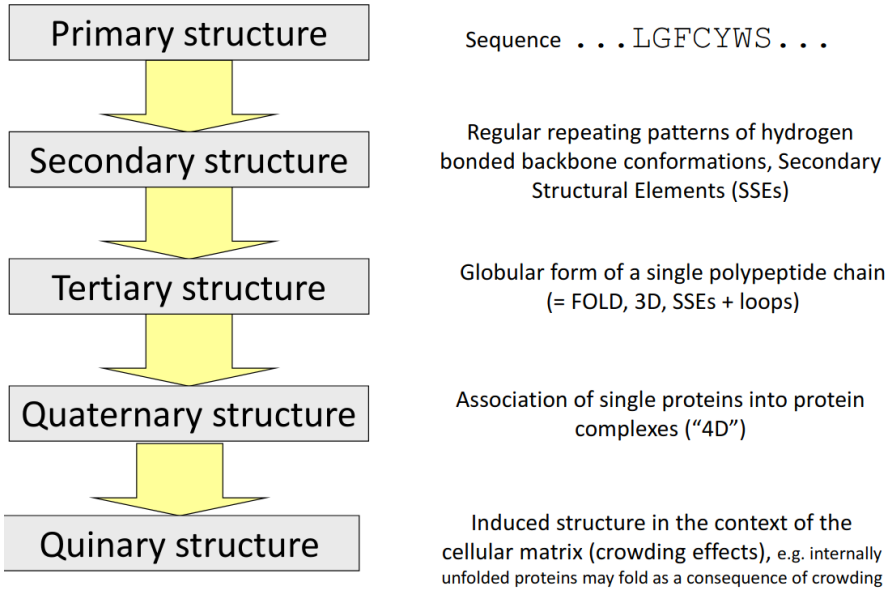
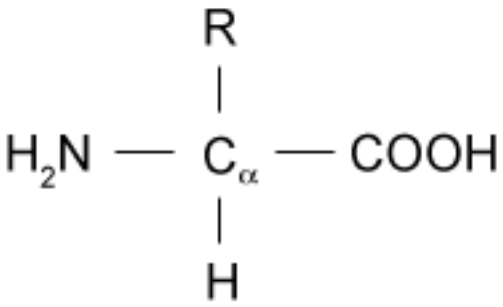
2.1 Principles of protein structure

Introduction



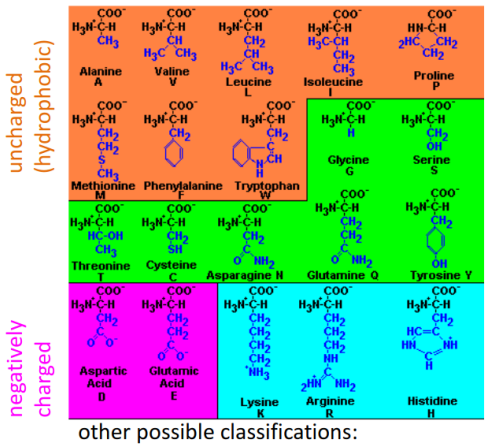
Proteins are biopolymers composed of amino acids in L form.



Primary structure

The peptide bond is formed by a condensation reaction between the OH of the carboxyl group and the amino group.

Proteins suffer often of post-translational modifications, which contribute to their diversity, and makes predicting their folded structure more difficult.



other possible classifications:

- size (Gly=small, Trp=large)
- aliphatic (Ala, Val, Ile, Leu)/aromatic (Phe, Tyr, Trp)
- sulfur containing (Met)

Amino Acid	3-Letter[80]	1-Letter[80]
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

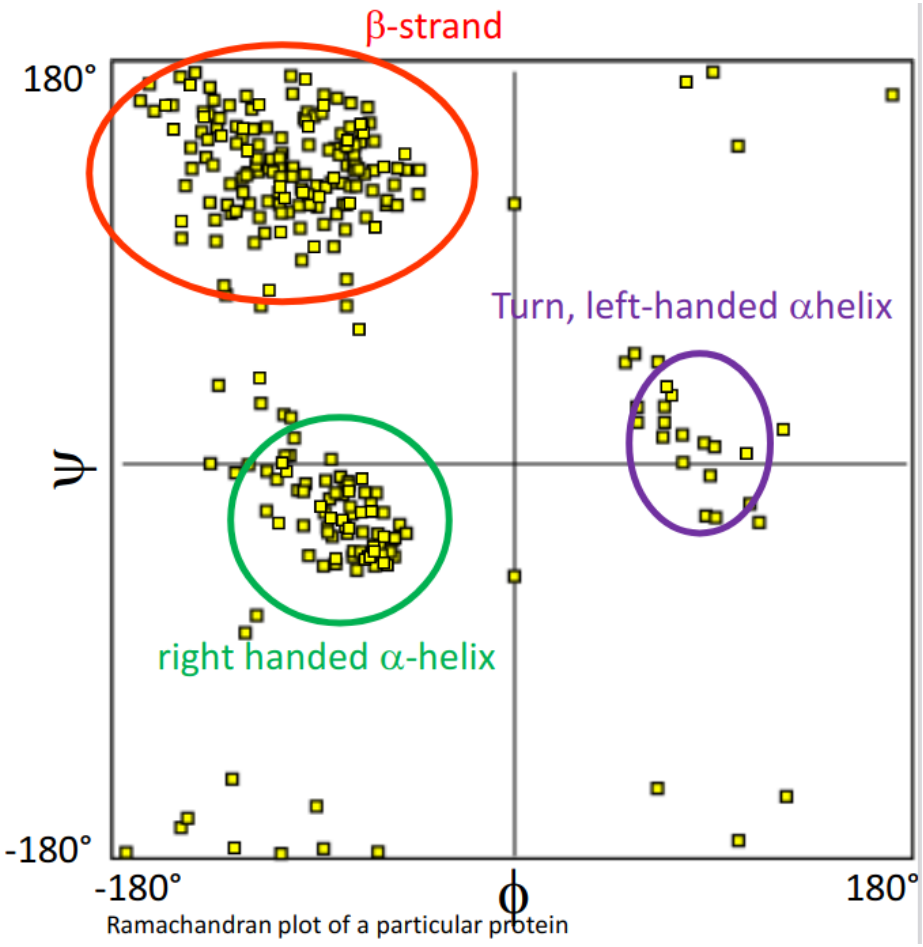
8

(Approximate) classification of polymers in function of the length of the sequence: - Oligopeptide: 1-10 amino acids - Polypeptide: 10-50 amino acids - Protein: +50 AA

Secondary structure

Rotation angles found in the backbone determine the geometry of the molecule and have different names: - ϕ Rotation around N-C_α. Not found at N-terminus - ψ Rotation around c_α-$c_{peptidic}$. Not found at C-terminus - ω around peptide bond: Partial double bond

Ramachandran plot: represents the angle combinations between ϕ, and ψ angles of each residue.



Hydrogen bonds: Proteins can form hydrogen bonds between mainchain atoms. These contribute to the formation of characteristic repetitve structural elements.

3_{10}-Helix

- H-bonds: i $\rightarrow$ i + 3
- 3 AA per turn
- rare, short

α-Helix

- Dihedrals: $(\phi, \psi) = (-60^o, -50^o)$
- H-bonds: $i \rightarrow i + 4$
- 5.4 Armstrong and 3.6 AA per turn

π-Helix

- 4.4 AA per turn
- Hbond (i $\rightarrow$ i + 5)

β-sheets

Non-local hydrogen bonds

3.5 A between strands

Parallel: - $(\phi, \psi) = (-120^o, 115^o)$

Antiparallel: - $(\phi, \psi) = (-140^o, 135^o)$

Turns

- H-bonds: i $\rightarrow$ i + 3
- 180 degrees turn

Secondary structure assignment

- DSSP (Kabsch & Sander, 1983)
- STRIDE (Frishman & Argos, 1995)
- DEFINE (Richards & Kundrot, 1988)
- P-Curve (Sklenar, Etchebest, Lavery, 1989)

DSSP

The so-called Dictionary of Secondary Structure of Proteins (DSSP) by Kabsch and Sander (1983a) performs its sheet and helix assignments solely based on the backbone-backbone hydrogen bonds. The DSSP method defines a hydrogen bond when the bond energy is below -0.5 kcal/mol from a Coulomb approximation of the hydrogen bond energy. The structure assignments are defined such that visually appealing and unbroken structures result. In case of overlaps, a-helix is given first priority, followed by b-sheet. This procedure does not affect the Coulomb approximation, rather the realisation of 'unbroken structures' addresses the step from individual hydrogen bonds to assigning macro-structures to groups of such bonds.

An a-helix assignment (DSSP state 'H') starts when two consecutive amino acids have i→ i+4 hydrogen bonds, and ends likewise with two consecutive i→ i-4 hydrogen bonds. This definition is also used for 310-helices (state 'G' with i → i+3 hydrogen bonds) and for p-helices (state 'I' with i→ i+5 hydrogen bonds) as well. The helix definition does not assign the edge residues having the initial and final hydrogen bonds in the helix. A minimal size helix is set to have two consecutive hydrogen bonds in the helix, leaving out single helix hydrogen bonds, which are assigned as turns for all three helices (state 'T').

b-sheet residues (state 'E') are defined as either having two hydrogen bonds in the sheet, or being surrounded by two hydrogen bonds in the sheet. This implies three sheet residue types: anti-parallel and parallel with two hydrogen bonds or surrounded by hydrogen bonds. The minimal sheet consists of two residues at each partner segment. Isolated residues fulfilling this hydrogen bond criterion are labelled as b-bridge (state 'B'). The recurring H-bonding patterns connecting the partnering strands in a b-sheet are occasionally interrupted by one or more so-called b-bulge residues. In DSSP these residues are also assigned as b-sheet 'E' and may comprise up to four residues on one strand and up to one residue on the partnering strand. These interruptions in the b-sheet H-bonding pattern are only assigned as sheet if they are surrounded by H-bond forming residues of the same type, i.e. either parallel or anti-parallel. The remaining two DSSP states 'S' and ' ' (space) indicate a bend in the chain and the unassigned/other state, respectively.

Packing in proteins

Proteins are densely packed, in fact close to the optimal sphere packing.

Cavities are frequently found in proteins, can be filled by wather or empty

Helix-helix packing, with optimal angles. Helix-sheet packing.

Supersecondary structure

Small substructures consisting of few SSE and their interactions. Several motifs can be combined to domains: - greek key motif - $\beta - \alpha - \beta$ Motif

Domains

A protein domain is a region of the protein's polypeptide chain that is self-stabilizing and that folds independently from the rest. They have most often specific functions within a protein.

Construction of automated domain identification is often difficult.

Tertiary structure

Spatial organization of SSEs

Mainly stabilized by: - Hydrophobic interactions - Ionic interactions - Disulfide bridges - Hydrogen bonds

Quaternary structure

Multiple interacting subunits, forming Dimers, trimers... Examples: - ATP synthase - Ion Channels - hemoglobin

2.2 Visualization of proteins

Visual representations of proteins

Different models can be used for visualization of proteins: - Line model - Space Filling model - Ball & sticks - Main chain trace - C-Alpha trace - Simple diagrams - Cartoon diagrams

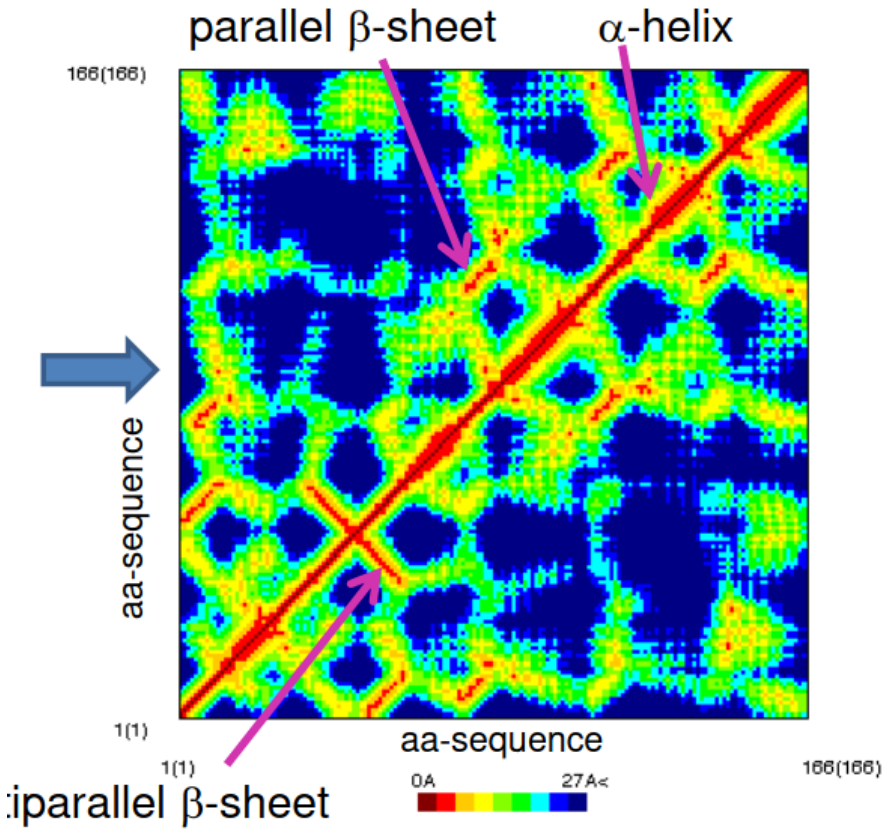
Molecular Surfaces

Can be represented using: - Van der waals surface: represents a imaginary hard sphere which can be used to model atoms as rigid bodies. - Solvent accesible surface: Van der Waals surface + 1.4A (surface not accesible is called Re-entrant or solvent excluded surface) - Dot Surface: Display points at a given distance from atom center (I.e Solvent accesible surface), and remove the interior dots - Gaussian surface (two parameters, can have different levels of smoothing)

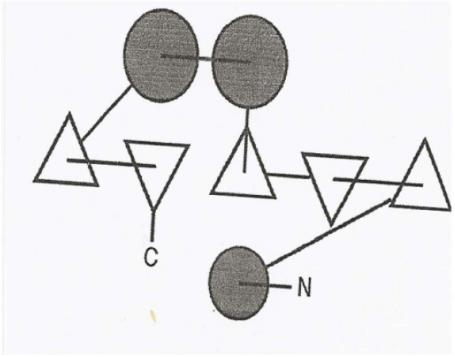
Marching cubes algorithm

Creates a grid, checks which points are inside the surface, which outside, and which in between, and assigns edges to the voxels that are between.

2-D Contact maps



TOPS diagram



- Represent both topology and interaction of structural elements in 2D (sequence of SSEs)
- Helices represented by circles
- β-strands represented by triangles
  - Upward-pointing triangle: strand pointing out of the page
  - Downward-pointing triangle: strand pointing into the page
- Connecting lines drawn to the centre of the symbol indicate connection to the top, and those drawn to the edge indicate connection to the base
- Loops connecting elements represented by lines
- Third dimension is implied

Radius of gyration

$R^2 = \frac{1}{N}\sum(r_k - r_{mean})^2 \rightarrow$ Squared eviation from the centroid along each axis