



Bent Petersen

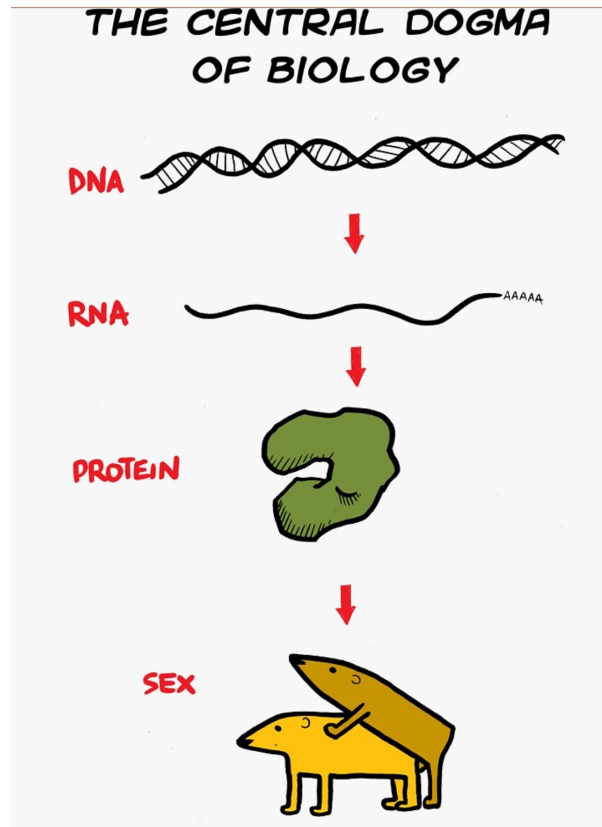
With slides by Carolina Barra Quaglia and others

Introduction to Protein Structure

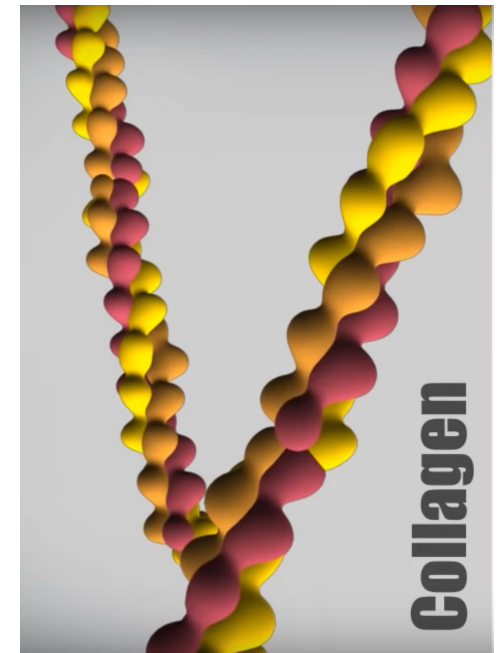
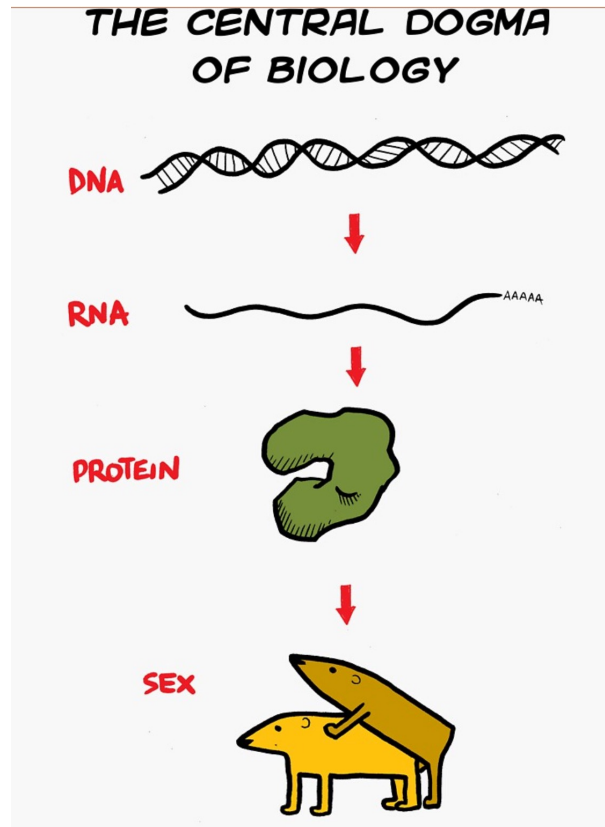
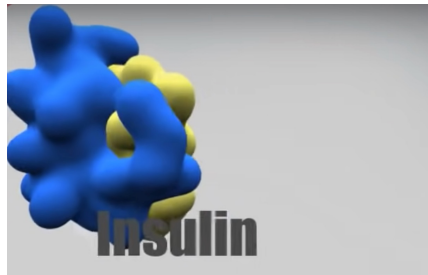
Learning Objectives

- 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**

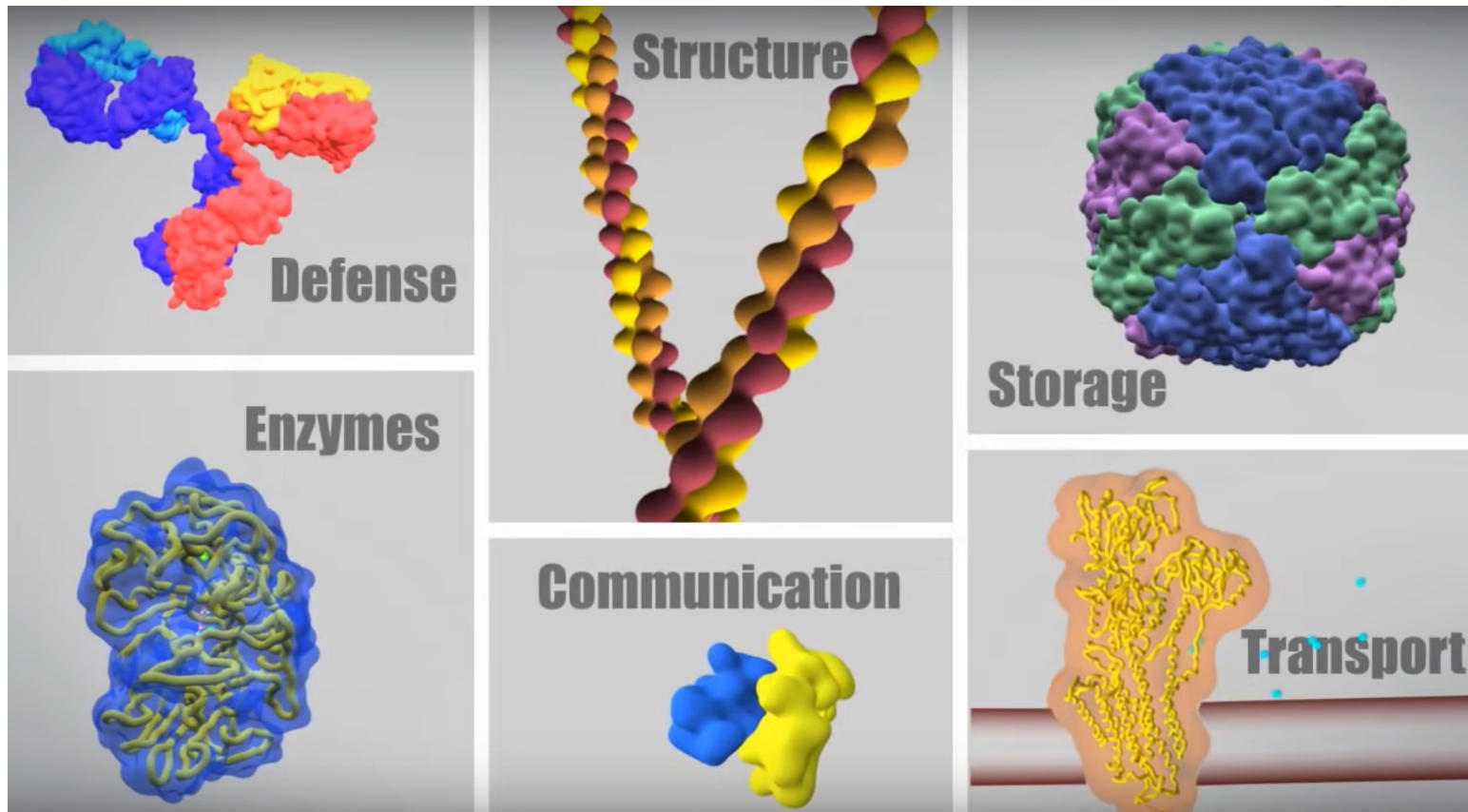
Why are proteins so interesting?



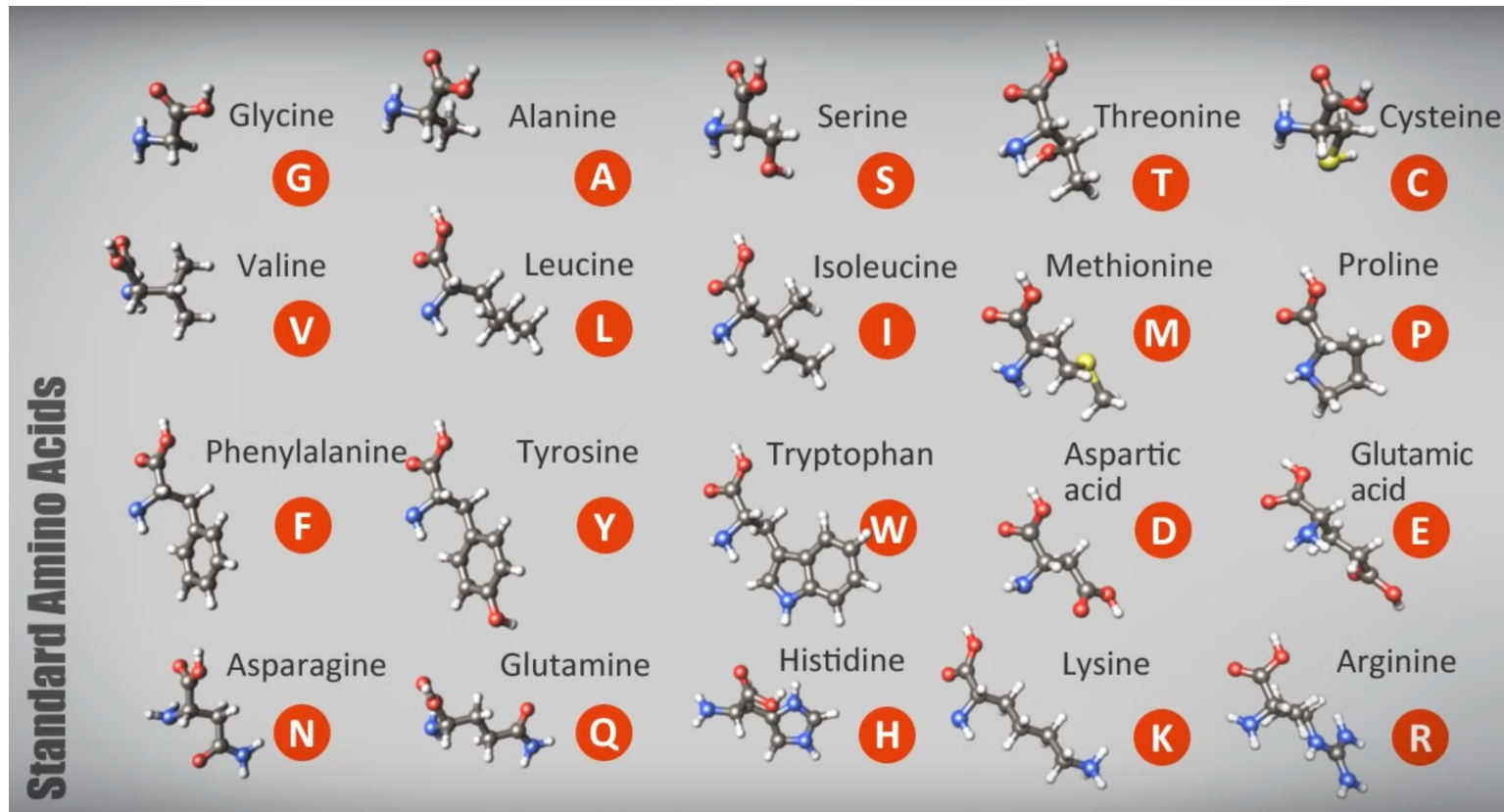
Why are proteins so interesting?



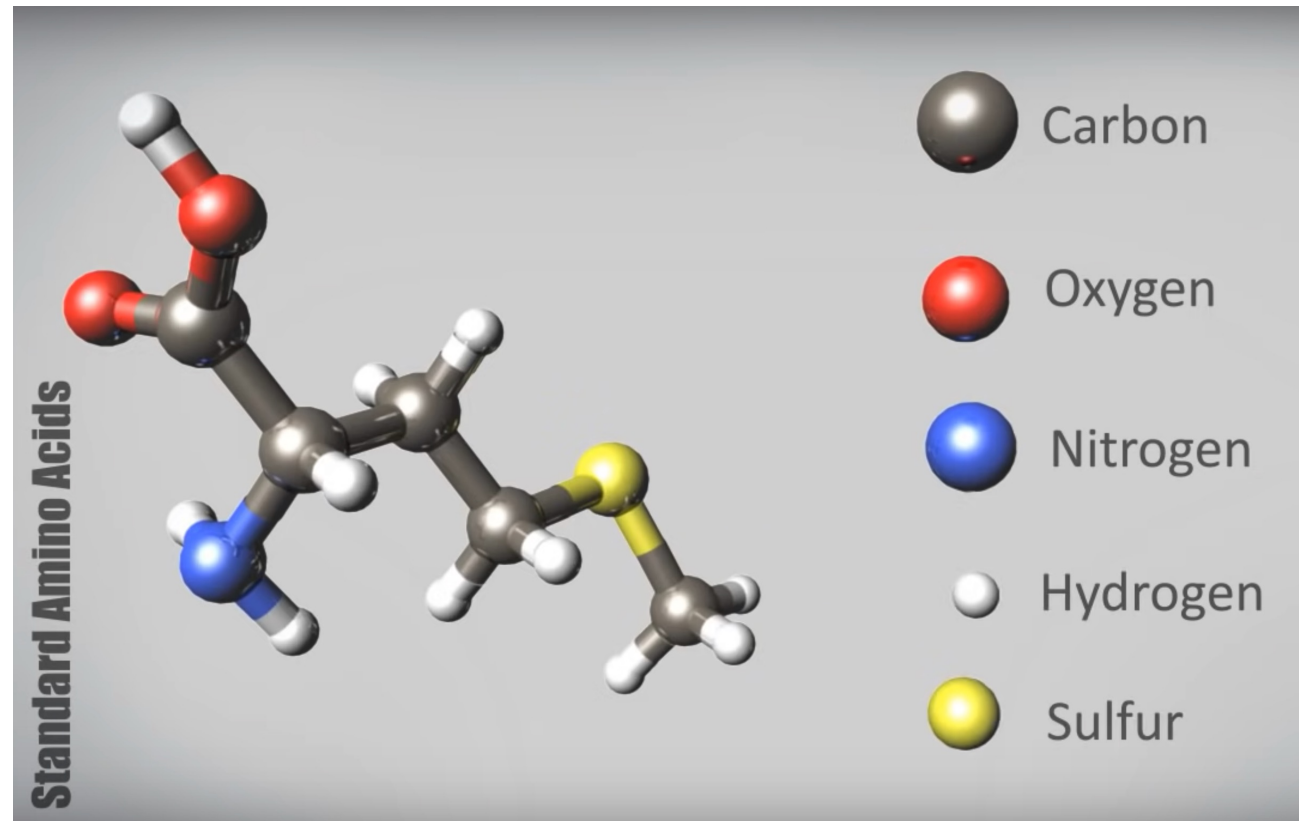
Why are proteins so interesting? **FUNCTION**



Protein's building blocks: the amino acids

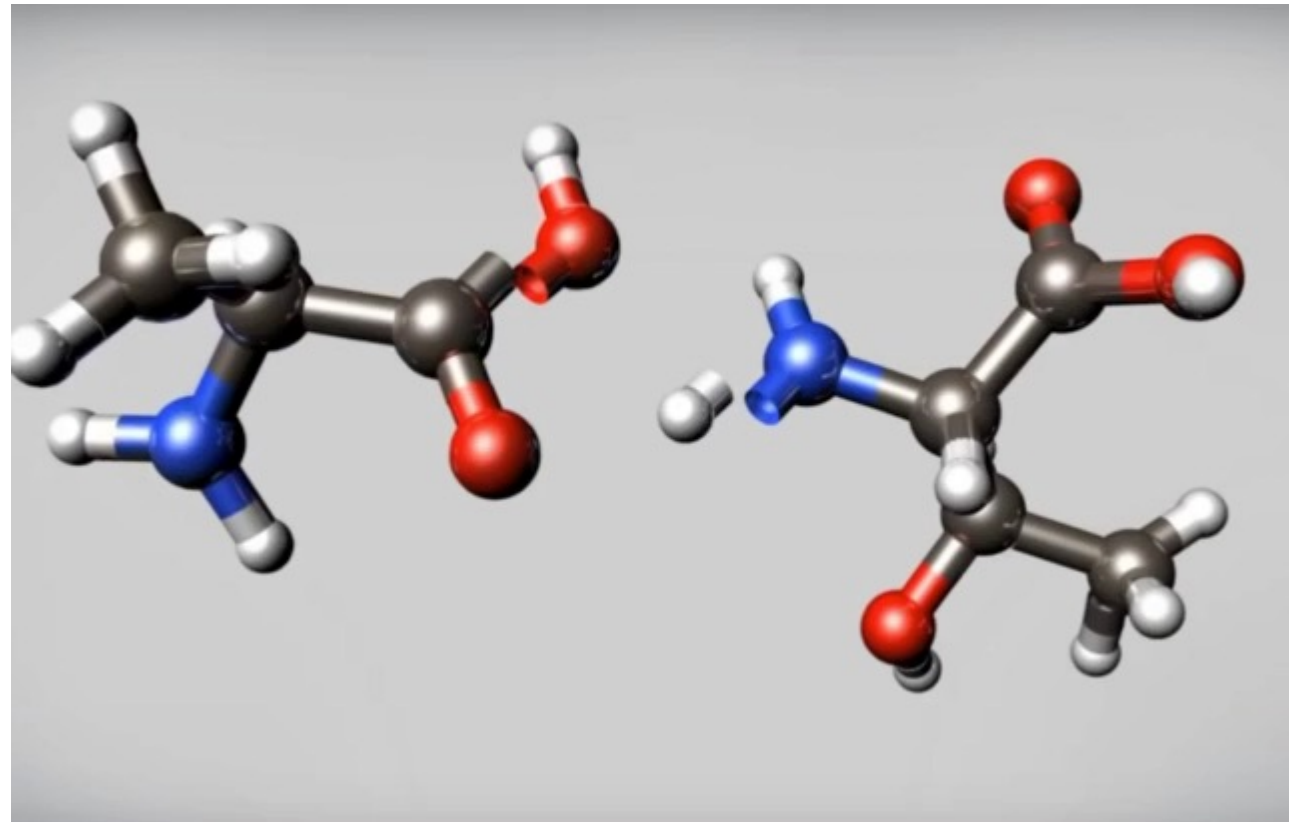
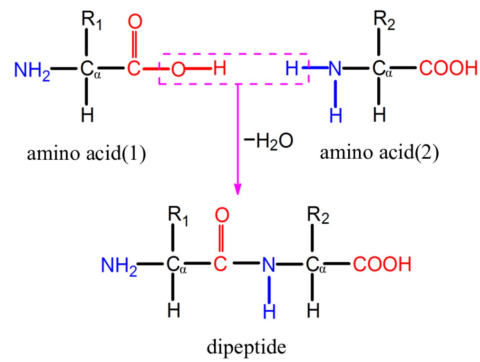


What are they made of?

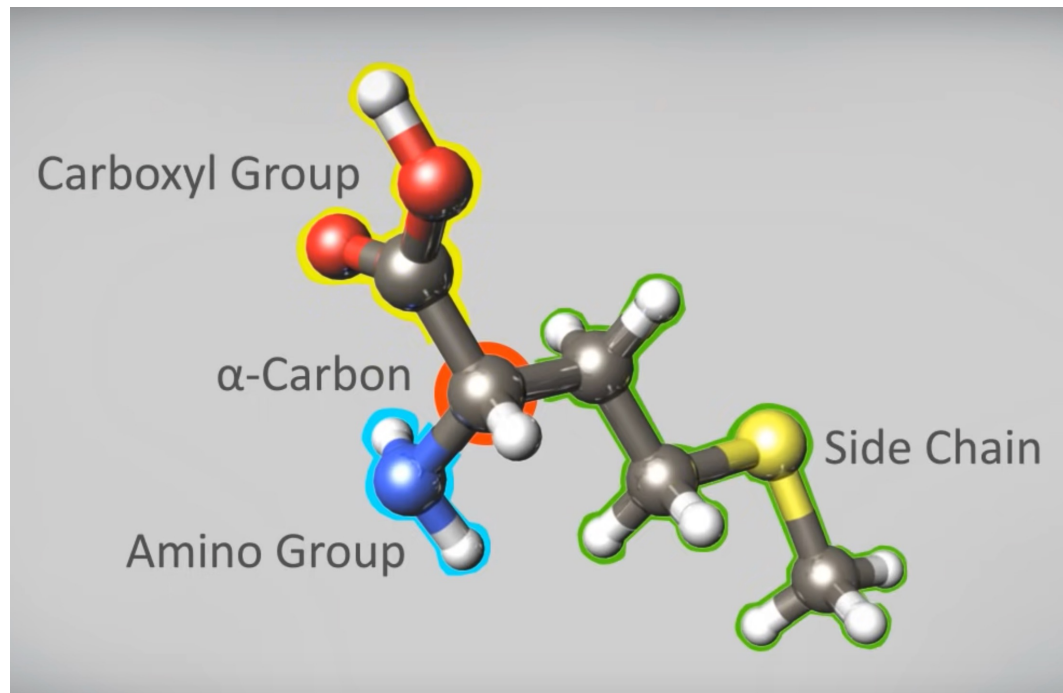


Primary structure

How polypeptide chains are formed?



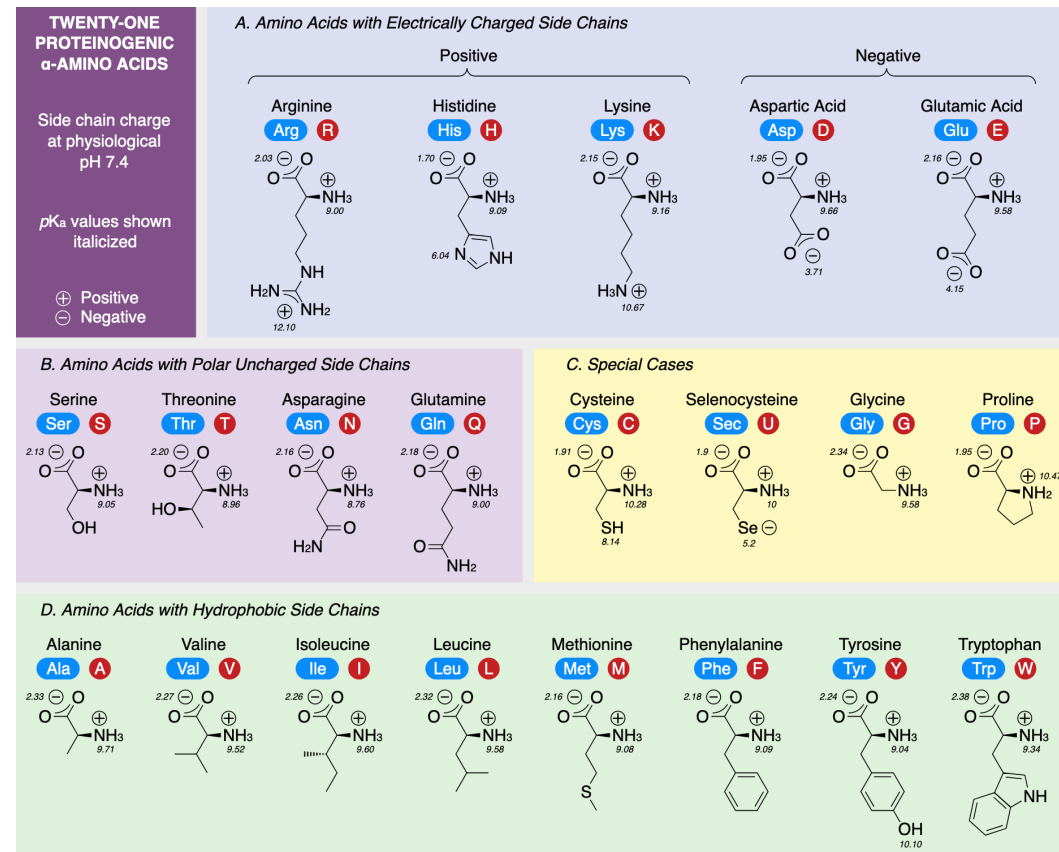
Amino acids are chiral (L - amino acids)



- Carboxyl group
 $-\text{COO}^- \rightleftharpoons -\text{COOH}$
- Amino group
 $-\text{NH}_2 \rightleftharpoons -\text{NH}_3^+$
- α -Carbon
- Side chain

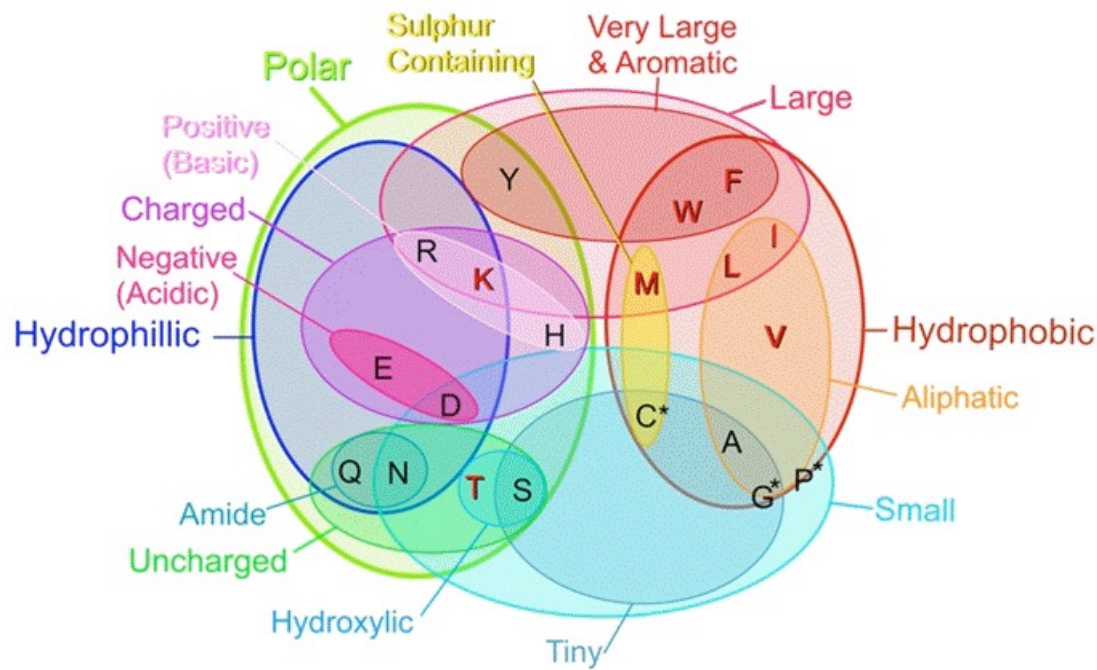
How to group the amino acids?

- 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

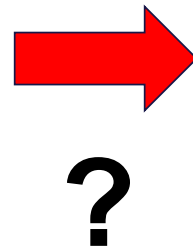
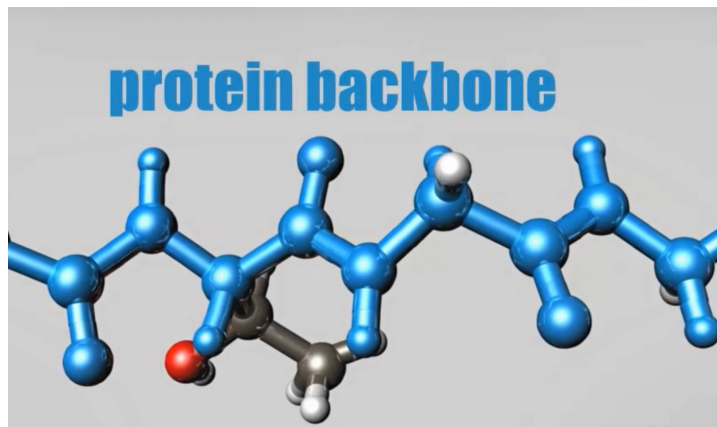
- A alanine (ala)
- C cysteine (cys)
- D aspartic acid (asp)
- E glutamic acid (glu)
- F phenylalanine (phe)
- G glycine (gly)
- H histidine (his)
- I isoleucine (ile)
- K lysine (lys)
- L leucine (leu)
- M methionine (met)
- N asparagine (asn)
- P proline (pro)
- Q glutamine (gln)
- R arginine (arg)
- 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

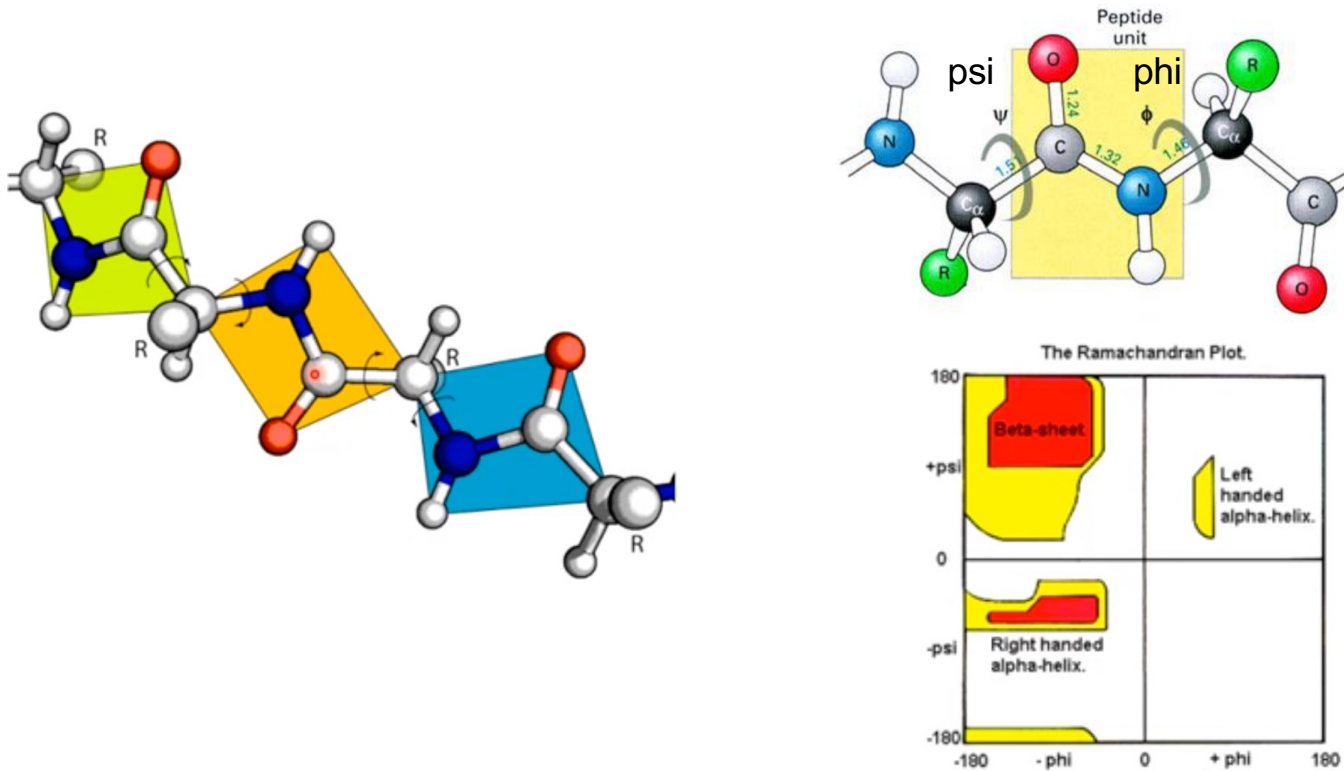
- Hydrophobicity (Entropy)
- Salt bridges
- H bonds
- Di-sulfide bridges
- Aromatic interactions

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

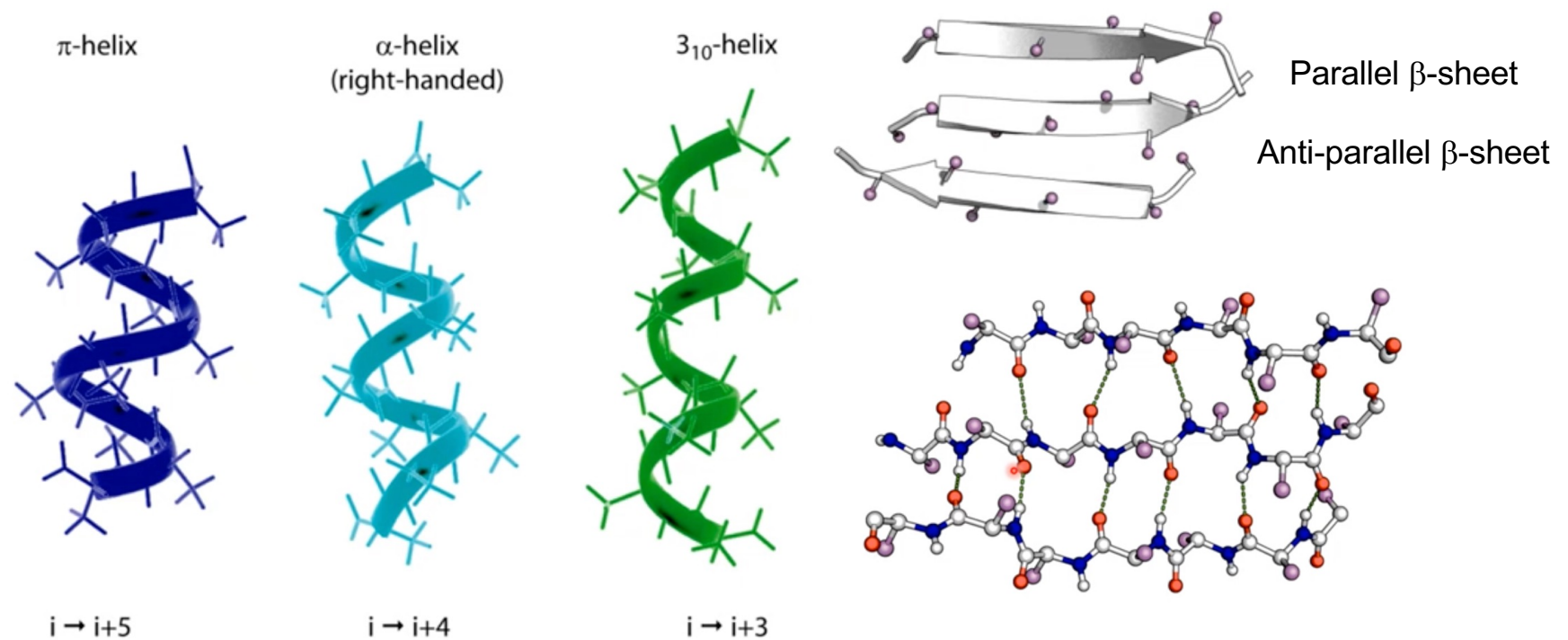
Proteins reach the conformation that minimizes their Free Energy



Can polypeptide chains adopt any conformation?

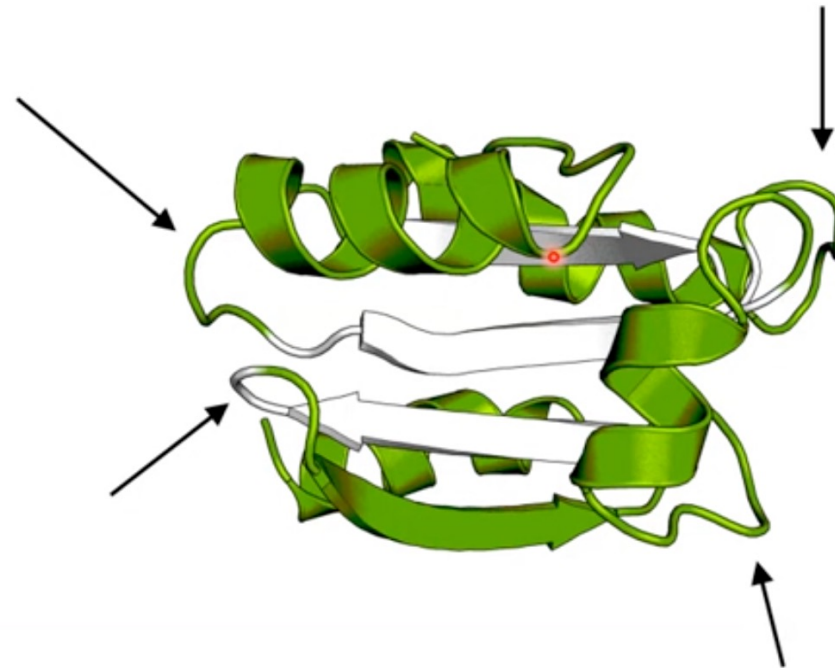


Secondary structure: α -helices and β -sheets



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

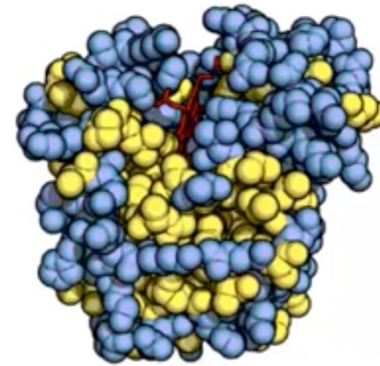
Secondary structure: Turn, loops and bends and disordered regions



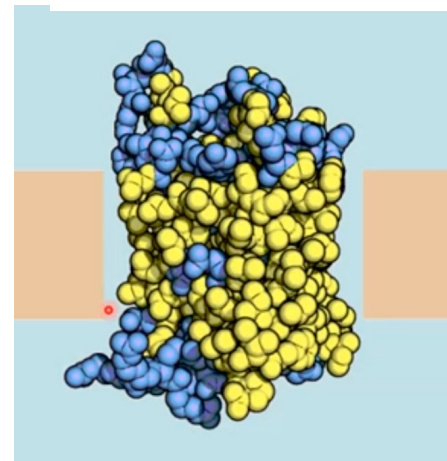
Disordered regions are regions with **non-stable** conformations

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

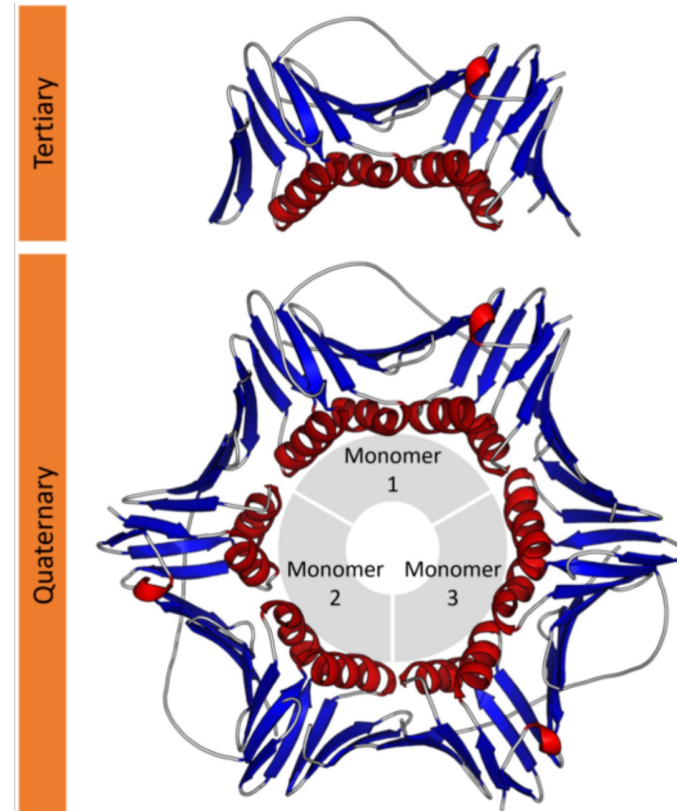


aquaporin

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?

RCSB PDB Protein Data Bank

195,565 Structures from the PDB

1,000,361 Computed Structure Models (CSM)

3D Structures

Enter search term(s), Entry ID(s), or sequence

Include CSM

Advanced Search | Browse Annotations

Help

PDB-101 PDB EMDataResource NUCLEIC ACID DATABASE wwPDB Foundation

NEW! Computed Structure Models (CSM) Learn more

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive
- Computed Structure Models (CSM)** from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

COVID-19 CORONAVIRUS Resources

Join the RCSB PDB Team

September Molecule of the Month

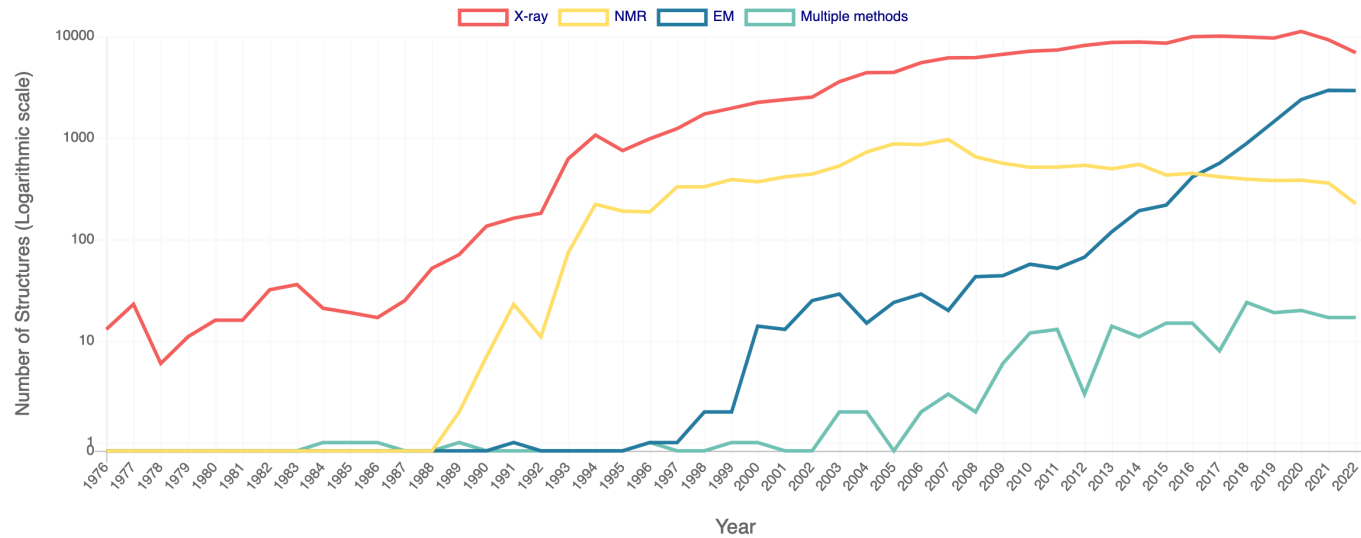
Respiratory Supercomplex

<https://www.rcsb.org/>

How do we experimentally determine the protein structure?

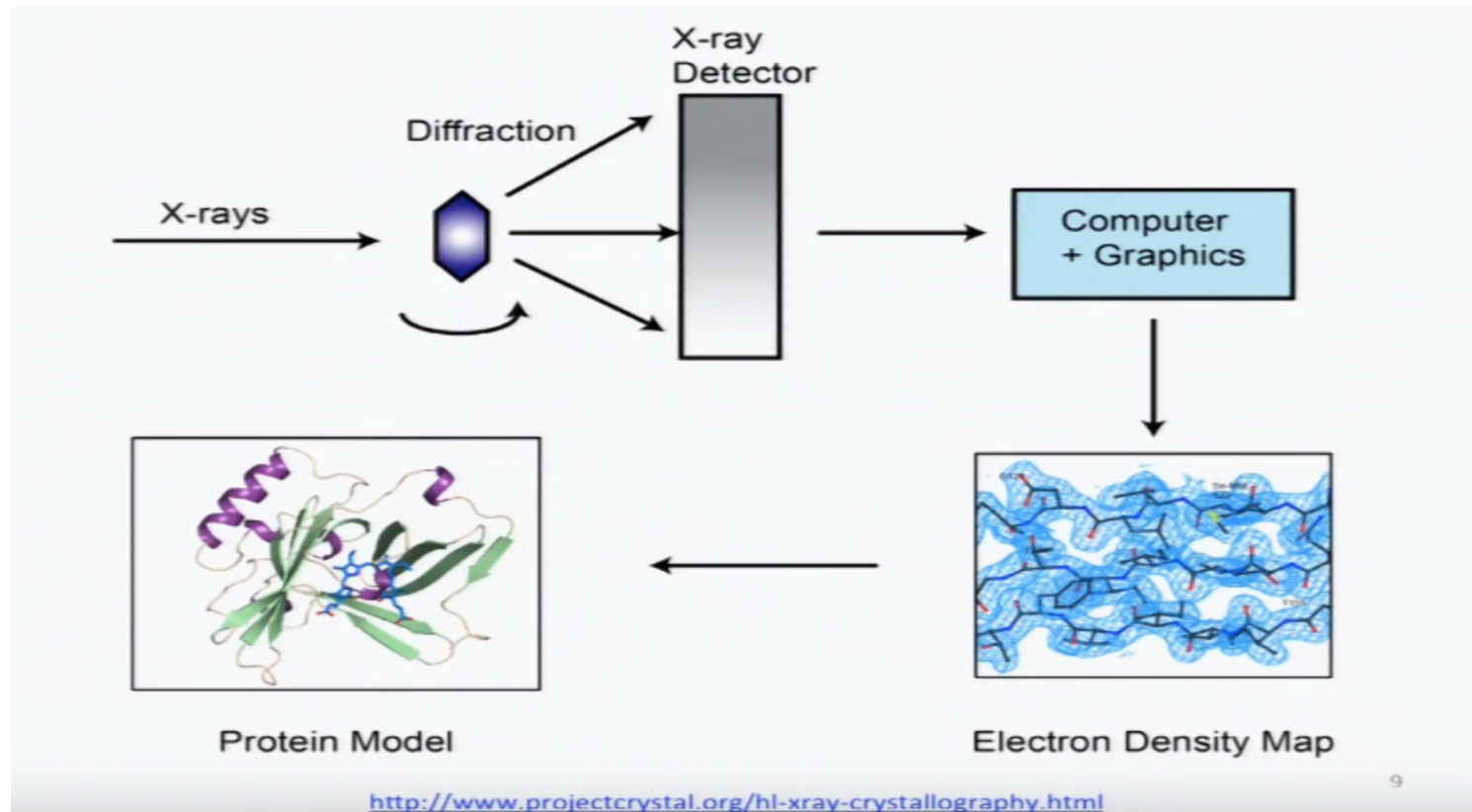
Number of Released PDB Structures per Year

All Statistics

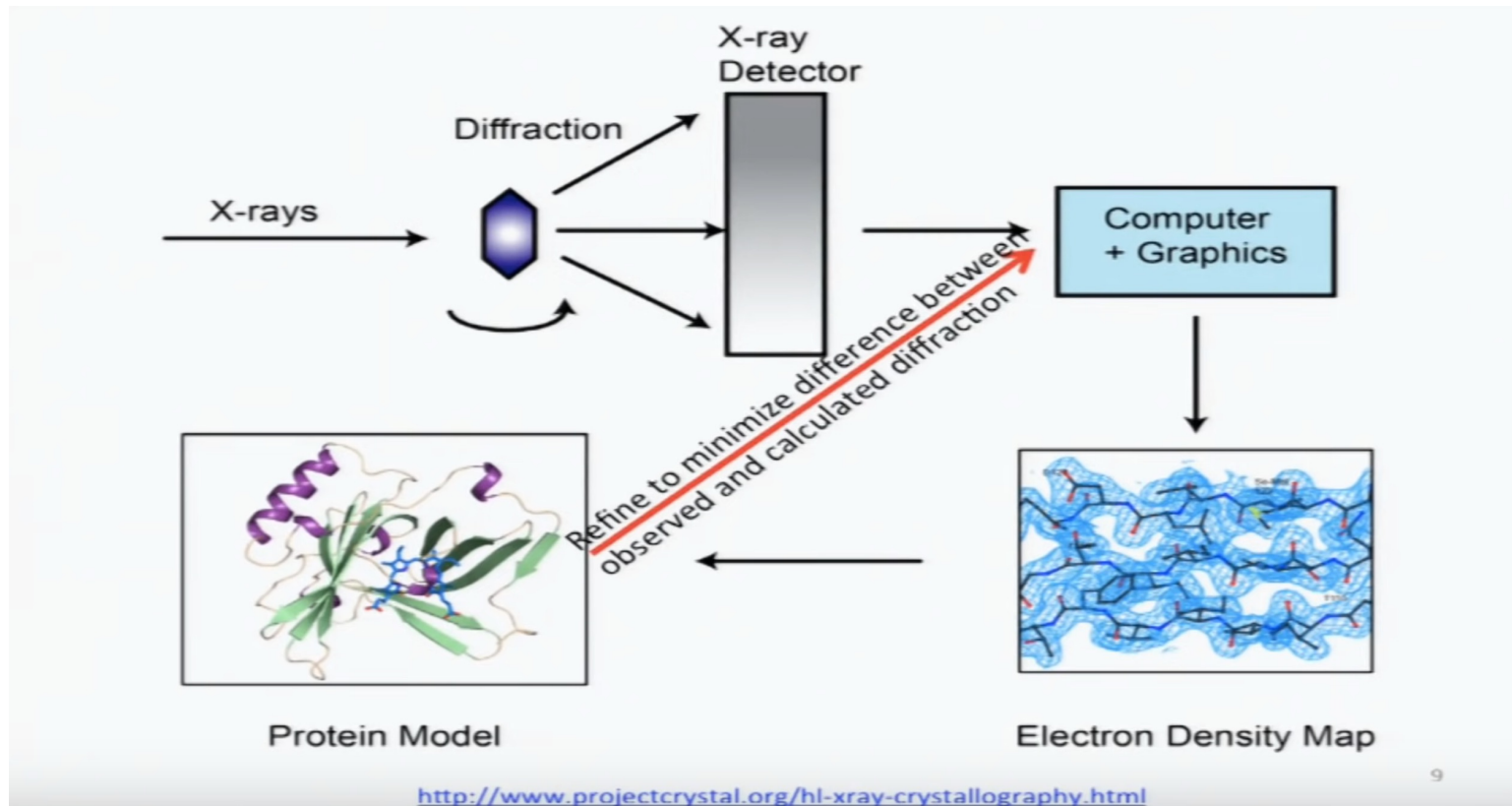


- X-ray: X-RAY DIFFRACTION, FIBER DIFFRACTION, or POWDER DIFFRACTION
- NMR: SOLUTION NMR or SOLID-STATE NMR
- EM: ELECTRON MICROSCOPY or ELECTRON CRYSTALLOGRAPHY or ELECTRON TOMOGRAPH.
- MULTIPLE METHODS: Multiple experimental methods. For example, if a structure is solved by X-RAY DIFFRACTION AND NEUTRON DIFFRACTION, it will be counted only in

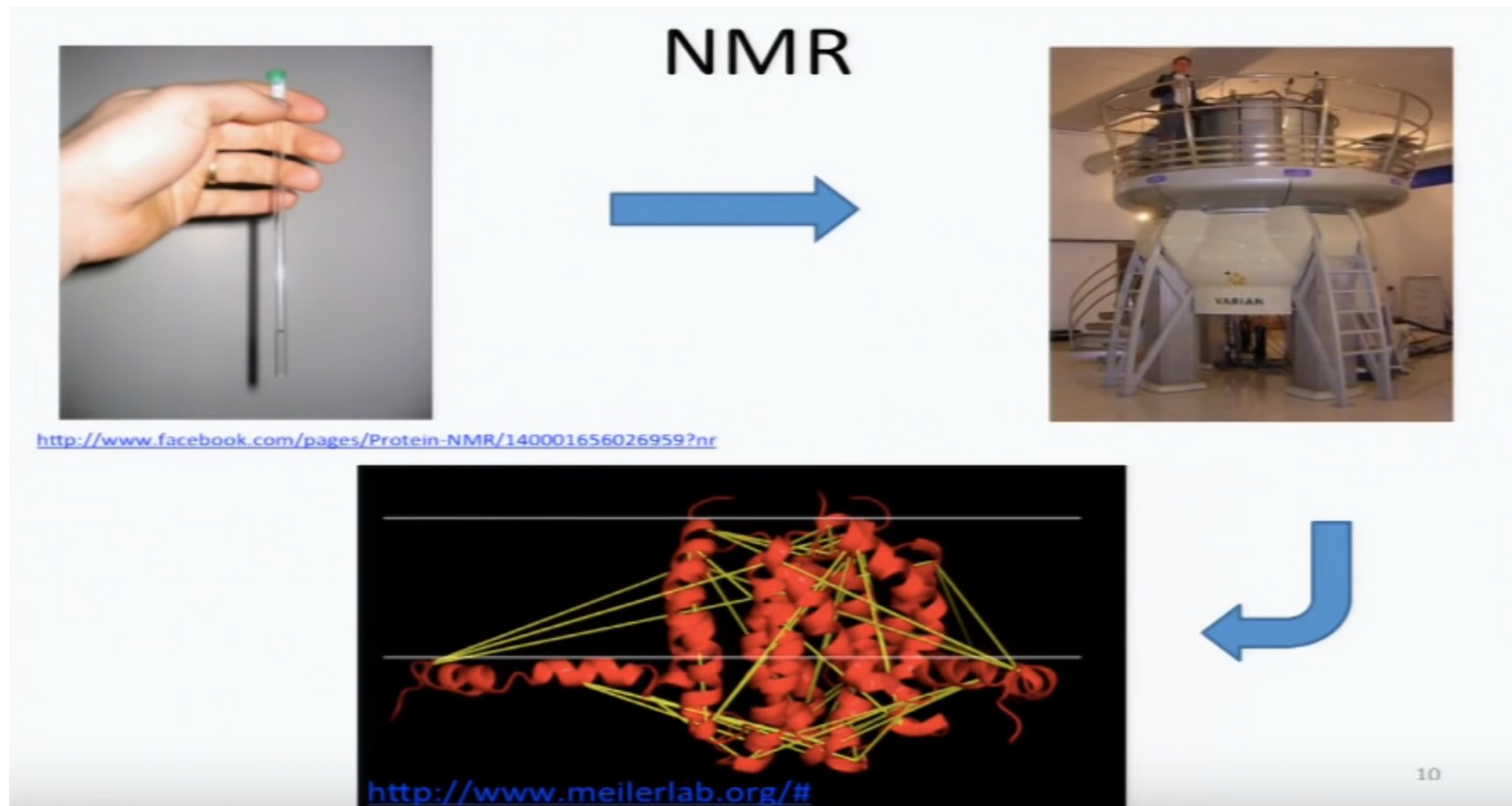
Overview of the X-ray crystallographic method



Overview of the X-ray crystallographic method

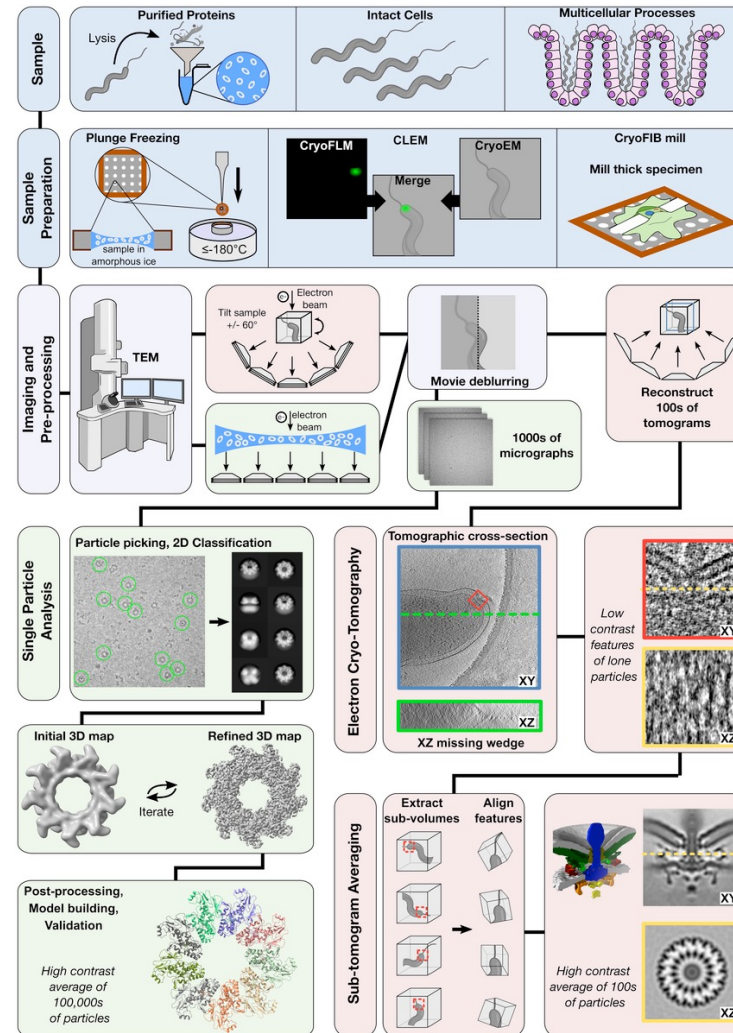


Nuclear magnetic resonance spectroscopy of proteins



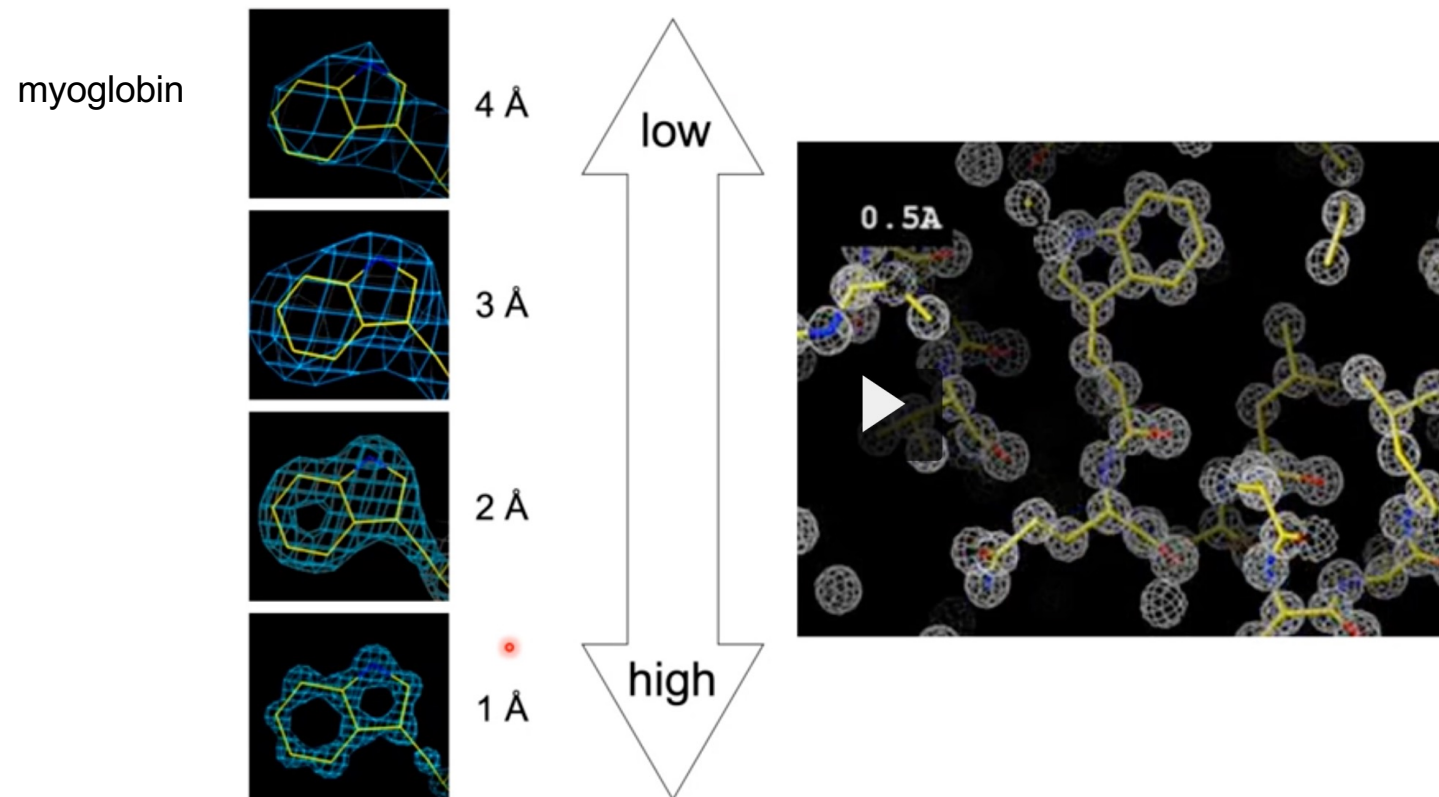
Cryo – Electron Microscopy EM

- 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



Quality of the protein structure: Goodness of model fitness

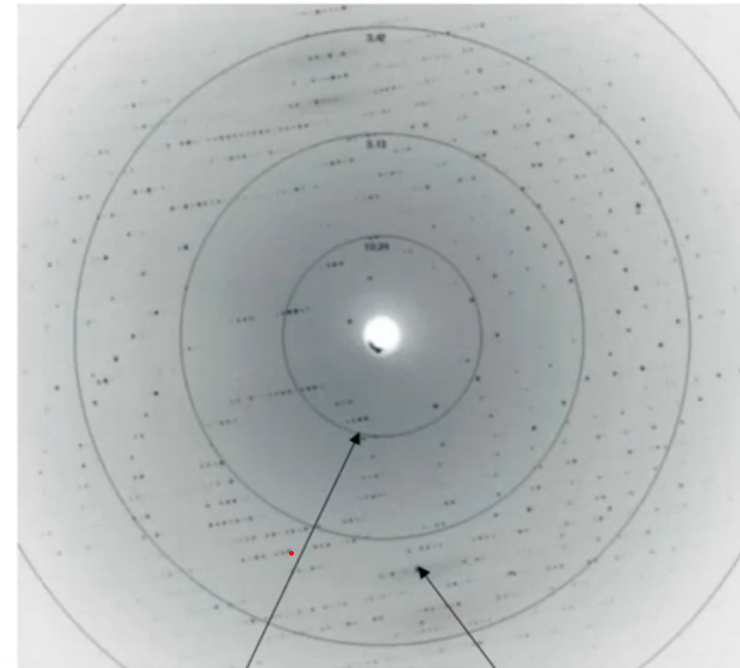
- Individual reflections

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

- R-factor:

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

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



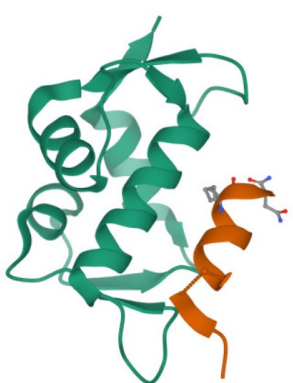
Resolution
rings

Reflection(s)

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](#)

6V4E

Crystal Structure Analysis of Zebra Fish MDM

PDB DOI: [10.2210/pdb6V4E/pdb](https://doi.org/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

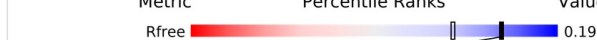


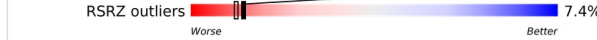
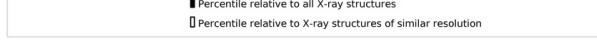
R-Value Free: 0.199

R-Value Work: 0.160

R-Value Observed: 0.162

wwPDB Validation ⓘ

3D Report
Full Report

Metric	Percentile Ranks	Value
Rfree		0.198
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		7.4%

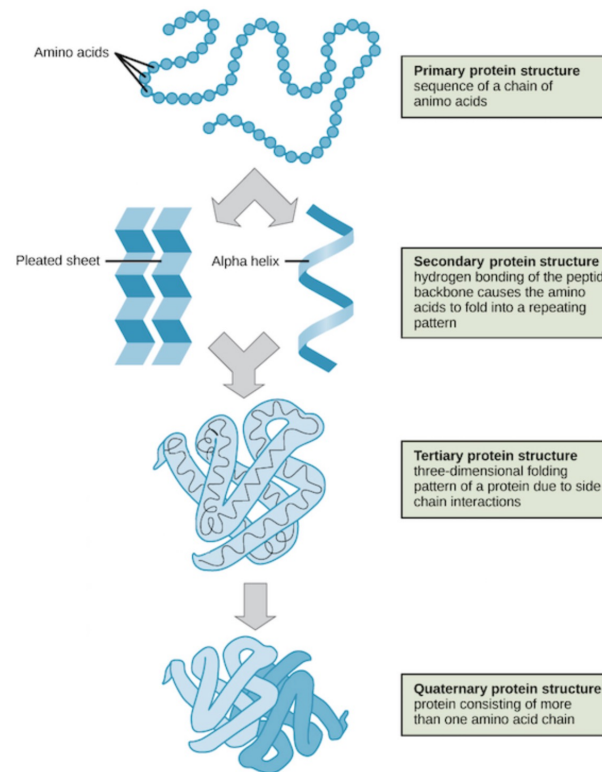
Percentile relative to all X-ray structures
 Percentile relative to X-ray structures of similar resolution

27 September 2022

DTU Bioinformatics

30

Summary: The four levels of protein structure



Video link to protein structure



[RCSBProteinDataBank](https://www.rcsb.org/)

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