

# **LSM2241**

## **Bioinformatics of protein structure and structure databases**

Greg Tucker-Kellogg  
[dbsgtk@nus.edu.sg](mailto:dbsgtk@nus.edu.sg)

14 October 2015

# Outline

Protein chemistry review

Basics of protein structure

Higher level protein structure

The Protein Data Bank

Structure visualization

Roundup and next week

# Topic

Protein chemistry review

Basics of protein structure

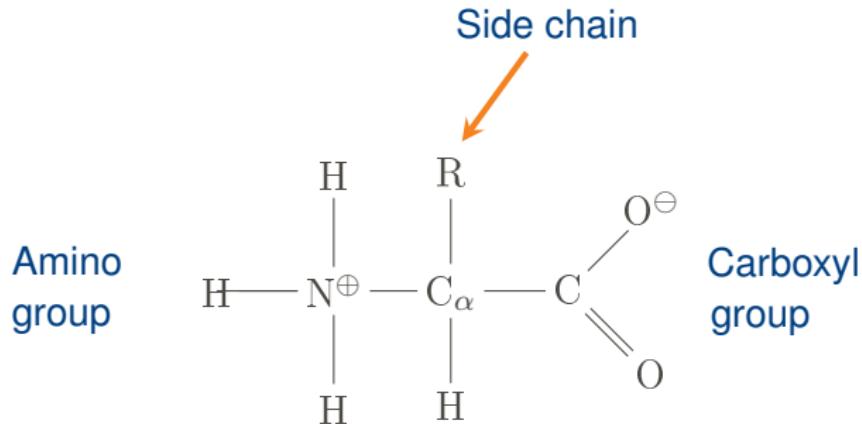
Higher level protein structure

The Protein Data Bank

Structure visualization

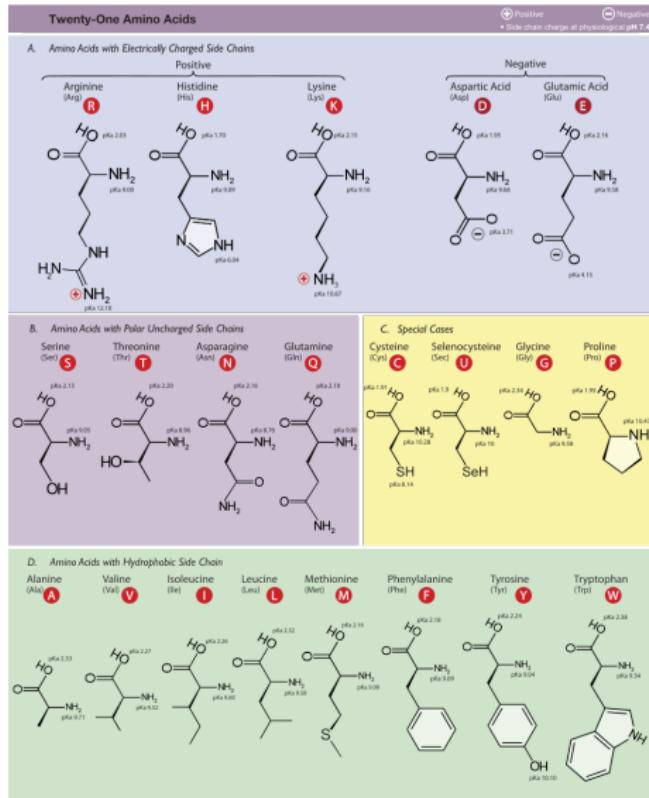
Roundup and next week

# The amino acid: the building block of proteins

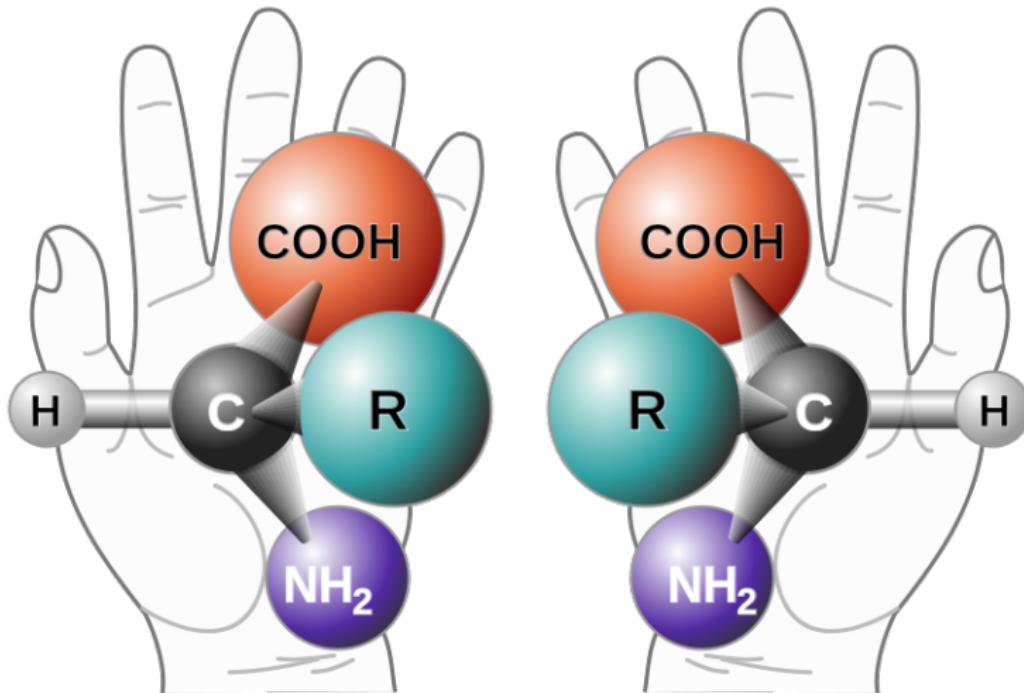


This is pictured as a *zwitterion* with no net charge but both positive and negative ionic states

# The 20+1 "proteinogenic" amino acids in eukaryotes



## Amino acid chirality

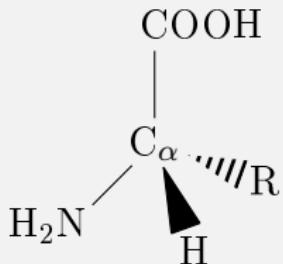


Amino acids come in two mirror image forms

# L and D forms of amino acids

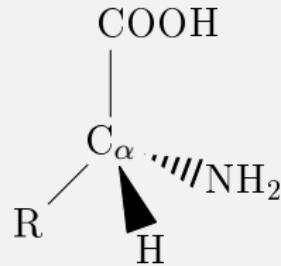
## L form

- All amino acids in translated proteins
- Spells "CORN" when looking down from the  $H_\alpha$  position



## D form

- Some natural products, parts of the cell wall



# Topic

Protein chemistry review

Basics of protein structure

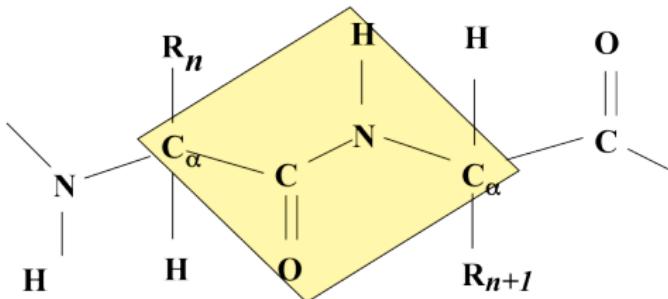
Higher level protein structure

The Protein Data Bank

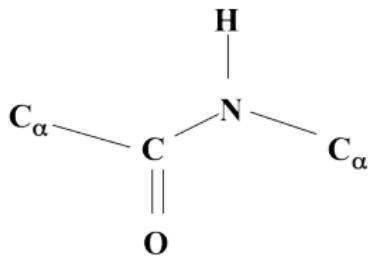
Structure visualization

Roundup and next week

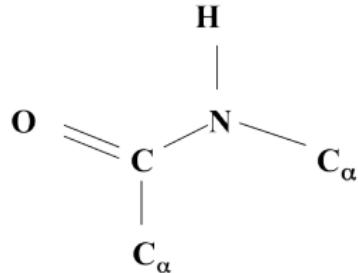
# The planar amide bond



The peptide bond is planar



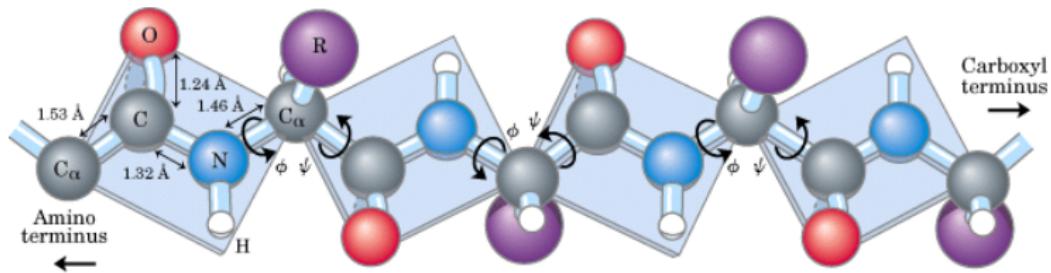
Conformation "Trans"



Conformation "Cis"

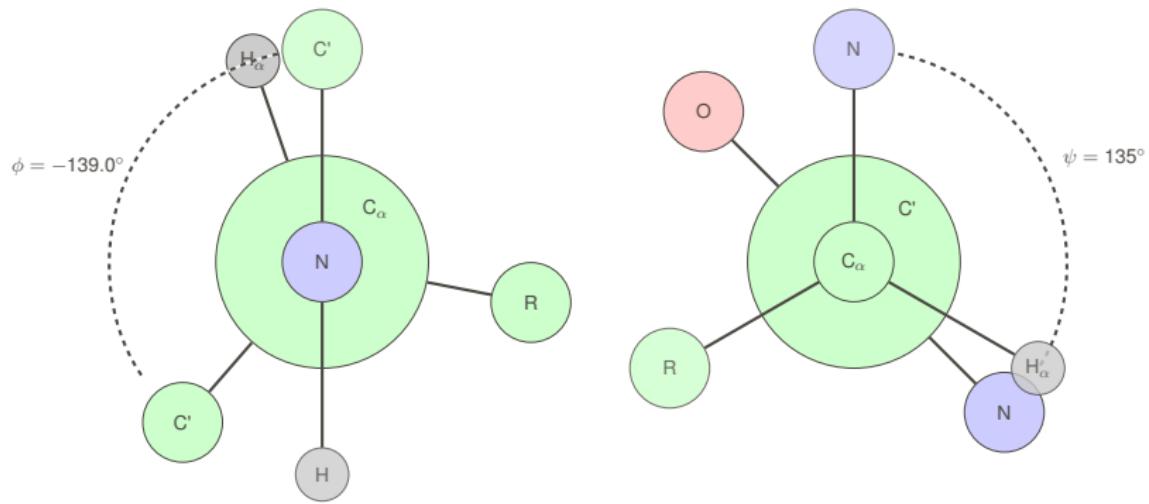
The planar peptide bond (click to activate in Adobe Reader)

# protein backbone structure is largely a matter of $\phi$ and $\psi$



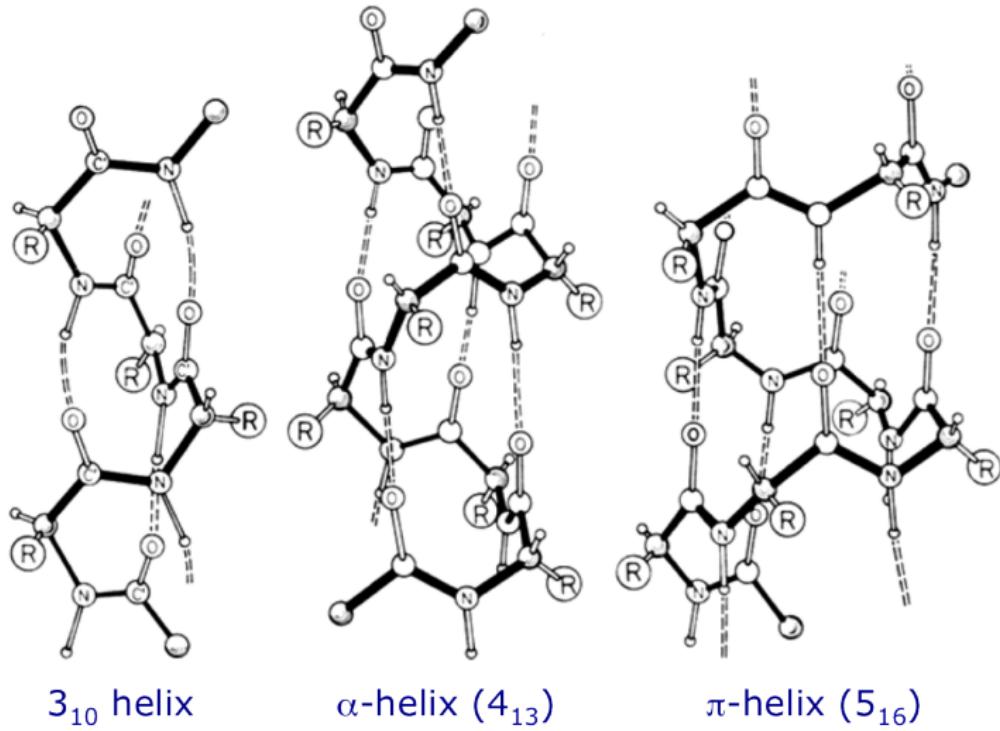
$\phi$ : C<sub>α</sub>-N torsion angle.  $\psi$ : C<sub>α</sub>-CO torsion angle. The only other torsion angle ( $\rho$ ) is largely fixed. See Richardson 1981

# Newman projections of $\phi$ and $\psi$ angles



$\phi$  and  $\psi$  angle Newman projections for an amino acid in a canonical  $\beta$  sheet

## Secondary structure: Helices



$3_{10}$  helix

$\alpha$ -helix ( $4_{13}$ )

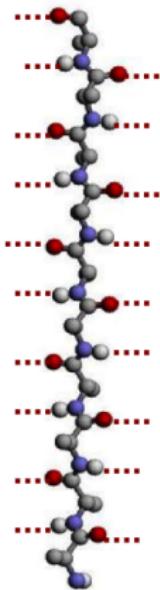
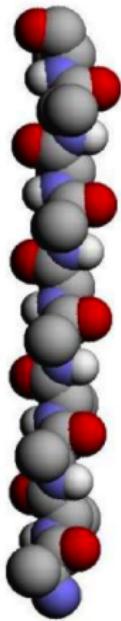
$\pi$ -helix ( $5_{16}$ )

click to activate

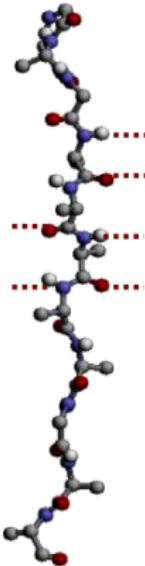
# What are these helices?

- |                                  |   |
|----------------------------------|---|
| <b>3<sub>10</sub> helix</b>      | 3.0 residues/turn (tight); with 10 atoms in the ring for each hydrogen bond. Roughly 4% of helices in structure databases |
| <b><math>\alpha</math> helix</b> | 3.6 residues per turn, the vast majority of helices   |
| <b><math>\pi</math> helix</b>    | 4.2 residues per turn (loose), not stable   |

## Secondary structure: Beta strands and sheets



NH-OC  
hydrogen bonds  
between  
strands



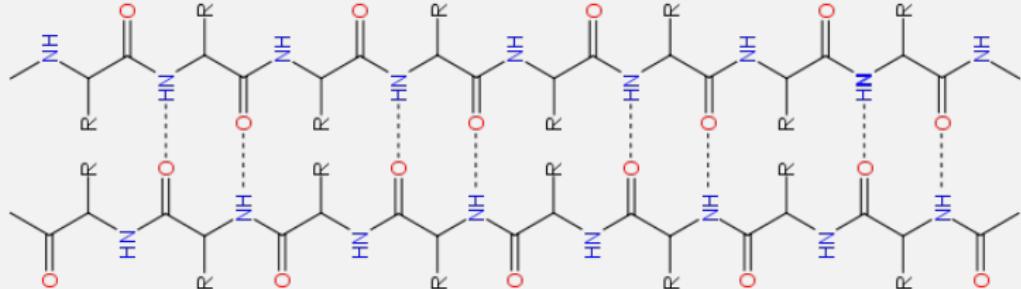
Extended chain is flat

Real strand is twisted

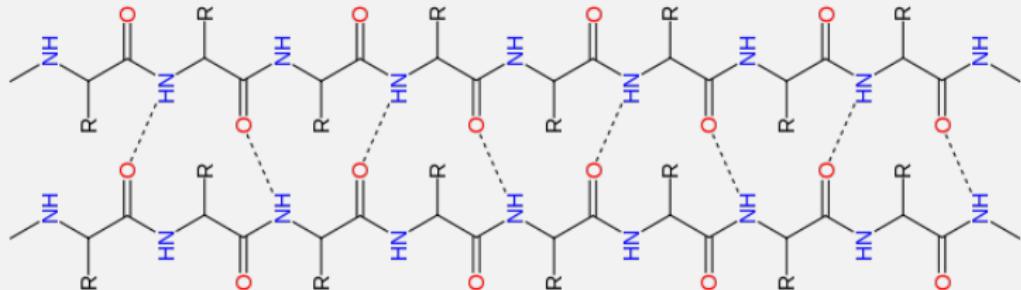
click to activate

# Two types of $\beta$ sheets

anti-parallel

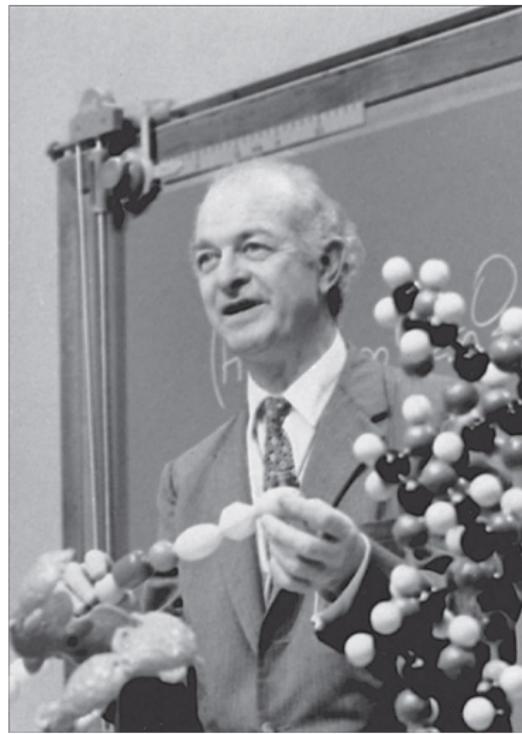


parallel



# Linus Pauling discovered protein secondary structure

- $\alpha$  helix,  $\beta$  sheet, planar peptide bond, electronegativity, VSEPR, fundamentals of X-ray crystallography, the hydrogen bond, mechanism of sickle cell disease, protein evolution, nature of enzymatic catalysis, many many others
- "The way to get good ideas is to get lots of ideas, and throw the bad ones away"



# **Turns connect secondary structure elements by reversing direction**

## Many contexts

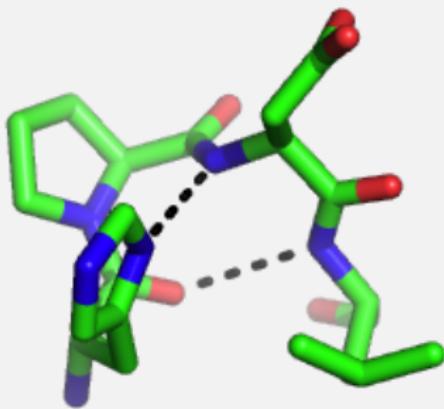
- helix-turn-helix
- antiparallel  $\beta$  sheets
- $\alpha$  helix –  $\beta$  strand
- $\beta$  strand –  $\alpha$  helix

## Characteristic types

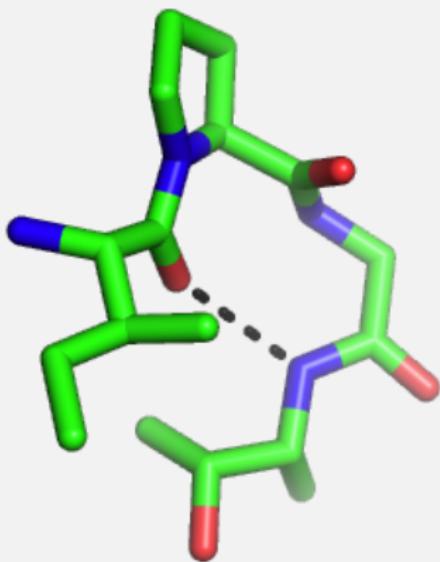
- $\alpha$  turn (4 peptide bonds between secondary structural elements)
- $\beta$  turn (3 peptide between secondary structural elements)
- Turns are characterized by backbone dihedral angles
- Most common is the  $\beta$  turn, with several types (two classical types, see Hutchinson and Thornton 1994).

## Types I and II $\beta$ turns

Type I turn



Type II turn



Note the proline in the second position

Also a proline at the second position, but a very different turn!

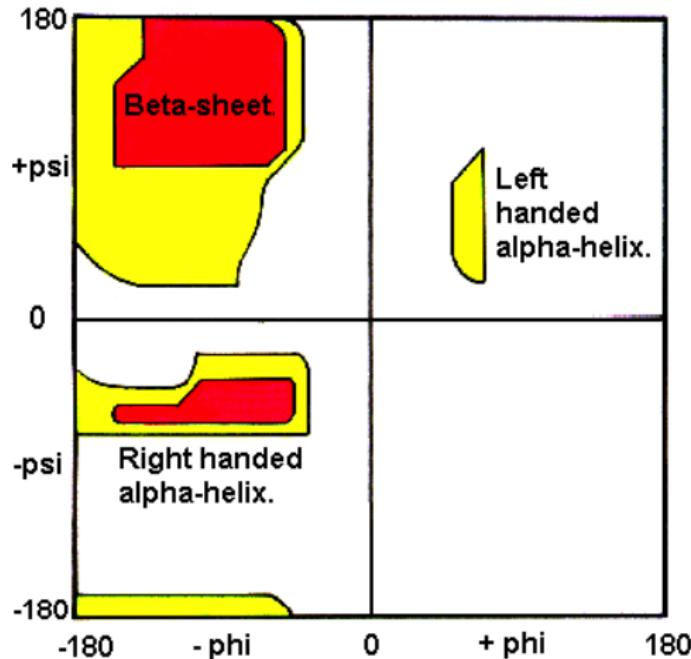
## Certain residues (amino acids) are favored or disfavored in $\beta$ turns

The two most common turn types, and their amino acid preferences, are shown below

$\beta$ turn type	Pos 1	Pos 2	Pos 3	Pos 4
I	Asp, Asn, Ser, Cys	Pro	Pro	Gly
II	Asp, Asn, Ser, Cys	Pro	Gly, Asn	Gly

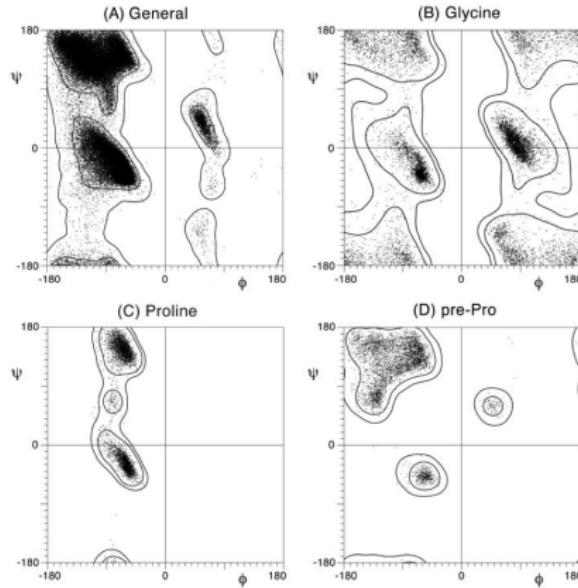
Favoured residues in black. Disfavoured residues in red.

## Secondary structures give characteristic Ramachandran plots



Regions of allowed  $\phi$  and  $\psi$  angles, with associated secondary structures

# Real proteins show the same constraints



Lovell et al. 2003, a classic paper from Jane Richardson lab

# Topic

Protein chemistry review

Basics of protein structure

Higher level protein structure

The Protein Data Bank

Structure visualization

Roundup and next week

# Protein tertiary structure

- Elements of secondary structure combine to create tertiary structure in proteins
- Various common topologies
- All  $\alpha$  proteins
- All  $\beta$  proteins
- $\alpha$  and  $\beta$  proteins in combination:
  - ▶  $\alpha/\beta$  proteins alternate  $\alpha$  and  $\beta$  segments
  - ▶  $\alpha + \beta$  proteins have larger chunks of each

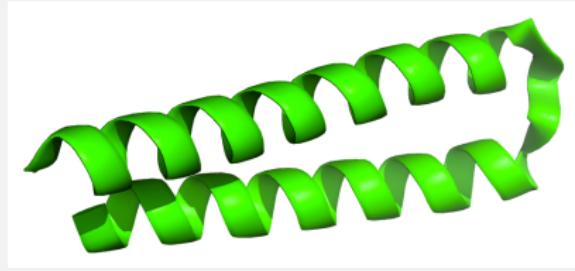
# All $\alpha$ proteins

The lone helix

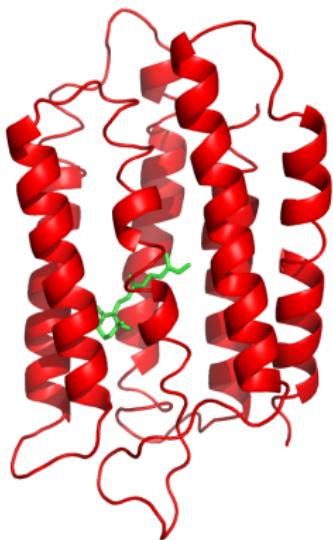
Glucagon (1GCN)



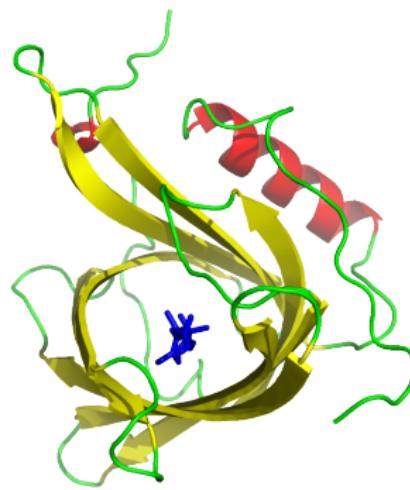
The helix turn helix motif (1ROP)



# Bacteriorhodopsin: a seven helix transmembrane protein



## Beta structures: $\beta$ barrels



The  $\beta$  barrel of human retinol binding protein

## Types of $\beta$ barrels

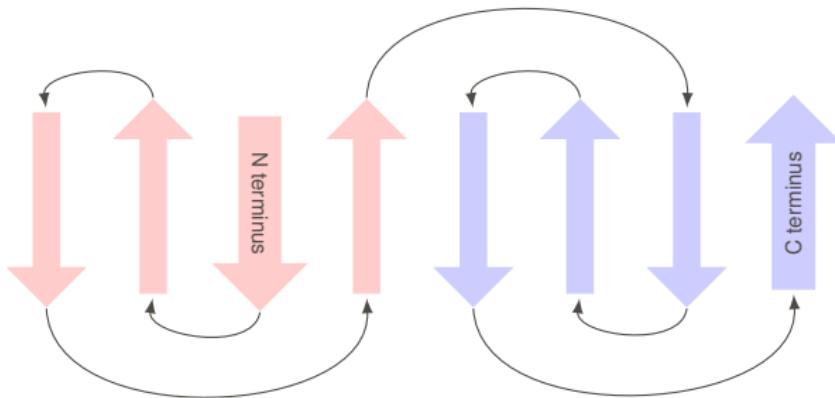
**Up and down** series of beta strands, hydrogen bonded to the preceding and following strands in primary sequence

**Greek key** containing a *Greek key* motif, with non-adjacent strands and a beta hairpin

**Jelly roll** a complex non-local form of Greek key

**others** We will see more of these and other structural motifs next week

# Greek keys motifs, a common motif in $\beta$ barrels



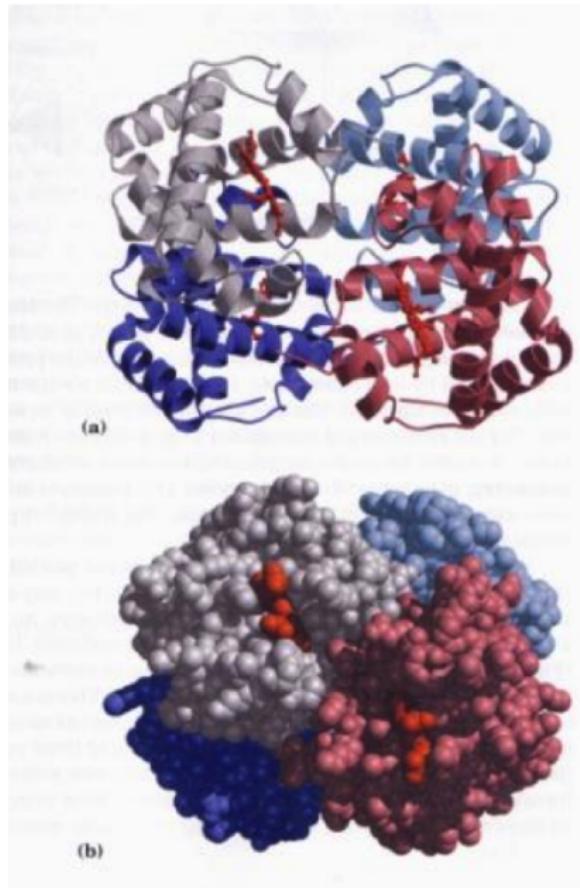
Each motif may fold independently

Motifs can be strung together into larger domains

One type of Greek key topology

# Quaternary structure

- The first two protein structures determined, myoglobin and hemoglobin, have homologous sequences, with several gene duplications
- hemoglobin is a tetramer composed of two  $\alpha$  subunits and two  $\beta$  subunits
- The *quaternary* structure refers to the arrangement of separate subunits



# Not all proteins fold: Intrinsically unstructured proteins

- For many years, it was believed that all proteins folded into stable structures under physiological conditions
- We now know this is not the case
- This challenges the *protein structure hypothesis* that sequence determines structure determines function

# Properties of intrinsically disordered proteins

- Biological
  - ▶ They can link other proteins
  - ▶ They can assume structure upon binding to other proteins or nucleic acids (coupling)
  - ▶ They have been implicated in disease (e.g., p53 and BRCA1 in cancers)
- Sequence
  - ▶ low levels of hydrophobic residues (unable to bury a hydrophobic core)
  - ▶ Often relatively low sequence complexity

# Topic

Protein chemistry review

Basics of protein structure

Higher level protein structure

The Protein Data Bank

Structure visualization

Roundup and next week

# The PDB itself

- Currently maintained by RCSB, the Research Collaboratory for Structural Bioinformatics
- The PDB is the longest maintained database in bioinformatics
- <http://www.rcsb.org/pdb/>
- 85,435 structures in October 2012
  - ▶ Proteins
  - ▶ nucleic acids
  - ▶ complexes

# experimental structure determination and the PDB

- All PDB structures are determined experimentally
  - ▶ X-ray crystallography
  - ▶ NMR spectroscopy
  - ▶ Electron microscopy

# Getting a structure out of the PDB

- Every entry has a four-character alphanumeric identifier
- The search interface is excellent
  - ▶ structure search
  - ▶ keyword search
  - ▶ blast against sequences in the database

# Topic

Protein chemistry review

Basics of protein structure

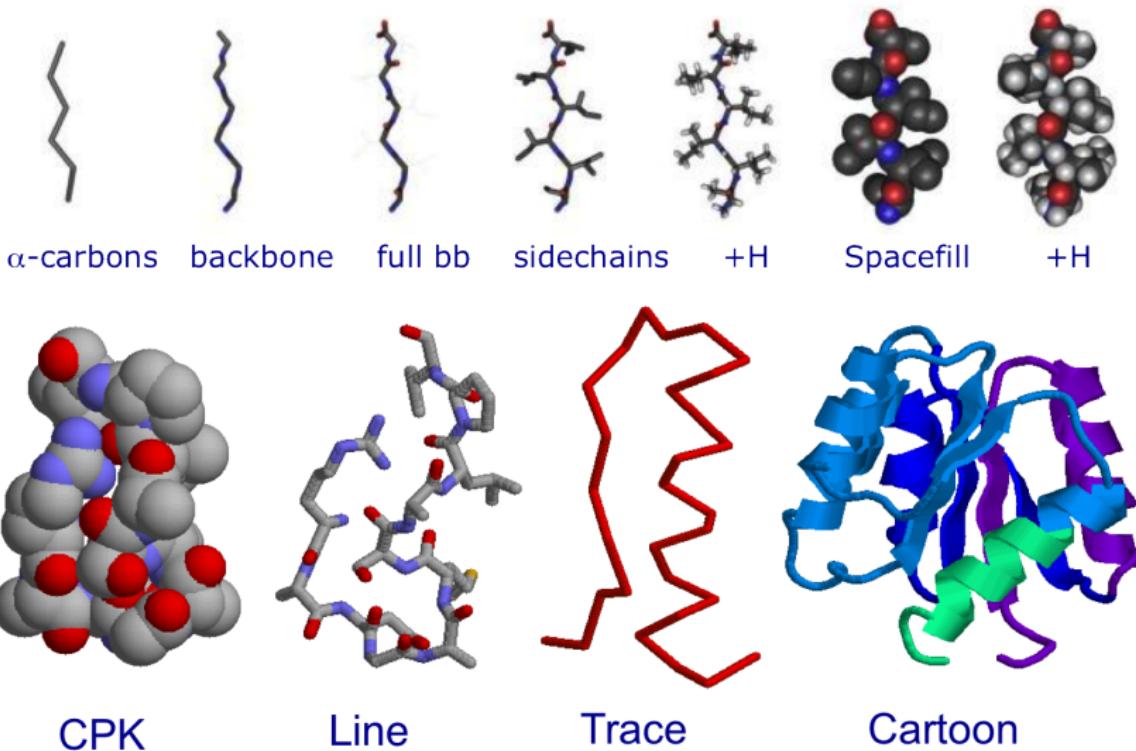
Higher level protein structure

The Protein Data Bank

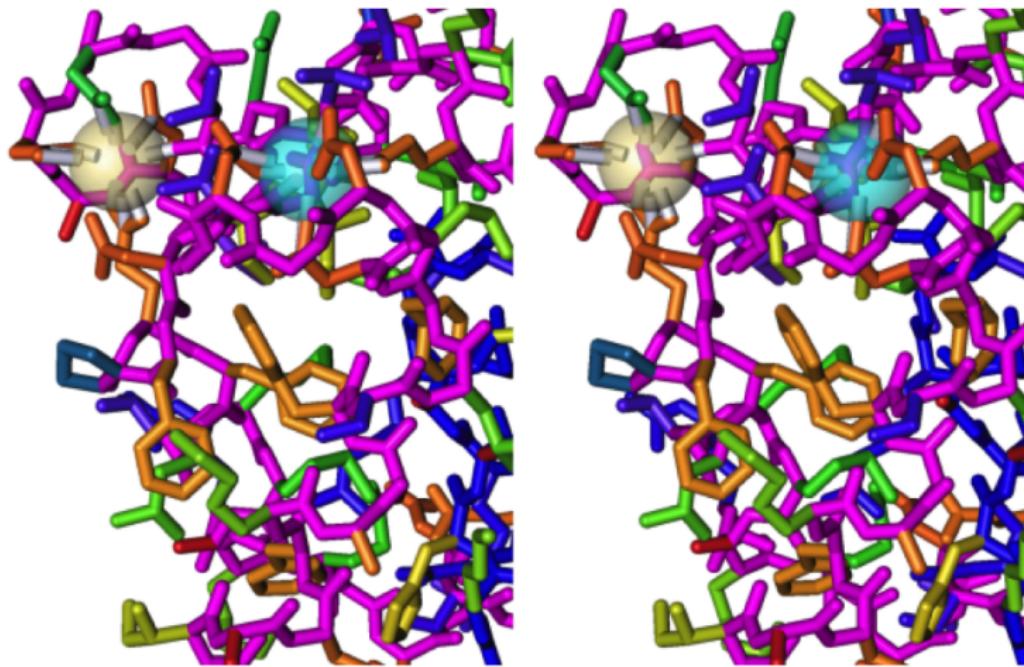
Structure visualization

Roundup and next week

# Many visual abstractions to aid interpretation of protein structure



## Stereo vision



protein 1CDL

## Alignment and measurement

- Good visualization tools are extremely powerful
- Visualization tools allow you to align structures to each other
- They allow you to measure distances in protein structures and perform calculations on them

## Visualization tools: JMOL

- A browser plug-in with the PDB
- Nothing needed except a capable browser (i.e., *not* IE)
- This will be the default at PDB

# Visualization tools: PyMol

- Open source (yet a commercial product)
- Extensible (using the Python scripting language)
- Multi-platform
- <http://www.pymol.org>

# Visualization tools: Cn3D

- NCBI standard
- Browser plugin linked to NCBI Entrez Structure database
- Nothing needed except a capable browser (i.e., *not* IE)
- `http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml`

# Topic

Protein chemistry review

Basics of protein structure

Higher level protein structure

The Protein Data Bank

Structure visualization

Roundup and next week

# What we learned I

- Proteins, as sequences of amino acids, adopt a range of folds that determine function
- Amino acids in proteins are all *L*, not *D*
- Amide bonds between adjacent amino acids are partially planar
- the dihedral angles  $\phi$  and  $\psi$  largely determine protein backbone structure
- Ramachandran plots show  $\phi$  and  $\psi$  angles, and can quickly characterize secondary structure
- Secondary structural elements are stabilized by characteristic hydrogen bonding patterns
- Higher order structure is characterized by *structural motifs*
- The PDB catalogs macromolecular structures

## What we learned II

- Structures in the PDB are determined by X-ray crystallography, NMR spectroscopy, or electron microscopy
- Structural motifs can be recognized and characterized in families
- Visualization tools allow detailed interaction with molecular structure

## Next week

- Next week, we will use sequence to model structure
- we will compare structures to each other and to families

# References

-  Hutchinson, E G and J M Thornton (1994). "A revised set of potentials for beta-turn formation in proteins." In: *Protein science : a publication of the Protein Society* 3.12, pp. 2207–16. DOI: [10.1002/pro.5560031206](https://doi.org/10.1002/pro.5560031206) (cit. on p. 17).
-  Lovell, Simon C et al. (2003). "Structure Validation by C  $\alpha$  Geometry : , and C Deviation". In: 450.August 2002, pp. 437–450 (cit. on p. 21).
-  Richardson, J S (1981). "The anatomy and taxonomy of protein structure." In: *Advances in protein chemistry* 34, pp. 167–339 (cit. on p. 10).