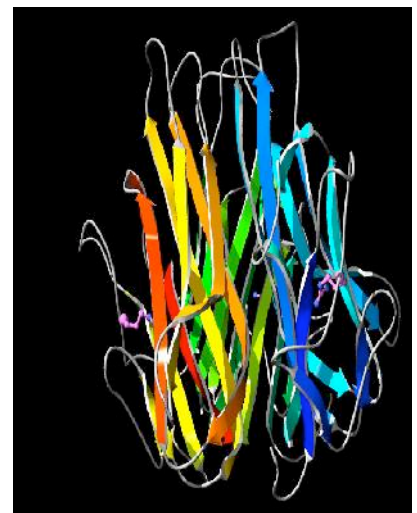
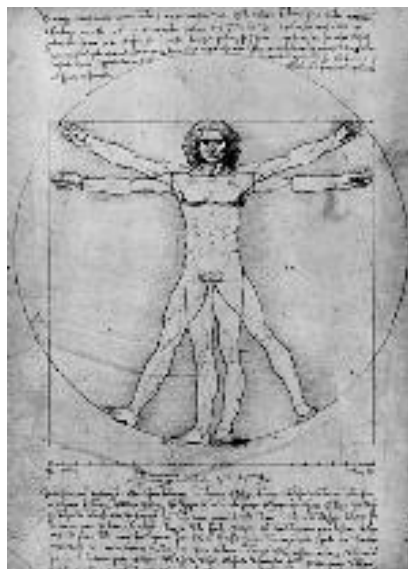
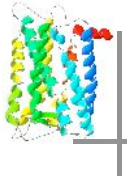


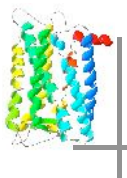
# Principles of protein structures



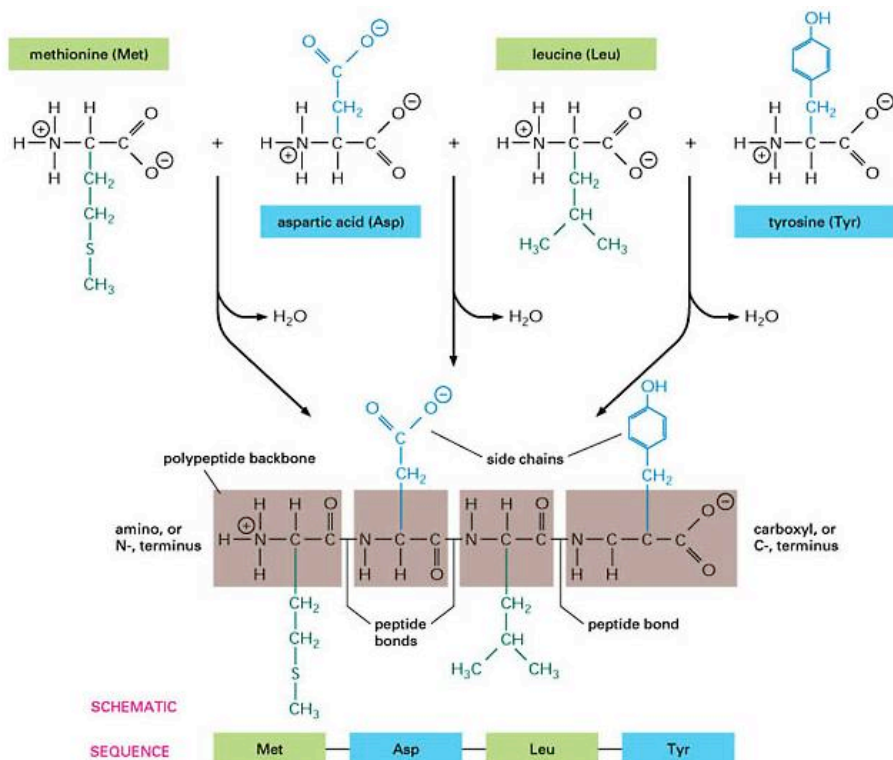


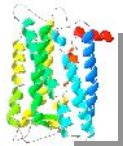
## Overview

- ❑ Amino Acids and Polypeptides (1)
- ❑ Physicochemical Properties of Amino Acid Side Chains (2)
- ❑ Protein Folding and Physicochemical Interactions (3)
- ❑ Principles of Protein Structure (4)
- ❑ Structure Comparison (5)
- ❑ Protein Structure Databases (6)



## Amino Acids and Polypeptides (1)

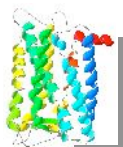




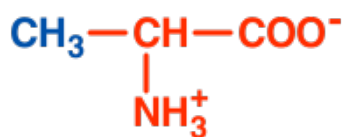
# Amino Acids

Three and one letter code:

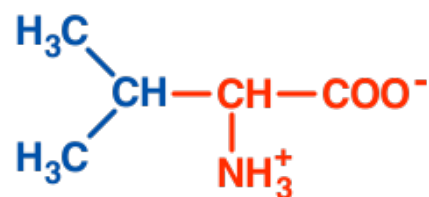
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V



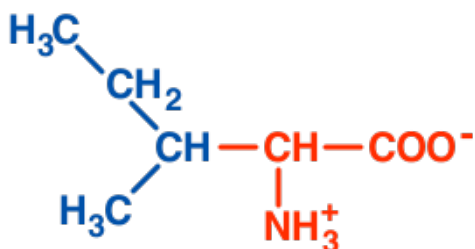
## Amino Acids with aliphatic Side-Chains



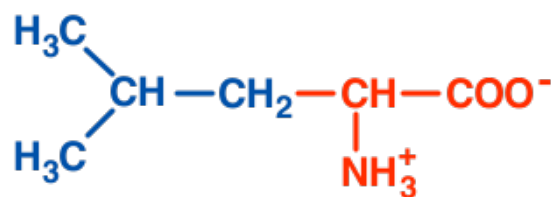
**Ala (A)**



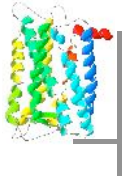
**Val (V)**



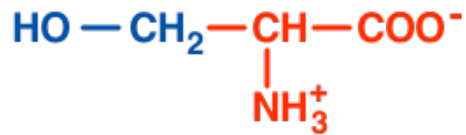
**Ile (I)**



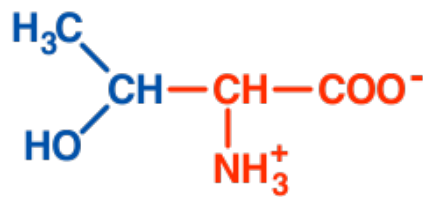
**Leu (L)**



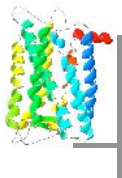
## Sidechains with hydroxyl (-OH) groups



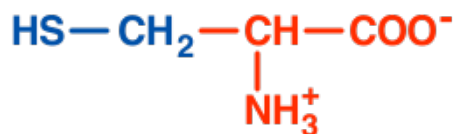
**Ser (S)**  
**pK<sub>a</sub>=13**



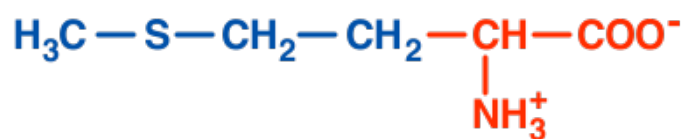
**Thr (T)**  
**pK<sub>a</sub>=13**



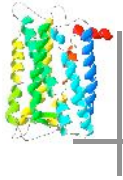
## Sidechains containing sulphur



**Cys (C)**  
**pK<sub>a</sub>=8.3**

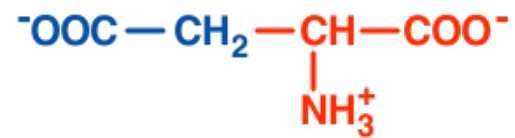


**Met (M)**

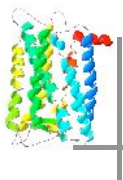
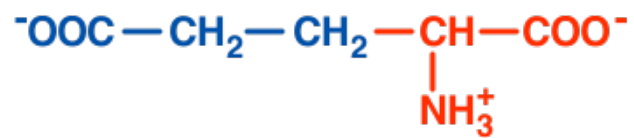


## Acidic amino acids

**Asp (D)**  
**pK<sub>a</sub>=3.9**

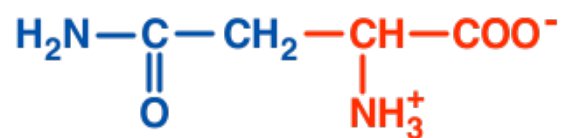


**Glu (E)**  
**pK<sub>a</sub>=4.1**

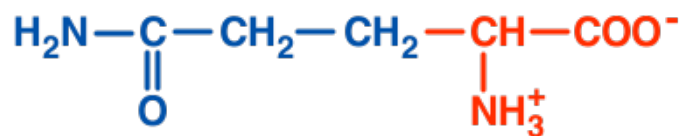


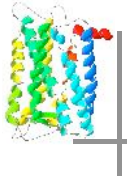
## Amides of acidic amino acids

**Asn (N)**



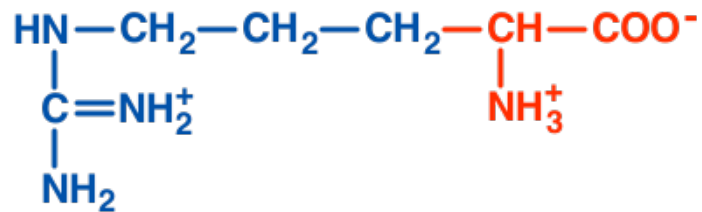
**Gln (Q)**



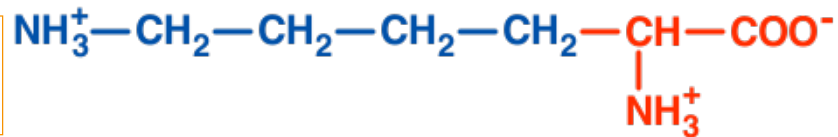


## Basic Amino Acids

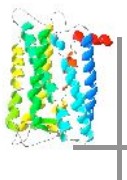
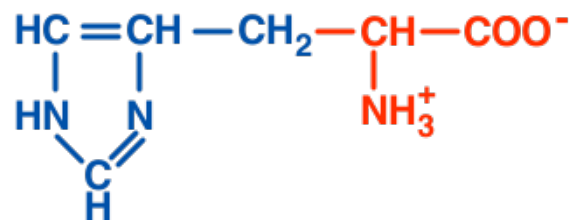
**Arg (R)**  
**pK<sub>a</sub>=12.5**



**Lys (K)**  
**pK<sub>a</sub>=10.8**

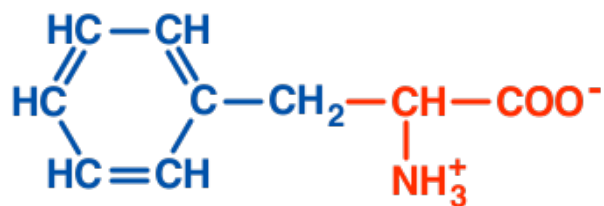


**His (H)**  
**pK<sub>a</sub>=6.0**

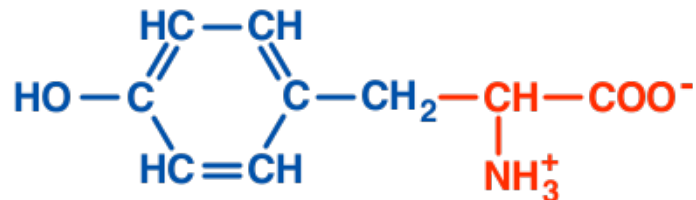


## Side-chains with aromatic rings

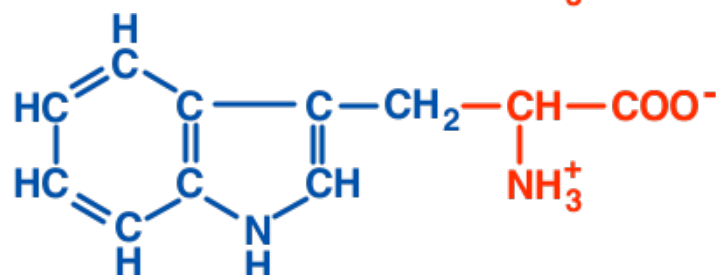
**Phe (F)**



**Tyr (Y)**  
**pK<sub>a</sub>=10.1**



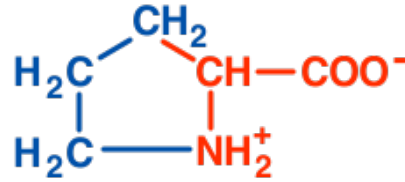
**Trp (W)**





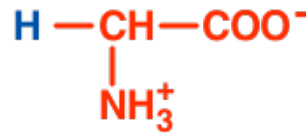
## Special cases ...

**Pro (P)**

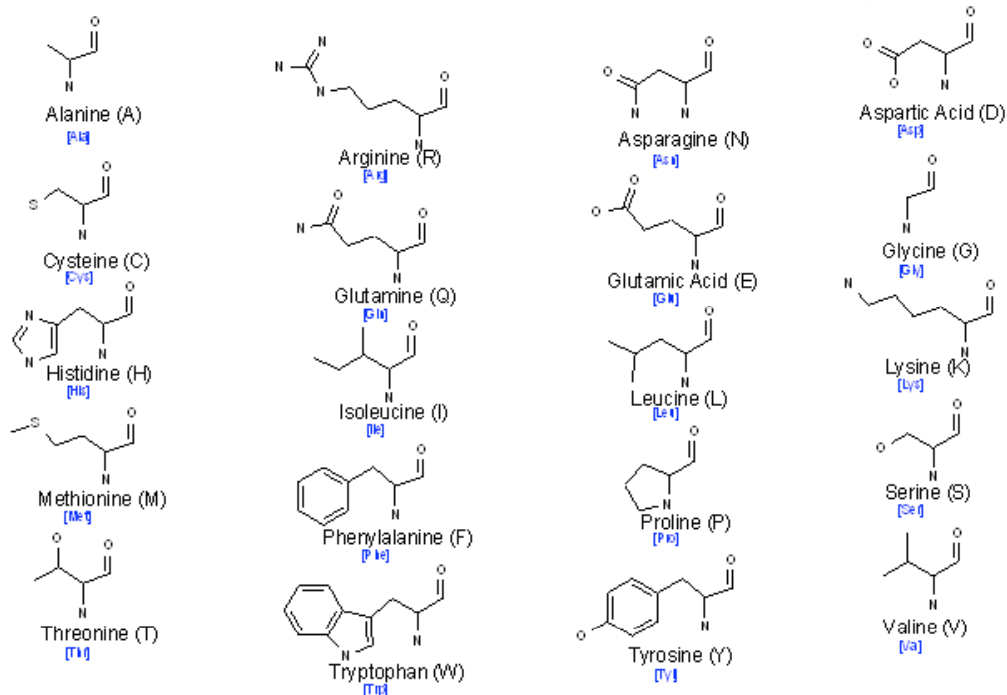


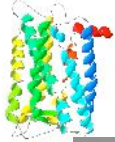
**Imino acid**

**Gly (G)**



## Side Chain Structures





## Physicochemical Properties of Amino Acid Side Chains (2)

### Neutral Hydrophobic

Alanine  
Valine  
Leucine  
Isoleucine  
Proline  
Tryptophane  
Phenylalanine  
Methionine

### Neutral Polar

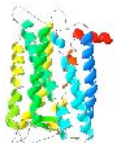
Glycine  
Serine  
Threonine  
Tyrosine  
Cysteine  
Asparagine  
Glutamine

### Basic

Lysin  
Arginine  
(Histidine)

### Acidic

Aspartic Acid  
Glutamic Acid



## pH and pKa

### ☐ pH

$$pH = -\log [H^+]$$

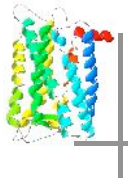
### ☐ Water ion product

$$K_w = [H^+] [OH^-] = 10^{-14}$$

$$\log[H^+] + \log[OH^-] = \log(10^{-14})$$

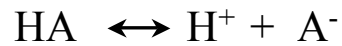
$$pH + pOH = 14$$





## pH and pKa

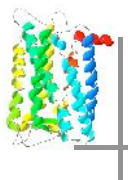
### □ Dissociation of weak acids



$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]} \quad [\text{H}^+] = K_a \frac{[\text{HA}]}{[\text{A}^-]}$$

$$\log [\text{H}^+] = \log K_a + \log \frac{[\text{HA}]}{[\text{A}^-]}$$

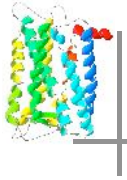
$$-\log [\text{H}^+] = -\log K_a + \log \frac{[\text{A}^-]}{[\text{HA}]}$$



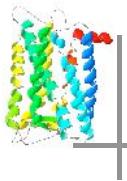
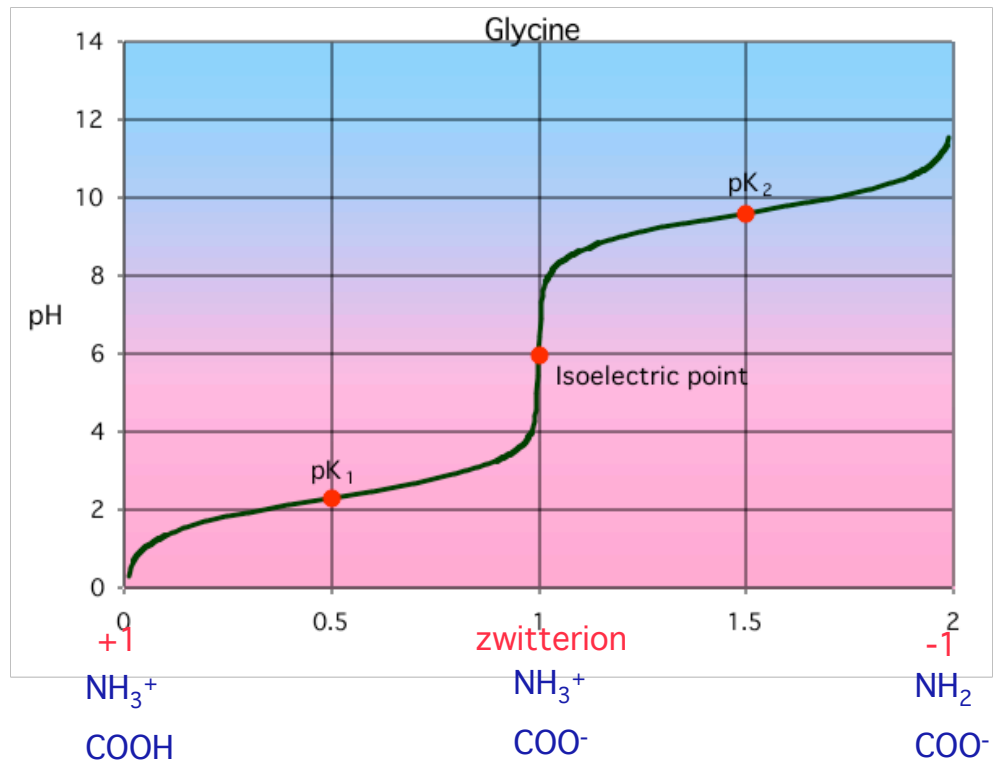
## pH and pKa

### □ Henderson - Hasselbach Equation

$$pH = pKa + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

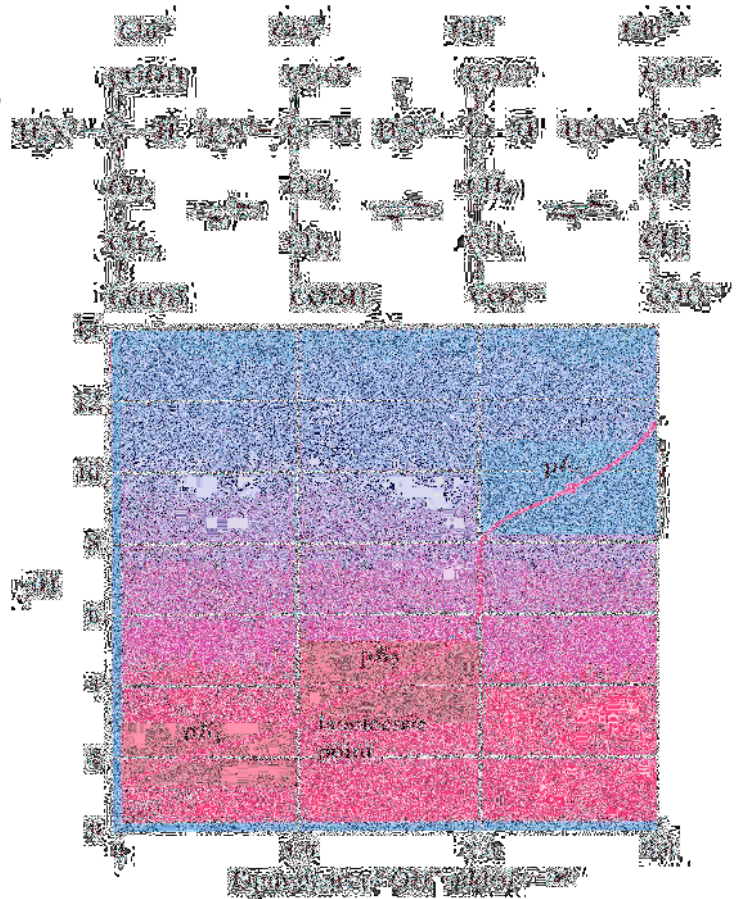


## pH and pKa



## pH and pKa

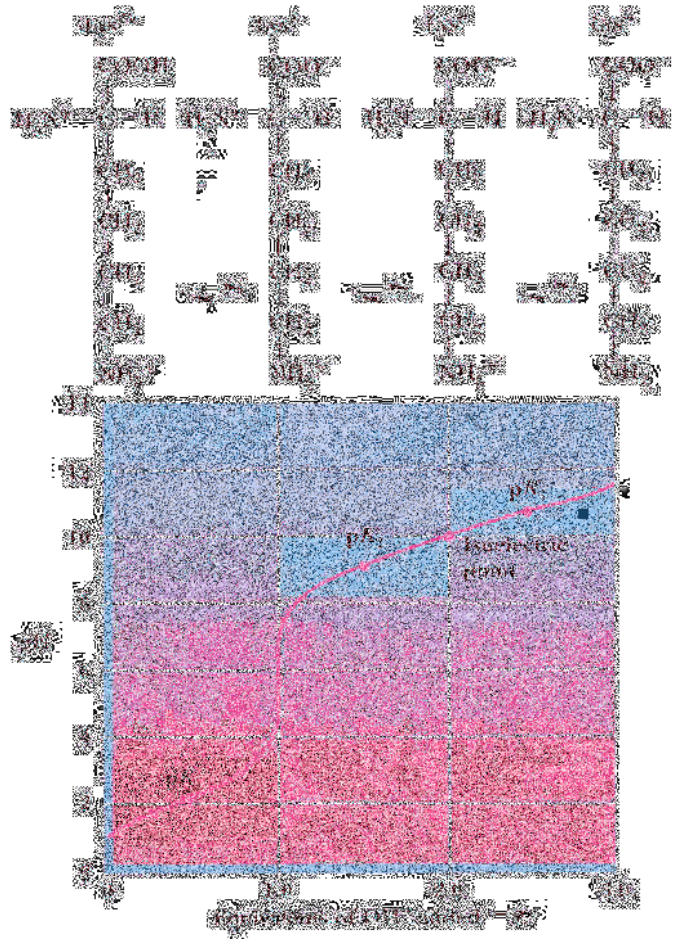
□ Glu





## pH and pKa

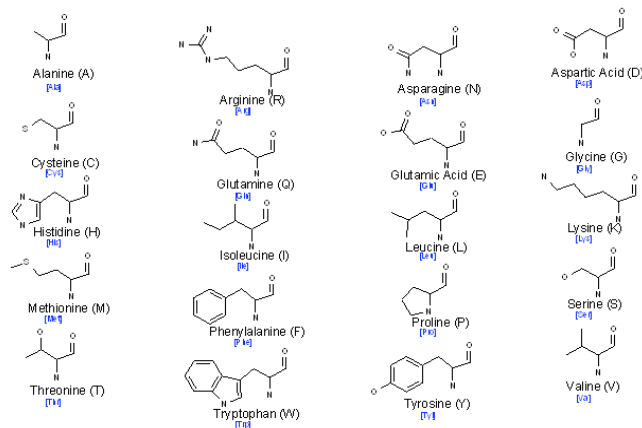
□ Lys

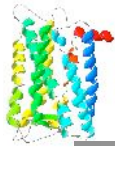


## pH and pKa

□ Enzymatic reactions often require proton transfer.

➤ Q: Which amino-acid(s) are able to change their protonation state under physiological conditions?

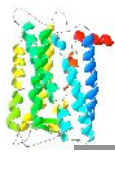
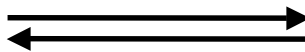




# Protein Folding and Physicochemical Interactions (3)

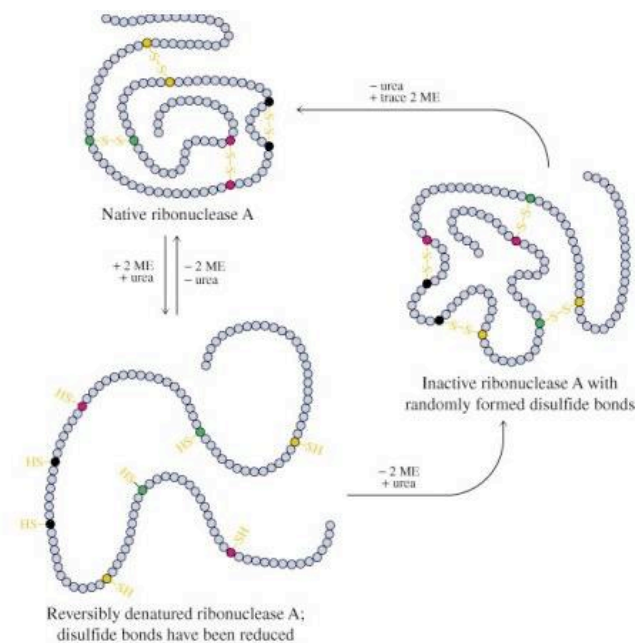
## Why do proteins fold?

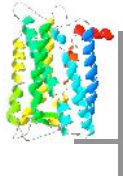
MNIFEMLRID EGLRLKIYKD TEGYYTIGIG  
 HLLTKSPSLN AAKSELDKAI GRNCNGVITK  
 DEAEKLFNQD VDAAVRGILR NAKLKPVYDS  
 LDAVRRCALI NMVFQMGETG VAGFTNSLRM  
 LQQKRWDEAA VNLAKSRWYN QTPNRAKRVI  
 TTFRTGTWDA YKNL



## Anfinson's paradigm

1957, Nobel Prize 1972

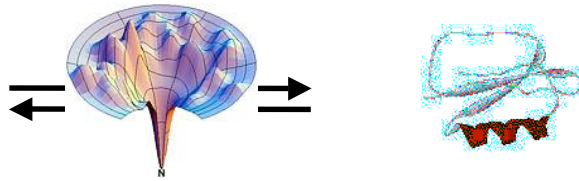




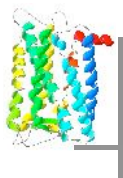
# Anfinson's paradigm

```

MNI FEMLRID  EGLRLKIYKD  TEGYYTIGIG
HLLTKSPSLN  AAKSELDKAI  GRNCNGVITK
DEAEKLFNQD  VDAAVRGILR  NAKLKPVYDS
LDAVRRCALI  NMVFQMGETG  VAGFTNSLRM
LQQKRWDEAA  VNLAKSRWYN  QTPNRAKRV
TTFRTGTWDA  YKNL
  
```



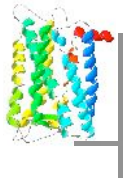
**All the necessary information for the 3-dimensional structure of an enzyme is contained in the primary structure or sequence of the amino acids.**



# Levinthal's Paradox (1968)

If a chain of a hundred amino acids is considered and it assumed each amino acid can exist in one of three conformations, extended, helical or loop, then there are  $3^{100}$  possible ways to arrange this chain.

This is roughly  $10^{48}$  conformations. Bond rotation can be estimated to occur at a rate of roughly  $10^{14} \text{ s}^{-1}$ . This means that search for the right conformation through random searching alone would take the order of  $10^{34} \text{ s}$  or  $10^{26} \text{ years}$ , several orders of magnitudes greater than the age of the universe!

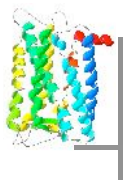


MNIFEMLRID EGLRLKIYKD TEGYYTIGIG  
 HLLTKSPSLN AAKSELDKAI GRNCNGVITK  
 DEAEKLFNQD VDAVRGILR NAKLKPVDYS  
 LDAVRRCALI NMVFQMGETG VAGFTNSLRM  
 LQQKRWDEAA VNLAKSRWYN QTPNRAKRV  
 TTFRTGTWDA YKNL

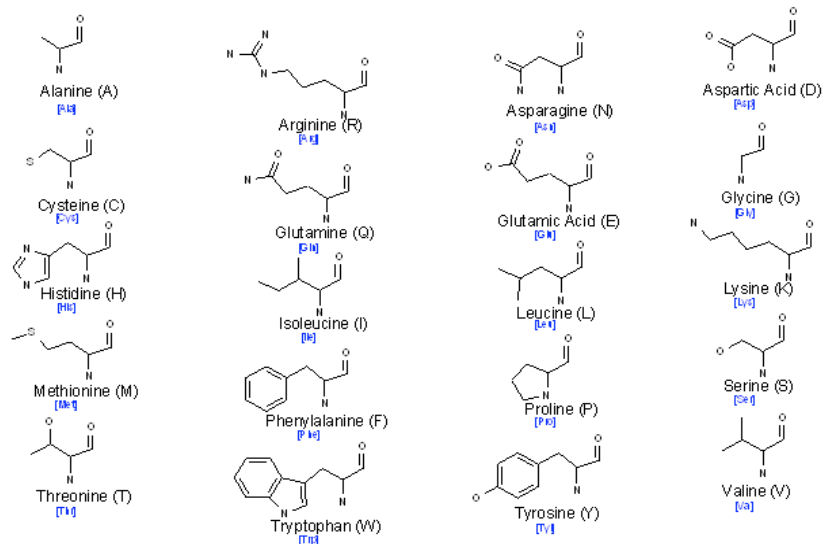


- Many proteins fold spontaneously to their native structure
- Protein folding is relatively fast (nsec – sec)
- Chaperones speed up folding, but do not alter the structure

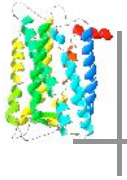
The protein sequence contains all information needed to create a correctly folded protein.



## Why do proteins fold?







## Side Chain Properties

### Neutral Hydrophobic

Alanine  
Valine  
Leucine  
Isoleucine  
Proline  
Tryptophane  
Phenylalanine  
Methionine

### Neutral Polar

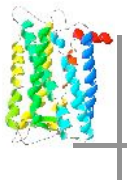
Glycine  
Serine  
Threonine  
Tyrosine  
Cysteine  
Asparagine  
Glutamine

### Basic

Lysin  
Arginine  
(Histidine)

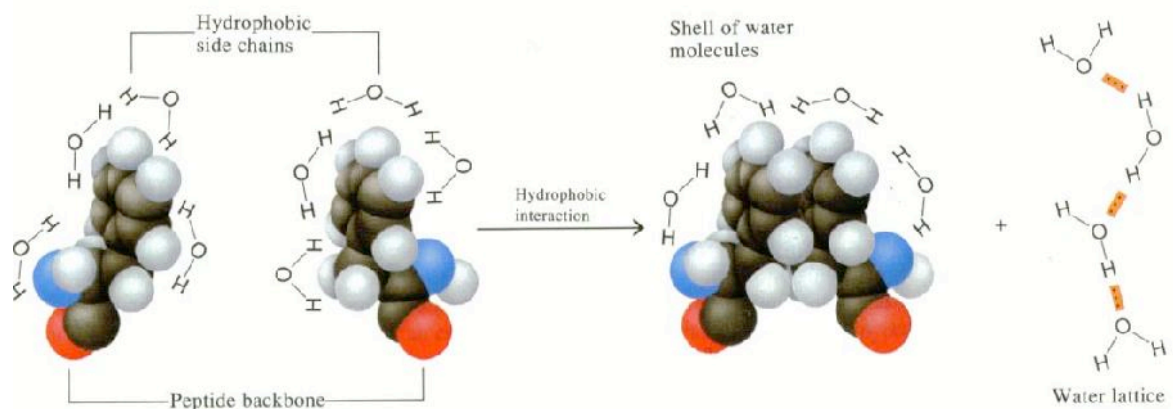
### Acidic

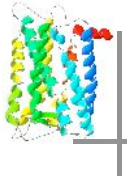
Aspartic Acid  
Glutamic Acid



## Hydrophobic Effects

- main driving force for protein folding

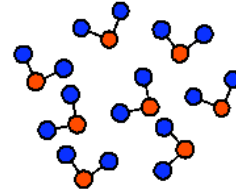




# Hydrophobic Effects

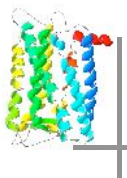
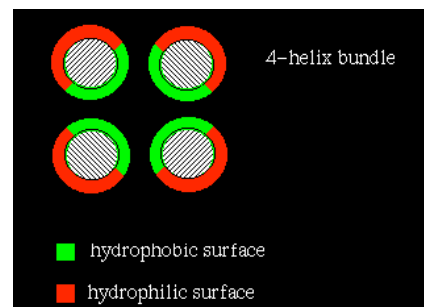
- main driving force for protein folding

Water molecules in bulk water are mobile and can form H-bonds in all directions.



Hydrophobic surfaces don't form H-bonds. The surrounding water molecules have to orient and become more ordered.

The entropy loss can be minimized by gathering the hydrophobic surfaces together in the core of a protein and separating them from the solvent.



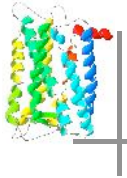
## 1D-Structure prediction

- Projection onto strings of structural assignments

➤ E.g. "Solvent Accessibility"

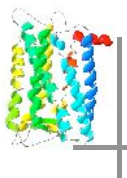
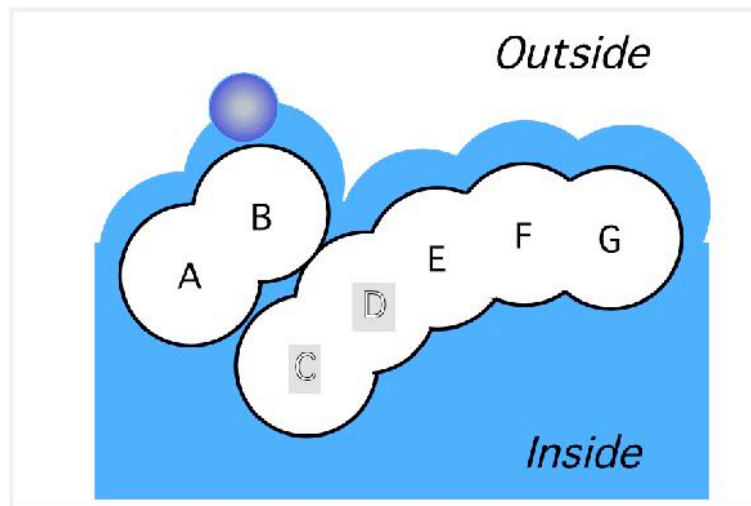
<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G...</b>
<b>e</b>	<b>e</b>	<b>b</b>	<b>b</b>	<b>e</b>	<b>e</b>	<b>e...</b>





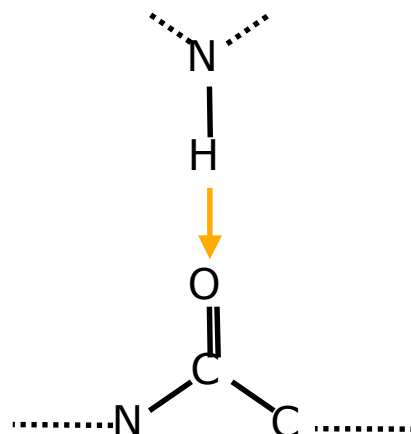
## Surface Definitions

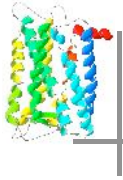
- ☐ Van der Waals Radius
- ☐ Molecular Surface
- ☐ Solvent Accessible Surface



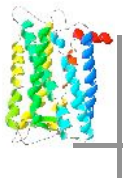
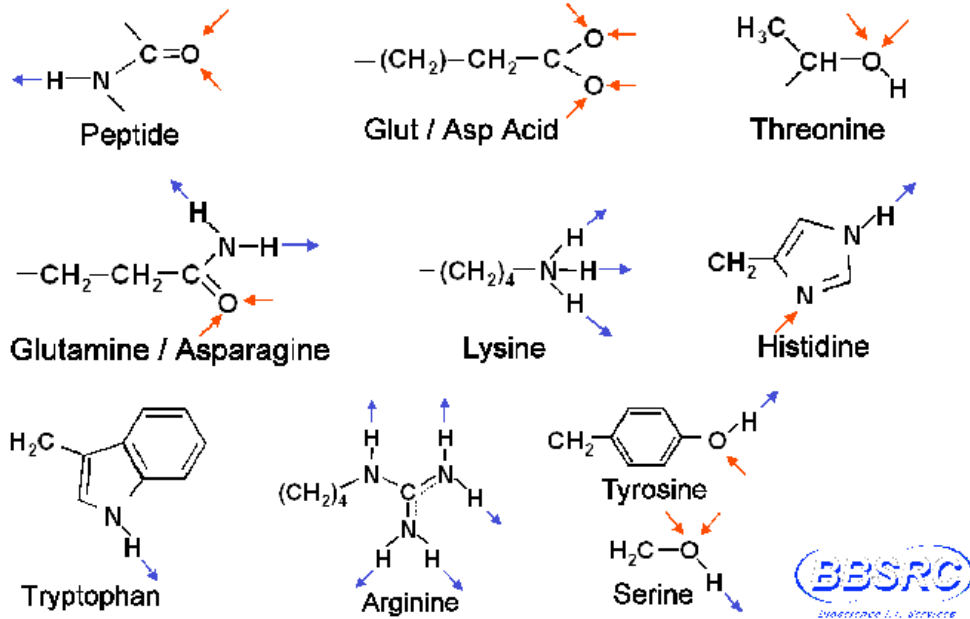
## Hydrogen Bonds

- ☐ H-atoms bound to electronegative atoms (e.g. N, O) are polarized and can form H-bonds
- ☐ H-bonding partners include:
  - main chain atoms
  - side chain atoms
  - water molecules
  - ligands, etc...





## Hydrogen bonding amino acids



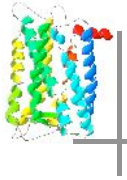
## Energetics of protein folding

$$\Delta G = \Delta H - T\Delta S$$

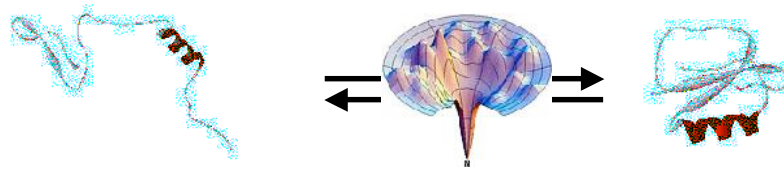
### Energetics of protein folding

H-bonds  
 hydrophobic effects  
 salt bridges  
 SS - bonds  
 loss of solvation  
 entropy change  
 dispersion / VdW contacts  
 conformational energy

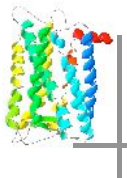
- ❑ Difference of two very large energetic terms
- ❑ Low overall stabilization energy



## Why do proteins fold?

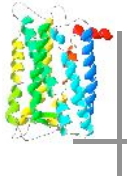


- ☐ Change of energy state from unfolded to folded
- ☐ Folded state must have overall lower energy
- ☐ Let's assume the folded state is the lowest possible state for this polypeptide



## Protein Sequence Space

- ☐ How many different proteins are theoretically possible?
- ☐ How many of these have been tested during evolution?



## Protein sequence space

- Assuming a peptide of length 100 aa

Possible combinations:  $n_c = 20^{100} \approx 1.27 * 10^{130}$

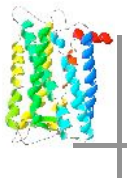
Volume of one peptide:

$$r_{\text{atom}} \approx 2 \text{ \AA}$$

$$v_{\text{atom}} \approx 35 \text{ \AA}^3$$

$$\text{packing} \approx 75\%$$

$$v_{\text{peptide}} \approx 1.3 * 10^5 \text{ \AA}^3$$



## Protein sequence space

- $1.27 * 10^{130}$  combinations. For comparison ...

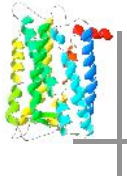
Volume of the Earth:

$$R \approx 6.4 * 10^3 \text{ km} \approx 6.4 * 10^{16} \text{ \AA}$$

$$V = \frac{4}{3} \pi R^3 \approx 1.1 * 10^{51} \text{ \AA}^3$$

Peptides/Earth:

$$n_p \approx \frac{1.1 * 10^{51}}{1.3 * 10^5} \approx 7.7 * 10^{45}$$



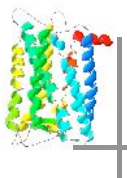
## Protein sequence space

- 1.27\*10<sup>130</sup> combinations. For comparison ...

Age of the Earth:  $3 * 10^9$  years  $\approx 2.6 * 10^{13}$  hours  
 $\approx 9.5 * 10^{16}$  sec  
 $\approx 9.5 * 10^{28}$  psec

- If the whole planet consisted of peptides, and peptides were renewed every psec...

$$n_T \approx (7.7 * 10^{45}) * (9.5 * 10^{28}) \approx 7.3 * 10^{74}$$



## Protein sequence space

- Assuming a peptide of length 100 aa

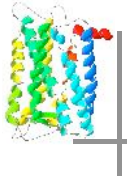
Possible combinations:  $n_c = 20^{100} \approx 1.27 * 10^{130}$

- If the whole planet consisted of peptides, and peptides were renewed every psec...

$$n_T \approx (7.7 * 10^{45}) * (9.5 * 10^{28}) \approx 7.3 * 10^{74}$$

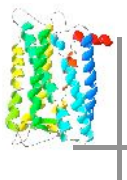
$$10^{130} - 10^{75} \approx 10^{130}$$

**?!?**

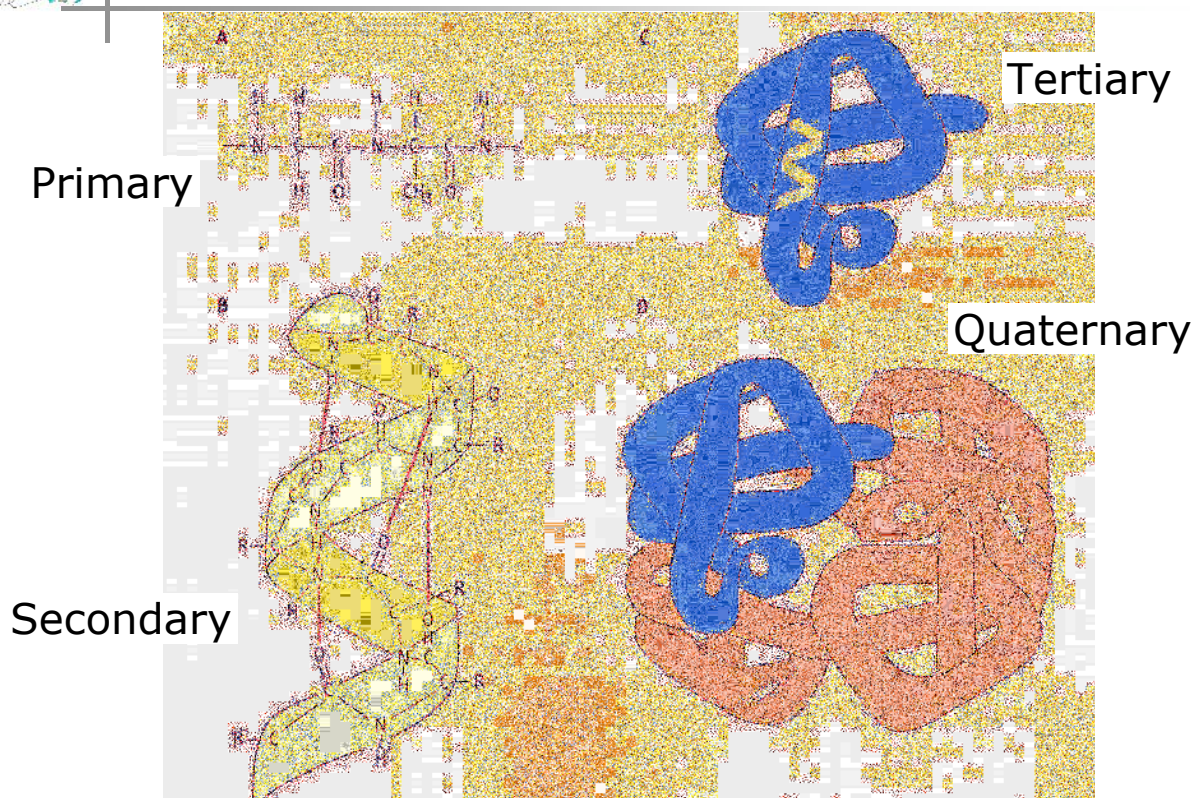


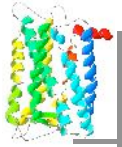
## Principles of Protein Structure (4)

- Primary Structure
- Secondary Structure
- Tertiary Structure
- Quaternary Structure

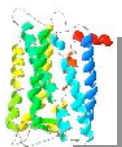
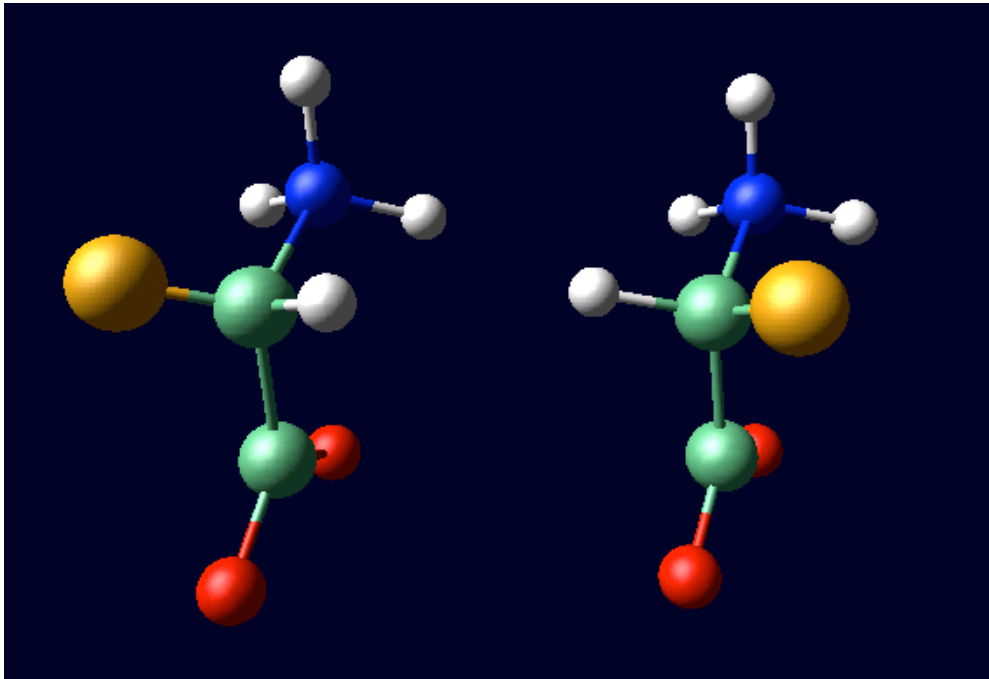


## Principles of protein structure

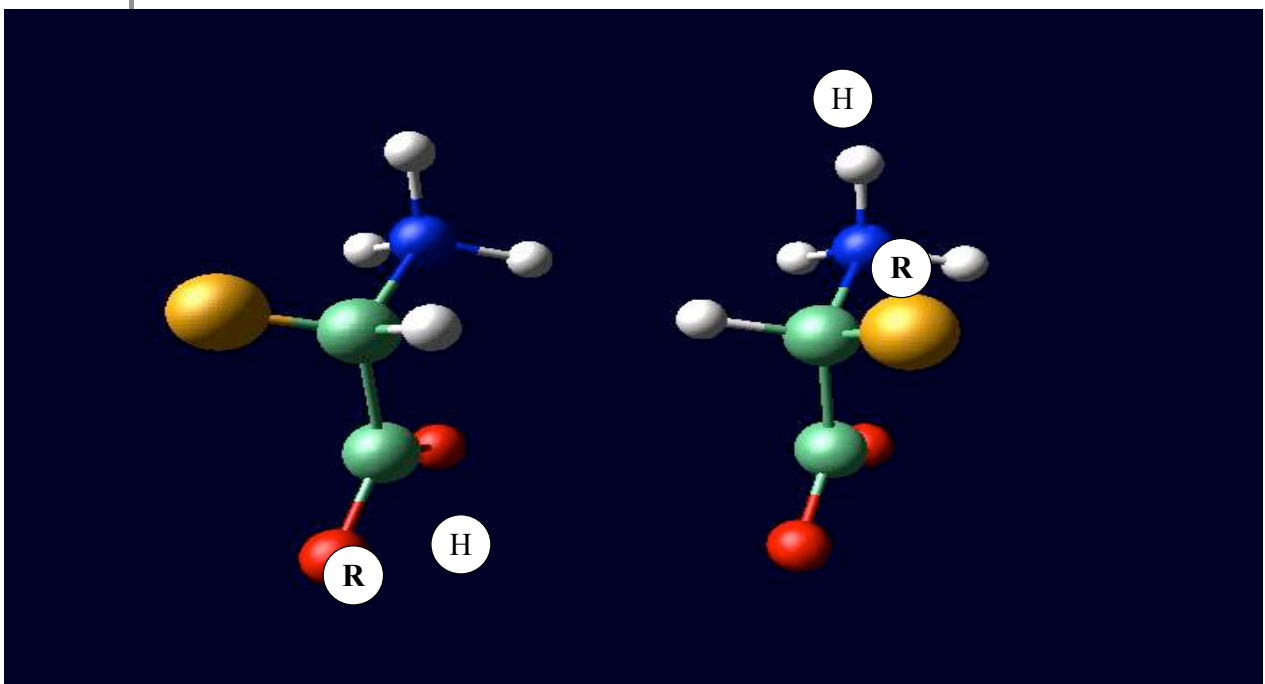




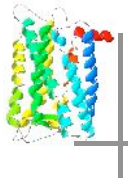
## Stereochemistry: L- and D-amino acids



## Geometry of a peptide bond



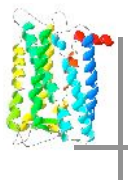
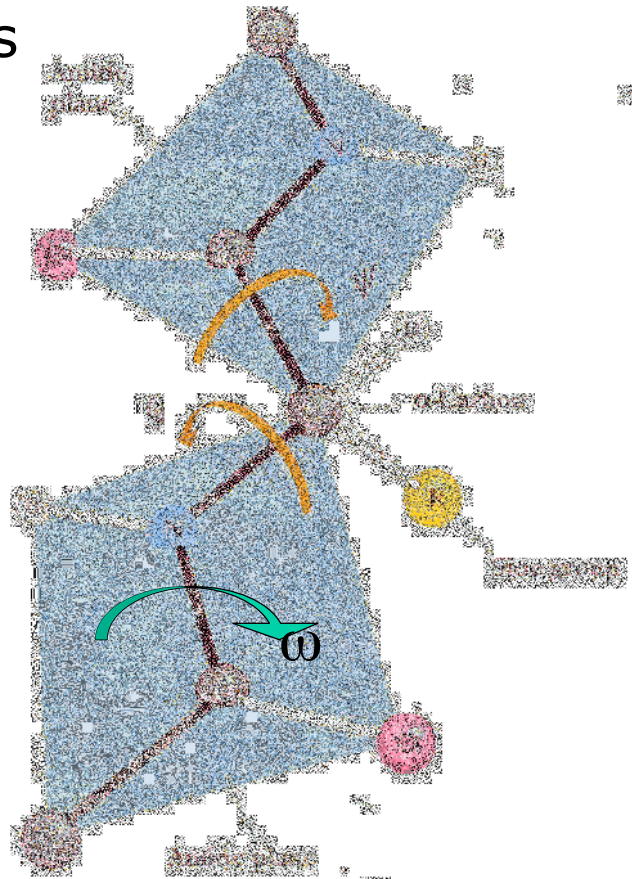




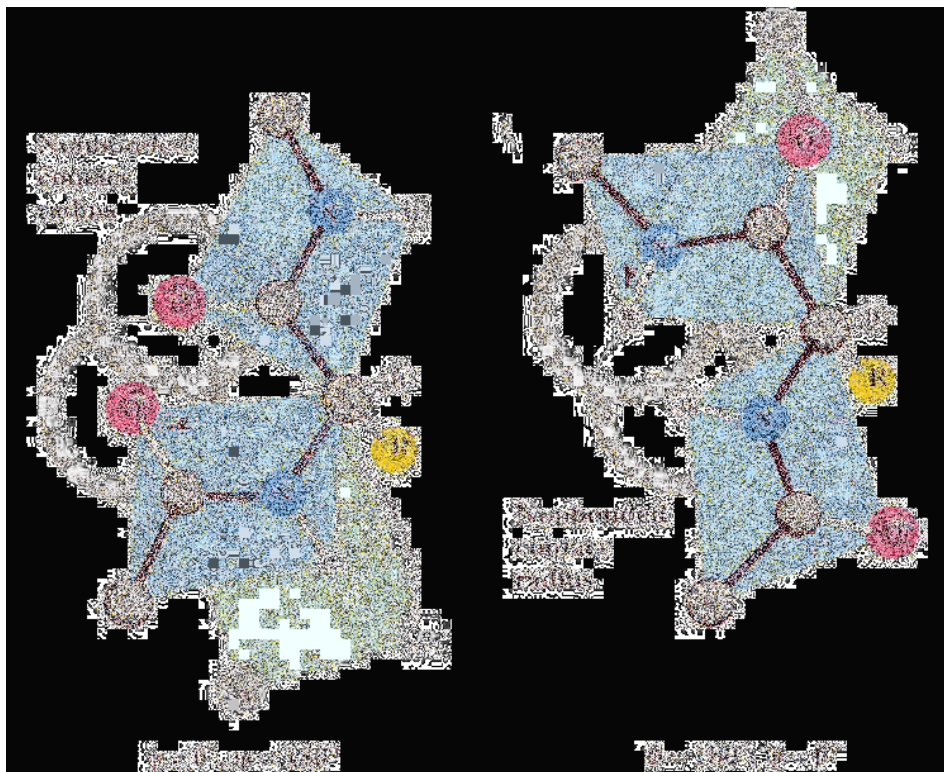
## Dihedral angles

$\Phi$ ,  $\Psi$ , and  $\omega$

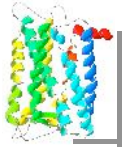
Q: Which values would you expect for  $\omega$ ?



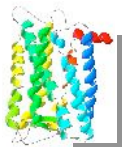
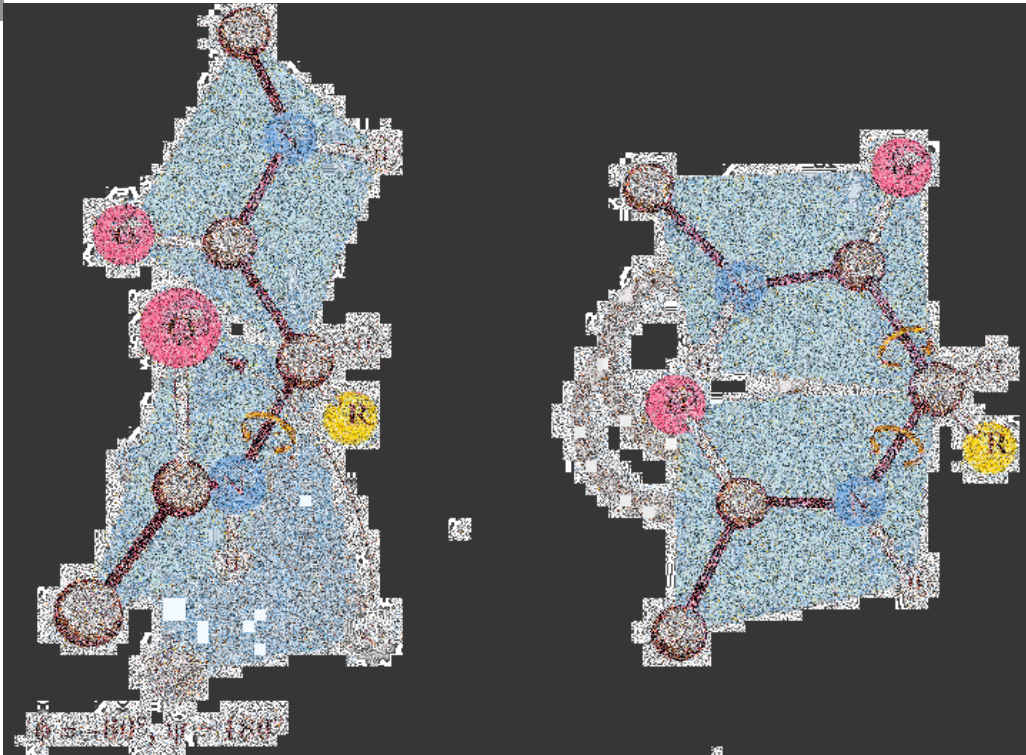
## Dihedral angles $\Phi$ and $\Psi$



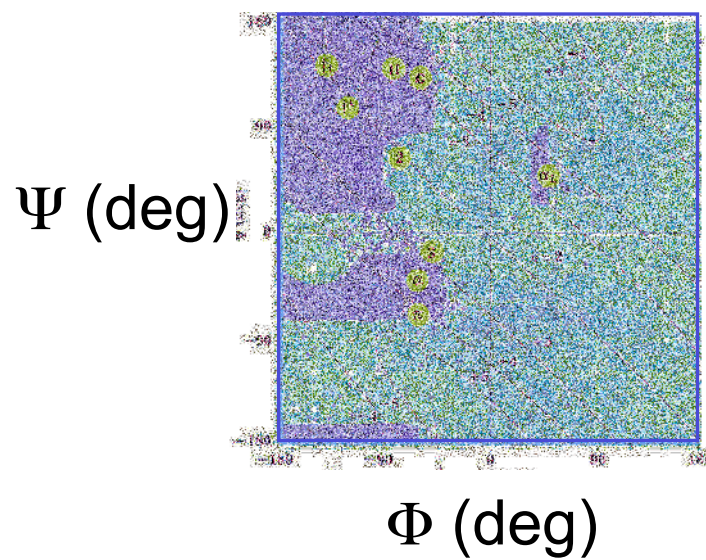


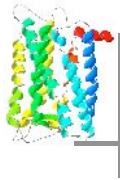


## Dihedral angles $\Phi$ and $\Psi$



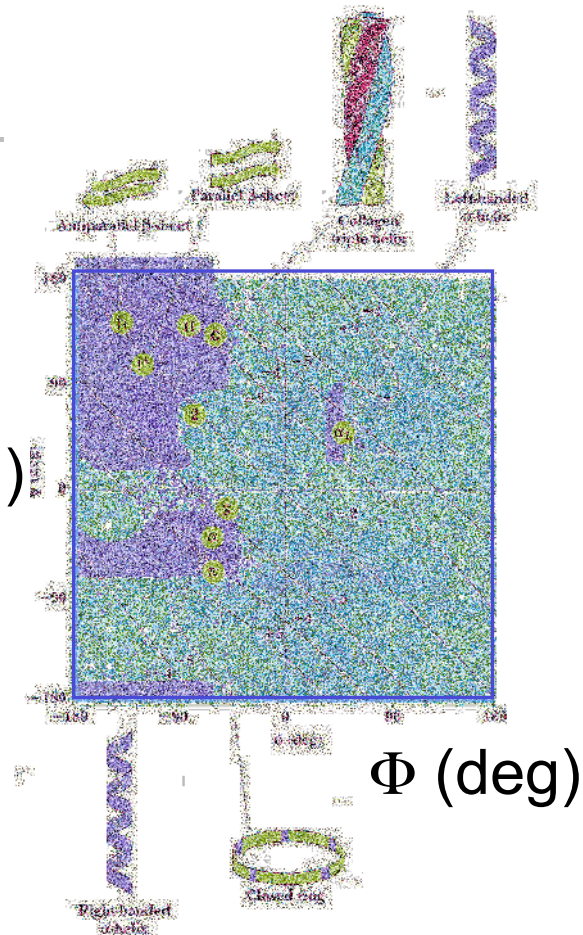
## Ramachandran Plots



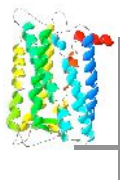


## Ramachandran Plots

$\Psi$  (deg)

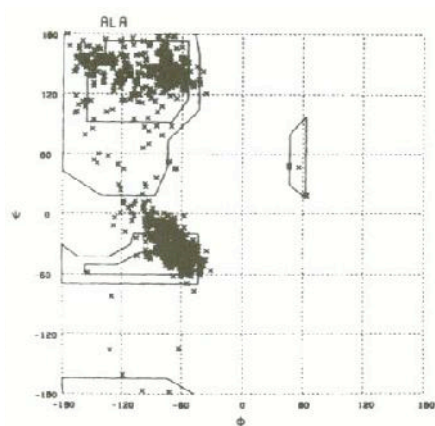


$\Phi$  (deg)

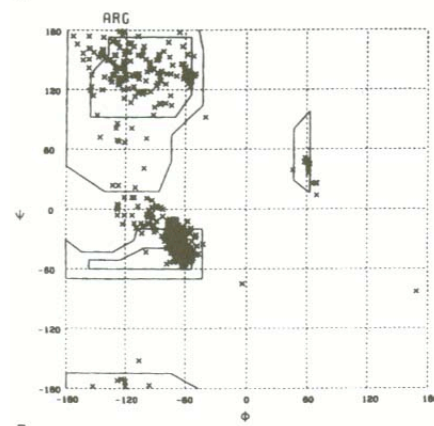


## Amino acid preferences

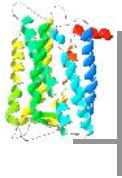
### Alanine and Arginine



**ALA**

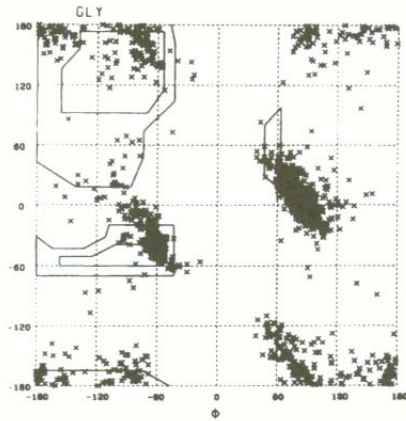


**ARG**

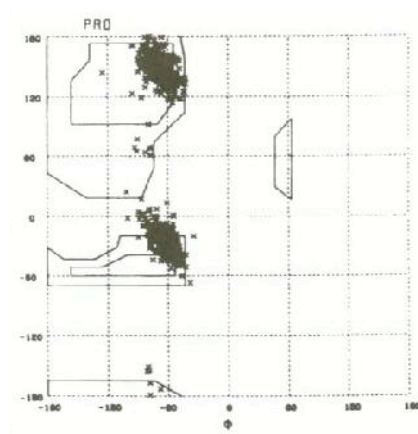


## Amino acid preferences

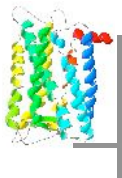
- Amino acid with special preferences:



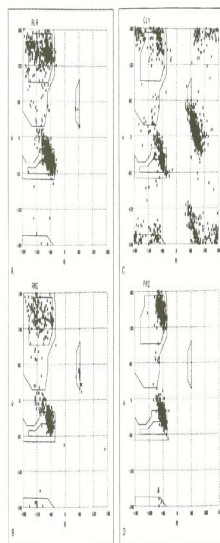
**GLY**



**PRO**



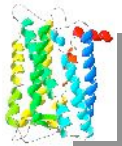
## Alpha helices



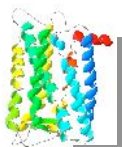
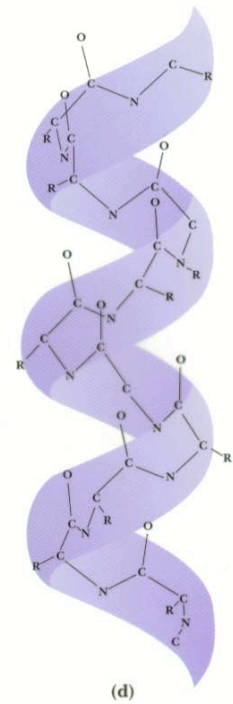
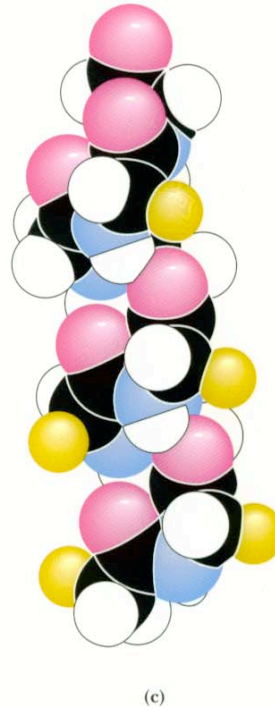
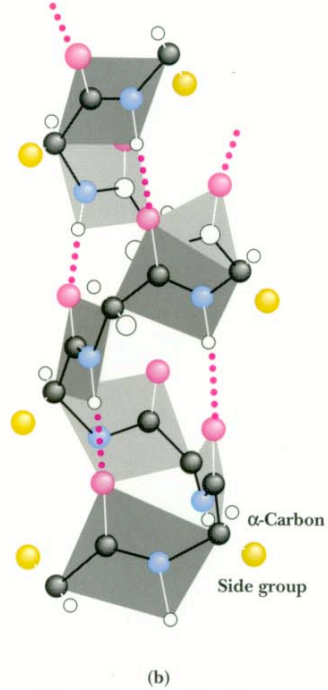
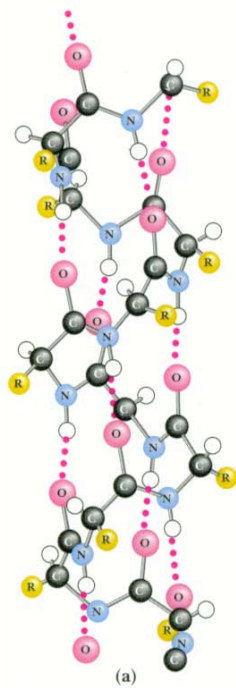
Right-handed (clockwise)  
helical conformation



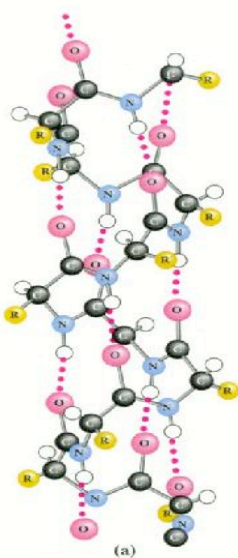




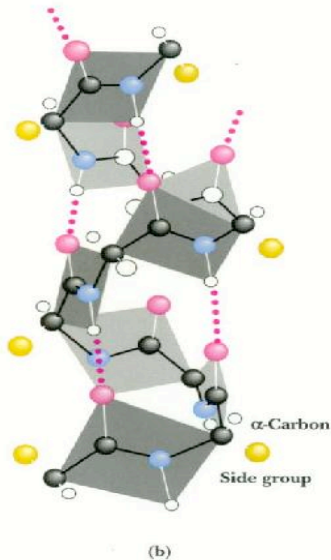
## Alpha helices



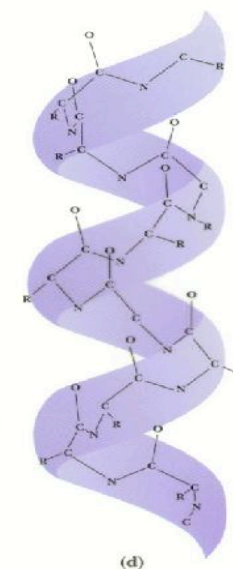
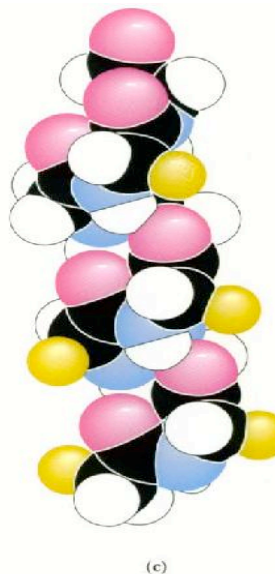
## Beta strands / beta sheets

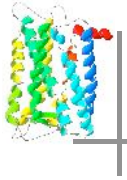


Hydrogen bonds stabilize the helix structure.

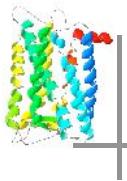
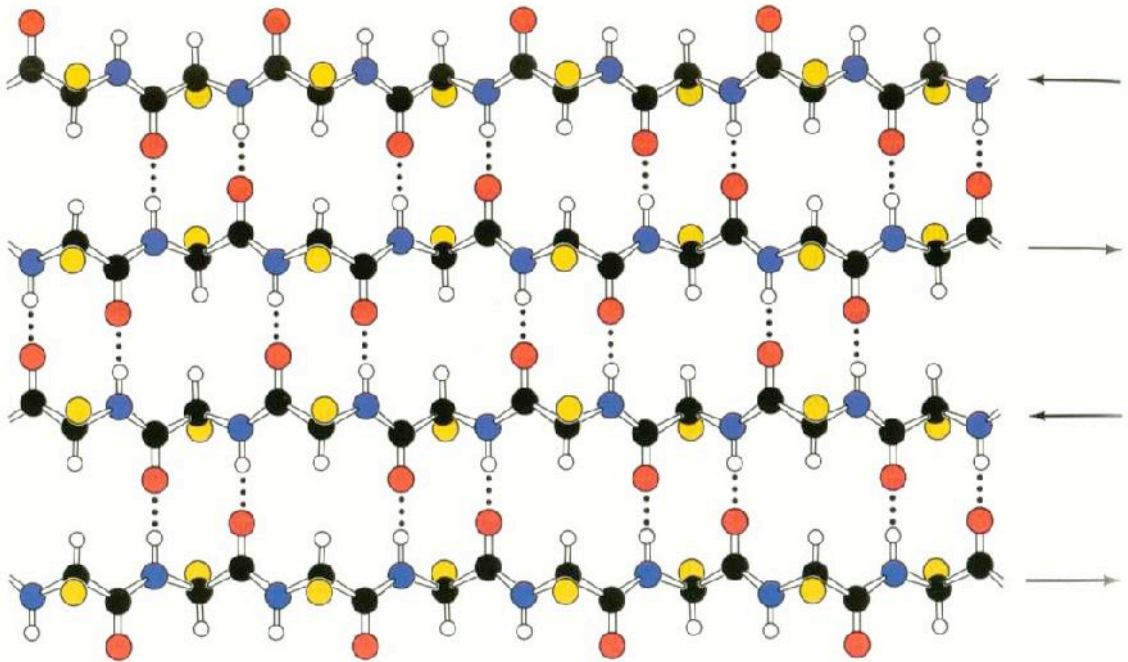


The helix can be viewed as a stacked array of peptide planes hinged at the  $\alpha$ -carbons and approximately parallel to the helix.

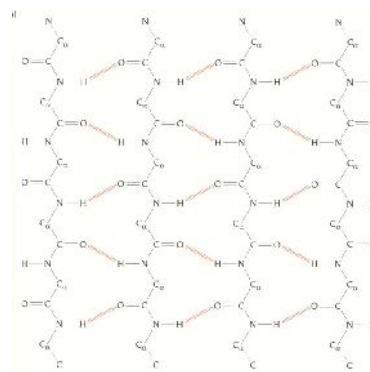
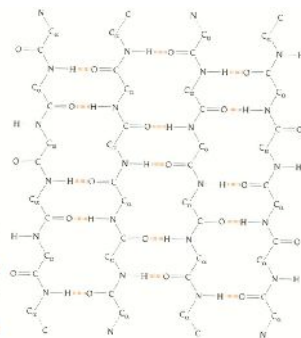




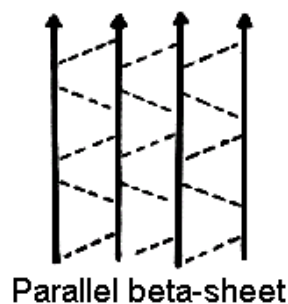
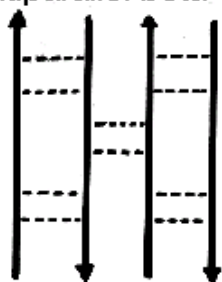
## Anti-parallel beta sheet



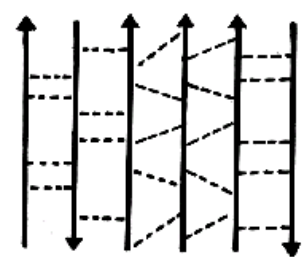
## Parallel and anti-parallel beta sheets



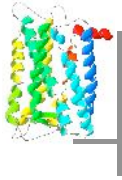
Antiparallel beta-sheet



Parallel beta-sheet

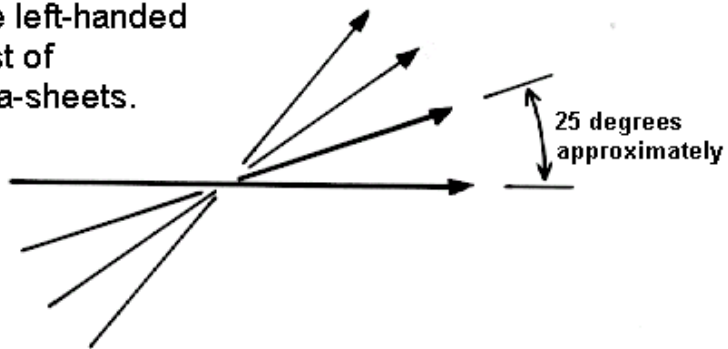


Mixed beta-sheet

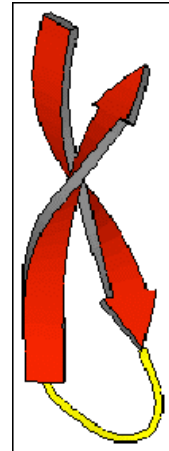


## Left-handed twist in beta-sheets

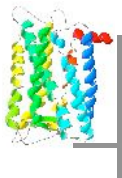
The left-handed twist of beta-sheets.



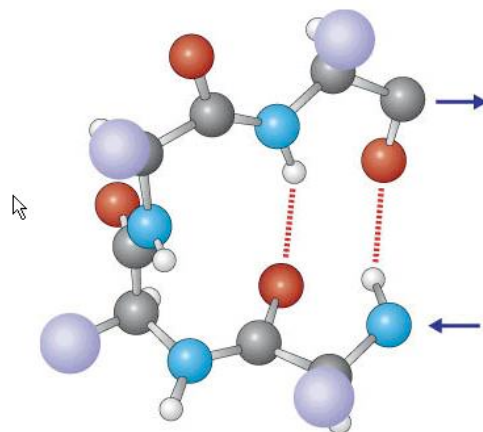
$0^\circ - 30^\circ$  per aa



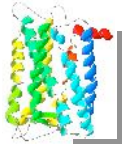
Bovine pancreatic trypsin inhibitor



## Turns and loops

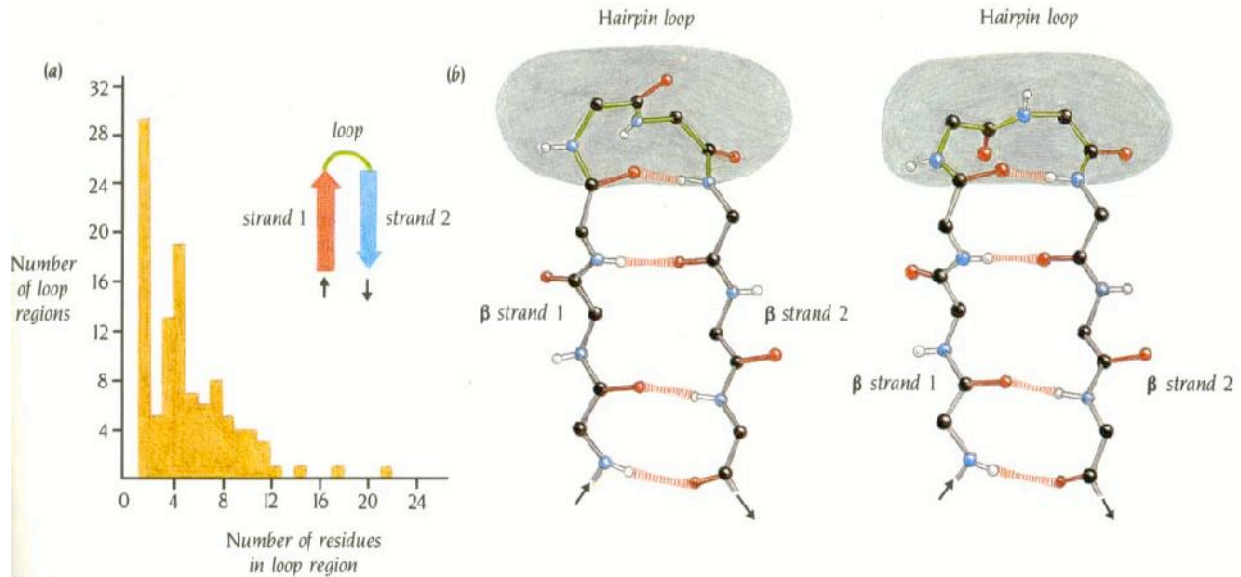


Schematic diagram showing the interresidue backbone hydrogen bonds that stabilize the reversal of the chain direction. Side chains are depicted as large light purple spheres. Due to the tight geometry of the turn, some residues are found more commonly in turns than others.



## Turns and loops

### □ Hairpin loops



## Conformational Preferences

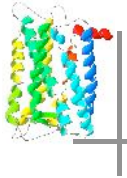
Conformational Preferences of the Amino Acids			
Amino acid	Preference		
	$\alpha$ -helix	$\beta$ -strand	reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.90
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

$\alpha$

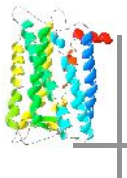
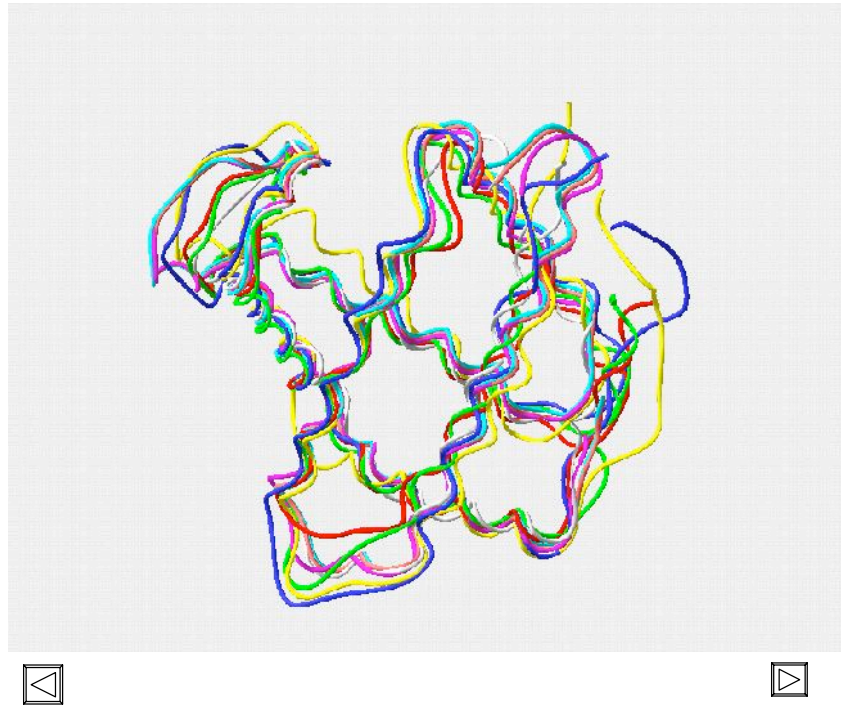
$\beta$

RT





## Pair Wise Structure Comparison (5)



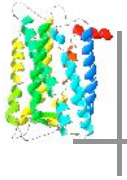
## Pair Wise Structure Comparison

### □ Root mean square deviation

- Comparing two structures A and B
- $R_{i,A}$  = Position of atom  $i$  in structure A
- $n$  = Number of equivalent atoms

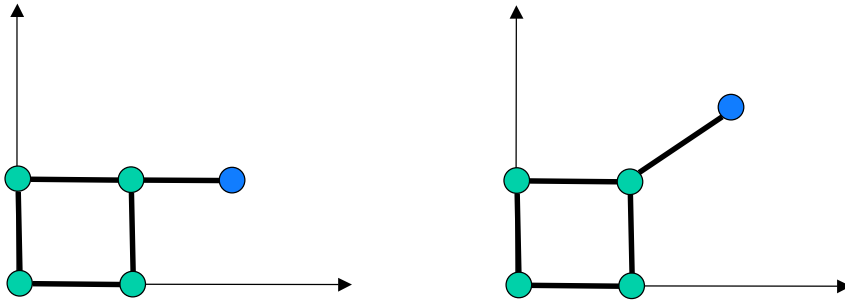
$$r.m.s.d. = \sqrt{\frac{\sum_{i=0}^n (R_{i,A} - R_{i,B})^2}{n}}$$



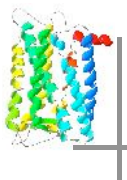


## RMSD

### □ Comparing two structures

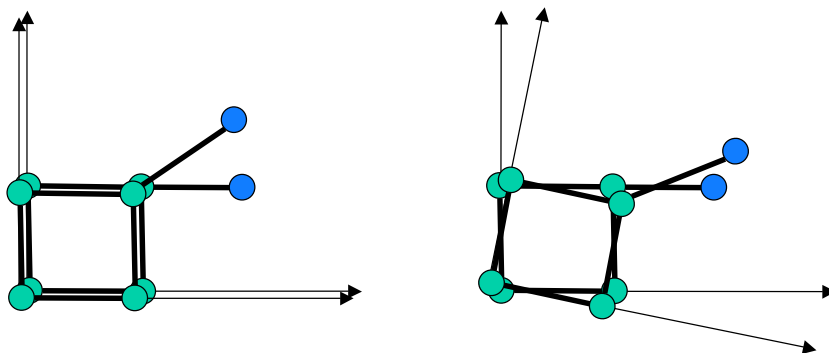


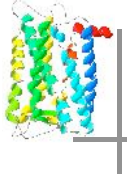
$$\text{Min: } r.m.s.d. = \sqrt{\frac{\sum_{i=0}^n (R_{i,A} - R_{i,B})^2}{n}}$$



## RMSD

### □ Comparing two structures





## Pair wise protein structure comparison methods

### •(Iterative) Least squares superposition (e.g. LSQ-Kabsch)

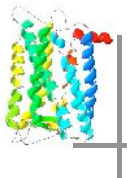
- uses RMSD (root mean square deviation) as similarity measure
- useful if sequence alignment is unambiguous
- only useful for similar structures (rmsd < 5 Å)

### •DALI (Distance Matrix Alignment)

- <http://www.ebi.ac.uk/dali/>
- Represents structures as 2-dimensional array of distances between C $\alpha$  positions
- Compare two structures by overlapping these arrays
- Similarity expressed as z-score vs. database background distribution

### •CE (Combinatorial Extension of the optimal path)

- <http://cl.sdsc.edu/ce.html>
- start by finding octameric fragments in both structures that can be overlapped
- extend aligned fragment pairs (AFPs) to find optimal alignment
- Similarity expressed as z-score vs. database background distribution

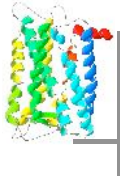


## Structural Alignments



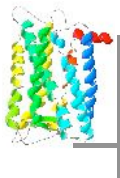
### ☐ **Protein Structure is better conserved than sequence**

- ☐ Structural alignments establish equivalences between amino acid residues based on the 3D structures of two or more proteins
- ☐ Structure alignments therefore provide information not available from sequence alignment methods
- ☐ Structural alignments can be used to guide sequence alignments (see: T\_COFFEE / SAP)



# Protein Structure Databases (6)

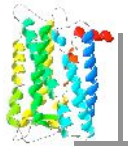
- ☐ PDB
- ☐ EBI-MSD
- ☐ SCOP
- ☐ CATH



The screenshot shows the RCSB Protein Data Bank (PDB) website in a Netscape browser window. The browser's address bar shows the URL <http://www.rcsb.org/pdb/>. The page has a blue header with the RCSB PDB logo and navigation links: [RCSB Home](#), [wwPDB Home](#), [Contact Us](#), and [Help](#). Below the header, a welcome message states: "Welcome to the PDB, the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data." The main content area is divided into several sections:

- DEPOSIT data**: Links to [DOWNLOAD files](#), [browse LINKS](#), [BETA TEST new features](#), and [BETA XML files](#).
- Current Holdings**: Displays "25115 Structures" and "Last Update: 13-Apr-2004". It includes a link to [PDB Statistics](#) and a 3D molecular model of a protein structure.
- Search the Archive**: A search interface with a text input field and a "Search" button. It offers search options: ☐ PDB ID, ☐ Authors, and ☒ Full Text Search. There are also checkboxes for ☒ match exact word and ☐ remove similar sequences. Below the search bar, there are links to [QuickSearch!](#), [SearchLite](#), [SearchFields](#), and [Status Search](#).
- PDB Mirrors**: A list of mirror sites with the note "\*\*Please bookmark a mirror site\*\*". The mirrors listed are: [San Diego Supercomputer Center, UCSD\\*](#), [Rutgers University\\*](#), [Center for Advanced Research in Biotechnology, NIST\\*](#), [Cambridge Crystallographic Data Centre, UK](#), [National University of Singapore](#), [Osaka University, Japan](#), and [Max Delbrück Center for Molecular Medicine, Germany](#). There is also a link to [OCA / PDB Lite](#) and a "MORE..." link.
- News**: A section with a "Complete News Newsletter" link and a "pdb-l Archive Subscribe" link. It features a news item dated "13-Apr-2004" about the "RCSB Poster Prize awarded at RECOMB".

At the bottom of the page, there is a section for citing the PDB, which includes a list of names and a reference to a publication in *Nucleic Acids Research*.



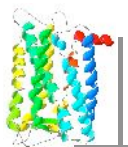
*This page contains:*

[PDB Holdings List](#): Breakdown of the contents of the PDB, following the latest update

[PDB Content Growth](#): Data on the growth in the number of structures in the PDB

### PDB Holdings List: 13-Apr-2004

		Molecule Type				
		Proteins, Peptides, and Viruses	Protein/Nucleic Acid Complexes	Nucleic Acids	Carbohydrates	total
Exp.	X-ray Diffraction and other	19696	969	727	14	21406
Tech.	NMR	3031	97	577	4	3709
	<b>Total</b>	22727	1066	1304	18	<b>25115</b>

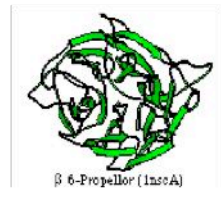
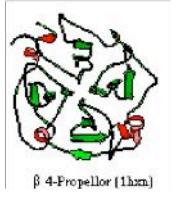
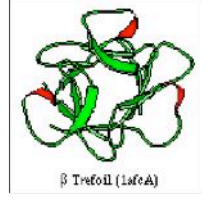
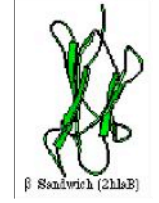
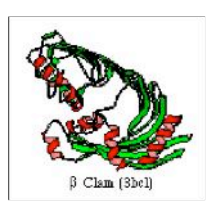
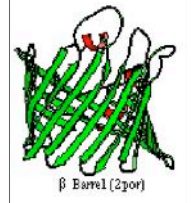
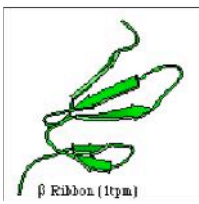
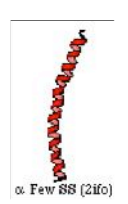
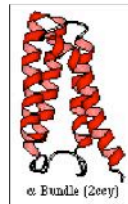
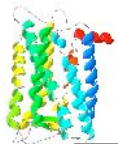


## Fold Classification Databases

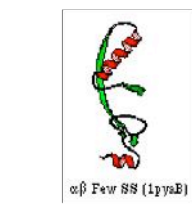
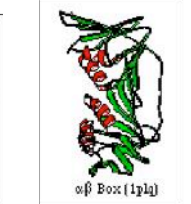
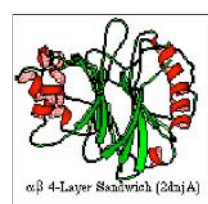
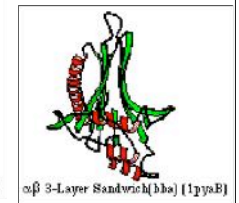
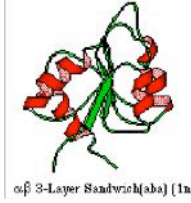
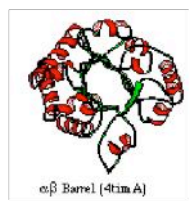
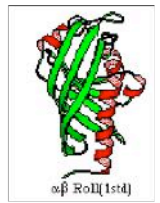
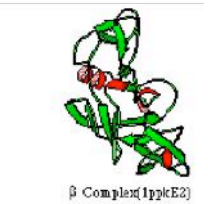
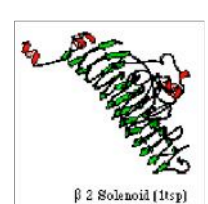
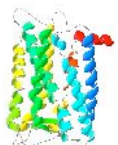
- Number of folds is limited
  - according to C. Chothia ca. 1000 – 5000 folds

Fold databases systematically classify protein structure folds

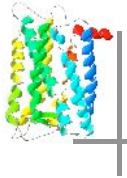
- **FSSP**
  - *Fold classification based on Structure-Structure alignment of Proteins*
  - <http://www2.ebi.ac.uk/dali/fssp/>
  - Contains structurally aligned proteins based in DALI
- **SCOP**
- **CATH**



Christine Orengo (Structures, 1997, 5, 1093-1108)



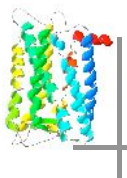
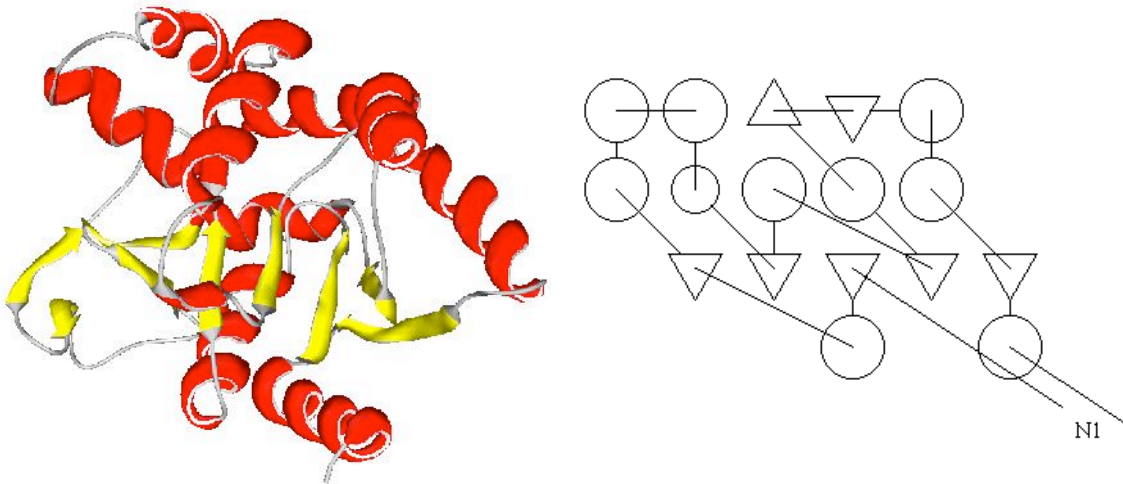
Christine Orengo (Structures, 1997, 5, 1093-1108)



## Topology Cartoons

### ☐ TOPS: Topology of Protein Strcuture

➤ <http://www.tops.leeds.ac.uk/>



## CATH - Protein Structure Classification

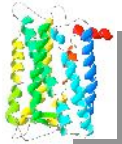
### ☐ Hierarchical classification of protein domain structures

### ☐ UCL, Janet Thornton & Christine Orengo

### ☐ clusters proteins at four major levels:

- Class(C)
- Architecture(A)
- Topology(T)
- Homologous superfamily (H)





## CATH - Protein Structure Classification



### □ **Class(C)**

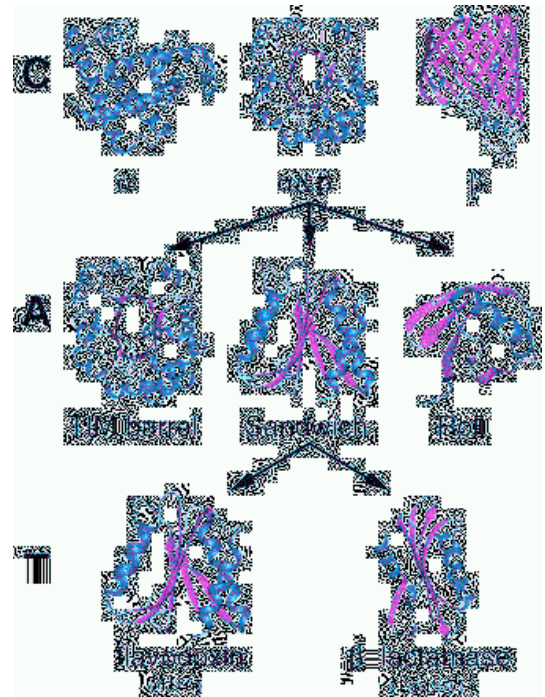
derived from secondary structure content is assigned automatically

### □ **Architecture(A)**

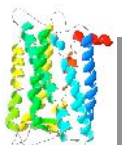
describes the gross orientation of secondary structures, independent of connectivity.

### □ **Topology(T)**

clusters structures according to their topological connections and numbers of secondary structures



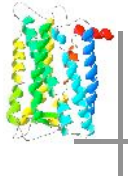
[ [http://www.biochem.ucl.ac.uk/bsm/cath\\_new/](http://www.biochem.ucl.ac.uk/bsm/cath_new/) ]



## SCOP - Structural Classification of Proteins

- MRC Cambridge (UK), Alexey Murzin, Brenner S. E., Hubbard T., Chothia C.
- hierarchical classification of protein domain structures
- created by manual inspection
- comprehensive description of the structural and evolutionary relationships
- organized as a tree structure
  - Class
  - Fold
  - Superfamily
  - Family
  - Species

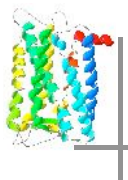
[ <http://scop.mrc-lmb.cam.ac.uk/scop/> ]



## Protein Structure / Fold Databases

---

- ❑ PDB: <http://www.pdb.org>
- ❑ EBI-MSD <http://www.ebi.ac.uk/msd/>
- ❑ SCOP <http://scop.mrc-lmb.cam.ac.uk/scop/>
- ❑ CATH [http://www.biochem.ucl.ac.uk/bsm/cath\\_new/](http://www.biochem.ucl.ac.uk/bsm/cath_new/)



## References

---

- ❑ I. Branden, J. Tooze. Introduction to Protein Structure, Garland Publishing.
- ❑ P.E.Bourne, H. Weissig. Structural Bioinformatics, Wiley-Liss and Sons.
- ❑ G.A. Petsko, D. Ringe. Protein Structure and Function, New Science Press.