Molecular Simulations: Applications in Biology

Structure of Biomolecules: An Overview

Dr. R. Sankararamakrishnan
Department of Biological Sciences & Bioengineering
Indian Institute of Technology, Kanpur

Understanding Molecular Simulations: Theory and Applications UMS2010
11th November 2010

Outline of this talk

- What are Biomolecules?
- Significance of knowing the structure of a biomolecule?

- Why simulate a biomolecule?

- What is the current status?

- Biomolecular simulation: An example

Biopolymers	Building Blocks
Proteins	Amino acids
Nucleic acids	Nucleotides
Carbohydrates	Sugars
Lipids	Fatty acids

Proteins play crucial roles in all biological processes

Trypsin, Chmytrypsin – enzymes

Hemoglobin, Myoglobin – transports oxygen

Transferrin – transports iron

Ferritin – stores iron

Myosin, Actin – muscle contraction

Collagen – strength of skin and bone

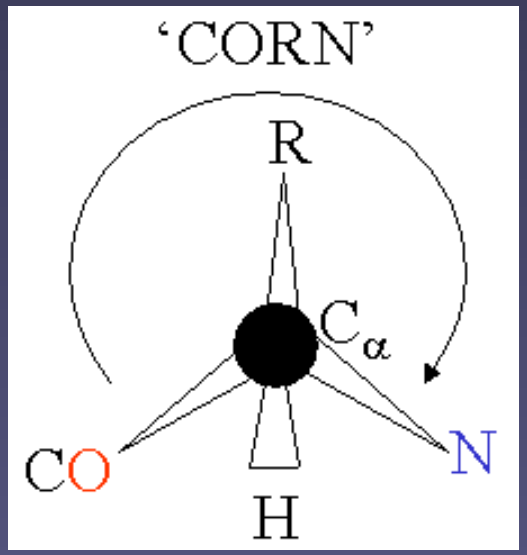
Rhodopsin – light-sensitive protein

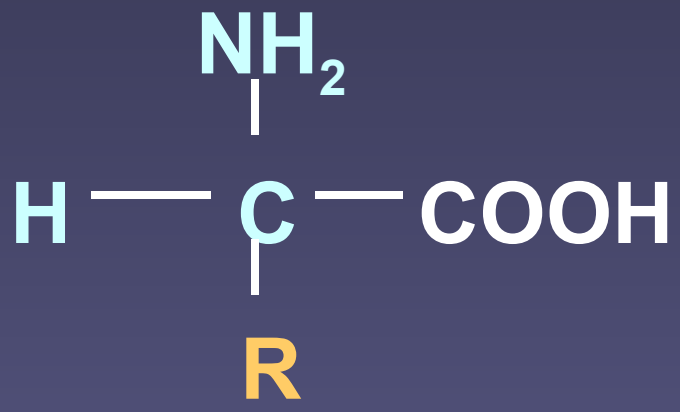
Acetylcholine receptor – responsible for transmitting nerve impluses

Antibodies – recognize foreign substances

Repressor and growth factor proetins

Proteins are made up of 20 amino acids

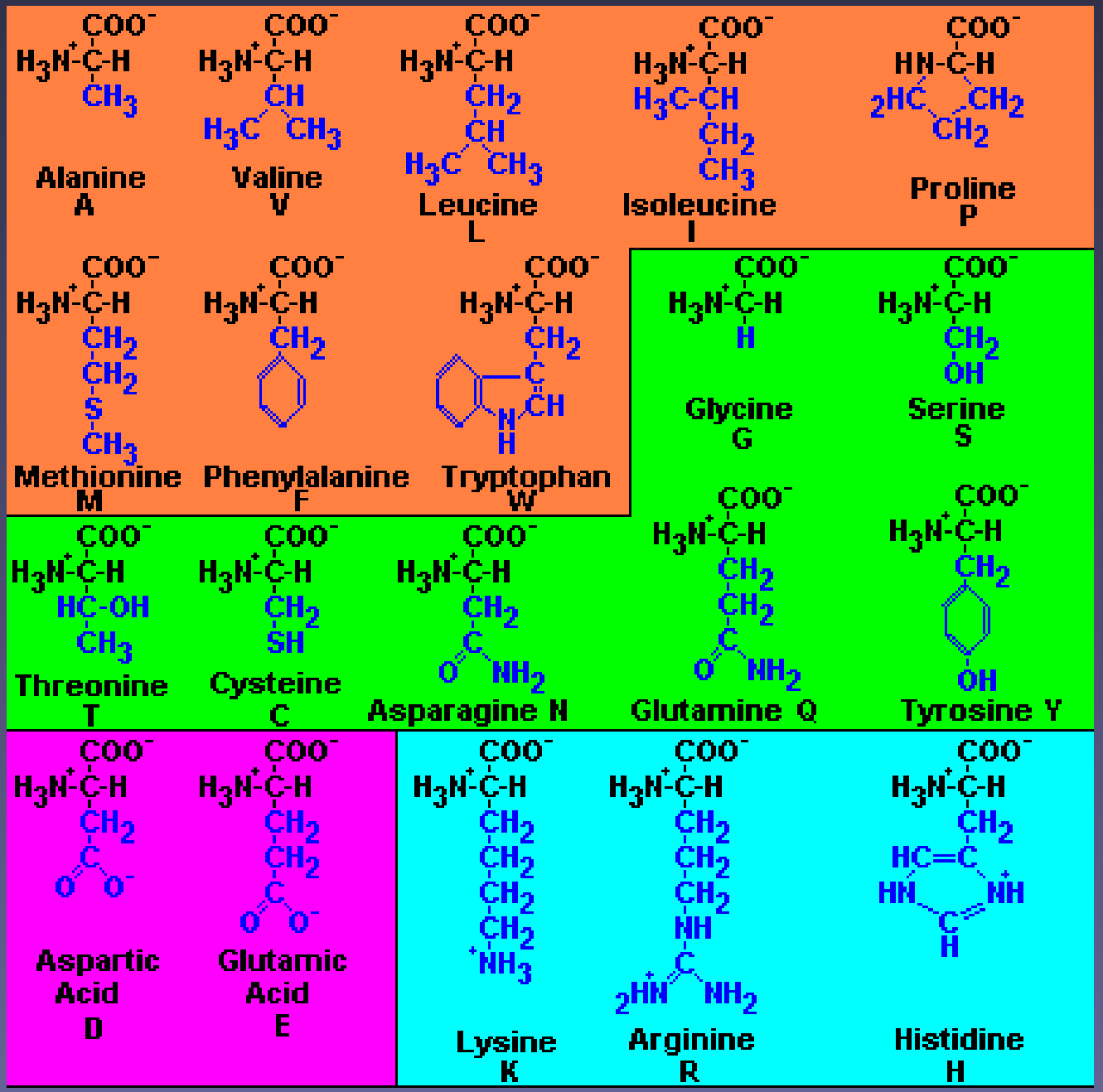




R varies in size, shape, charge, hydrogen-bonding capacity and chemical reactivity.

Only L-amino acids are constituents of proteins

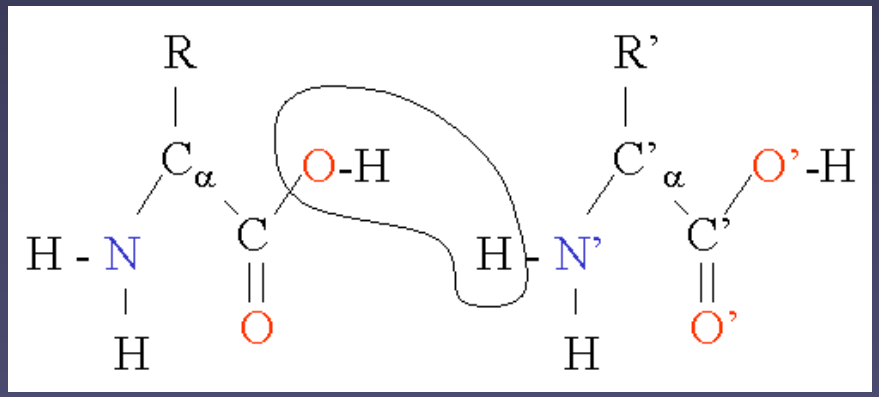
Nonpolar and hydrophobic

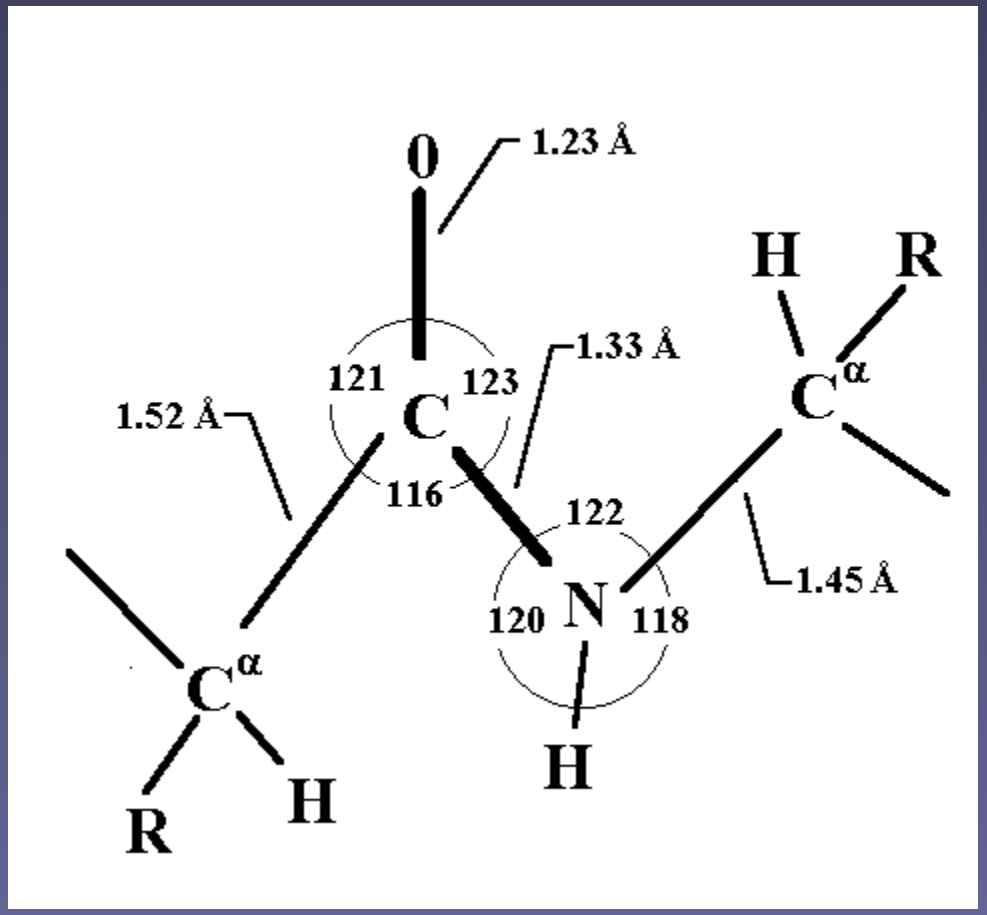


Acidic

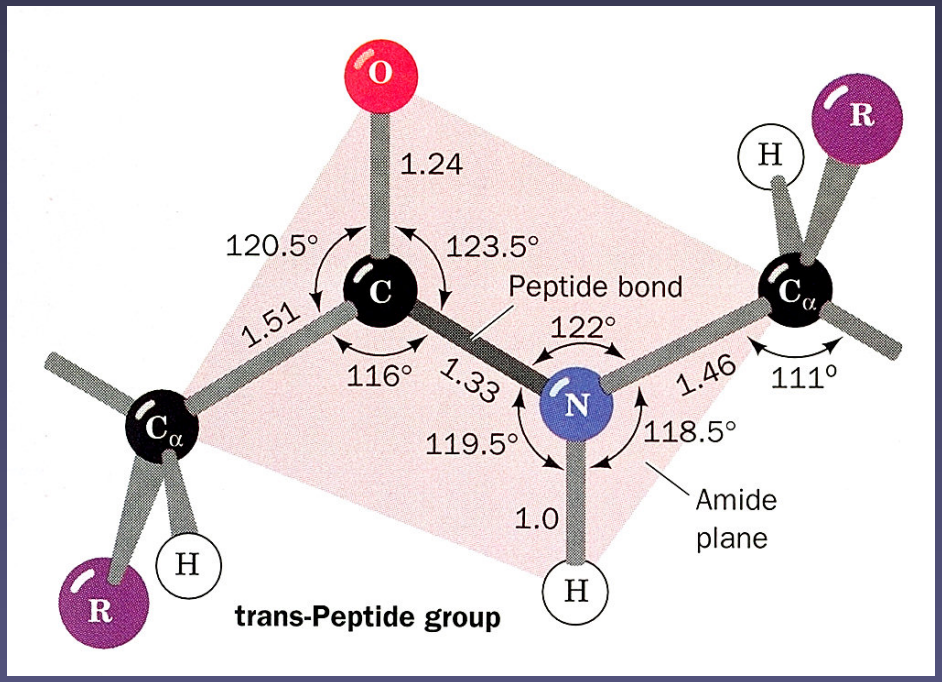
Basic

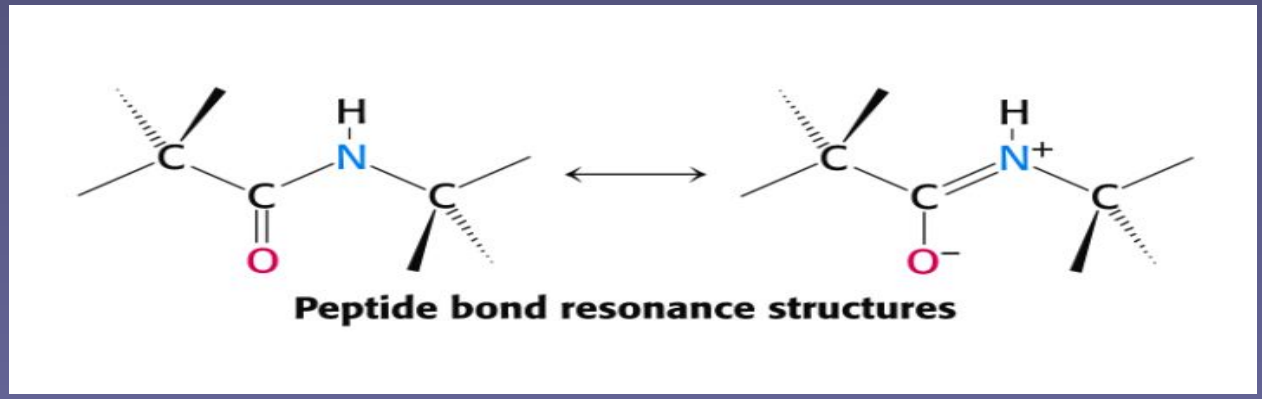
20 amino acids are linked into proteins by *peptide bond*





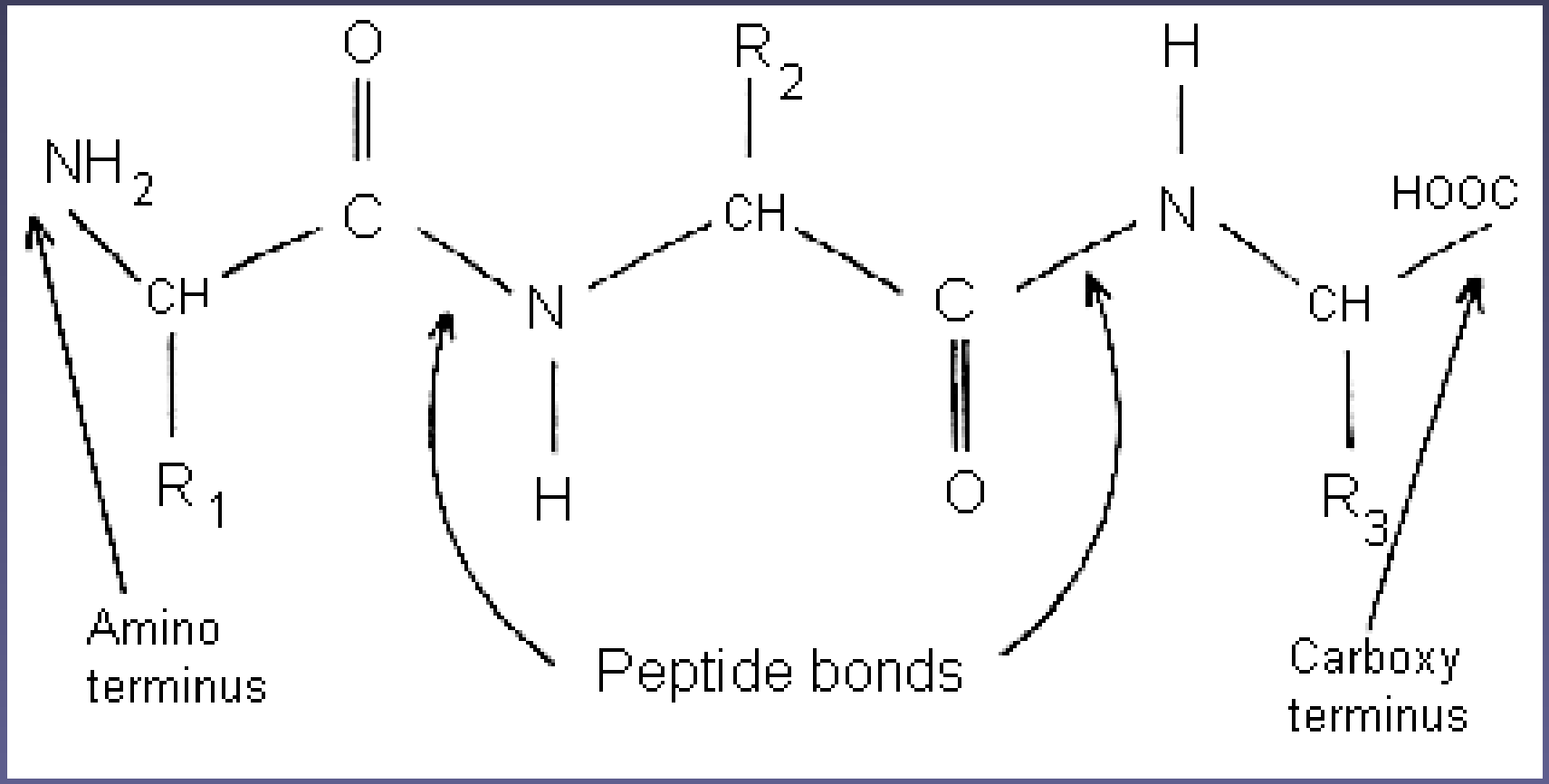
Peptide bond has partial double-bonded character and its rotation is restricted.





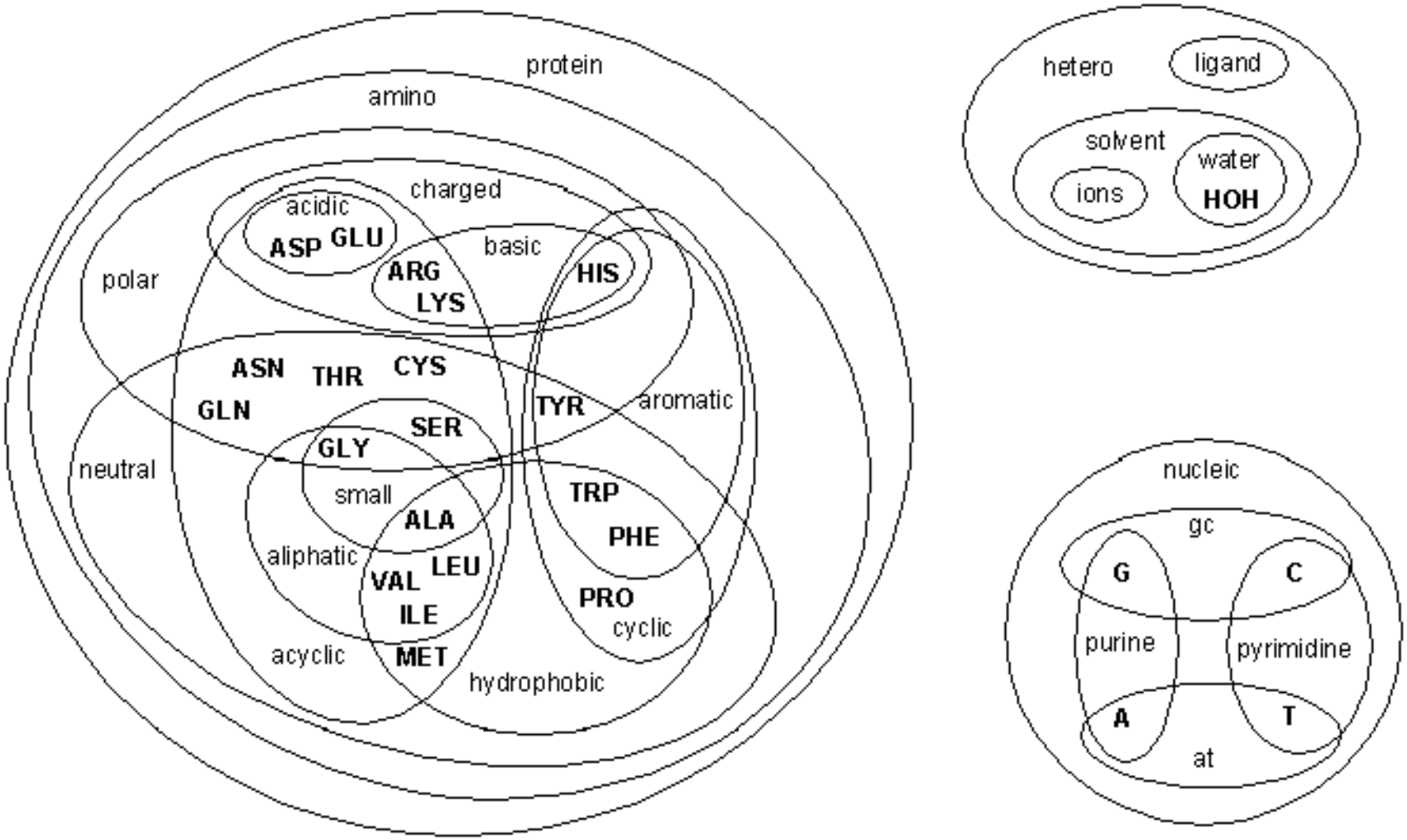
Peptide bond resonance structures

Polypeptide backbone is a repetition of basic unit common to all amino acids



Frequently encountered terms in protein structure

- Backbone
- Side chain
- Residue



RasMol/Chime Venn diagram v1.2
Drawn by Kurt Giles (kurt@inn-prot.weizmann.ac.il) for the Israeli National Node of EMBnet

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

One letter and three-letter codes for amino acids

Proteins can exist in two types of environments

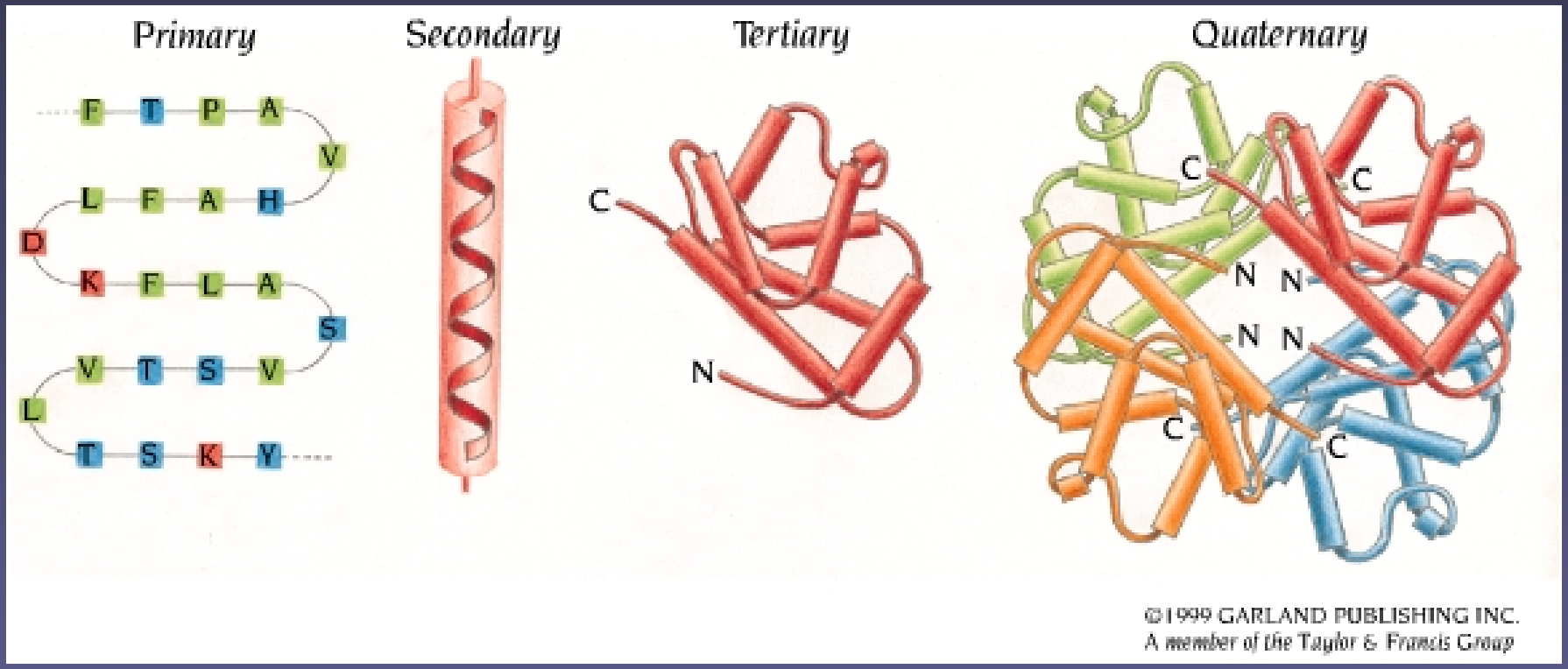
Globular proteins

Membrane proteins – Dr. Satyavani

Each protein has a characteristic three-dimensional structure which is important for its function

Protein Structure: Four Basic Levels

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


Primary
Secondary
Tertiary
Quaternary
F T P A V
L F A H
D
K F L A S
V T S V
L
T S K Y
C
N
C
C
N N
N N
C
C
©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

Protein – Primary Structure

- Linear amino acid sequence
- Determines all its chemical and biological properties
- Specifies higher levels of protein structure (secondary, tertiary and quaternary)

Most proteins contain between ~200 to ~500 residues

Histone (human)

SETVPPAPAASAAPEKPLAGKKAKKPAKAAAAASKKKPAGPSVSELIVQAASSSKER
GGVSLAALKKALAAAGYDVEKNNSRIKLGIKSLVSKGTLVQTKGTGASGSFKLNK
KASSVETKPGASKVATKTKATGASKKLKKATGASKKSVKTPKKAKKPAATRKSSK
NPKKPKTVKPKKVAKSPAKAKAVKPKAAKARVTKPKTAKPKKAAPKKK

Rhodopsin (human)

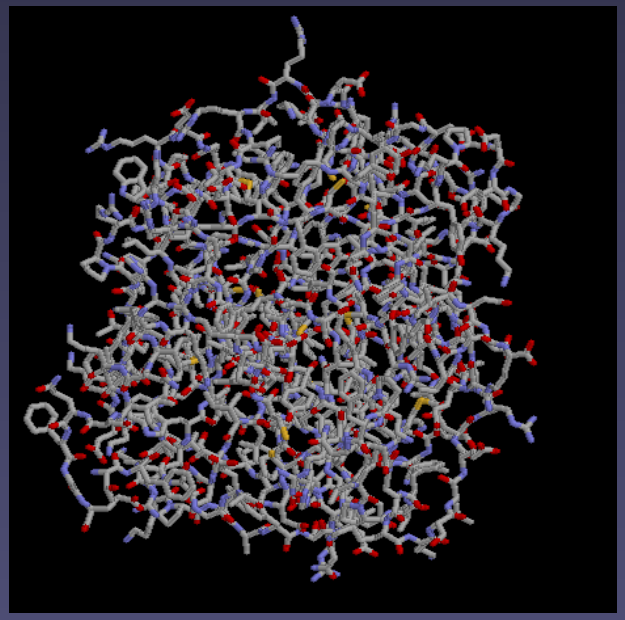
MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPI
NFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGC
NLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVM
ALACAAPPLAGWSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIII
FFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMIIMVIAFLICWVPYASVAF
YIFTHQGSNFGPIFMTIPAFFAKSAAIYNPVIYIMMNKQFRNCMLTTICCGKNP
LGDDEASATVSKTETSQVAPA

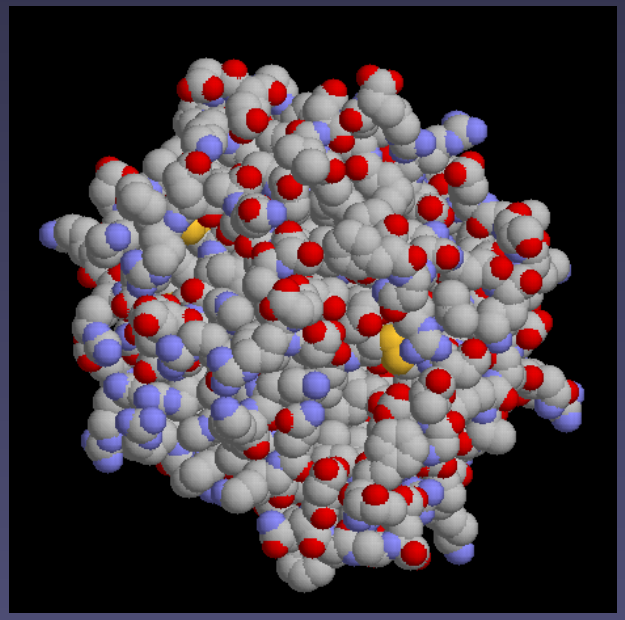
Thrombin

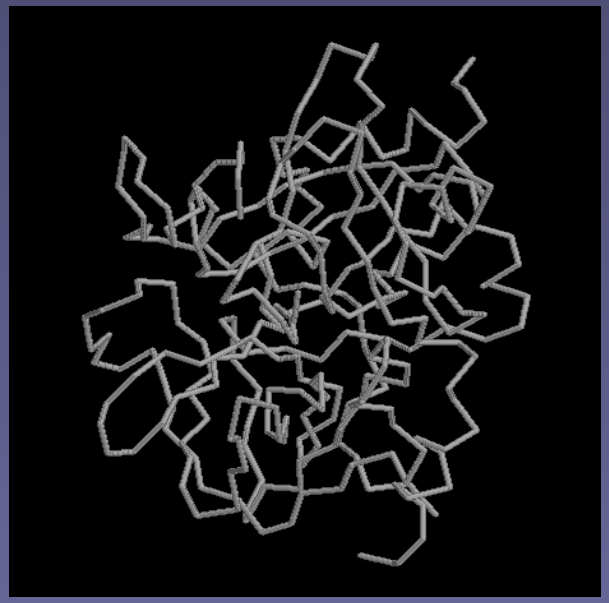
Heavy chain:
IVEGSDAEIGMSPWQVMLFRKSPQELLCGASLISDRWVLTAAHCLLYPPW
DKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIHPRYNWRENLDRDIAL
MKLKKPVAFSDYIHVCLPDRETAASLLQAGYKGRVTGWGNLKETWTANVG
KGQPSVLQVVNLPIVERPVCKDSTRIRITDNMFCAGYKPDEGKRGDACEGDS
GGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYGFY
THVFRLKKWIQKVIDQFGE

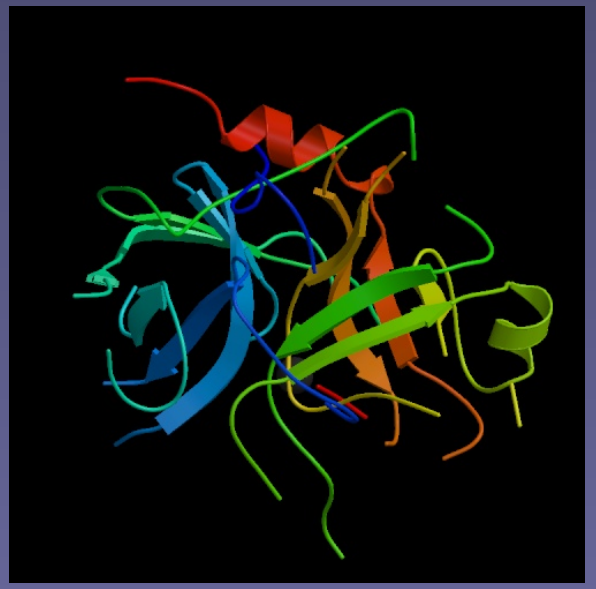
Light Chain: TFGSGEADCGLRPLFEKKSLEDKTERELLESYIDGR

Thrombin Structure

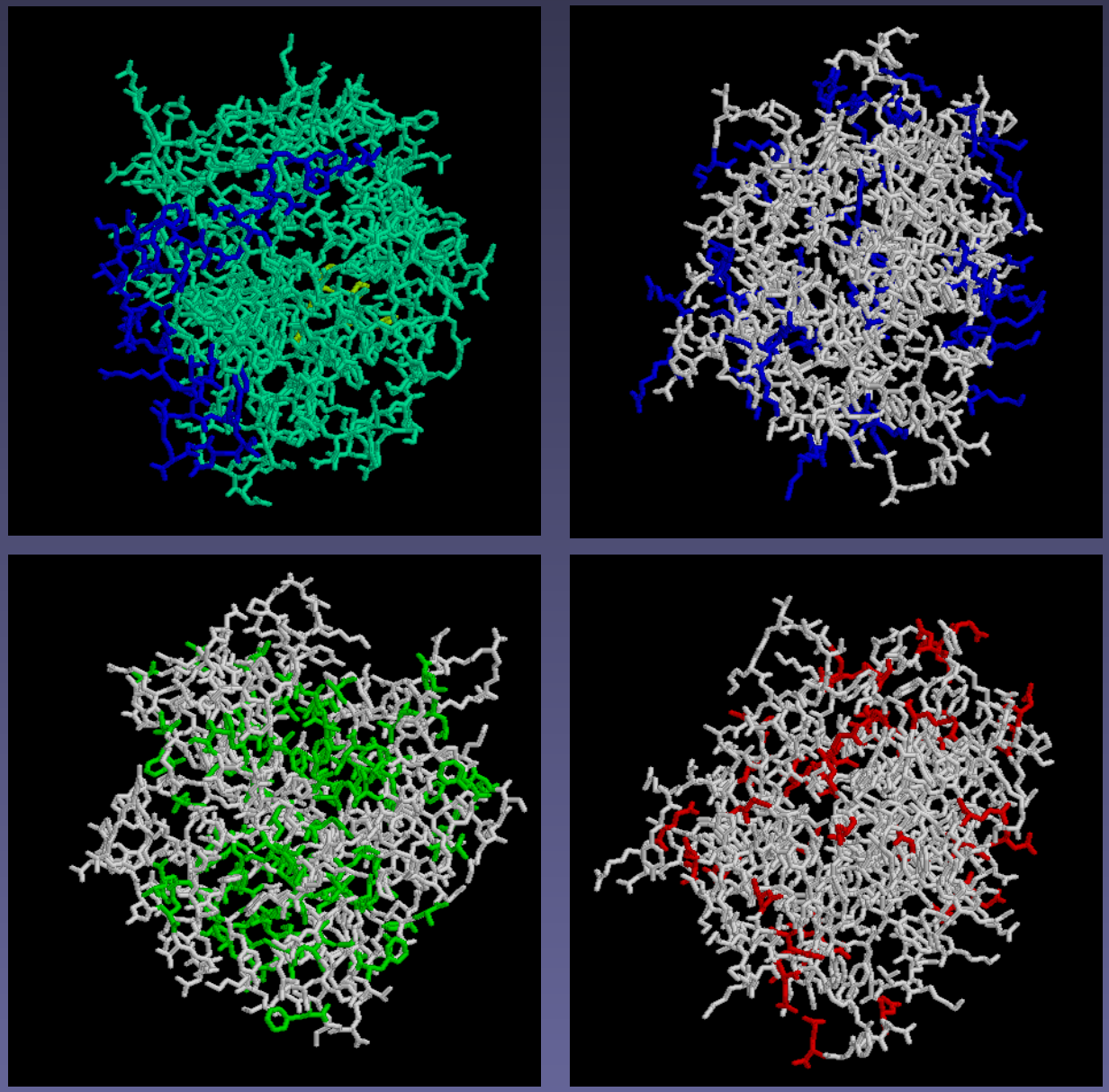






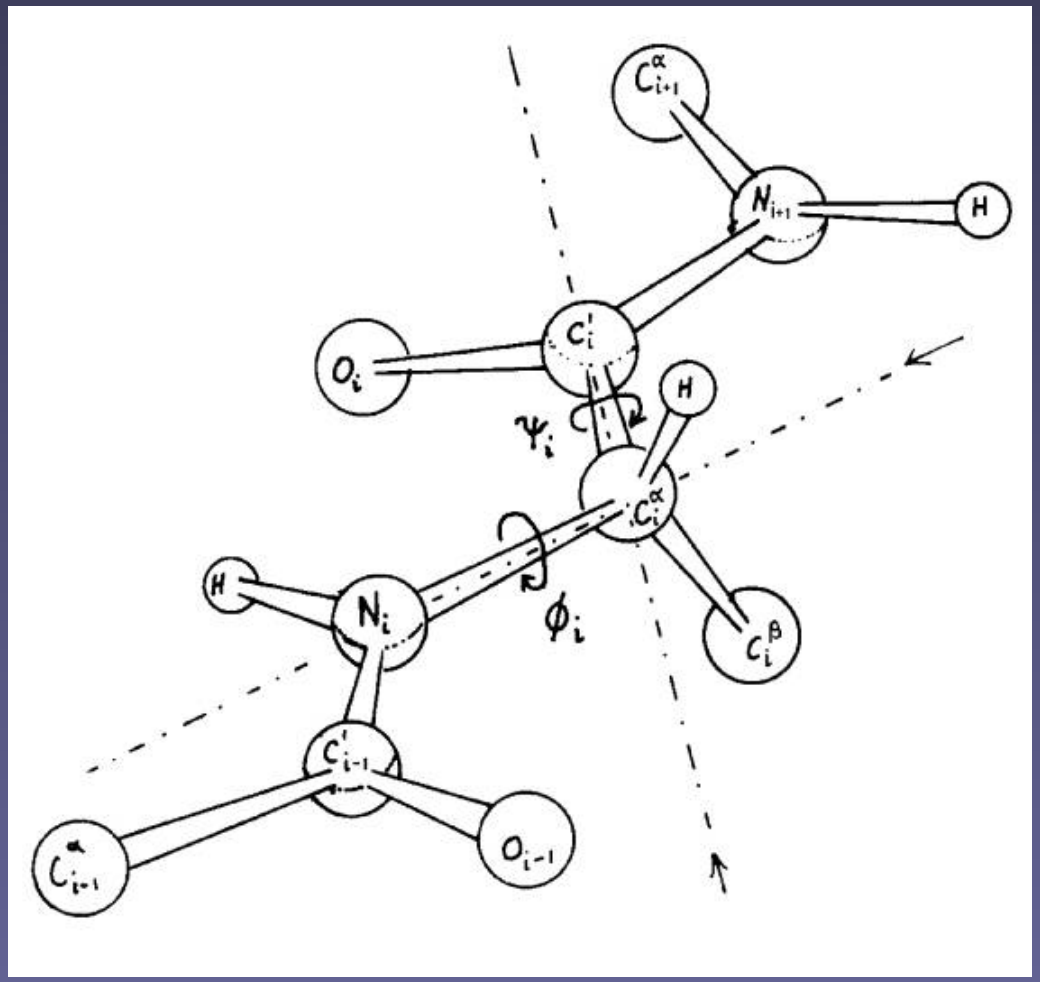


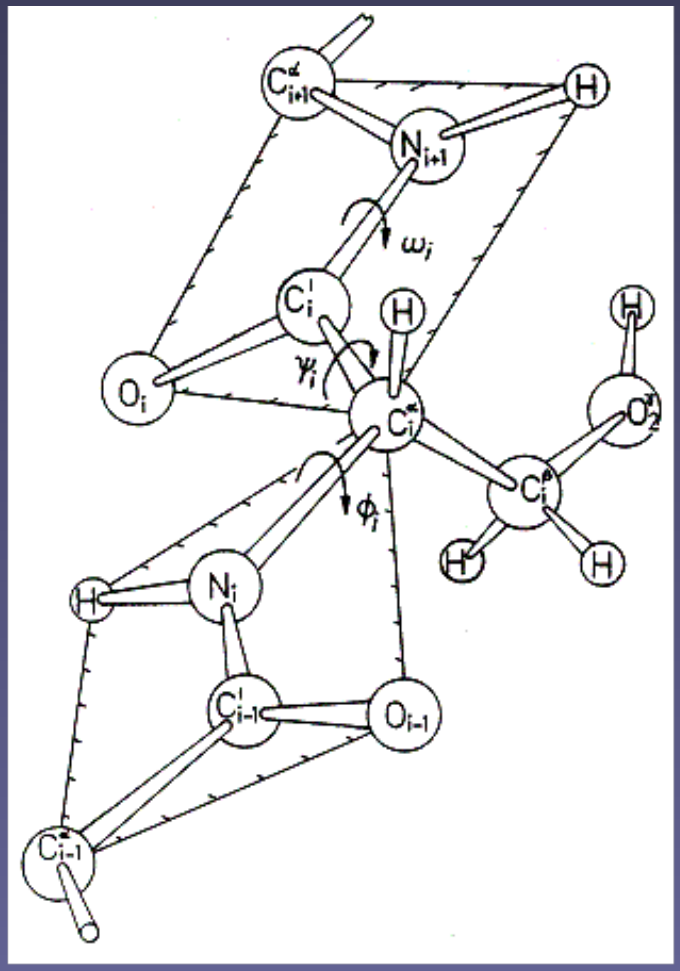
Thrombin Structure



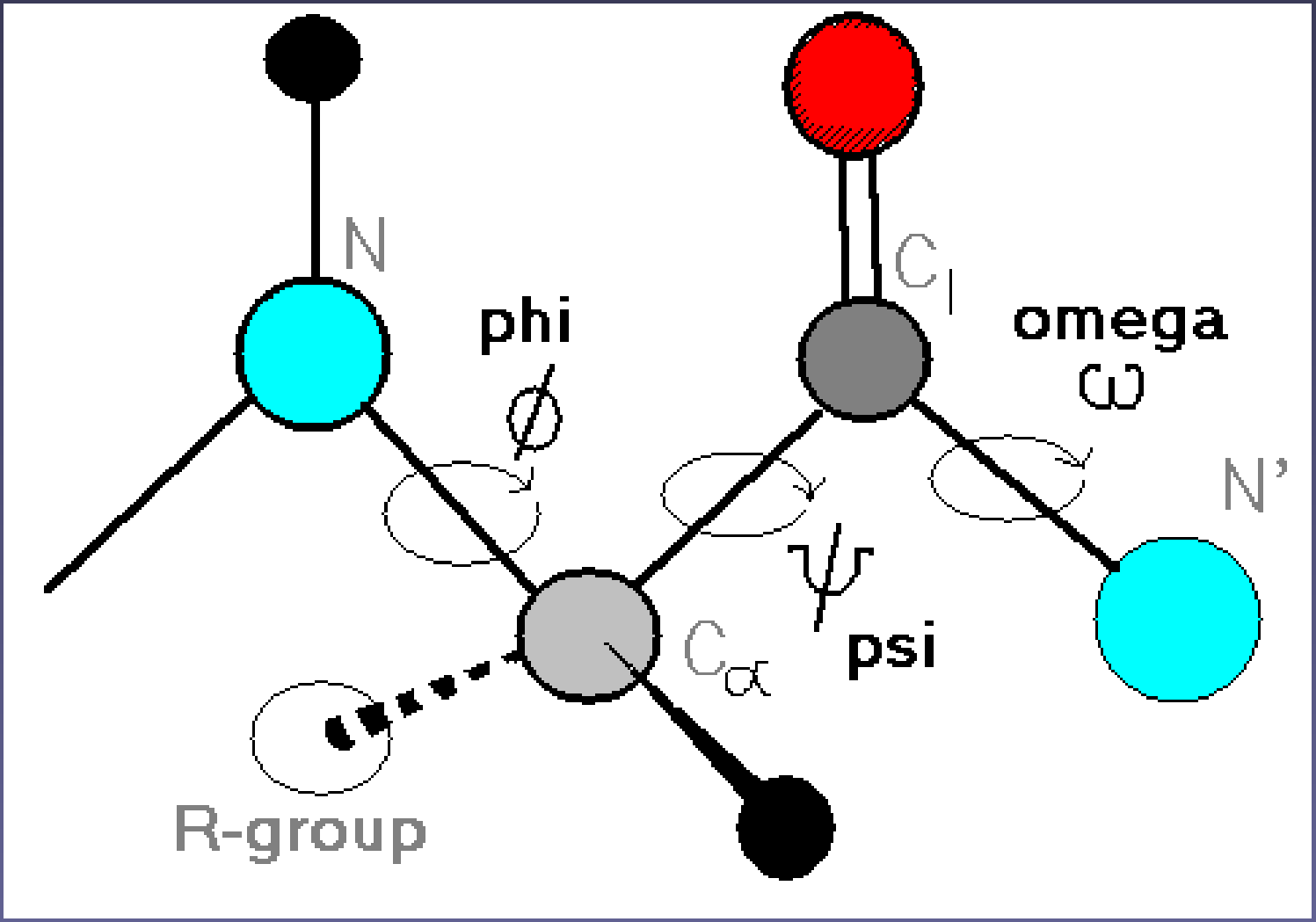
Primary to Secondary structure

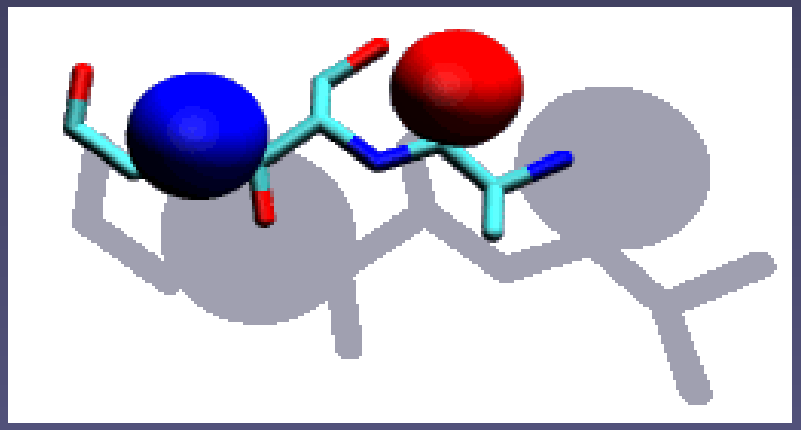
Importance of Dihedral Angle



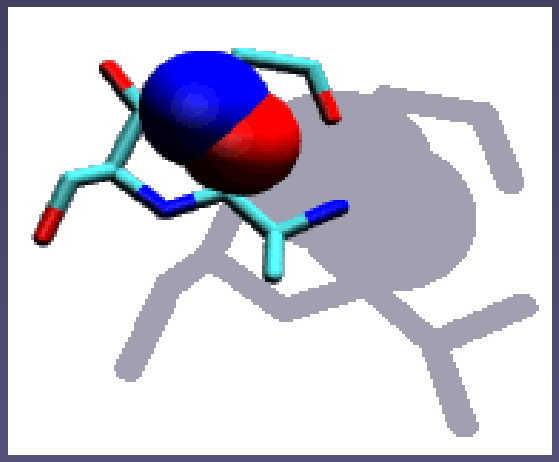


Dihedral angles ϕ, ψ and ω

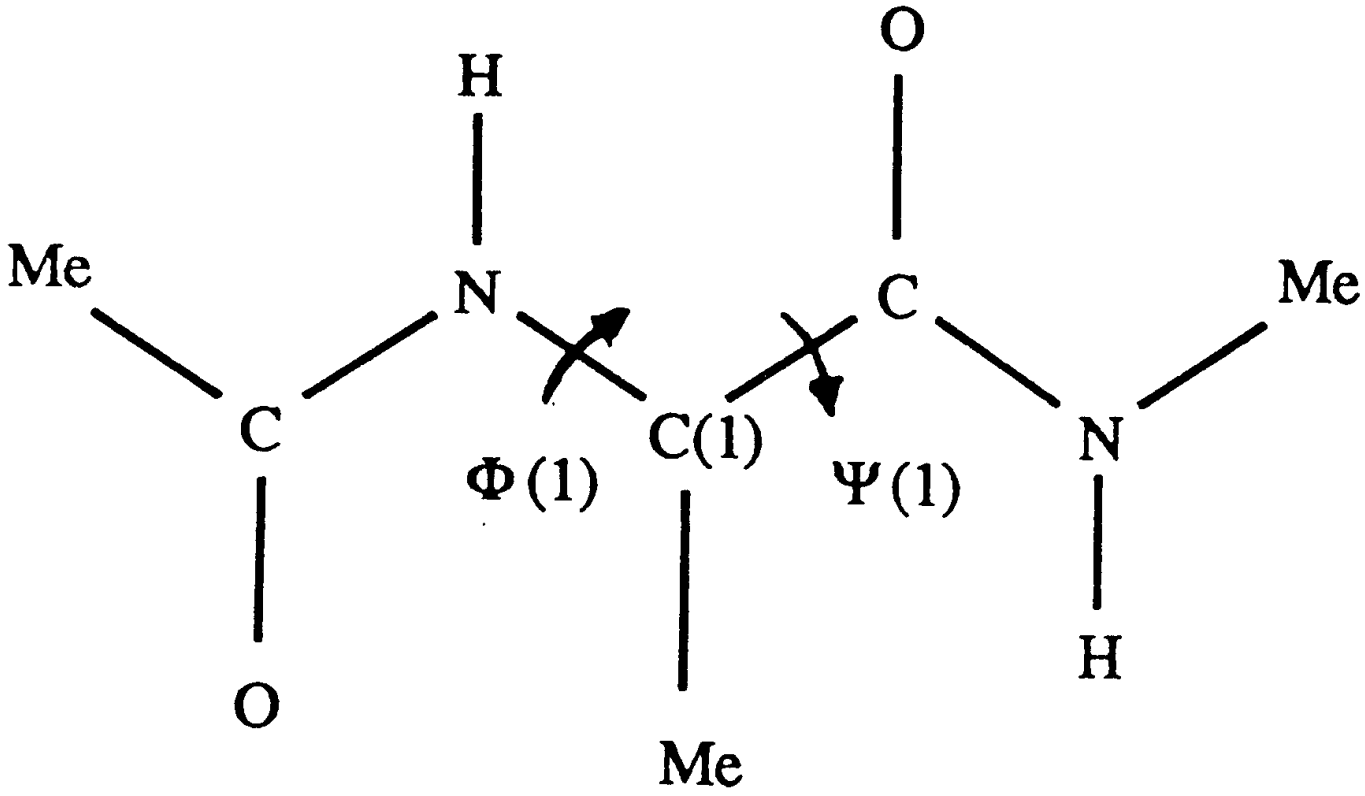




$\phi = 180°; \psi = 180°$



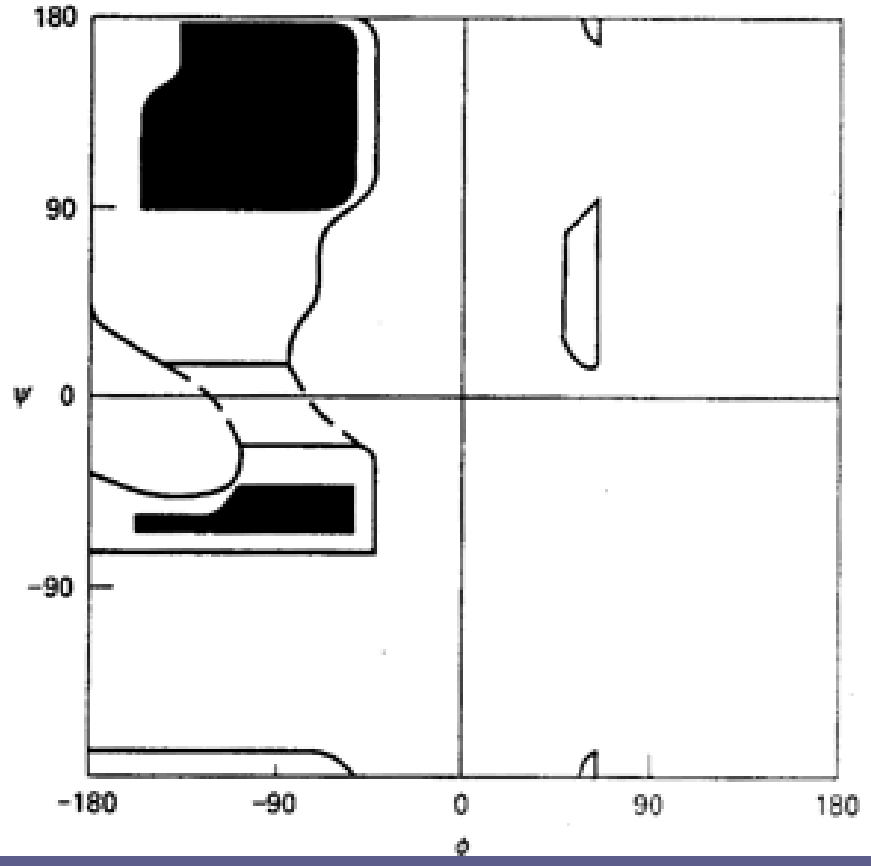
$\phi = 0°; \psi = 0°$


O
H
Me
N
C
Me
C
C(1)
N
Φ(1)
Ψ(1)
O
H
Me

Limiting distances for various interatomic contacts

Types of contact	Normal Limit	Extreme Limit
H…H	2.0	1.9
H…O	2.4	2.2
H…N	2.4	2.2
H…C	2.4	2.2
O…O	2.7	2.6
O…N	2.7	2.6
O…C	2.8	2.7
N…N	2.7	2.6
N…C	2.9	2.8
C…C	3.0	2.9
C…C(H)	3.2	3.0
C(H)…C(H)	3.2	3.0

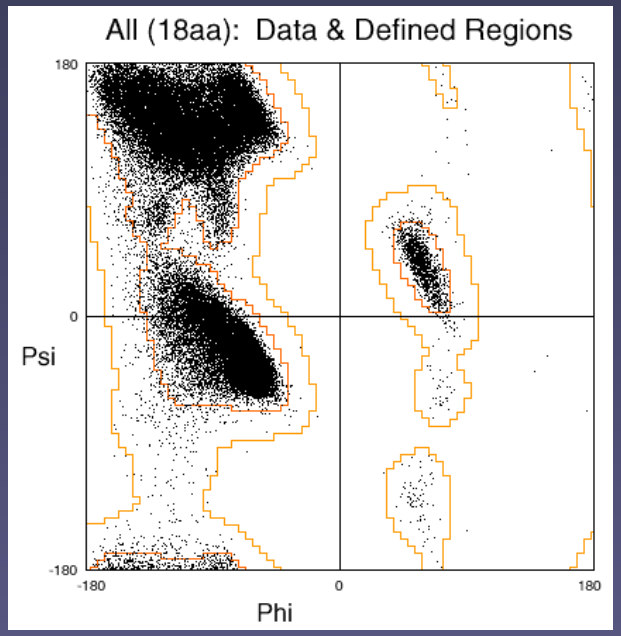
Ramachandran & Sasisekharan (1968) Adv. Protein Chem.

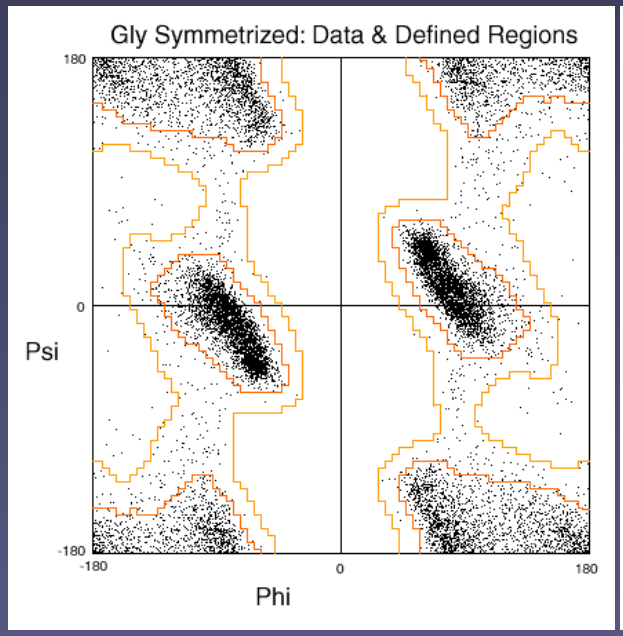


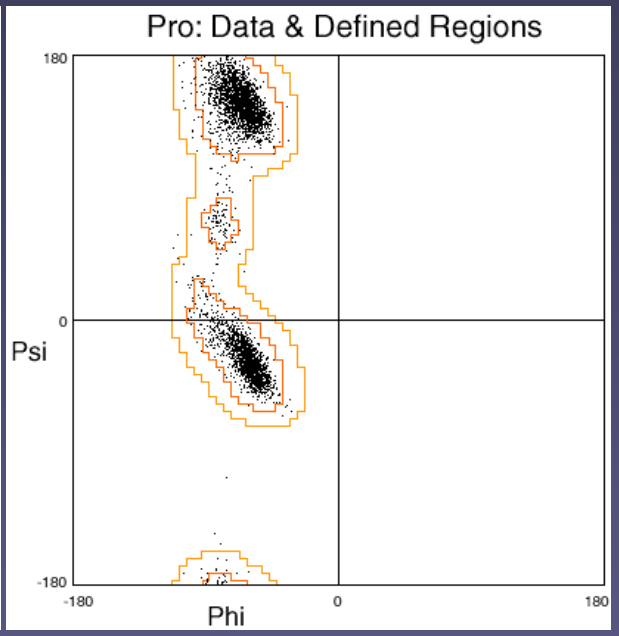
Ramachandran Plot

Ramachandran Plot

Data from 500 high-resolution proteins

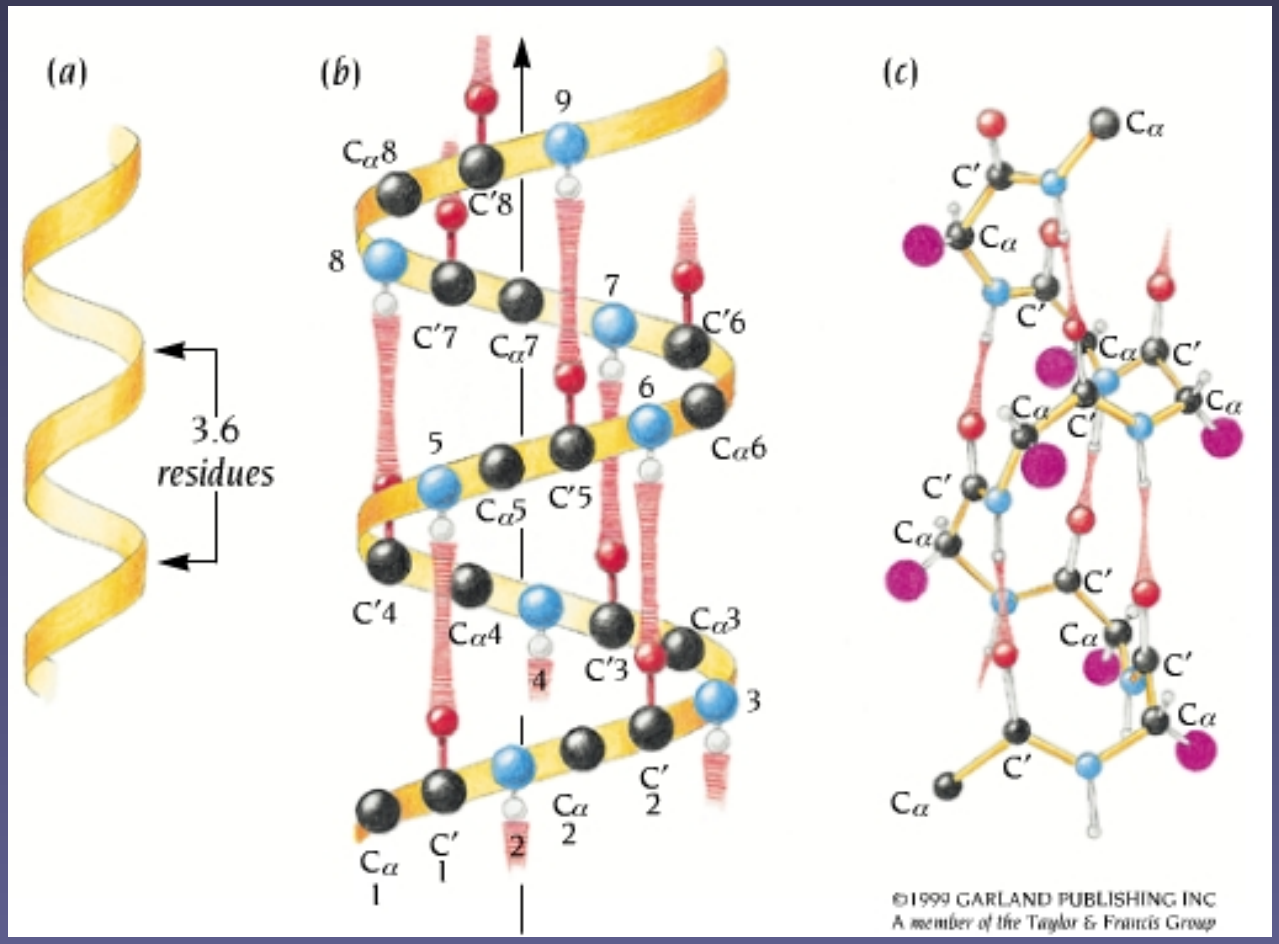


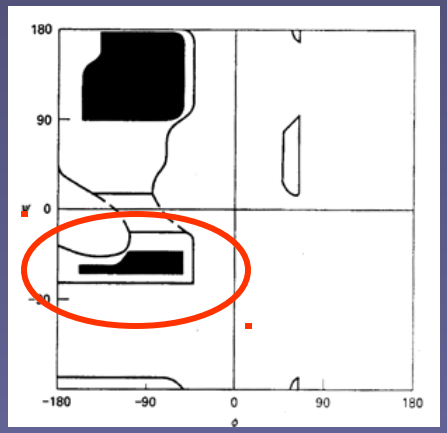




Secondary Structure

α-helix

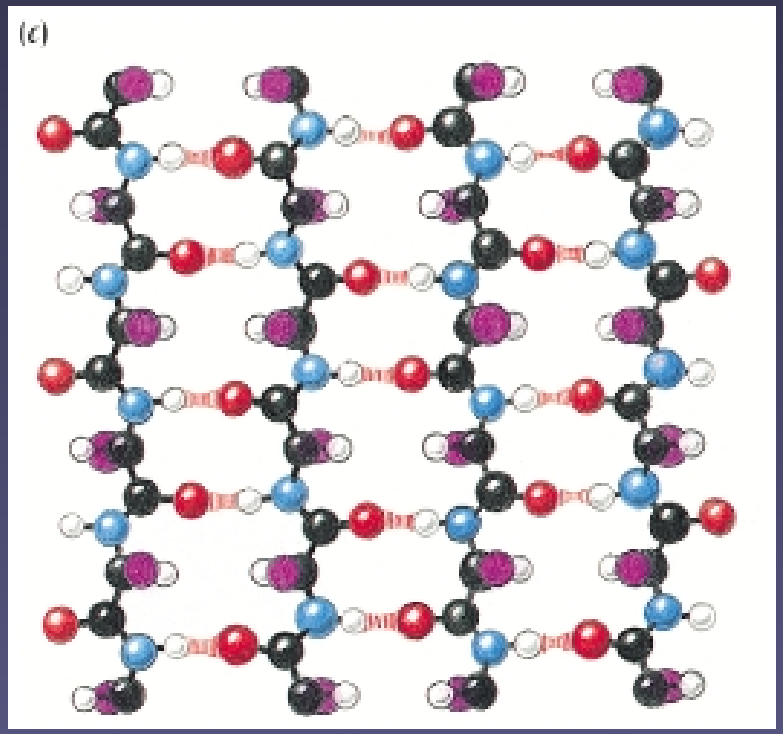


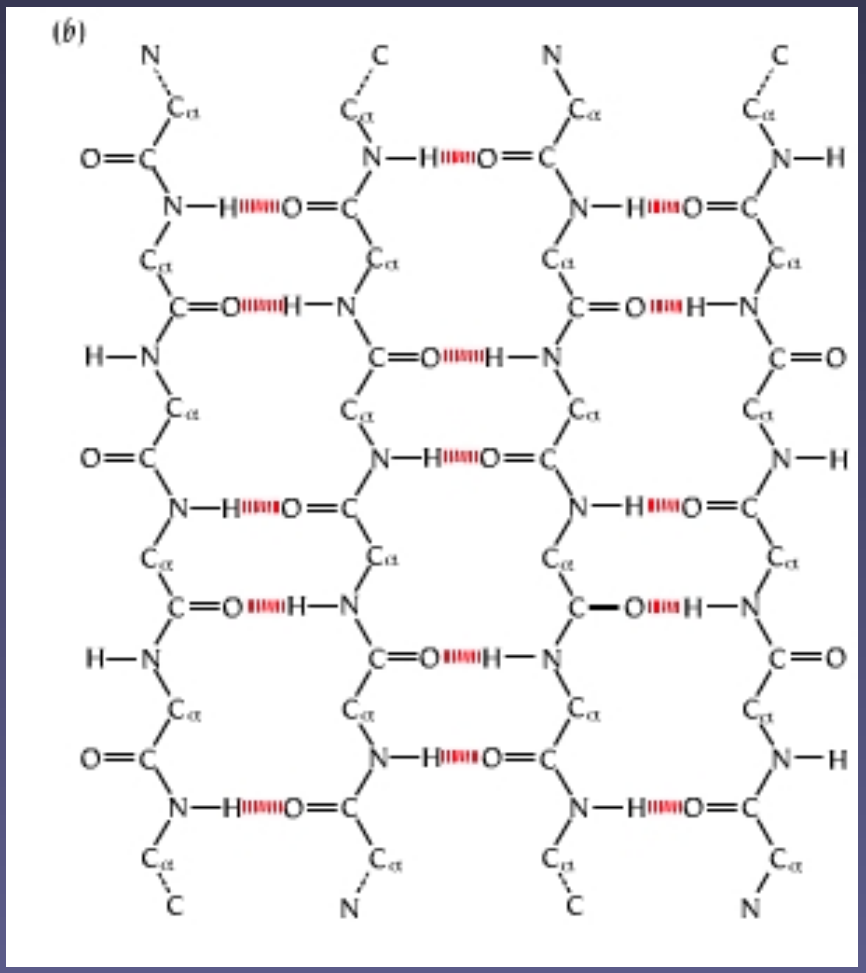


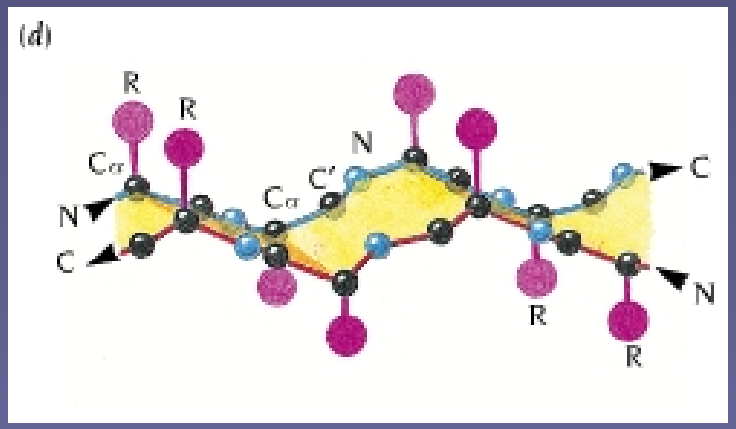
α-helix

- 3.6 residues per turn
- Translation per residue 1.5 Å
- Translation 5.4 Å per turn
- C=O (i) … H-N (i+4)
- ϕ = -57°; ψ = -47° (classical value)
- ϕ = -62°; ψ = -41° (crystal structures)
- Preference of residues in helix
- Can proline occur in a helix?
- Average helix length ~ 10 residues

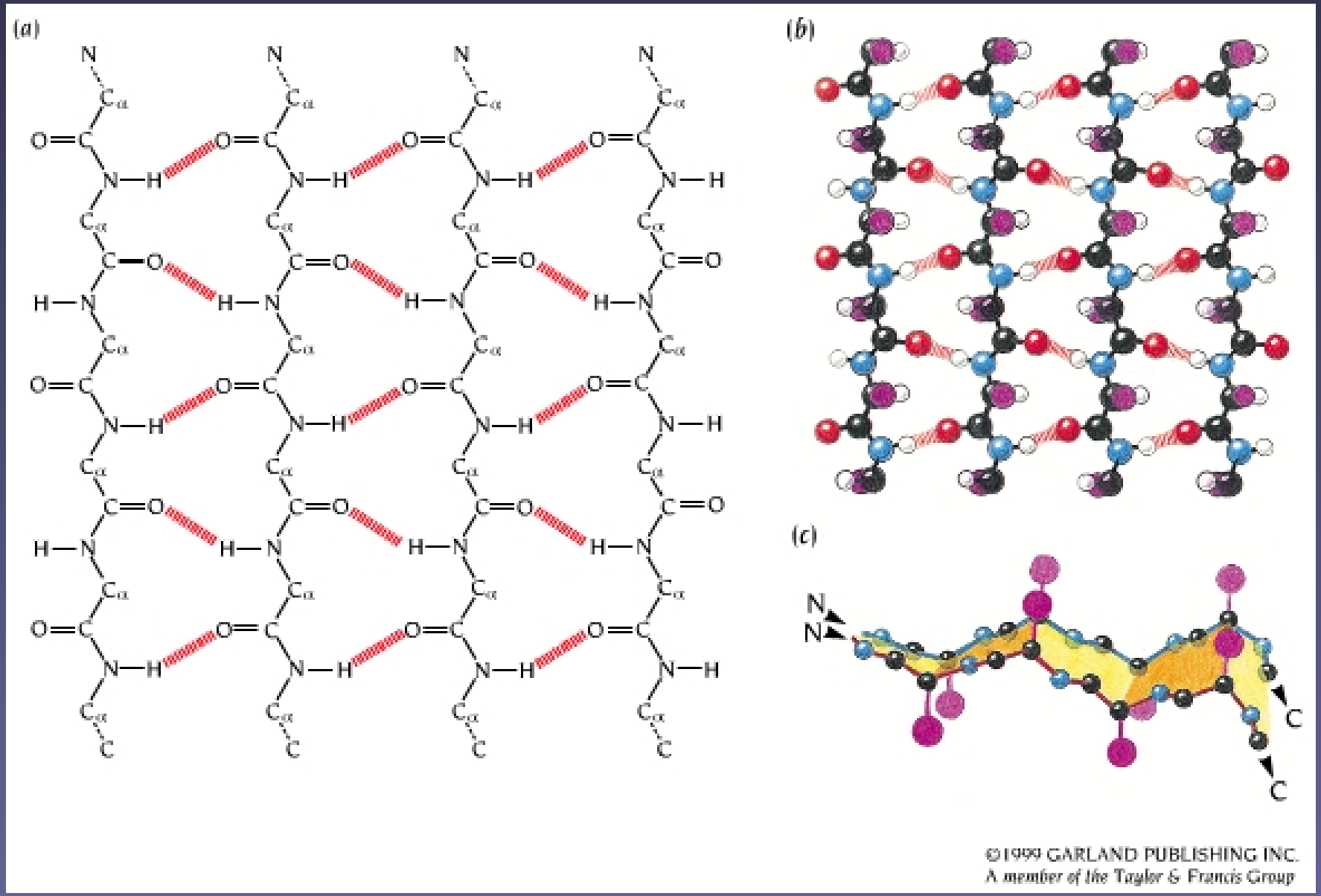
Antiparallel β-sheet





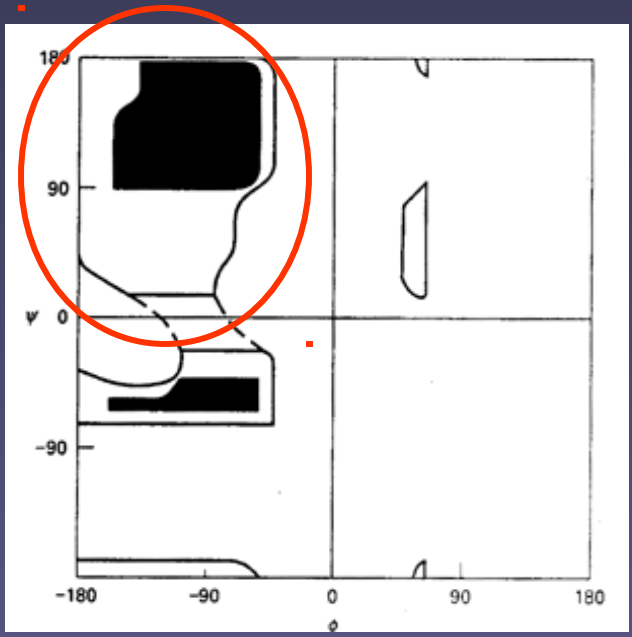


Parallel β-sheet



©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

β-strand

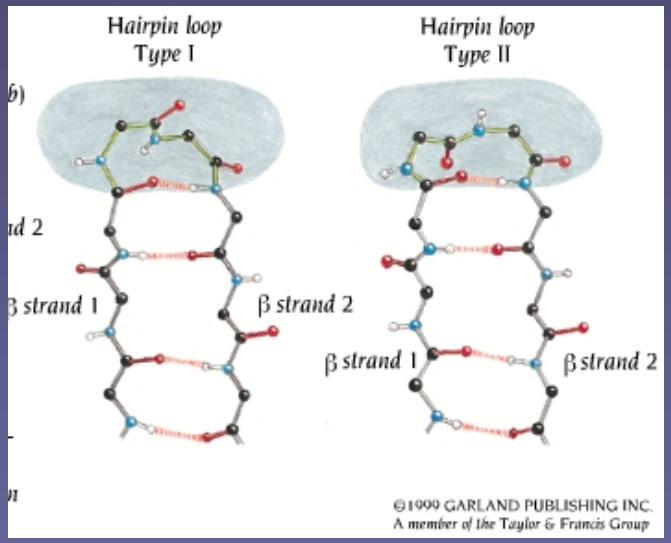


- Polypeptide fully extended
- 2.0 residues per turn
- Translation 3.4Å per residue
- Stable when incorporated into a β-sheet
- H-bonds between peptide groups of adjacent strands
- Adjacent strands can be parallel or antiparallel

Turns

- Secondary structures are connected by loop regions
- Lengths vary; shapes irregular
- Loop regions are at the surface of the molecule
- Rich in charged and polar hydrophilic residues
- Role: connecting units; binding sites; enzyme active sites
- Loops are often flexible; adopt different conformations

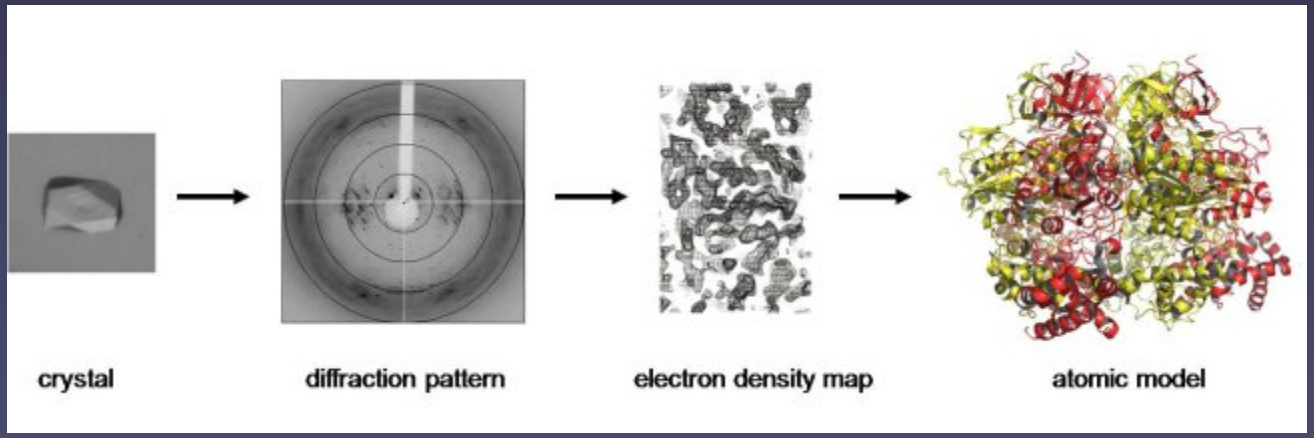
- β-turns: Type I, Type II etc.
- γ-turns; classical, inverse



G.D. Rose et al., Adv. Protein Chemistry **37** (1989) 1-109

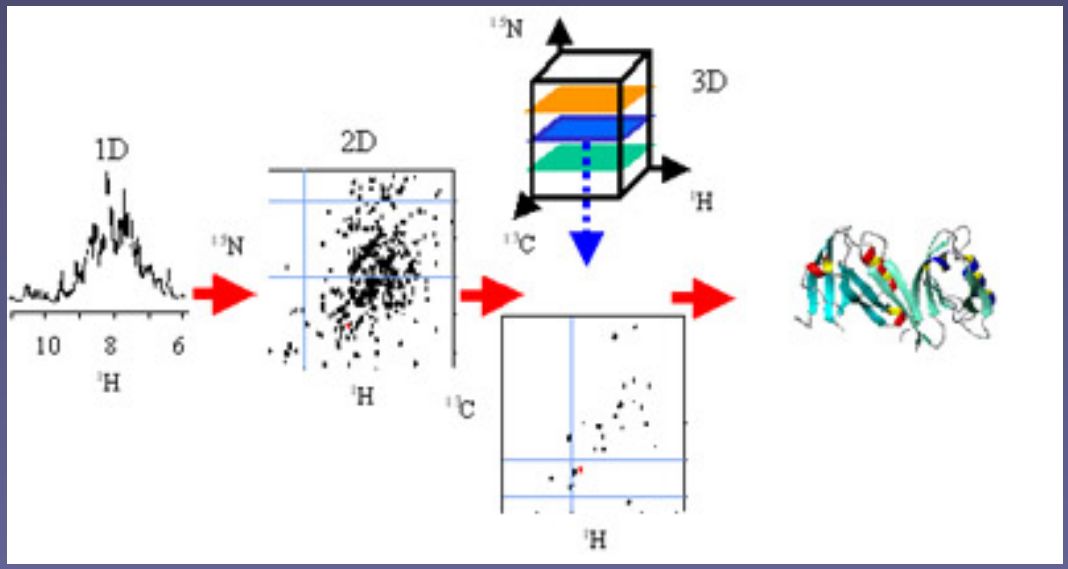
Structure Determination: Experimental Methods

X-ray crystallography



http://www.uni-duesseldorf.de/home/Fakultaeten/math_nat/Graduiertenkollegs/biostruct/Research/BioStruct_Groups/AG_Groth/expertise.html

NMR



http://www.dbs.nus.edu.sg/staff/henry.htm

RCSB Protein Data Bank - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.pdb.org/pdb/home/home.do Go Links

MyPDB Login

A MEMBER OF THE PDB

RCSB PDB PROTEIN DATA BANK

An Information Portal to Biological Macromolecular Structures

As of Tuesday Aug 10, 2010 at 5 PM PDT there are 67131 Structures | PDB Statistics

HELP | PRINT PDB ID or Text Search | Advanced Search

Customize This Page

Home Hide

- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- External Links
- New Website Features

Deposition Hide

- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamline
- Related Tools

Search Hide

- Advanced Search
- Latest Release
- Latest Publications
- Sequence Search
- Chemical Components
- Unreleased Entries

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

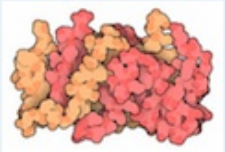
The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Hide Welcome Message

Featured Molecules (Previous Features: MOM | PSI) Hide

Molecule of the Month:

Interferons



Our cells have many defenses against viruses. When cells are infected, they build enzymes that slow protein synthesis, and thus also slow down viral growth, and they build enzymes to chop up double-stranded RNA, which is made primarily by viruses. Infected cells also alert the immune system by displaying pieces of the virus on their surfaces. In the worst cases, infected cells make the ultimate sacrifice and destroy themselves by apoptosis.

New Features Hide

Latest features released:

Website Release Archive:

News Hide

Weekly | Quarterly | Yearly

Statement on Retraction of PDB Entries

2010-08-10

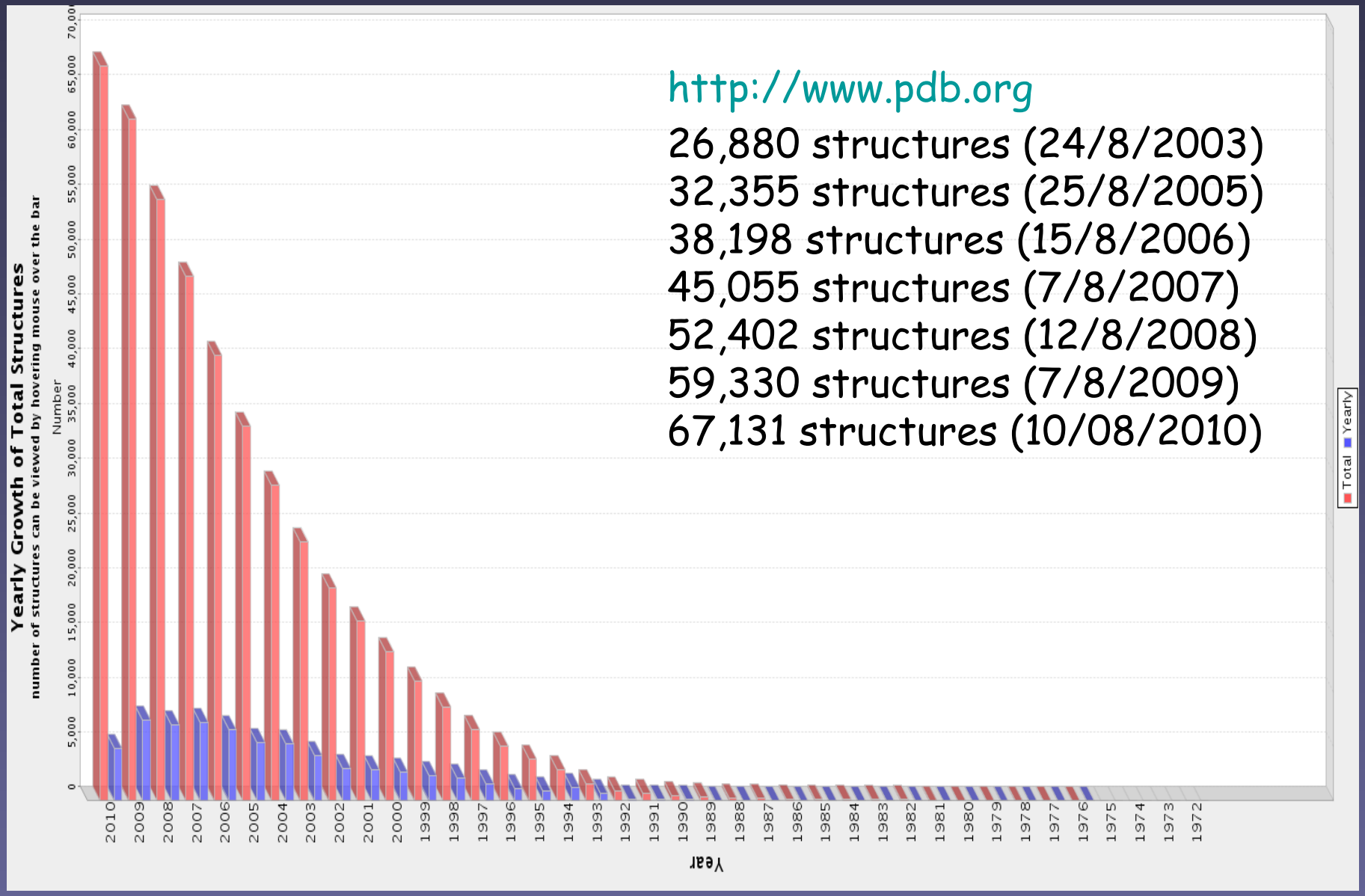
ADIT 2.0 Offers Improved Deposition Process

AD IT! Auto Dep Input Tool

A new and improved version of ADIT, a tool for validating

Done Internet

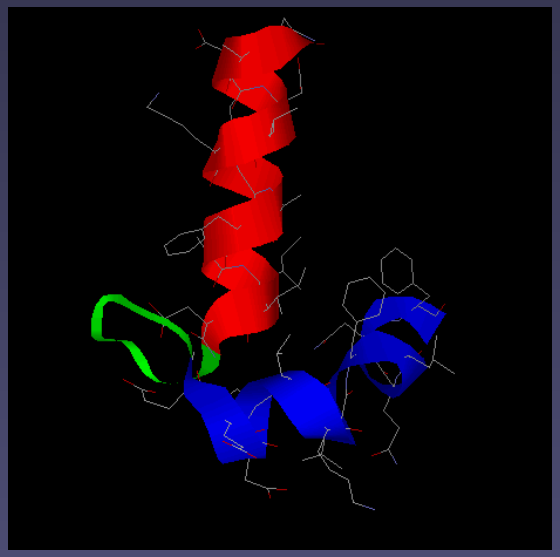
Growth of Protein Data Bank

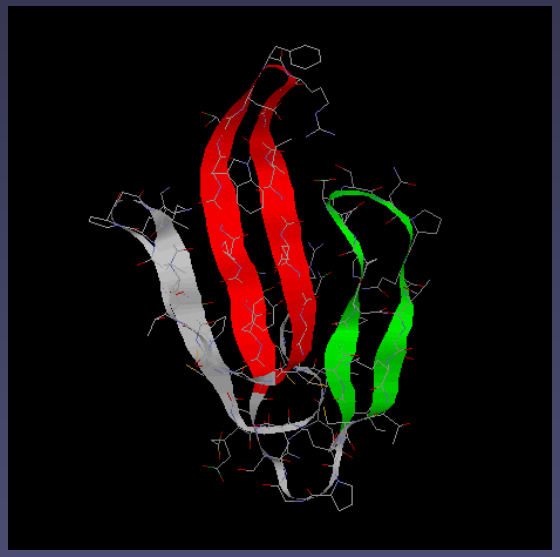


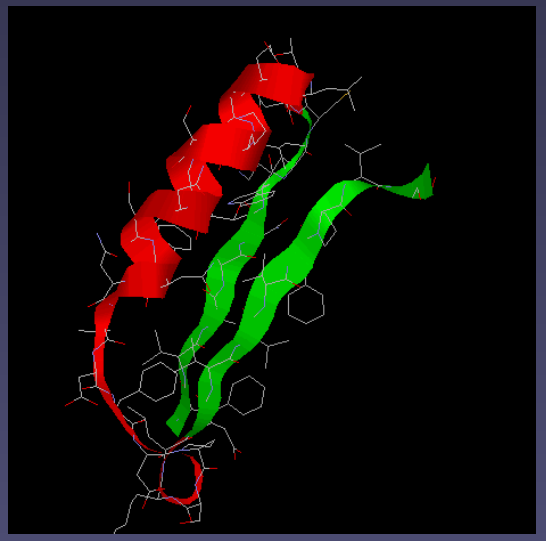
http://www.pdb.org

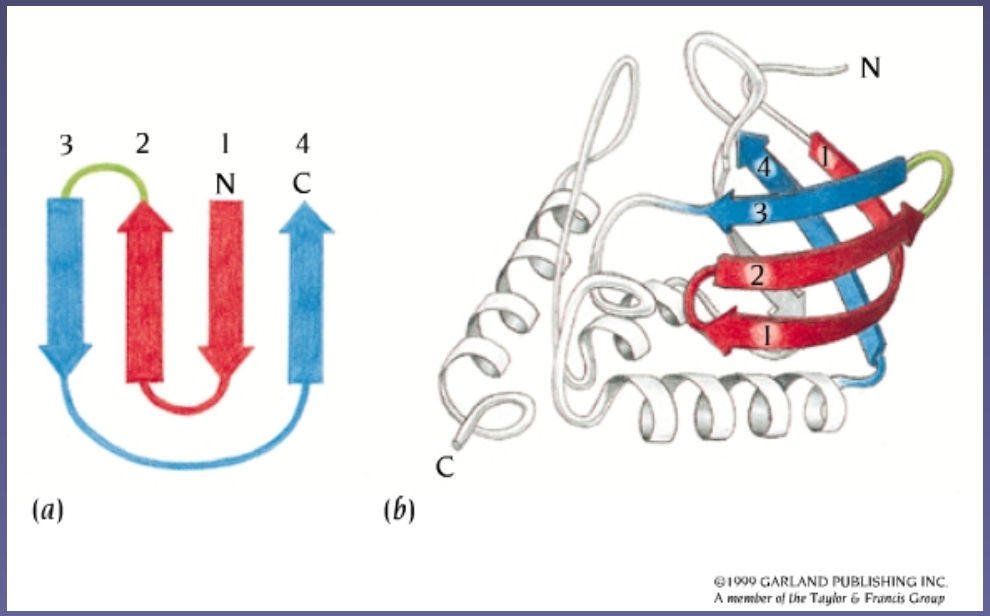
26,880 structures (24/8/2003)
32,355 structures (25/8/2005)
38,198 structures (15/8/2006)
45,055 structures (7/8/2007)
52,402 structures (12/8/2008)
59,330 structures (7/8/2009)
67,131 structures (10/08/2010)

Motifs

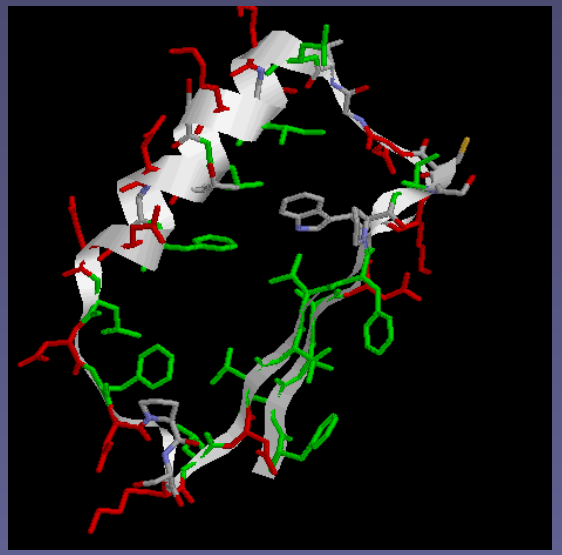








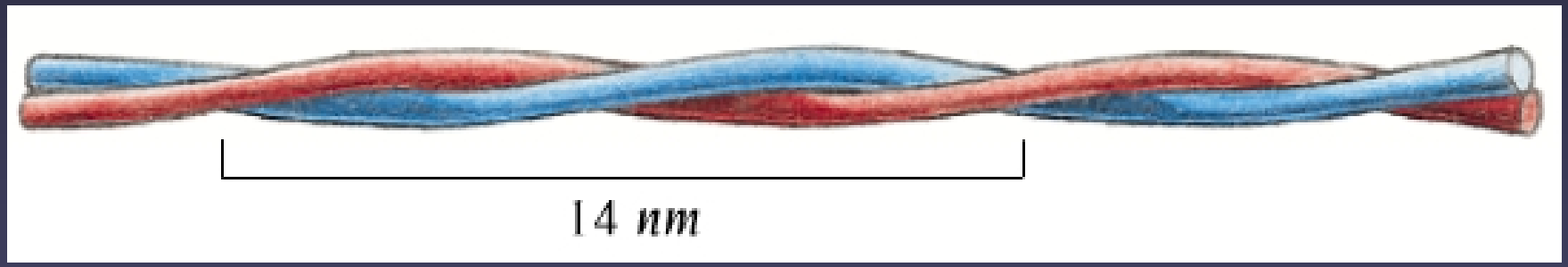
©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

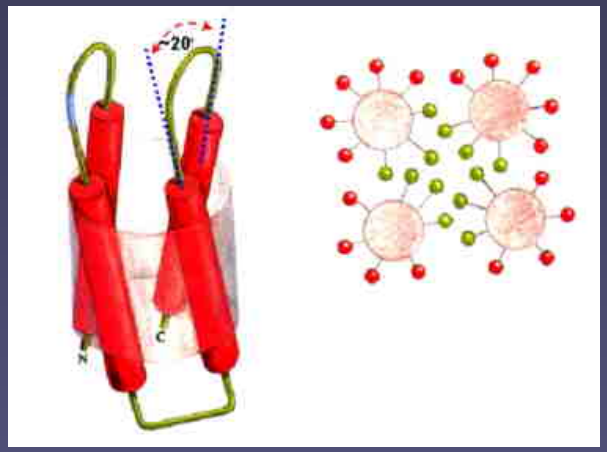


Main Classes of Protein Structures

- ✓ α domains — α-helices
- ✓ β domains — Antiparallel β-sheets
- ✓ α/β domains — Combinations of β-α-β motifs
- ✓ α + β domains — Discrete α and β motifs
- ✓ Disulfide bonds/metal atoms

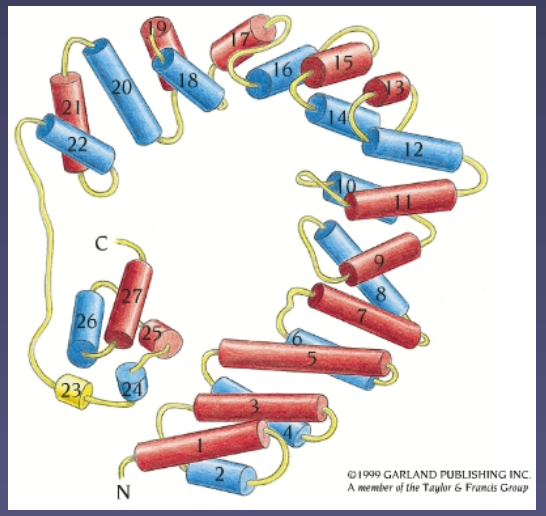
Coiled-coil



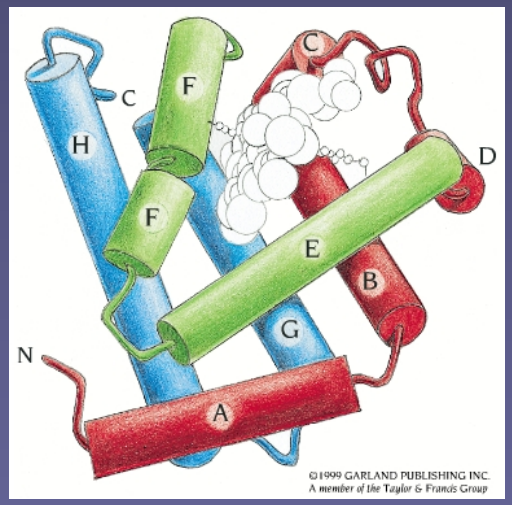


Four-helix bundle

Alpha-domain



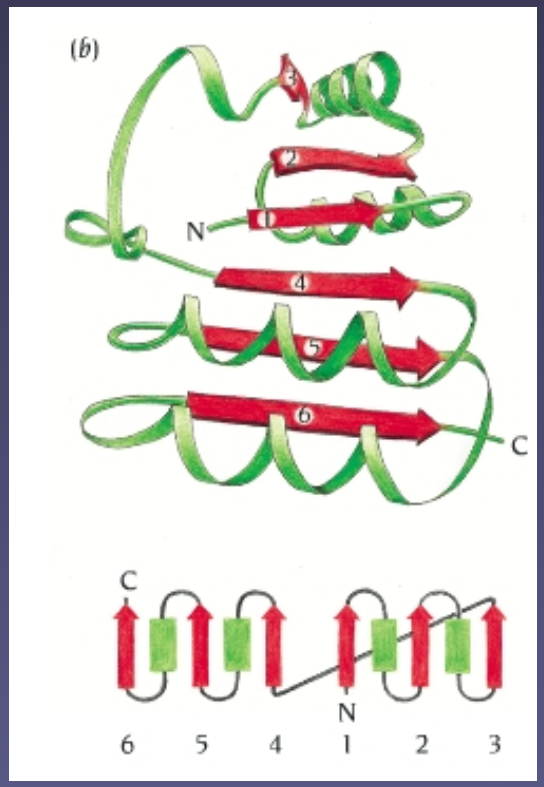
Large alpha-helical domain

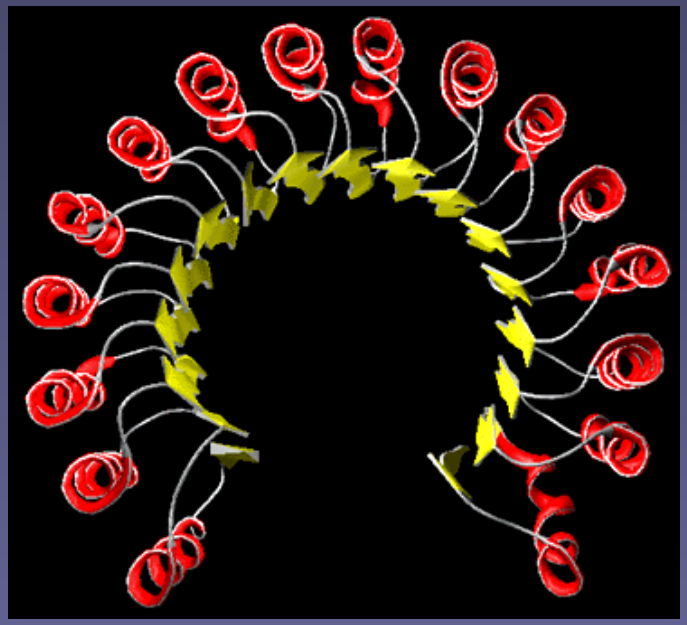


Globin fold

TIM-barrel

Rossman fold

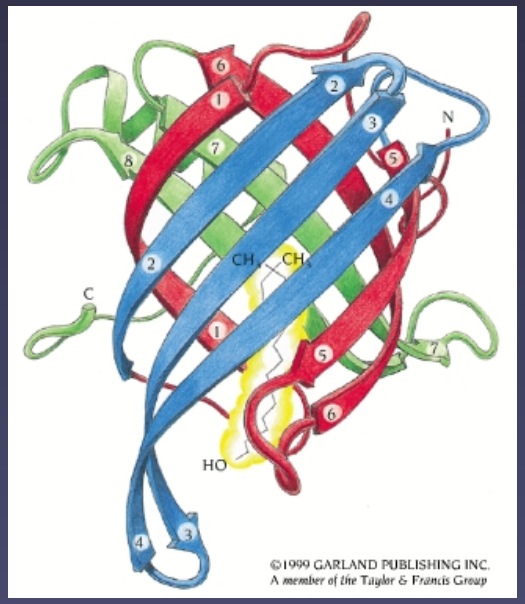


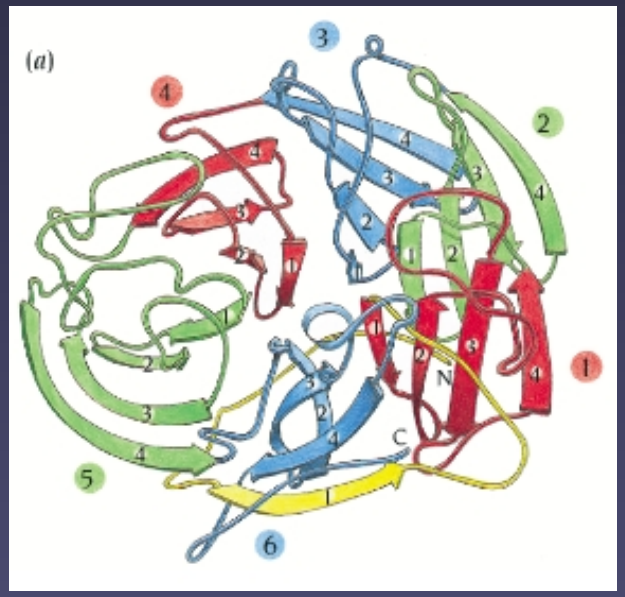


Horseshoe fold

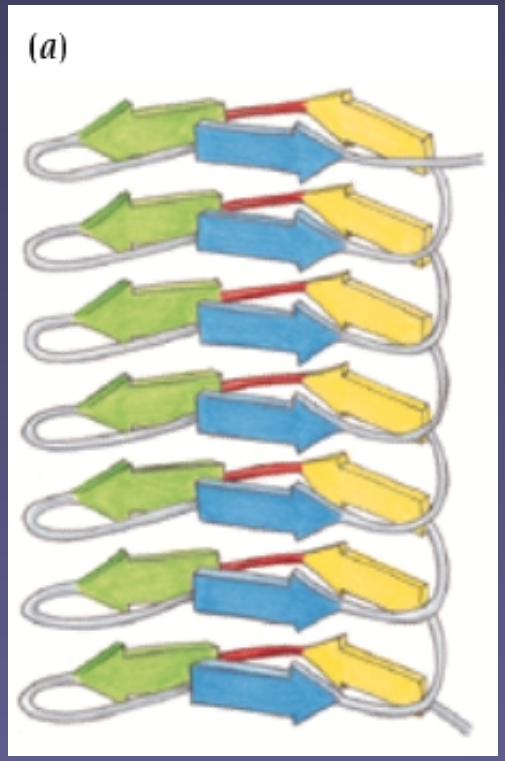
α/β structures

β-domain

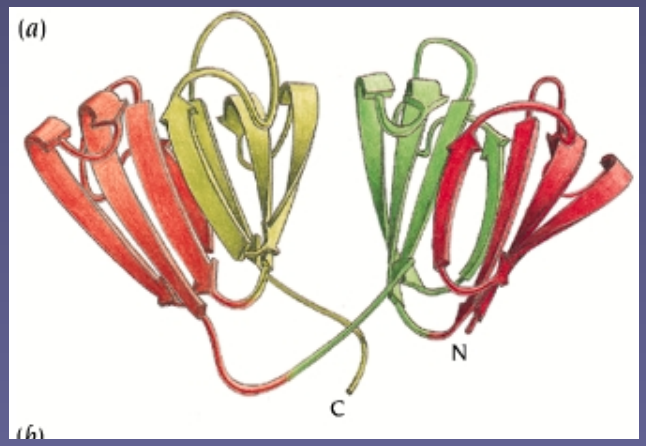




Up-and-down beta-barrel



Beta-helix

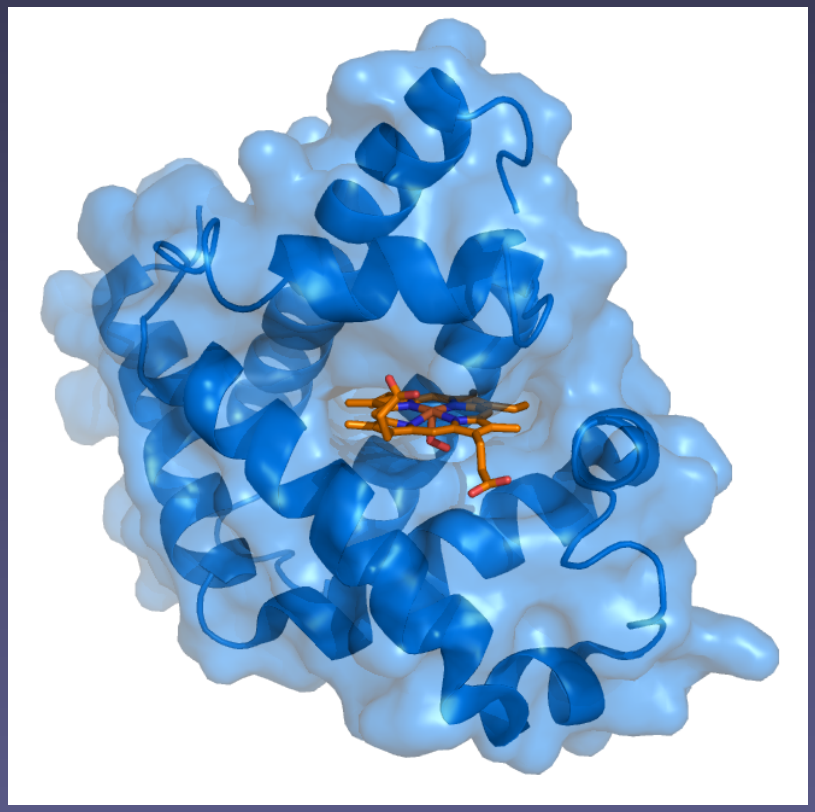


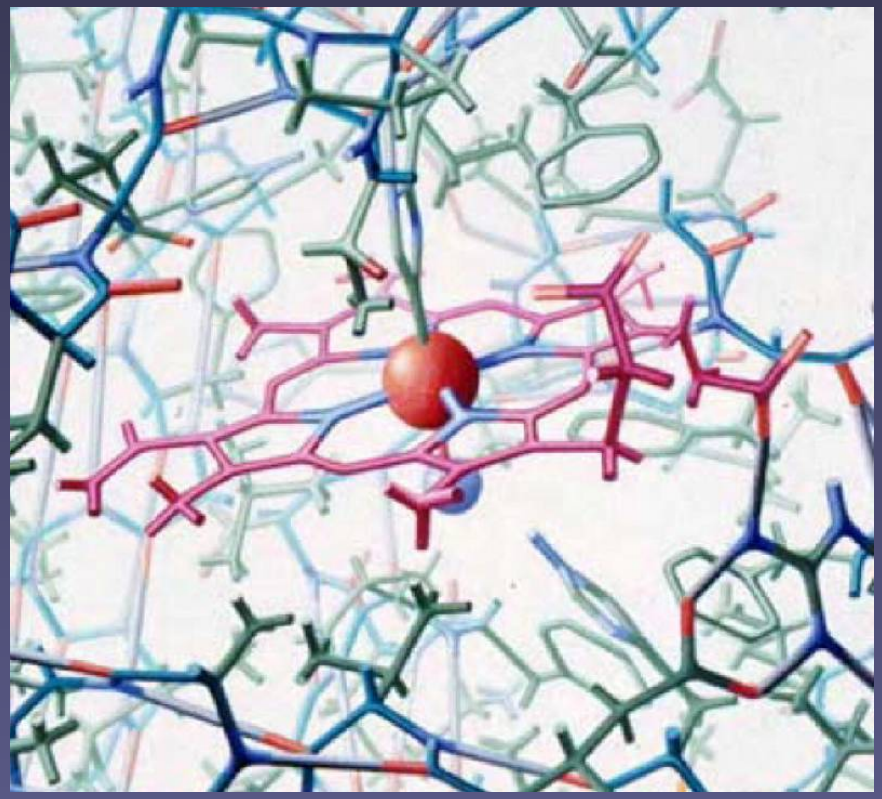
Greek-key

Is knowledge of 3-D structure enough to understand the function?

What we don't know?

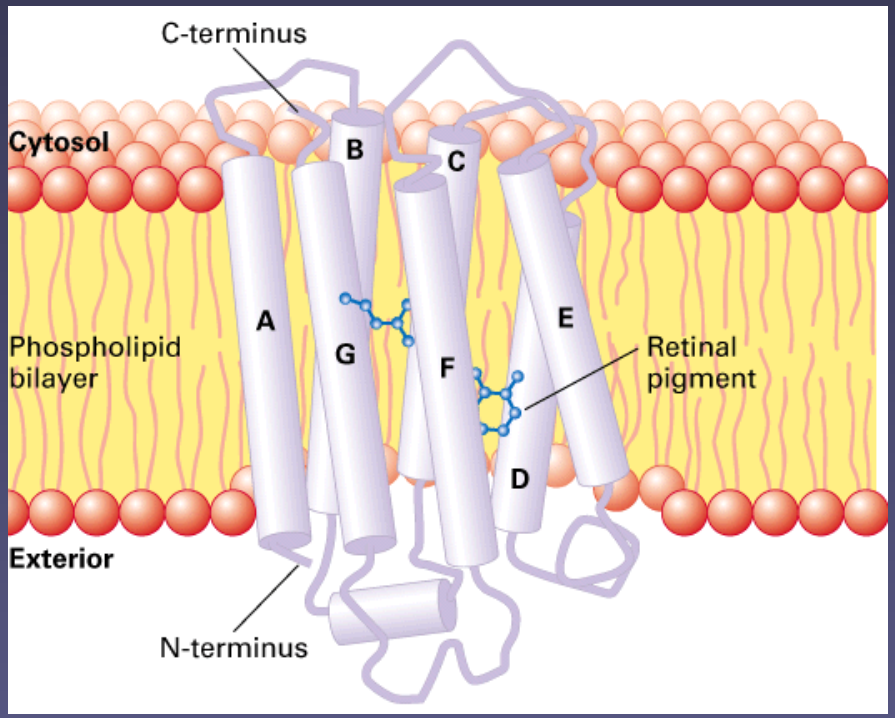
Example 1: Myoglobin

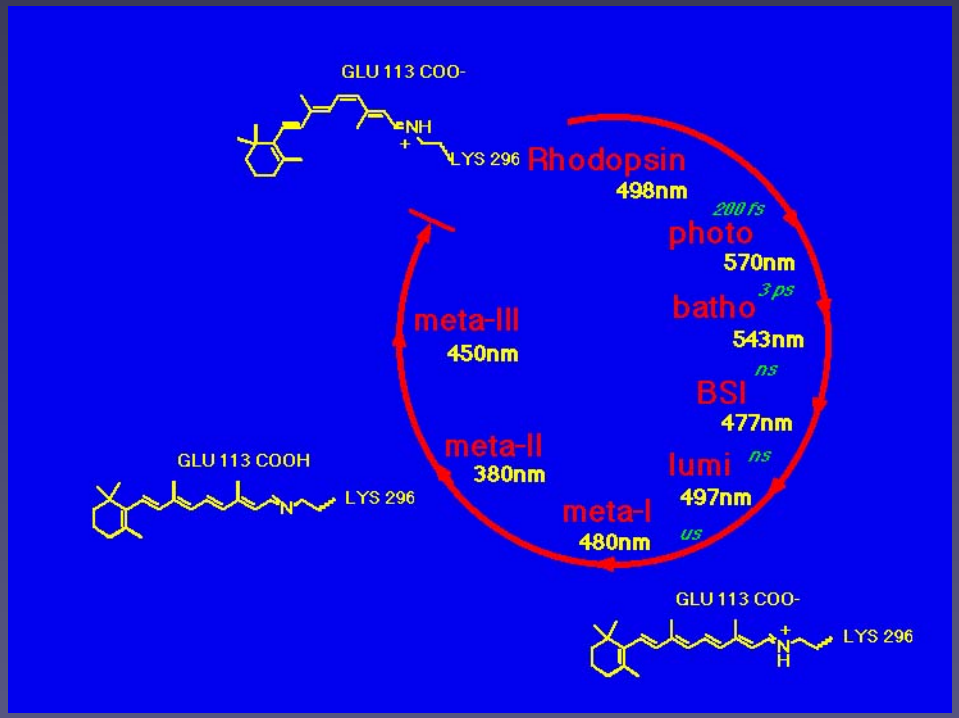




Breathing motions in myoglobin opens up pathways for oxygen atoms to enter its binding site or diffuse out

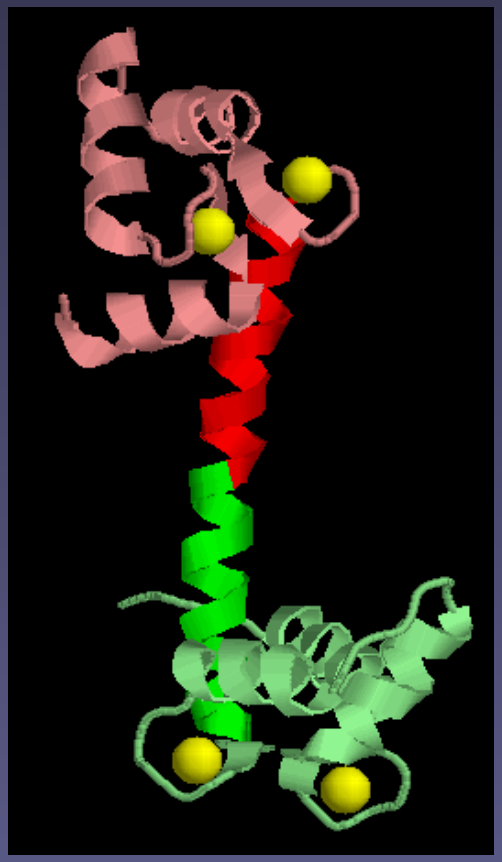
Example 2: Rhodopsin

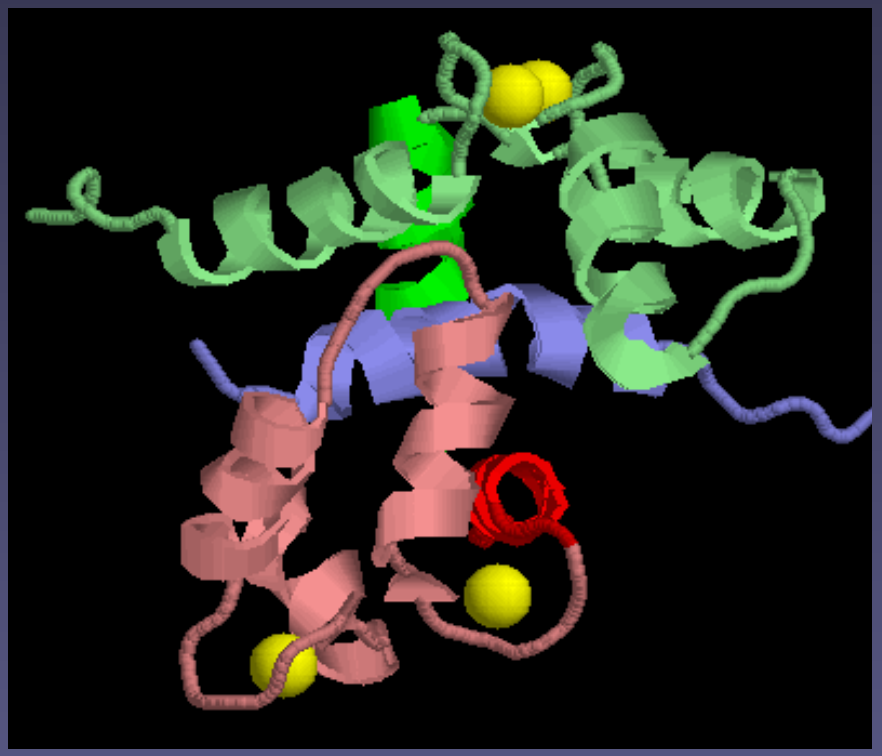




GPCRs like rhodopsin undergo conformational changes during signal transduction

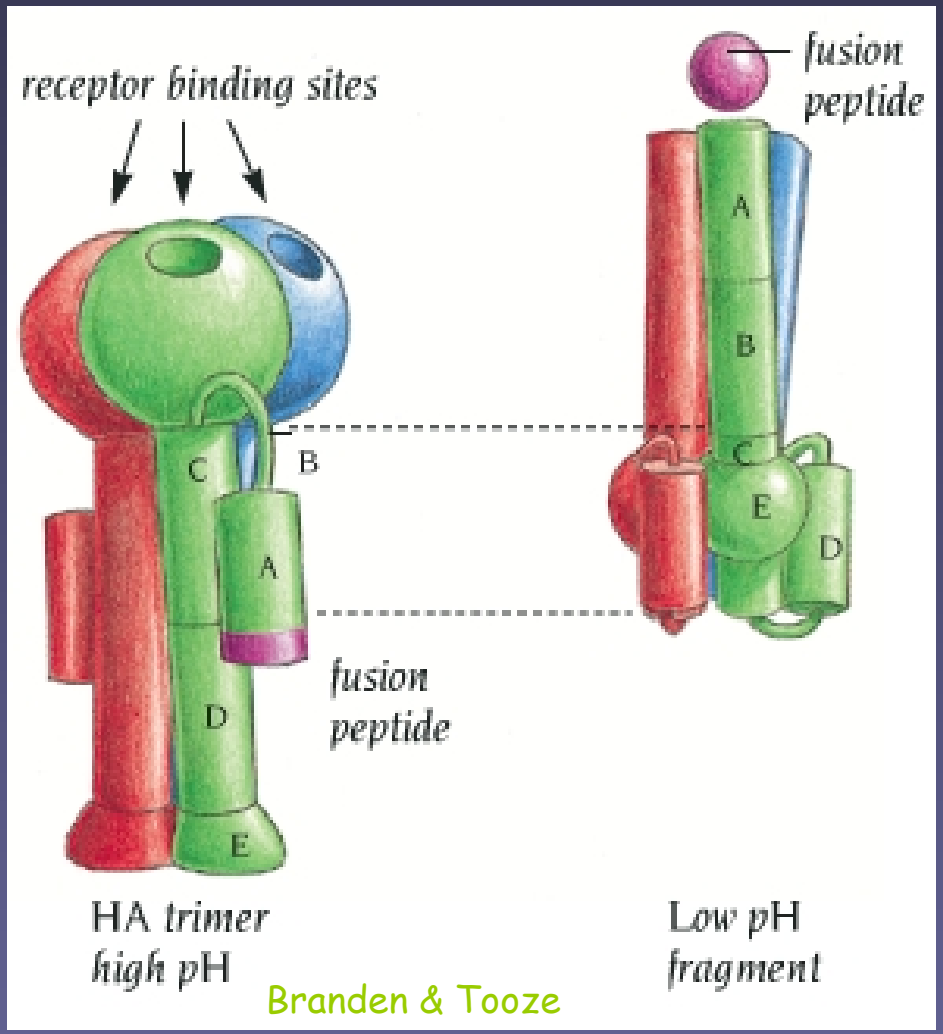
Example 3: Calmodulin





Largest ligand-induced interdomain motion known in proteins

Example 4: Hemagglutinin



Hemagglutinin from influenza virus undergoes large conformational changes

At low PH, the N-terminal helix moves 100 Å to bring the fusion peptide closer to the host cell membrane

Why Molecular Dynamics?

- Experimentally determined structures are static
- They represent the average structure of an ensemble of structures
- They do not provide the dynamic picture of a biomolecule
- Molecular dynamics is one way to understand the conformational flexibility of a biomolecule and its functional relevance

Biological molecules exhibit a wide range of time scales over which specific processes

- Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)
  - Atomic fluctuations
  - Sidechain Motions
  - Loop Motions
- Rigid Body Motions (1 to 10Å, 10^{-9} to 1s)
  - Helix Motions
  - Domain Motions (hinge bending)
  - Subunit motions
- Large-Scale Motions (> 5Å, 10^{-7} to 10^{4} s)
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding

http://cmm.info.nih.gov/modeling/guide_documents/molecular_dynamics_document.html

Potential Energy Function (Equations)

- Potential Energy is given by the sum of these contributions:

$$V_{bonded}(R) = \sum_{bonds} k_l (l - l_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{impropers} k_\omega (\omega - \omega_0)^2 + \sum_{torsions} A_n [1 + \cos(n\phi - \phi_0)]$$

$$V_{nonbonded}(R) = \sum_{i<j} (\varepsilon_{ij} [(\frac{r_{ij}^{\min}}{r_{ij}})^{12} - 2(\frac{r_{ij}^{\min}}{r_{ij}})^{6}] + \frac{q_i q_j}{4\pi\varepsilon_r \varepsilon_0 r_{ij}}$$

Molecular Dynamics

- Calculate Energy 'E' using the potential Energy function
- Calculate Force by differentiating the potential Energy
- Calculate Acceleration 'a' using Newton's second Law
- Calculate Velocity at a later time 't+dt'
- Calculate Position at a later time 't+dt'
- Calculate Energy at new position.
- Create a Trajectory by repeating the above steps 'n' number of times.

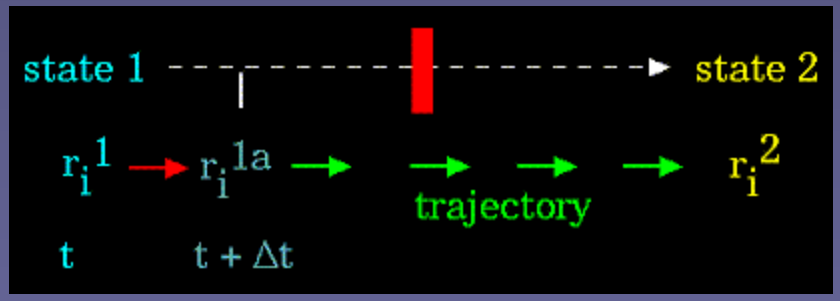
$$-\frac{dE}{dr_i} = F_i$$

force = mass x acceleration ($F_i = m_i a_i$)

$$a_i = \frac{dv_i}{dt}$$

$$v_i = \frac{dr_i}{dt}$$

$$-\frac{dE}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$



http://cmm.info.nih.gov/modeling/guide_documents/molecular_dynamics_document.htm

Some Popular Simulation Force Fields

- AMBER (Assisted Model Building with Energy Refinement)
- CHARMm (Chemistry at HARvard Macromolecular Mechanics)
- CVFF (Consistent-Valence Force Field)
- GROMOS (GROningen MOlecular Simulation package)
- OPLS (Optimized Potentials for Liquid Simulations)

First Biomolecular simulation was performed in 1977

Dynamics of folded proteins - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.nature.com/nature/journal/v267/n5612/abs/267585a0.html Go Links

npg nature publishing group

nature.com | about npg | news@nature.com | naturejobs | natureevents | help | site index

nature

my account | e-alerts | subscribe | register

SEARCH JOURNAL Go

Journal Home
Current Issue
AOP
Archive

THIS ARTICLE

Download PDF
References

Export citation
Export references

Send to a friend

More articles like this

Table of Contents
| Next >

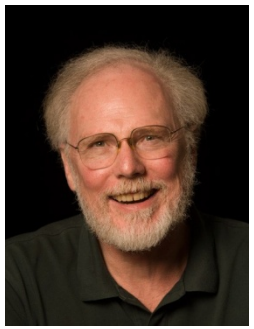
article

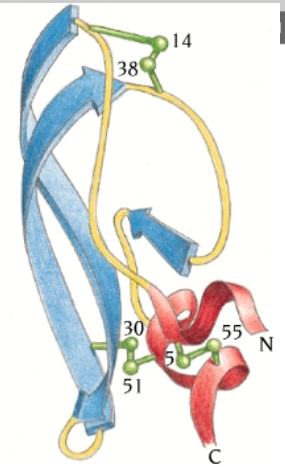
Nature **267**, 585 - 590 (16 June 1977); doi:10.1038/267585a0

Dynamics of folded proteins

J. ANDREW MCCAMMON, BRUCE R. GELIN & MARTIN KARPLUS

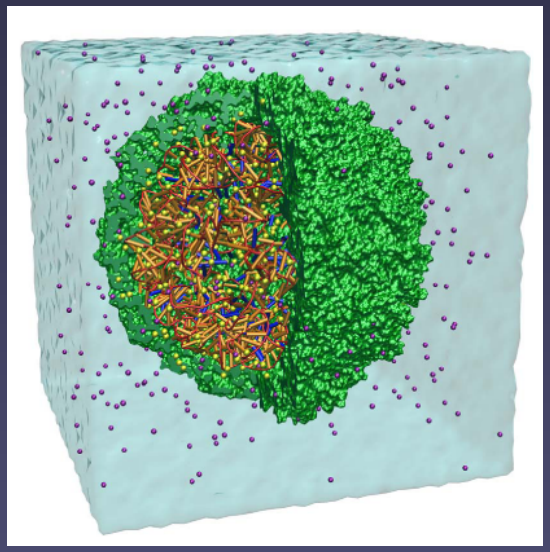
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138



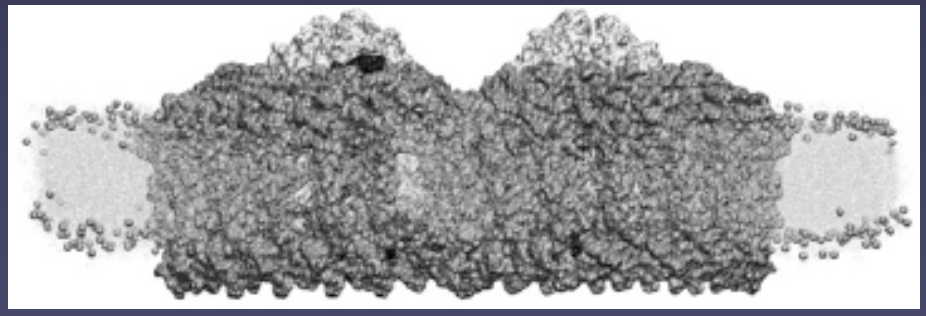


The dynamics of a folded globular protein (bovine pancreatic trypsin inhibitor) have been studied by solving the equations of motion for the atoms with an empirical potential energy function. The results provide the magnitude, correlations and decay of fluctuations about the average structure. These suggest that the protein interior is fluid-like in that the local atom motions have a diffusional character.

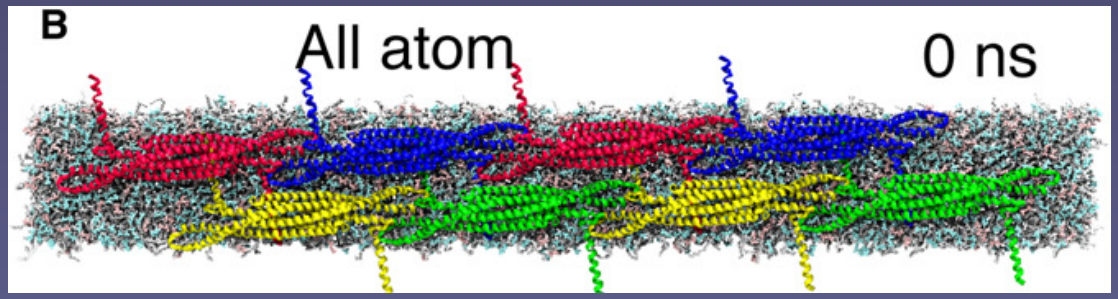
Simulations reaching the million-atom mark



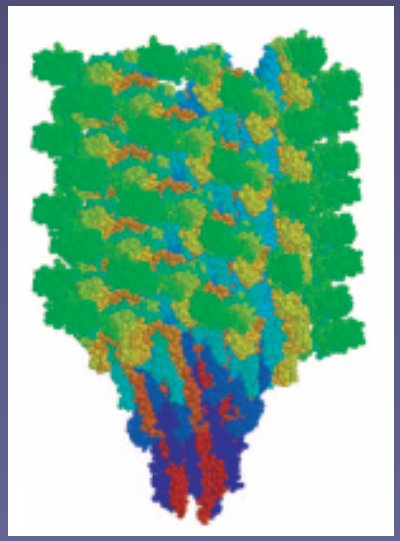
Complete virus: 1 million atoms (Freddolino et al., 2006)



Arrays of light-harvesting proteins – 1 million atoms (Chandler et al., 2008)

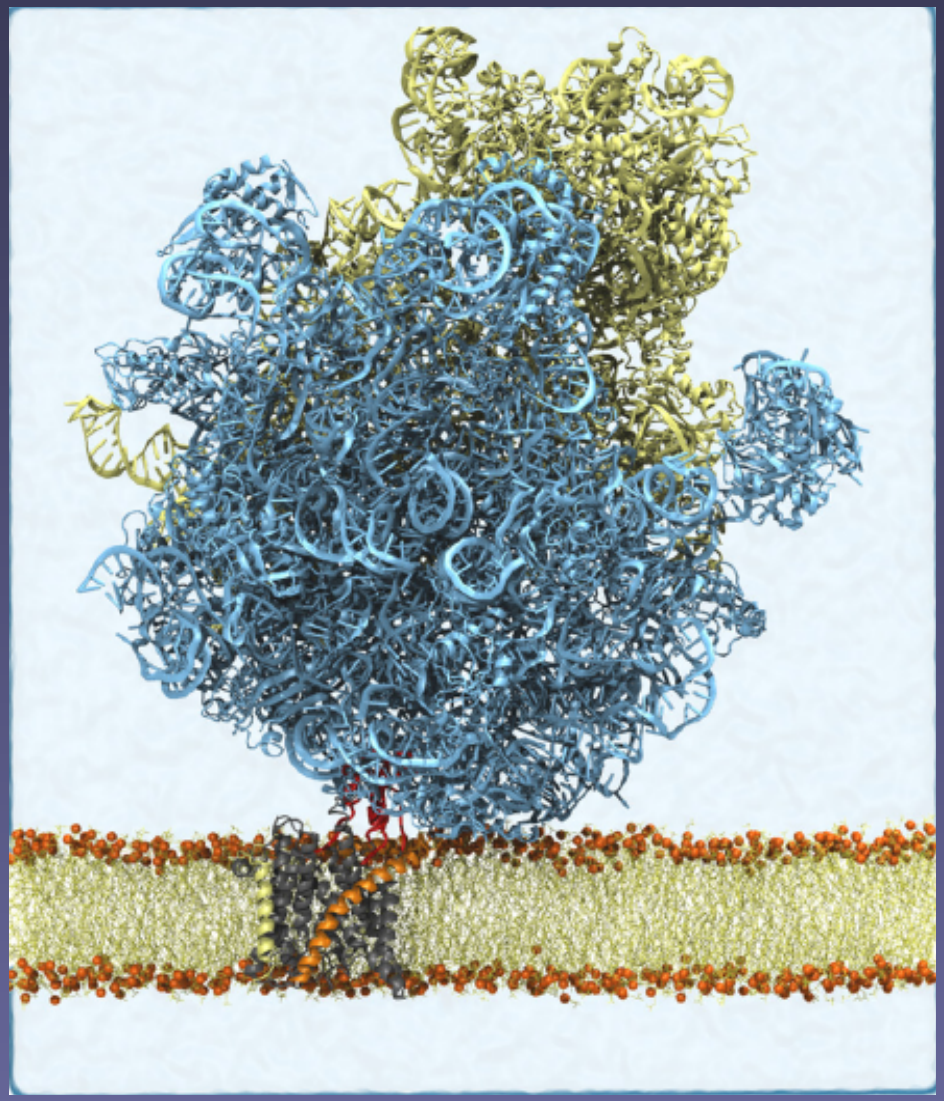


BAR domain proteins – 2.3 million atoms (Yin et al., 2009)



The flagellum – 2.4 million atoms (Kitao et al., 2006)

MD of protein-conducting channel bound to ribosome



Bacterial ribosomes are important targets for antibiotics

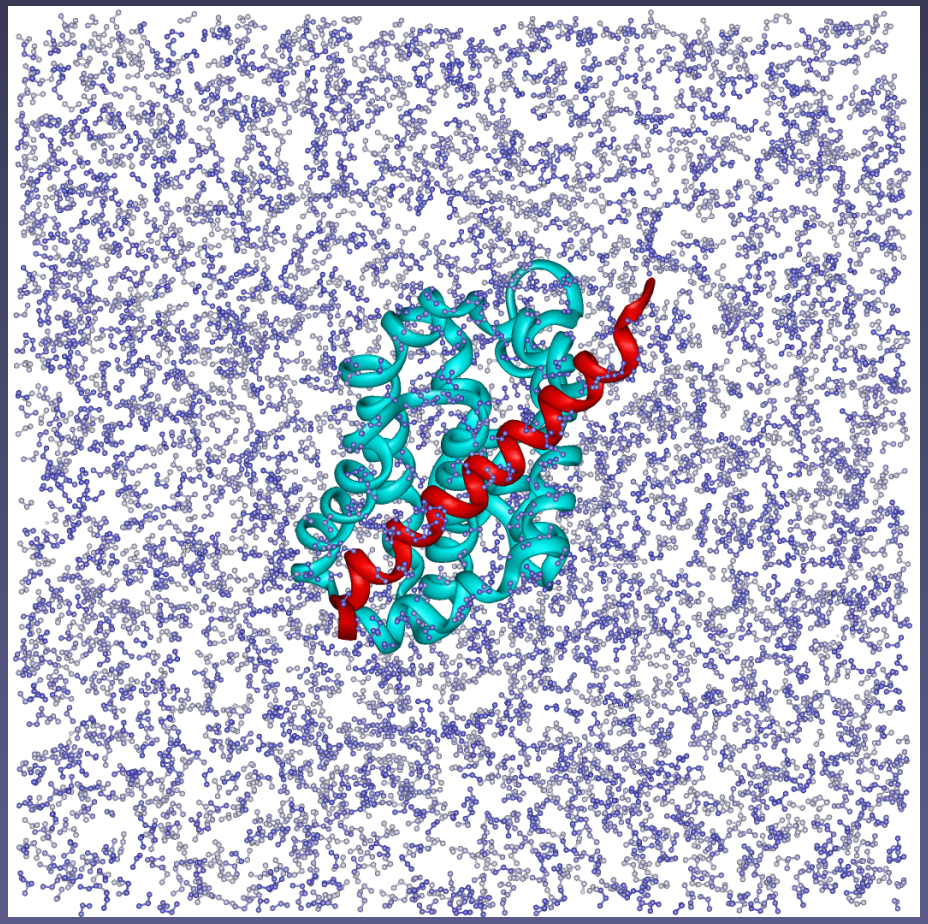
2.7 million atoms

50 ns simulation

Largest system simulated to date

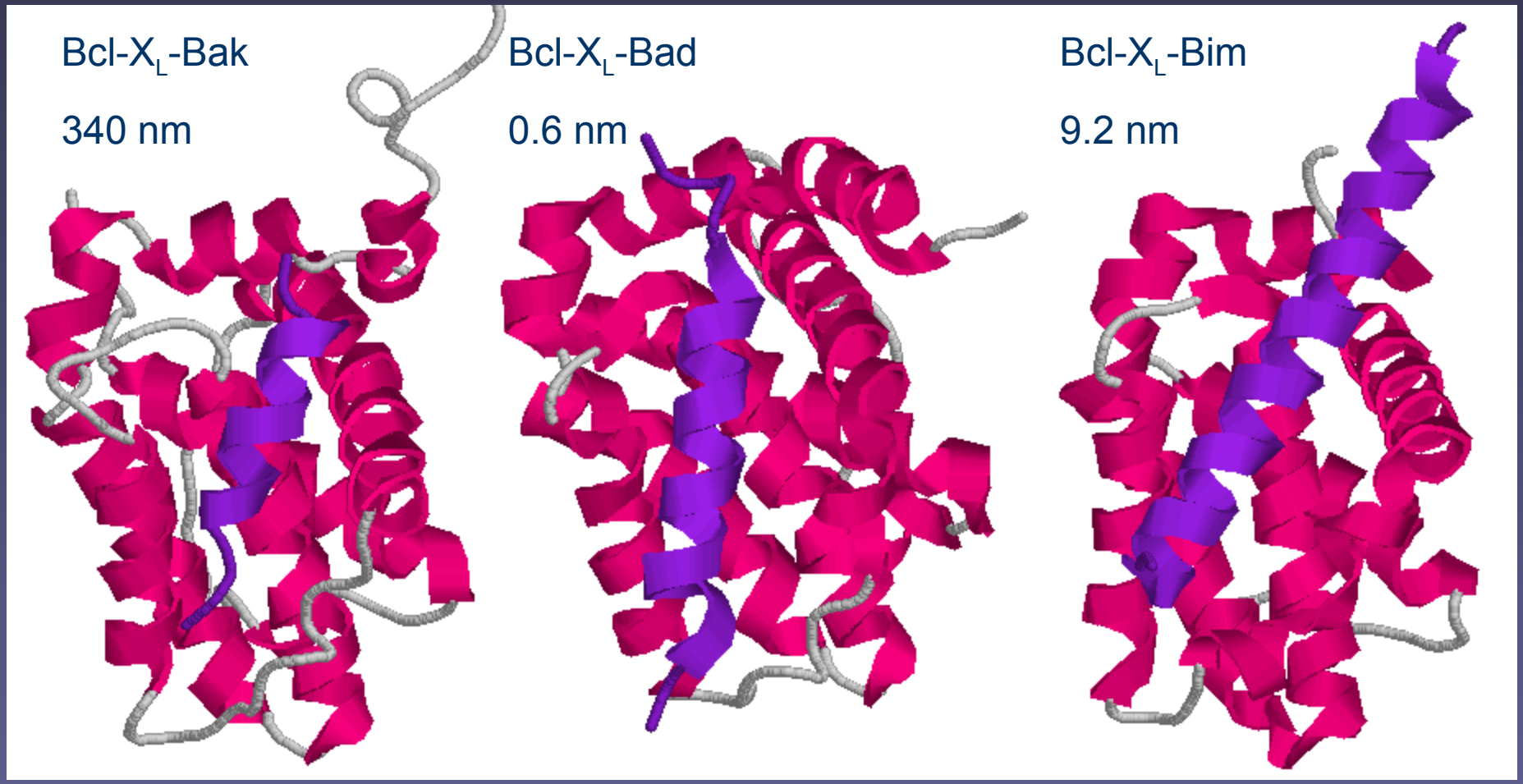
Gumbart et al. (2009)

Biomolecular structures should be simulated under native environment



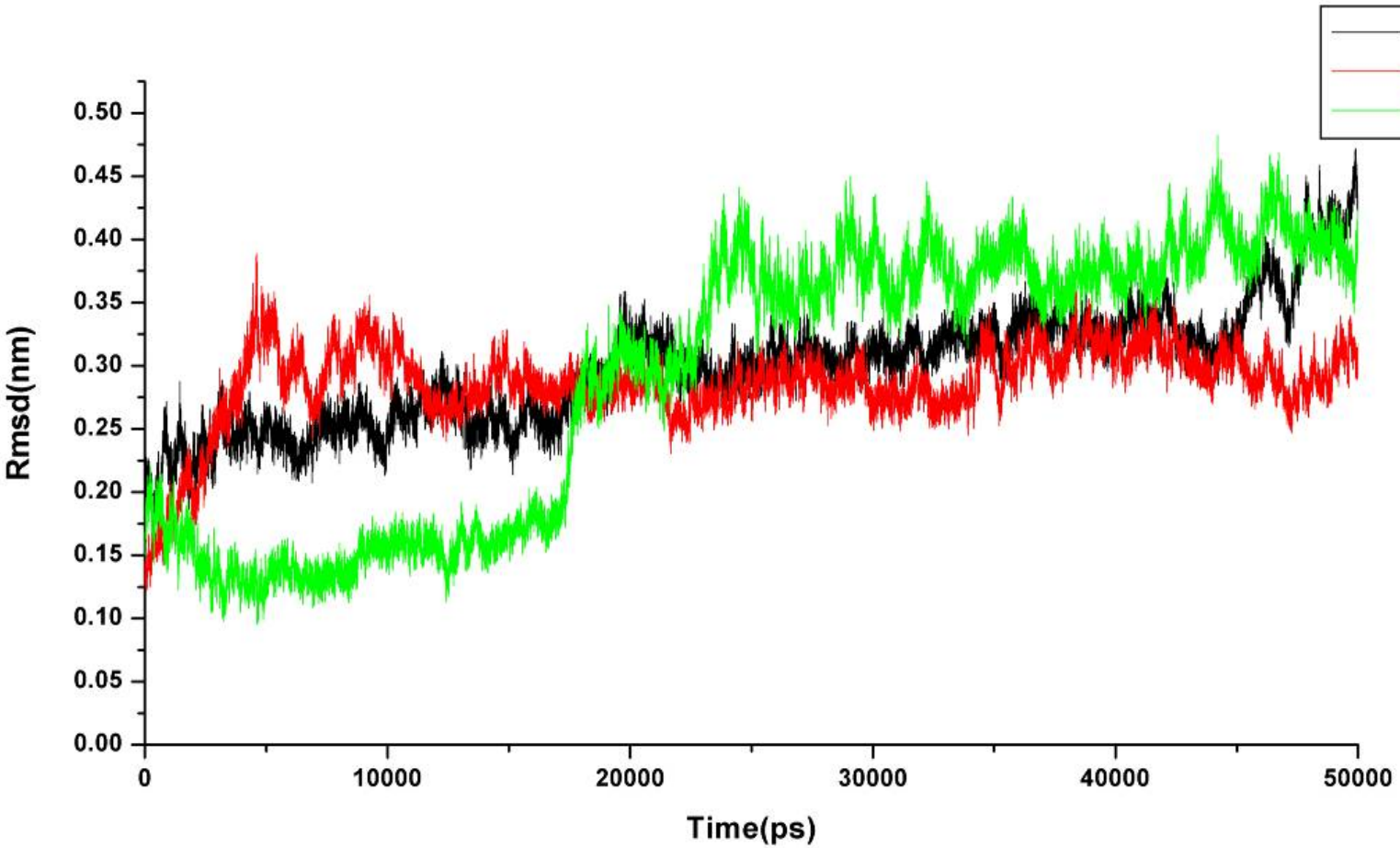
Simulation conditions should be similar to that observed under physiological conditions

Bcl-X_L protein has different affinities for different BH3 pro-apoptotic peptides



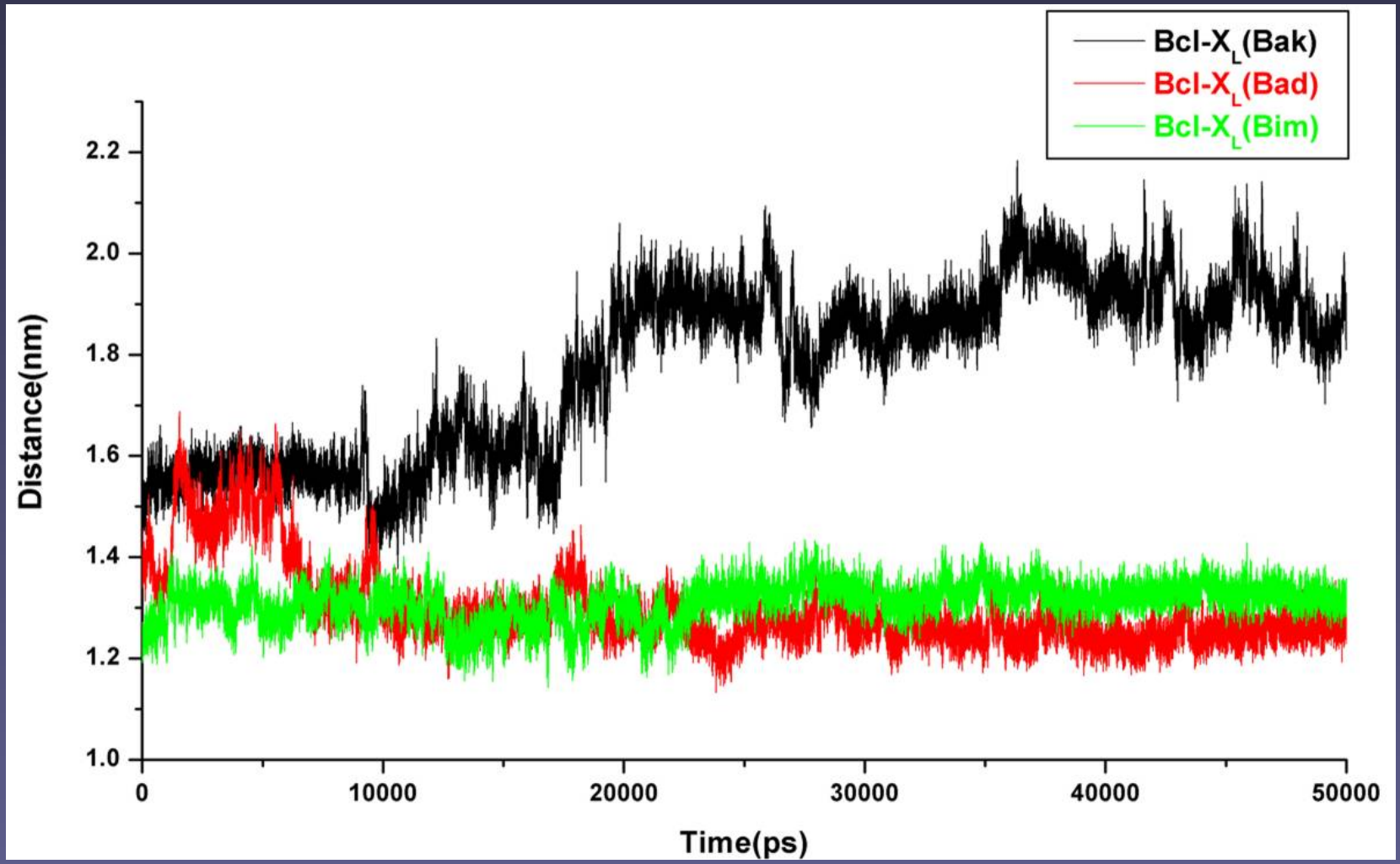
What are the factors that contribute to the different affinities of Bcl-X_L?

RMSD Analysis



Lama and Sankararamakrishnan, *Proteins* (2008)

Distance between helix H3 and the BH3 peptide



Bak peptide moves away from helix H3

Lama and Sankararamakrishnan, *Proteins* (2008)

Protein-peptide interactions

Bcl-X_L residue[a]	Bad residue[b,c]
A89 (H2A)	F162 – 3.64 (85)
E92 (H2)[d]	F162 – 3.47 (97)
A93 (H2)[d]	F162 – 3.63 (88)
F97 (H2B)	F158 – 3.53 (91)
R100 (H2B)	M154 – 3.73 (85); S155 – 4.11 (71)
Y101 (H2B)	L151 – 3.78 (78); M154 – 3.71 (86)
A104 (H2B)	Y147 – 3.21 (98); E150 – 3.89 (76); L151 – 3.92 (71)
F105 (LB)	L151 – 3.79 (82)
S106 (LB)	Y147 – 3.22 (95)
Q111 (LB)	N140 – 3.83 (64); A143 – 3.55 (90); ***Y147 – 3.11 (92)***, W142 – 3.41 (94)
L112 (LB)	A144 – 4.03 (62)
H113 (LB)	N140 – 4.20 (56)
S122 (H3)	N140 – 3.53 (78); A144 – 3.52 (96)
Q125 (H3)	L141 – 3.58 (87); A144 – 3.63 (89); Q145 – 3.19 (99)
V126 (H3)	A144 – 3.37 (98); G148 – 3.68 (87); L151 – 4.13 (50)
E129 (H3)	Q145 – 3.63 (81); G148 – 3.84 (72); R149 – 3.37 (93); ***R152 – 3.07 (99)***
L130 (H3)	G148 – 3.75 (75); L151 – 3.70 (90); R152 – 3.40 (94)
N136 (LC)	***D160 – 3.25 (95)***
W137 (LC)	V159 – 3.36 (97)
G138 (LC)	F158 – 3.62 (79); V159 – 3.78 (80); ***D160 – 3.72 (71)***
R139 (LC)	R152 – 3.62 (80)
A142 (H4)	L151 – 3.71 (86)
E193 (H6)	K164 – 3.64 (80)
L194 (H6)	K163 – 3.27 (99); K164 – 3.31 (95)
Y195 (C-Ter)	***D160 – 3.01 (93)***

Lama and Sankararamakrishnan, *Proteins* (2008)

Acknowledgements

- Anjali Bansal
- Dilraj Lama
- Alok Jain
- Tuhin Kumar Pal
- Priyanka Srivastava
- Vivek Modi
- Ravi Kumar Verma
- Krishna Deepak
- Phani Deep

DST, DBT, CSIR, MHRD