Protein and RNA Review

Proteins: Introduction

- Proteins are the versatile building blocks and active molecules that form the basis of living systems.
- Function follows structure
  - We study protein structure and its dynamic changes so that we can better understand protein function.
  - These studies will involve mathematical modeling of protein structure using geometric analysis, statistics, machine learning, …

Protein Function (1)

Protein function includes:

- Molecular recognition
  - Proteins show a specificity in their interactions. This includes small ligands and other macromolecules.
- Molecular movement
  - Changes in protein conformation produces mechanical movement, be it large muscle movement, or the small flagellar motors of a bacterium.

Protein Function (2)

- Enzymatic catalysis
  - Almost all chemical reactions in a cell are facilitated by a catalytic mechanism involving a protein.
  - Reactions are accelerated, increasing the rate, sometimes by as much as a factor of 10^{17}.
  - Enzymes may use any of the following strategies:
    - Provide an alternative reaction route.
    - Stabilize reaction intermediates.
    - Facilitate alignment of reacting groups.

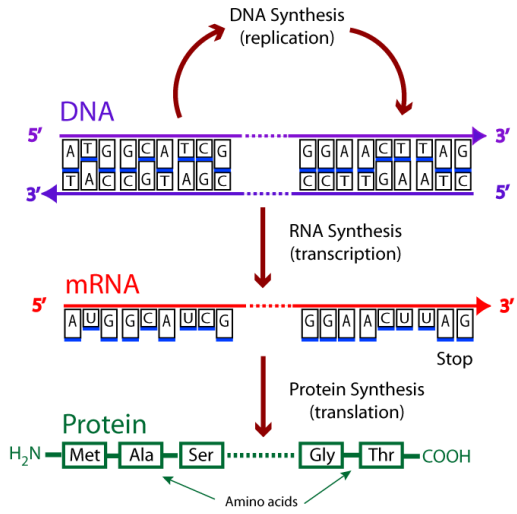
Protein Function (3)

- Structural systems
  - Proteins act as the building blocks of living systems.
  - In large complex systems they contribute to the formation of large structures such as bone, hair (collagen), nails (keratin), and many other tissues.
  - Smaller structures include strands of silk, viral coats, and frameworks for the movement of molecules within cells.

Protein Function (4)

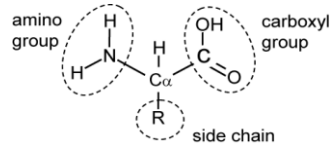
- Signal transmission
  - Information can be moved from one part of a biological system to another.
    - For example, neurotransmitters, hormones, and cytokines.
      - Cytokines act as signals in immune systems and during embryogenesis.
      - They typically bind to a cell-surface receptor and this initiates a signal cascade resulting in an up-regulation or down-regulation of gene expression.

Central Dogma

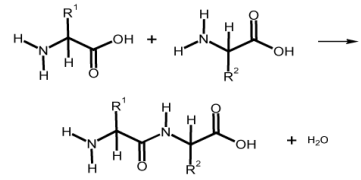


Amino Acids

- The monomers of proteins are amino acids.
  - They have the following general form:



  - A peptide bond is formed by a condensation reaction:

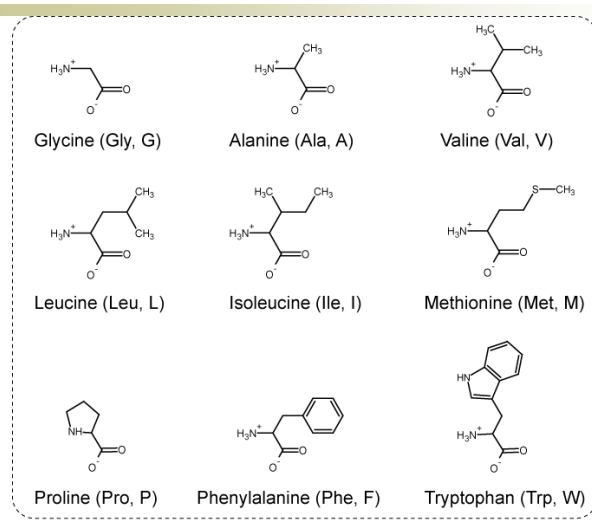


Amino Acid Classifications

- Amino acids are classified according to the nature of their side-groups.

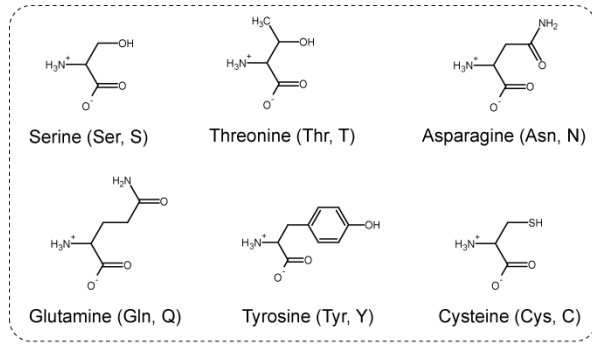
- Classifications:
  - Nonpolar (hydrophobic)
  - Polar uncharged
  - Polar charged (+)
  - Polar charged (–)

Nonpolar and Hydrophobic AA's



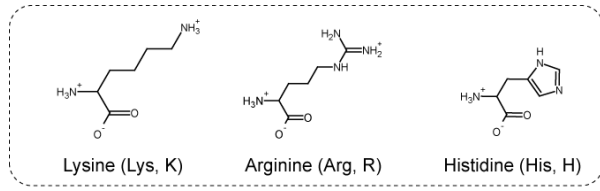
Nonpolar (Hydrophobic) Residues

Polar Uncharged AA's



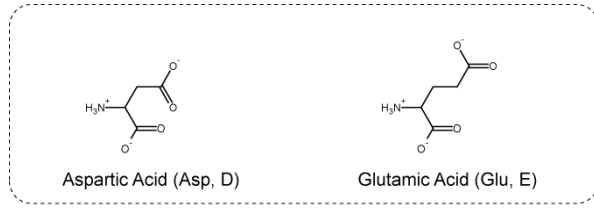
Polar Uncharged Residues

Polar charged (+) AA's



Polar Charged (+) Residues

Polar charged (–) AA's



Polar Charged (-) Residues

Special Features: Proline

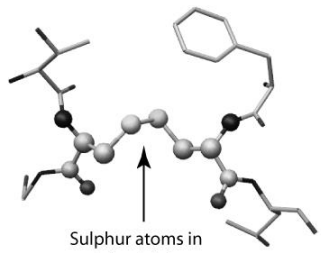
- Proline has a peculiar structure due to the cyclic binding of its three-carbon side chain to the nitrogen atom that is normally part of the amine group in the other amino acids.
  - This binding places limits on the flexibility of the backbone and so proline will often cause a bend in the backbone.
  - It is more likely to appear at the edges of helices and beta strand secondary structures, or in loops that go between them.

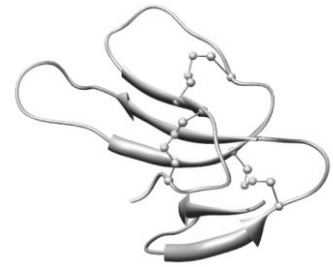
Special Features: Cysteine

- Two cysteine residues can form a disulfide bridge by means of an oxidative reaction:

$$R\text{-}SH + SH\text{-}R' \rightarrow R\text{-}S\text{-}S\text{-}R' + 2H^{+} + 2e^{-}.$$

  - This typically confers extra stability on the protein.
    - Note that bridging helps steady loops in bucain.

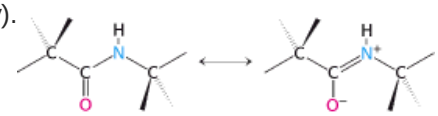




Cysteine bridges in bucain 2H8U (a krait venom)

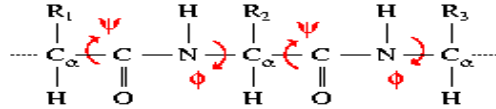
Peptide Bonds Are Planar

- Planar peptide bond
  - The C-N bond has about 40% double bond character (usually).
    - Note the “trans” configuration.
    - Some cyclic peptides do not follow these rules.



Peptide bond resonance structures

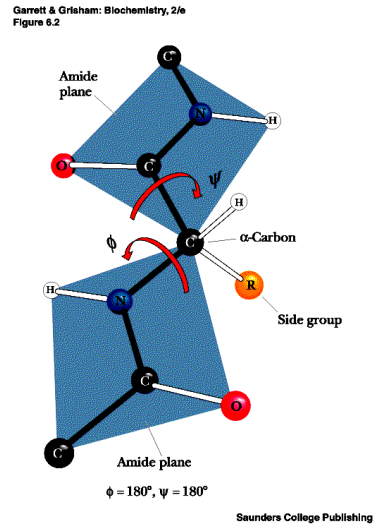
- Free rotation for attached chains:



- Rotation on either side of the alpha carbon does take place but is dependent on the rest of the molecule.
  - Can be probabilistically characterized by knowing the attached chains on either side.

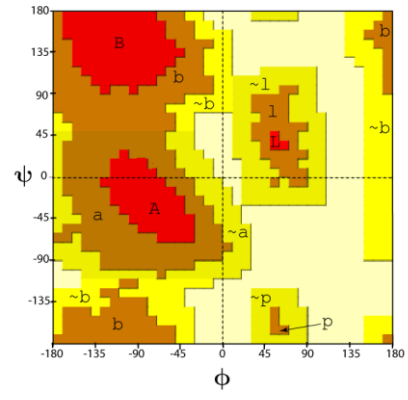
The Amide Plane

- Phi, Psi angles are illustrated in this figure.
- Later we will develop the equations that compute these angles given the coordinates of the atoms.



Ramachandran Plots

- G. N. Ramachandran plotted the distribution of ϕ and ψ angles in proteins containing polyalanine residues.
- The intensity of the shaded areas represents the likelihood of the most favorable ϕ and ψ angles.
- Many of the ϕ and ψ combinations are actually sterically forbidden.
- This corresponds to the white areas.



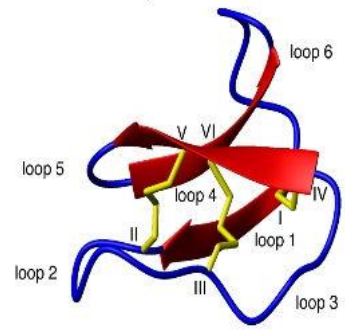
Digression: Cyclotide protein as art

- Kalata, 2004
  - Stainless steel, length 50" (1.30 m)
  - The protein Kalata is a small cyclic protein that has been recently found to be the utero-active component in a traditional African herbal medicine used to accelerate labor in childbirth.

http://www.julianvossandreae.com/works

Digression: Cyclotides

- Cyclotides are being studied at David Craik's lab in Australia.
  - They also work with venoms from cone snails, spiders, and frogs.
  - They want to develop therapeutic agents and insecticides.



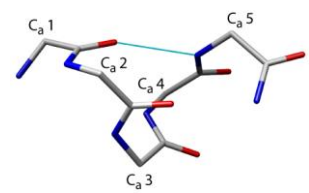
Kalata B1

Levels of Structure for Proteins

- Primary Structure
  - The primary structure is simply the sequence of amino acids.
- Secondary Structure
  - Secondary structure is described by categorizing the amino acids as being part of alpha helices, beta-sheets, or loops.
- Tertiary Structure
  - The helices and sheets combine to form a definite three dimensional conformation of the molecule.
- Quaternary Structure
  - Quaternary structure is specified by combining multiple tertiary structures (molecules) to form a working unit.
    - Not all proteins do this.

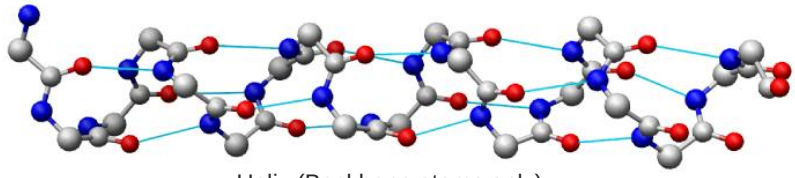
The Alpha Helix

- The consecutive peptide planes twist into a helix.
- The side chains typically point outside the helix.
- The ideal alpha helix has 3.6 residues for every complete turn of the helix.
- Note the hydrogen bond between the H on the nitrogen atom and the double bonded oxygen of the downstream carbon atom.

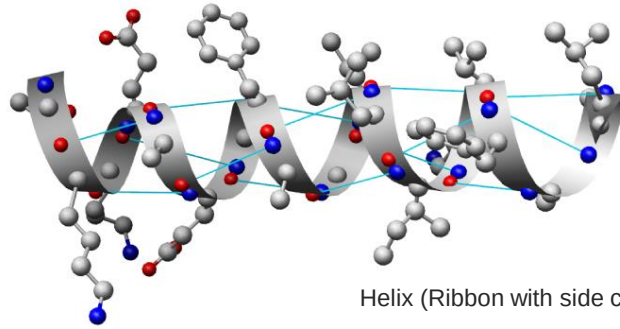


Most of the figures in this unit were created using USDF Chimera: http://www.cgl.ucsf.edu/chimera/

Hydrogen Bonds in Helices



Helix (Backbone atoms only)

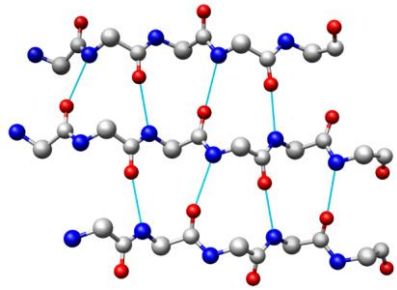


Helix (Ribbon with side chains)

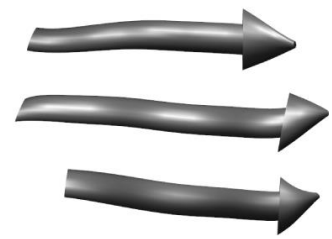
23

The Beta Sheet

- The beta-sheet is formed when peptide planes tend to align and form hydrogen bonding.
- Note the typical hydrogen bond between the H on the nitrogen atom and the double bonded oxygen of a carbon atom that is much more distant in the sequence.

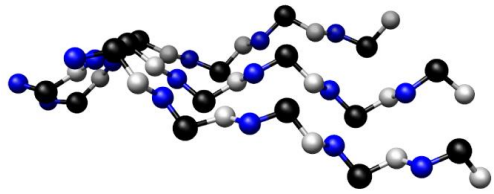


Parallel Beta Sheet

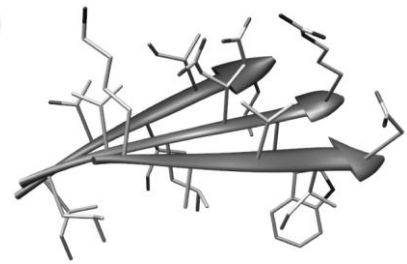


The Beta-Sheet (cont.)

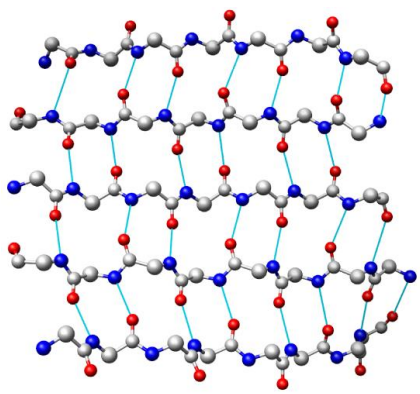
- Pleated conformation of a beta-sheet:

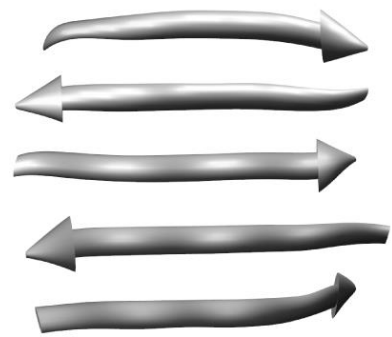


Same sheet
with side chains:



The Beta-Sheet (cont.)

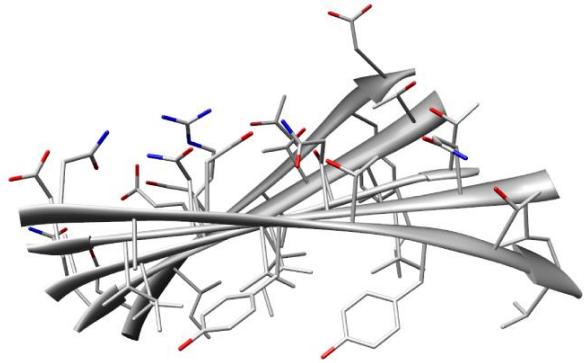




Anti-Parallel Beta Sheet

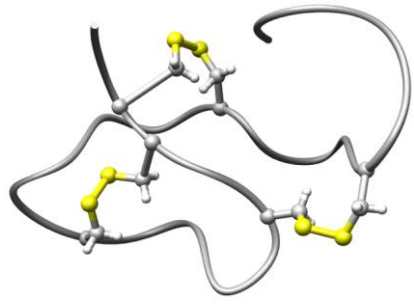
Sheet Twist

- Sheet twist can be observed for both parallel and anti-parallel beta strands.
  - Researchers have collected empirical evidence that such deformations are typically larger for anti-parallel sheets than for parallel sheets.



Loops

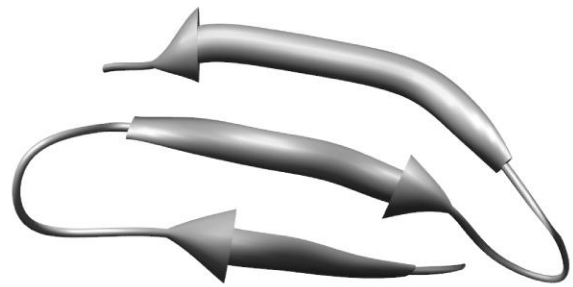
- Loops are chains of amino acids that have no particular hydrogen bonding patterns with other parts of the protein.



1ANS: a neurotoxin from the sea anemone, Anemonia sulcata.

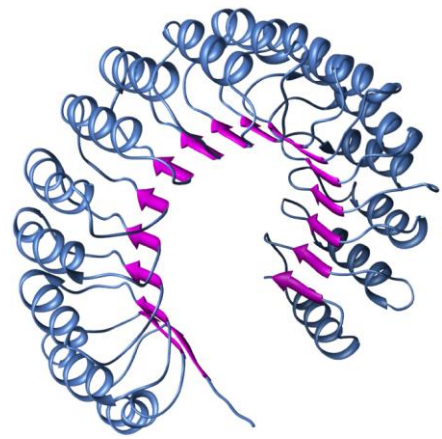
Loops (cont.)

- Proteins that are “all loop” are fairly rare.
  - Most proteins will have beta sheets and helices forming a hydrophobic core and these secondary structures will be interconnected by loop segments.



Tertiary Structure

- Helices, Sheets, and Loops combine to give a complete molecule in a three dimensional conformation:

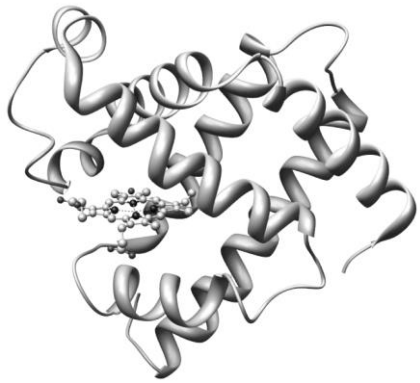


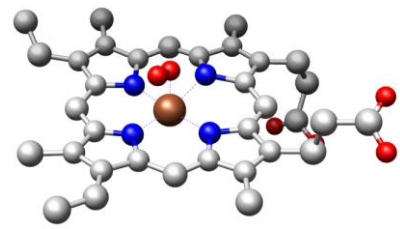
Chain D of 1A4Y

Tertiary Structure: Myoglobin (1)

- A globin fold:
  - 1MBN: 153 residues forming 8 helices with short loops forming a hydrophobic pocket containing a heme group.

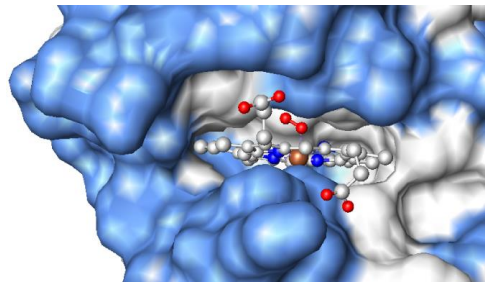
(A good example of structure supporting function).

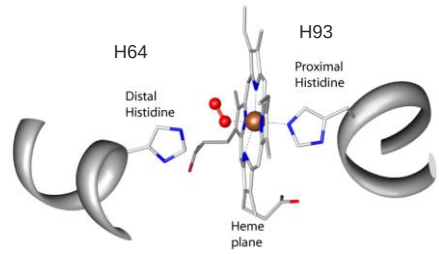




Tertiary Structure: Myoglobin (2)

- Heme group in the globin pocket:





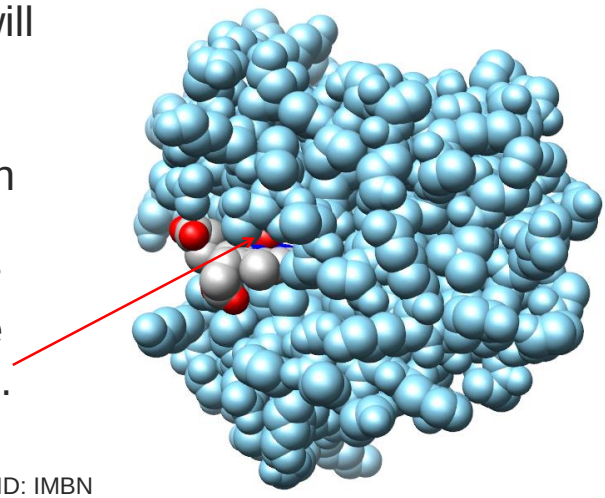
Tertiary Structure and Function

- The tertiary structure may be quite extensive while the binding site is comparatively small.
- What is the overall structure doing?
  1. Controlling the correct folding pathway.
  2. Forming the binding site to support a chemical reaction.
  3. Providing specific types of flexibility (Eg.: Allostery).
  4. Maintaining stability of the protein and allowing eventual degradation.
     - Not always fully successful (consider prions).

  In summary: the tertiary structure has specific requirements over the entire “life cycle” of the protein.

Flexibility & Functionality

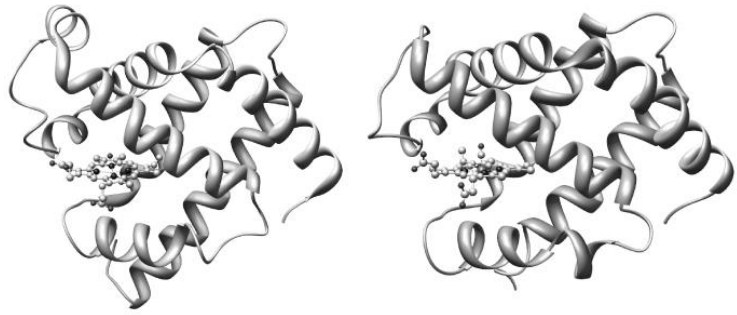
- Myoglobin will slightly change conformation when accepting or donating the O_2 molecule.



PDB ID: IMBN

Conservation of Structure (1)

- The globins are an excellent example of structure being more conserved than sequence.



Myoglobin (1MBN) One chain of hemoglobin (1JEB)

Conservation of Structure (2)

- Tertiary structures can be very similar, even when primary sequences have a relatively low similarity.
  - In the globin family, similar folds have been seen for very low sequence identities (less than 15%).

```
1MBN   VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEA
1JEB   SLTKTERTIIVSMWAKISTQADTIGTETLERLFLSHPQTKTYFPHFDLHPGSA

1MBN   EMKASEDLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFI
1JEB   QLRA------HGSKVVAAVGDAVKSIDDIGGALSKLSELHAYILRVDPVNFKLL

1MBN   SEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
1JEB   SHCLLVTLAARFPADFTAEAHAAWDKFLSVVSSVLTEKYR------
```

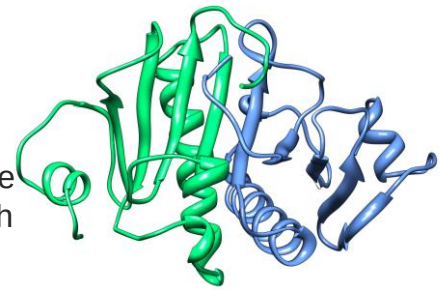
Distal H

Proximal H

Glutamic acid (E mutated to V in Sickle-cell anemia)

Quaternary Structure

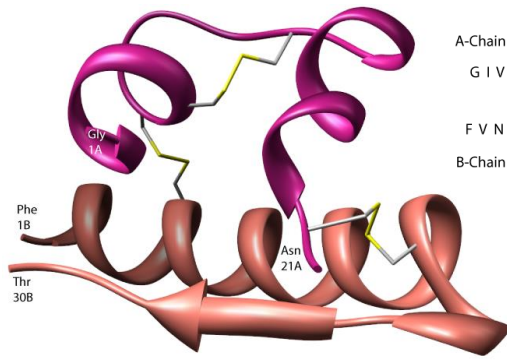
- Chains may combine to give a higher level structure.
  - Here we have the complete protein: Kinase C Interacting Protein (both chain A and chain B).
    - Often the tertiary components are replicates.

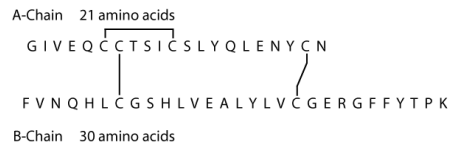


1KPA

Quaternary Structure: Insulin (1)

- Insulin is a trimer of dimers.
  - We do a bottom-up construction starting with the insulin monomer (two chains connected by disulfide bridges):





Quaternary Structure: Insulin (2)

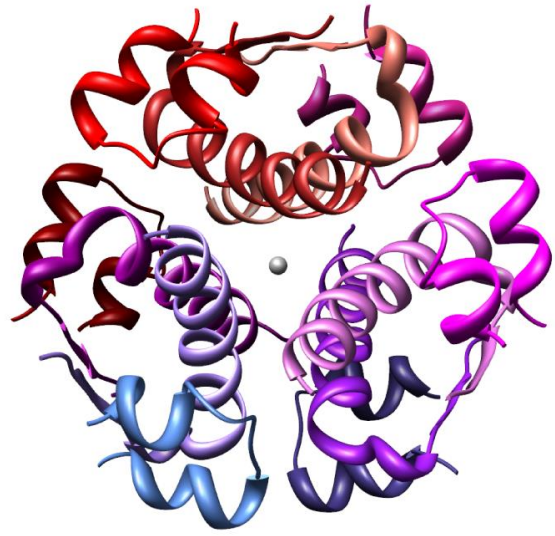
- Two of these monomers make up the insulin dimer by having their beta strands combine in an anti-parallel fashion to make a small beta sheet.

Quaternary Structure: Insulin (3)

- To understand how the dimers combine to make a hexamer, imagine taking the previous dimer and tilting the entire assembly so that your line of view is such that the darker zinc atom is exactly behind the brighter zinc atom.
  - This will put the dimer into the exact orientation that the upper third of the hexamer shows in the next figure.
    - Notice that in this orientation, the beta sheets are seen on edge and so they are somewhat difficult to discern.

Quaternary Structure: Insulin (4)

- Hexamer structure of insulin.
  PDB ID: 1znj



Protein Domains (1)

- A domain is an independently folded region of a protein with its own stable hydrophobic core.
  - The observation of independent folding is due to experiments that have shown that the subsequence of amino acids in a domain can often fold into the final conformation even when the rest of the protein is absent.

Protein Domains (2)

- Domains can be structurally similar even though they are in proteins that are otherwise quite different.
  - Such domains usually show a high sequence similarity.
- Domains tend to be compact and globular.
  - Linkages between domains are often loop structures and hardly ever helices or beta strands.
- Domains have distinct solvent accessible surfaces and are typically separated by water molecules.

Protein Domains (3)

- Residues within a domain will contact other residues in the same domain.
  - There is very little contact, if any, between residues in different domains.
- Usually a domain is formed from a contiguous residue sequence of the protein.
  - In rare cases a domain is made up of two or more regions of protein that are sub-sequences from one or more polypeptide chains.
  - This is why expert visual inspection is often required to determine domains.

Protein Domains (4)

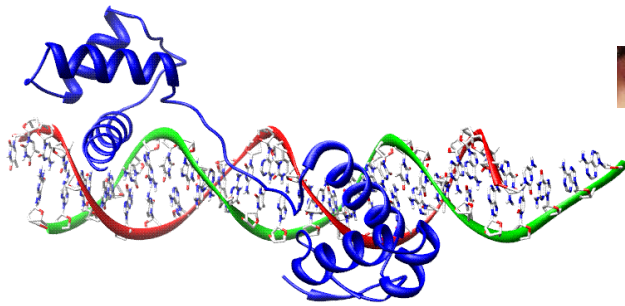
- Multiple domains within a single protein often cooperate to accomplish related tasks.
  - For example, reverse transcriptase protein of HIV has various domains that work together to convert viral RNA to DNA followed by destruction of the viral RNA.
  - So, by studying the domain structure of a protein we can strive to understand how design principles facilitate particular types of functionality.
  - Although this is an important goal, the mechanisms of cooperative function are not fully understood in many cases.

Protein Domain Example: 6PAX (1)

- The next figure presents a simple example of the 6PAX protein that uses two domains to interact with a segment of DNA.
  - This protein is part of a family called the homeodomain proteins.
  - They interact with DNA to provide transcription regulation and they are important in development and cell differentiation.

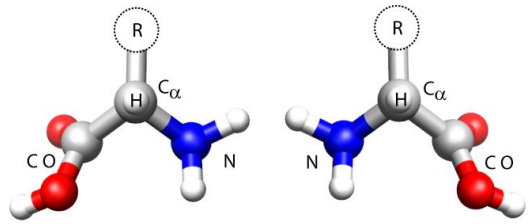
Protein Domain Example: 6PAX (2)

- When the 6PAX protein interacts with DNA the event initiates a cascade of protein interactions that leads to the development of various eye structures.
  - In the figure it is clear that the protein has two hydrophobic cores each responsible for interacting with a particular DNA segment.
  - The two cores are linked by a rather lengthy loop structure.



Mutations in PAX 6 lead to aniridia (no iris).

Protein Chirality



L-form D-form

- Looking down the H-C_{alpha} bond (from H) the L-form can be read clockwise as “CORN”.
  - The D-form can be read as “NRCO”.
- Most protein amino acids have the L-form.
  - The D-form can be found but is very rare.

RNA: Introduction

- RNA functionality:
  - RNA has information storage capability.
  - RNA can act as an enzyme.

- Types of RNA:
  - mRNA
  - tRNA
  - rRNA
  - ncRNA

Types of RNA

- Messenger RNA (mRNA)
  - Carries genetic information from DNA to the ribosome where it directs the biosynthesis of polypeptides.
- Transfer RNA (tRNA)
  - Transfers a particular amino acid to the growing polypeptide when its anticodon region recognizes the corresponding codon in mRNA.
- Ribosomal RNA (rRNA)
  - Found in ribosomes. Acts as the catalytic agent in protein synthesis.
- Non-coding RNA (ncRNA)
  - Active in biological processes such as regulation of transcription and translation, replication of eukaryotic chromosomes, RNA modification and editing, mRNA stability and degradation, and protein translocation.

Types of ncRNAs

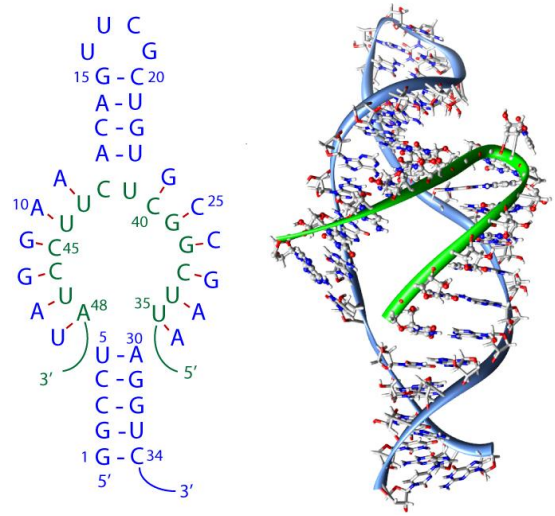
- tmRNA
  - tmRNA is best described as a combination of tRNA and mRNA.
- snRNA
  - Small nuclear RNA molecules are part of nuclear ribonucleoprotein particles responsible for splicing of eukaryotic mRNAs.
- snoRNA
  - Small nucleolar RNA molecules are involved in rRNA modification.

RNA Functionality Example (1)

- Interaction between H/ACA snoRNP and its RNA substrate:
  - There is a bulge in the RNA stem that forms the so called "ψ pocket".
  - This pocket combines with a loop from a different RNA, in this case, S14.
  - This is done by establishing 12 base pairs between S14 and the ψ pocket.

RNA Functionality Example (2)

- The reason for this binding is to put S14 in a precise position relative to a Cbf5 protein (not shown in the figure) which performs a conversion of the uridine at position 41 to pseudouridine (designated as ψ).
- Thus, H/ACA snoRNA does sequence recognition of a loop in S14 followed by a structural orientation.

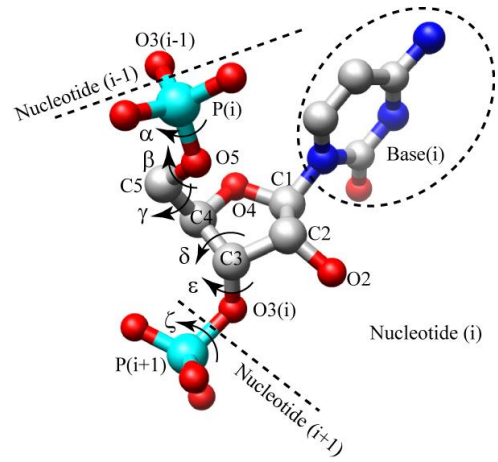


RNA: Primary Sequence (1)

- RNA is formed from nucleotides.
  - There are four different nucleotide *bases* in RNA: guanine, cytosine, adenine, and uracil. These are typically designated with the single letter codes G, C, A, and U respectively.

RNA: Primary Sequence (2)

- Each nucleotide unit has a backbone comprised of six atoms.
  - Starting just below the uppermost dashed line we see the phosphorous atom P(i) for the i^{th} nucleotide and just below that an oxygen atom labeled as O5.
  - Continuing along the backbone we encounter C5 of the ribose group followed by C4, C3 in the ribose ring, and finally O3.



RNA: Primary Sequence (3)

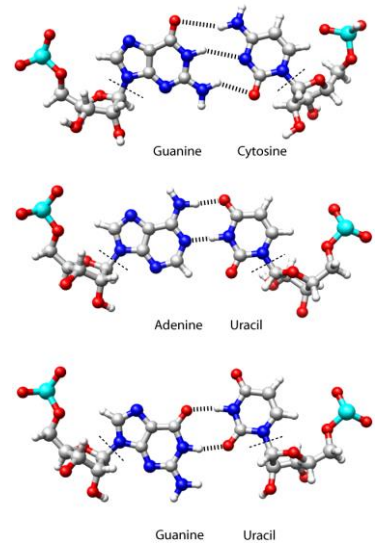
- RNA directionality:
  - The chain has a direction that is specified by the C5 and C3 atoms of the ribose group.
  - Consequently, moving along the backbone in the direction of nucleotides with increasing index means that we are going in the 3′ direction.
  - Moving in the opposite direction of decreasing nucleotide indexes is considered going towards the 5′ end.

RNA: Primary Sequence (4)

- There are six angles of rotation for each nucleotide.
  - These are labeled as α, β, γ, δ, ε, and ζ.
  - The “floppiness” of the molecule is reduced by means of extensive hydrogen bond formation between base pairs.

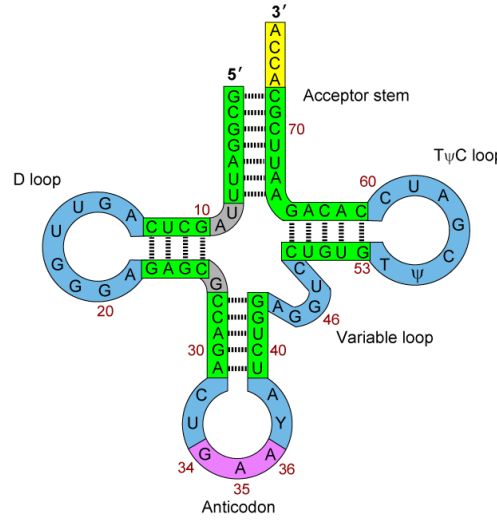
RNA Secondary Structure (1)

- RNA folds back on itself to form stem structures.
  - Stems are formed by complementary base pairing: GC and AU (sometimes GU).
    - GU base pairs are often called *wobble* base pairs.



RNA Secondary Structure (2)

- RNA base pairing can be illustrated using a two dimensional diagram that is drawn to show the pairs that are in association due to hydrogen bonding.
  - The figure does this for the yeast phenylalanine tRNA (PDB ID: 1EHZ).
  - This 2D drawing of the tRNA is done in the "cloverleaf" format.

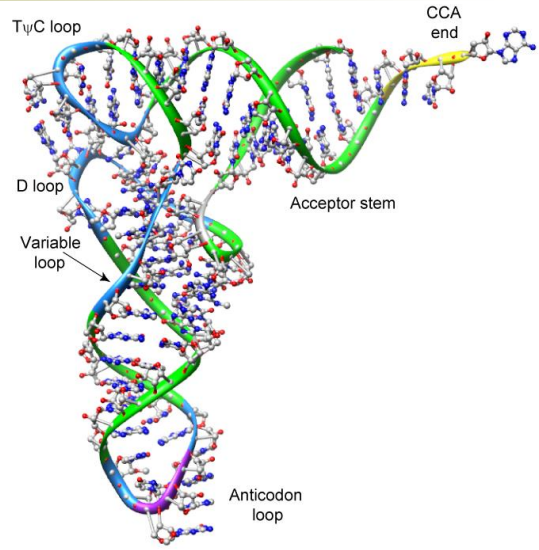


RNA Tertiary Structure (1)

- The stem structures of RNA tend to form short helical structures that are locally similar to the much longer double helix structure of DNA.
  - This gives a lot of stability to the molecule and helps to counteract the floppiness arising from the six torsional angles between consecutive nucleotides mentioned earlier when we introduced the structure of a nucleotide.

RNA Tertiary Structure (2)

- The figure to the right illustrates the beautiful 3D conformation of phenylalanine tRNA (PDB ID: 1EHZ).
  - The ribbon backbone has been given the same coloration as that used in the secondary structure.



_1

RNA Tertiary Structure (3)

- Pseudo-knots:
  - There is additional hydrogen bonding that can occur outside of the stem regions.
    - It is also possible to have extra base pairing between nucleotides that are in different loop regions.
    - With a careful inspection of the last figure one can see this occurring between nucleotides in the TψC loop and the D loop.
    - This type of base pairing is called a *pseudo-knot*.
      - These pairings are important structural features because they help to increase the conformational stability of the molecule.

RNA Tertiary Structure (4)

- Pseudo-knots in $tRNA^{Phe}$:

[U(8)::A(14)]	[G(10)::G(45)]	[G(15)::C(48)]
[G(18)::ψ(55)]	[G(19)::C(56)]	[G(22)::G(46)]
[A(9)::A(23)]	[G(26)::A(44)]	[T(54)::A(58)]

  - The symbol ψ represents pseudouridine, a post-transcriptionally modified ribonucleotide.

RNA Tertiary Structure (5)

- Pseudo-knots complicate base pairings:

Note:
A(23) and U(12) are in the $tRNA^{PHE}$ D stem!

SO:
If a nucleotide is in a base pair, it can still be involved in additional pairing when pseudo-knots occur.