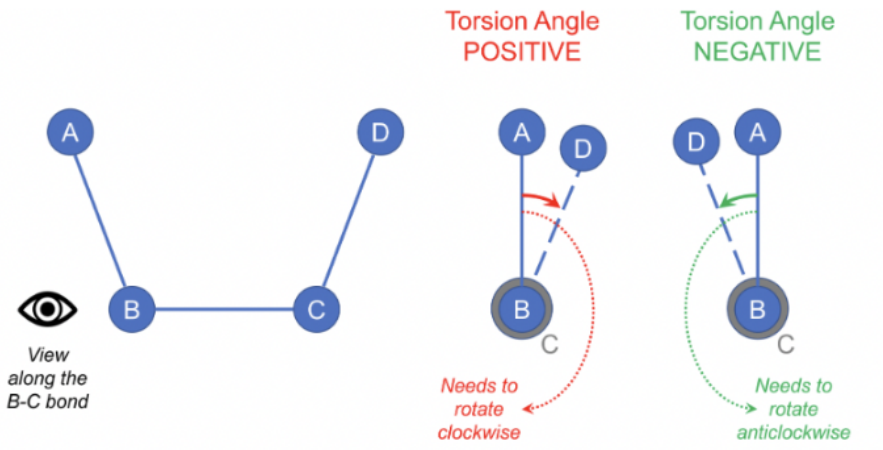
Protein Structure

Course	🍖 Essential Protein Structure and Function
Confidence	Not Confident
Next Review	@April 23, 2024
Last Edited	@May 28, 2024 10:42 PM

Amino acids - the structural alphabet - diverse chemical roles

- Six hydrophilic amino acid sidechains
  - Serine Ser S, Threonine Thr T, Tyrosine Tyr Y, Asparagine Asn N, Glutamine Gln Q, Cysteine Cys C
  - Hydrogen bonds with water, usually found at protein surfaces
- Two acidic amino acids
  - Aspartate Asp D, Glutamte, Glu, E
  - Hydrogen bond with water, contain ionisable groups
- Three basic amino acids
  - Lysine Lys K, Arginine Arg R, Histidine His H
  - Hydrogen bond with water, contain ionisable groups
  - The side chain of arginine can donate up to five hydrogen bonds.
- Nine hydrophobic amino acid sidechains
  - Glycine Gly G, Alanine Ala A, Valine Val V, Leucine Leu L, Isoleucine Ile I, Methionine Met M, Phenylalanine Phe F, Tryptophan Trp W, Proline Pro P (imino acid)
  - No hydrogen bonds with water, Buried in proteins
  - The amino acid proline lacks a backbone amide proton when polymerised into a polypeptide chain.

Dihedral torsion angle



- The atoms used to define the backbone peptide bond (omega) angle for the peptide bond between residues i and i+1 are Cα(i)—C′(i)—N(i+1)—Cα(i+1).
- The atoms used to define the backbone phi angle of residue i are C′(i-1)---N(i)---Ca(i)---C′(i) .
- The atoms used to define the backbone psi angle of residue i are N(i)---Ca(i)---C'(i)---N(i+1).

The peptide bond is rigid

- There are three bonds between every R group. One of these is a partial double bond and cannot rotate - it is rigid (omega is fixed), the ther two bonds (phi and psi) can rotate

Cis- and trans-peptide bonds

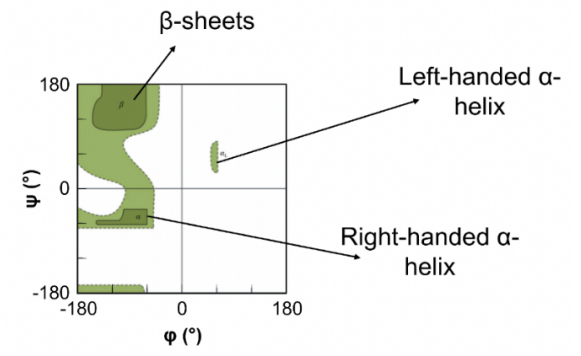
- The vast majority of peptide bonds are in the trans-configuration, presumably becauce of steric interactions between sequential side chians
  - Usually not allowed except of proline
- For Xaa-Pro peptide bonds, the trans-configiration is only marginally more stable than the cis- form

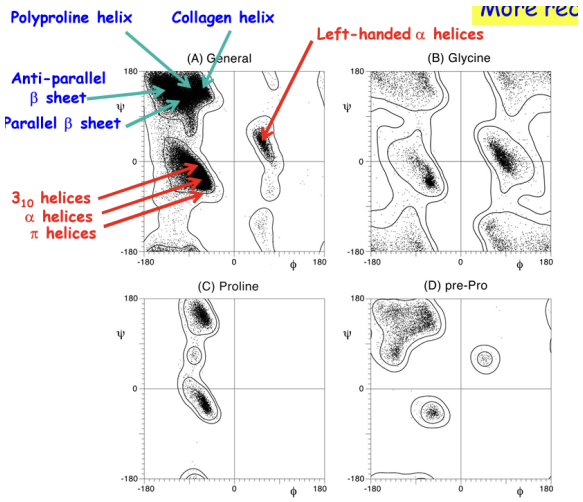
The polypeptide chain conformation is defined by 2 angles per residue: phi and psi

- Each peptide bond hehaves as a planar unit (omega, w=180)

Ramachandran plot

- Many psi and phi angles are not allowed, and those that are allowed show up as clusters





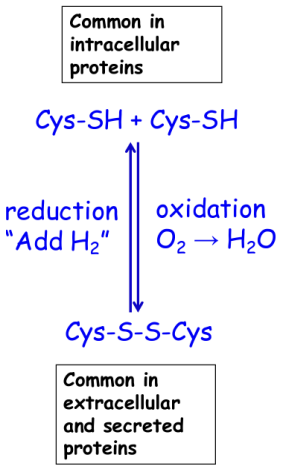
- Glycine can populate both sides of the Ramachandran plot

Properties of amino acid side chains

Metal ions in proteins

- Fe in ribonucleotide reductase
- Zn in alcohol dehydrogenase

Cysteine - disulphide bonds



- Stabilise protein
- Intracellular: more acidic, thus reduce the bond
- Extracellular: more basic, thus oxidise and form the bond

- Glycine - very tiny - hence conformationally mobile
- Histidine - the only amino acid with a pKa close to physiological
- Acid and basic amino acid - ion bridges, catalytic sites, etc.
- Proline - imino acid - different conformational properties
- Grouping amino acids
  - Tiny: GAS
  - Aliphatics: IVLM
  - Basic: RKH
  - Acidic: DE

Secondary structure

- Secondary structure backbones are the most obvious feature of a protein structure - hydrogen bonds
- Hydrogen bonding defines the secondary structure - and also the structure of water
- There are many different ways of representing a protein structure
  - More typical representations of protein folds: ribbon diagrams

Alpha-helices

- 3.6 residues per turn - 5.4Å per turn, 1.5Å per residue
- Right-handed (i.e. clockwise from N terminus)
  - Bacterial: left-handed
- Full set of H-bonds between C=O and N-H parallel to helix
  - NH (i+4)—C′O (i)
- N-H groups point back along helix direction
- a-helix variations (both right-handed)
  - NH(i+3)—C′O(i): 3(10) helix - overwound helix
  - NH(i+5)—C′O(i): pi helix - underwound helix

Side chains on an a-helix

- Side chains project away from a-helix
  - Creating the “helical wheel”
- Tend to occur as pairs
  - e.g. two hydrophilic 2 and 3 folloed bout two hydrophobic 4 and 5

a-helix dipole

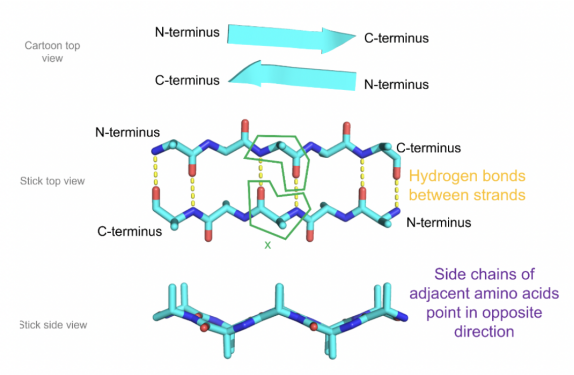
- Net polarisation of bonds is thought to stabilise the coordination of negatively charged groups at N-terminal end
  - e.g. phosphates

Beta-strands and beta-sheets

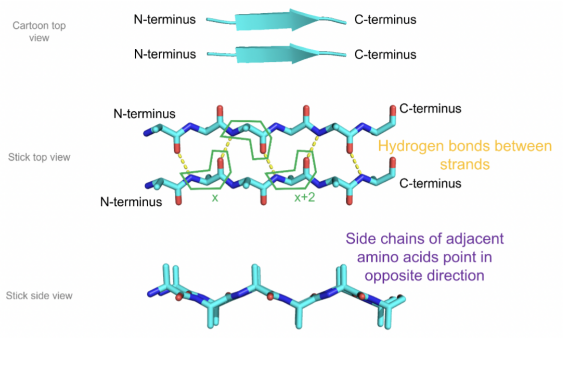
- A b-strand is close to the most extended form of the polypeptide chain possible
- N-N distance ~3.6Å (compared to C(a)-C(a) ~3.8Å)
- b-sheets are made up of H-bonded b-strands from different regions of the protein
- Side chains alternate along a single strand and point to either side of the sheet
- Side chains on neighboring strands point to the same side

Anti-parallel b-sheet

Parallel beta-sheet



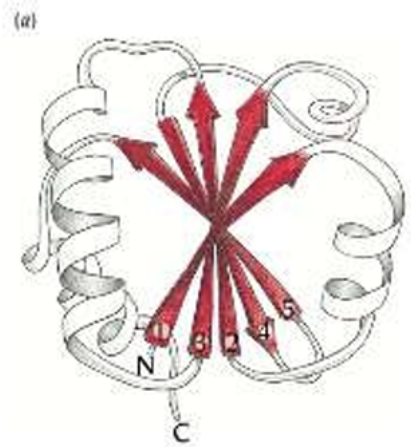
- Neighbouring strands extend in opposite directions
- Side chains tend to occur in alternation, i.e., one hydrophilic, followed by hydrophobic
- Note that H-bonds are anti-parallel too
- Two hydrogen bond between a pair of amino acid

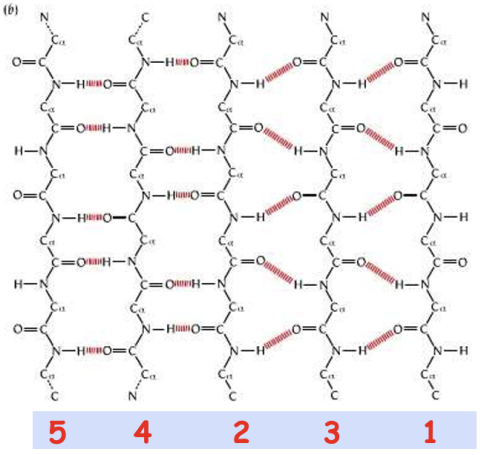


- Polypeptide strands run in the same direction (*i.e.* from N- to C-terminus)
- Note that H-bonds are not parallel
- One amino acid bond to 2 residues on the adjacent strand

Mixed beta-sheet

- Beta sheets are not flat
- Each strand has a right-handed twist
    - Thioredoxin, note that the arrow heads can be either way

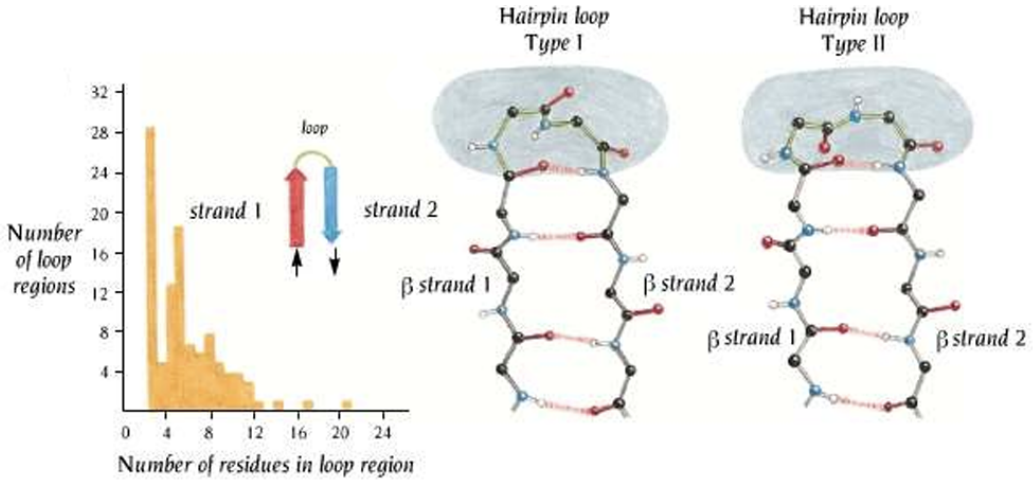




Beta-sheet configurations are often described by topology diagrams

The topologies can be described mathematically and therefore can be useful for assessing structural similarity of different proteins in an automated fashion

Loops



- Strands in anti-parallel beta sheets are often connected by short loops
  - Important in structural stabilisation
- The shortest loops (2 residues) are referred to as ‘tight turns’ or hairpin loops.
  - The confirmation of the loops has been classified by analysis of the phi/psi angles
  - Some types of hairpin use a positive phi angle to make the turn - hence the preponderance of Gly residues in such structure

Suprasecondary Structures

- There are now about 1350 unique protein folds and 148,000 structures
  - There are so many ways of pernuting a-helices and b-strands in domains up to 250 amino acids (limited by the number of amino acid in protein fold
  - There is an exponential growth in the released structure and is saturating

Four common motifs

- Protein architectures are comprised of simple structural motifs

EF-hand motif

- Subset of helix-turn-helix motif
- E.g. troponin-c: 4 EF motifs, 2 with calcium

Beta-hairpin

- E.g. bovine trypsin inhibitor
- Snake venom erabutoxin

Greek key

- Staphylococcal nuclease
- Perhalps the Greek key motif is an easily accessible extension of a b-hairpin, which makes this motif common
- There are 24 different ways to combine 2 beta-hairpin motifs into a 4 strand beta-sheet
    - Yet, examples have been observed only for 8 of. them, and the vast majority are the Greek keys (i) and (v)

Beta-alpha-beta

- The ends of strands in a parallel b-sheet get connected by a cross-over a-helix
- There are two different b-a-b motifs possible
    - Right handed (helix above)
    - Left handed (helix below)
    - Only **right handed** b-a-b motifs are observed with unclear reason

Simple motifs combine to form more complex motifs

Protein folds are created from motifs

Proteins are constructed as hierarchical structures

Secondary structure → motifs → domains → tertiary structure → quaternary structures

Large proteins are made up of multiple domains, each with between 40-50 aa up to 200-250 aa

Protein folds are classified into groupings

All-alpha class (helical structures)

alpha/beta class (parallel beta-sheets with b-a-b motifs - 2 types

Beta-cass (anti-parallel beta sheets)

alpha+beta class (minor class - discrete alpha- and beta-subdomains)

Small domains held together by multiplr S-S bonds metal ions (e.g. zinc fingers)

Unstructured (not all parts of all proteins are ordered)

Collagen triple helices as well

All- or mainly-alpha class protein folds

The a/b-class of protein folds

- B-a-b motifs as building blocks

Triose phosphate isomerase (TIM) barrel

- 8 strands will close the barrel
- Helices on one side of the sheet, enzyme sites at the terminal end

Rossman (open twisted) fold

- Helices on both side of the sheet, binding crevice at the C-terminal end

The beta-class of protein folds

- 2 anti-parallel beta-sheets
- Superoxide dismutase - 8 antiparallel beta-strands plus two metals
- Muramidase (flu) - 6 small beta-sheets with 4 anti-parallel beta strands