to the tertiary contacts between the different parts of the protein that are far away along the sequence (think of the β-sheet contacts, which had H-bonding between the β -strands, and the β -strands are located at different parts of the polymer.

3.4. The packing together of the secondary structural elements, each of which are often only marginally stable, are classified as “super secondary structures”. Even β -sheets, stacked on top of each other, can only be done in certain ways. Coiled coils are common, and only certain angles between the helices are preferred.

3.5. There seem to be “families” of tertiary structures (20-40 of them), and these “families are found in many different, at first seemingly unrelated, structures. This may make the job of understanding the organizational properties of proteins easier that earlier thought.

4. **Quaternary** structure

4.1. Interacting subunits, where, each subunit is a complete folded polypeptide structure.

Globular proteins (and enzymes are always globular proteins) fold into compact structures, and the length of the secondary “modules” are limited and connected by “loops” in order to form these globular structures.

In general, the way proteins fold is thought to proceed in steps and along certain paths so that the catastrophe of the Leventhal paradox does not have a chance to happen.

There are differences in opinion in how this happens, and which parameters and mechanisms are important.

Note in a globular protein there are many intramolecular contacts between different residues that are separated by many AA-positions on the one-dimensional sequence of the polymer. These “non-local” interactions are very difficult to calculate and to model, *a priori*.

Some important general considerations about protein structure:

We have seen that proteins are made up of **α-helices,** β **-ribbons** and connecting stretches of the polypeptide chain that are usually located between two segments of these two major secondary structures; these **connecting segments** usually **turn** the “running direction” of the secondary structures, and often also include important interactions for the protein structure.

We have also seen that the sides of the proteins that **project** towards the **interior** of the protein are usually **hydrophobic**, and those that face the solvent are **hydrophilic**. The structures are **tightly packed globular structures**, where the **volume** of the **amino-acids inside** the proteins corresponds to their **volumes** in **crystals** of **free amino acids**.

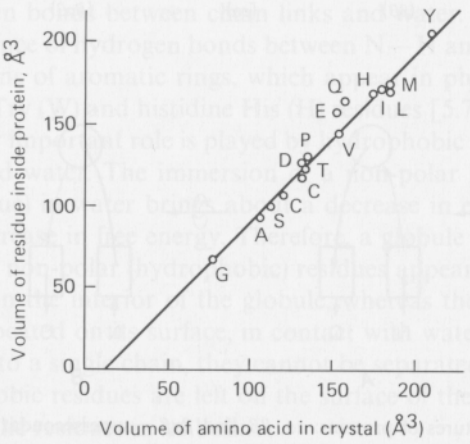

Fig. 5.2. The average volume of amino-acid residues inside protein as a function of the volume of amino-acid residues in crystals [5.9]

From: Clothia,C. (1984) **Ann.Rev.Biochem 53** 537-572

There seem to be four main classes of protein structure:

1) Proteins built from α-helices

2) β -proteins, which mostly contain collinear or orthogonal β -layers.

3) ᾶ β proteins, where α - and β -segments nearly alternate along the chain (these proteins usually contain β͂ ᾶ β clusters).

4) (α+β) proteins, which are heterogeneous structures with an irregular arrangement of units of α- and β-structures.

This classification of protein structures has come from detailed X-ray structural analysis of many proteins of over 200 different proteins.

Example of an α-helix protein (myoglobin – one of the most studied proteins by physical techniques):

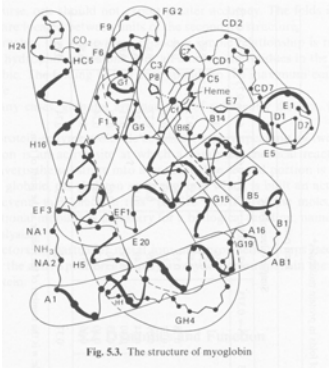

Fig. 5.3. The structure of myoglobin

** For a tutorial on the web dealing with the basics of the different structures, we advise going to http://www-structure.llnl.gov/Xray/101index.html.

To get an overview and review of what we covered in class relating to basics about protein structures, go to the bottom of the list on the left-hand side, and click on "Protein Structure Basics". This site is also an **excellent review of x-ray crystallography**, which we will not cover.

So, if you have questions about this, go to this site. It is very well done and concisely explained, and the techniques how it is practiced in biophysics are also well explained.

Proteins fold very differently from random polymers, and we want to see why this is true. But we just mention a few things now.

The root mean square **end-to-end distance** of a Gaussian freely jointed chain is:

$$\mathbf{\langle h^2 \rangle^{1/2} = Z^{1/2} b},$$

where b = the length of the segment, and Z = the number of freely jointed segments.

The **volume** of the polymer is: proportional to $\mathbf{h^3 = Z^{3/2} b^3}$

The **average concentration of the segments** is therefore =

$$\mathbf{c = Z/Z^{3/2} b^3 = Z^{-1/2} b^{-3}}$$

Such a polymer is a very “open” structure, and we will show later that the fluctuation in the radius of the Gaussian freely jointed polymer is as large as the average size of the coil.

If there are electrostatic interactions between the segments this loose coil can collapse into a sort of “molten globule” state, but this will be very different from the “molten globule” of proteins, because there is still only random structure in a random polymer, as opposed to a much more constrained, fairly unique three-dimensional structure of a protein.

The distribution distances in proteins are different from random polymers (see next). Proteins are actually very compact.

We might say that the proteins have a sort of **linear memory** (linear regarding their sequence), and they interact specifically with distal parts (distal according to the linear sequence). The structure of a protein is very dependent on the exact amino acid sequence, and it depends on the physical properties of the amino acids along the sequence.

Also, the structure of a protein results from the interactions between distal parts of the protein (electrostatic, salt bridges, coordination with metal ions, van der Waals interactions, hydrophobic interactions, and hydrogen bonds), and these distal interactions are often very specific, especially in the active site.

Fluctuations: There will be fluctuations in a protein structure, but unlike the fluctuations of the random polymer, these fluctuations are **small in extent** and they are **local** (the whole structure does not expand or contract). It is interesting to ask what effect these fluctuations have on the enzymatic activity of proteins.

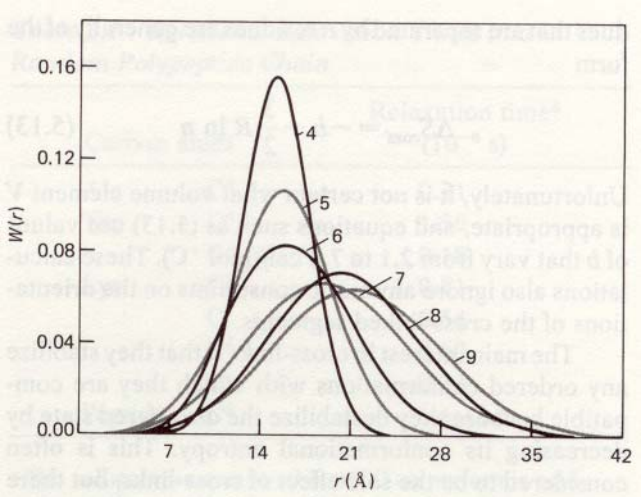

FIGURE 5.5
Radial distribution function of the distances between naphthalene and dansyl groups attached to the ends of peptides of 4–9 residues of *N*-hydroxyethyl-Gln, measured by fluorescence energy transfer. (From E. Haas et al., *Proc. Natl. Acad. Sci. USA* 72:1807–1811, 1975.)

Distribution of the radius of gyration of different conformations of polymers:

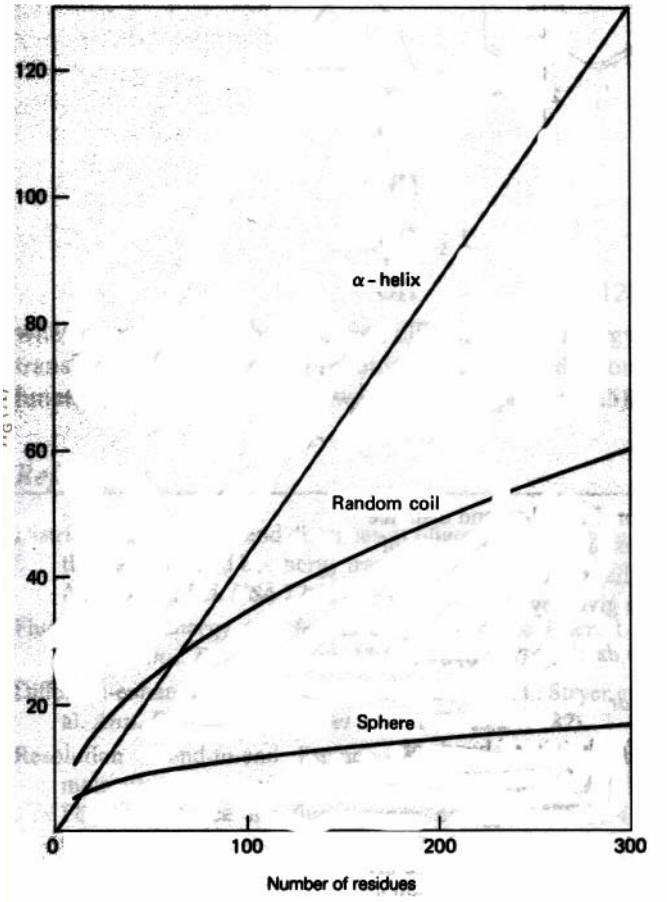

FIGURE 5.3
The radius of gyration R_G for polypeptide chains of various lengths in α-helical, random coil, and compact spherical conformations.

$\langle R_g \rangle_0^2 = nl^2 = \langle r_0^2 \rangle / 6$

radius of gyration = “end-to-end distance” /6

Look at this link for some general introduction to this concept
polystat2002.pdf

A short overview of the forces and energetics that are active in constraining proteins to specific structures.

- What are the properties of proteins that lead to their structure?

- And what are the forces acting between the individual parts that leads to protein folding (DNA structure, membrane formation), and to aggregates of macromolecules?

- Remember we have the following stages of structures:

1) The basic structure and properties of the amino acids
2) The specific amino-acid sequence (primary structure)
3) The formation of the peptide bond and its properties – for instance the constraints on the molecular geometry of the planar peptide bond and the geometry of (rotation about) the tetrahedral symmetry of the C_α-carbons,

- An aside on some basic energetics:

We use kcal/mole.

23.06 kcal/mole = 1 eV;
1 calorie = 4.184 joules.

k_BT = 0.617 kcal/mole at 298 K (25 ° C)

- **Important weak interactions are 2-6 kcal/mole.**

This is only 4-10 times **k_BT!**

- This level of energetic interactions allows **stability and flexibility** at the same time.

Let's go through the interactions that are important. We will give an overview at this point. Use the link on previous lectures to go into the details of each interaction.

1) **Covalent bonds**. **50-250 kcal/mole** (**2-10 eV**)

The only way to break them during the normal functioning of a cell is by the action of an enzyme.

2) **van der Waals interactions**.
Typical depth of energy minimum is about k_BT. The interactions are ubiquitous, and happen between all atoms; so adding up many interactions can add up to very large energies.

A) ***Long range***: QM fluctuations of atomic dipoles; attractive and varies as $1/R^6$.
B) ***Short range***: Repulsive; overlapping of electron clouds, Pauli exclusion principle, electrostatic repulsion of nuclei. Increases much more rapidly than the attractive forces.

These interactions set the hard-sphere steric constraints of intermolecular interactions. Typical values of these hard sphere distances are 1.2 Å to 2.2 Å. These steric effects reduce drastically the number of conformations available to the polymer.

Remember the Ramachandran Map (see other notes)

3)

Electrostatic effects.

A) ***Two sources of separate charges in protein***s.
1) H-ionization reactions – sensitive to pH. There are 5 AAs that ionize:
2) negatively charges AAs: aspartic acid and glutamic acid
3) positively charged AAs: lysine, arginine, histidine.

Often these groups are exposed to the solvent, or involved in "salt bridges" (in the core of the protein structure). Charges on these side groups are sensitive to the pH of the surrounding solution, and depend on the pKs. The pKs can be a function of the environment, so they are not the same for all positions of these AAs.

B) ***Unequal electron distribution in a covalent bond***. ***Electronegativity***.

- Partial charges lead to dipole moments, which have a large effect on the protein structures.

- Peptide bond has a dipole moment of 3.7 Debye (D); water has a dipole of 1.85 D.

- 1 Debye = 10^{-10} esu cm. A dipole moment of an electron and a proton separated by 1 Å = 4.8 D.

- There are 6 AA-side groups that have significant dipoles: serine, threonine, cysteine, tyrosine, asparagine, glutamine.

- The energies of the electrostatic interactions are:

$$E_{ij} = q_i q_j / \varepsilon r_{ij}$$

What is the dielectric constant inside a protein?

Static for water: $\varepsilon_{water} = 78.5$
Amide polymers: $\varepsilon_{amide\ polymer} = 4$

Notice the large effect of the medium on the electrostatic energies.

The potential energy of two charges separated by 4 Å of water is about 1 kcal/mole.

The potential energy of two charges separated by 4 Å of protein interior is abouț 20 kcal/mole. Remember that the dielectric constant is a bulk property.

The appropriate treatment of $\varepsilon_{protein}$ is a very difficult and challenging problem, and there is a fair amount of disagreement on this.

4) **Hydrogen bonds**:
H connected to N, O and S

$O^{\delta-}$ — $H^{\delta+}$ This is the hydrogen donor
$O^{\delta-}$ — $H^{\delta+}$ ----- $O^{\delta-}$=$C^{\delta+}$

- It is mainly an electrostatic effect. Strength of a H-bond is **1 – 7 kcal/mole**.

- Strongly dependent on geometry.

- A strong H-bond requires interatomic distances of <3.5 Å, and should be linear.

- The backbones of proteins have many opportunities for H-bonding.

- Interacting groups and substrates (in active sites) satisfy strict geometries and patterns.

5) <u>Water</u>

- 1.85 D dipole. Has a major effect on protein structure. Bulk water forms local structures of tetrahedral coordination.

- These structures last for about 10^{-12} to 10^{-9} seconds (Frank 1958).

- These are sometimes called clathrate cage-like structures near proteins.

- There is a primary layer of water around each protein (hydration layer).

6) **<u>Hydrophobic effect</u>**.

- Water-water interactions are more favorable than water-protein (non-polar) interactions.

- Hydrophilic interactions include interactions between water and charged and polar chemical groups.

- The larger the nonpolar AA surface area, the stronger the effect per mole. 8 AAs are nonpolar: glycine, alanine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan.

- The strength of the hydrophobic effect is proportional to the surface area. It can be as high as **2 kcal/mole of residue**.