Protein Structure  (I1-I3)
First, Something New for This Year: HELM

- HELM:
  - Heirarchical Editing Language for Macromolecules

  • (What do those terms mean?)
RNA polymer
8 nucleotides because of the 7 “.”s
7 modified monomers, because of 7 “[]”s
Nucleotide 2 has a modified sugar [mR] and linker [sP]
Nucleotide 4 has a modified base [5meC]
Nucleotides 5-8 have modified sugar [dR]
Analogous to the natural AUGCTTTT
First, Something New for This Year: HELM

HELM:

- Heirarchical Editing Language for Macromolecules
  - (What do those terms mean?)

- Let's check openhelm.org
  - Also the notation link
    - ...and the notation presentation link
There are three types of protein structure:

- Fibrous proteins
- Globular proteins
- Integral membrane proteins
Types of Protein Structures I

- **(i) Fibrous**, (ii) globular, (iii) integral membrane

**Fibrous**
- Made of long, stringy molecules that clump
- Clumps are fiber-like (duh!)
- Usually inert
- Their physical properties tend to count
- Found in e.g. muscles, ligaments, tendons, bones
- Example proteins:
  - Collagen (in skin, into gelatin)
  - Keratin (in hair, in nails)
- Animals only (why not plants?)
Types of Protein Structures II

- Fibrous, **globular**, integral membrane
- Globular
  - What shape do they tend to be? Do they use fractal globule folding?
  - Not inert:
    - function as enzymes, antibodies, for signalling, for transport, and many other things
  - Their working environment is **water**
  - Their shape is partly due to interaction with water molecules
    - (What is distinctive about water molecules?)
Globularity and Water

- Water molecules are quite polar
- What is polarity in this context?
- Different amino acids have side chains that are
  - Polar, or
  - Nonpolar
- Which will be attracted to water molecules?
- Are the hydrophilic AAs polar or nonpolar?
Globularity and Water II

- Water molecules are quite polar
- What is polarity in this context?
- Different amino acids have side chains that are
  - Polar
  - Nonpolar
- Which will *not* be attracted to water?
- Are the *hydrophobic* AAs polar or nonpolar?
Globularity and Water III

- Are most globular proteins linear sequences of AAs, or highly branched ball-like shapes?
- According to dictionary.com:
  - Antibodies tend to be Y-shaped. How do they do it?
  - They bind to antigens, thereby helping immunity
    - Antigens – anything that stimulates production of, and binds to, an antibody; they are “bad”
  - *Some* globular proteins are antibodies
    - Are *these* linear sequences or highly branched shapes?
Indeed,

- globular proteins fold into ball-like shapes
- We’re talking ball-like, not perfect, tiny beach balls!

Hydrophobic AAs tend to be on the inside

Hydrophilic AAs tend to be on the outside

- Why and why?

The ball shape tends to minimize free energy

- That’s why they fold into that shape
- Imagine a stretched spring trying to shrink
Types of Protein Structures III

- Fibrous, globular, **integral membrane**
- Their working environment is the *cell membrane*
- Integral membrane protein functions include:
  - Signalling (signal transduction); membrane transport
    - (e.g. pumping ions across a cell membrane)
- Rhodopsin has 7 membrane-spanning segments
  - (see next slide)
- Membrane is a lipid bilayer (except in a few *archaea*)
- Typically, parts are in the membrane and parts project out
  - Which parts are polar? Nonpolar?
Fig. 9. Structural model of rhodopsin showing seven transmembrane components and the attachment site for retinal.

Protein Structure – another way

- Fibrous, globular, and integral membrane is one way to do it
- Here is another...
Protein structure, another way

- Primary, secondary, tertiary, quaternary

- **Primary**:
  - Detailed description of composition
  - For proteins, AA sequence is typically used
    - How about DNA? RNA?
  - Cross-links may also exist
  - Chemical modifications may also exist
Abstracting primary protein structure

- Peptide bond
  - Link between 2 AAs
- Dipeptide
  - 2 linked AAs
- Tripeptide, quadrapeptide...dodecapeptide
  - E.g. nonapeptide has 95,000 Google hits
    - But what is it?
- Polypeptide – about 10-100 AAs
Protein structure, secondary

- Primary, secondary, tertiary, quaternary
- **Secondary**
  - Composition in terms of segments
  - Main segment types are
    - Alpha-helices
    - Beta-sheets/beta-strands/beta-pleated sheets
      - Strands suitable for folding into sheets
  - Coils/random coils
    - They connect the other secondary structures
    - Not particularly “coiled,” more like connecting strings
Alpha Helix

http://www.uic.edu/classes/bios/bios100/lecturesf04am/alphahelix.jpg
Are alpha-helices a type of DNA?
Alpha Helix

http://student.ccbcmd.edu/~gkaiser/biotutorials/proteins/images/alphahelix.jpg
Alpha Helix

http://opm.phar.umich.edu/images/proteins/2irg.gif

“Amphipathic alpha-helix of apolipoprotein”
Alpha Helix

Amino acid residues

- Amino acid molecules connect into sequences.
- When so doing, water molecules are released.
- An amino acid molecule minus
  - a hydrogen atom from the amino end, or
  - a hydrogen and oxygen atom from the carboxyl end, or
  - both
- ...is called an amino acid residue.
Residue

- What does the word ‘residue’ mean?

- What does it mean in the phrase “amino acid residue”? 
Alpha Helix

http://www.brooklyn.cuny.edu/bc/ahp/LAD/C4b/graphics/C4b_alphaHelix.GIF
Is this a zig-zag shape?
Alpha Helix
http://osulibrary.orst.edu/specialcollections/coll/pauling/dna/pictures/alphahelix.html
Alpha Helix

http://wiz2.pharm.wayne.edu/biochem/nsphelix1.jpg

How many residues make a complete turn (on ave.)

What is that?
Alpha Helix

Alpha Helix


Is this a ring?
How big can an alpha helix get? W-i-d-t-h? Length?
Alpha Helix

Beta Sheet

- Both cross-bond between C and N.
- What are C and N?
- Could you turn (b) into (a) by shifting one strand a little bit sideways?
Want More Beta Sheets?
See your favorite image search engine!
Tertiary protein structure

- Primary, secondary, **tertiary**, quaternary

**Tertiary**

- Fold configuration of the protein
  - I.e., “the fold”
  - Why do proteins fold up?
- 3-D position of every atom

- PDB has tertiary structure information

  - PDB=Protein Database
    - (Westhead p. 33)
Protein structure, quaternary

- Primary, secondary, tertiary, quaternary

- Quaternary
  - The arrangement of subunits, monomers, polypeptide chains that connect together to form a complete protein
  - Some proteins are just proteins
  - Others are complexes of multiple proteins
    - Because proteins can bind together!
      - E.g. an antibody and antigen protein bound together
    - These have quaternary structure
Protein Structure: Pentenary Huh?!
Protein Structure: a 3\textsuperscript{rd} Classification Method
Protein Domains (a *third* way!)

- Proteins are often modular
  - Different pieces (modules) do different things
  - A software module can be called a module
  - A protein module is called a “domain”
  - Some proteins have a homeo domain
    - This domain binds to DNA
Protein Domains II

- Domain boundaries
  - A domain has a beginning and end
  - Some think they bracket a
    - Functional piece of the protein
    - A geometrically distinct piece
    - A piece found in varied proteins
  - Types: as with proteins themselves...
    - Globular, integral membrane, fibrous

- A protein may have domains of different types
  - Receptors can have integral membrane anchors and globular sensors
    - with docking sites outside the membrane on the receptor or anchor?
The Function of Structure

- Protein shape can depend on environment
  - Hydrophilic AAs are pulled to the outside
    - in a water environment
    - not in, say, air
  - Normal function requires proper structure
    - Active sites must be on the surface, e.g.
- Some examples follow...
Abstract Example

(www.smi.tu-berlin.de/story/Intro.htm)
Abstract Example II

(www.columbia.edu/.../lectures/lec06_05.html)
Abstract Example III

(www.blobs.org/science/enzyme/index.shtml)
(View in presentation mode for animation)
“The dUTPase Active Site. View looking into active site with protein surface colored according to charge (positive - BLUE; negative - RED). Binding of substrate dUTP and Mg2+ ions is illustrated.”
More than Just Active Sites

- Enzymes are important
- So are other kinds of proteins
- A key function of surface shape is:
  - Recognition
    - (Enzymatic) *catalysis* is just one example
    - *Transportation* is another
    - *Immune* system activities are another
    - *Signaling*, complex formation,...etc.
Shape is Just One Relevant Property

- For two molecules to fit together –
  - Shape is important
    - (think of jigsaw puzzle pieces)
  - Charge is also important
    - Negative attracts positive, repels negative
    - Analogously for positive
  - Other physico-chemical properties matter
    - What were those?
Some nucleotide mutations are likelier than others (why?)

Some amino acid mutations are likelier than others, e.g.
- Similar physico-chemical properties
- Similar size

Some DNA areas mutate easier than others

Some protein areas mutate easier than others
- Why and why?
Protein Evolution II

- Some protein **areas** mutate easier than others
  - Globular
    - Globular cores mutate slowly
      - Tight packing facilitates stable molecules
      - Mutations tend to reduce packing compactness
      - ...leading to less structural stability
      - ...leading to less effective function
    - Globular surface AAs mutate faster
      - Often one hydrophilic AA is as good as another
  - Secondary structures (α-helices, β-sheets)
    - tend to be in cores
  - Loops connecting secondary structures
    - tend to be on surfaces
    - **So, loop AAs tend to mutate faster**
Some protein **areas** mutate easier than others

- Globular loop segments mutate fast
  - (see previous slide)

- Integral membrane proteins
  - Membrane-spanning segments mutate slowly
  - Connecting loops on cell surface mutate faster

**So, loop AAs tend to mutate faster**

- True for both globular and integral membrane proteins/protein domains
Why Some AAs are More Likely to Mutate

- Are changes more likely if they:
  - increase or decrease structural stability?
  - Consider hereditary Creutzfeldt-Jacob Disease (CJD)...
    - increase or decrease functional efficiency?
- Are these rules absolute?
Some AAs are Especially Important

- A particular AA at a particular location
  - may be key to structural stability
    - It will be conserved over time
  - may be key to function
    - It will also be conserved over time

- Example:
  - S, H, and D in active site in serine protease
    - What are they?
      - See e.g. Westhead Fig. 1, p. 130

- Usually, do surface loops evolve faster or slower than secondary structures?
  - What might be an exception?
Example: insertions/deletions are in loops; Cs are all conserved due to structural stability of their (covalent) disulfide bonds

**Fig. 1.** A multiple alignment of lysozyme and $\alpha$-lactalbumin sequences. Sequence names starting LYC are lysozymes and those starting with LCA are $\alpha$-lactalbumins. The KEY RES lines indicates key structural residues (disulfide-forming cysteines) with the symbol ‘$\$’; and key lysozyme catalytic residues with the symbol ‘$!$’. Key structural residues are conserved in all sequences, but the lysozyme functional residues are not conserved in the $\alpha$-lactalbumin sequences. The CONSERV line indicates the degree of conservation in a particular column (‘*’—identically conserved, ‘.’ contains only very conservative substitutions, ‘,’ contains conservative substitutions). The SS line shows secondary structure (taken from one sequence of known 3D structure).
Structure and MSA

- MSA?
- Suppose we don’t know the structure
- Given the MSA, we can predict things about structure
  - Conserved Cs may participate in disulfide bonds
  - Highly varying and rather unvarying segments...
    - ...suggest surface loops and secondary structures
Conservation of Structure: Global Protein Properties

Rules suggest mutability at the
- Nucleotide level
- Amino Acid level
- Segment level (secondary structures)
- Segment level (domains)
- Entire protein level

A heuristic for natural proteins:
- \( t(\mathcal{L}) = 290.15 \mathcal{L}^{-0.562} \)

If AAs shared is at least \( t \) then basic structure is probably conserved
- What if \( \mathcal{L} = 80? 20? 200? \)
Conservation of Structure *Without* Conservation of Sequence

Globin family proteins are found in animals, plants

- For example:
  - Hemoglobin (blood), leghemoglobin (plants)

- Structure and function are conserved
  - For example: they carry oxygen

- Sequence similarities can be < 20%
  - So they can’t be homologous?!?

- Evolution preserved
  - Function (why?)
  - Structure (why?)

- Homologs can sometimes be found from structure...
  - even when sequences don’t suggest it
Function and AA Conservation

- AAs may be critical to
  - structure
  - function
  - neither

- Homologous proteins may have completely different functions (why?)
  - Then, structure-preserving AAs conserve
  - Function-preserving AAs don’t conserve
  - Example:
    - Figure 1, p. 133 (Westhead et al.) shown earlier