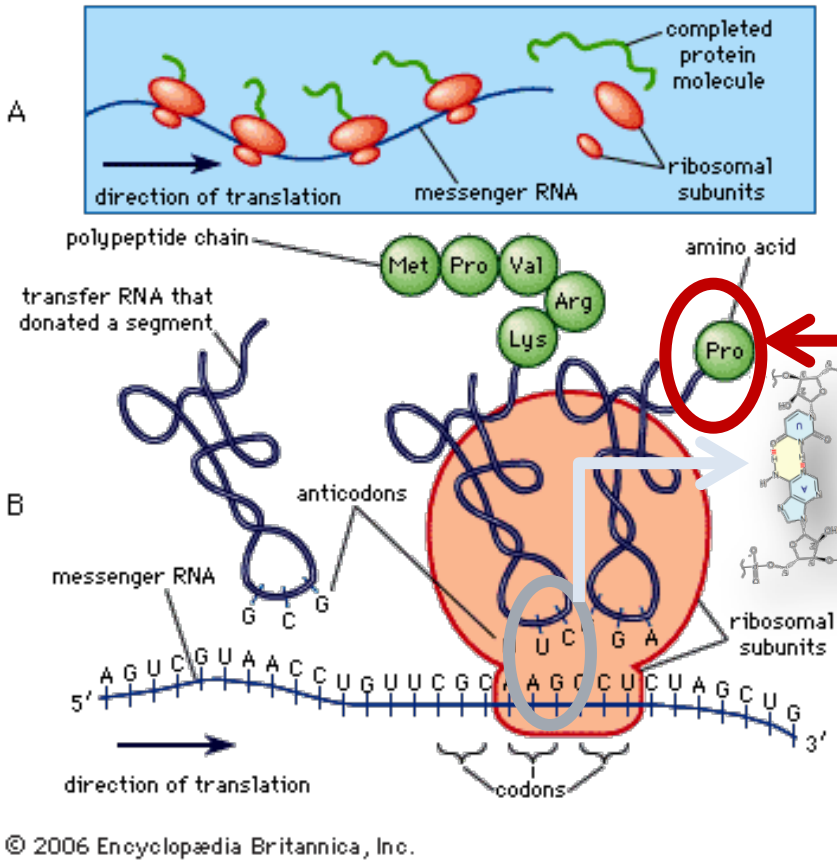
Instructor: Sael Lee
CS549 Spring – Computational Biology

Lecture 15:
Analyzing Protein Structure and Dynamics

Resources:

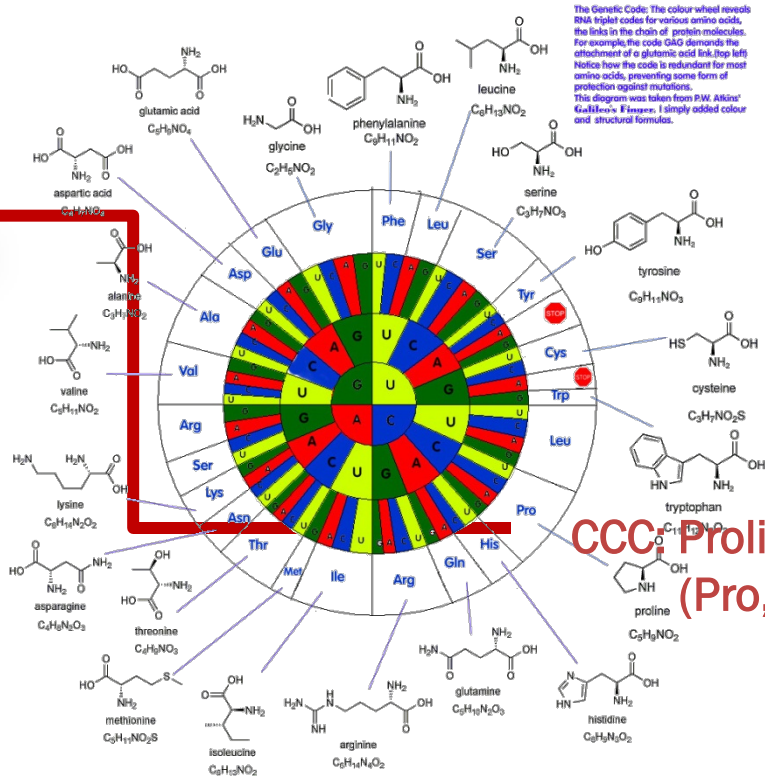
- Slide 9 of Protein Bioinformatics, Spring 2013 Daisuke Kihara
- Wikipedia

Translation process



http://content.answcdn.com/main/content/img/BritannicaConcise/images/780.gif

Codon: Three nucleic acid coding one of 20 amino acid (alphabet of 20 size) + START & STOP CODEN



CCC: Proline (Pro, P)

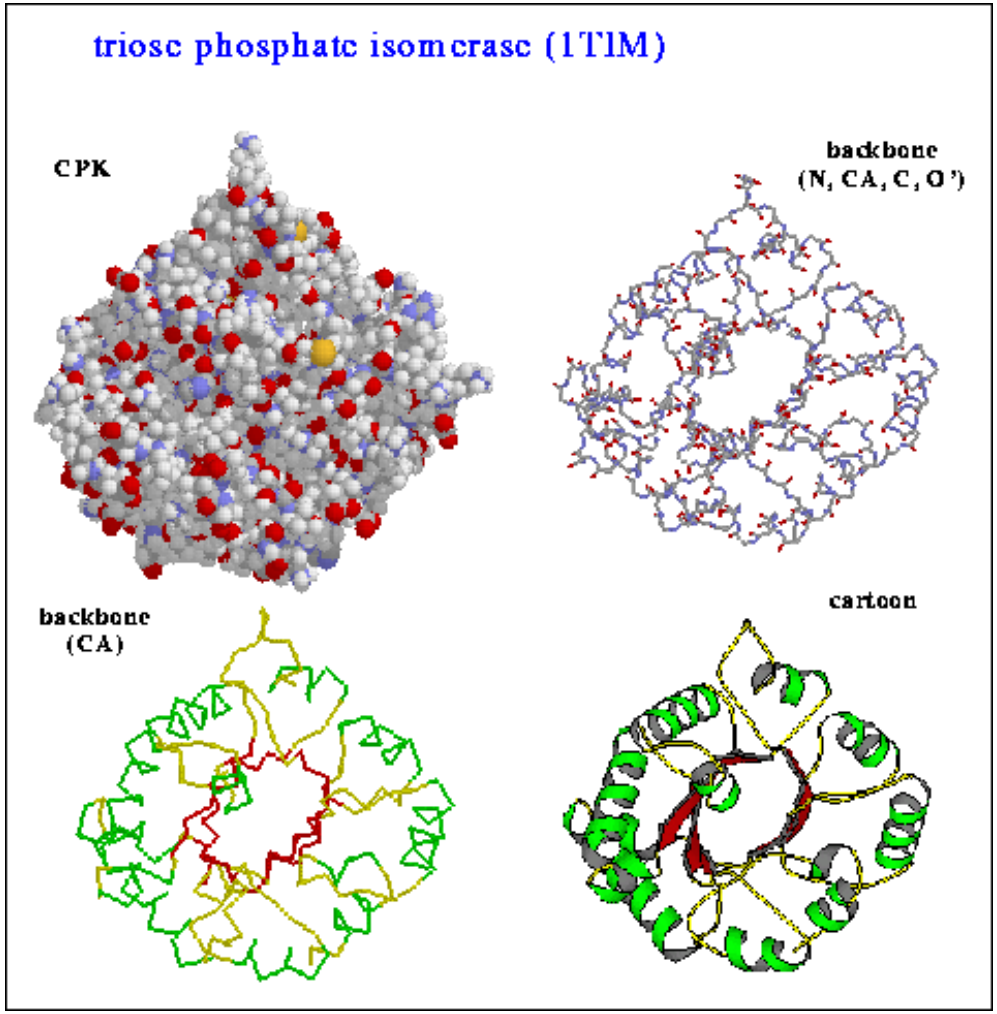
Start codon: AUG (also Methionine (Met, M))
Stop codon: UAA, UAG, UGA

Protein Structure

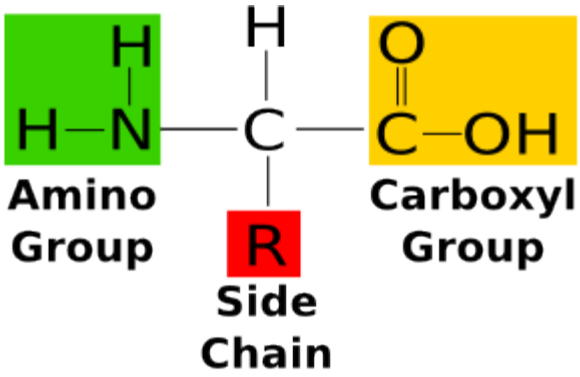
Amino Acid Sequence

```
APRKFFVGGNWKMNGDKKSLGELIHTLNGAKL
SADTEVVCGAPSIYLDFARQKLDAKIGVAAQN
CYKVPKGAFTGEISPAMIKDIGAAWVILGHSE
RRHVFGESDELIGQKVAHALAEGLGVIACIGE
KLDEREAGITEKVVFEQTKAIADNVKDWSKVV
LAYEPVWAIGTGKTATPQQAQEVHEKLRGWLK
SHVSDAVAQSTRIIYGGSVTGGNCKELASQHD
VDGFLVGGASLKPEFVDIINAKH
```

=



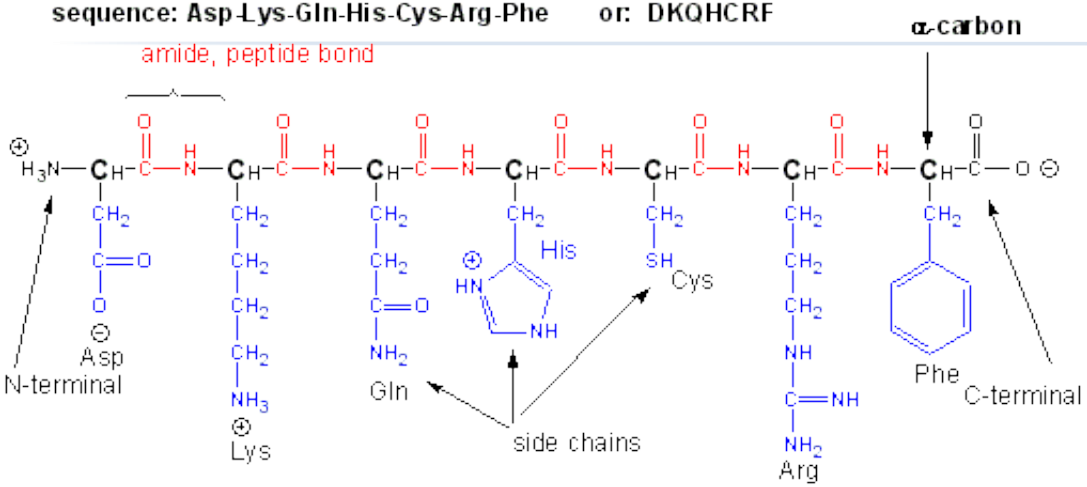
General Structure of AA



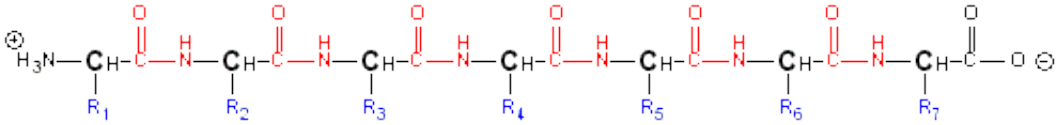
DIFFERENT REPRESENTATIONS OF A POLYPEPTIDE

1. Full structure

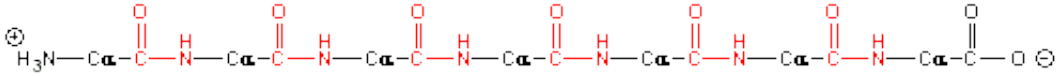
sequence: Asp-Lys-Gln-His-Cys-Arg-Phe or: DKQHCRF



2. Abbreviated Side Chains



3. Backbone



4. Main Chain Trace

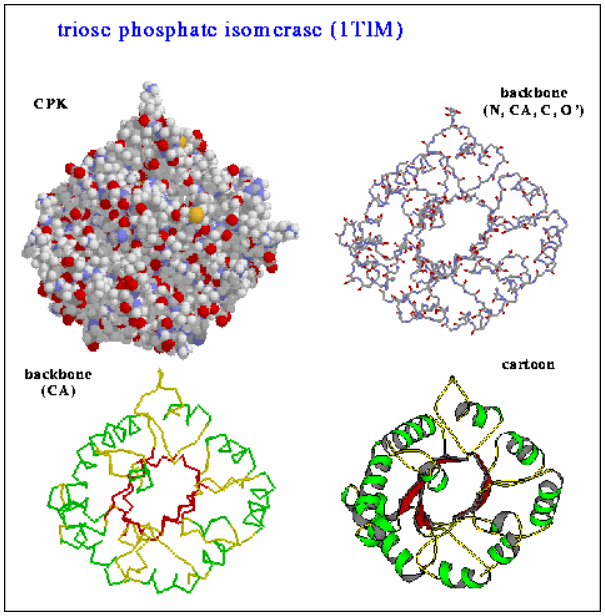
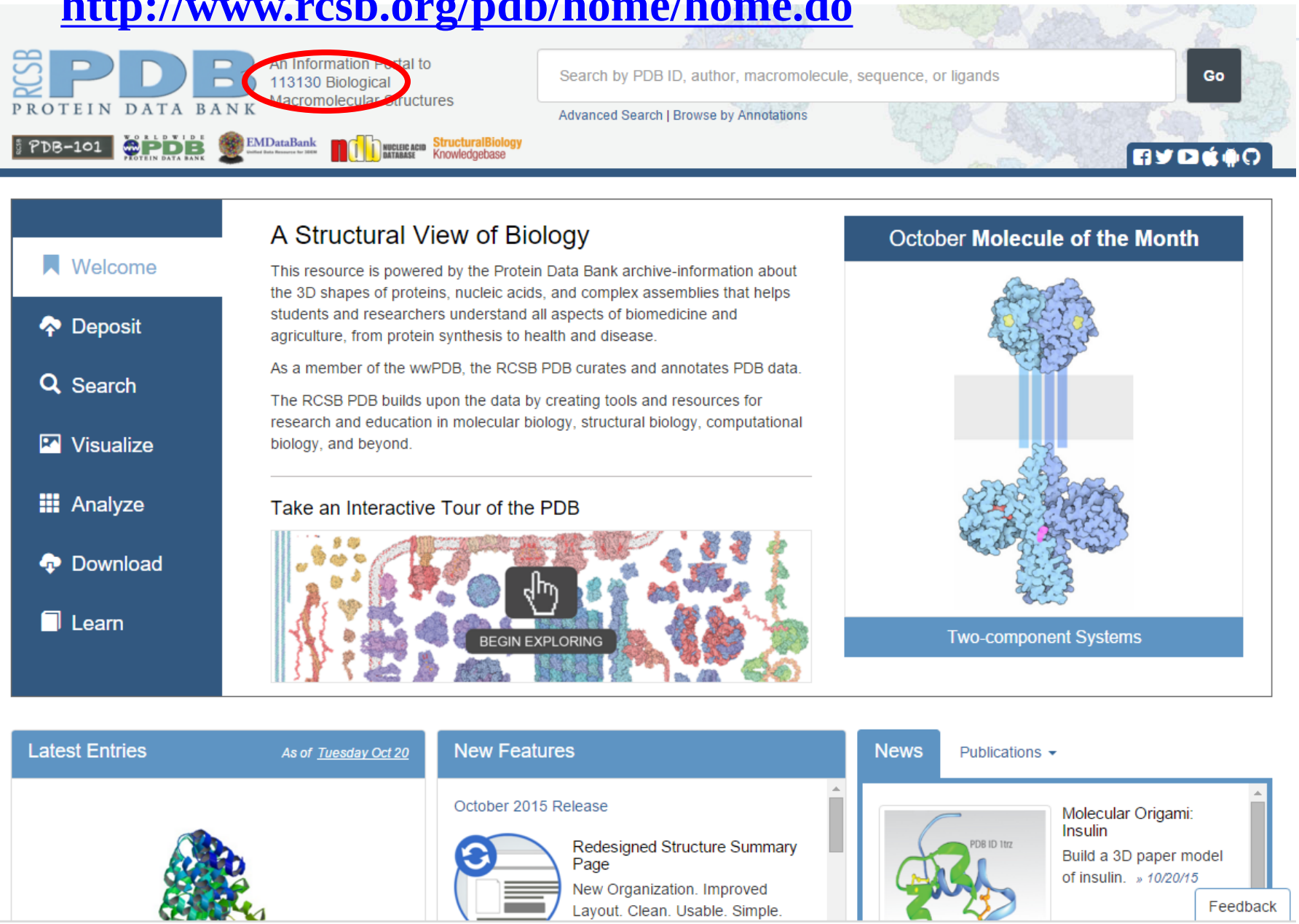


Figure from : http://employees.csbsju.edu/hjakubowski/classes/ch331/protstructure/olprotein-aminoacid.html

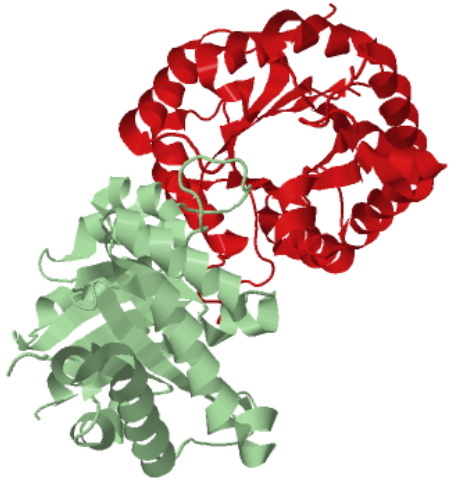
PDB database

http://www.rcsb.org/pdb/home/home.do



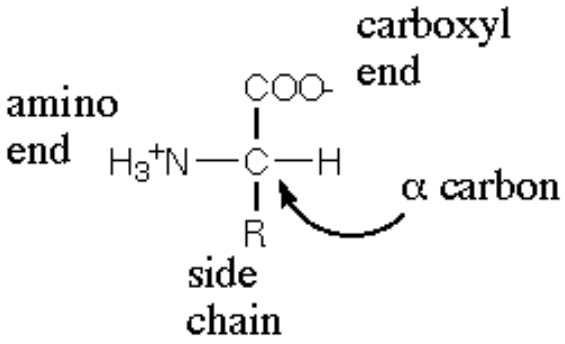
Coordinates (PDB file: 1tim)

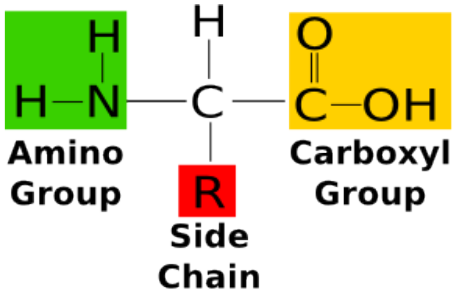
```
HEADER    ISOMERASE(INTRAMOLECULAR OXIDOREDUCTSE) 01-SEP-76   1TIM
TITLE     STRUCTURE OF TRIOSE PHOSPHATE ISOMERASE FROM CHICKEN MUSCLE
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: TRIOSEPHOSPHATE ISOMERASE;
COMPND   3 CHAIN: A, B;
COMPND   4 EC: 5.3.1.1;
COMPND   5 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: GALLUS GALLUS;
SOURCE   3 ORGANISM_COMMON: CHICKEN;
SOURCE   4 ORGANISM_TAXID: 9031
KEYWDS    ISOMERASE, ISOMERASE(INTRAMOLECULAR OXIDOREDUCTSE)
EXPDTA    X-RAY DIFFRACTION
AUTHOR    D.W.BANNER,A.C.BLOOMER,G.A.PETSKO,D.C.PHILLIPS,I.A.WILSON
 ...
ATOM      1  N   ALA A   1      43.240  11.990  -6.915  1.00  0.00           N
ATOM      2  CA  ALA A   1      43.888  10.862  -6.231  1.00  0.00           C
ATOM      3  C   ALA A   1      44.791  11.378  -5.094  1.00  0.00           C
ATOM      4  O   ALA A   1      44.633  10.992  -3.937  1.00  0.00           O
ATOM      5  CB  ALA A   1      44.722  10.051  -7.240  1.00  0.00           C
ATOM      6  N   PRO A   2      45.714  12.244  -5.497  1.00  0.00           N
ATOM      7  CA  PRO A   2      46.689  12.815  -4.561  1.00  0.00           C
ATOM      8  C   PRO A   2      46.042  13.601  -3.411  1.00  0.00           C
ATOM      9  O   PRO A   2      46.030  13.141  -2.267  1.00  0.00           O
ATOM     10  CB  PRO A   2      47.640  13.732  -5.359  1.00  0.00           C
ATOM     11  CG  PRO A   2      47.006  13.820  -6.760  1.00  0.00           C
ATOM     12  CD  PRO A   2      46.056  12.615  -6.882  1.00  0.00           C
ATOM     13  N   ARG A   4      45.521  14.773  -3.763  1.00  0.00           N
ATOM     14  CA  ARG A   4      44.872  15.621  -2.730  1.00  0.00           C
ATOM     15  C   ARG A   4      46.092  16.358  -2.094  1.00  0.00           C
ATOM     16  O   ARG A   4      46.091  16.680  -0.916  1.00  0.00           O
ATOM     17  CB  ARG A   4      44.264  14.706  -1.647  1.00  0.00           C
ATOM     18  CG  ARG A   4      42.729  14.802  -1.637  1.00  0.00           C
ATOM     19  CD  ARG A   4      42.158  13.894  -0.528  1.00  0.00           C
ATOM     20  NE  ARG A   4      42.592  12.508  -0.754  1.00  0.00           N
```

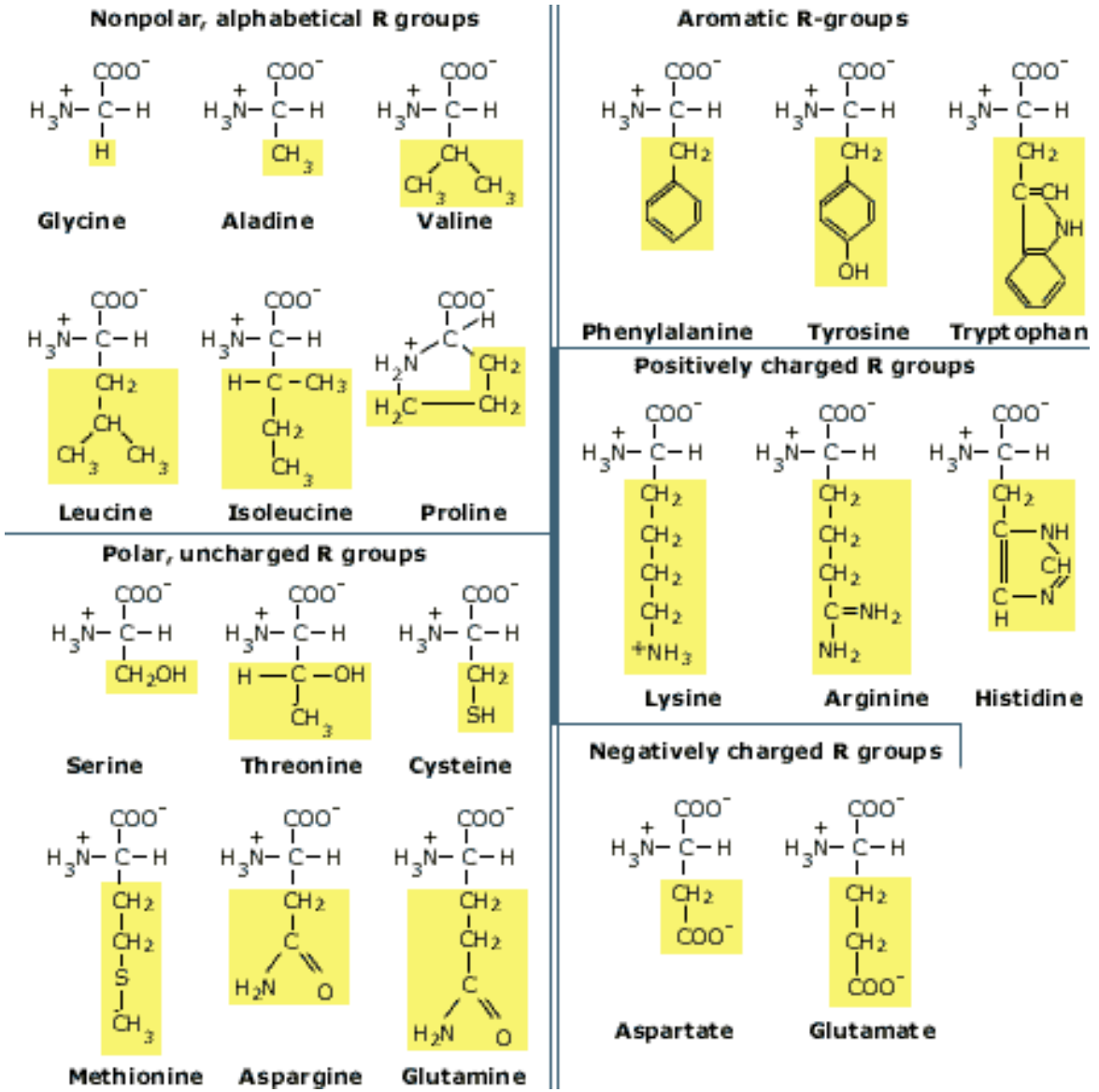


20 AA found in biological systems

All amino acids have the same general formula



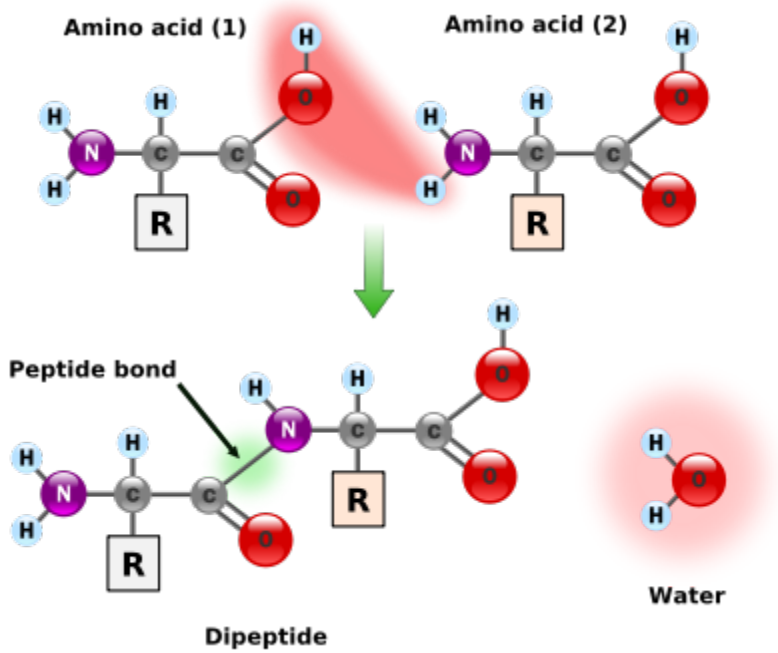




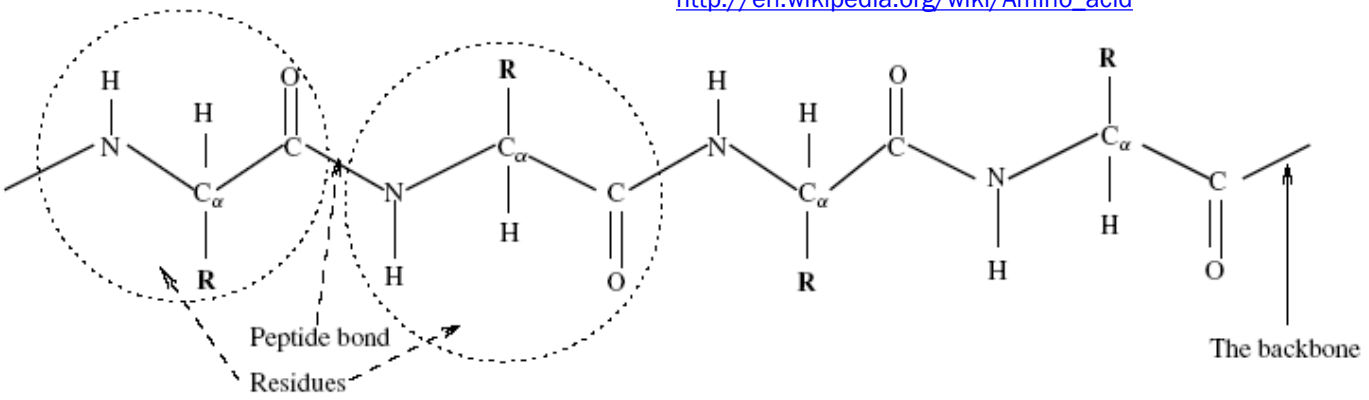
Dr. M.O. Dayhoff's resources on amino acids

- http://www.biology.arizona.edu/biochemistry/problem_sets/aa/aa.html#Essentialaa

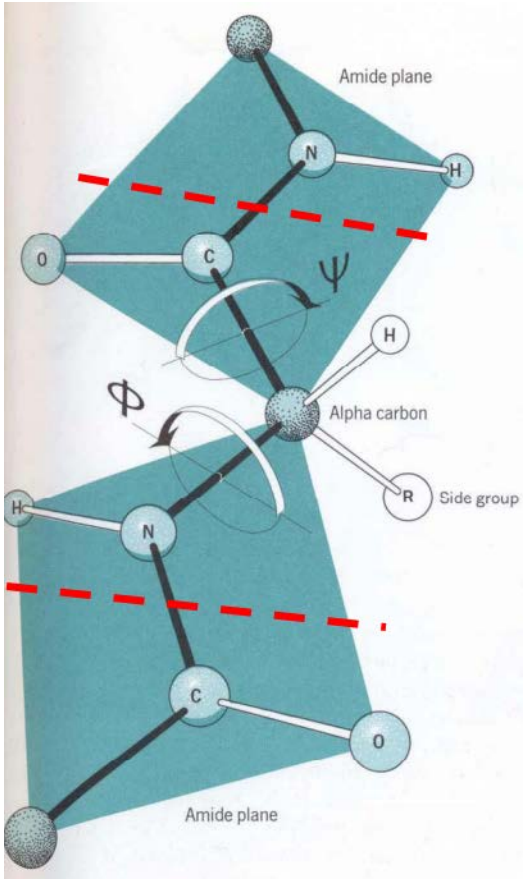
Amino acid and main chain



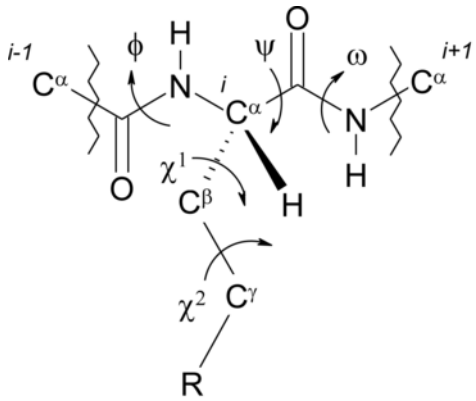
http://en.wikipedia.org/wiki/Amino_acid

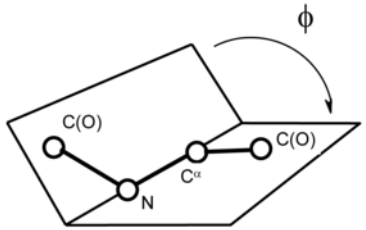


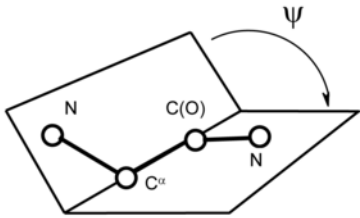
Dihedral Angles



- Dihedral Angles (Torsion angles):
  Angels between two planes.



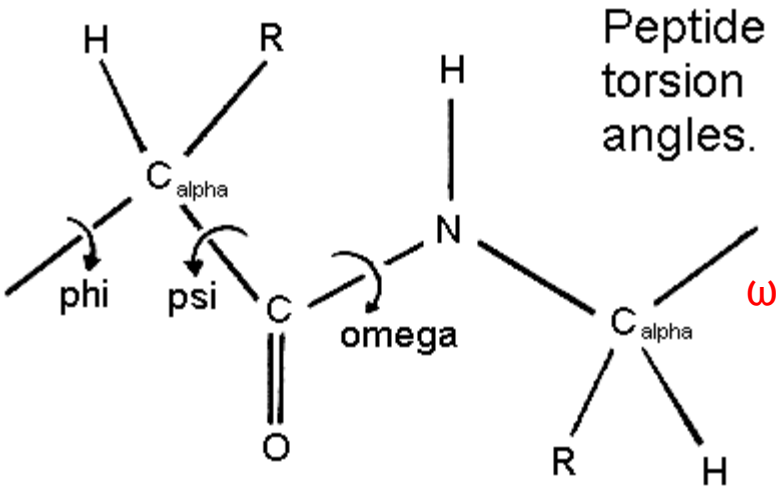


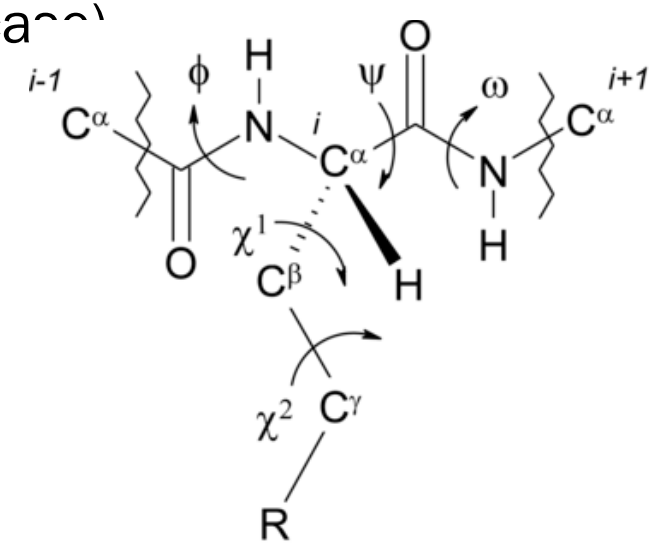


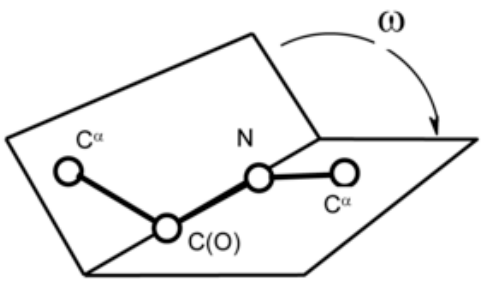
- φ (*phi*, involving the backbone atoms C'-N-C^{α}-C‘)
- ψ (*psi*, involving the backbone atoms N-C^{α}-C'-N)
- φ controls the C'-C' distance, ψ controls the N-N distance
- rotations about φ and ψ angles are the softest

Dihedral Angles

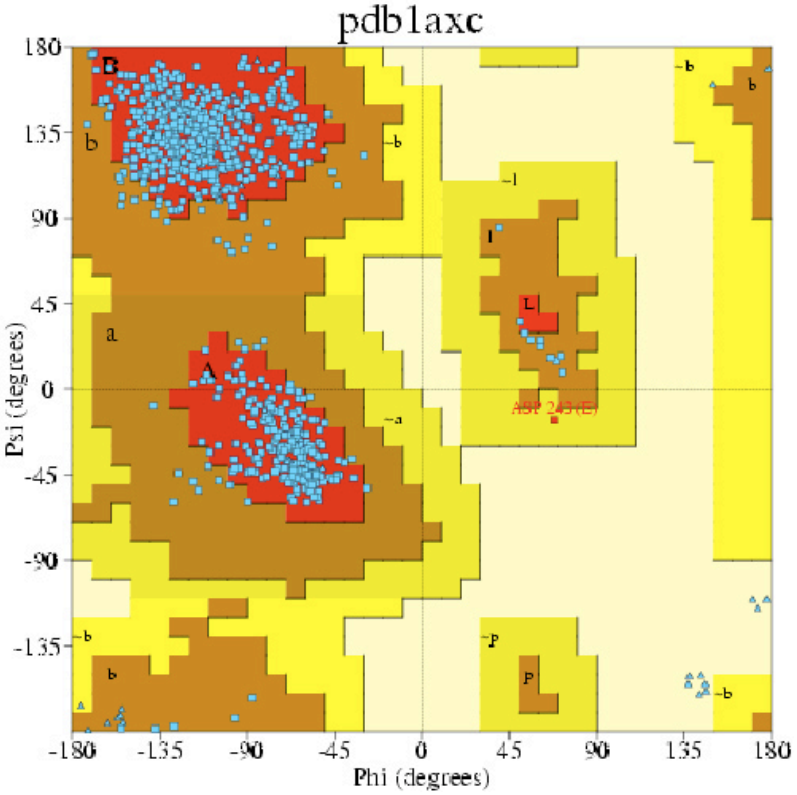
- ω (*omega*, involving the backbone atoms C^{α}-C'-N-C^{α}).
- ω controls the C^{α}-C^{α} distance
- Peptide bond usually restricts ω to be 180˚ (the typical trans case) or 0˚ (the rare cis case)







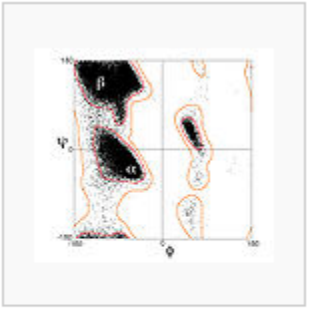
Ramachandran plot



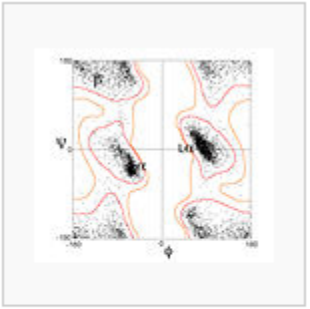
The red, brown, and yellow regions represent the favored, allowed, and "generously allowed" regions as defined by ProCheck

A Ramachandran plot

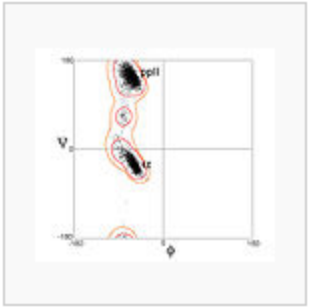
Is a visualization tools for visualizing backbone dihedral angles ψ against φ of amino acid residues in protein structure.



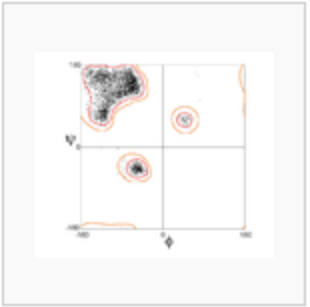
Ramachandran plot for the general case; data from Lovell 2003



Ramachandran plot for Glycine



Ramachandran plot for Proline



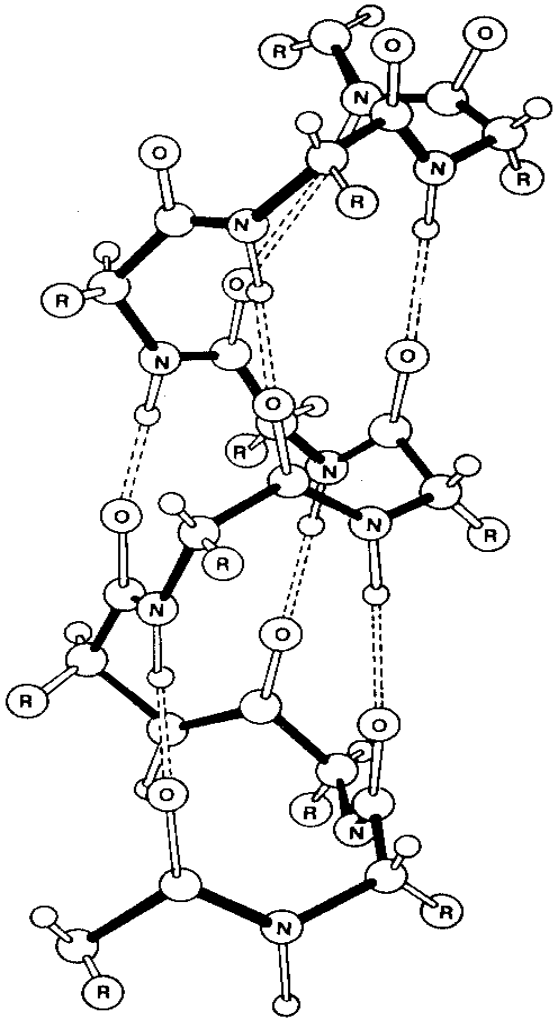
Ramachandran plot for pre-Proline

https://en.wikipedia.org/wiki/Ramachandran_plot

Protein secondary structures

- Proteins packs the hydrophobic side chains inside the molecule.
- Proteins have hydrophobic kernel and hydrophilic surface.
- The backbone is polar, hence hydrophilic.
- To neutralize this hydrophility there are hydrogen bindings between NH and CO on the backbone.
- This is done by constructing regular *secondary structures*
  - *Helices, alpha most usual*
  - *Beta sheets*

Alpha Helix



Alpha-helix:

- Right-handed helix
- 3.6 residues per helix turn
- Hydrogen bond between n and n+4

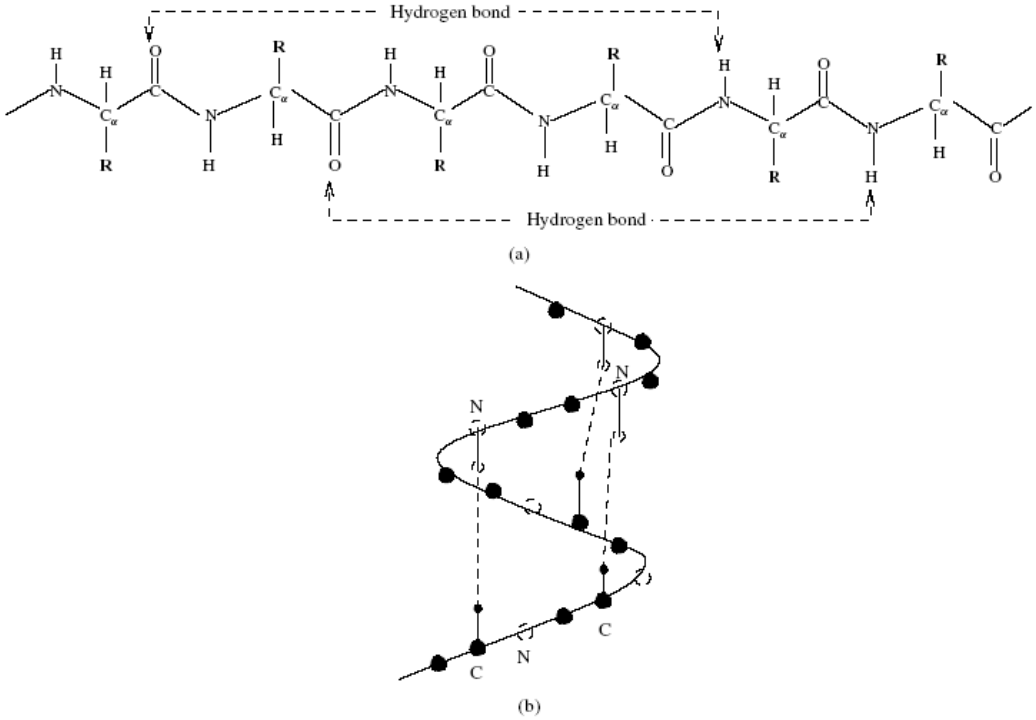


Figure B.6 (a) Schematic of the hydrogen bonding forming an α-helix. (b) For the hydrogen bonding to take place, the sequence must be formed as a helix in the space.

Beta Sheets

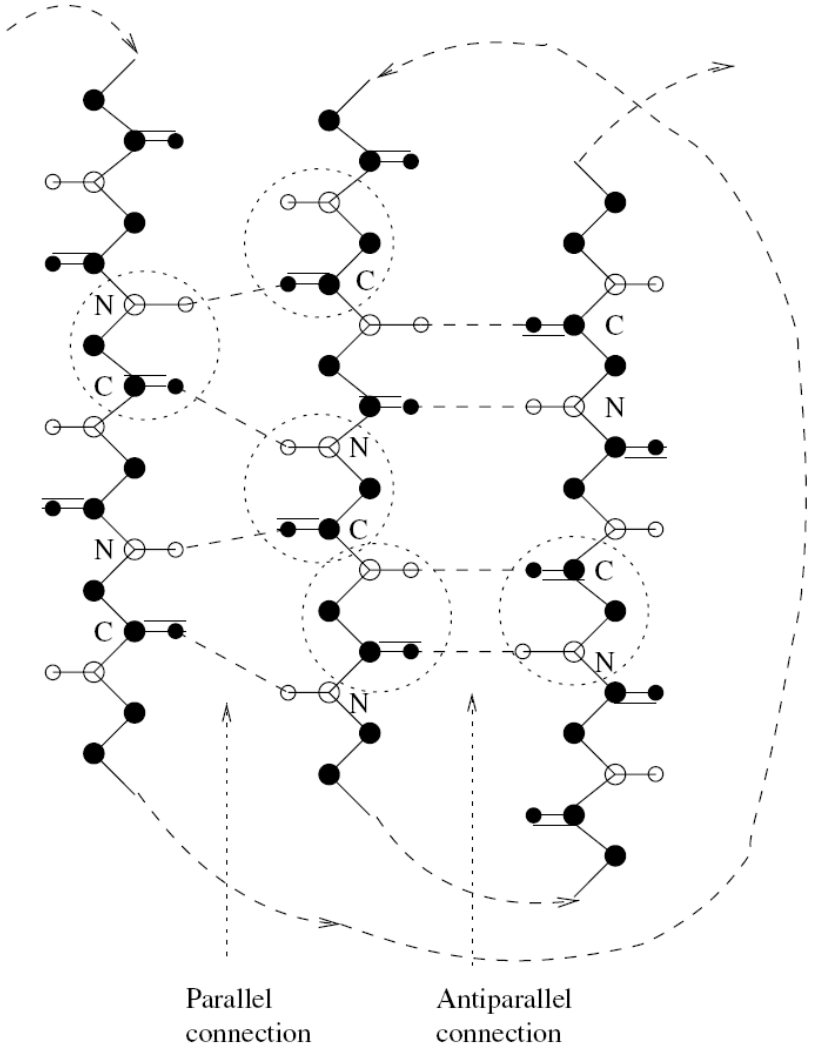
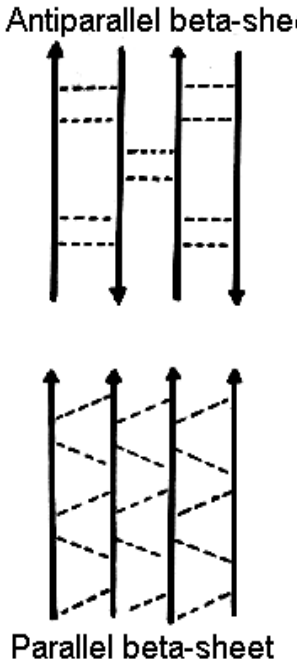
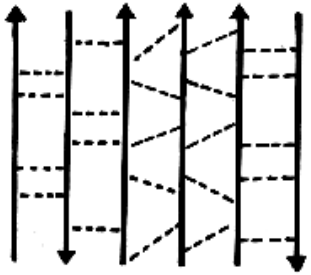


Figure B.7 A β-sheet formed of three β-strands, with one parallel and one antiparallel set of H-bonds. Note that strands near in space do not need to be near in sequence.



The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.



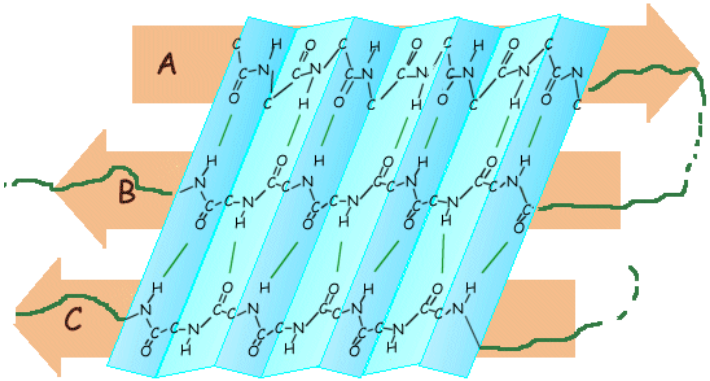
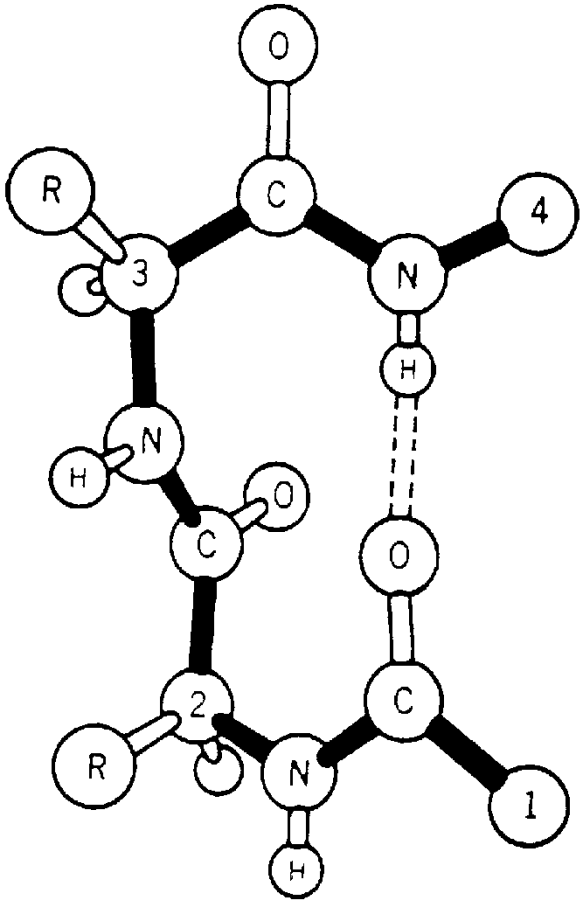
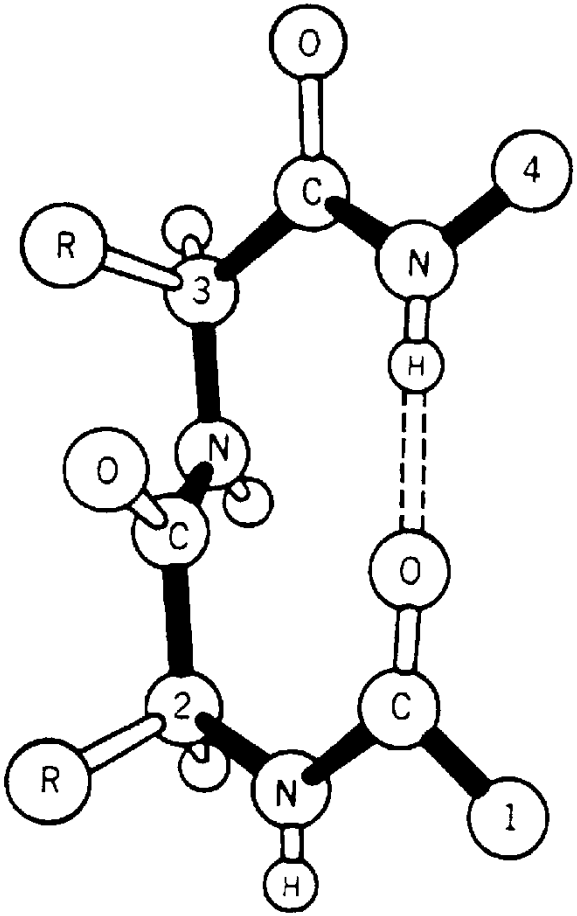


Diagram 1: Beta pleated sheet. The lateral groups (R) are not shown.

Beta Turn

- 4 residues in length
- Enables structure to have an 180 degree turn



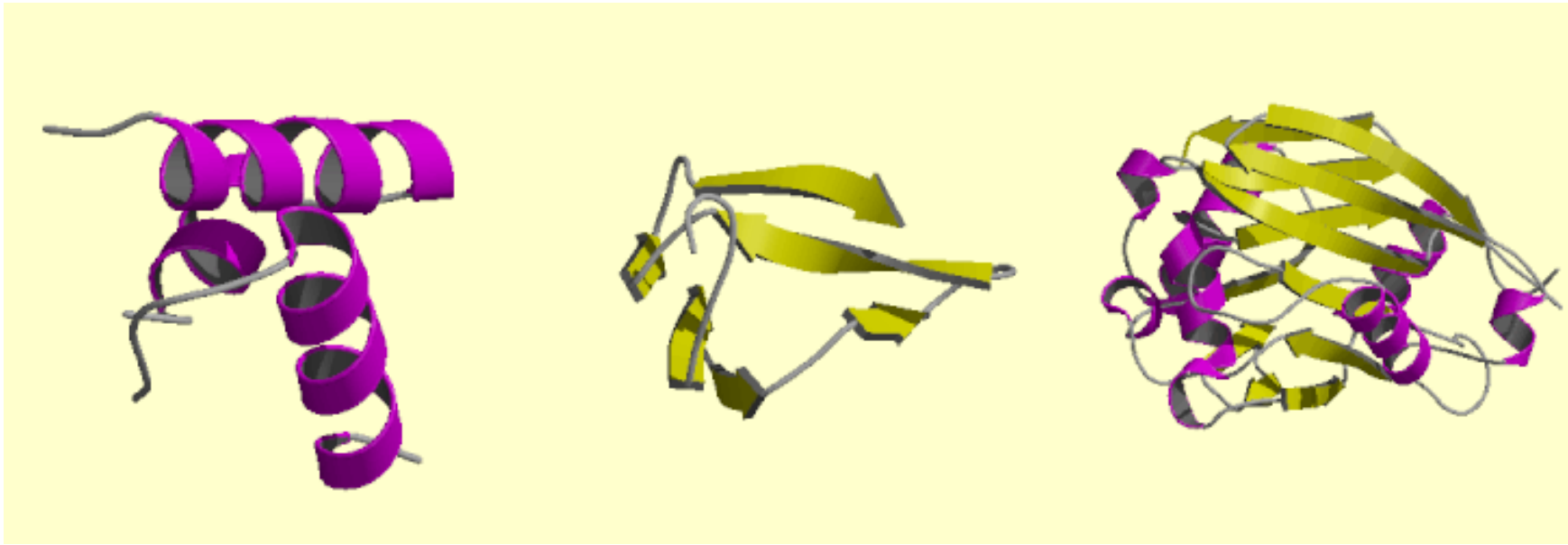


imtech.res.in

Protein Tertiary Structure

Driving force for folding:

- Hydrophobic effect
- Electrostatic
- Hydrogen bond
- Disulfide bond



Protein Structure Classification

SCOP Classification

- SCOP: Structural Classification of Proteins

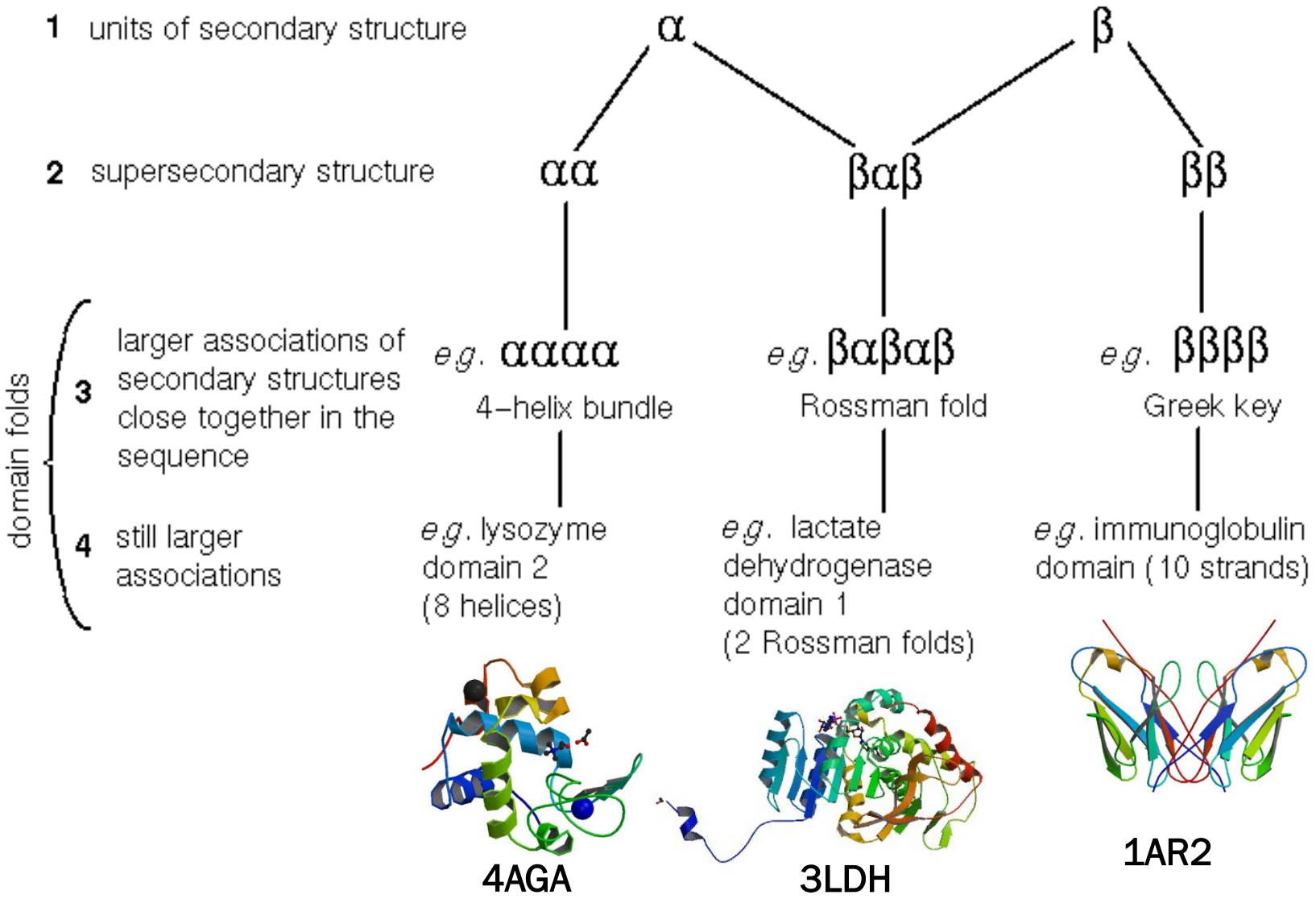
- **Classes:**
  - All alpha proteins (126)
  - All beta proteins(81)
  - Alpha and beta proteins (a/b) (87)
    Mainly parallel beta sheets (beta-alpha-beta units)
  - Alpha and beta proteins (a+b) (151)
    Mainly antiparallel beta sheets (segregated alpha and beta regio ns)
  - Multi-domain proteins (alpha and beta) (21)
    Folds consisting of more than one domain of different classes

Protein Structure Classification

SCOP Classification

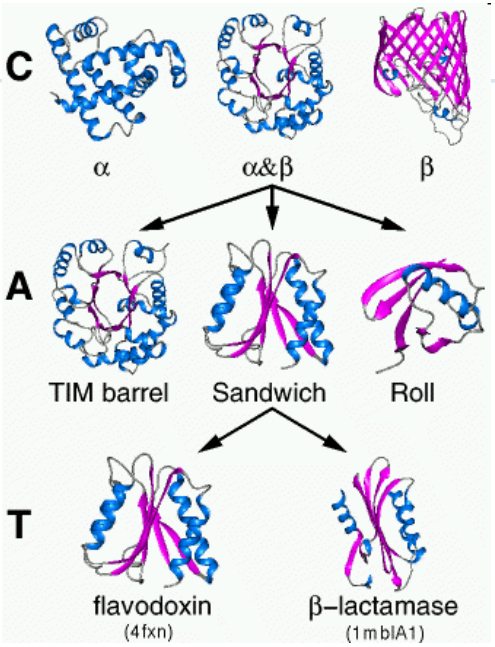
- **Classes cont.:**
  - Membrane and cell surface proteins and peptides (10)
    Does not include proteins in the immune system
  - Small proteins (44)
    Usually dominated by metal ligand, heme, and/or disulfide brid ges
  - Coiled coil proteins (4)
  - Low resolution protein structures (4)
  - Peptides (61)
    Peptides and fragments
  - Designed proteins (17)
    Experimental structures of proteins with essentially non-natura l sequences

SCOP cont.


1 units of secondary structure
α
β
2 supersecondary structure
αα
βαβ
ββ
domain folds
3 larger associations of secondary structures close together in the sequence
4 still larger associations
e.g. αααα
4–helix bundle
e.g. βαβαβ
Rossman fold
e.g. ββββ
Greek key
e.g. lysozyme domain 2 (8 helices)
e.g. lactate dehydrogenase domain 1 (2 Rossman folds)
e.g. immunoglobulin domain (10 strands)
4AGA
3LDH
1AR2

Protein Structure Classification

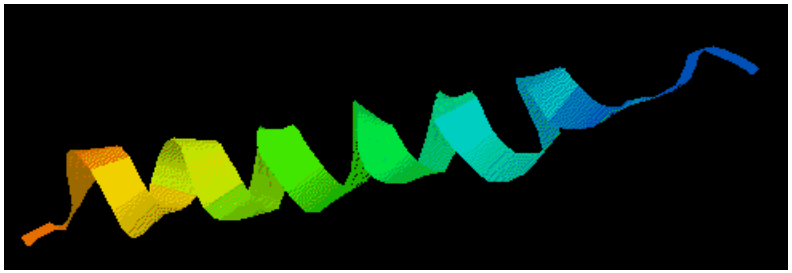
CATH database



- Class, Architecture, Topology, Homology
  - Architecture: the global spatial arrangement of 2ndary structure segments
  - Topology: connectivity of the 2ndary structure segments is also counted
- Protein structure comparison program, SSAP is used

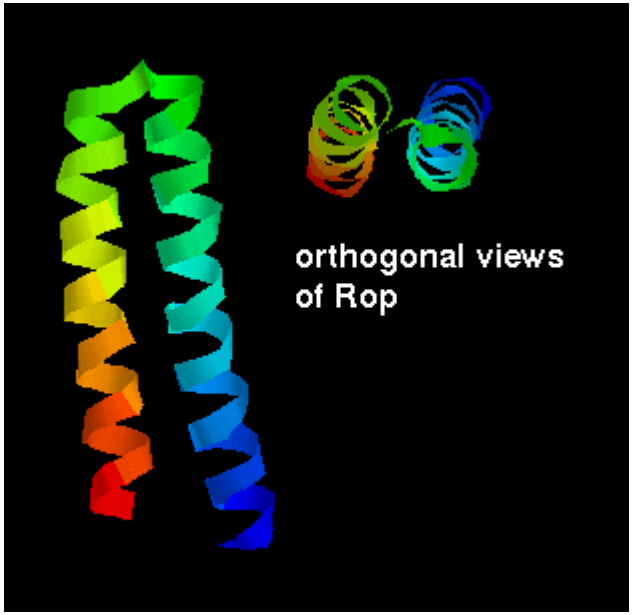
Class	Architecture	Topology	Homologous Superfamily	S35 Family	S60 Family	S95 Family	S100 Family	Domains
1	5	376	839	2763	3571	4679	9217	32396
2	20	228	514	2514	3573	5668	9824	39140
3	14	577	1082	5849	8381	10626	21900	79038
4	1	101	114	204	253	352	547	2346
Total	**40**	**1282**	**2549**	**11330**	**15778**	**21325**	**41488**	**152920**

All-Alpha



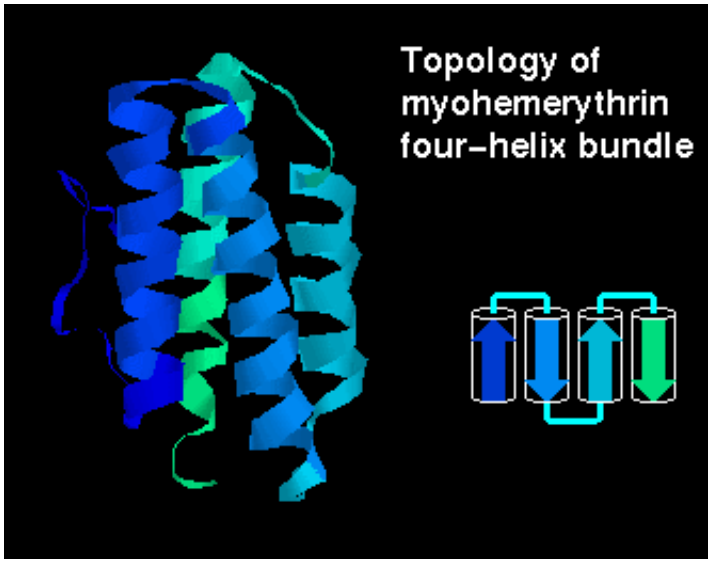
- **The Lone Helix:** small proteins (or peptides) which consist of little more than a single helix. Example: **glucagon**, a hormone involved in regulating sugar metabolism in mammals (as is insulin).

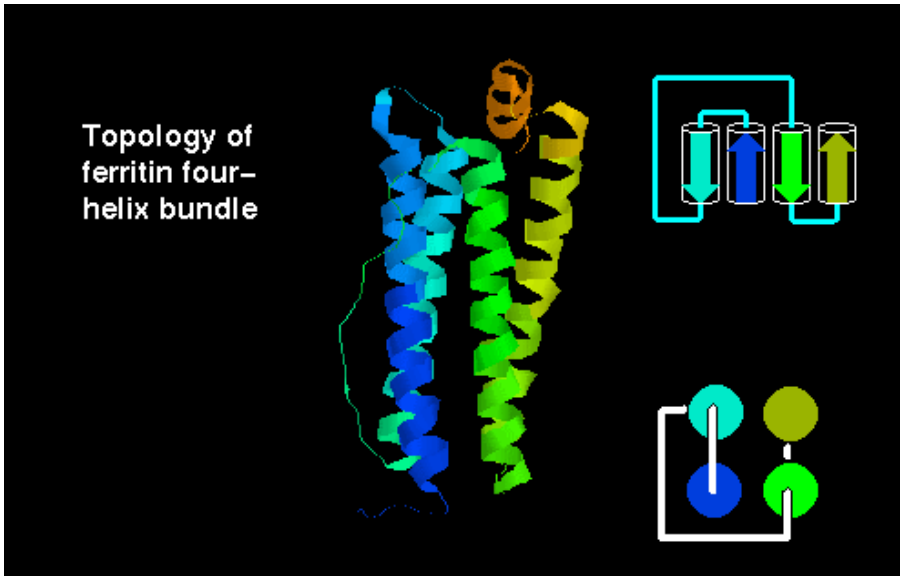
- **Helix-turn-helix motif**: The simplest packing arrangement of a domain of two helices is for them to lie *antiparallel*, connected by a *short loop*.



Four Helix Bundle

- **Topology:** The four helices may be arranged in a simple up-and -down topology, or more complex arrangement.

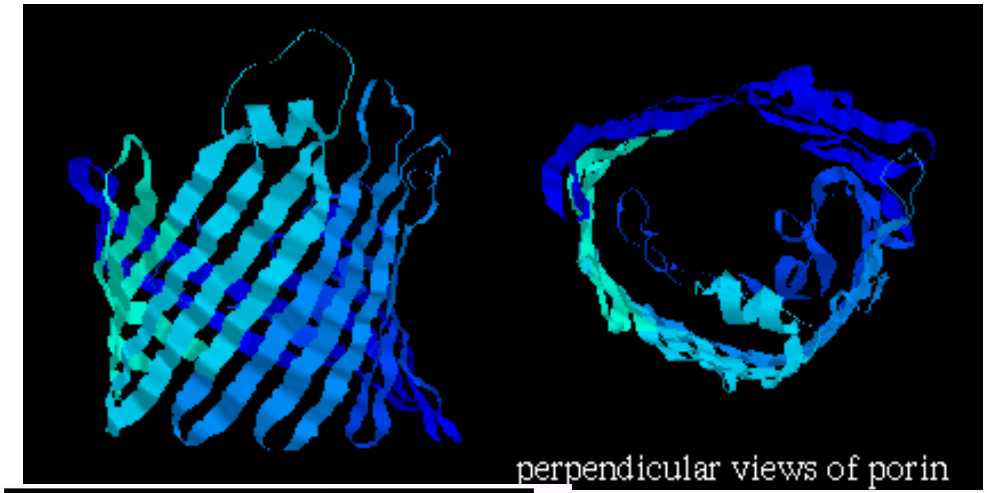




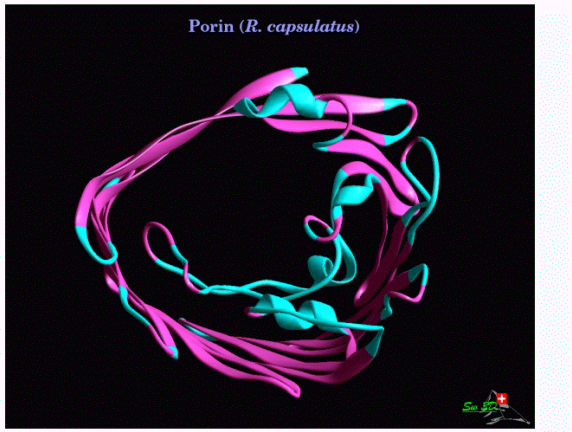
All-Beta

- **Beta Sandwiches and Beta Barrels**: In the immunoglobulin fold, the strands form two sheets packed against each other, forming a "**beta sandwich**".

Beta Barrels

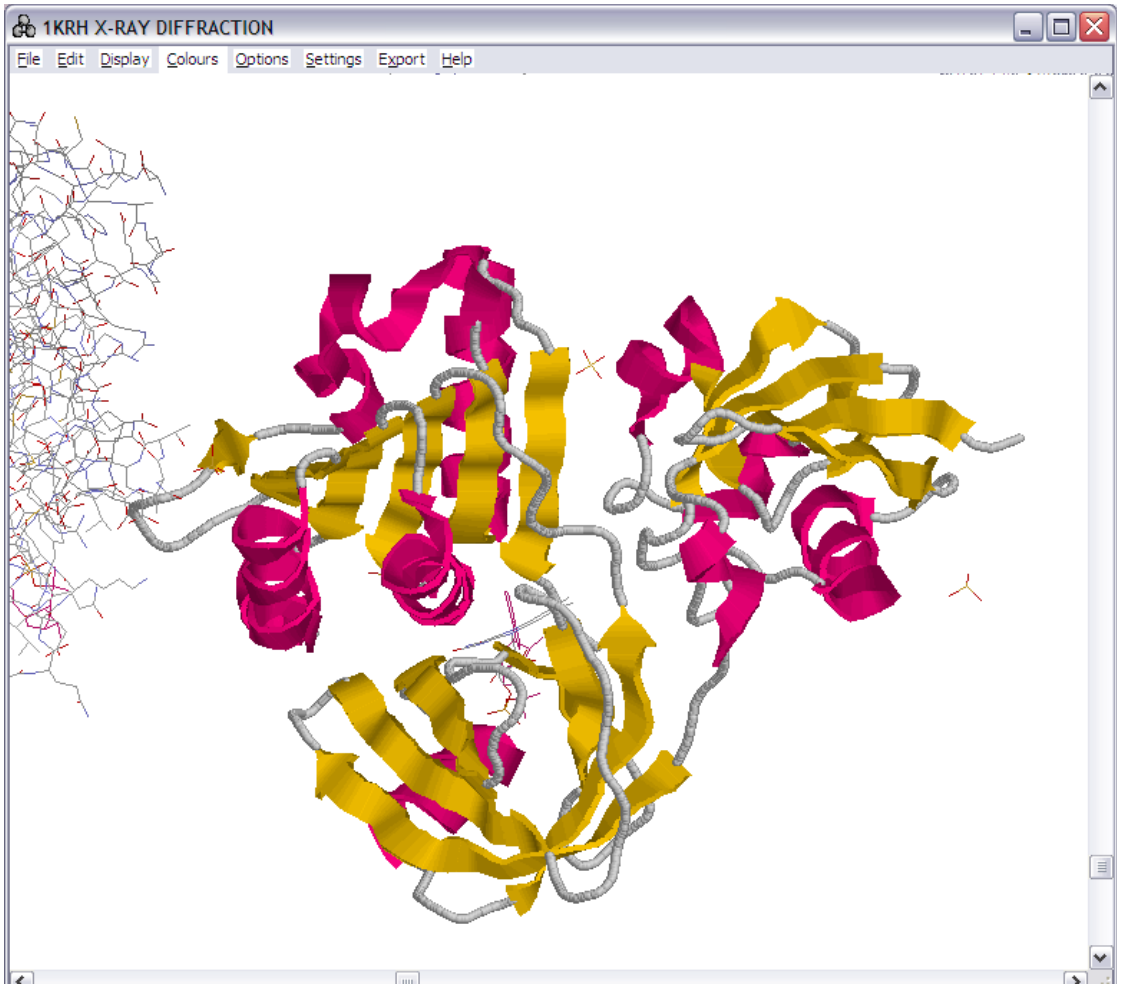


- **beta barrels**: Some antiparallel beta sheet domains are better described as **beta barrels** rather than beta sandwiches



- eg. Porin: beta-sheets in a 16-stranded **beta-barrel** formation and forms a pore in the membrane 1.7 - 2.5 nm in diameter

α/β protein (Rossmann fold, 1krh)



Cannot use pure dynamic programming for structure comparison

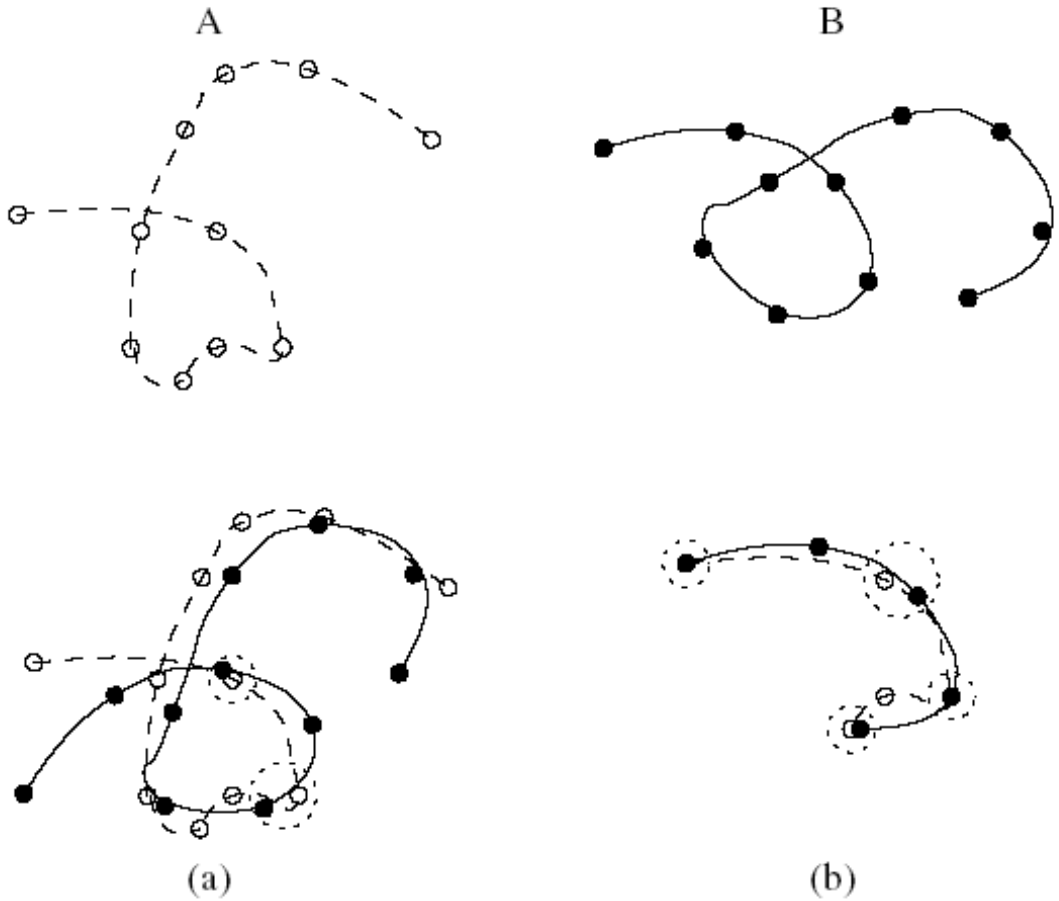
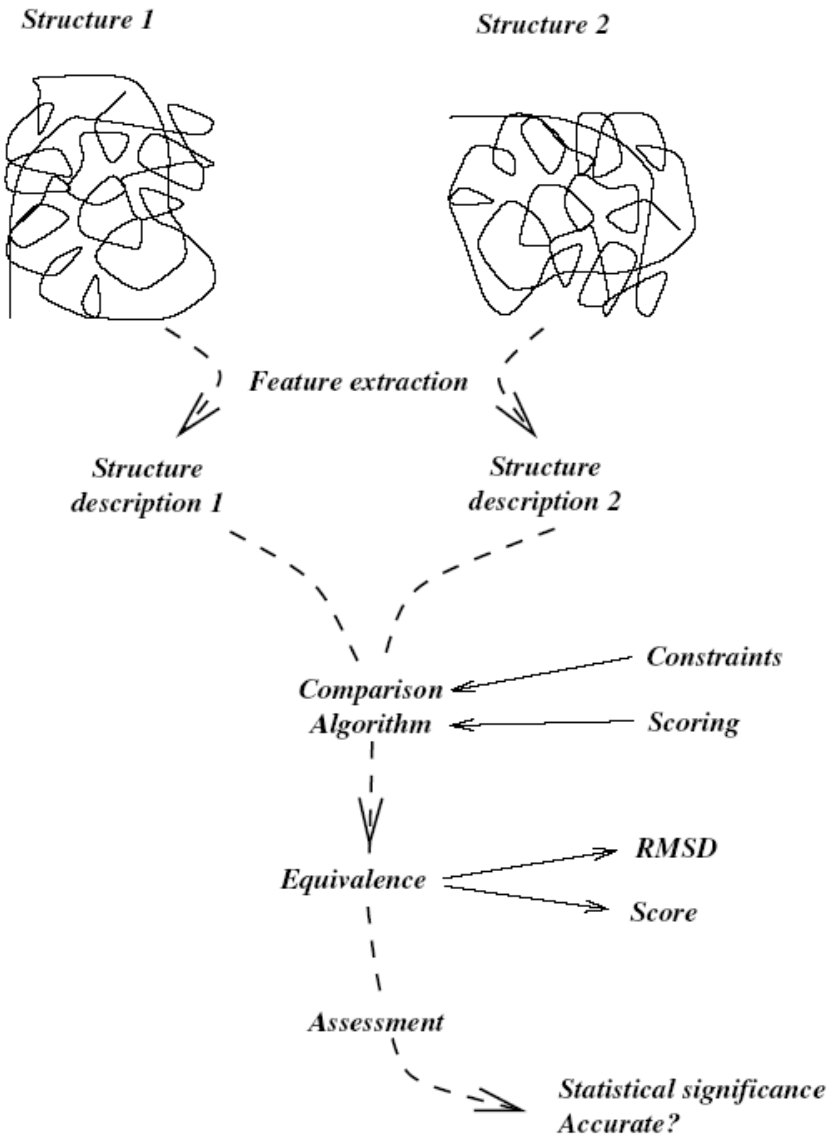


Figure 8.18 Illustration that dynamic programming cannot be used directly for structure alignment (see the text).

Framework for pairwise structure comparison

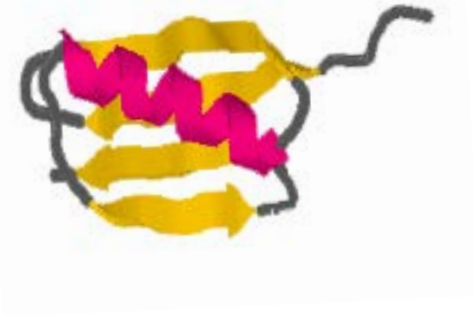


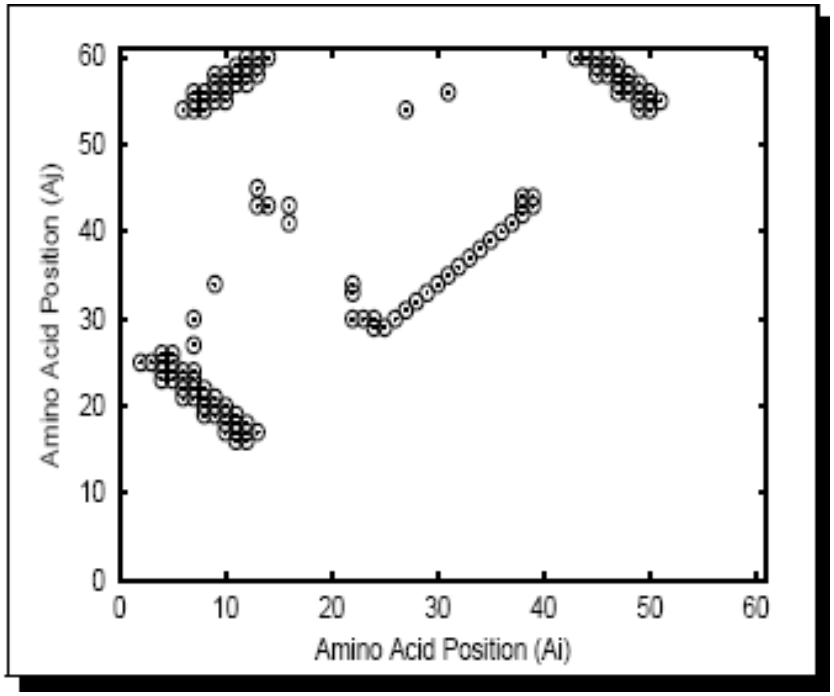
Distance Matrices

```
   123456789A123456789B123456789C123456789D123456789E123456789
 1 0469.....................................................
 2 40469....................................................
 3 640479...................................................
 4 9640479..................................................
 5 .9740469.....778.........9................................
 6 ..9740468..8.869.9.......8779.............................
 7 ...9640469857669..........989.............................
 8 ....964045547769.........9857..88..........99.............
 9 .....8640466..9............7...98..9......967.............
10 ......9540469..............89..76.97.......79.............
11 ......856404688.............98..89..........................
12 .....8546640456.............99..............................
13 ......77.9640468.............9..............................
14 ....7867..8540469............8..............................
15 ....76669.866404679.......8658.9.............................
16 ....8999....864047.........968.............................9..
17 ..............964047.......98559.........................8868.
18 .....9........774046.....9564579.....................98676..
19 ..............9.740469..8855645788..................66578..
20 ..................64046985546888.99.................85578...
21 ..................6404765678..9.78..9.....8....9.676579....
22 ..................964034579.....................9966568.....
23 ..................9730469.....................897779......
24 ..................86440469..............9.98.7.89........
25 ..................8555640469.........................988.....
26 .....8.9.........98567964046....9.........9.89.............
27 ....9798......8.955479.96404789.9...................999....
28 .....7857899..6986568...9640477669.....................
29 .....997.98998565468......7404657.....................9....
30 ..............885548......87404689..................99698..
31 ................97589.....9764046579................798....
32 .......8978...9..97........6564046679.....................
33 .......8869........897....9967864045569....88........9.....
34 ..................898......9.9564045569.9........9.97.....
35 .........9.....................7654046679.................
36 ........97.....................97554046667.88..............
37 ....................9..........9656404544768....88.........
38 ..................................966640466.................
39 ...............................97654046....................
40 ..................................9646404788................
41 ..................................9.746640468..8867.........
42 .....................................7..74047874457.........
43 ........9...........8..9.9......8..86..86404554577.........
44 .......967.........................8..88..887404669.........
45 .......979..............9.8.................8540469.........
46 ......................8.9................75640469..........
47 ..........................................8446640479.........
48 ....................9987................84599640469........
49 ......................99..............8...657..974047.........
50 ....................6678.........9..8...777...9640468.......
51 ....................7679.........................974047.......
52 ..................8657.9.........9..............64047.....
53 ................965569.8.9..97.97..............874047....
54 ................86578..8.9..99.....................740469..
55 ..............86579.....9.968.....................740469.
56 ..............8778.........9.......................64047.
57 ..............9668..........8.....................964047
58 ..............8.....................................97404
59 ..........................................................740
```

re 8.5 A distance matrix for the first 59 residues of PDB entry 1chc, where each residue presented by its C_α atom. The distances are rounded to integers, and distances larger than are represented by dots.

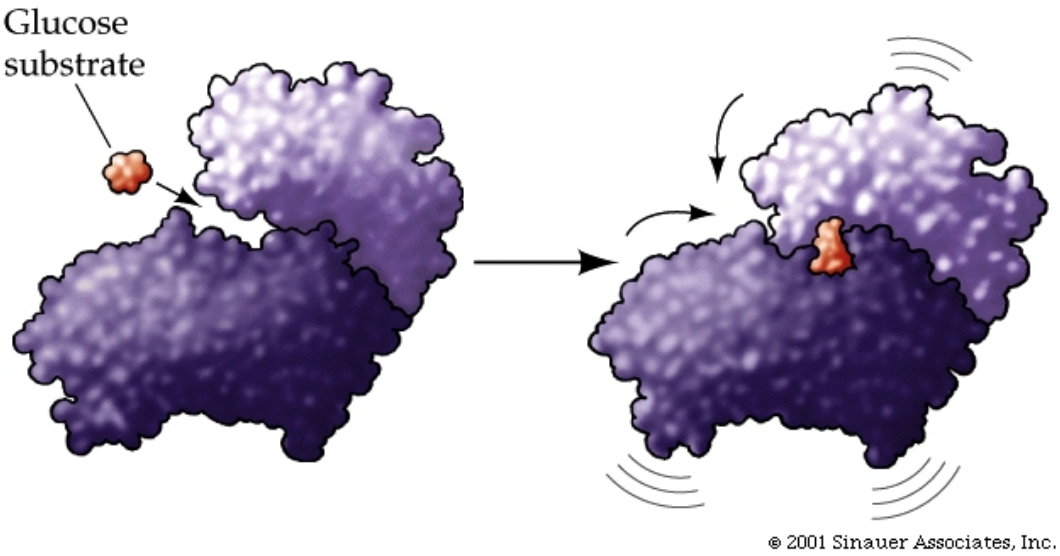
CONTACT MAP PATTERNS



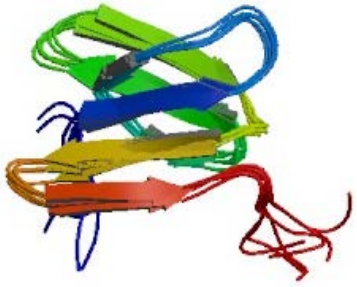


Protein Dynamics

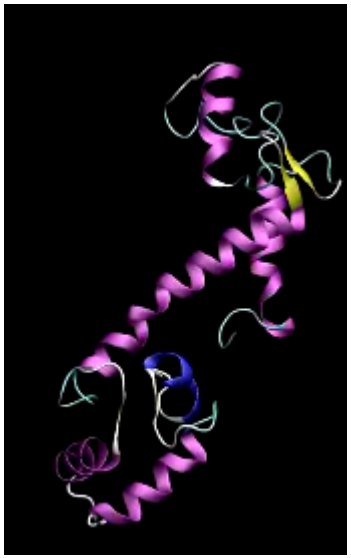
Induced fit model:



1OQK



1AHR



Molecular Dynamics Extended Library:
http://mmb.pcb.ub.es/MoDEL/ :
test searching **1e5w** & **1AHR**

How/why does a molecule move?

- Among the 3N-6 internal degrees of freedom, **bond rotations** (i.e. changes in dihedral angles) are the softest, and mainly responsible for the functional motions

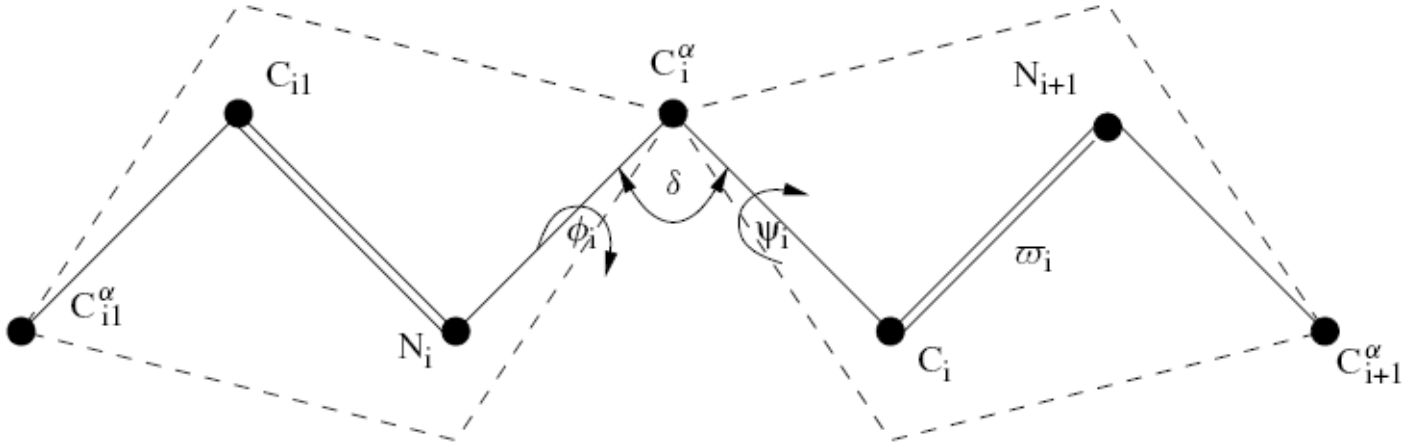
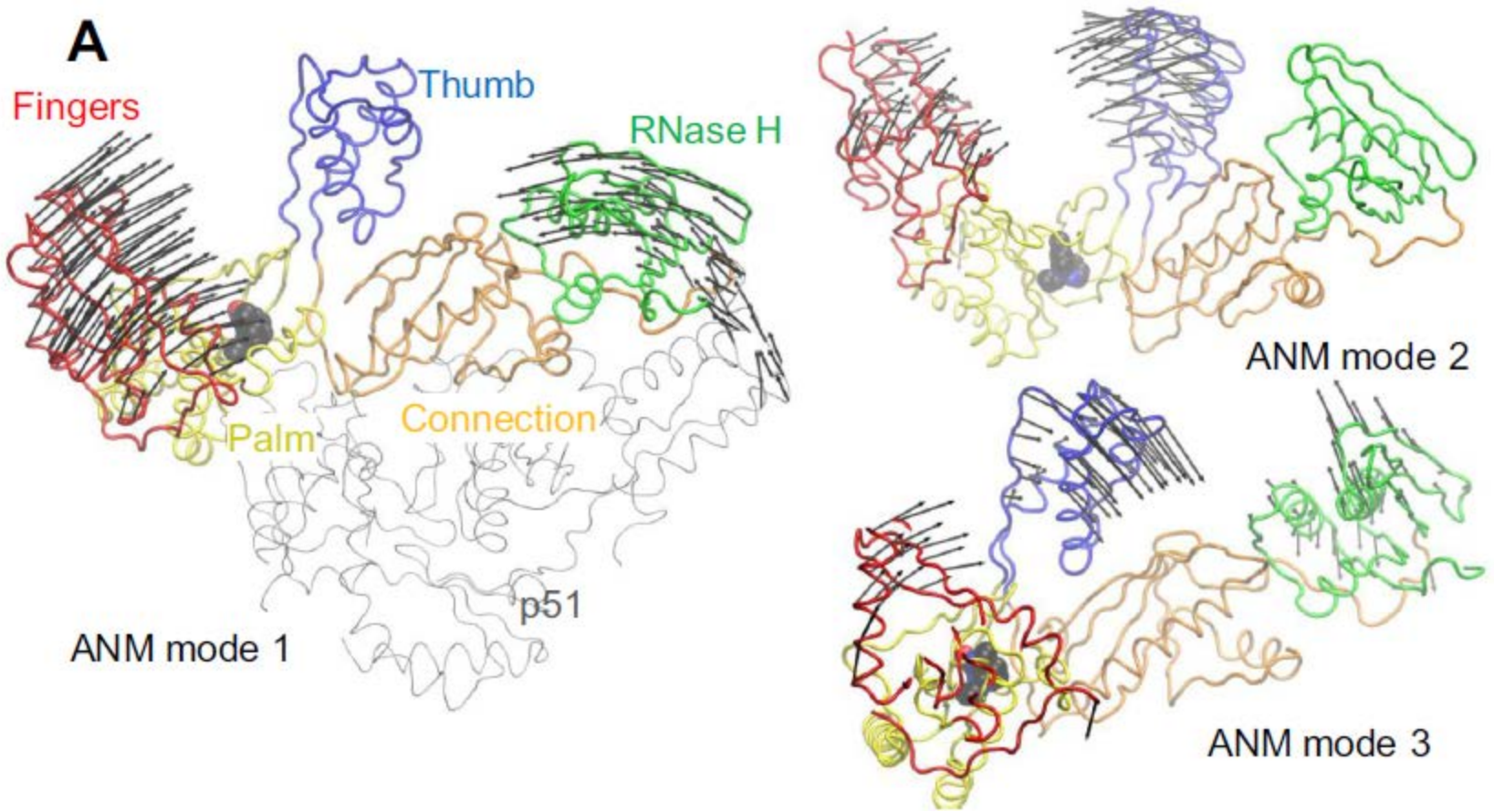


Figure 8.7 Figure illustrating the torsion angles.

The intrinsic dynamics of enzymes



Bakan, A., & Bahar, I. (2009). The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *PNAS*, *106*(34), 14349–54.

Feature extraction

Dimension Reduction

reducing the number of random variables under consideration

Find a subset of the original variables

Feature Selection

Feature Extraction

Transforms the data in the high-dimensional space to a space of fewer dimensions

Feature Extraction Methods

- **Principal component analysis**
- Semidefinite embedding
- Multifactor dimensionality reduction
- Multilinear subspace learning
- Nonlinear dimensionality reduction
- Isomap
- **Kernel PCA**
- Multilinear PCA
- Latent semantic analysis
- Partial least squares
- Independent component analysis
- Autoncoder

Reading

- Chapter 12 of PRML