Bent Petersen

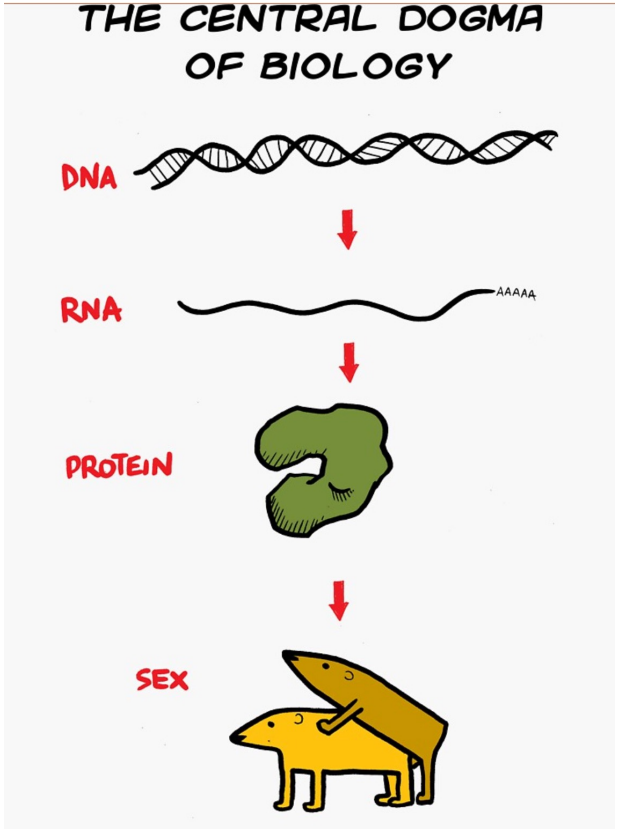
With slides by Carolina Barra Quaglia and others

Introduction to Protein Structure

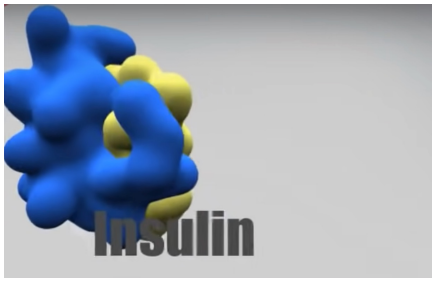
Learning Objectives

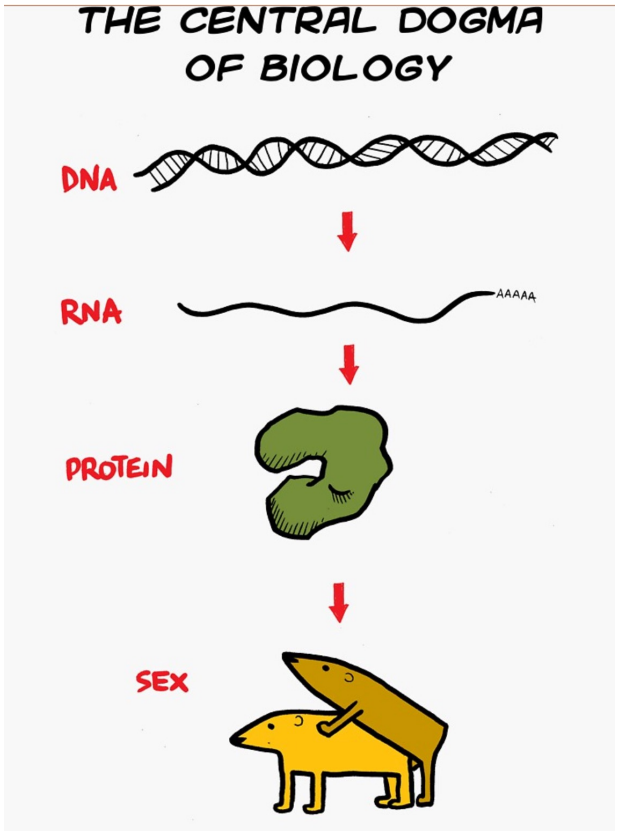
- Identify the **four different levels** of protein structure
- Discuss experimental assays to obtain protein structures: **X-ray crystallography, NMR, CryoEM**
- Interpret relevant parameters for **evaluating the quality** of protein structure determined by X-ray crystallograph
- Identify protein structure **databases**
- Visualize and manipulate protein structures using **PyMol**

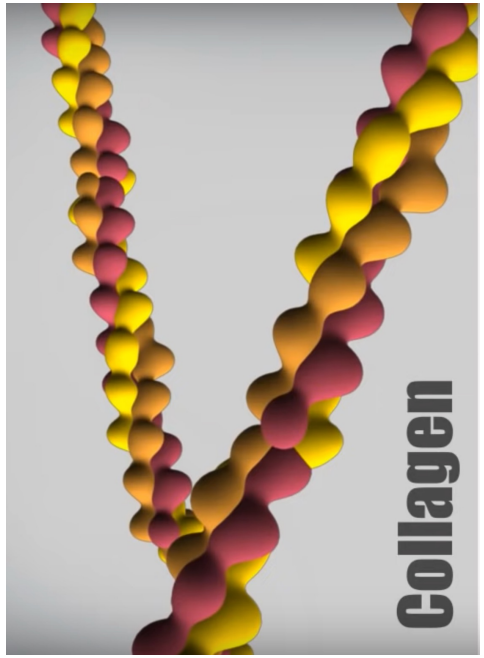
Why are proteins so interesting?



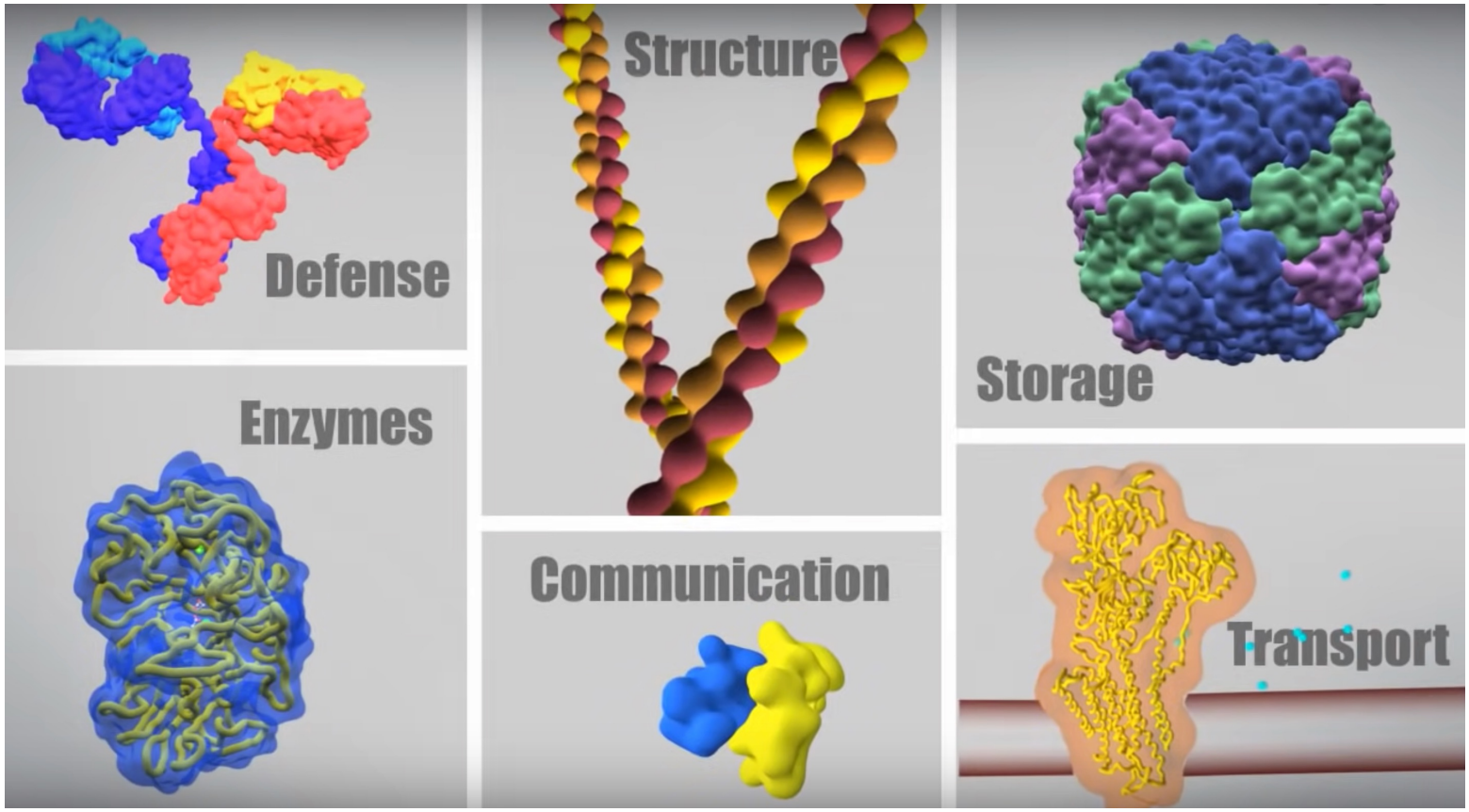
Why are proteins so interesting?



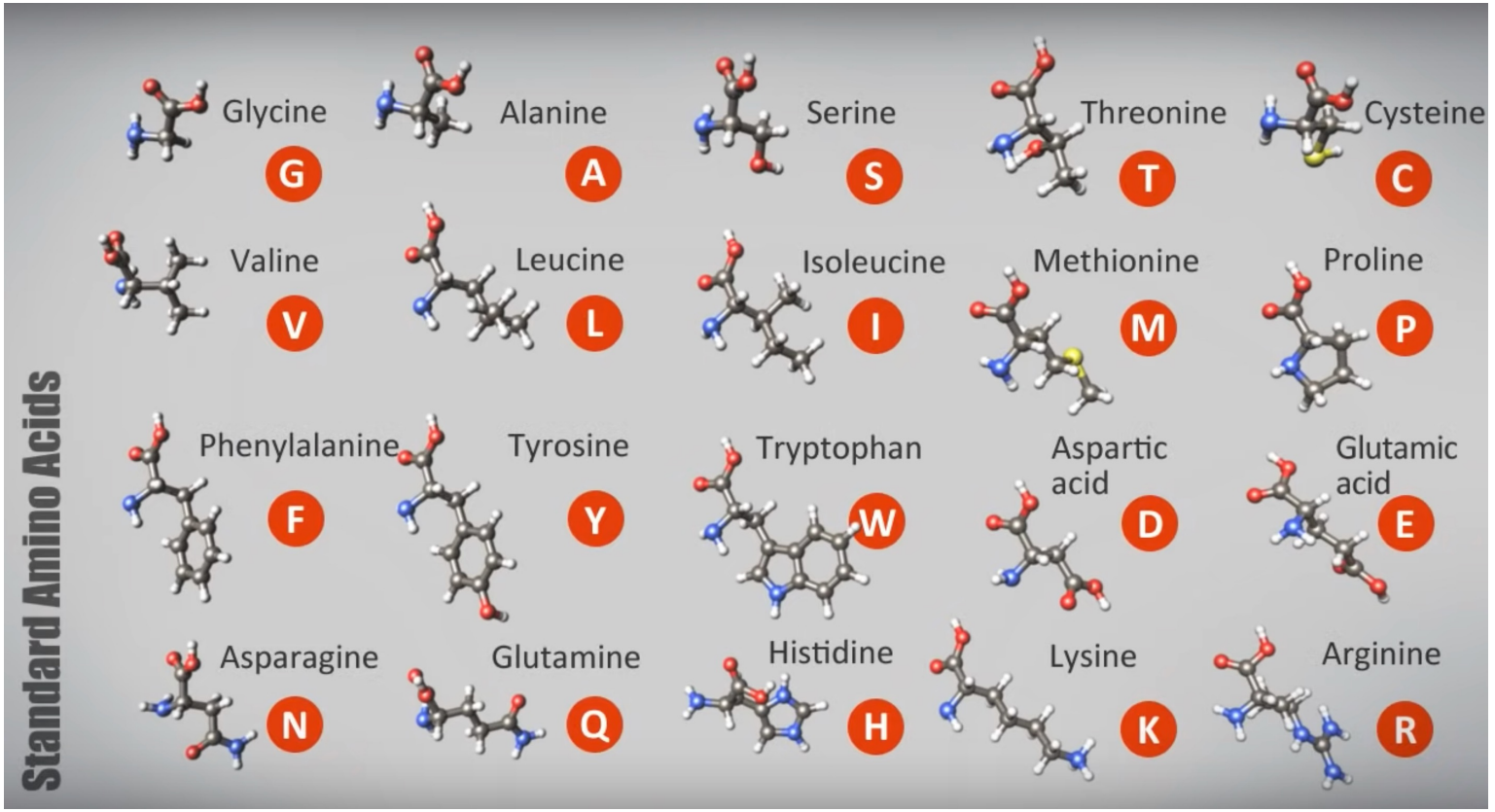




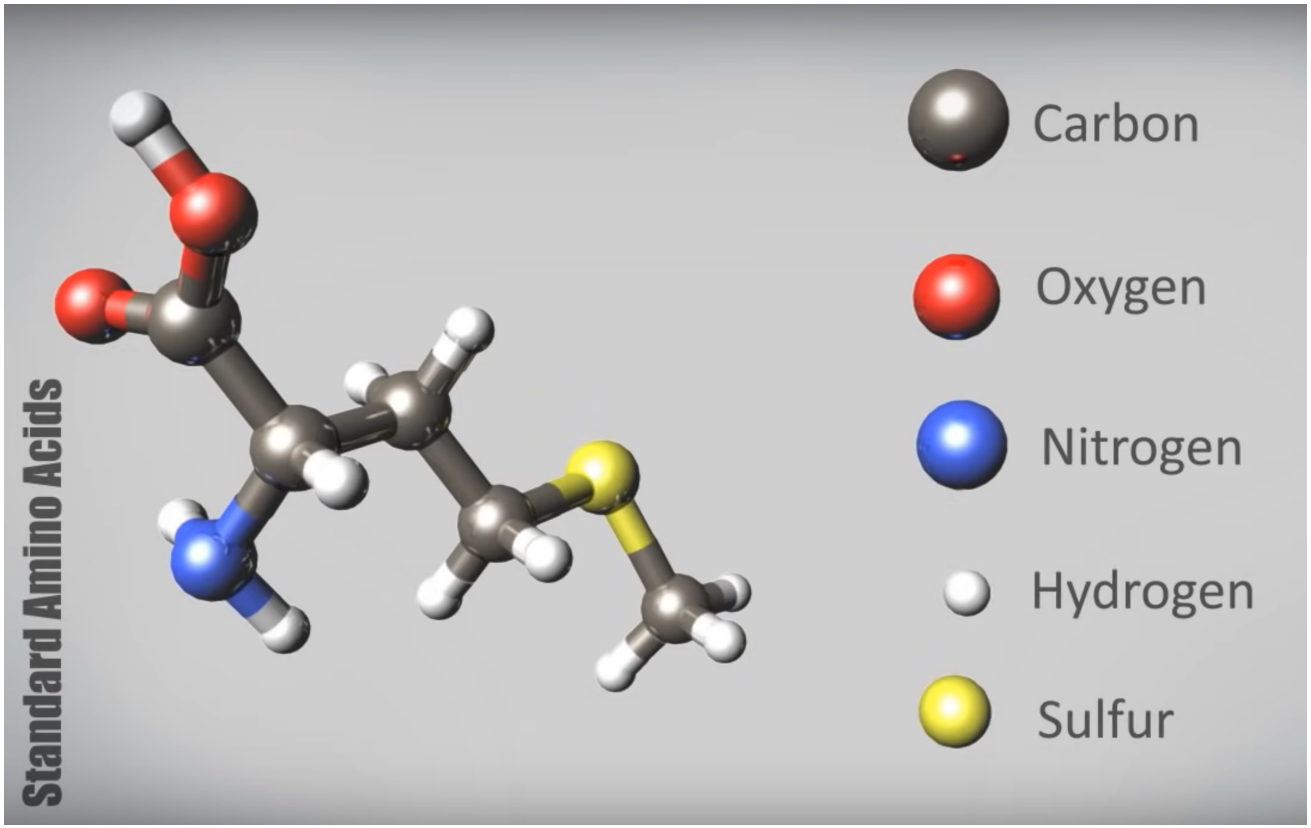
Why are proteins so interesting? FUNCTION



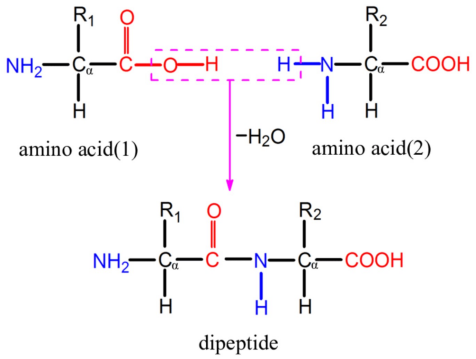
Protein's building blocks: the amino acids

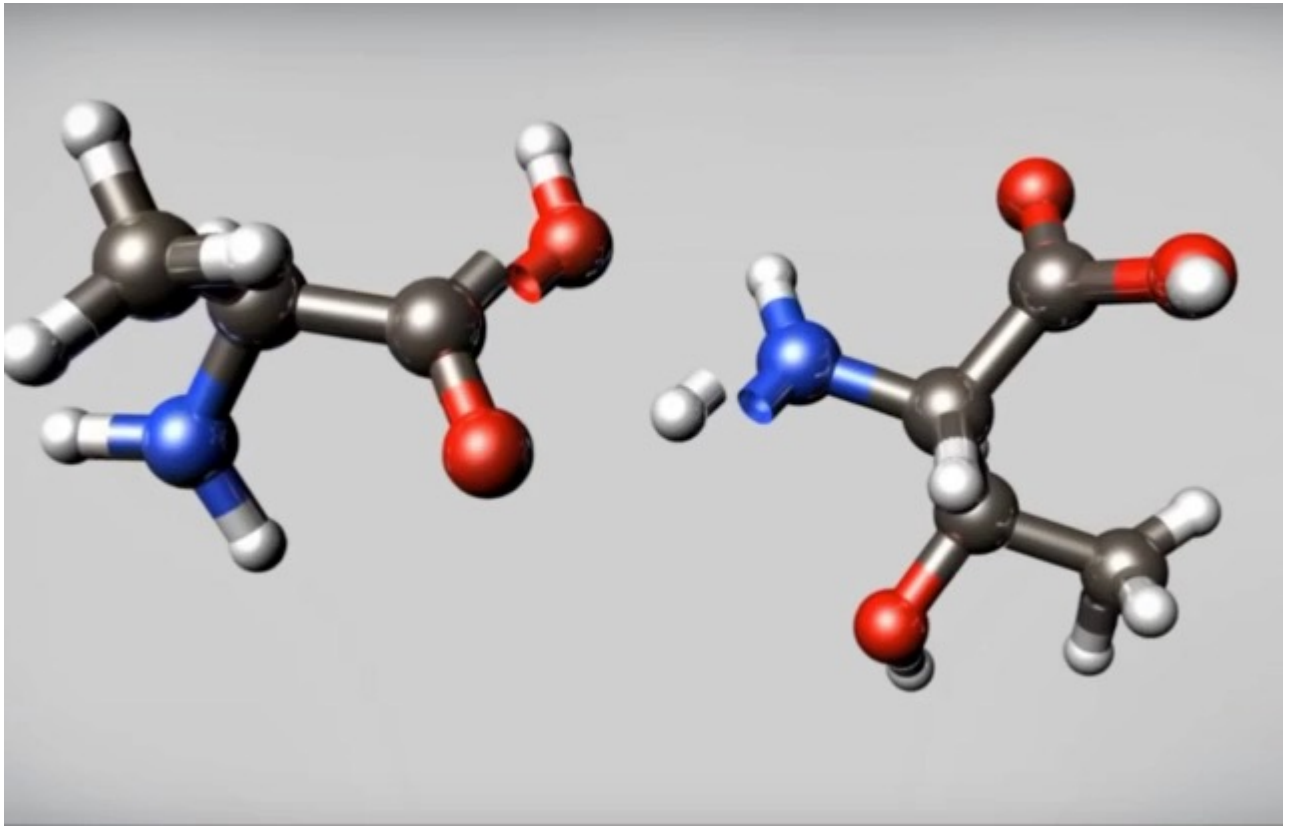


What are they made of?

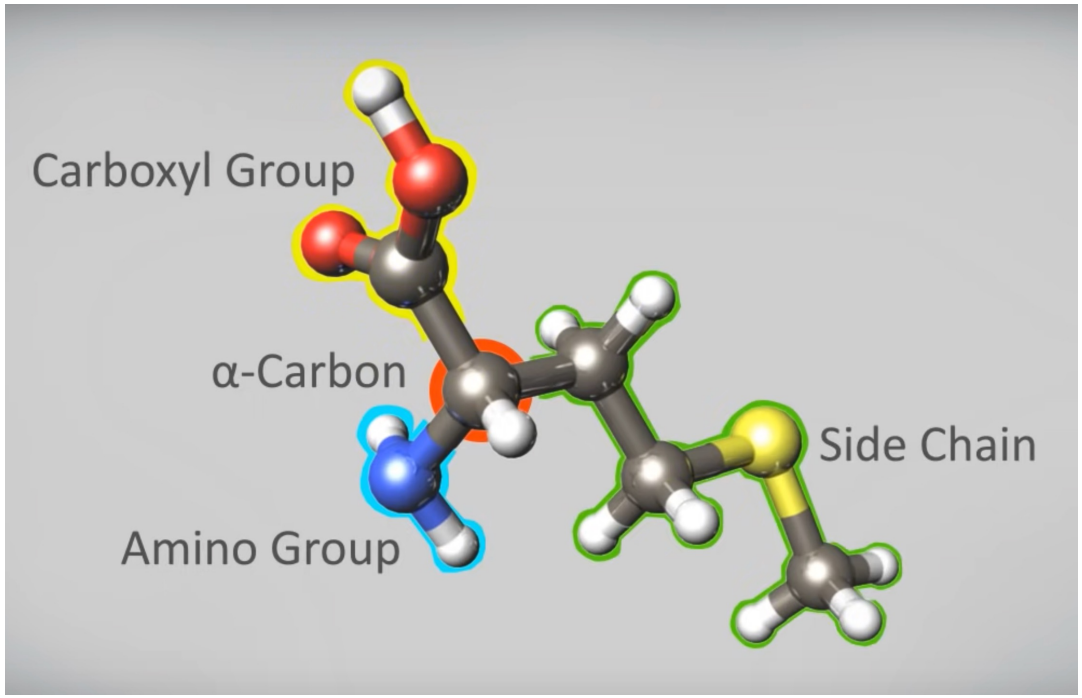


Primary structure
How polypeptide chains are formed?





Amino acids are chiral (L - amino acids)



- Carboxyl group

$-COO^- <==> -COOH$

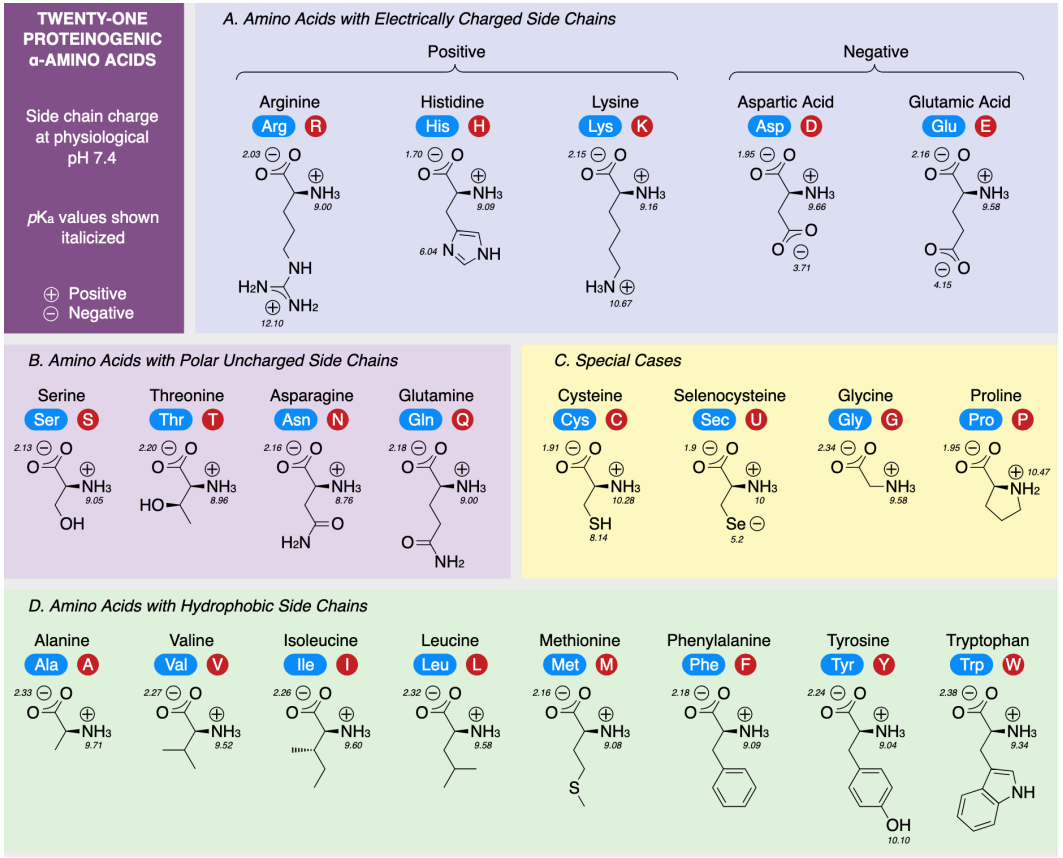
- Amino group

$-NH_2 <===> -NH_3^+$

- α–Carbon
- Side chain

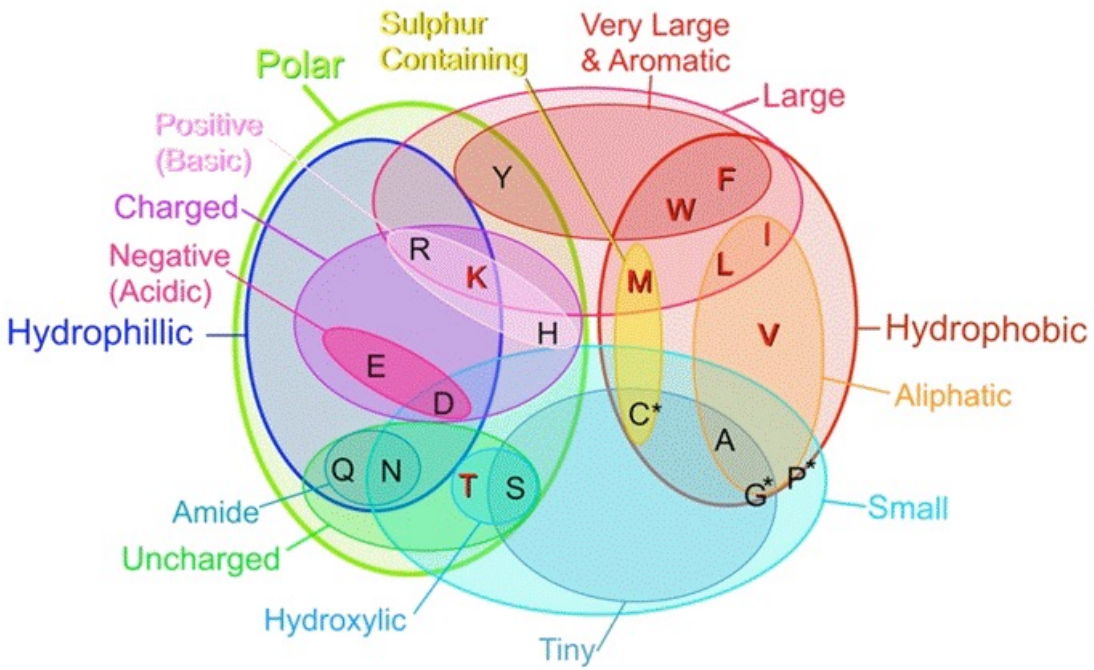
How to group the amino acids?

- Charge +/-
  - Acid vs Basic
- Polarity (polar/non-polar)
  - Type, distribution
- Size
  - Length, weight, volume,
  - surface area
- Type (Aromatic/Aliphatic)



https://upload.wikimedia.org/wikipedia/commons/4/4f/ProteinogenicAminoAcids.svg

How to group the amino acids?



Amino Acids

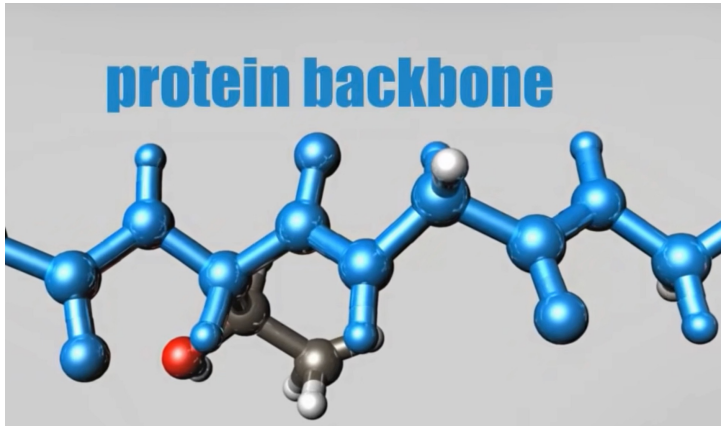
A alanine (ala)
C cysteine (cys)
D aspartic acid (asp)
E glutamic acid (glu)
F phenylalanine (phe)
G glycine (gly)
H histidine (his)
I isoleucine (ile)
K lysine (lys)
L leucine (leu)
M methionine (met)
N asparagine (asn)
P proline (pro)
Q glutamine (gln)
R arginine (arg)
S serine (ser)
T threonine (thr)
V valine (val)
W tryptophan (trp)
Y tyrosine (tyr)

* Unlike some other sources, J. Mol. Recognit., 17, 17-32 (2004) lists G and P as 'Neutral' (not hydrophobic), and does not include them in the group of Aliphatics. C is sometimes listed as polar, uncharged and weakly acidic.

Essential amino acids are shown in red. These cannot be synthesized by the human body, and must be obtained from food.

http://betarhythm.blogspot.com

The protein folding problem



?

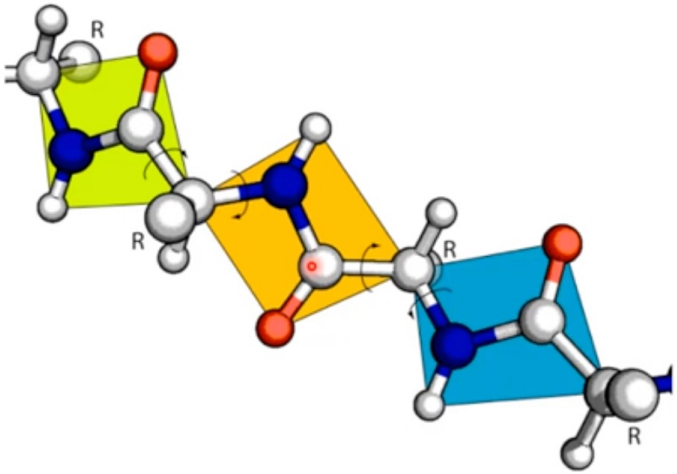
Amino acids are held together by different forces

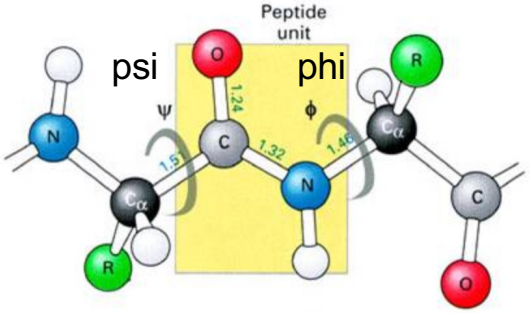
- Hydrophobicity (Entropy)
- Salt bridges
- H bonds
- Di-sulfide bridges
- Aromatic interactions

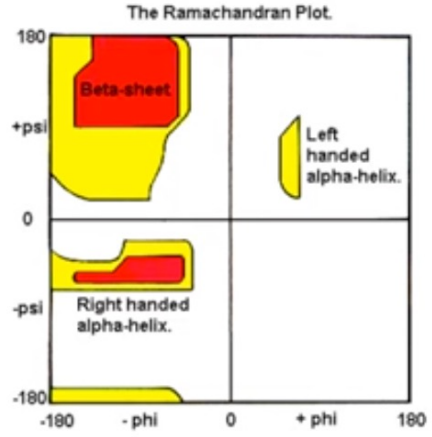
Mostly proteins are held together by **polar interactions and entropy**

Proteins reach the conformation that minimizes their Free Energy

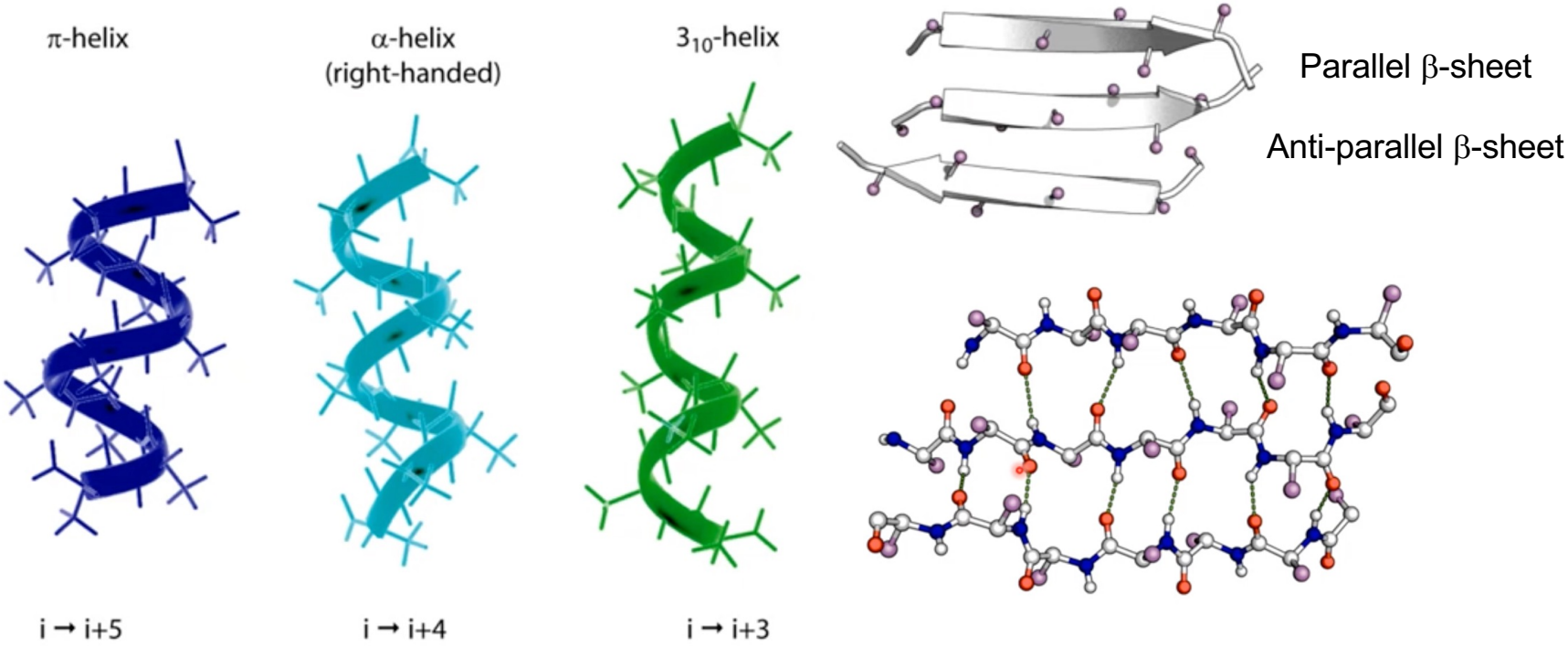
Can polypeptide chains adopt any conformation?





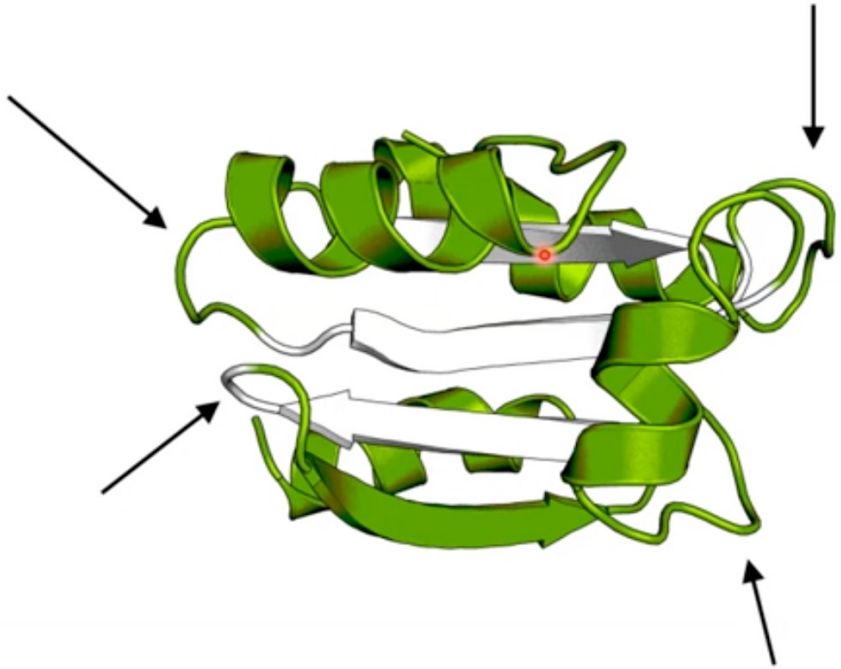


Secondary structure: α–helices and β–sheets



Right handed α-helix i+4 are the most common α-helices

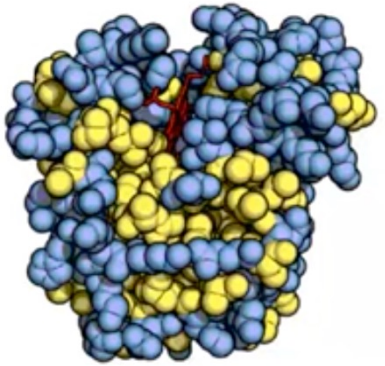
Secondary structure: Turn, loops and bends and disordered regions



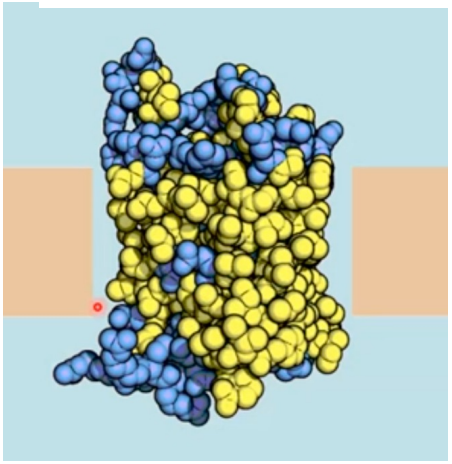
Disordered regions are regions with **non-stable** conformations

Hydrophobicity and hydrophilicity

Hydrophobic side chains such as in **Val, Iso, Leu, Met, Phe, Tyr** go into the core of the protein or into the membrane



myoglobin

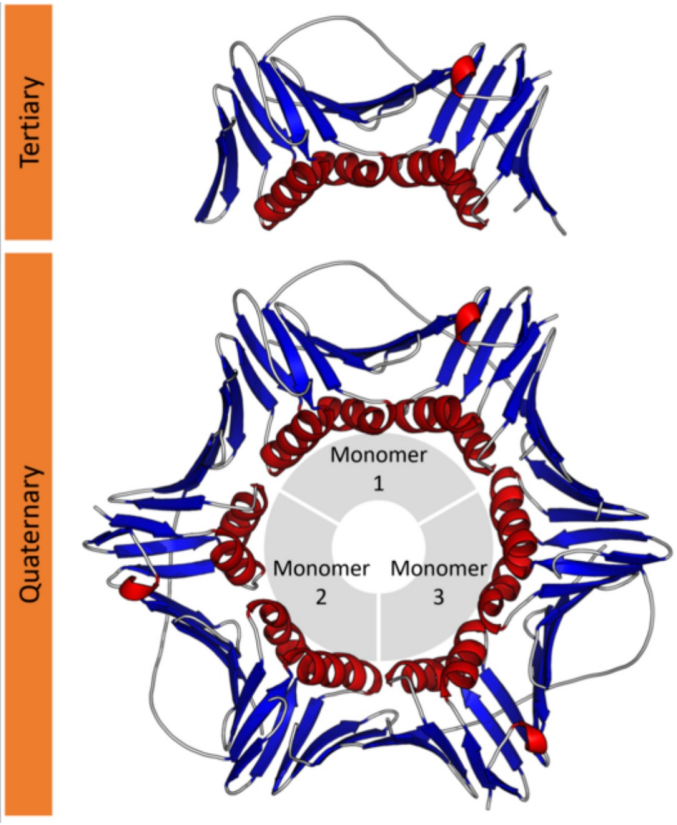


aquaporin

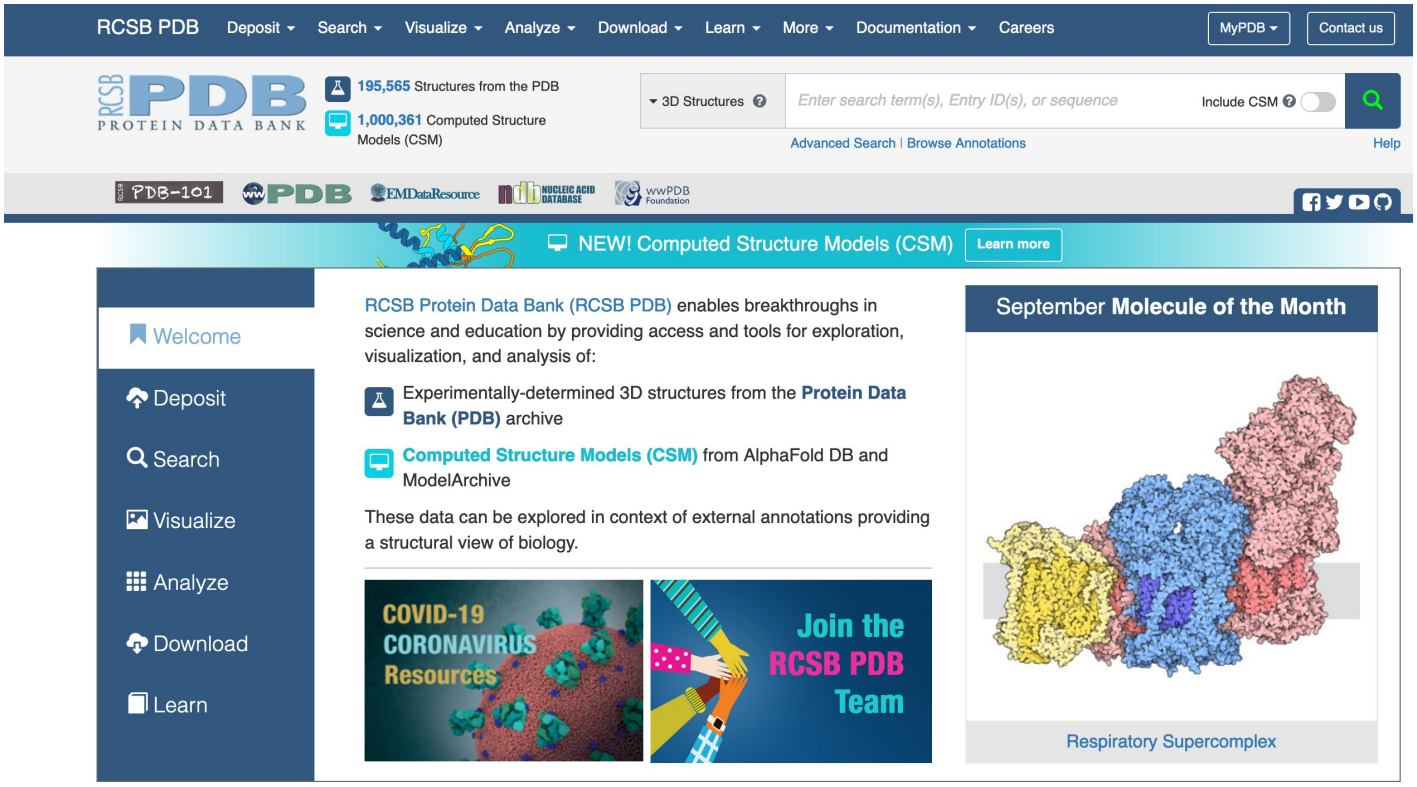
Tertiary and Quaternary protein structure

Tertiary structures are arrangements of secondary structure elements within one protein chain

Quaternary structures are assemblies of multi-chain complexes

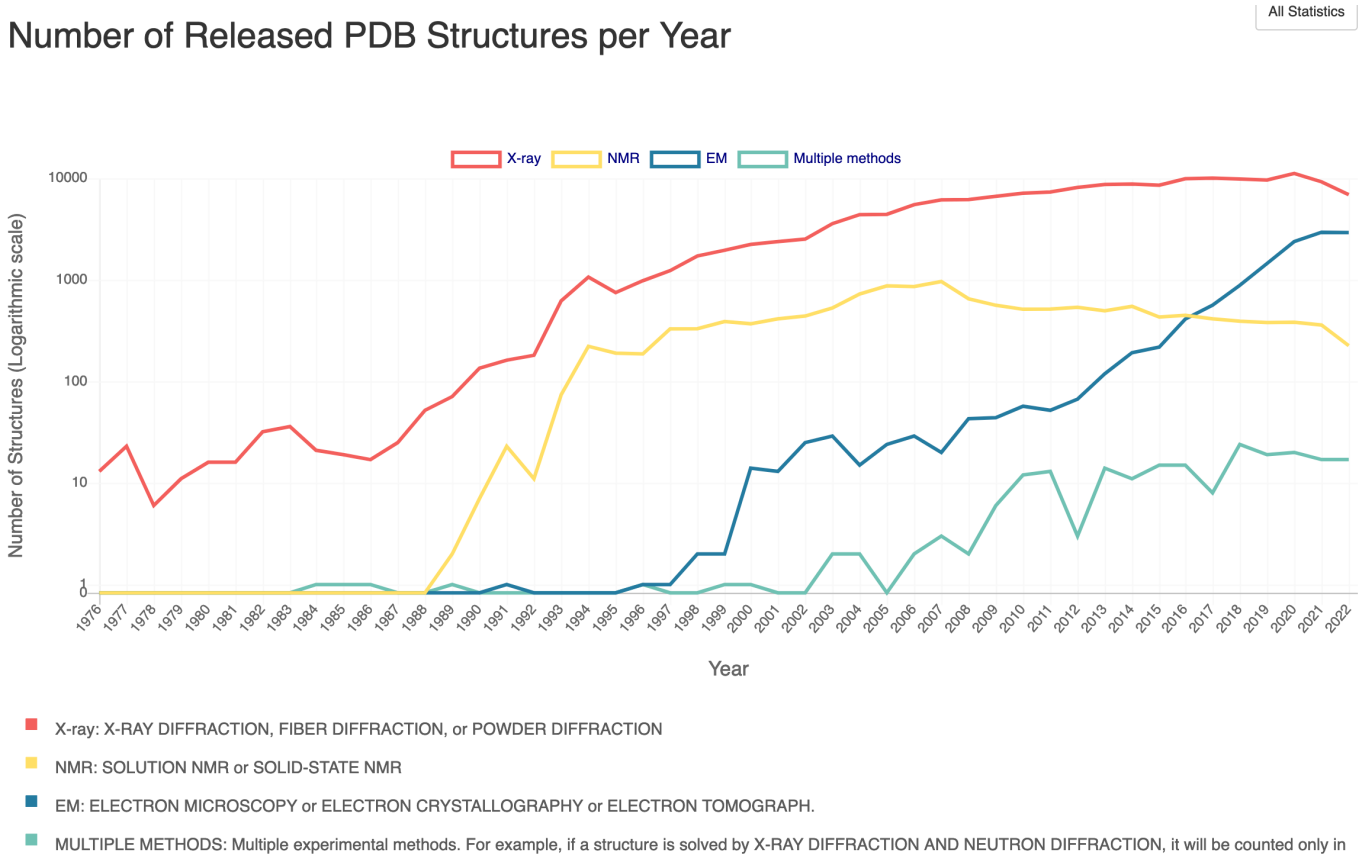


Where do we find protein structure data?

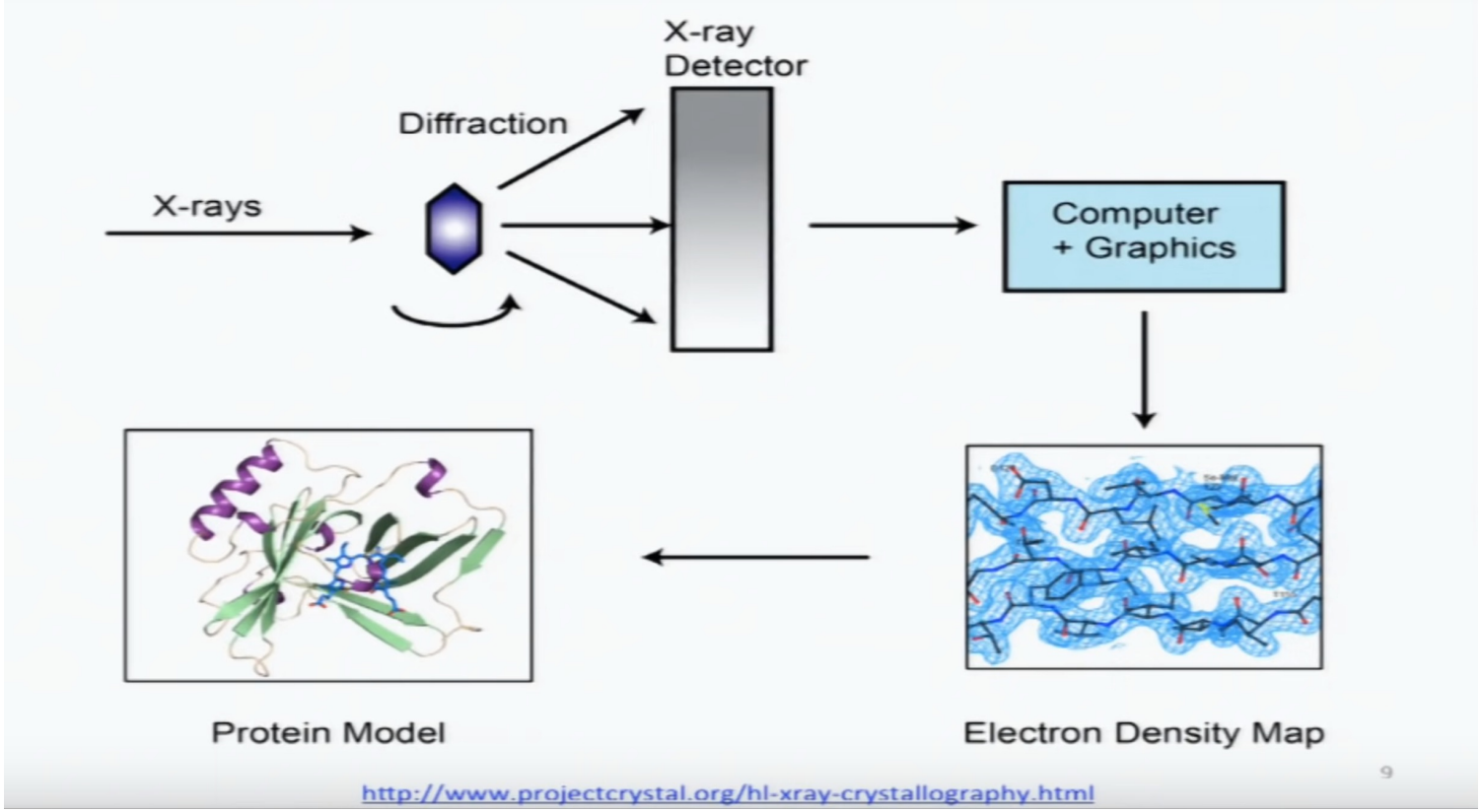


https://www.rcsb.org/

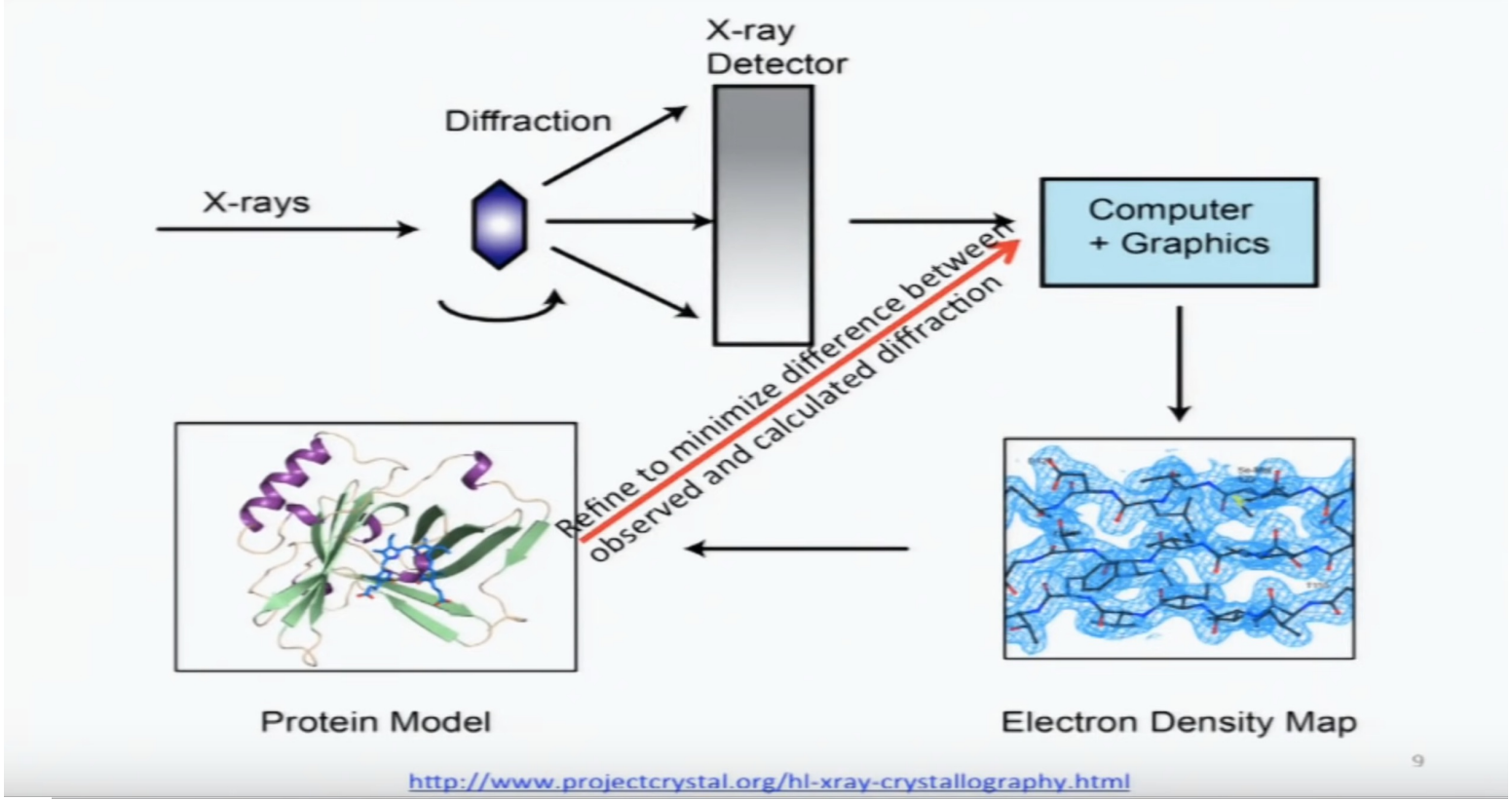
How do we experimentally determine the protein structure?



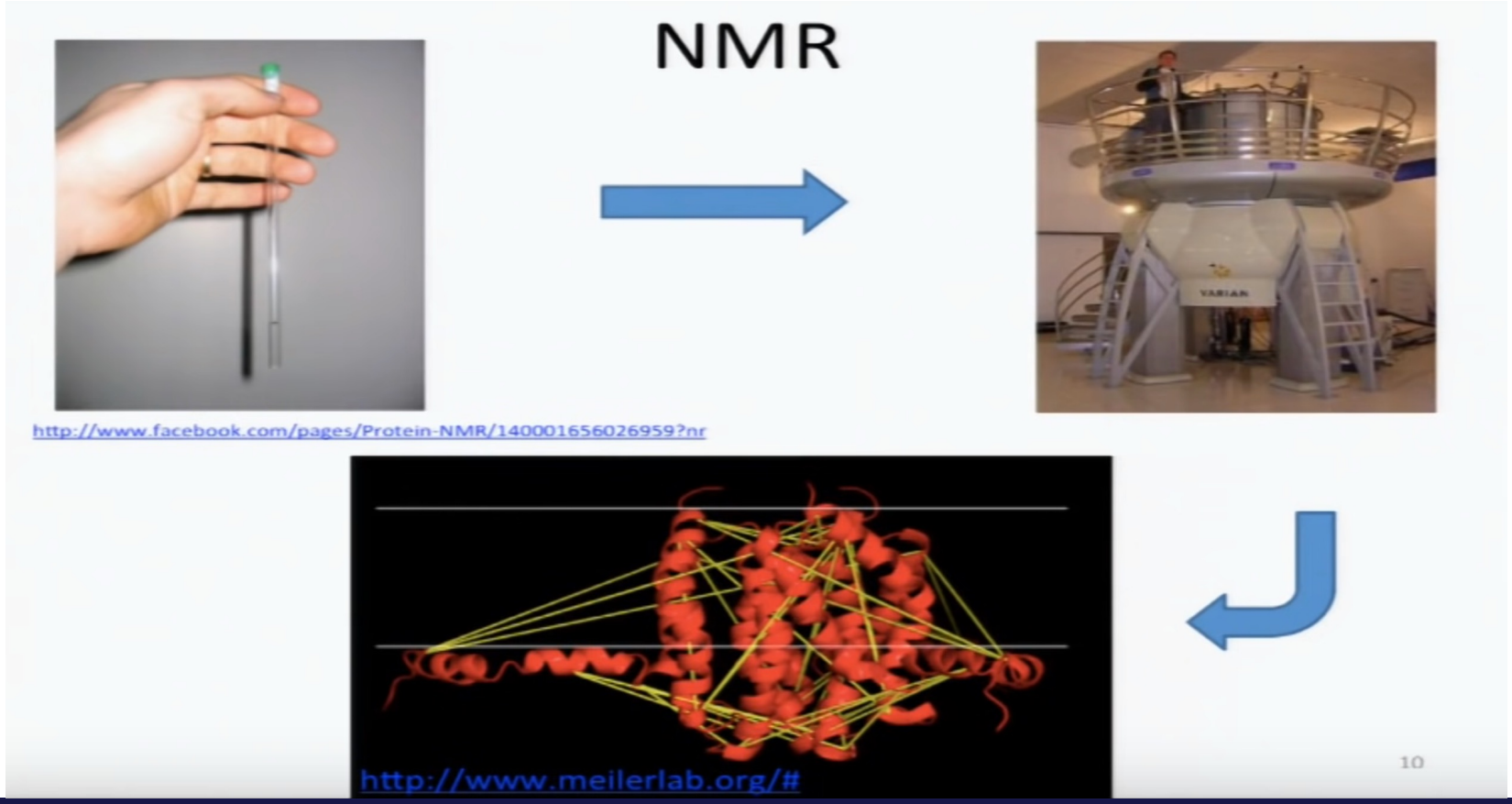
Overview of the X-ray crystallographic method



Overview of the X-ray crystallographic method

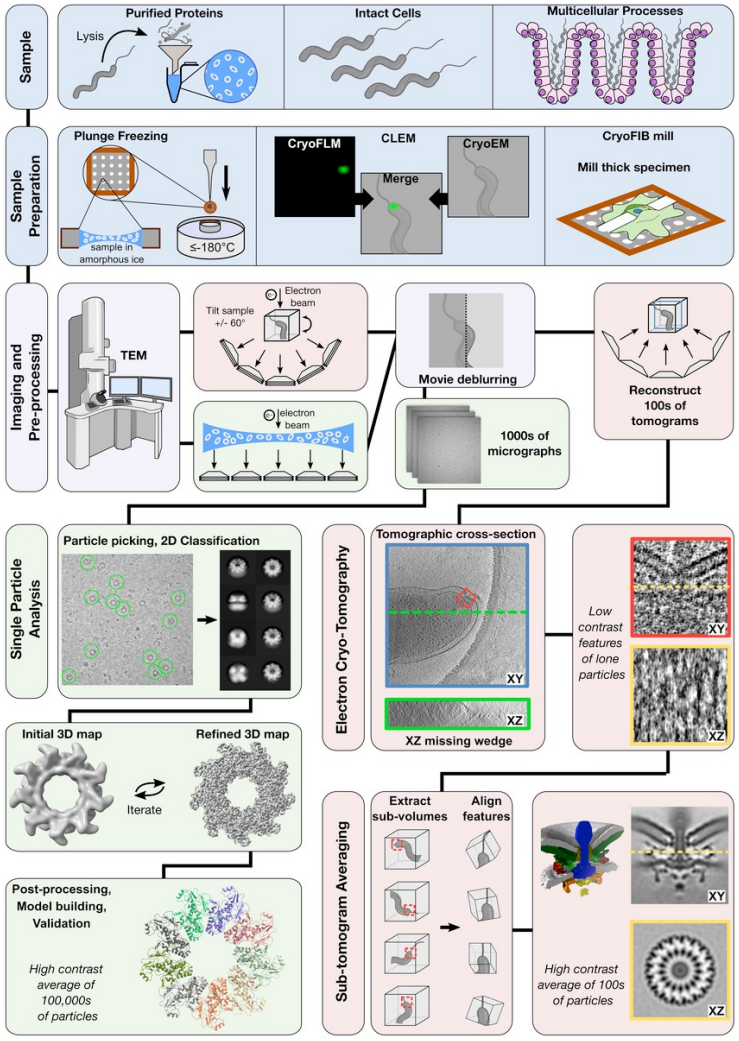


Nuclear magnetic resonance spectroscopy of proteins



Cryo – Electron Mycorscopy EM

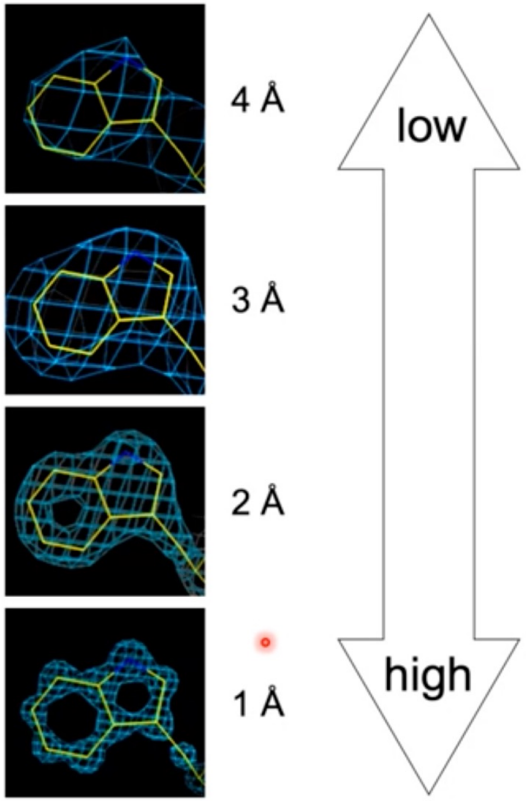
- FAST process
- More native structures
- Ideal for large proteins, or protein complexes
- Now at high resolution

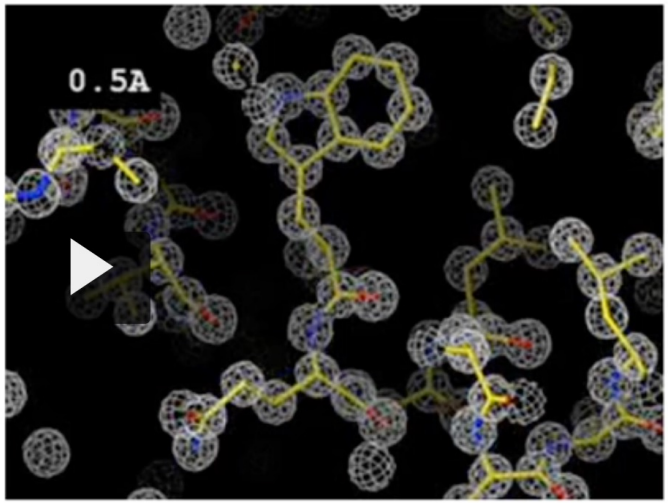


Umrekar et al. Molecular Microbiology, 2020

The importance of protein structure resolution

myoglobin





Quality of the protein structure: Goodness of model fitness

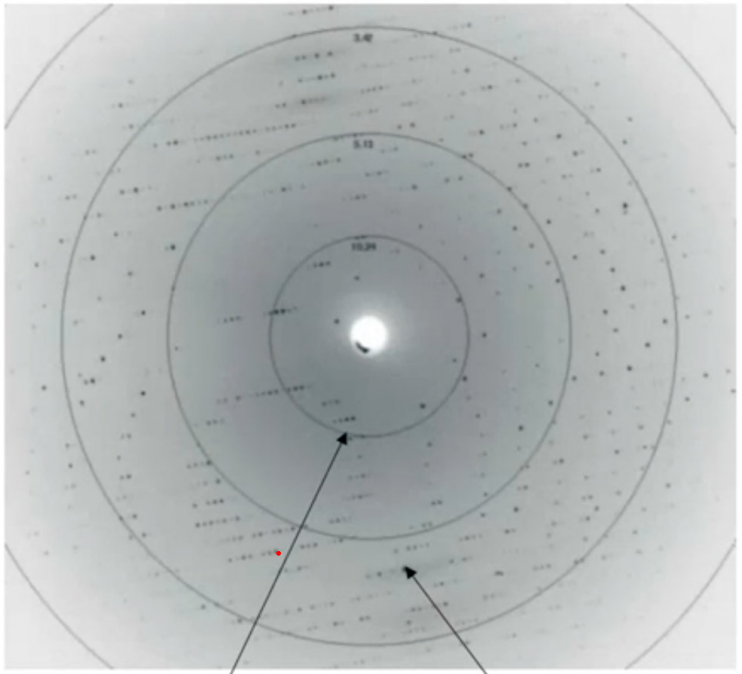
- Individual reflections

$$I_{hkl} \propto \left|F_{obs}(hkl)\right|^2$$

- R-factor:

$$R = \frac{\sum \left| \left|F_{obs}\right| - \left|F_{calc}\right| \right|}{\sum \left|F_{obs}\right|}$$

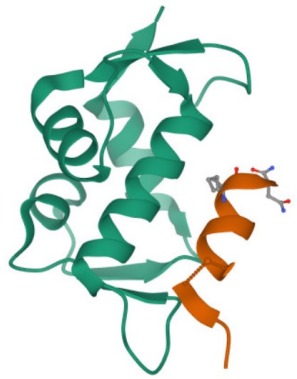
- R_{free}:
  - Like R-factor
  - Unbiased measure.
  - Calculated on 5-10% of data **not** included in refinement.



Where do we find the information in PDB?

Structure Summary | 3D View | Annotations | Experiment | Sequence | Genome | Versions

Biological Assembly 1



3D View: Structure | 1D-3D View | Electron Density | Validation Report

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Hetero 2-mer - A1B1

Find Similar Assemblies

Display Files | Download Files

6V4E

Crystal Structure Analysis of Zebra Fish MDM

PDB DOI: 10.2210/pdb6V4E/pdb

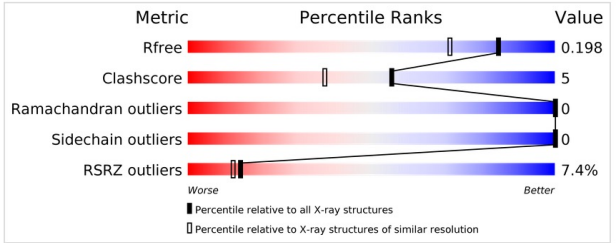
Classification: APOPTOSIS
Organism(s): Danio rerio, Homo sapiens
Expression System: Escherichia coli
Mutation(s): Yes

Deposited: 2019-11-27 **Released:** 2020-04-22
Deposition Author(s): Seo, H.-S., Dhe-Paganon, S.
Funding Organization(s): National Institutes of Health/National Cancer Institute (NIH/NCI)

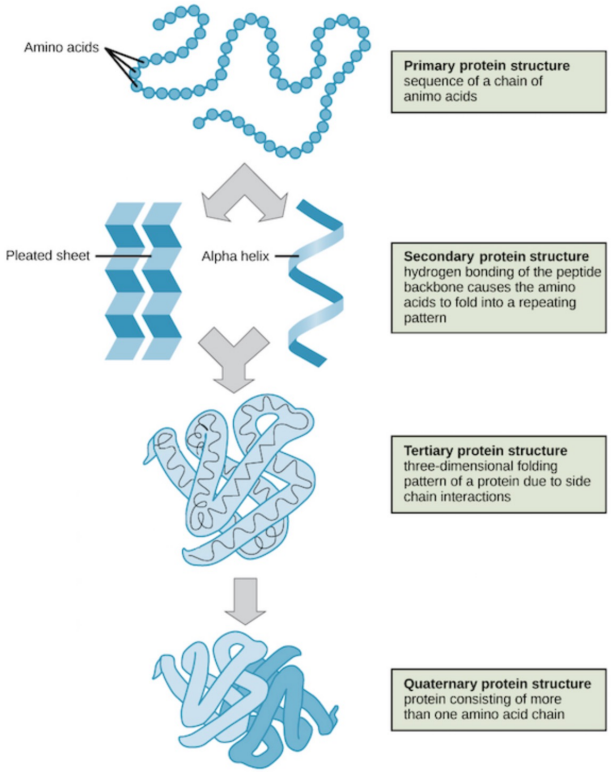
Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.62 Å
R-Value Free: 0.199
R-Value Work: 0.160
R-Value Observed: 0.162

wwPDB Validation | 3D Report | Full Report



Summary: The four levels of protein structure



Video link to protein structure

RCSBProteinDataBank

https://www.youtube.com/watch?v=qBRFIMcxZNM