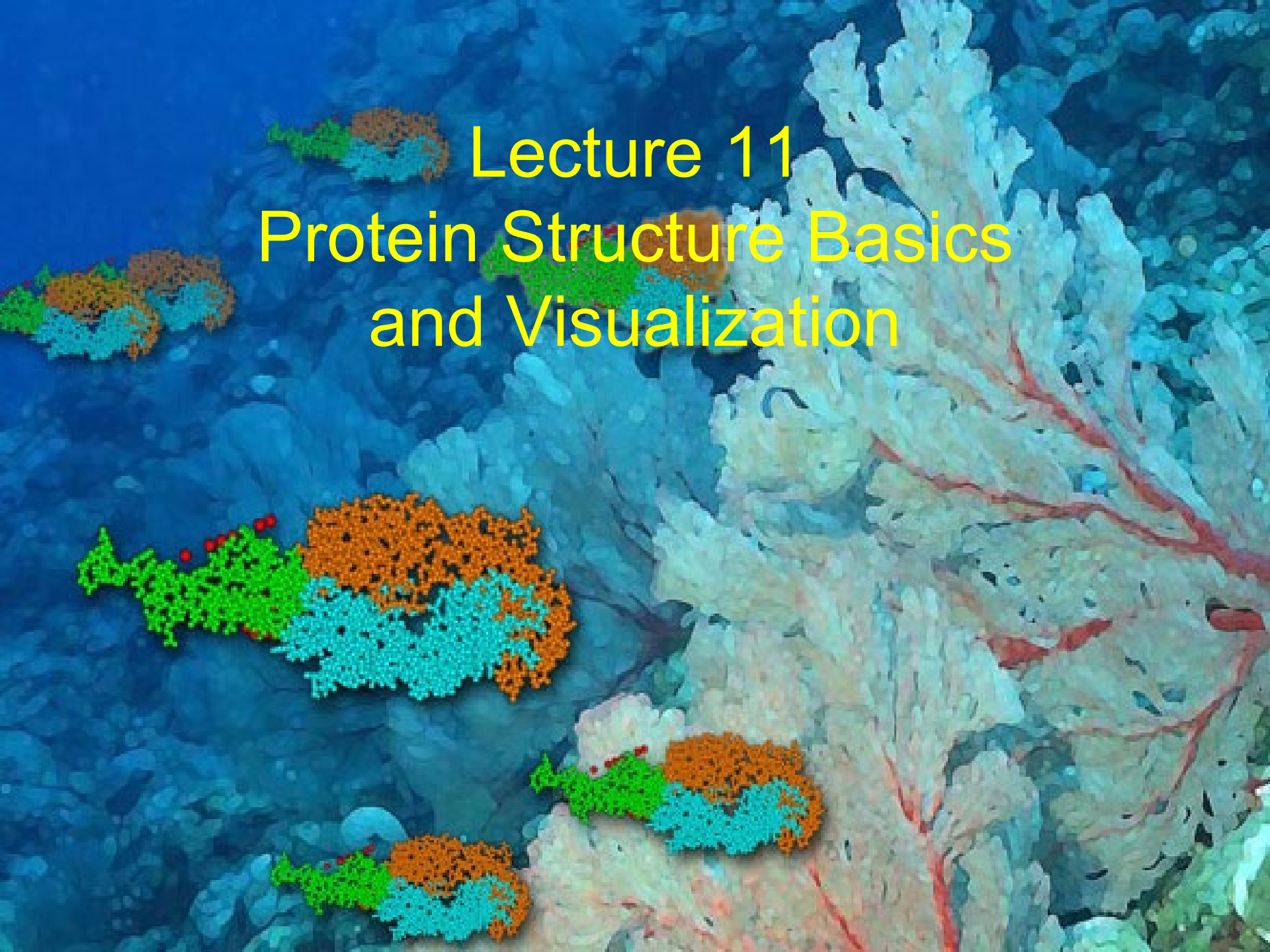

Lecture 11
Protein Structure Basics
and Visualization

Overview

1. Protein structure and basic conception

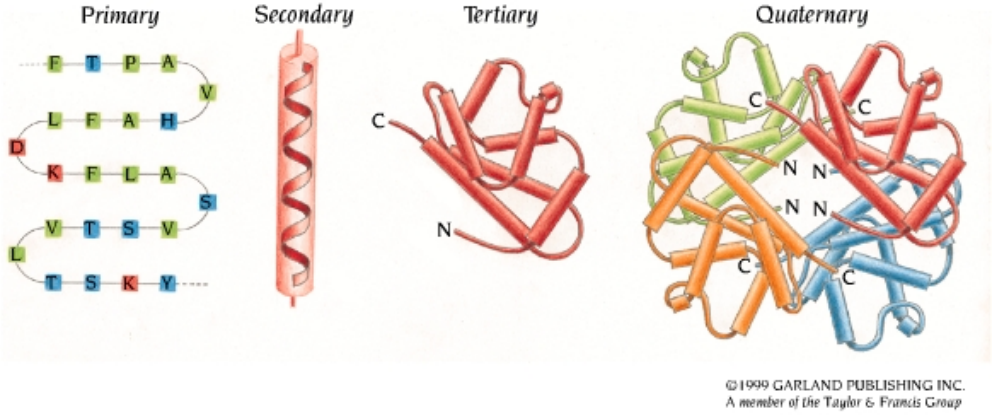
2. Protein structure visualization and classification

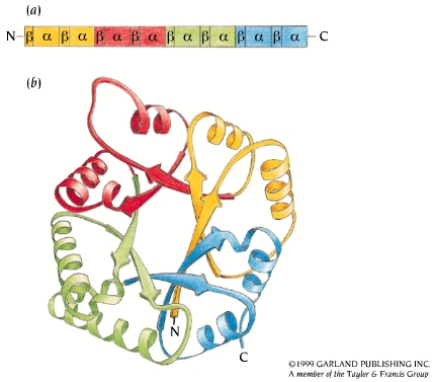
1. Why protein structure?

- In the factory of the living cell, proteins are the workers, performing a variety of tasks
  - *Each protein adopts a particular folding pattern that determines its function*
  - The 3D structure of a protein brings into close proximity residues that are far apart in the amino acid sequence

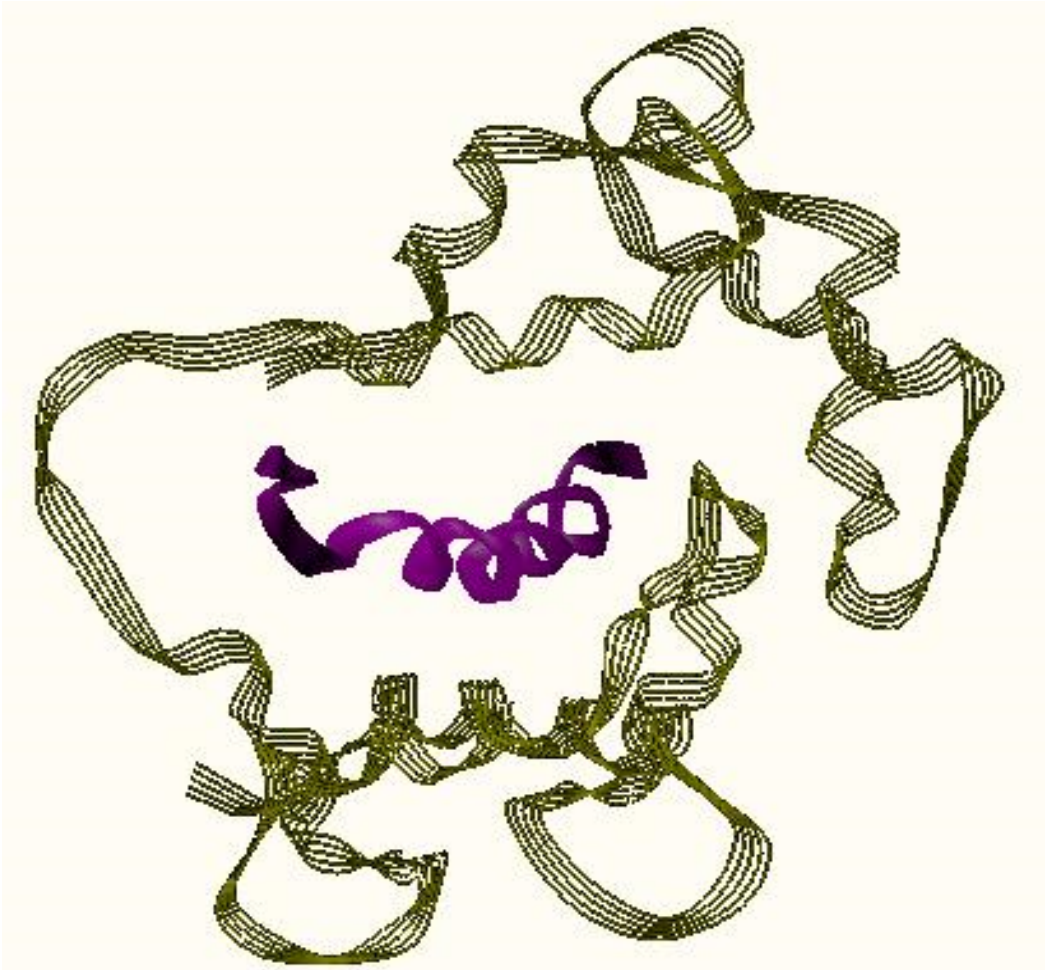
Proteins

- Proteins play a crucial role in virtually all biological processes with a broad range of functions.
- The activity of an enzyme or the function of a protein is governed by the three-dimensional structure



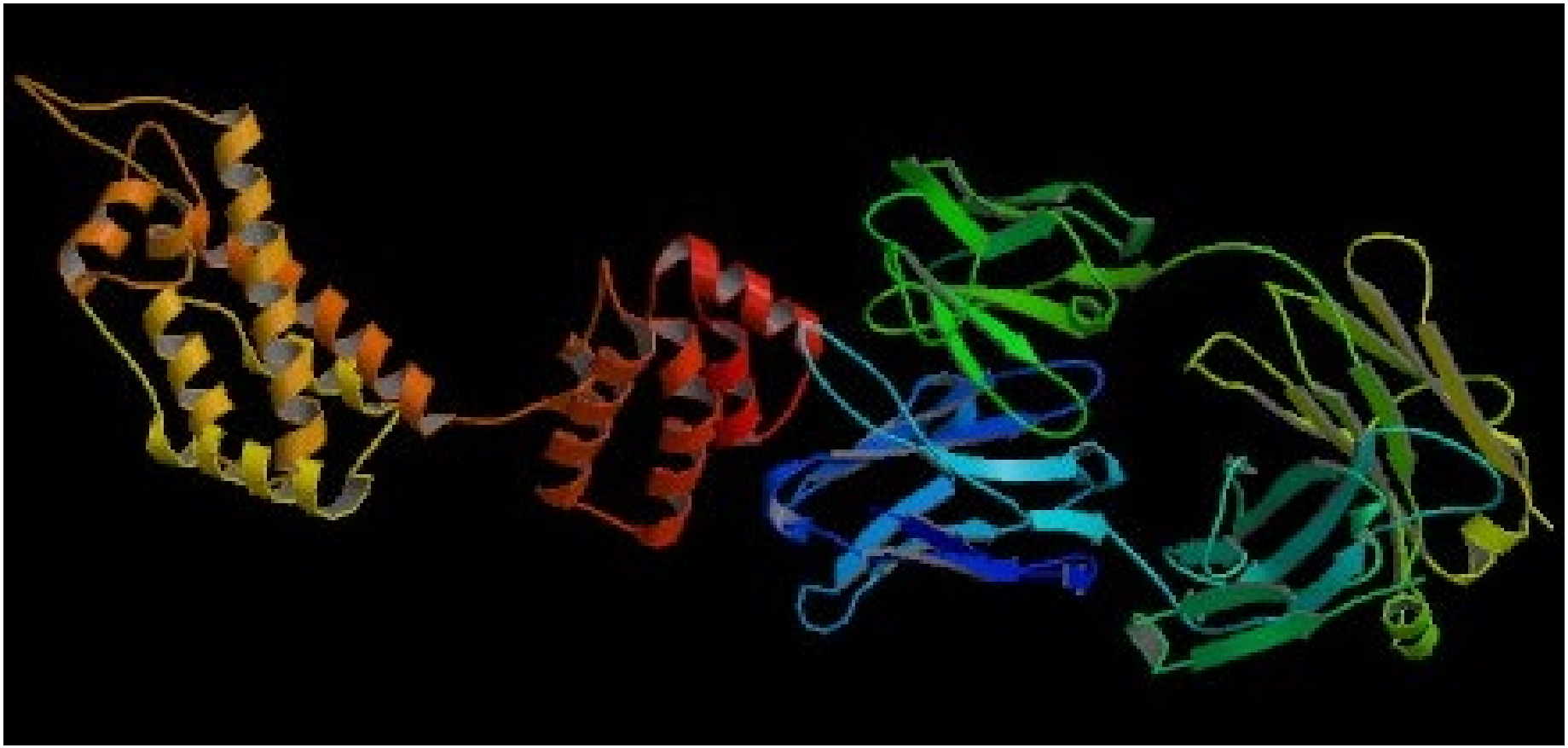


Protein-ligand/substrate interaction



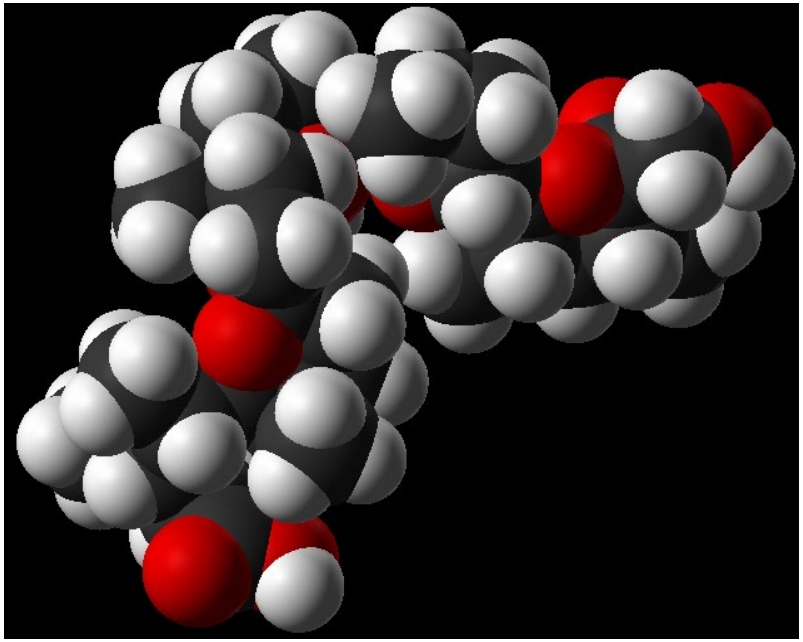
CaM complex with peptide ligand
(PDB code is 2BBM)

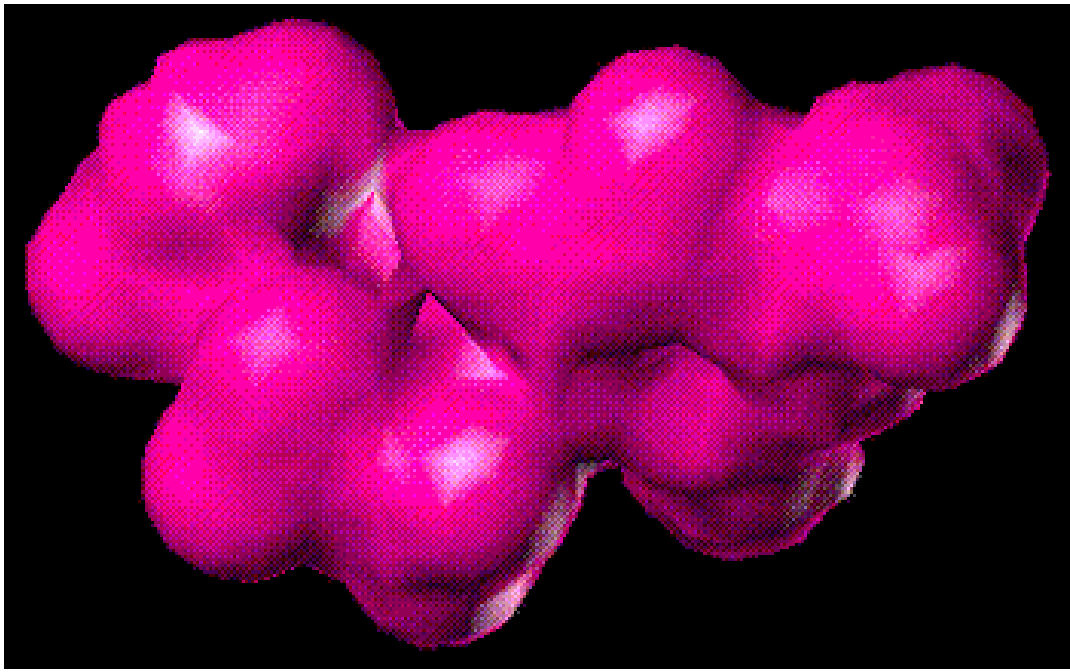
Protein-Protein Interaction



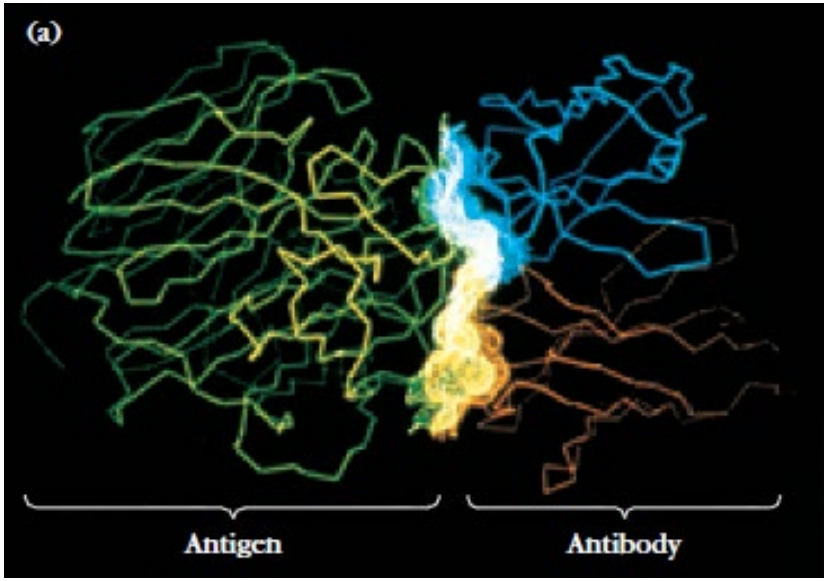
Structural organization of a molecule

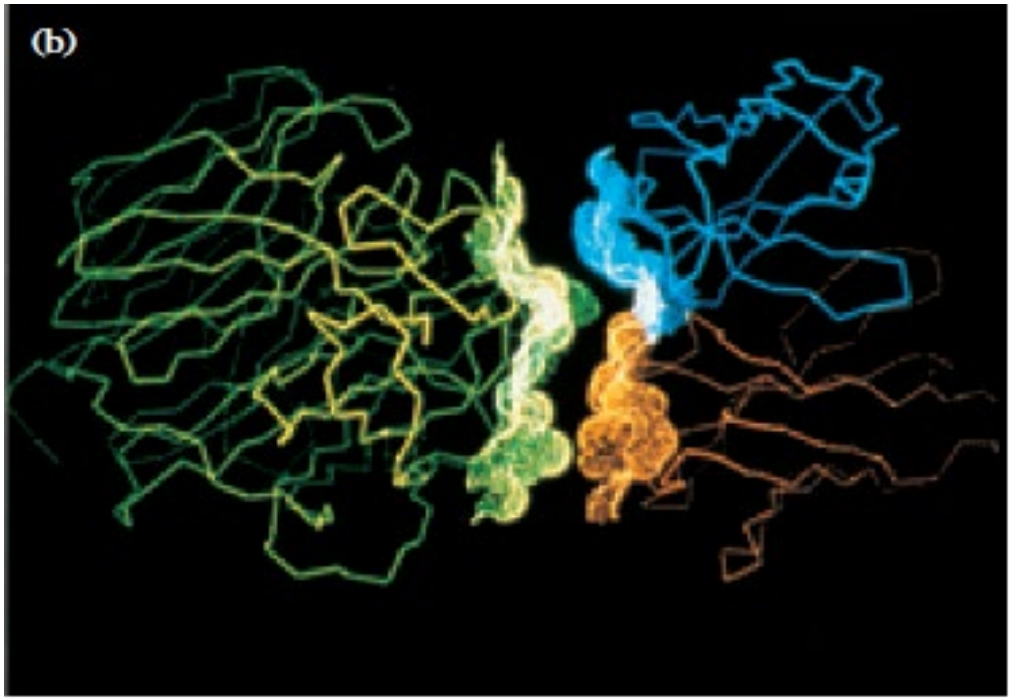
- Surface features (cavities, grooves where other molecules can bind to).





Antibody-antigen complex





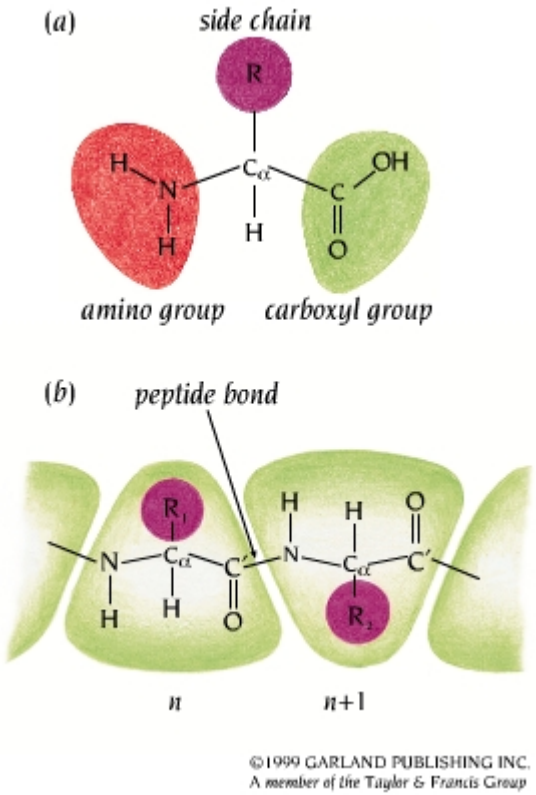
How does a protein fold?

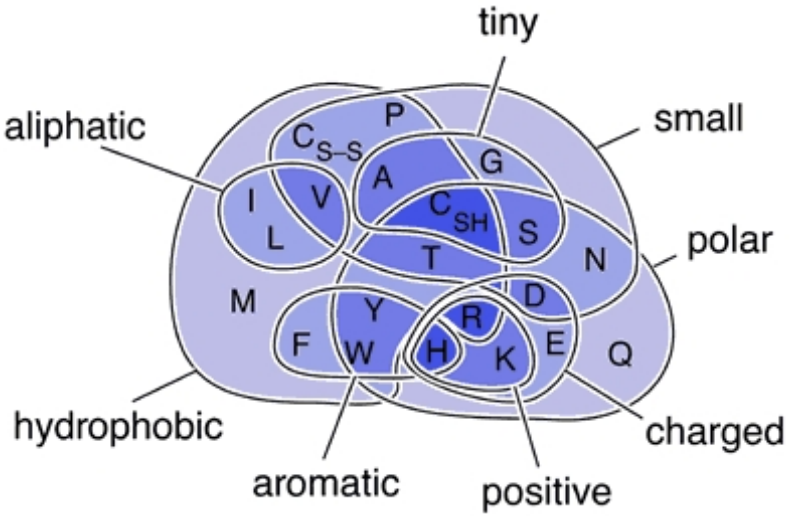
- Most newly synthesized proteins fold without assistance!
  - *Ribonuclease A: denatured protein could refold and recover its activity (C. Anfinsen -1966)*
  - "Structure implies function"
    - *The amino acid sequence encodes the protein's structural information*

The basics

- Proteins are linear heteropolymers: one or more polypeptide chains
  - *Repeat units: 20 amino acid residues*
  - Range from a few 10s-1000s
  - *Three-dimensional shapes ("folds") adopted vary enormously*
  - Experimental methods: X-ray crystallography, electron microscopy and NMR (nuclear magnetic resonance)

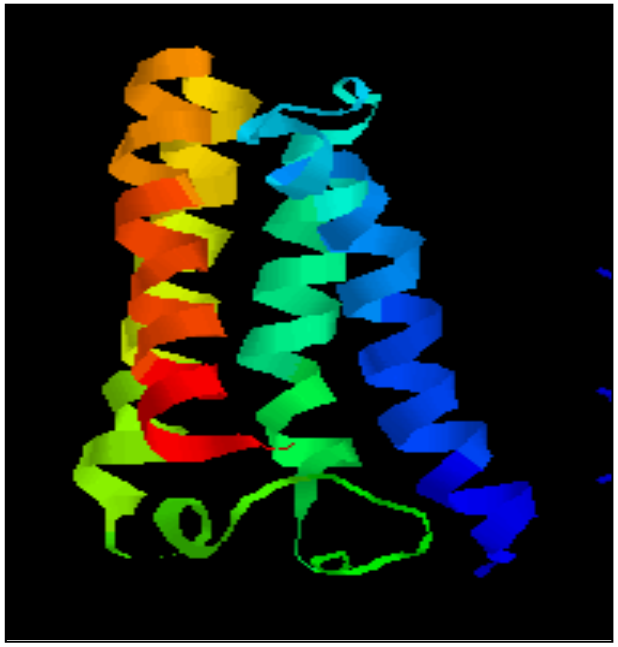
20 amino acids - the building blocks

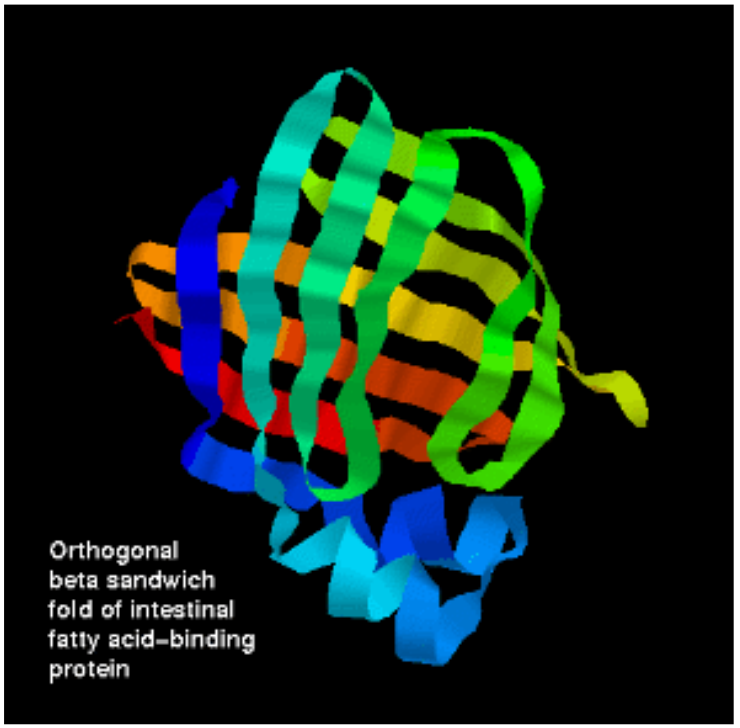




Structural Hierarchy

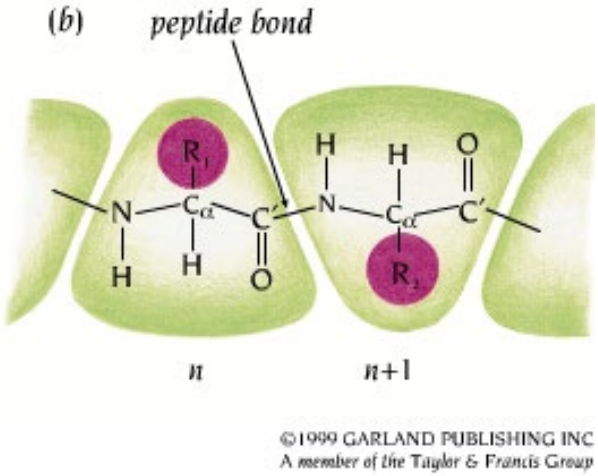
- Primary structure:-→ Secondary structure:-→ Tertiary structure: → Quaternary structure.

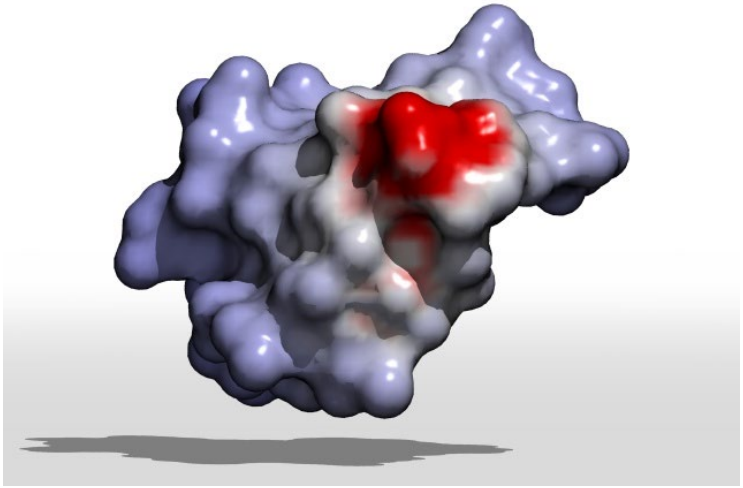




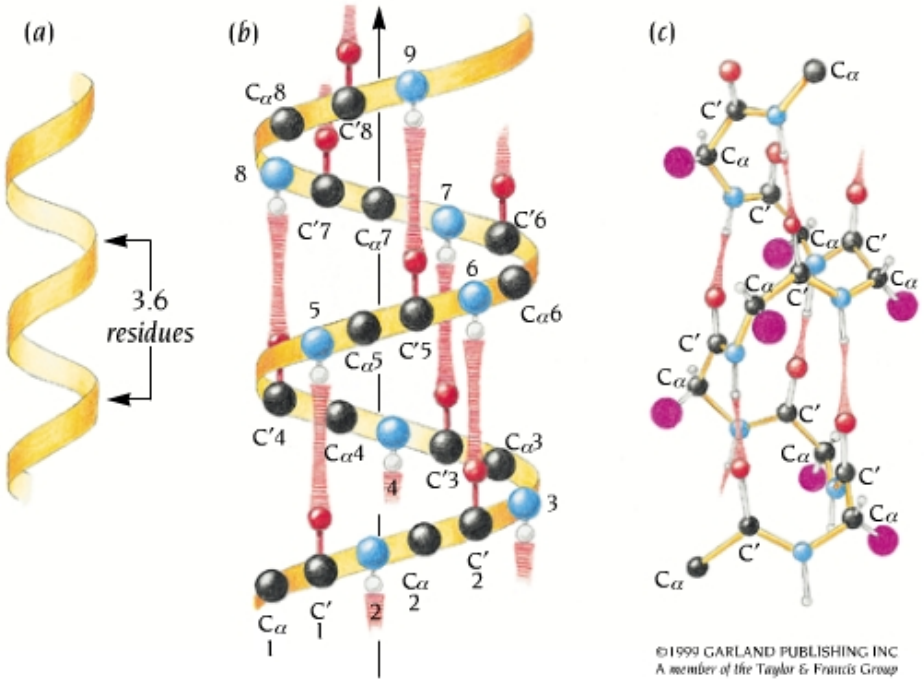
Hydrophilic or hydrophobic..?

- Virtually all soluble proteins feature a hydrophobic core surrounded by a hydrophilic surface
- But, peptide backbone is inherently polar ?
- Solution :neutralize potential H-donors & acceptors using ordered secondary structure





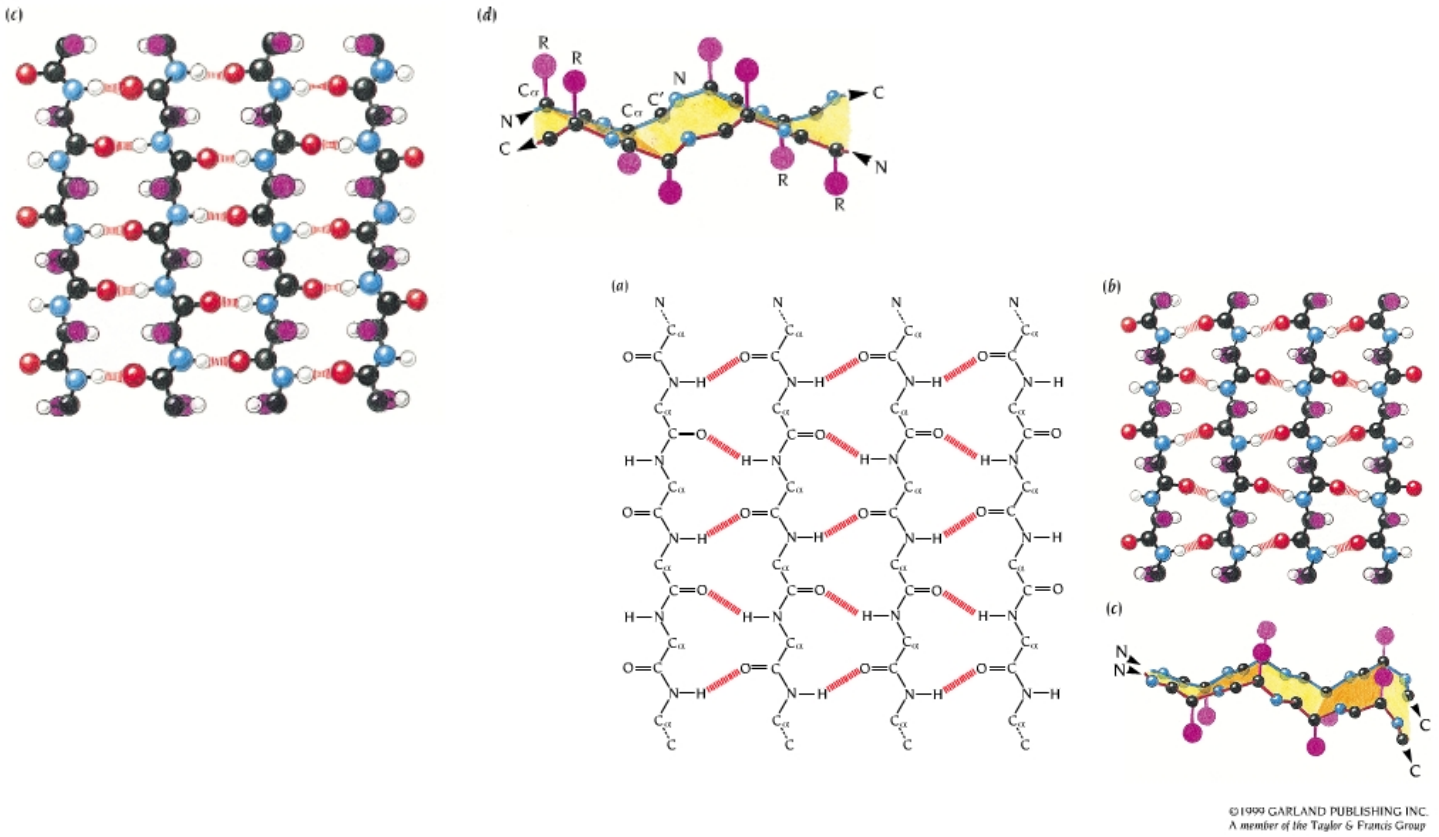
Secondary Structure: α-helix



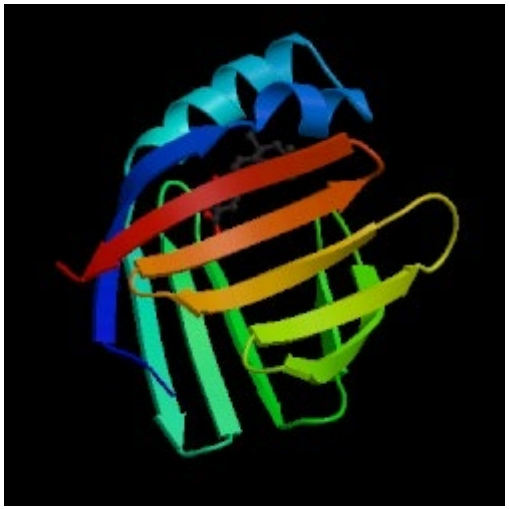
Secondary Structure: α-helix

- 3.6 residues / turn
- Not Proline & Glycine
- Protein surfaces

Secondary Structure: β-sheets



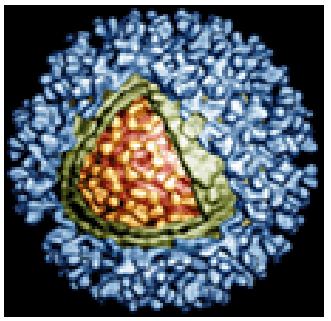
Secondary Structure: β-sheets



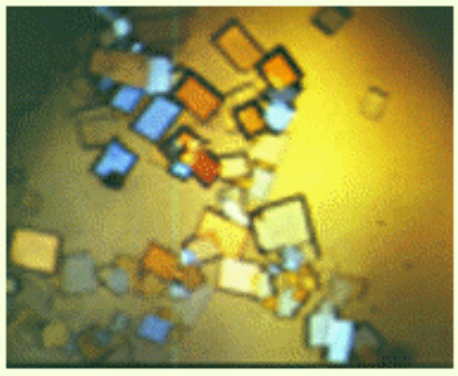
- Parallel or antiparallel
- Alternating side-chains
- No mixing
- Loops often have polar amino acids

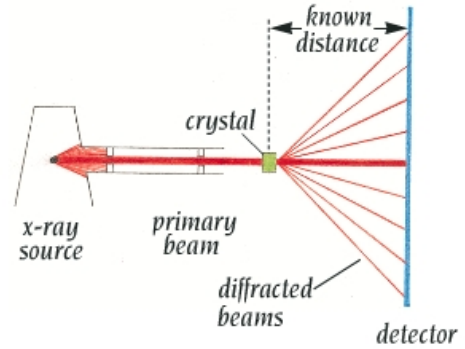
Experimental techniques for structure determination

- X-ray Crystallography
- Nuclear Magnetic Resonance spectroscopy (NMR)
- Electron Microscopy/Diffraction

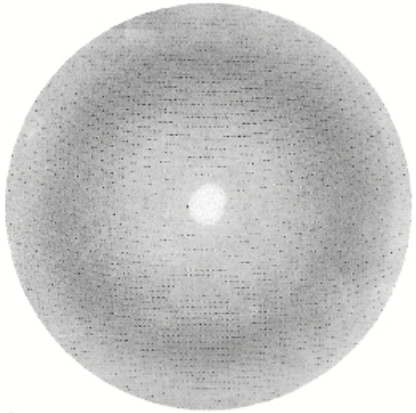


X-ray Crystallography

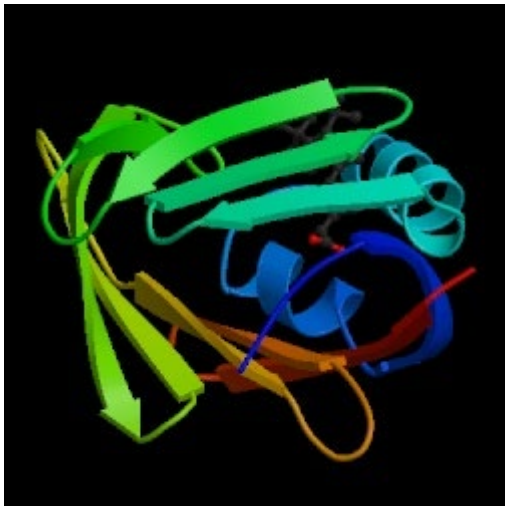




(a)

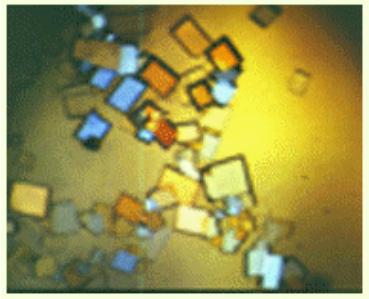


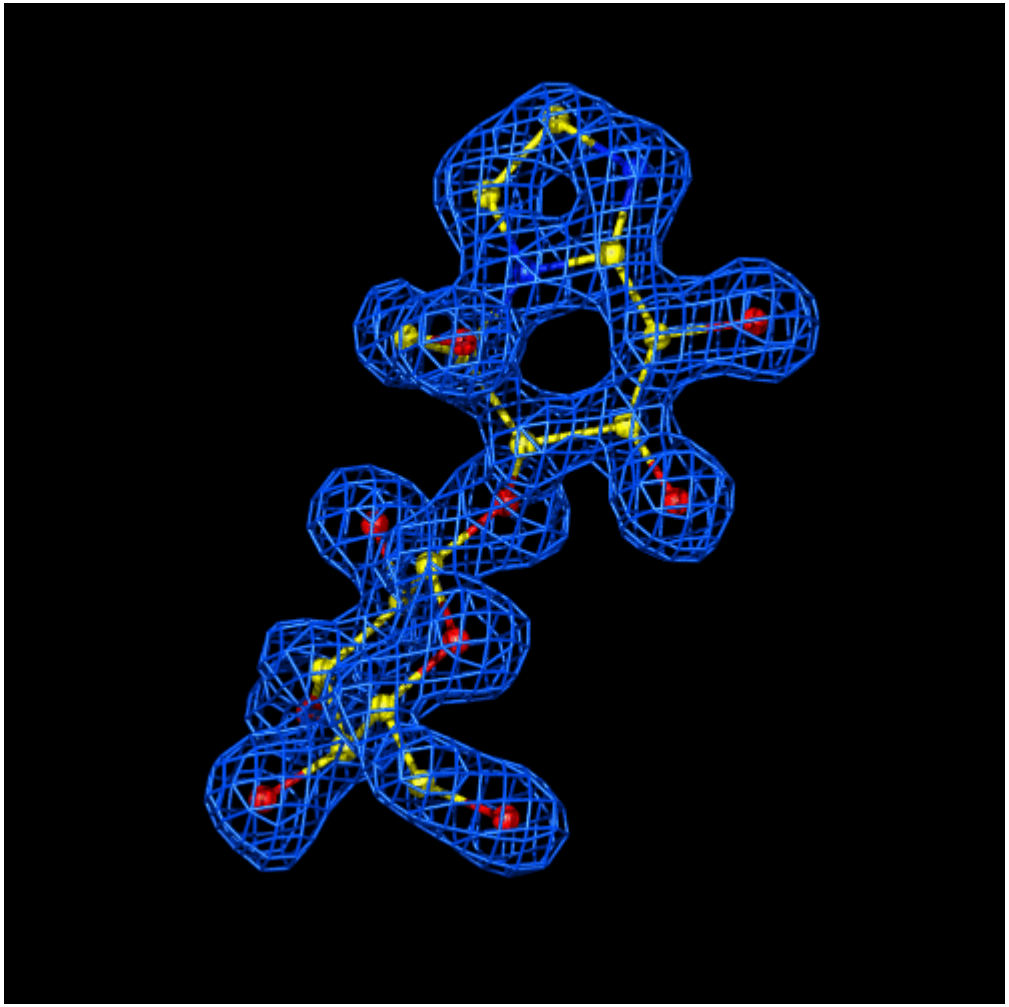
(b)



X-ray Crystallography..

- From small molecules to viruses
- Information about the positions of individual atoms
- Limited information about dynamics
- Requires crystals



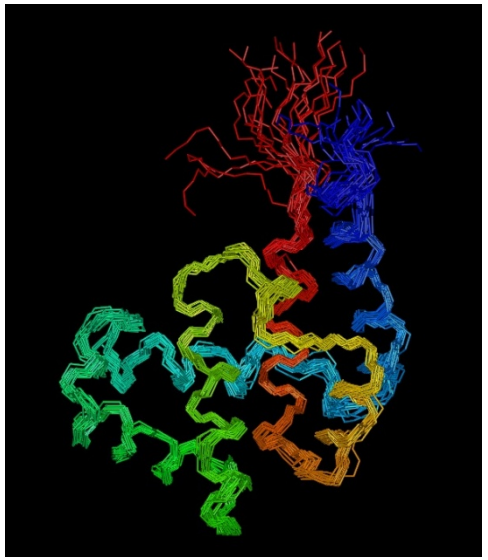


上海同步辐射光源

中国最大的大科学装置，在浦东张江高科技园区建成并投入使用，总投资约12亿元。辐射装置的电子储存环电子束能量为3．5GeV（35亿电子伏特).“上海光源”的能量居世界第四，仅次于日本、美国、欧洲的有关设施.

NMR

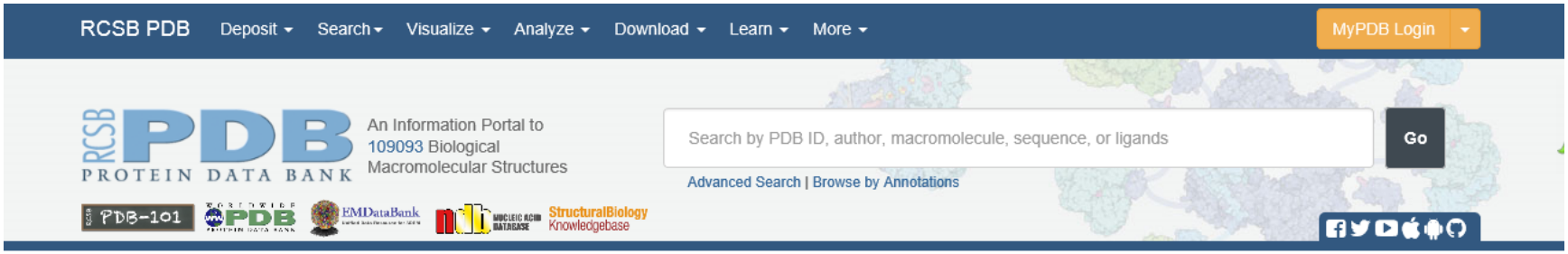
- Limited to molecules up to ~50kDa (good quality up to 30 kDa)
- Distances between pairs of hydrogen atoms
- Lots of information about dynamics
- Requires soluble, non-aggregating material
- Assignment problem

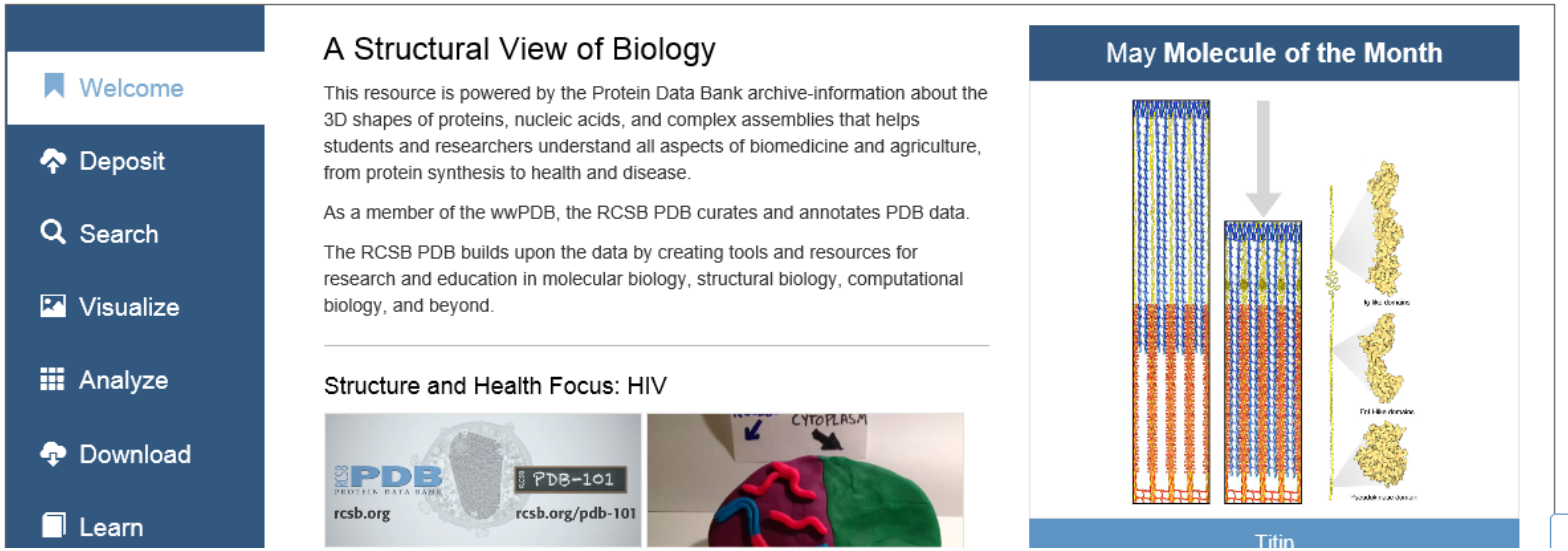


2. Structural visulization

PDB: Protein data bank

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





PDB

- Also contains structures of Protein/Nucleic Acid Complexes, Nucleic Acids, Carbohydrates
- Each entry in PDB is identified by a unique 4-letter code, such as 1SHA.
- Method: mainly X-ray, NMR
- The PDB file has two parts:
  - HEADER
  - Data – 3D coordinates

PDB Header Details

- Identifies the molecule, any modifications, date of release of PDB entry

```
HEADER    PHOSPHOTRANSFERASE                      18-AUG-92   1SHA      1SHA   2
COMPND    V-SRC TYROSINE KINASE TRANSFORMING PROTEIN (PHOSPHOTYROSINE   1SHA   3
COMPND   2 RECOGNITION DOMAIN SH2) (E.C.2.7.1.112) COMPLEX WITH         1SHA   4
COMPND   3 PHOSPHOPEPTIDE A (TYR-VAL-PRO-MET-LEU, PHOSPHORYLATED TYR)   1SHA   5
SOURCE    ROUS SARCOMA VIRUS (SCHMIDT-RUPPIN STRAIN A)                  1SHA   6
AUTHOR    G.WAKSMAN,J.KURIYAN                                           1SHA   7
```

- organism, keywords, method
- Authors, reference,
- resolution if X-ray structure
  - Smaller or bigger better?

The Data Itself

- Coordinates for each heavy (non-hydrogen) atom from the first residue to the last

```
ATOM      1  N   ALA A   2      40.757  22.808  12.014  1.00 61.89      1SHA  65
ATOM      2  CA  ALA A   2      39.528  23.448  12.431  1.00 59.98      1SHA  66
ATOM      3  C   ALA A   2      38.513  23.693  11.308  1.00 56.31      1SHA  67
ATOM      4  O   ALA A   2      37.607  24.536  11.413  1.00 64.00      1SHA  68
ATOM      5  CB  ALA A   2      39.882  24.777  13.140  1.00 56.35      1SHA  69
ATOM      6  N   GLU A   3      38.694  22.905  10.238  1.00 40.05      1SHA  70

ATOM    878  OXT LEU B 205      61.380  28.054   2.998  1.00 62.30      1SHA 942
TER     879      LEU B 205
```

- Any ligands (starting with HETATM) follow the bio-macromolecule
- O atoms of water molecules at the end

- H atom?
- Missing parts?

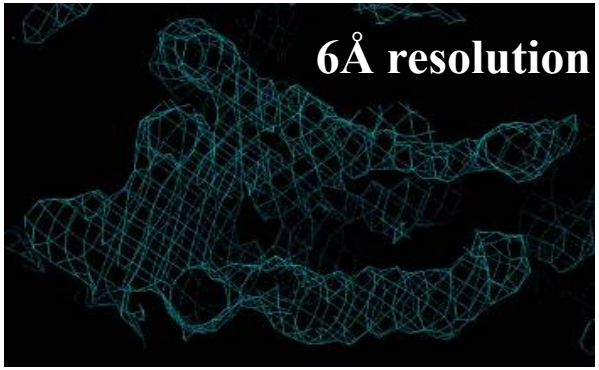
晶体质量与衍射仪决定数据分辨率

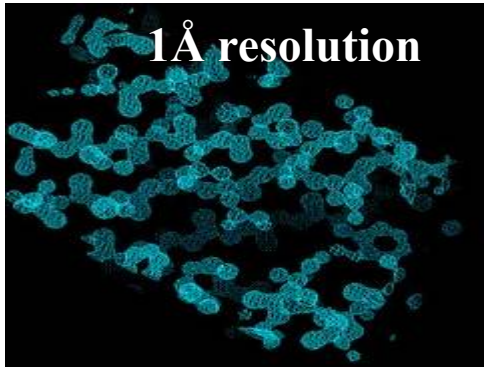
1.0 Å：可分辨单个原子（N比C大）

2.5 Å：可分辨环状状结构

3.0 Å：不易可分辨环状状结构

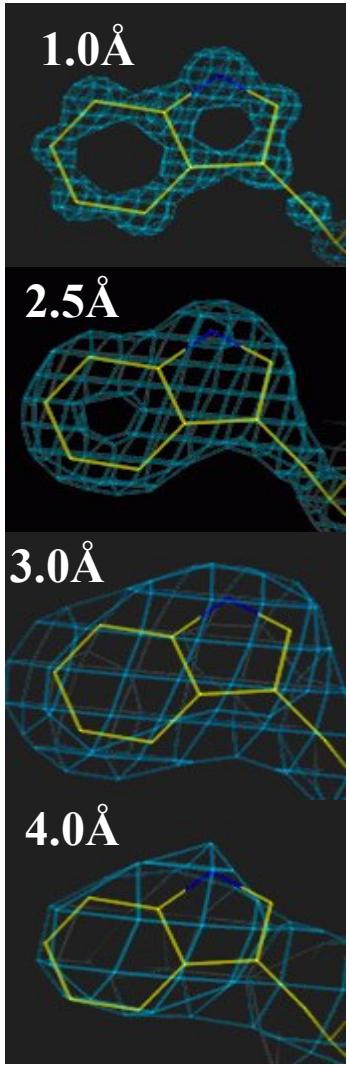
4.0 Å：很难确定是否环状结构.





6.0 Å：可分辨 *a* -helix，不能分辨*b* sheet

≥8 Å：只能分辨分子



Program for Visualization

- RASMOL is one of the most frequently used software for structure visualization.
  - Downloadable at http://www.OpenRasMol.org/
  - Available for most of computer systems
    PC/Windows, Macintosh, Unix
  - Easy to operate and generate nice pictures.

- Swiss PDB Viewer (authored by Nicolus Guex, etc)
  - Downloadable at http://tw.expasy.org/spdbv/
  - Complex but provides more computational functions.

Structure Visualization

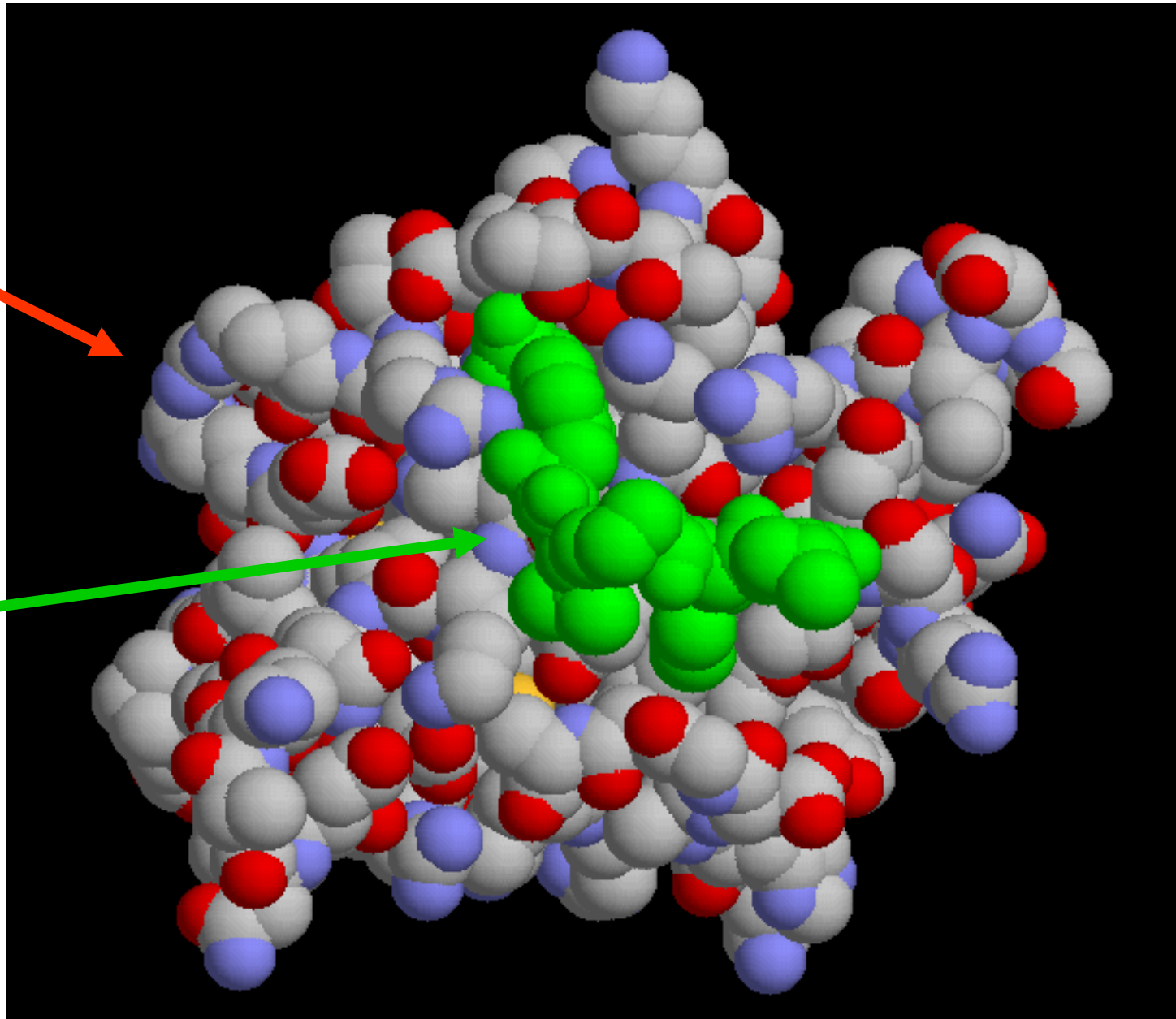
- Structure displaying mode
  - Space-fill
  - Ball and stick
  - Cartoons

Space-fill || Color by CPK

Protein:
Phosphotyrosine Recognition Domain Sh2

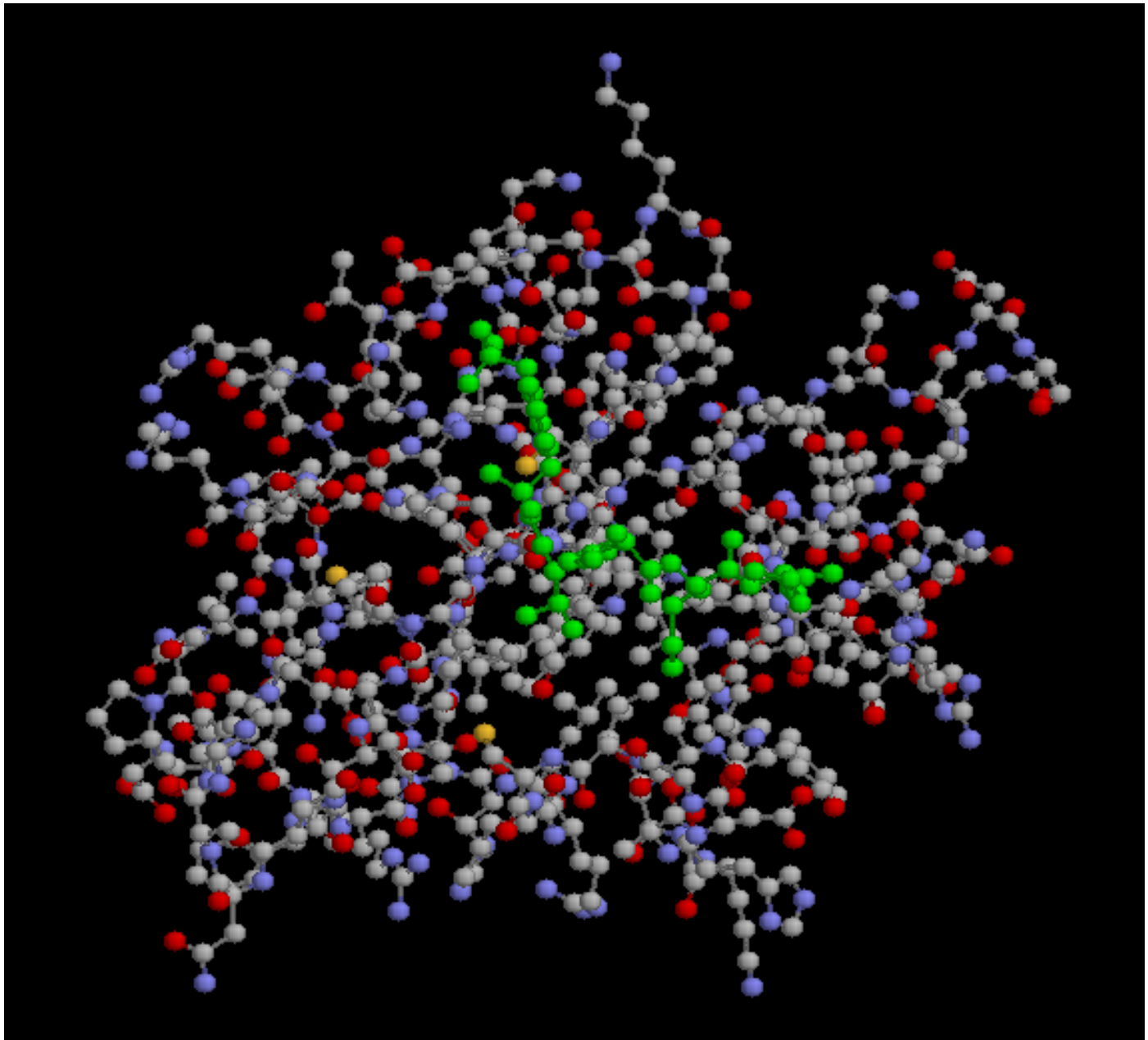
PDB:1SHA

Ligand:
tyrosine-phosphorylated peptides



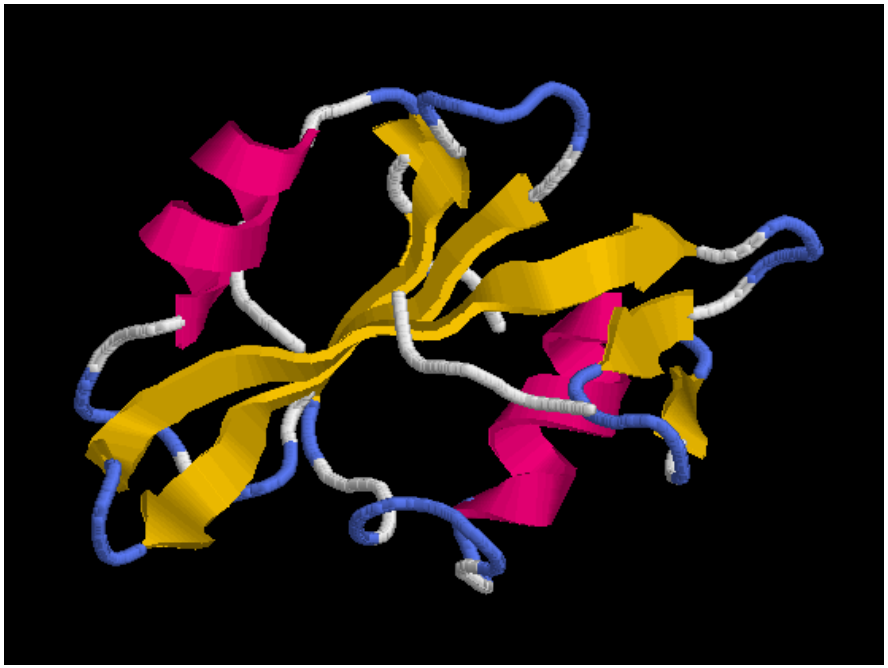
BACK

Ball and stick || Color by CPK

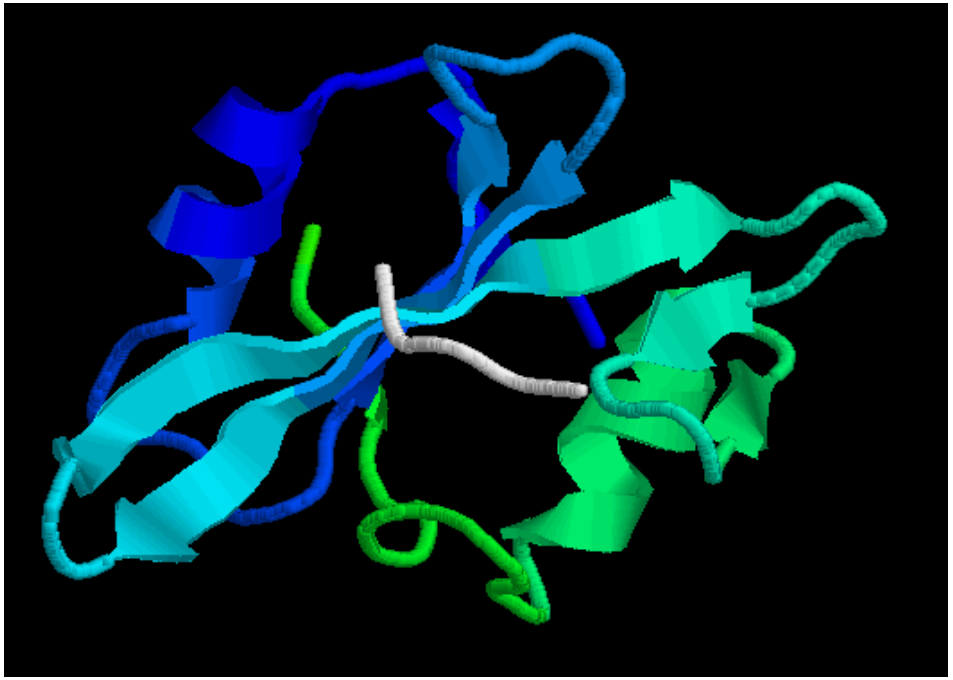


BACK

Cartoon || Color by Structure



Cartoon || Color by Group



BACK

Structural classification

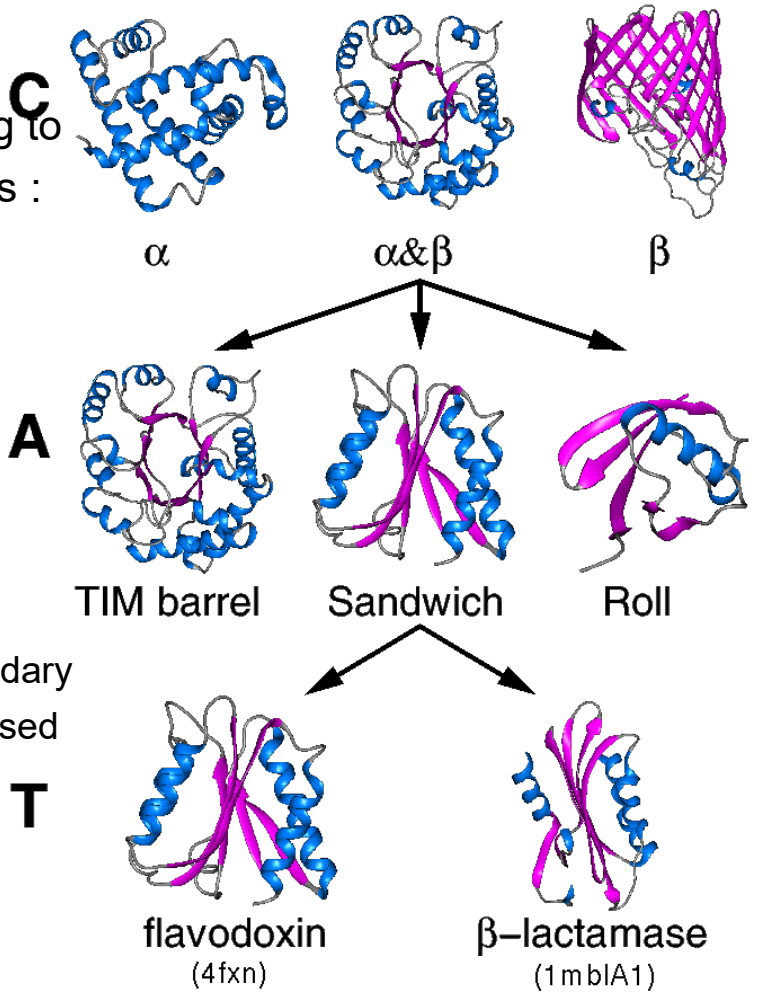
- Databases
  - SCOP, ’Structural Classification of Proteins’, manual classification
  - http://scop.mrc-lmb.cam.ac.uk/
  - CATH, ’Class Architecture Topology Homology’, based on the *SSAP* algorithm
  - http://www.cathdb.info/

CATH

- The CATH database is a hierarchical domain classification of protein structures in the Protein Data Bank. Protein structures are classified using a combination of automated and manual procedures. There are four major levels in this hierarchy:

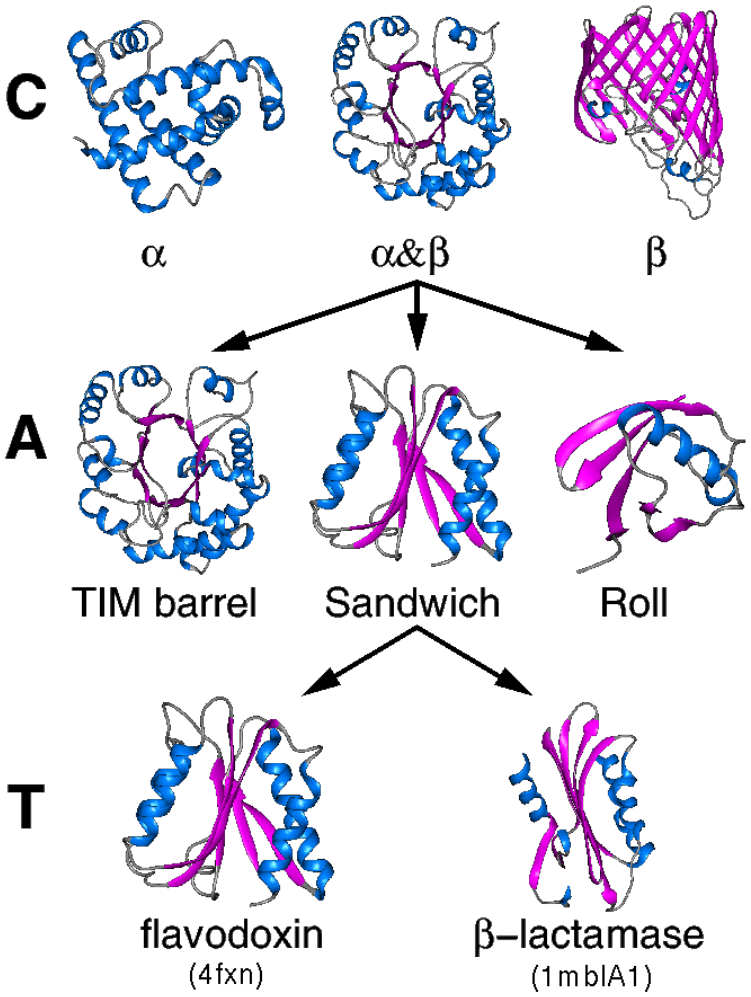
Structural classification, CATH

- Class: structures are classified according to their secondary structure composition four types :
  - Mainly α
  - α / β structures
  - Mainly β
  - few secondary structure
- Arhitecture : information on the secondary structure arrangement in three-dimensional space is used for assignment

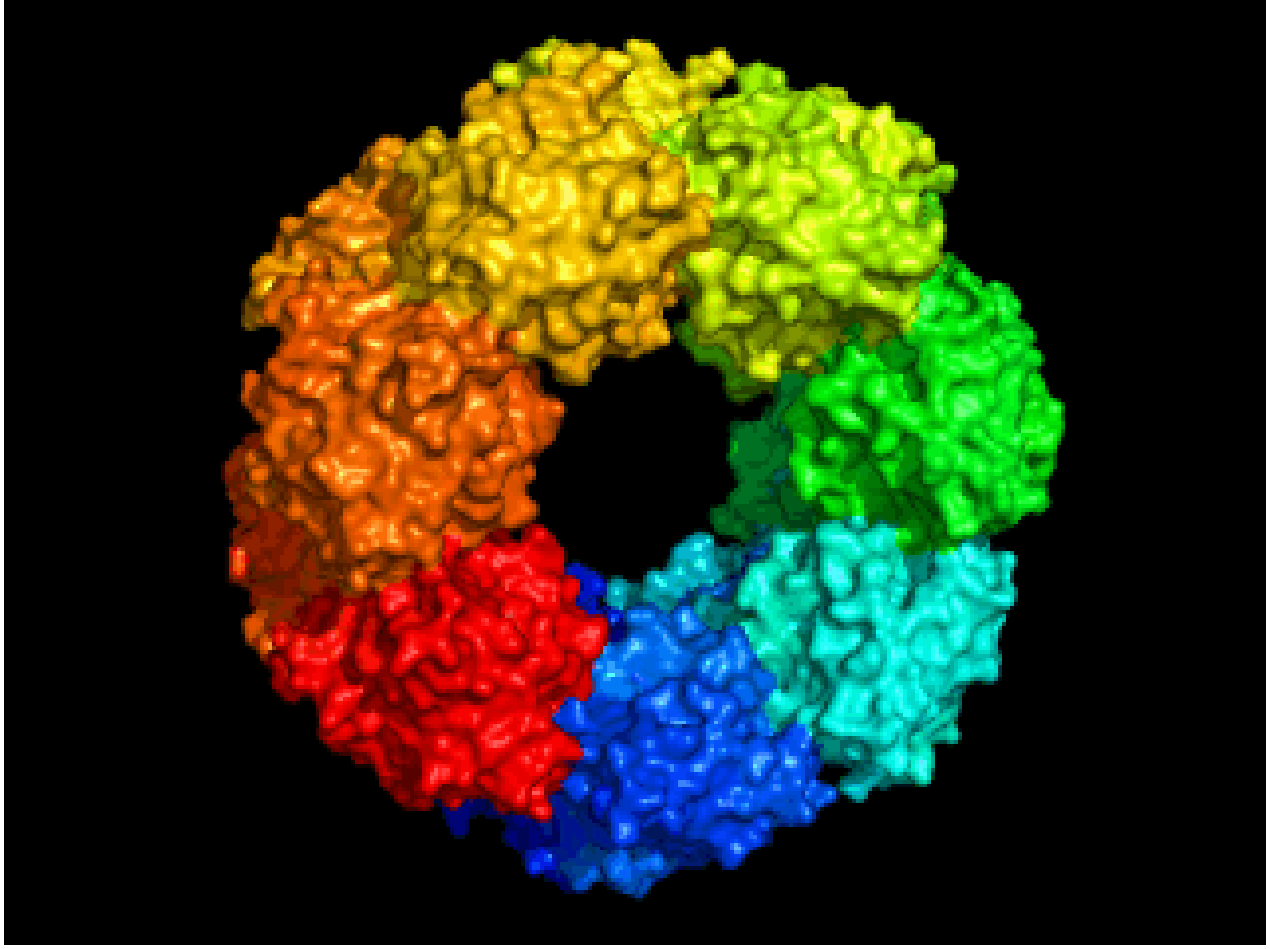


Structural classification, CATH

- Topology (fold family) - structures are grouped into fold groups at this level depending on both the overall shape and connectivity of the secondary structures.
- Homologous superfamily - this level groups together protein domains which are thought to share a common ancestor and can therefore be described as homologous.



Animation of (a mutant form of) the human enzyme GLO1 (PDB accession code 1BH5), which is involved in the detoxification of methylglyoxal.



This protein (3FLP) made of two stacked cyclic molecular aggregations (CMA), with each CMA having either 7 or 8 domains, depending on the organism in which it is found. 3FLP is classified as a sugar binding protein and is an important effector protein of the hemolymph immune system.

课堂作业

- 了解上海同步辐射光源

- Databases
  - SCOP, 'Structural Classification of Proteins', manual classification
  - http://scop.mrc-lmb.cam.ac.uk/
  - CATH, 'Class Architecture Topology Homology', based on the *SSAP* algorithm
  - http://www.cathdb.info/