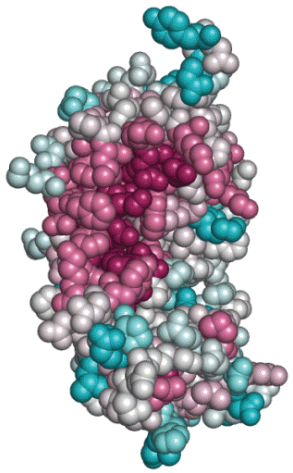
Bioinformatics for molecular biology

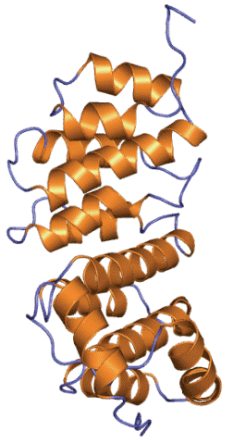
Structural bioinformatics tools, predictors, and 3D modeling – Structural Biology Review

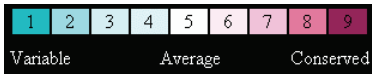
Dr Jon K. Lærdahl, Research Scientist

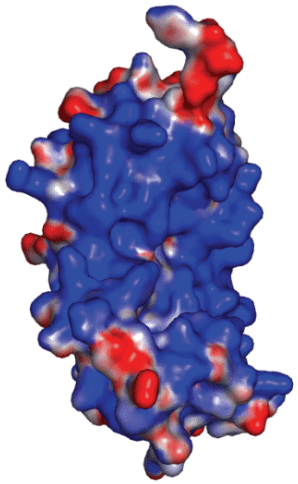
Department of Microbiology,
Oslo University Hospital - Rikshospitalet &
Bioinformatics Core Facility/CLS initiative,
University of Oslo

E-mail: jonkl@medisin.uio Phone: +47 22844784
Group: Torbjørn Rognes
(http://www.ous-research.no/rognes)
CF: Bioinformatics services
(http://core.rr-research.no/bioinformatics)
CLS: Bioinformatics education
(http://www.mn.uio.no/ifi/english/research/networks/clsi)
Main research area:
Structural and Applied Bioinformatics









Oslo universitetssykehus

Overview

Jon K. Lærdahl,
Structural Bioinformatics

Now:

- Protein Structure Review
  - Amino acids, polypeptides, secondary structure elements, visualization, structure determination by X-ray crystallography and NMR methods, PDB

Later…

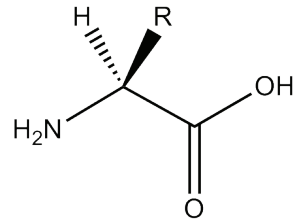
- Structure comparison and classification (CASP & SCOP)
- Predictors
- 3D structure modeling
  - *Ab initio*
  - Threading/fold recognition
  - Homology modeling
- Practical exercises
  - PyMOL & visualization
- Practical Exercises
  - Homology modeling of influenza neuraminidase (Tamiflu resistance?)
  - Other homology modeling
  - Threading
  - Your own project?

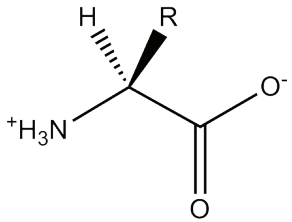
Stop me and ask questions!!

Amino acids – the building blocks of proteins

Jon K. Lærdahl,
Structural Bioinformatics

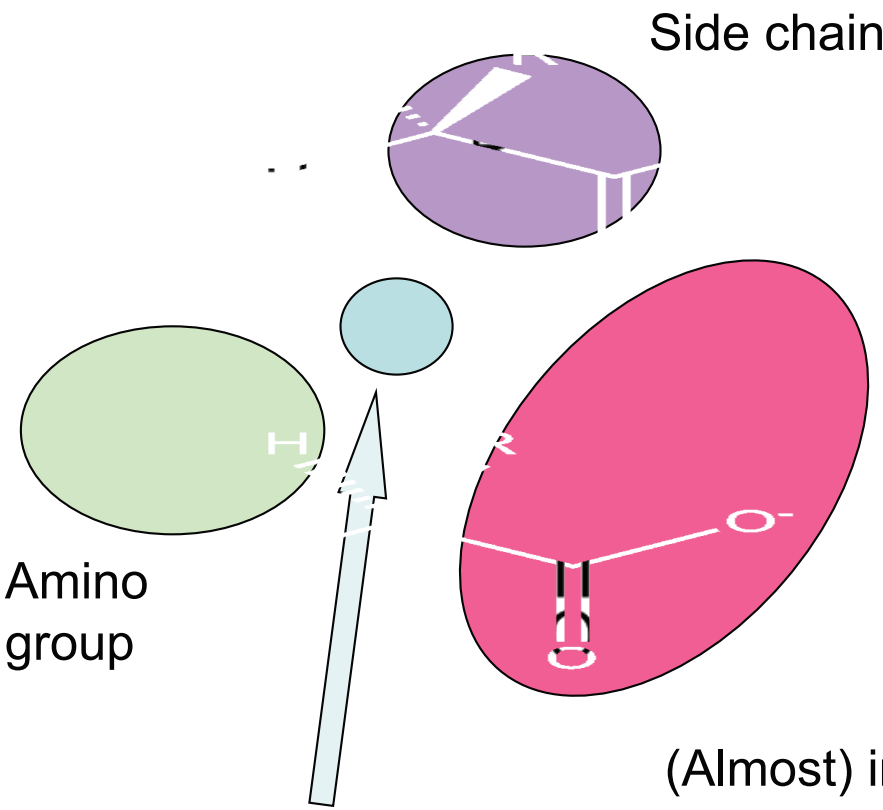
Proteins are built from 20 naturally occurring amino acids. They have an amino (-NH_2) and acidic (-COOH) functional group





Zwitterionic form found at physiological pH

The side chain group (R) determines the properties of the amino acid



(Almost) invariably enantiomeric L-form (or *S*-form)

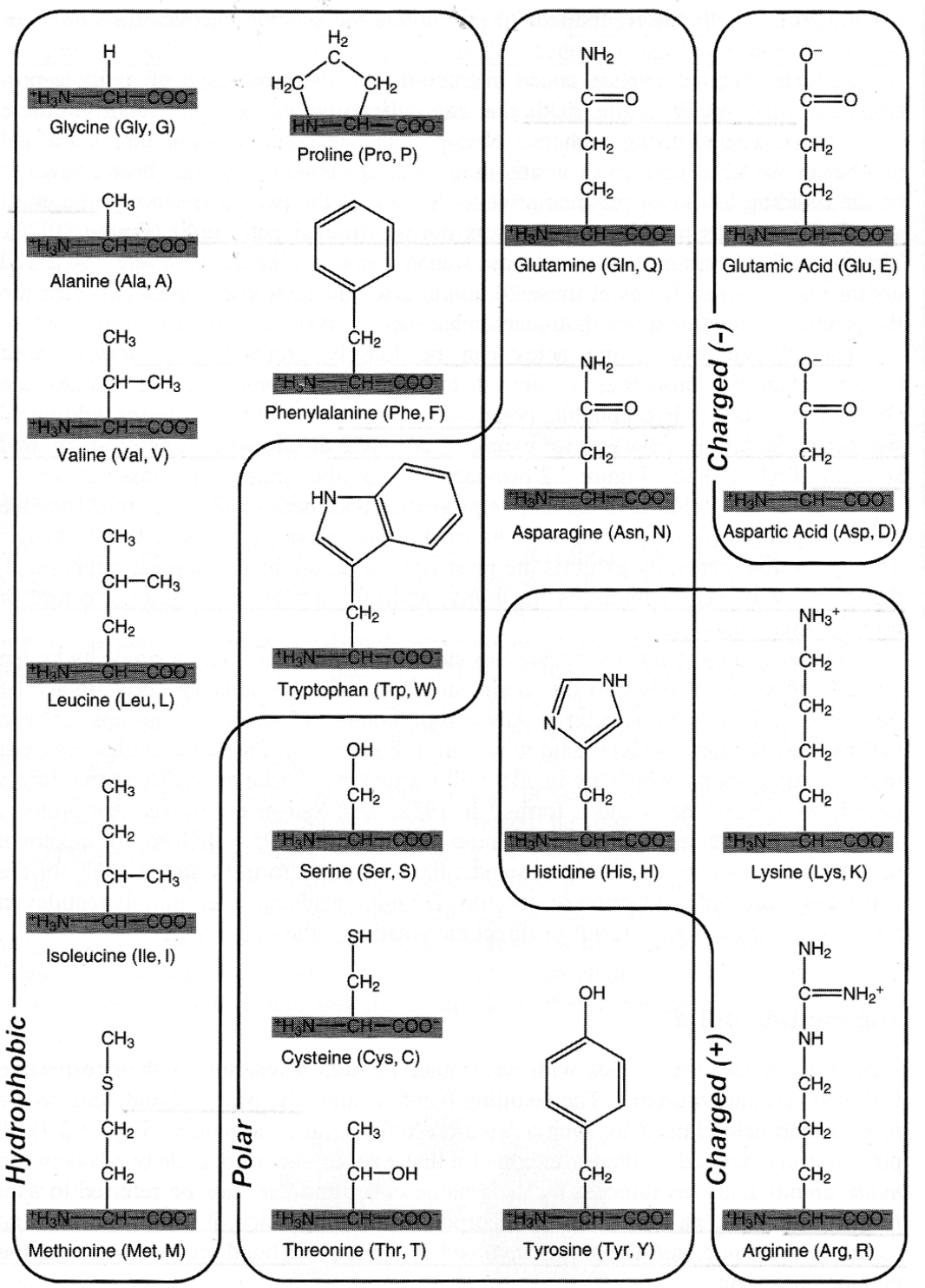
Amino acids

R-group properties:

- Large
- Small

- Hydrophobic
  - Aliphatic
  - Aromatic
- Polar
- Charged
  - Positive/negative charge

Increasing hydrophilicity/higher water (solvent) affinity

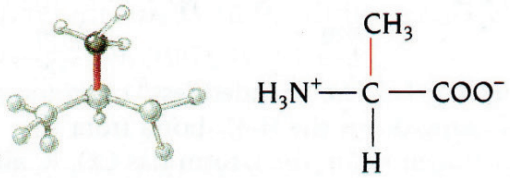


Structural Bioinformatics,
Eds. P.E. Bourne & H. Weissig
(Wiley, Hoboken, NJ, 2003)

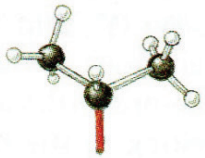
Amino acids

Introduction to Protein Structure, C. Branden & J. Tooze (Garland, New York, 1998)

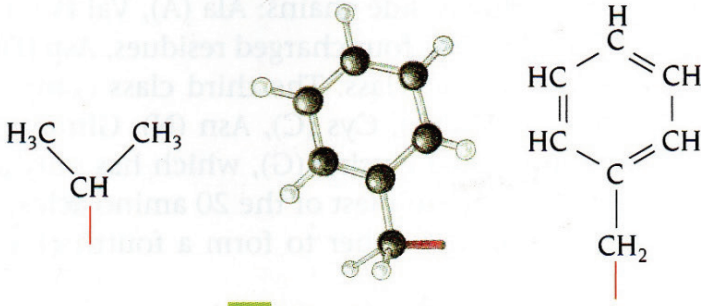
(a) Hydrophobic amino acids



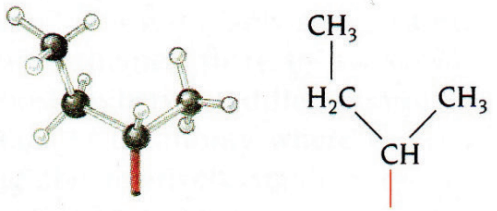
A Ala, Alanine



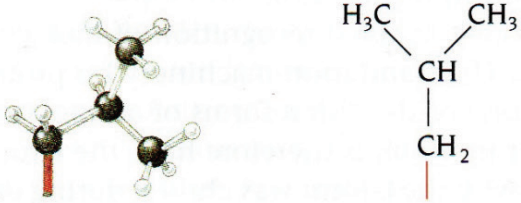
V Val, Valine



F Phe, Phenylalanine

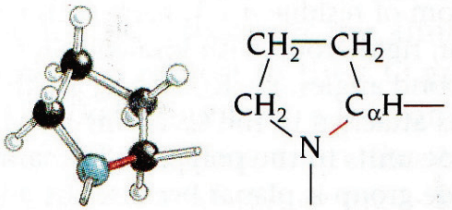


I Ile, Isoleucine

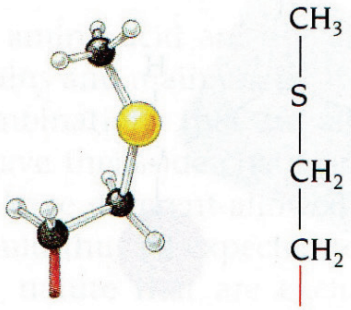


L Leu, Leucine

- Hydrophobic
  - Aliphatic
  - Aromatic



P Pro, Proline

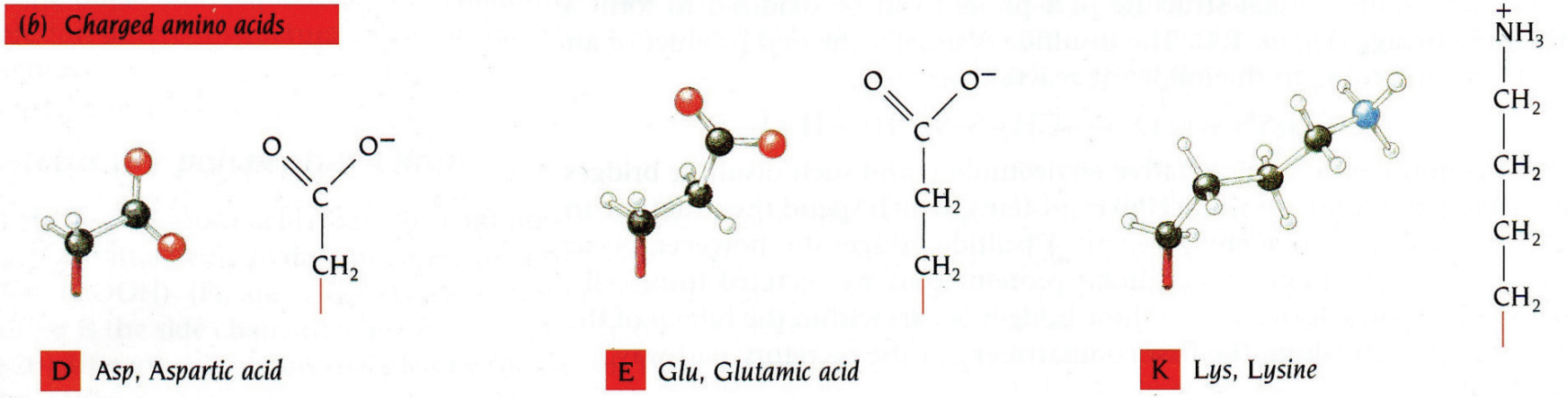


M Met, Methionine

- 3-letter code
- 1-letter code

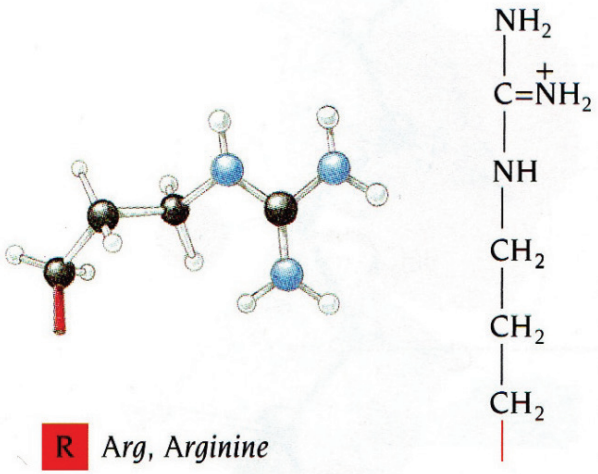
Amino acids

Introduction to Protein Structure, C. Branden & J. Tooze (Garland, New York, 1998)

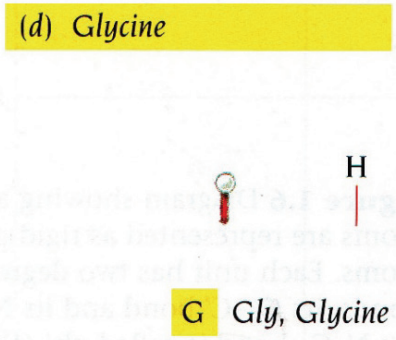


Aspartate

Glutamate



- Hydrophilic
  - Positive charge/basic
  - Negative charge/acidic

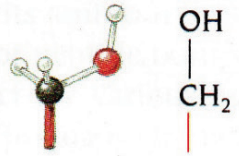


Amino acids

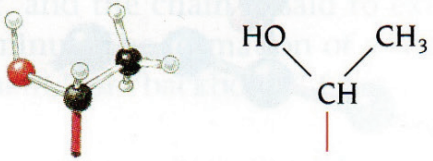
Introduction to Protein Structure, C. Branden & J. Tooze (Garland, New York, 1998)

(c) Polar amino acids

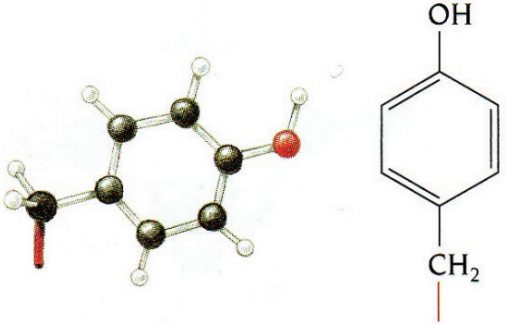
- Hydrophilic



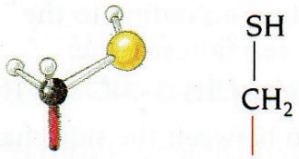
S Ser, Serine



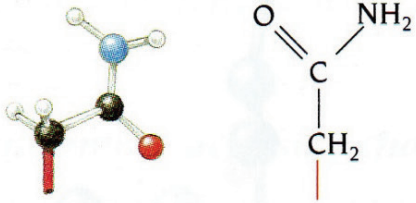
T Thr, Threonine



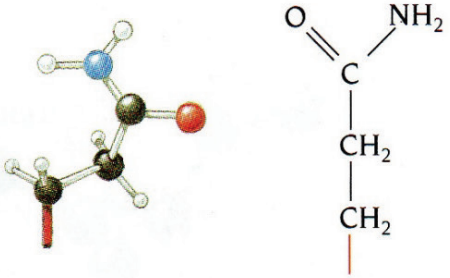
Y Tyr, Tyrosine



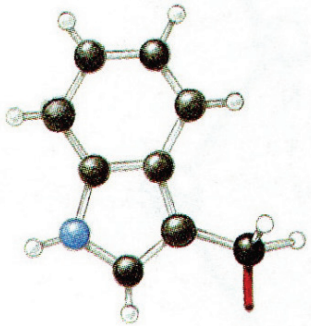
C Cys, Cysteine

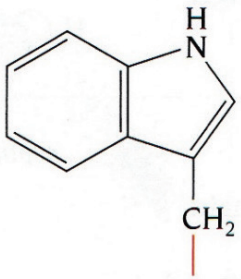


N Asn, Asparagine

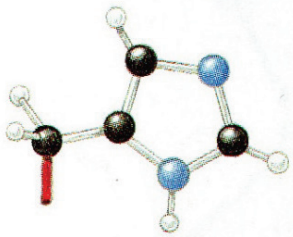


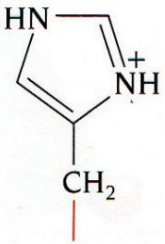
Q Gln, Glutamine





W Trp, Tryptophan

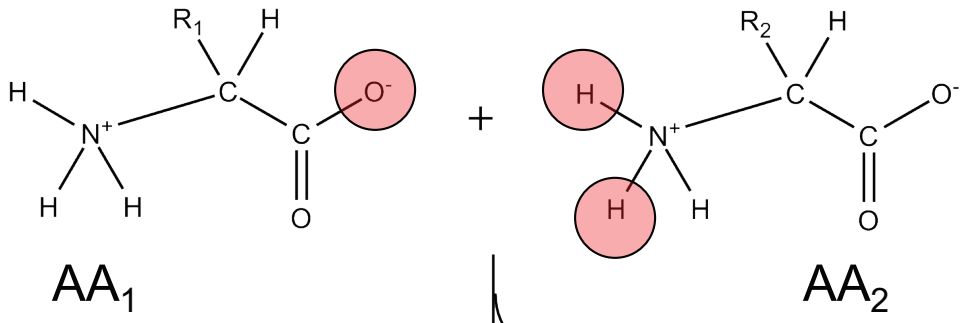




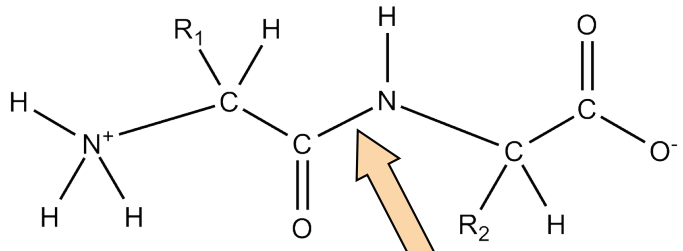
H His, Histidine

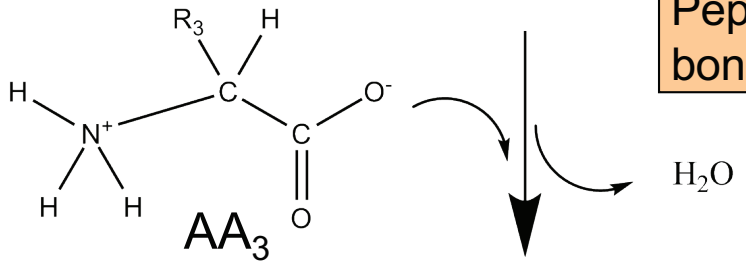
Polypeptides

Jon K. Lærdahl,
Structural Bioinformatics



Proteins are polypeptides, *i.e.* many amino acids connected by peptide bonds



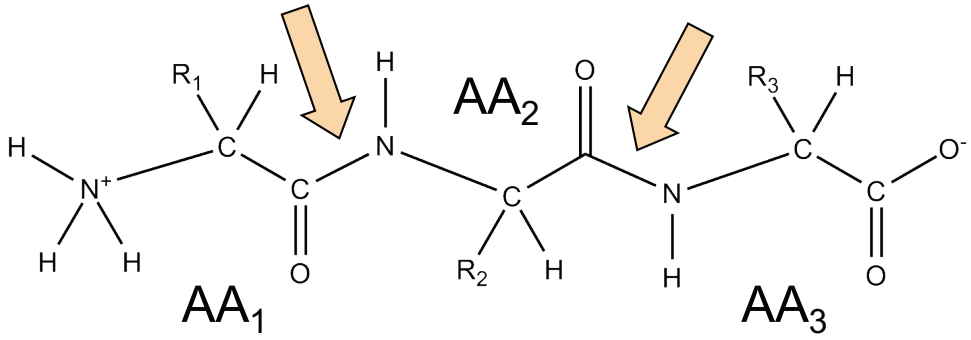


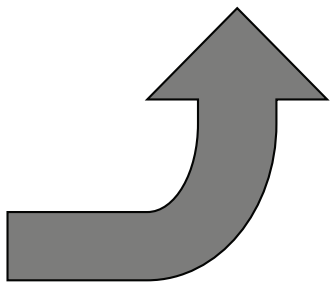
Peptide bonds

N-terminus

C-terminus

H_3N^+--AA_1--AA_2--AA_3--AA_4--…..--$AA_N$$COO^-$



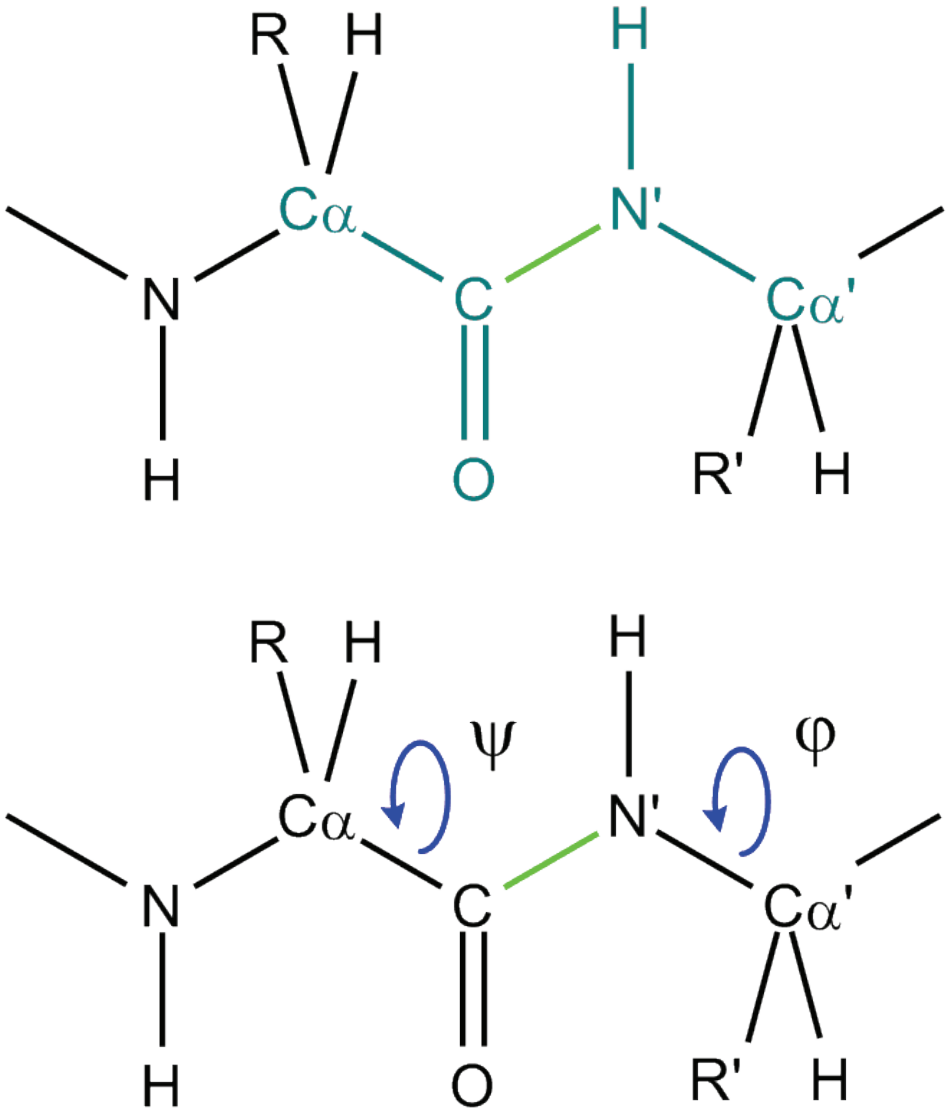


Amino acid *residue*

Dihedral angles

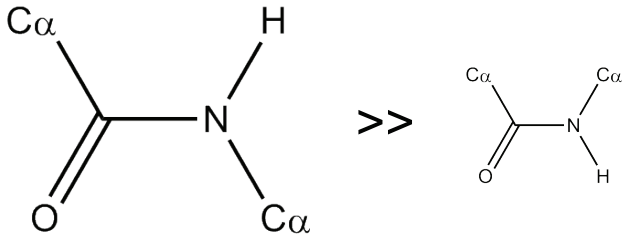
Jon K. Lærdahl,
Structural Bioinformatics

Proteins are polypeptides, *i.e.* many amino acids connected by peptide bonds



The peptide bond (light green) is a partial double bond and is fixed at ~180º, *i.e.* the green part is flat

Cis-form for peptide bond is extremely rare except for prolines (~25%).

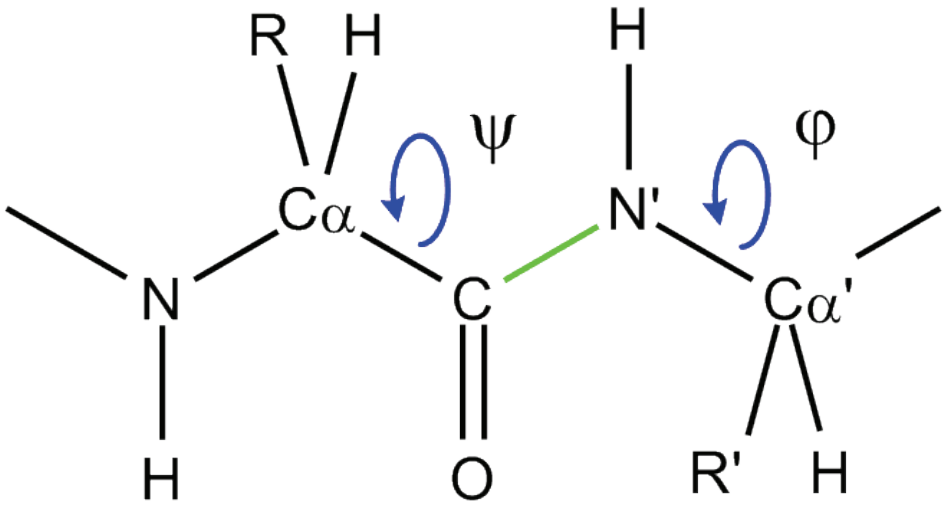


The dihedral angles phi (φ) and psi (ψ) determines the conformation of the peptide backbone

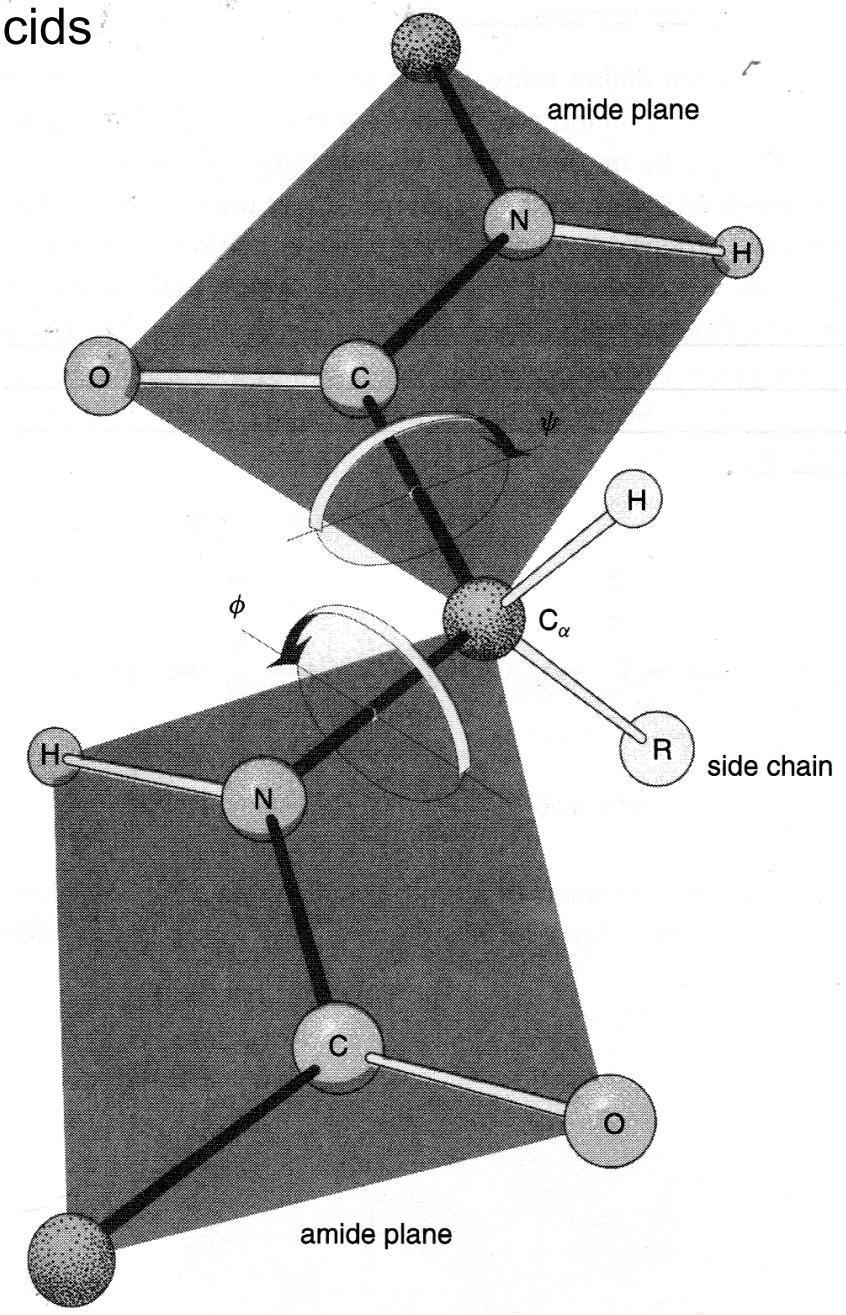
Dihedral angles

Jon K. Lærdahl,
Structural Bioinformatics

Proteins are polypeptides, *i.e.* many amino acids connected by peptide bonds



One (φ,ψ) pair for each residue in a protein



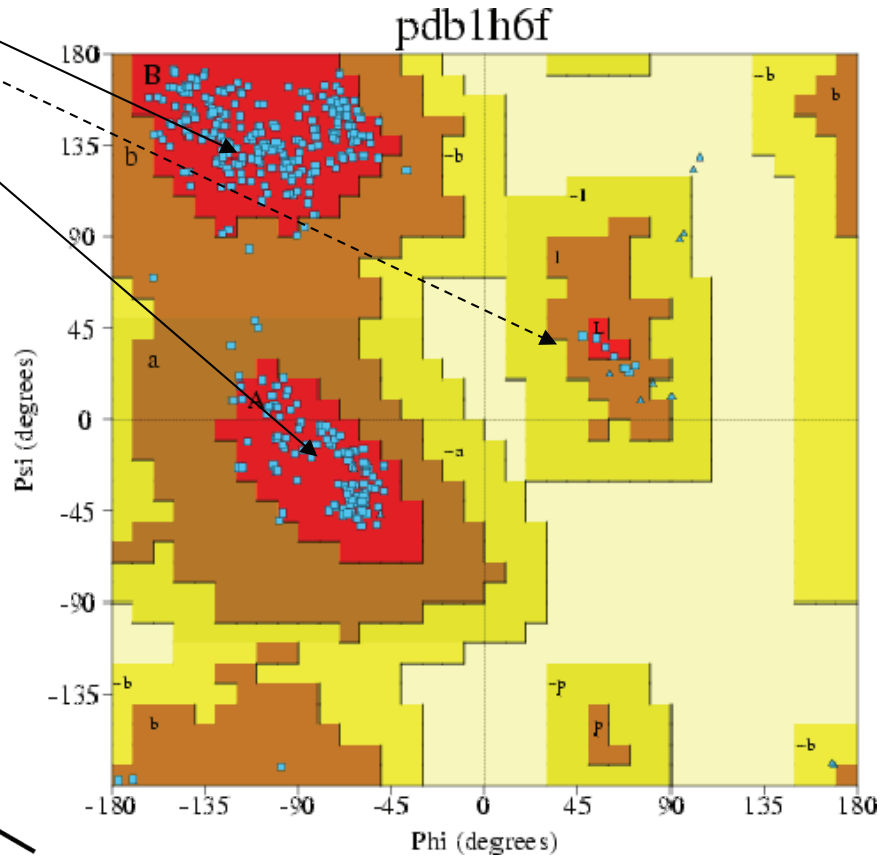
Structural Bioinformatics,
Eds. P.E. Bourne & H. Weissig
(Wiley, Hoboken, NJ, 2003)

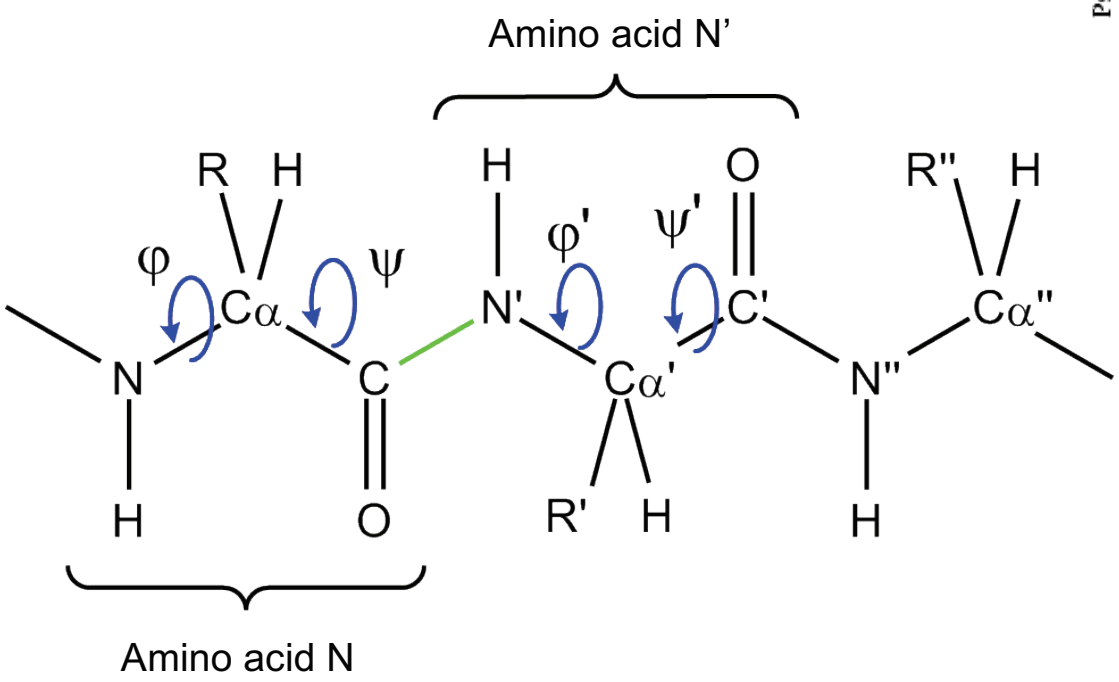
Jon K. Lærdahl,
Structural Bioinformatics

Ramachandran plot

- Dihedral angles
  - Phi (φ)
  - Psi (ψ)
- Plot of (φ,ψ) angle pairs for each residue in a protein: *Ramachandran plot*

Most (φ,ψ) pairs in two (three) regions

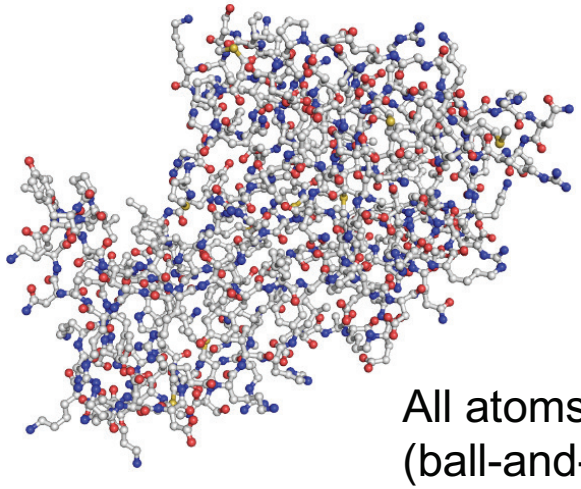




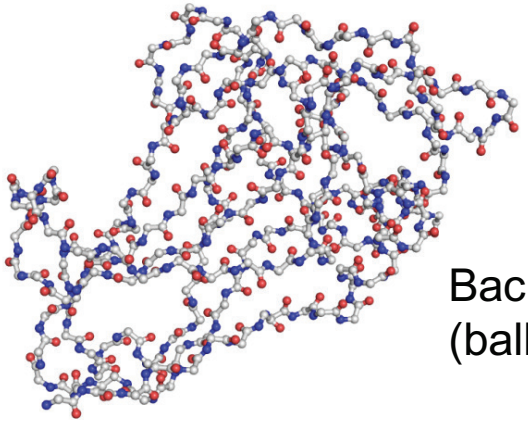
One point (blue spot) for each of the 184 residues in this protein (1H6F) (a human a transcription factor)

Jon K. Lærdahl,
Structural Bioinformatics

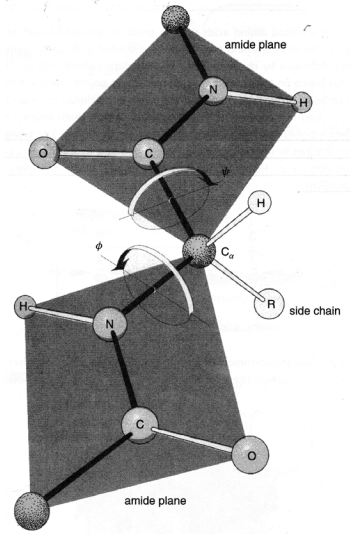
Ramachandran plot



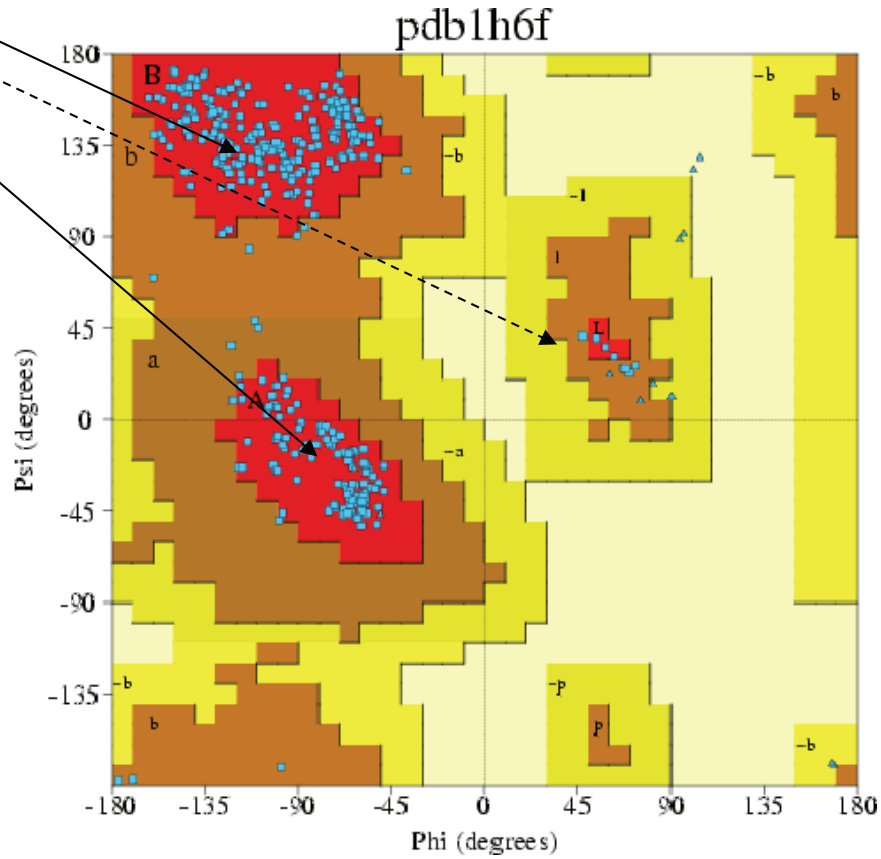
All atoms
(ball-and-stick)



Backbone atoms only
(ball-and-stick)



Most (φ,ψ) pairs in
two (three) regions

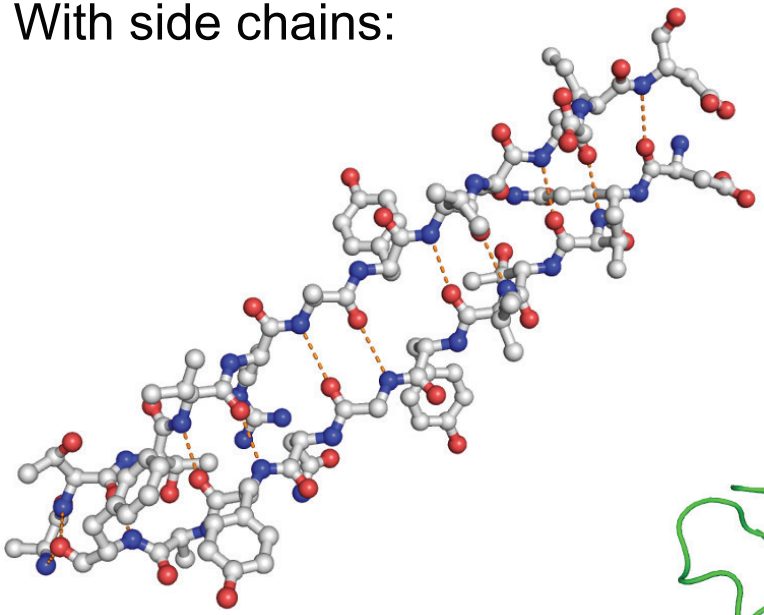


One point (blue spot) for each of the 184 residues in this protein (1H6F) (a human a transcription factor)

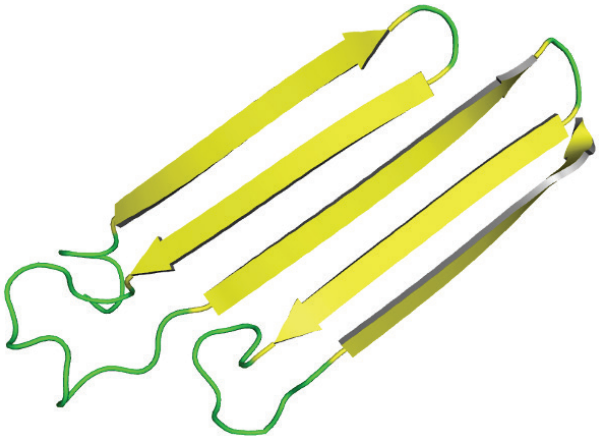
Jon K. Lærdahl,
Structural Bioinformatics

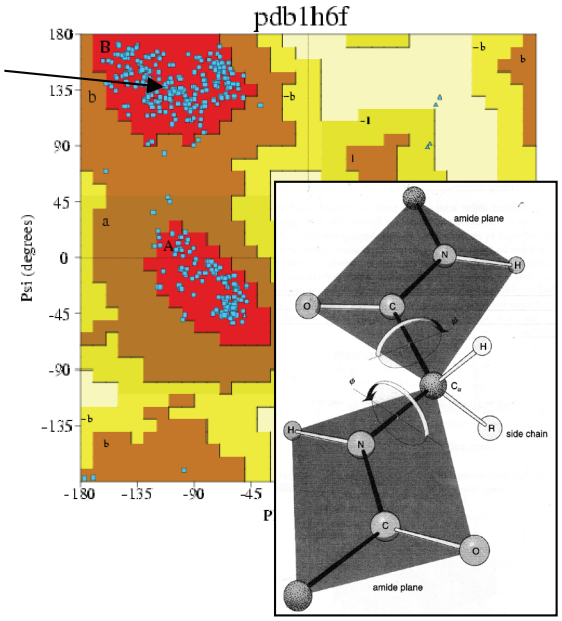
Secondary structure – β-sheets

With side chains:



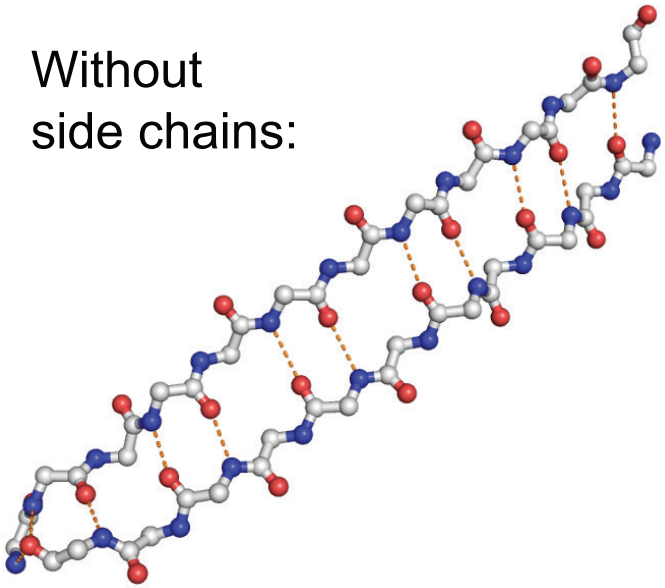
β-strands & β-sheets

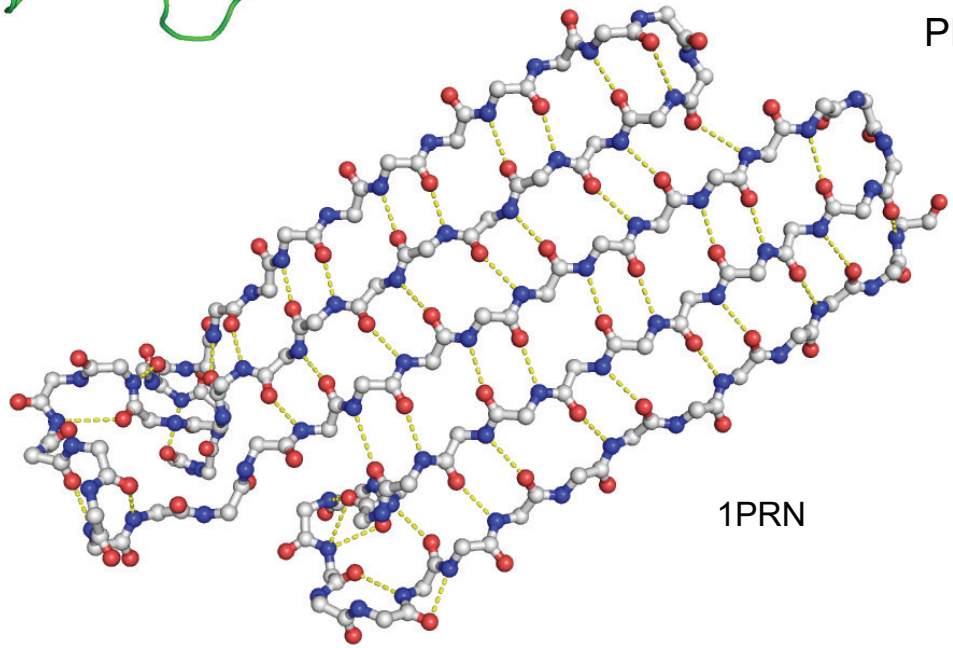




Psi ~ 135º
Phi ~ -100º

Without
side chains:



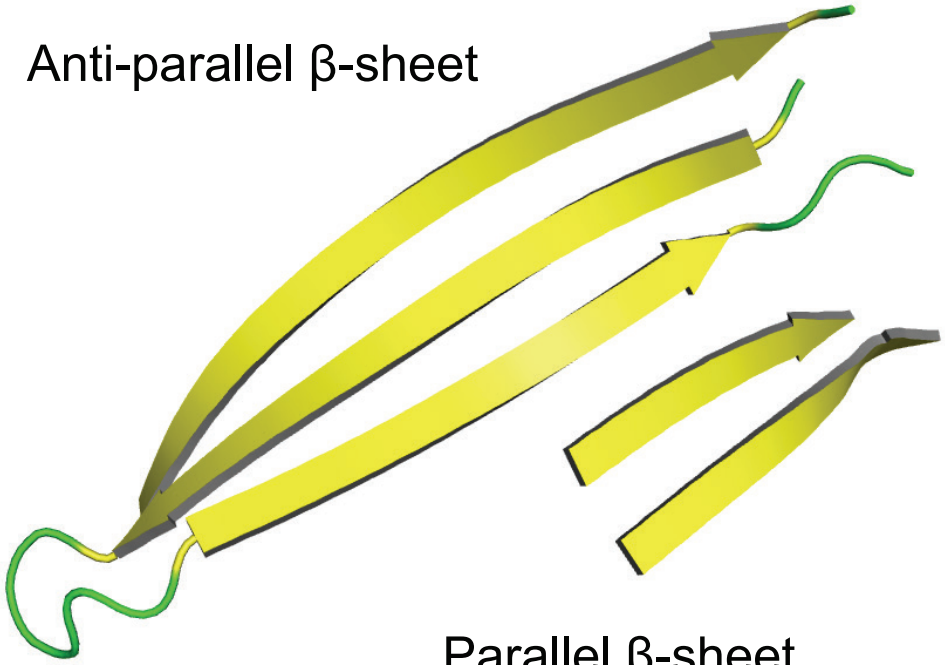


1PRN

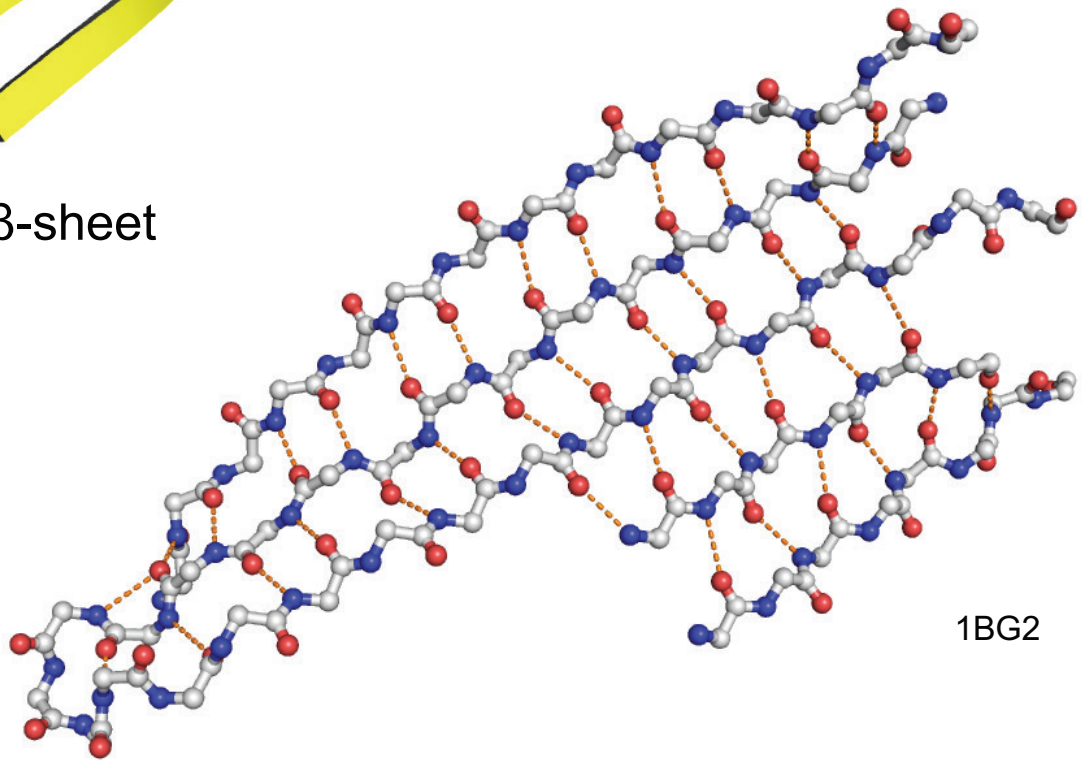
Jon K. Lærdahl,
Structural Bioinformatics

Secondary structure – β-sheets

Anti-parallel β-sheet



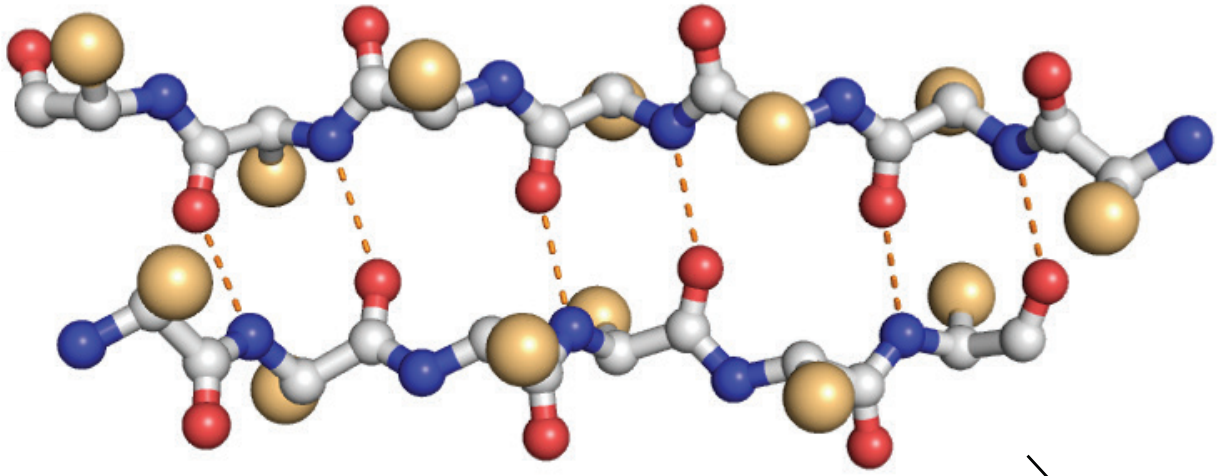
Parallel β-sheet



1BG2

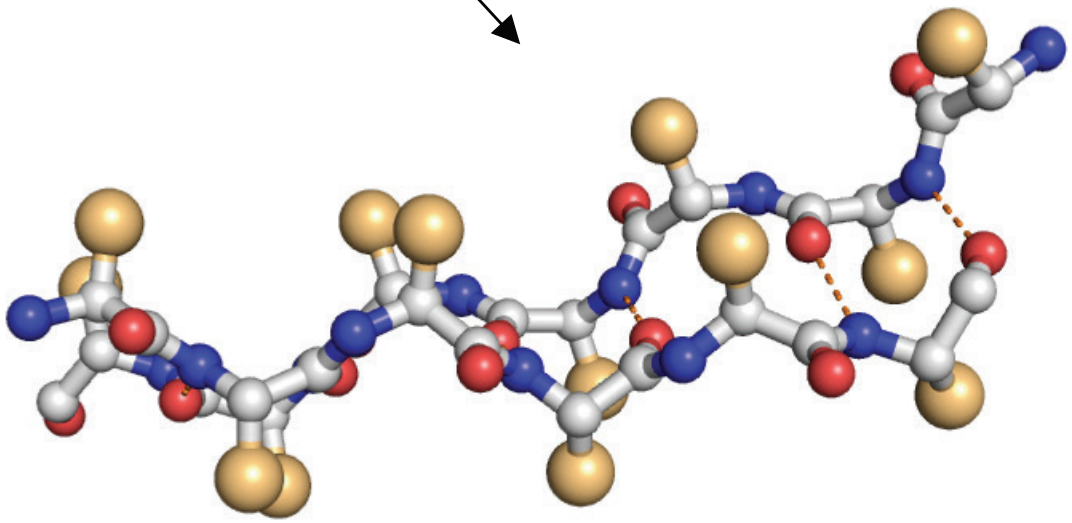
β-sheets can be parallel, anti-parallel or mixed

Secondary structure – β-sheets



90º rotation

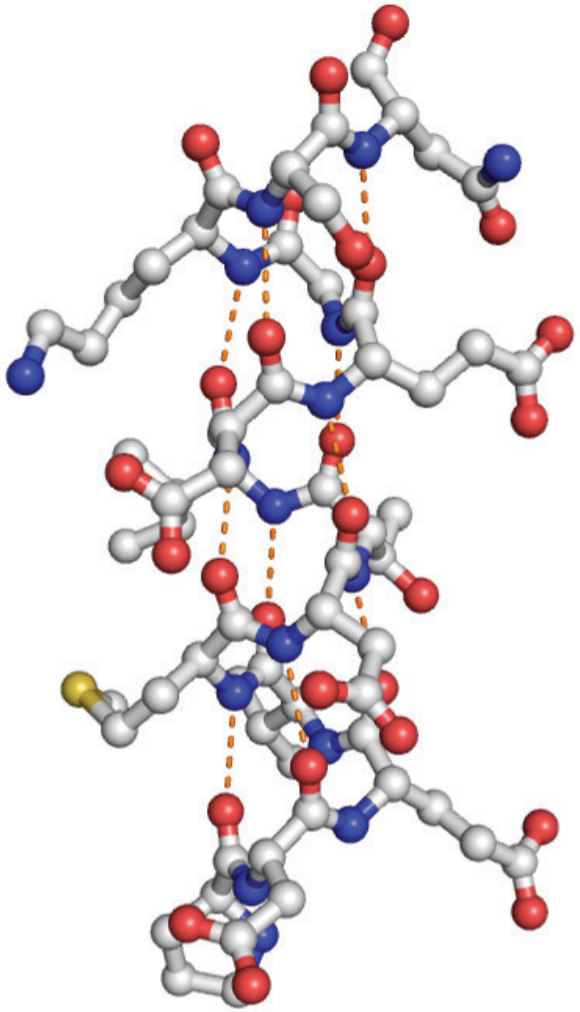
In β-sheets each side chain R-group is alternately on opposite sides of the plane of the sheet



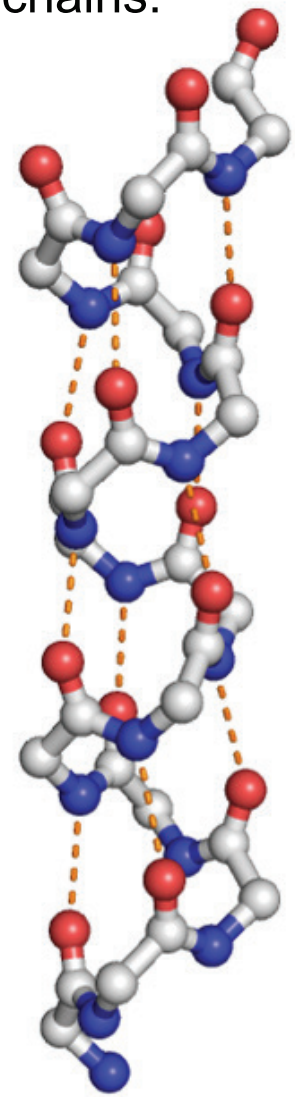
Jon K. Lærdahl,
Structural Bioinformatics

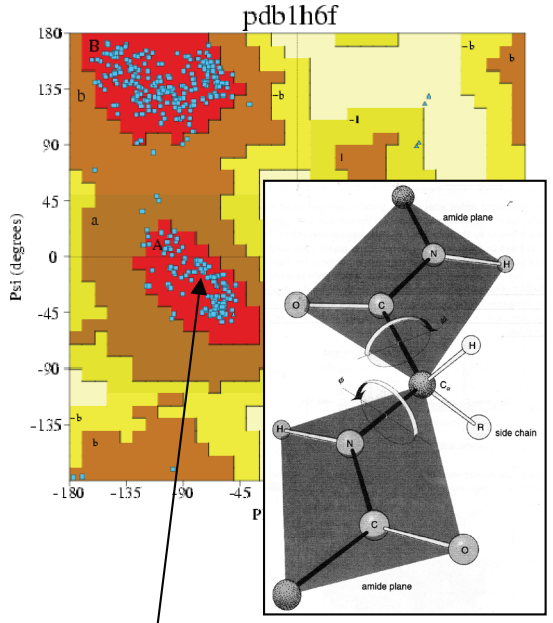
Secondary structure – α-helices

With side chains:



Without side chains:





α-helix

3.6 amino acids/turn

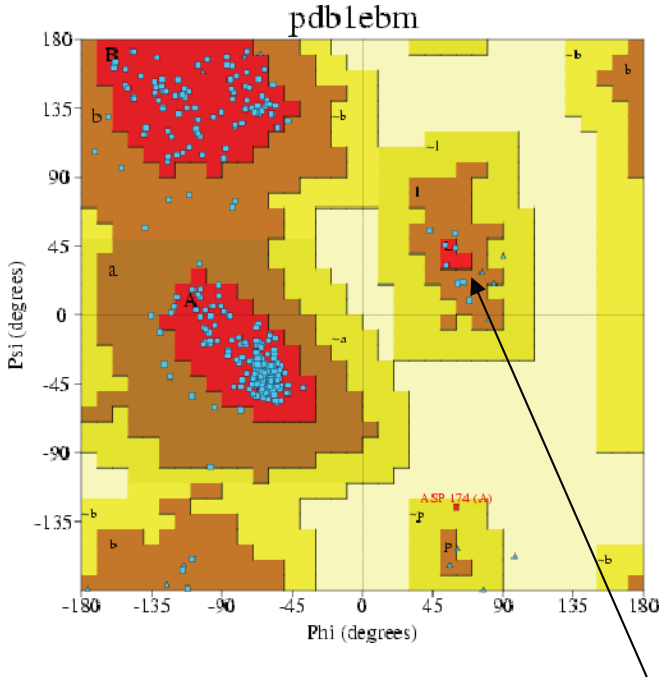
H-bonds between amino acids n & n+4

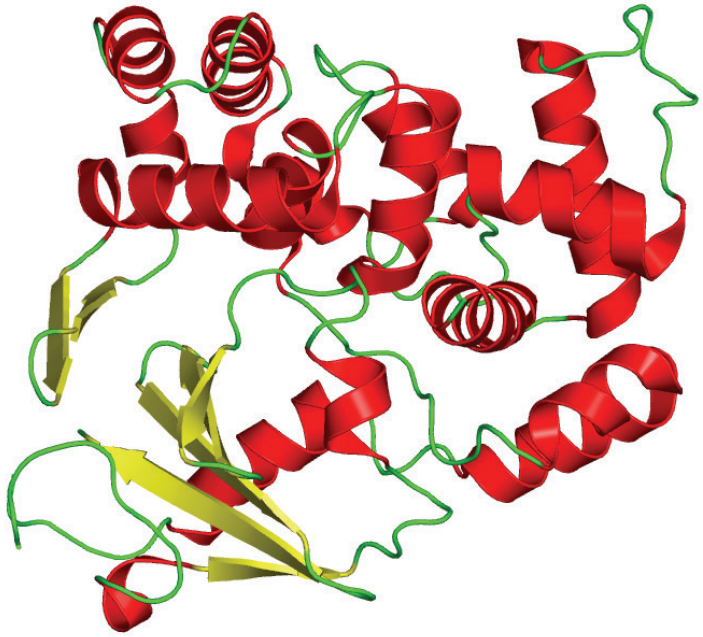
Partial positive charge at N-terminus and negative charge at C-terminus, *i.e.* it is a *dipole*

Jon K. Lærdahl,
Structural Bioinformatics

Secondary structure – 3 states

Three ”states”:
α-helices (H)
β-sheets (E)
Loops/coils (C)





Loops/coils:
• Loops may be hairpins or sharp turns
• Random coils/irregular loops
• Often “allowed” with insertions/deletions, *i.e.* evolutionary variable regions

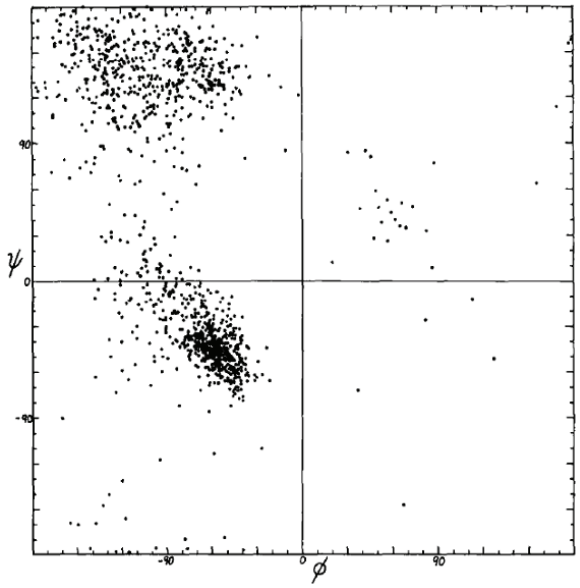
Coil here: “Everything that is not helix or sheet”

Coil often means: “Everything that is not helix or sheet or some characteristic loops”

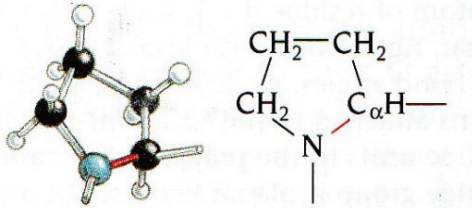
Often contains Gly (to give flexibility) or Pro (to “break up” secondary structure elements)

Left-handed helices

Secondary structure – Gly & Pro

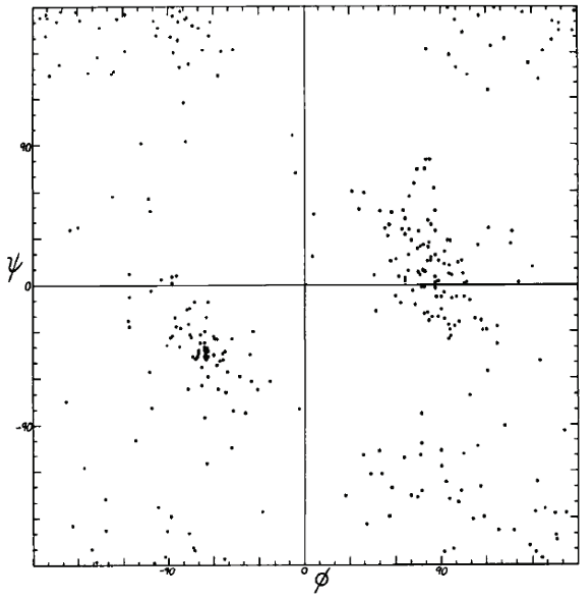


Non-glycine residues are mainly in α-helices and β-sheets

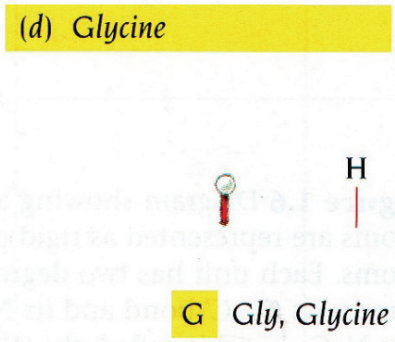


P Pro, Proline

Proline has very little flexibility in the backbone (disruptive to normal secondary structure)



Glycine has no side chain and a more flexible backbone



J. Richardson, *Adv. Prot. Chem.* **34**, 167 (1981)

Protein structure

Jon K. Lærdahl,
Structural Bioinformatics

• Primary structure: Linear amino acid sequence

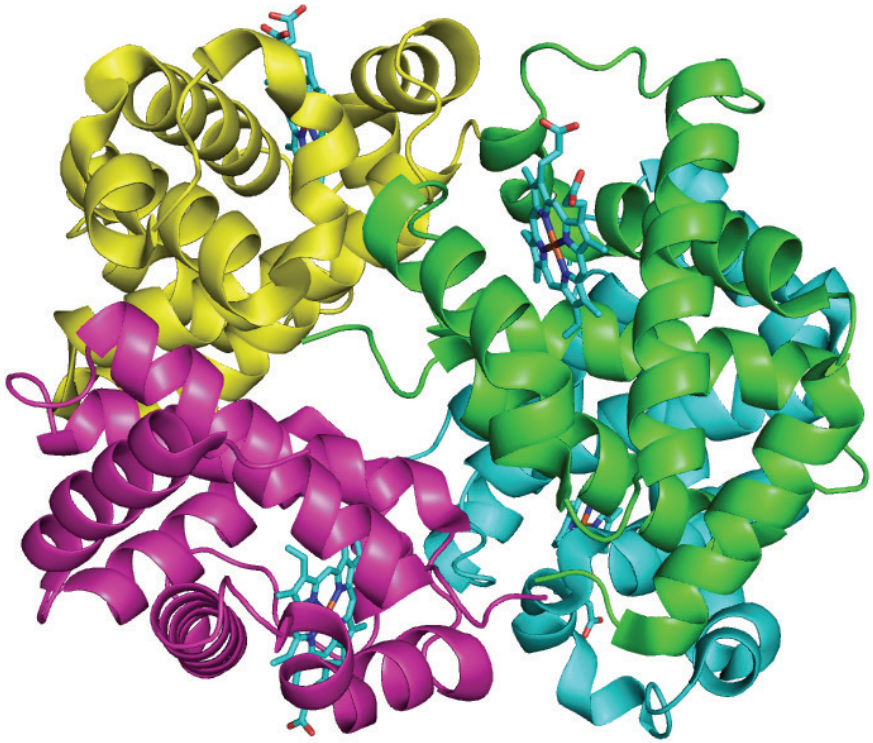
Met-Ala-Leu-Asp-Asp-…

• Secondary structure: Local conformation of the peptide chain:
  • α-helix
  • β-sheet

• Tertiary structure: The full 3D structure

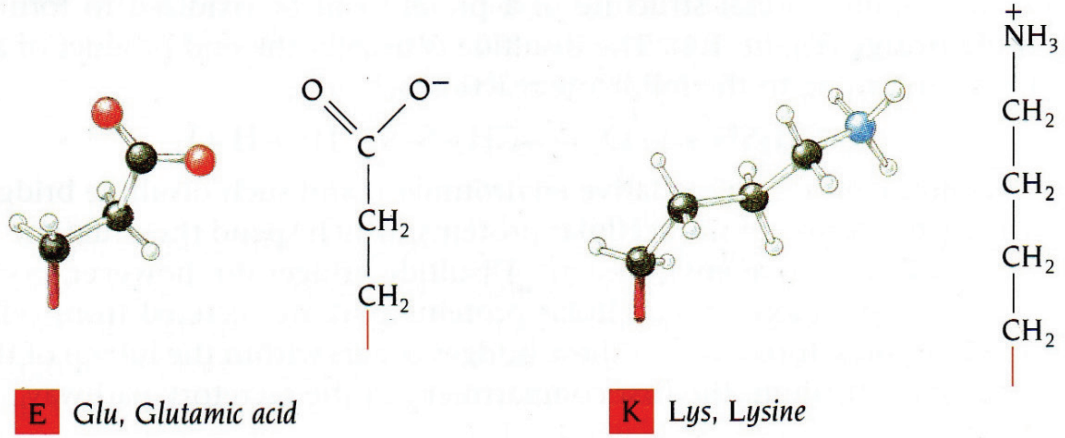
• Quaternary structure: Association of several proteins/peptide chains into protein complexes

Hemoglobin, 1GZX



Jon K. Lærdahl,
Structural Bioinformatics

Residue properties



E Glu, Glutamic acid

K Lys, Lysine

pK_a depends on local environment

e.g. Glu close to negatively charged moiety: higher pK_a
Glu close to Lys is more willing to give off H^+, *i.e.* lower pK_a

$pK_a = 4.25$ $pK_a = 10.53$

Free amino acid

N-terminal amino group, $pK_a \sim 7.4$
C-terminal acidic group, $pK_a \sim 3.9$

In a protein

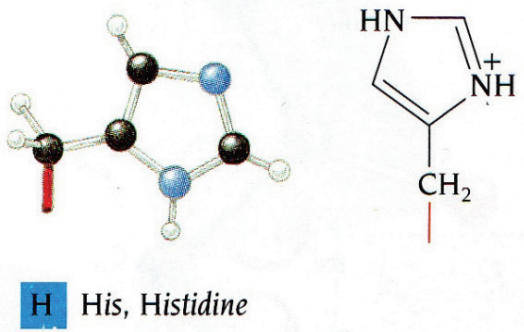
Table 1.2 *Intrinsic pK_a Values of Ionizable Groups Found in Proteins*

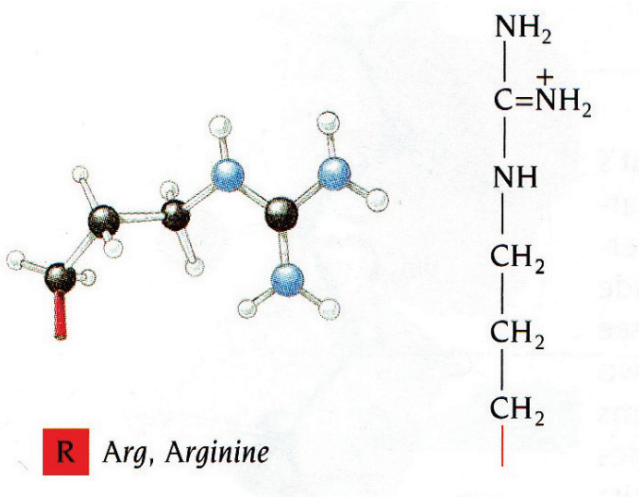
Group	Observed pK_a[a]
α-Amino	6.8–8.0
α-Carboxyl	3.5–4.3
β-Carboxyl (Asp)	3.9–4.0
γ-Carboxyl (Glu)	4.3–4.5
δ-Guanido (Arg)	12.0
ϵ-Amino (Lys)	10.4–11.1
Imidazole (His)	6.0–7.0
Thiol (Cys)	9.0–9.5
Phenolic hydroxyl (Tyr)	10.0–10.3

Proteins, T.E. Creighton (Freeman, New York, 1997)

Jon K. Lærdahl,
Structural Bioinformatics

Residue properties





Arg is "always" positively charged with pKa close to 12

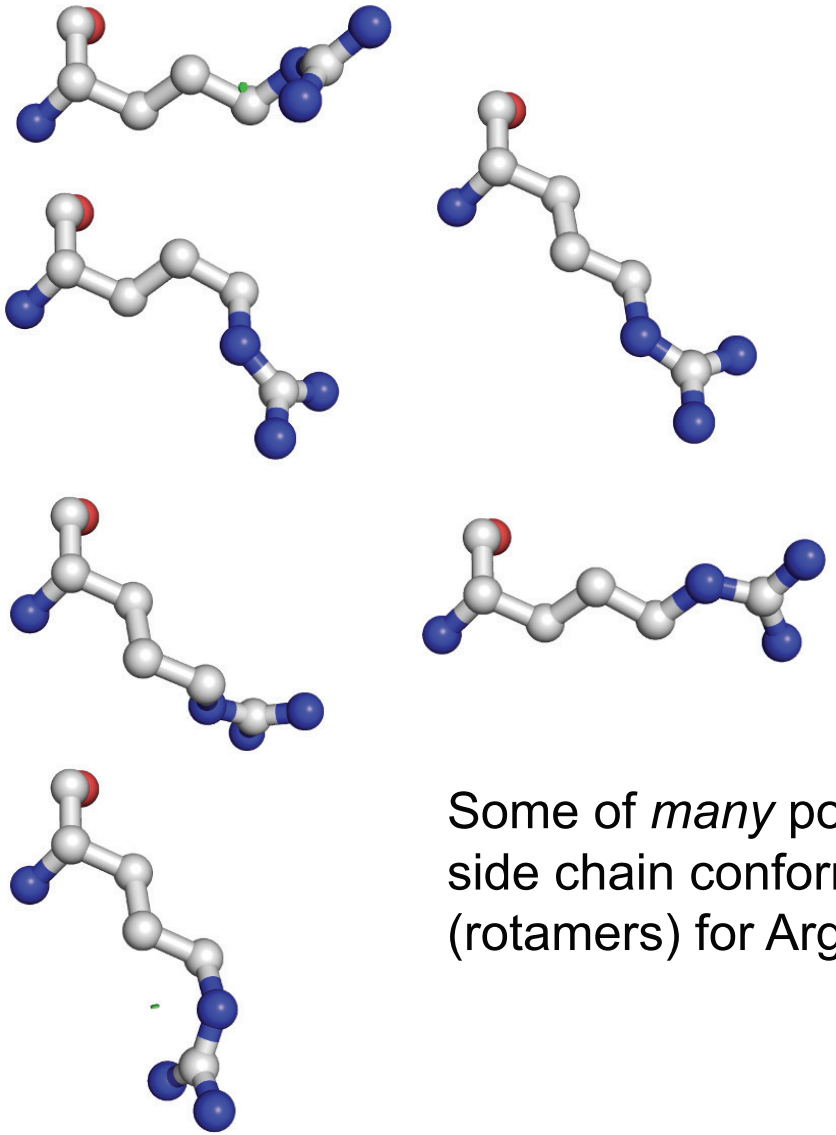
His has pK_a close to 7 and the local environment is often tuned to to give correct acid/base chemistry. Strong base at neutral pH/Strong nucleophile. Often a catalytic residue.

Table 1.2 *Intrinsic pK_a Values of Ionizable Groups Found in Proteins*

Group	Observed pK_a[a]
α-Amino	6.8–8.0
α-Carboxyl	3.5–4.3
β-Carboxyl (Asp)	3.9–4.0
γ-Carboxyl (Glu)	4.3–4.5
δ-Guanido (Arg)	12.0
ϵ-Amino (Lys)	10.4–11.1
Imidazole (His)	6.0–7.0
Thiol (Cys)	9.0–9.5
Phenolic hydroxyl (Tyr)	10.0–10.3

Proteins, T.E. Creighton (Freeman, New York, 1997)

Side chain conformations (Rotamers)



Some of *many* possibly side chain conformations (rotamers) for Arg

Analysis of many structures have shown that residues prefer one or a few conformations. These are called *rotamers* and are collected and distributed in *rotamer libraries*

These libraries are used in computational modeling of protein 3D structure.

Very simply put:

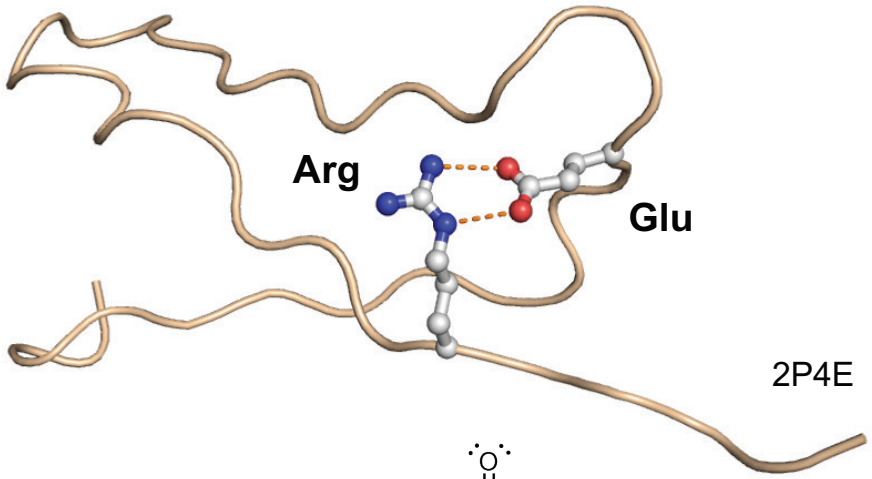
1. Determine overall 3D structure of backbone
2. Add side chains
3. Optimize side chains using conformations from rotamer libraries

Stabilizing forces

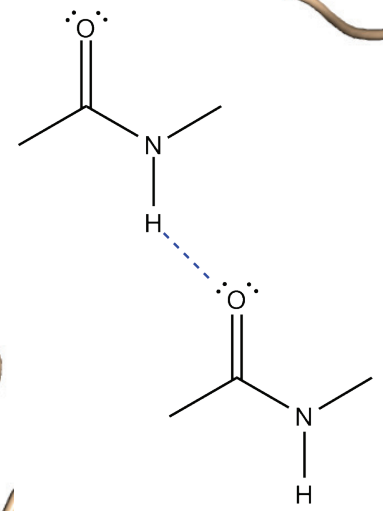
Jon K. Lærdahl,
Structural Bioinformatics

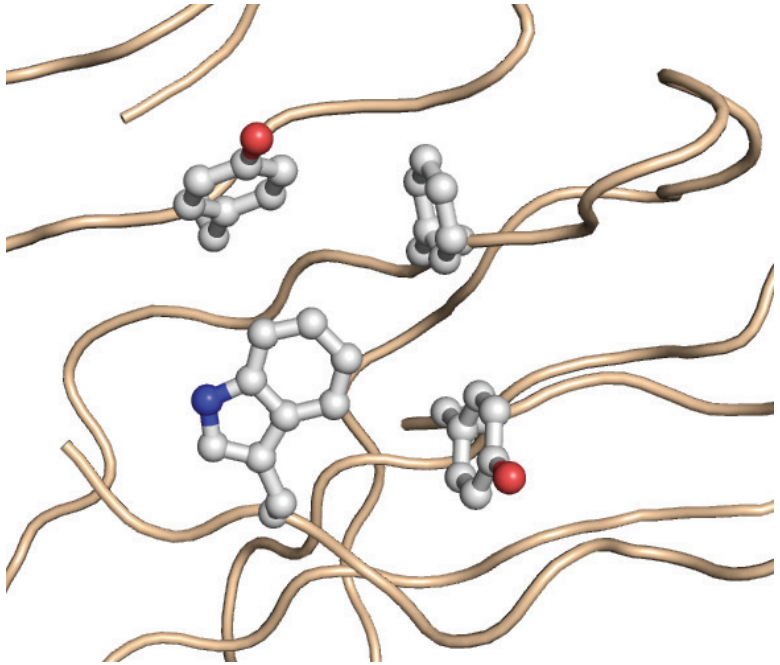
What is making proteins fold and associate into a well-defined 3D structure?

- Electrostatic interactions (salt bridges)
- Hydrogen bonds (H-bonds)
- van der Waals forces (weak)
- IMPORTANT: Hydrophobic interaction forces (minimizing the surface area of hydrophobic side chains exposed to solvent)



2P4E



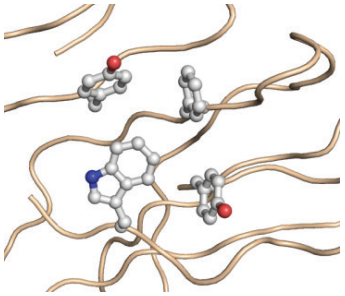


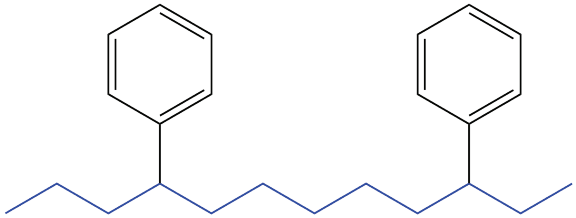
2P4E

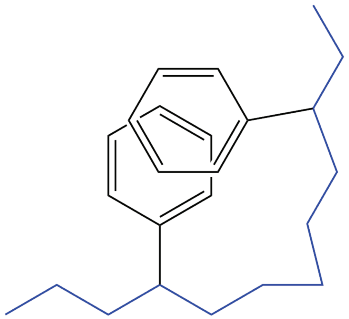
Stabilizing forces

Jon K. Lærdahl,
Structural Bioinformatics

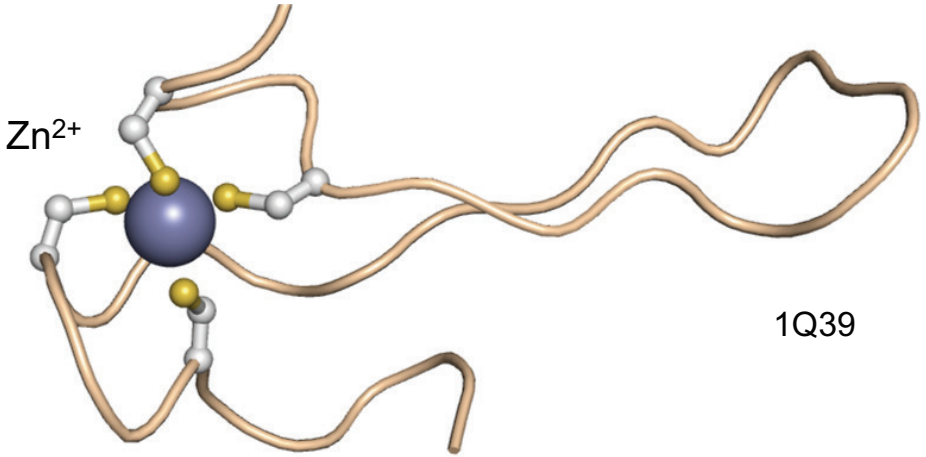
IMPORTANT: Hydrophobic interaction forces (minimizing the surface area of hydrophobic side chains exposed to solvent)

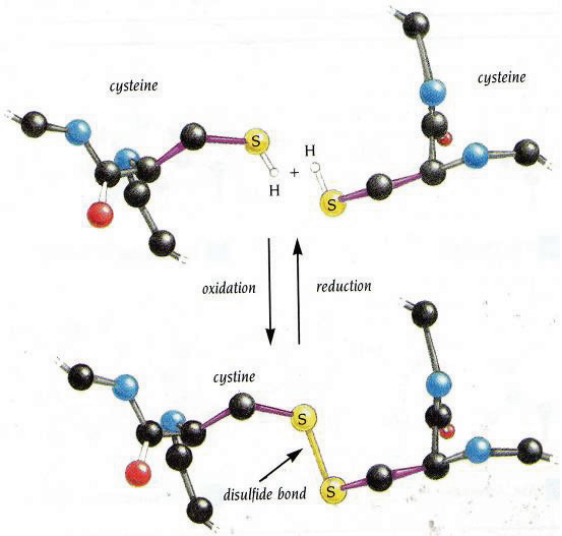






Reduced surface area exposed to solvent (water) for the hydrophobic side chains





Covalent Cys-Cys disulfide bonds

Introduction to Protein Structure,
C. Branden & J. Tooze
(Garland, New York, 1998)

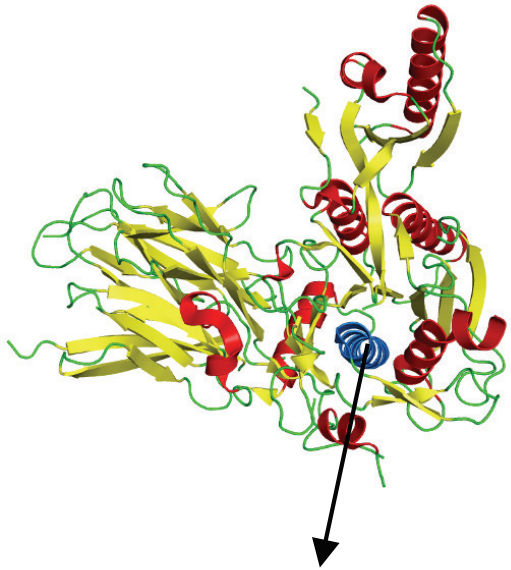
Metal ions may stabilize the protein structure (e.g. in zinc fingers)

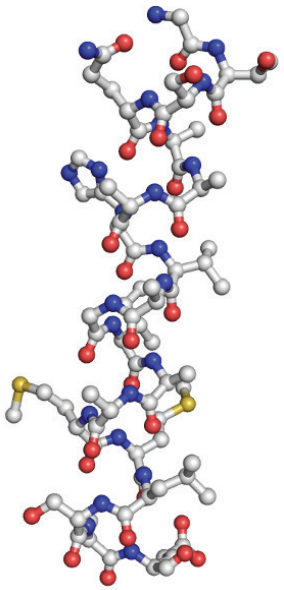
Protein folding

Jon K. Lærdahl,
Structural Bioinformatics

What is making proteins fold and associate into a well-defined 3D structure?

- Proteins are often found in water and both protein-protein and protein-water interactions must be taken into account (*i.e.* interactions in folded vs. denatured state)
- *Dominant* forces responsible for tertiary structure are (believed to be) the hydrophobic interaction forces
    - Residues with hydrophobic side chains are packed in the interior of the protein
    - Charged and polar residues tend to be on the protein surface
    - Polar backbone in the protein interior is “hidden” by building secondary structure elements
- Polar residue side chains in the core must be “neutralized” by interacting with other residues, e.g. in H-bond donor-acceptor pairs
- Charged residue side chains in the core must be “neutralized” by interacting with other residues through salt bridges



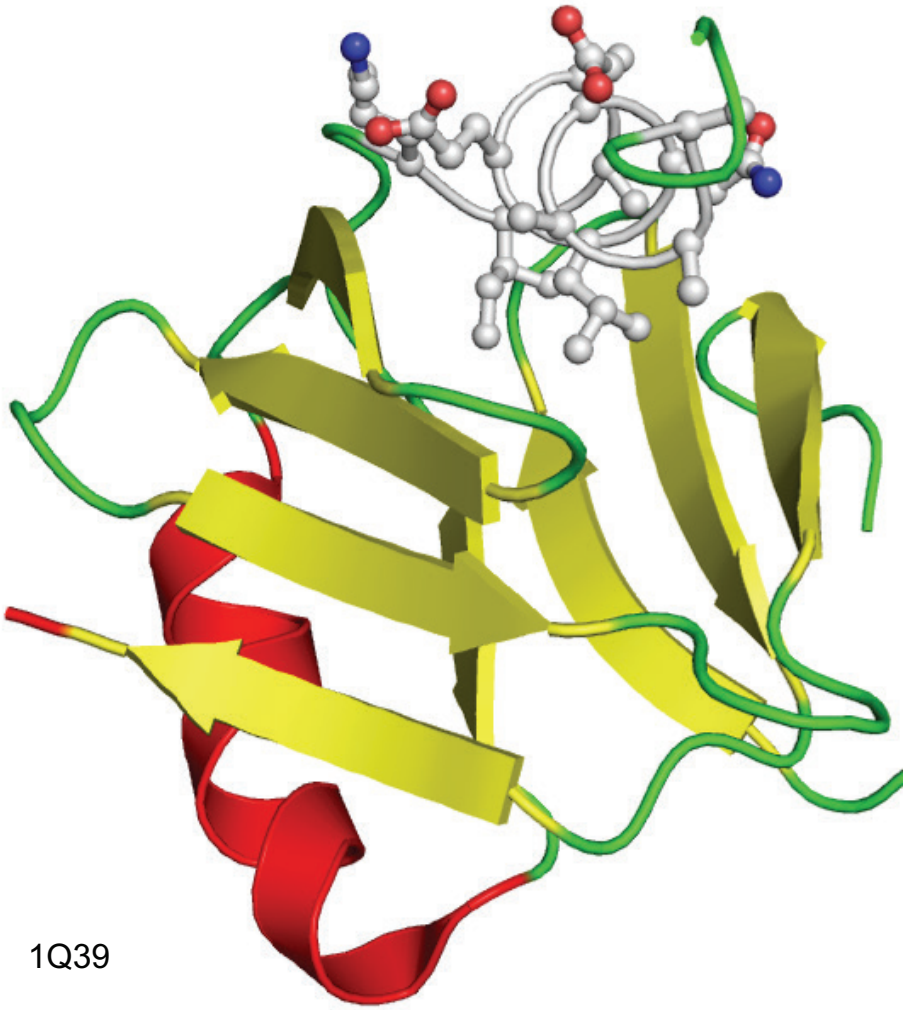


2P4E

Protein folding

Jon K. Lærdahl,
Structural Bioinformatics

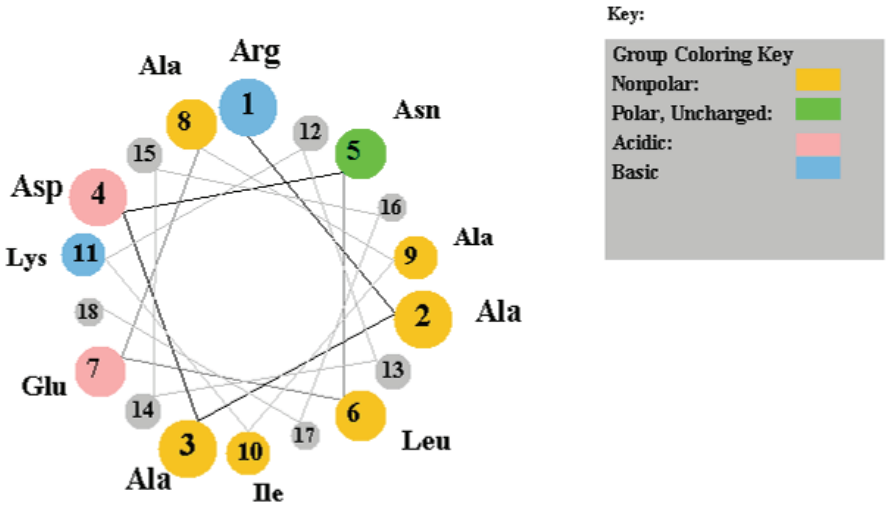
Secondary structure elements (α-helices & β-sheets) on the surfaces of proteins are often amphipathic (one hydrophilic and one hydrophobic side)



1Q39

"Pattern" of every 3-4 residues hydrophobic

Patterns can be used for predictions by computational methods, *e.g.* predict secondary structure from primary sequence



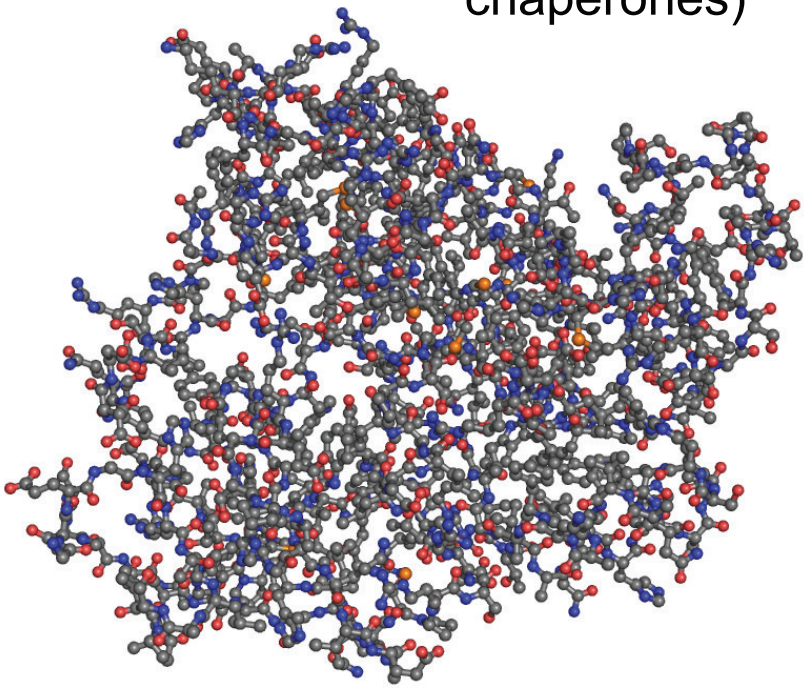
http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html

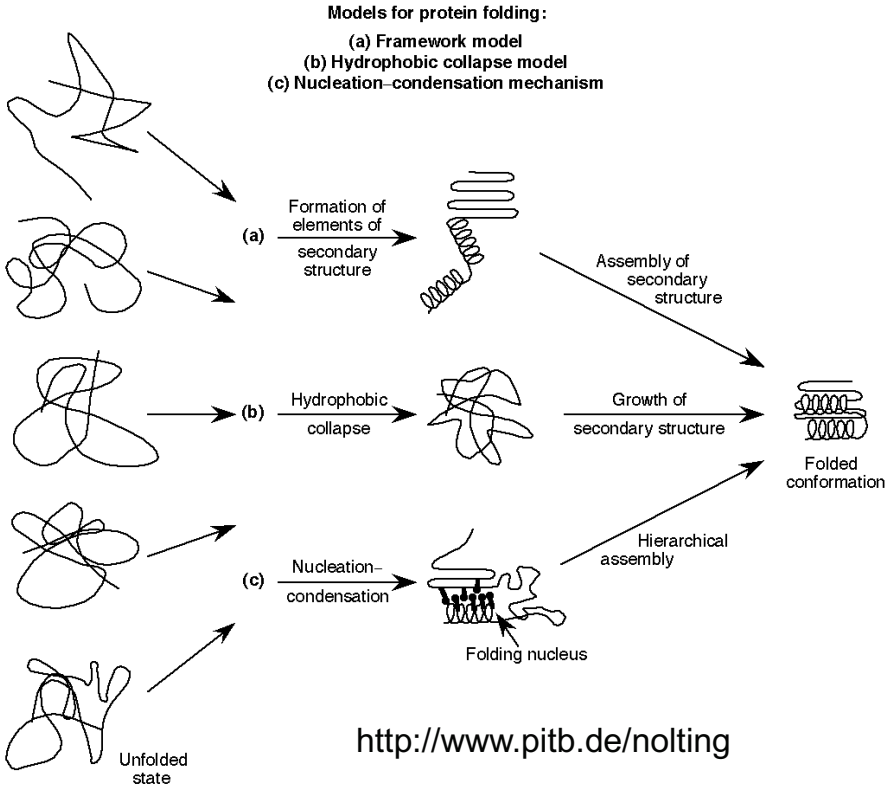
Protein folding

Jon K. Lærdahl,
Structural Bioinformatics

TLASTPALWASIPCPRSELRLDLV LPSGQS

Folding is spontaneous in the cell (but often with helper molecules, chaperones)





http://www.pitb.de/nolting

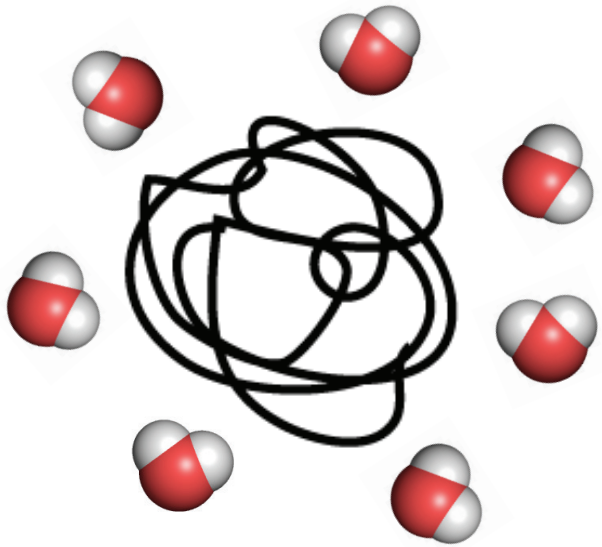
Put *very* simply:

1. Secondary structure forms transiently
2. Hydrophobic collapse, formation of stable secondary structure
3. Folding completes, formation of tertiary interactions

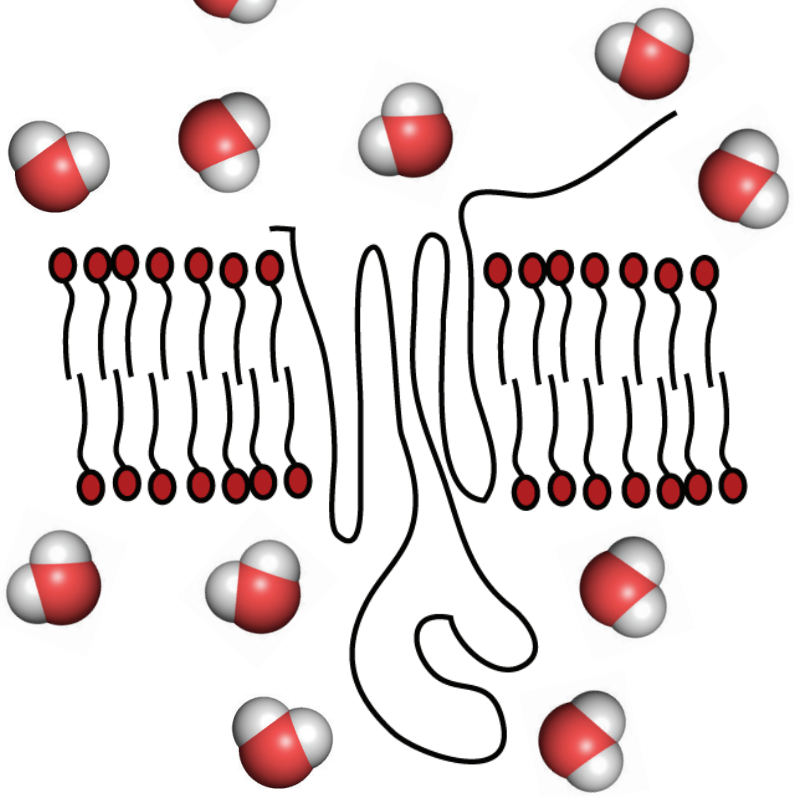
Globular vs. membrane proteins

Jon K. Lærdahl,
Structural Bioinformatics

Globular proteins
- Soluble
- Surrounded by water

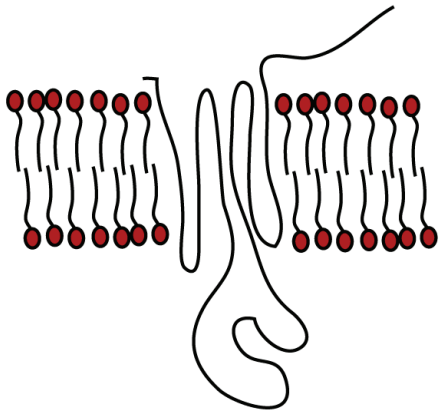


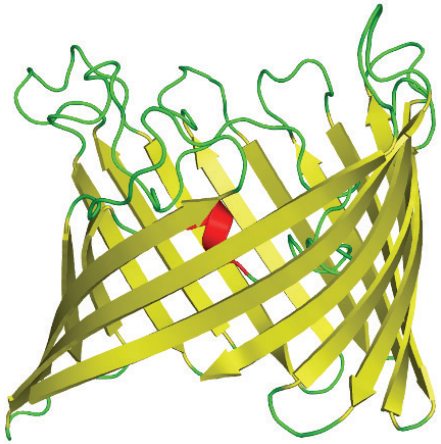
Membrane proteins
- In lipid bilayers
- Hydrophobic surface facing membrane interior



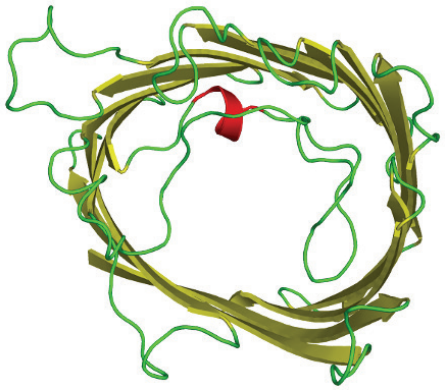
Membrane proteins

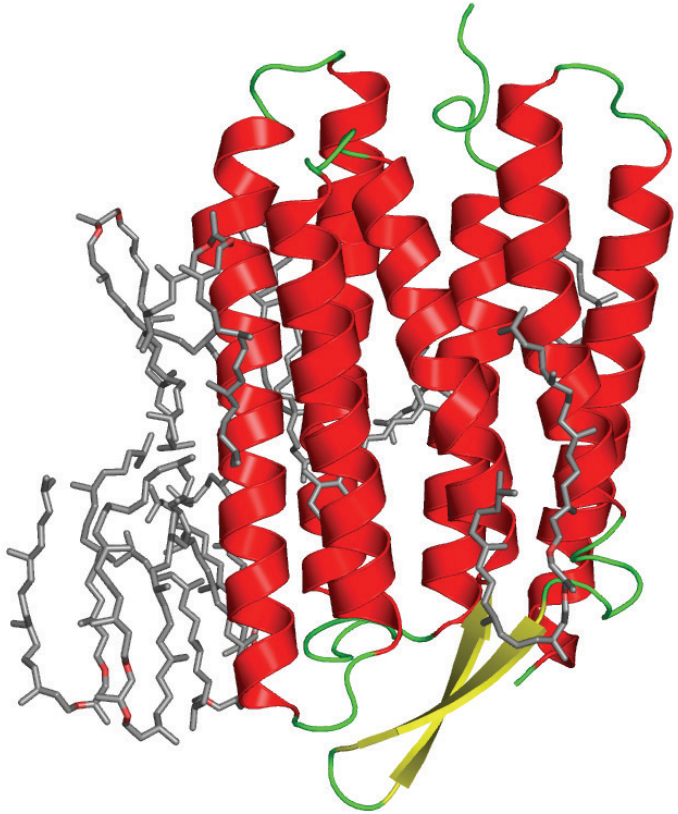
Jon K. Lærdahl,
Structural Bioinformatics





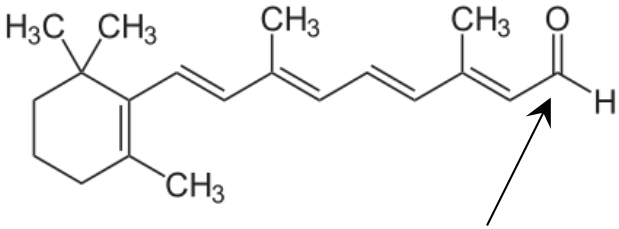
Beta-barrel porin (1PRN)





Rhodopsin (1QHJ)

Co-factor/prosthetic group retinal:



Covalent (Schiff bond) linkage to protein Lys residue

Many apo-proteins need co-factors/prosthetic groups to become functional