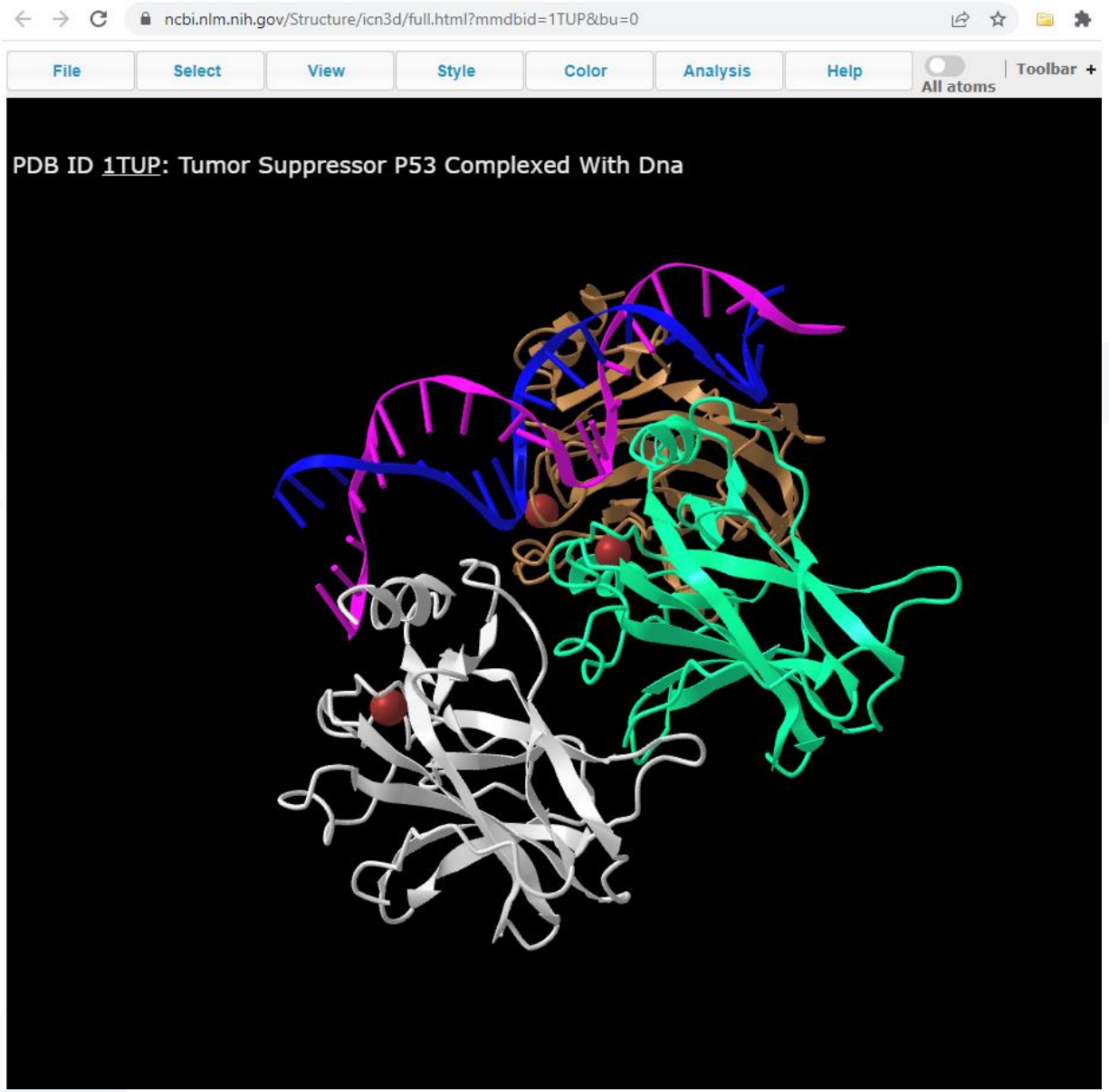
Exploring 3D Molecular Structures with iCn3D

Alexa M. Salsbury, Ph.D.

National Library of Medicine
National Center for Biotechnology Information

Overview

- Background
- iCn3D Fundamentals
  - Selection
  - Coloring
  - Style
- Individual work & office hours



https://www.nlm.nih.gov/ncbi/workshops/ASBCB_2023-04_3d-molecular-structures/workshop_details.html

Structural Biology

1952-1953- Pioneering DNA structure work by Wilkins, Franklin, Watson, & Crick.

1956-1960- Rich & Davies' structural experiments showed how information could be transferred from DNA to RNA.

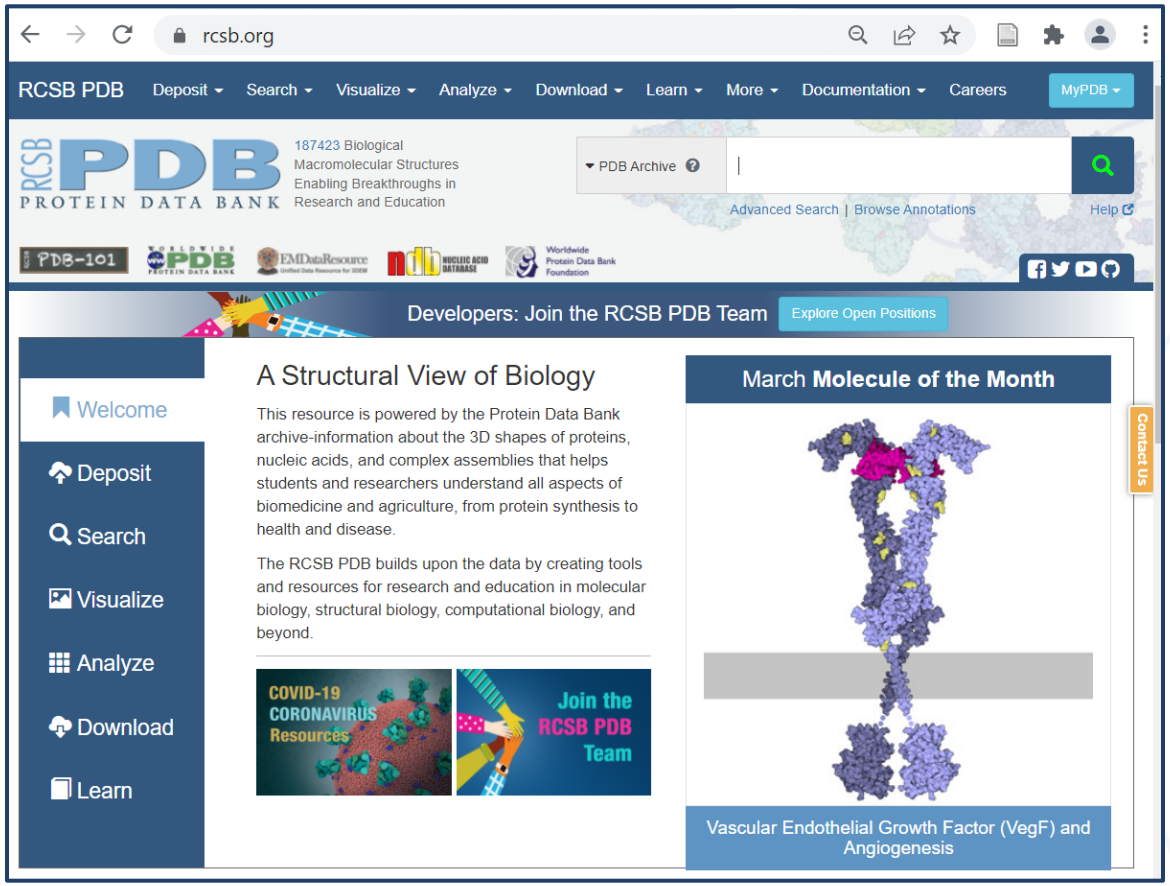
1957- The first protein with a crystal structure was solved in by Kendrew and co-workers

Now- over 175,000 structures are publicly available and structure prediction is improving!

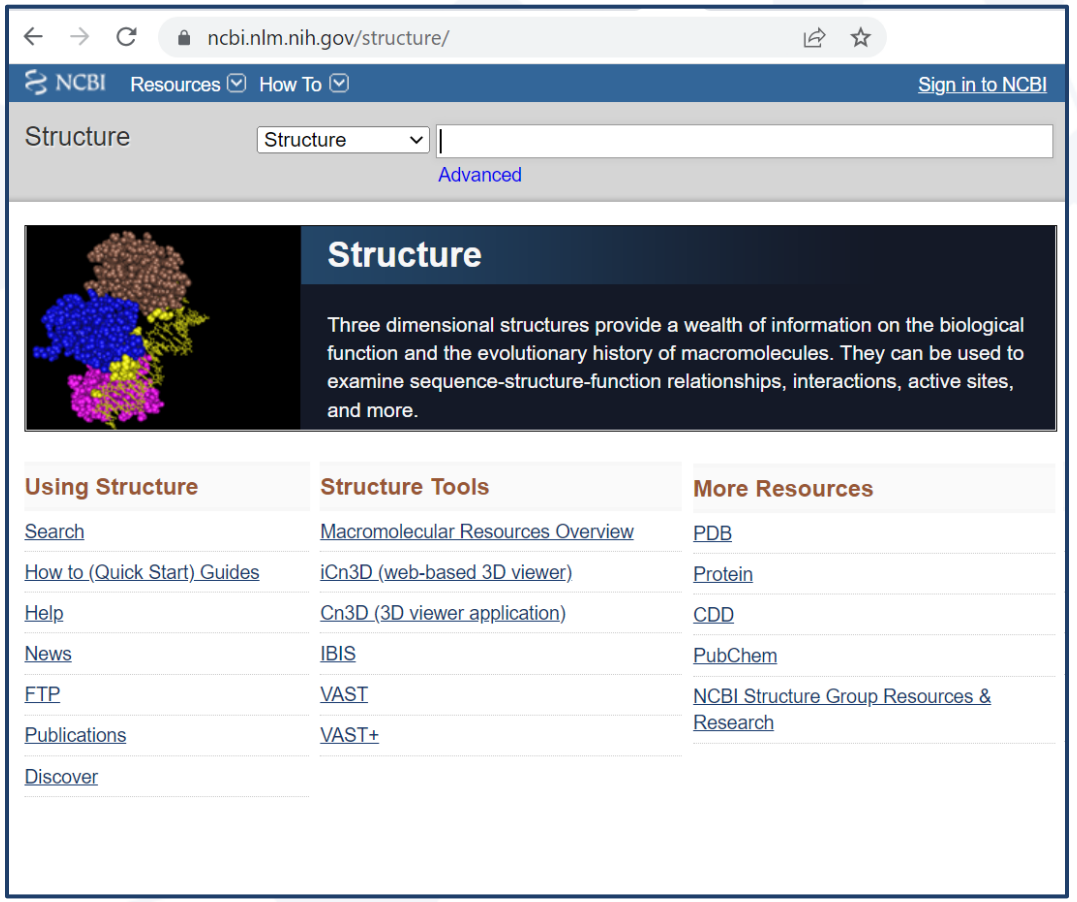
Experimental techniques

	Advantages	Disadvantages
X-ray crystallography	• Well developed • High resolution • Broad molecular weight range	• Difficult sample prep • Static crystalline state
NMR	• High resolution • 3D structure in solution • Good for dynamic study	• Difficult sample prep • High sample purity needed • Static crystalline state captured
Cryo-EM	• Simple sample prep • Structure in native state • Small sample size needed	• Lower resolution • Works best for samples with high molecular weight • Equipment can be expensive, but costs are decreasing

Where do I find experimentally determined structures?



RCSB Protein Data Bank



NCBI Structure Database

Protein Data Bank (PDB)

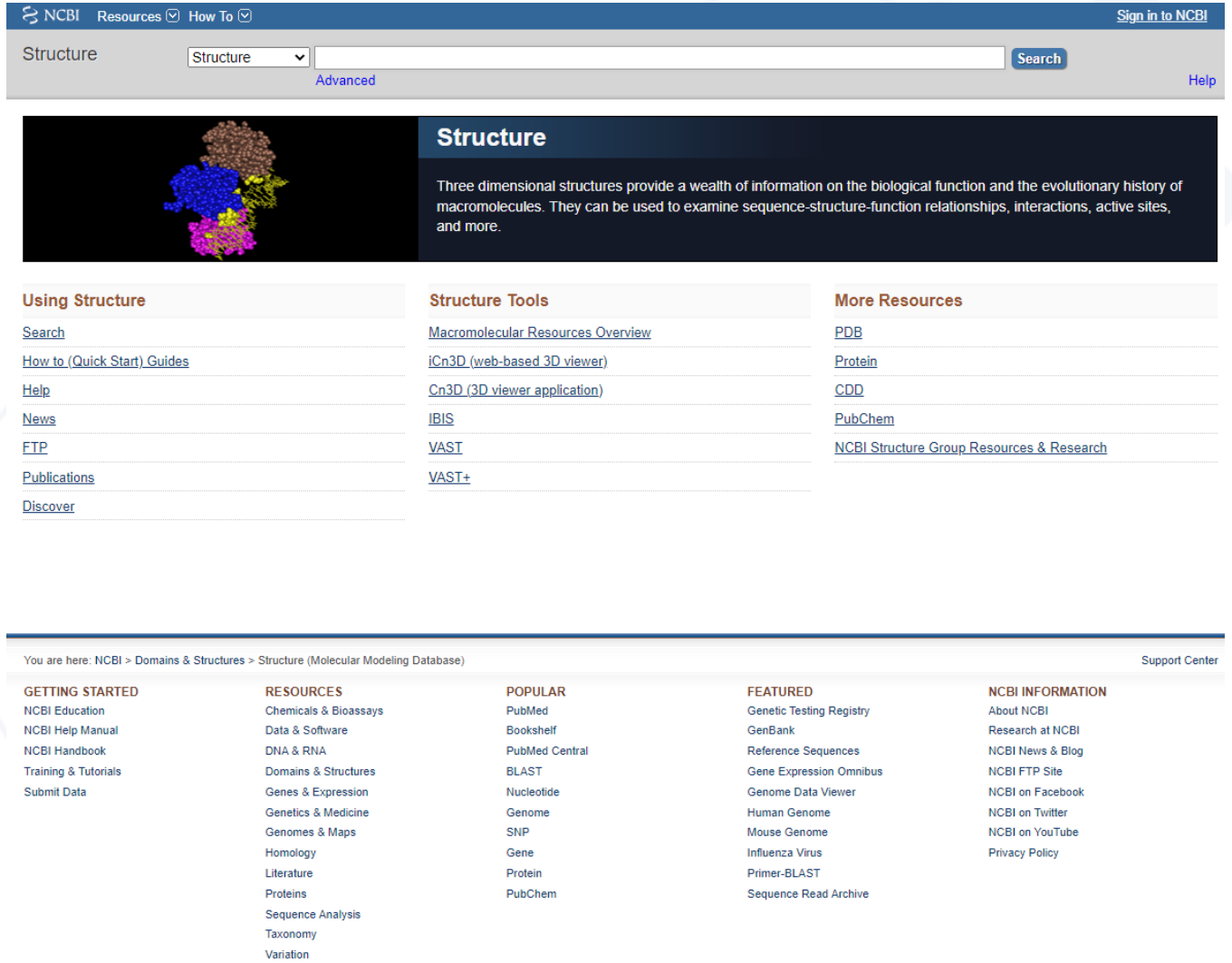
- New Structures are deposited daily

Each structure contains:

- 3D atomic coordinates
- Mandatory Metadata
  - Author Information
  - Primary citation
  - Experimental Data
  - Polymer sequence(s)- proteins, DNA, RNA
  - Small Chemical component structures- ligands, inhibitors, etc.

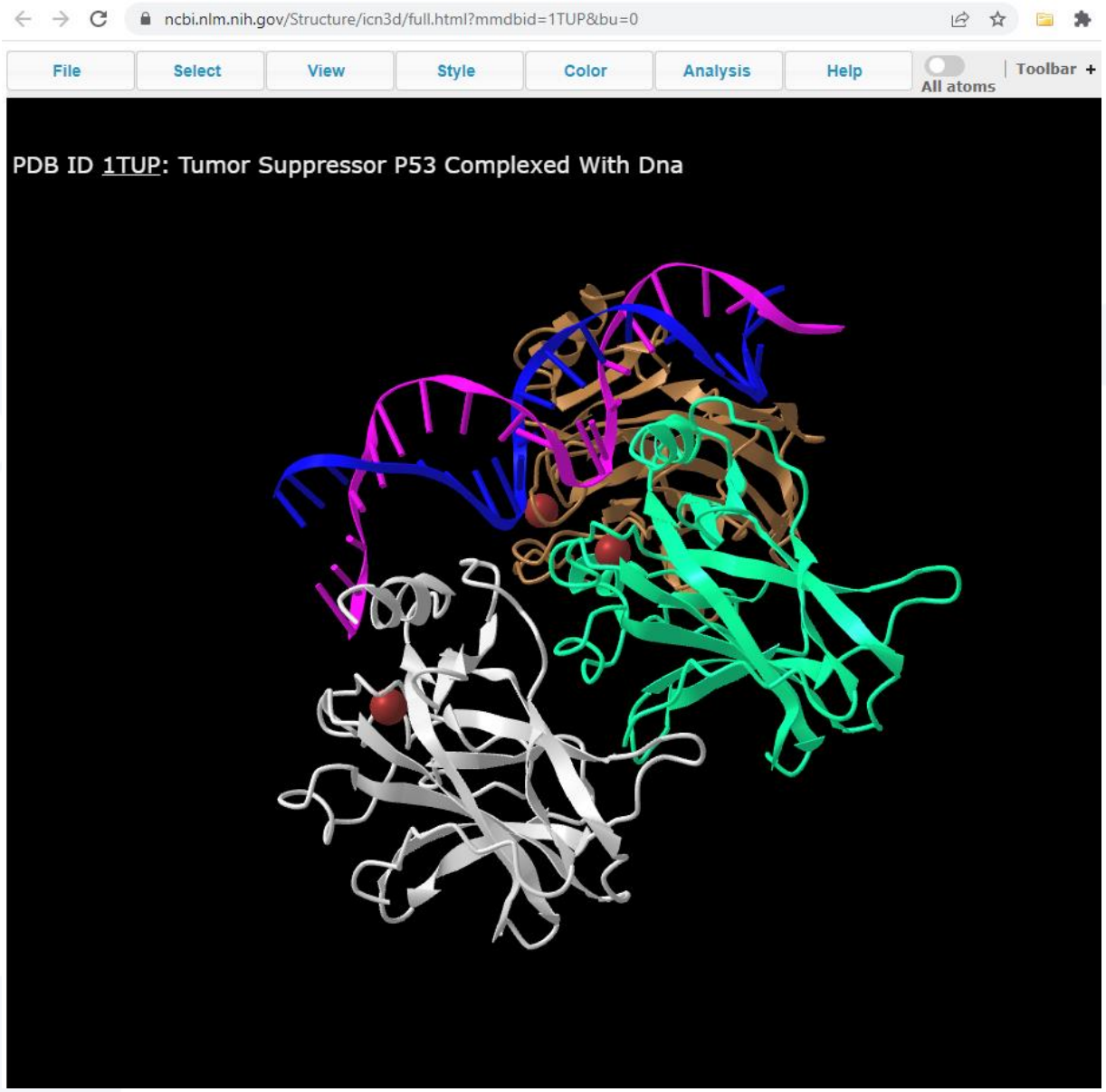
NCBI's Structure Database

- Updated monthly
- Derived from PDB records
- Additional information added, including:
  - Explicit chemical graph information
  - Validation (secondary structure elements)
  - Includes taxonomy
- Connects 3D to associated literature, molecular data, chemical data, and other NCBI tools



iCn3D

- Interactive, web-based 3D structure viewer
  - No installation needed!
- Users can
  - Visualize structure in 1D, 2D, and 3D
  - View sequence and structure alignments
  - Probe perturbations
  - And more!



3D Viewer Feature Comparison

	Web-based	1D Sequence	2D Diagram	Annotation	Align	Share Link	Script	Jupyter Notebook	Virtual Reality	3D Printing
iCn3D	✓	✓	✓	✓	✓[a]	✓[b]	✓[c]	✓[d]	✓	✓
Mol*	✓	✓	Web	Web						
Aquaria	✓	✓		✓					✓	
Chimera		✓		✓					✓	✓
PyMol		✓		✓			✓			
Cn3D		✓	Web	✓	✓					

[a]: iCn3D aligns structures (PDB or AlphaFold) based on structures or sequences.
[b]: iCn3D sharable links could be a short URL or a URL containing the address of an iCn3D PNG Image
[c]: iCn3D supports command-line analysis with either Python scripts or Node.js scripts
[d]: iCn3D can also be used in Jupyter Notebook

iCn3D Features of Interest

- iCn3D aligns structures (PDB or AlphaFold) based on structures or sequences.
- iCn3D sharable links (https://structure.ncbi.nlm.nih.gov/icn3d/share.html?XCxR6fSTmXHxR3o1A)
- iCn3D supports command-line analysis with either Python scripts or Node.js scripts
- iCn3D can also be used in Jupyter Notebook (https://pypi.org/project/icn3dpy)
- 3D printing: structure.ncbi.nlm.nih.gov/icn3d/share.html?wt4TDqzhC2rhCYTD7
- Contact map: structure.ncbi.nlm.nih.gov/icn3d/share.html?rnMbe26tNsAjJLGK9
- Precalculated symmetry: structure.ncbi.nlm.nih.gov/icn3d/share.html?bGH1BfLsiGFhhTDn8
- Symmetry dynamically: structure.ncbi.nlm.nih.gov/icn3d/share.html?6NvhQ45XrnbuXyGe6
- Electron density map: structure.ncbi.nlm.nih.gov/icn3d/share.html?QpqNZ3k65ToYFvUB6
- EM map: structure.ncbi.nlm.nih.gov/icn3d/share.html?L4C4WYE85tYRiFeK7
- Transmembrane protein: structure.ncbi.nlm.nih.gov/icn3d/share.html?jMN16mJyR9STUx6E6
- Solvent Accessible Area: structure.ncbi.nlm.nih.gov/icn3d/share.html?xKSyfd1umbKstGh29

iCn3D Fundamentals Demo

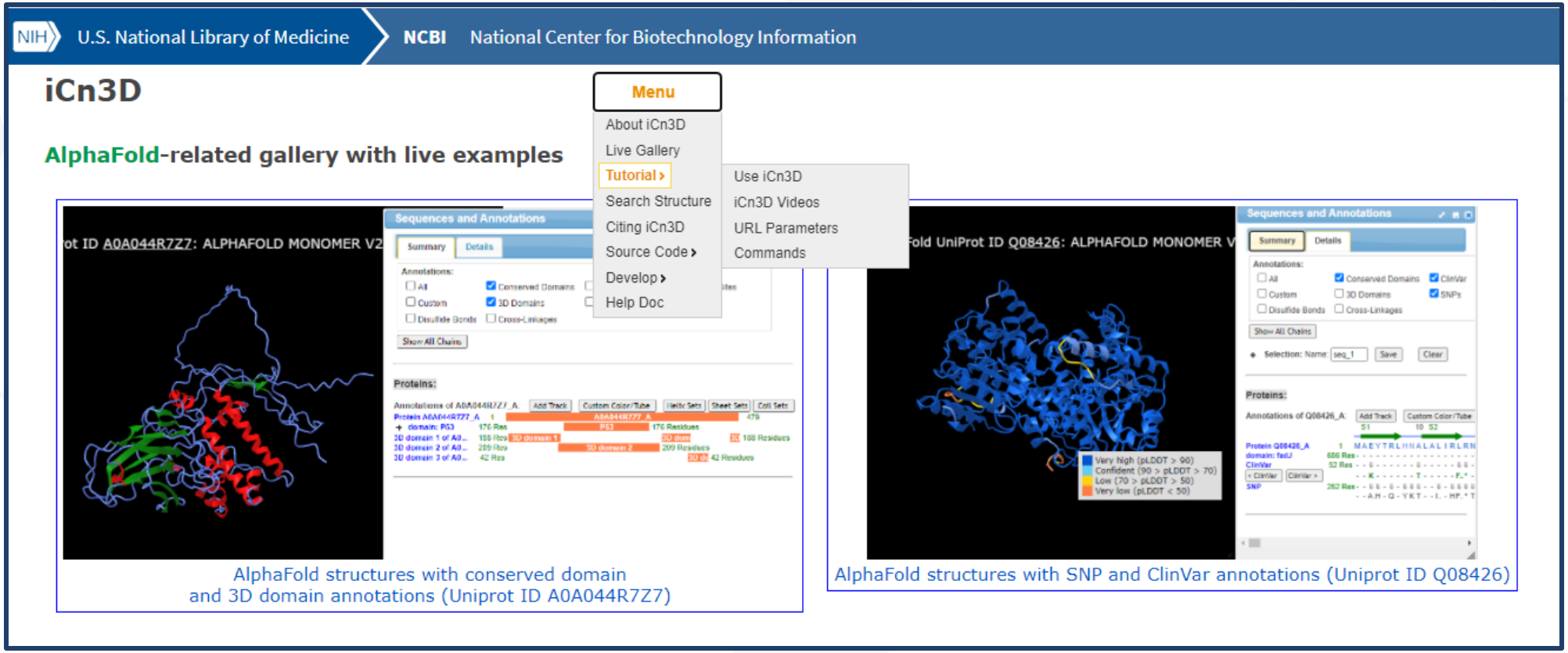
https://www.nlm.nih.gov/ncbi/workshops/ASBCB_2023-04_3d-molecular-structures/icn3d_fundamentals.html

Codeathon Exercise

https://www.nlm.nih.gov/ncbi/workshops/ASBCB_2023-04_3d-molecular-structures/codeathon_exercise.html

Continue learning about iCn3D

Tutorials and help documents are available here:



AlphaFold structures with conserved domain and 3D domain annotations (Uniprot ID A0A044R7Z7)

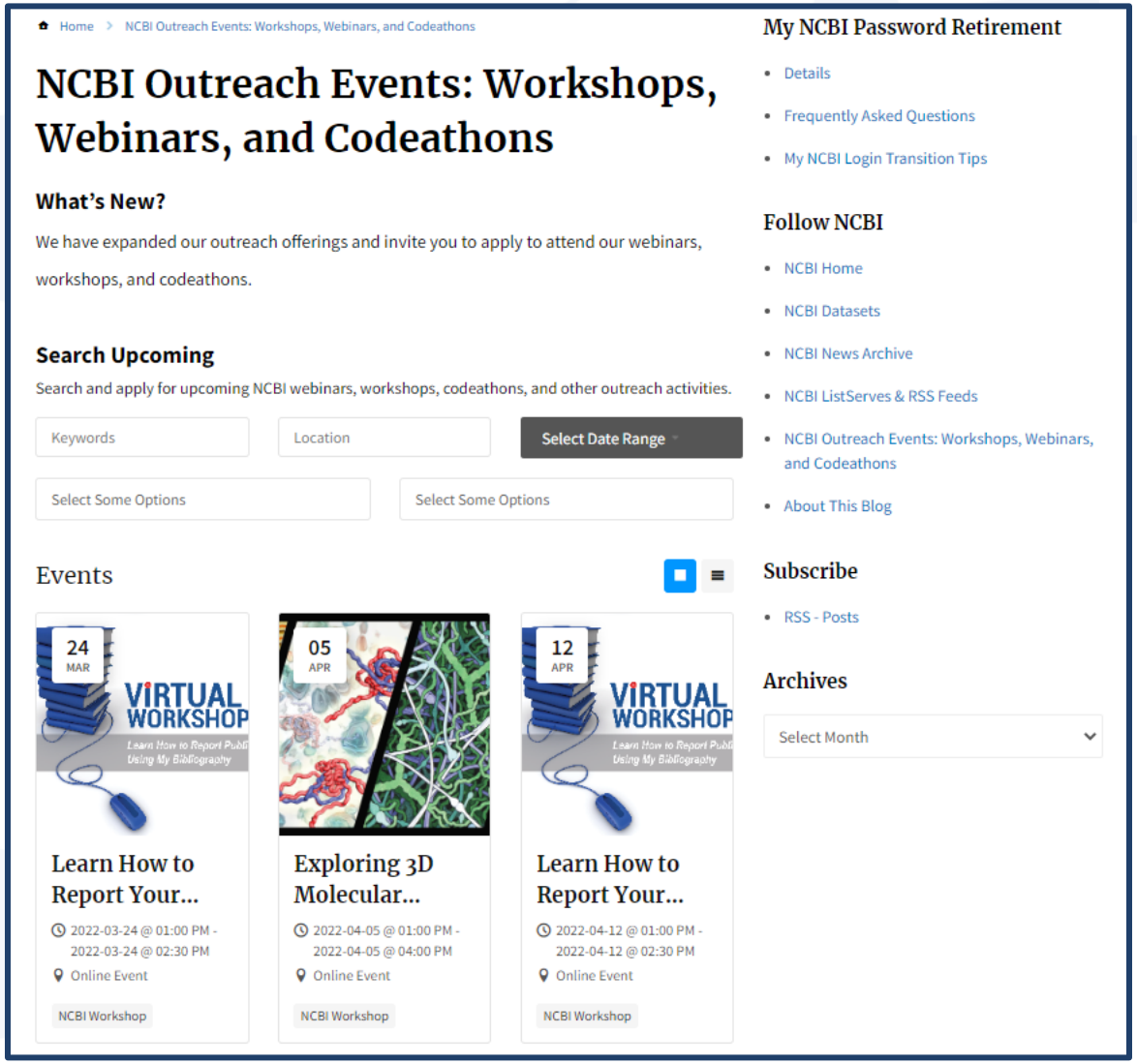
AlphaFold structures with SNP and ClinVar annotations (Uniprot ID Q08426)

Continue learning about NCBI Resources

- Join us for workshops, webinars, or codeathons!

  NCBI Insights Blog

- Follow us on social media:
  - Twitter
  - LinkedIn
  - Facebook



Questions & Discussion

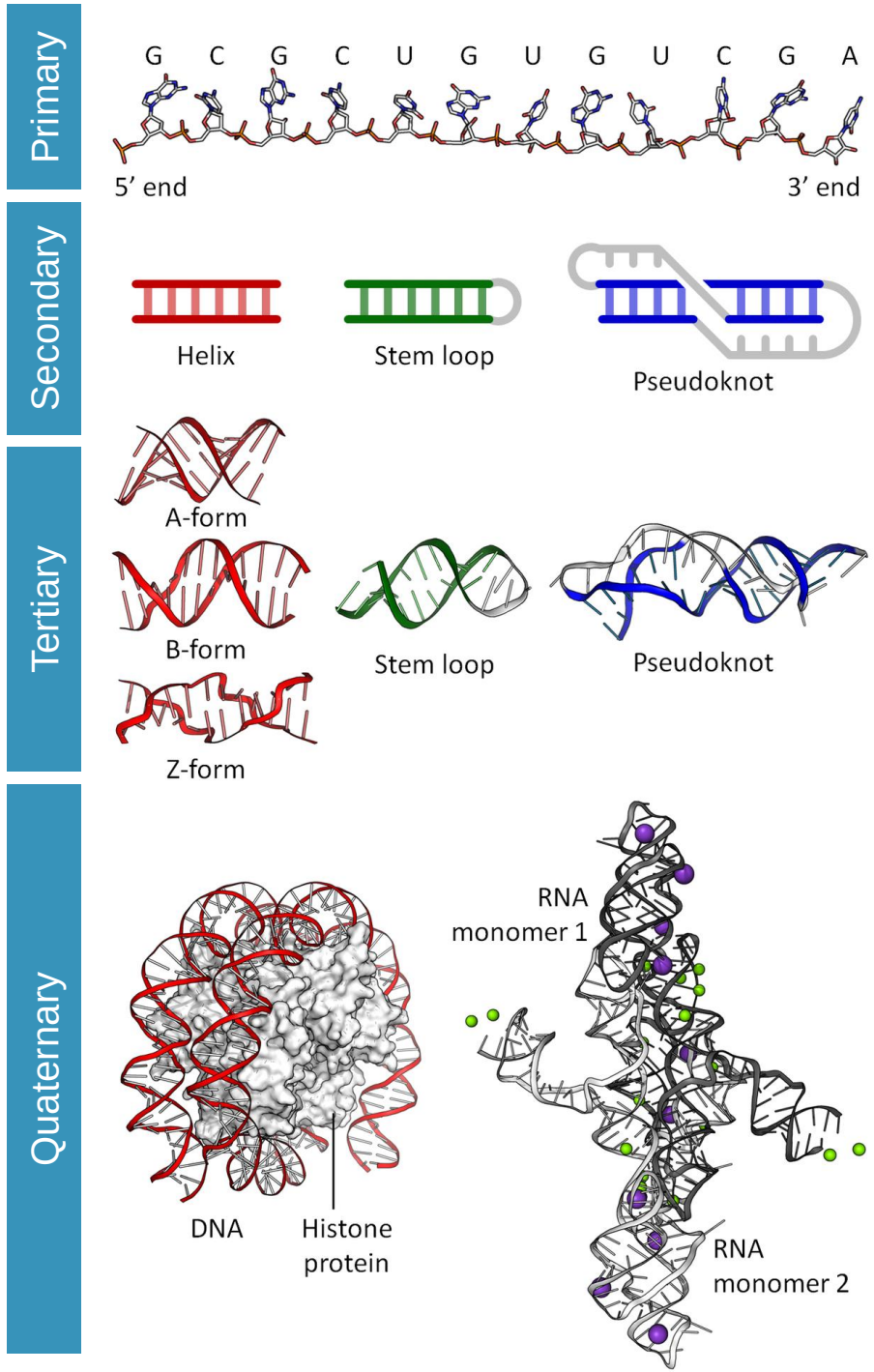
Exploring 3D Molecular Structures with iCn3D Supplemental Learning Materials

Alexa M. Salsbury, Ph.D.

National Library of Medicine
National Center for Biotechnology Information

Nucleic Acid Structure

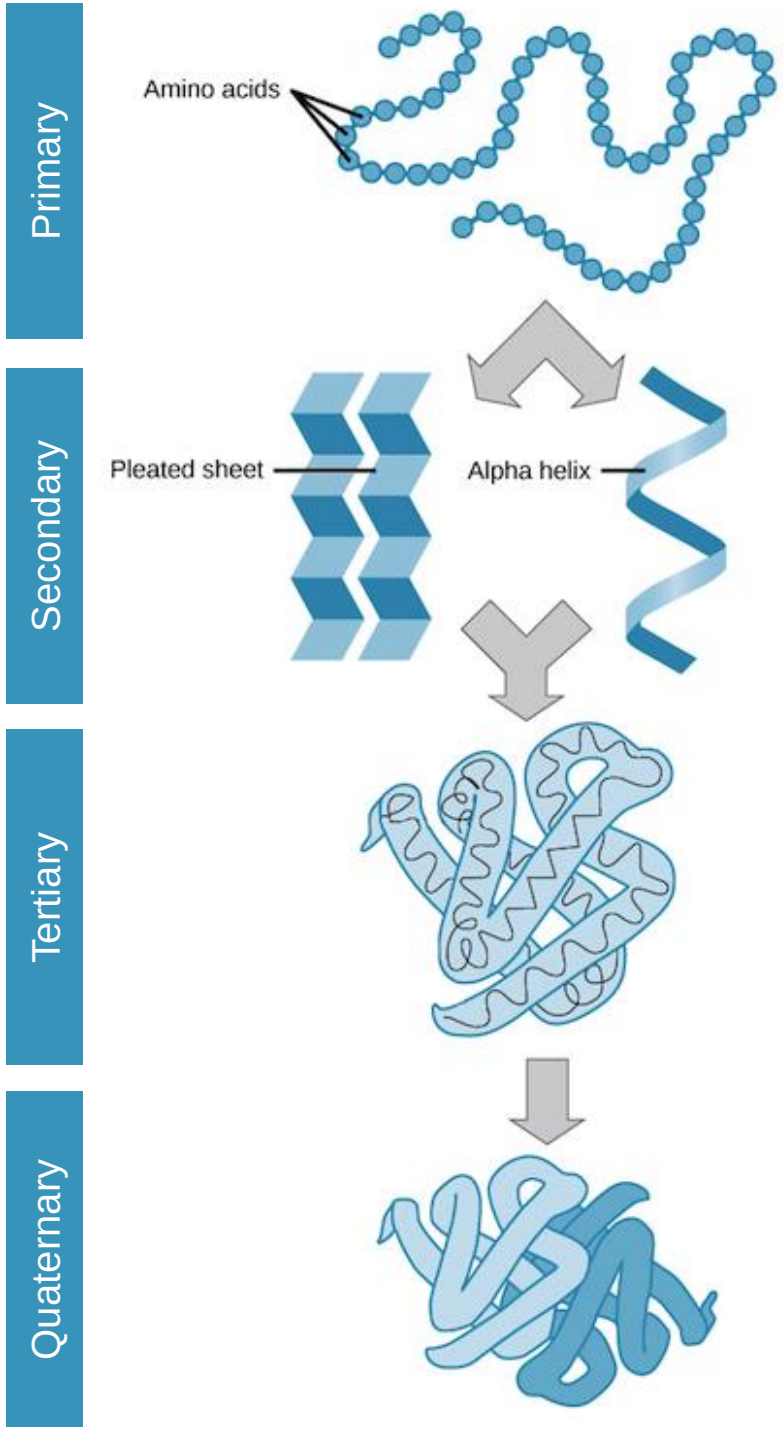
- **Primary-** sequence of nucleotides
- **Secondary-** base pairing interactions between polymers (DNA) or within a single polymer (RNA)
- **Tertiary-** 3D folding pattern
- **Quaternary-** interactions of nucleic acids with other molecules (DNA, RNA, or Protein)



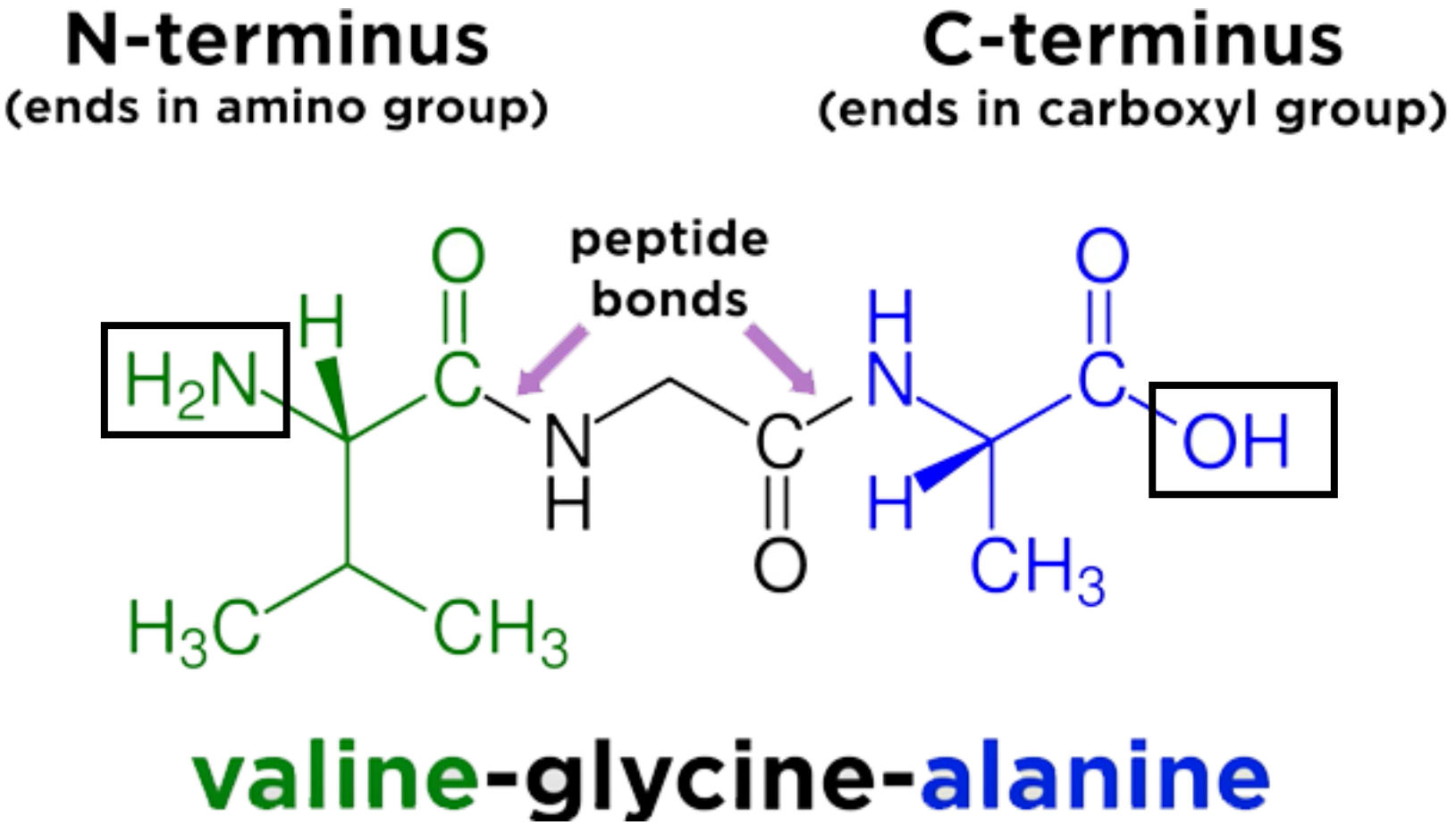
NIH National Library of Medicine
National Center for Biotechnology Information

Protein Structure

- **Primary-** sequence of amino acids
- **Secondary-** hydrogen bonding of the peptide backbone that causes amino acids to fold into a repeating pattern
- **Tertiary-** 3D folding pattern of a protein due to side chain interactions
- **Quaternary-** protein consisting of more than one polypeptide

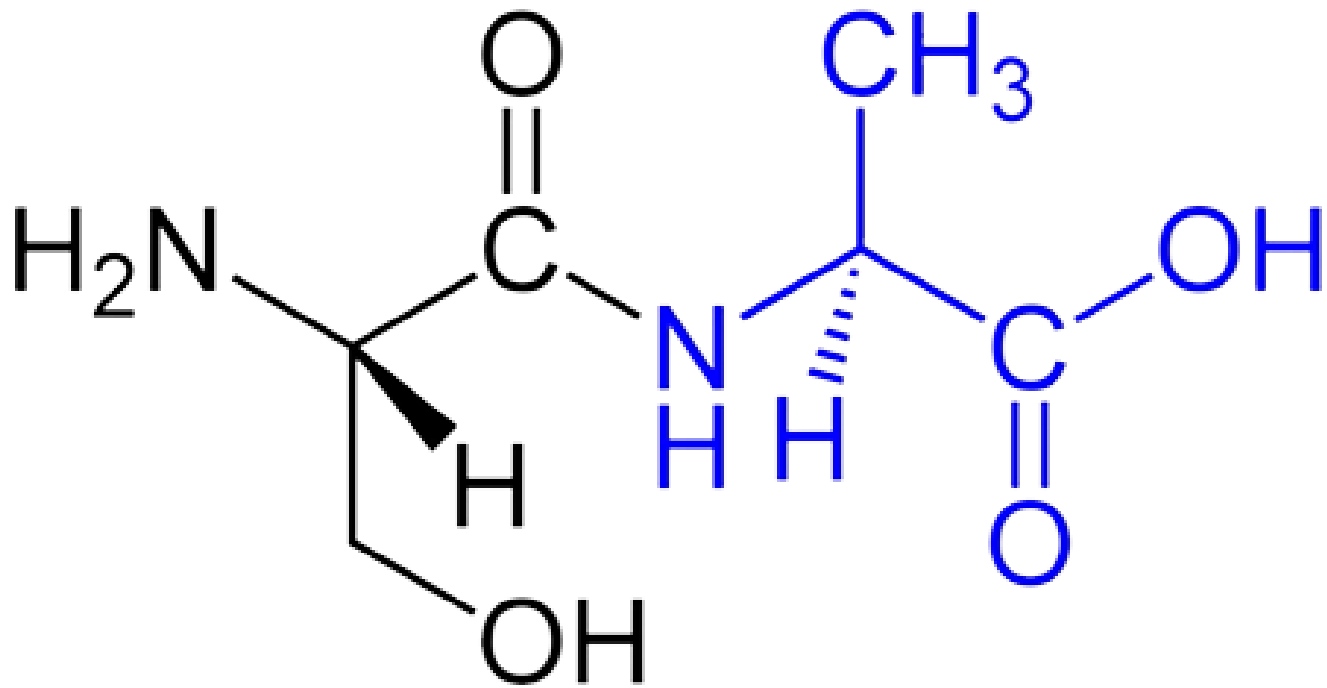


Terminology



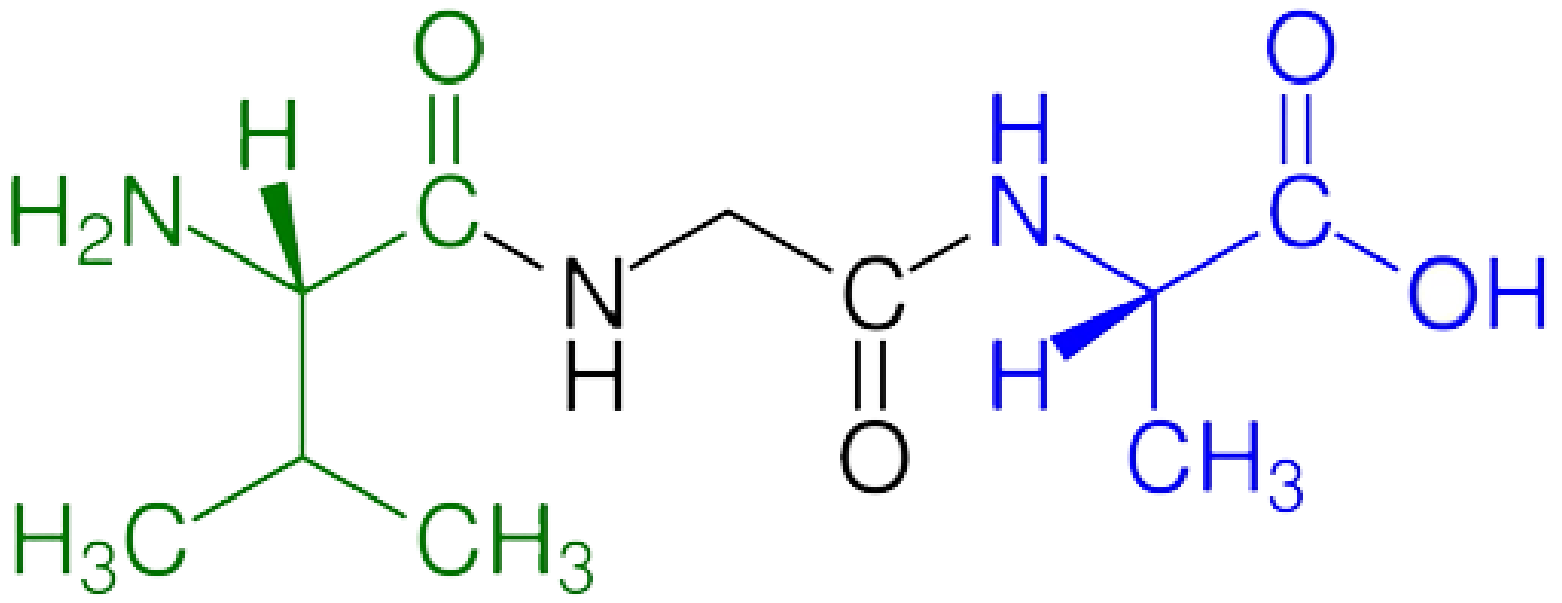
Terminology

dipeptide (2 amino acids)

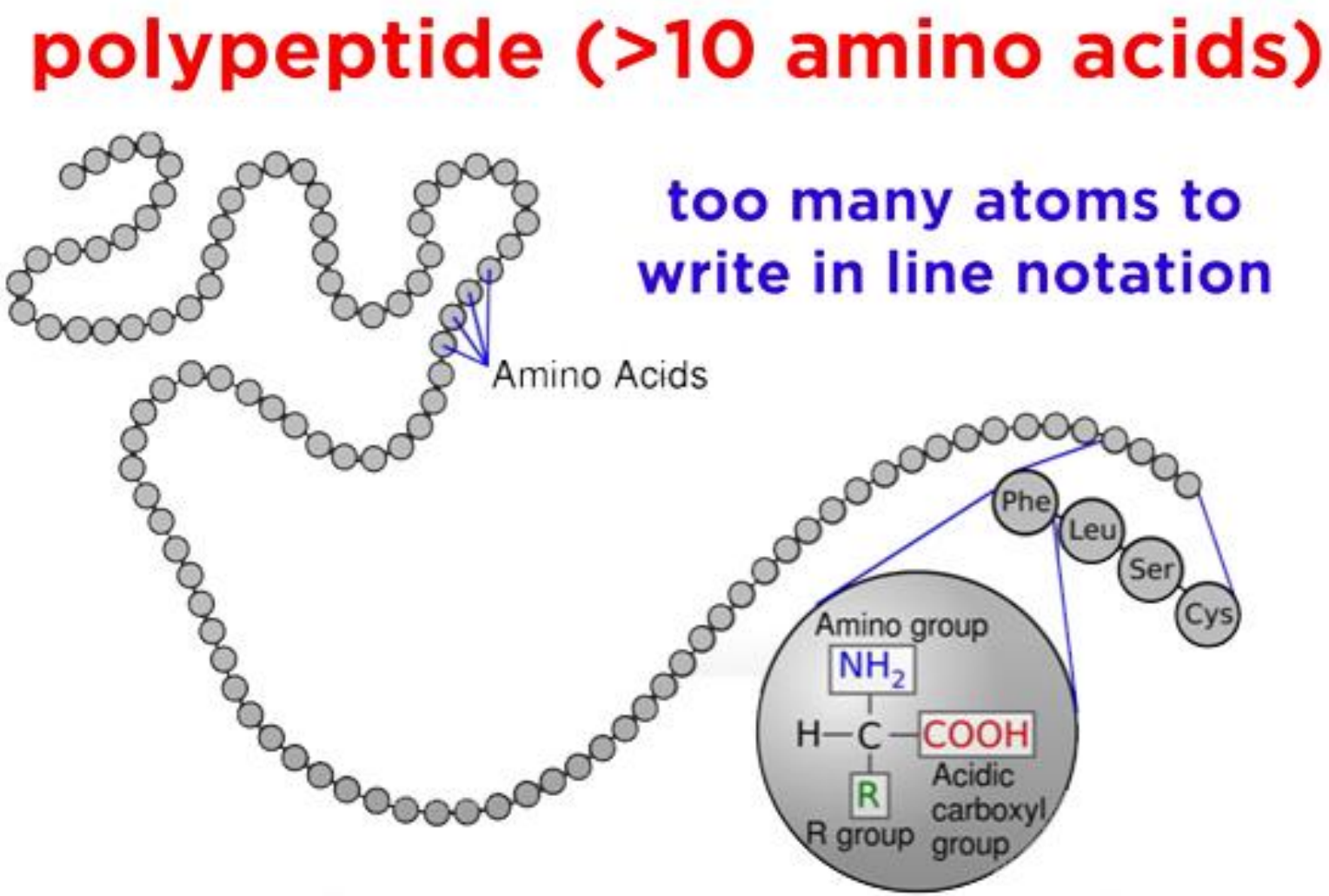


Terminology

oligopeptide (3-10 amino acids)

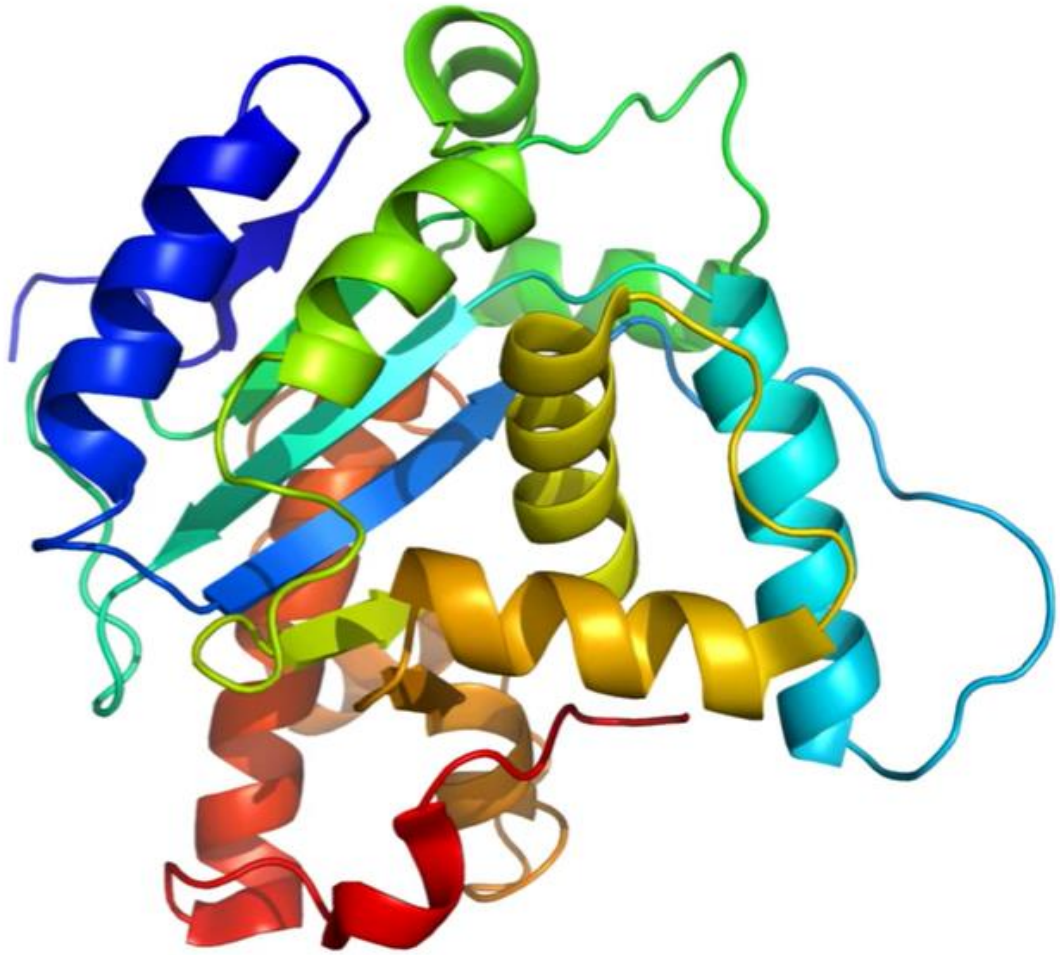


Terminology

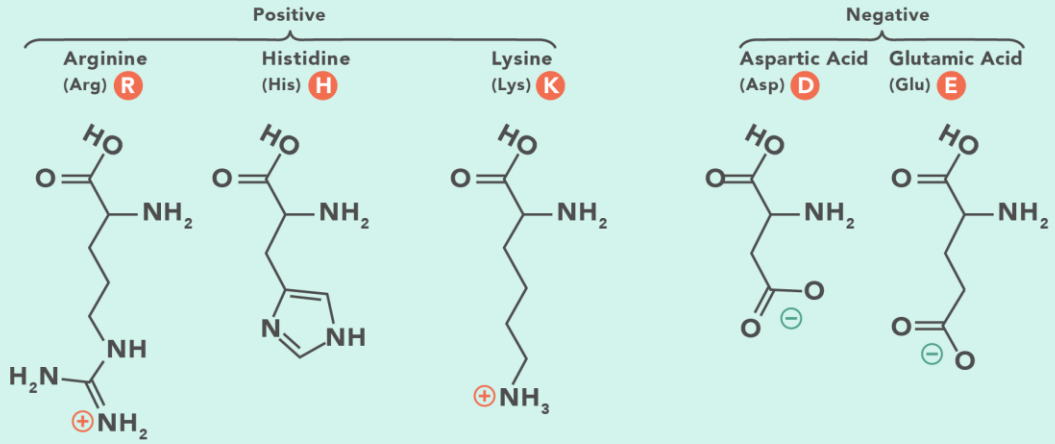


Terminology

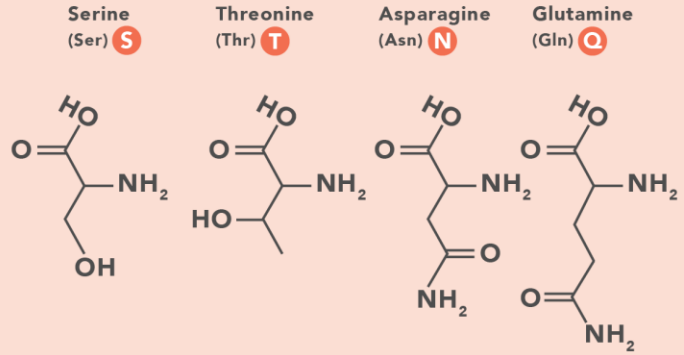
protein (generally 300-1000 amino acids)



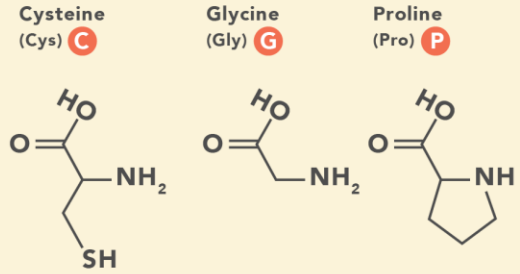
A. Amino Acids with Electrically Charged Side Chains



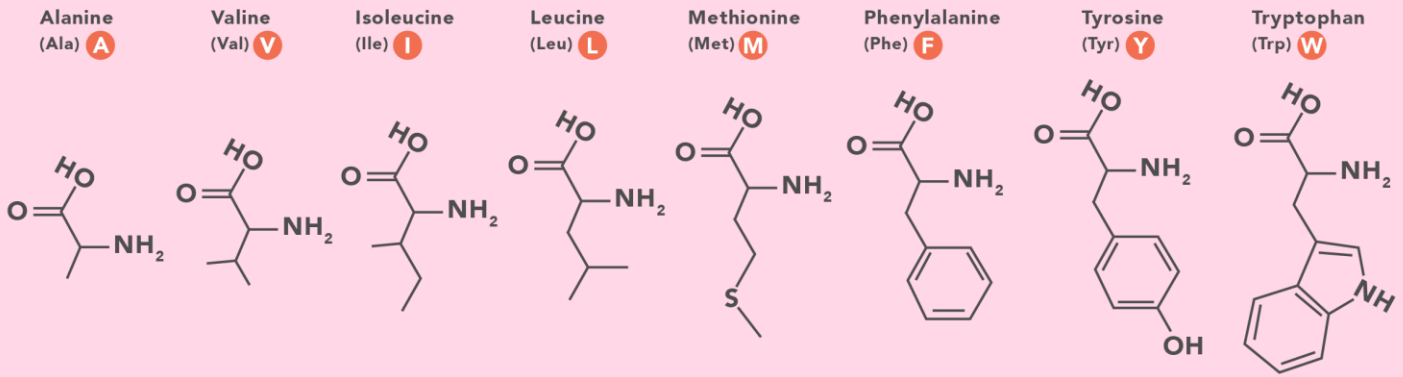
B. Amino Acids with Polar Uncharged Side Chains



C. Special Cases



D. Amino Acids with Hydrophobic Side Chains



Functional definition:

Enzymes: Accelerate biochemical reactions

Structural: Form biological structures

Transport: Carry biochemically important substances

Defense: Protect the body from foreign invaders

Structural definition:

Globular: Complex folds, irregularly shaped tertiary structures

Fibrous: Extended, simple folds -- generally structural proteins

Cellular localization definition:

Membrane: In direct physical contact with a membrane; generally water insoluble.

Soluble: Water soluble; can be anywhere in the cell

Experimental techniques

Single crystal X-ray diffraction (SC-XRD)

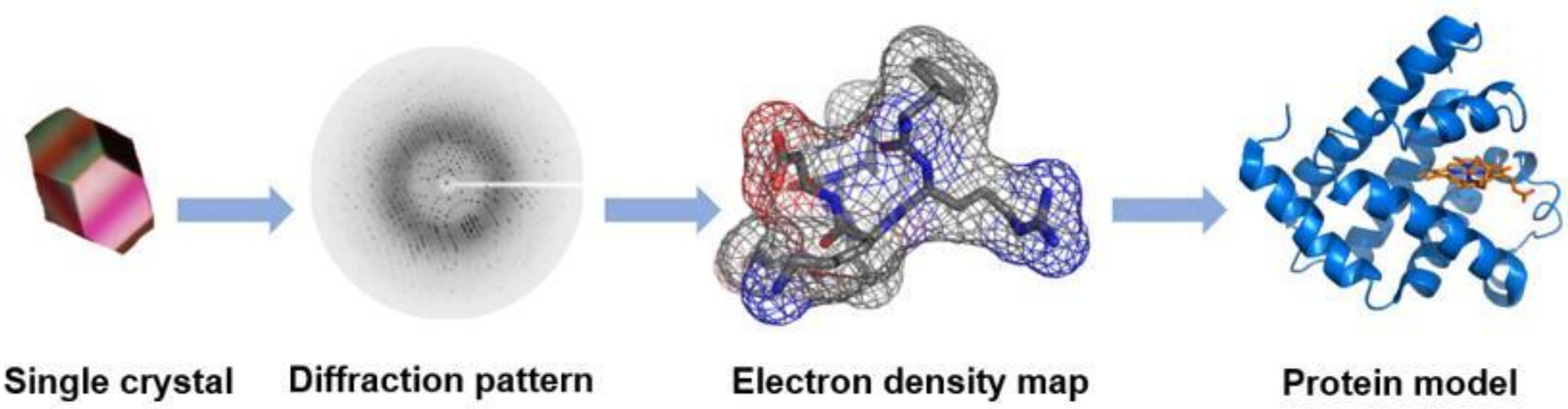
Nuclear magnetic resonance (NMR)

Cryo-electron microscopy (Cryo-EM)

Three main research techniques for structural biology.
According to the statistics of PDB (https://www.rcsb.org/)

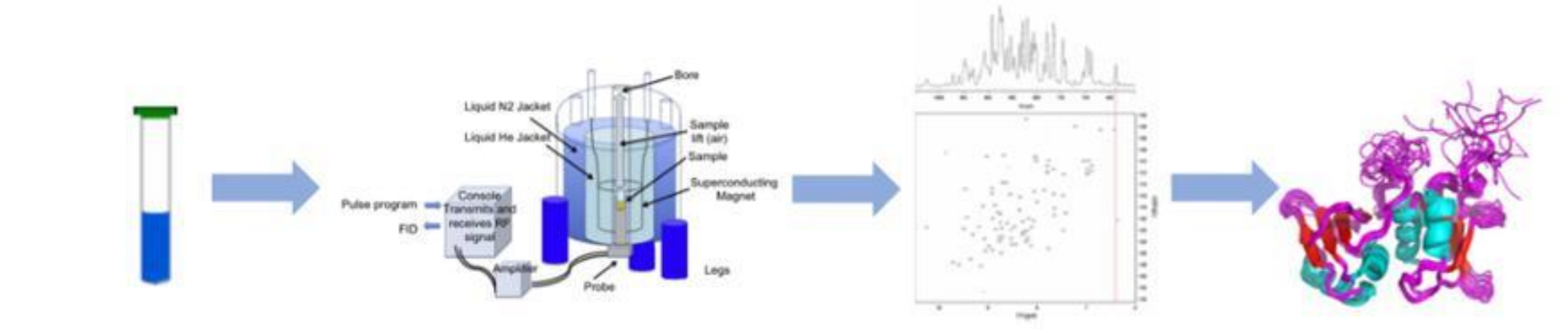
X-ray Crystallography

- Requires crystals, which can be hard to make
- Can handle very large proteins and complexes (e.g. ribosome)
- Provides a “flash picture” with little or no data about motions
- Can include packing artifacts from crystallization



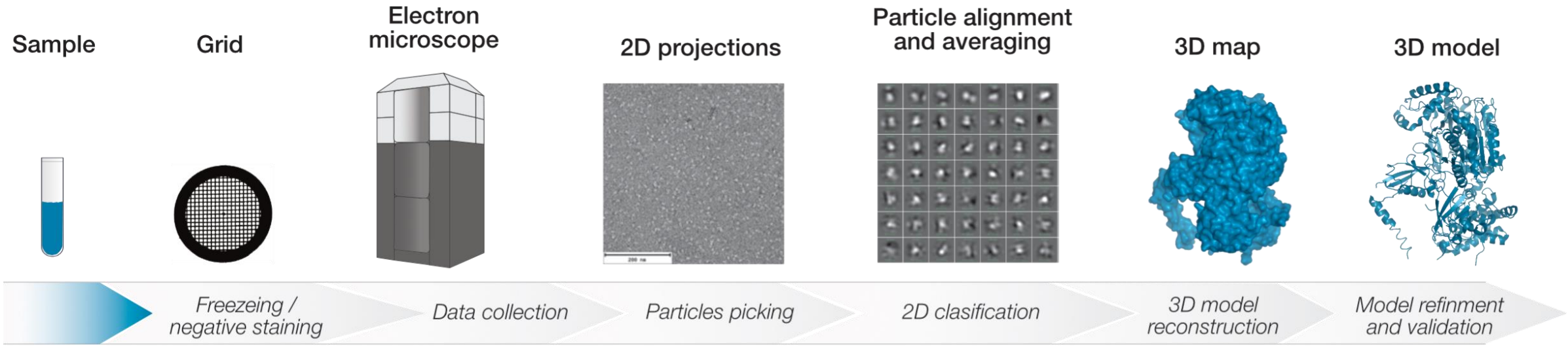
Nuclear Magnetic Resonance

- Requires highly concentrated, C13/N15-labeled protein solutions
- Limited to relatively small proteins (<30 kDa)
- Sensitive to molecular motions
- High protein concentrations may induce non-biological binding



Cryo-electron microscopy

- Requires expensive equipment
- Only small amount of sample
- Rapid freezing sample allows sample to maintain a closer-to-native state
- Useful for biomolecules with high molecular weight



NCBI Structure Database Search Tips

Entrez is a molecular biology database system that provides access to a wealth of NCBI data

- More Entrez Help is available on the NCBI website

Finding structures with Entrez

```
"term1"[field1] AND/OR/NOT "term2"[field2] AND/OR/NOT …
```

- Use field limits and Boolean operators
- Put phrases in quotes

NCBI Structure Database Search Examples

Useful Search Fields
Organism Ex. “Homo sapiens”[orgn]
Experimental Method Ex. “NMR”[exp]
Chemical Name “zinc”[chemical name]
PDB Description Ex. “Tumor Suppressor p53”[title]
[Filter] Ex. “Complex DNA”[filter]

More Search Field Options

```
term1[field1] AND/OR/NOT term2[field2] AND/OR/NOT …
```

```
“Homo sapiens”[orgn] AND “X-ray diffraction”[exp]
```

Search results

Items: 1 to 20 of 47803

```
“Homo sapiens”[orgn] AND “X-ray diffraction”[exp]AND “zinc”[chemical name]
```

Search results

Items: 1 to 20 of 6092

```
“Homo sapiens”[orgn] AND “X-ray diffraction”[exp] AND “zinc”[chemical name] AND “Complex DNA”[filter]
```

Search results

Items: 1 to 20 of 288

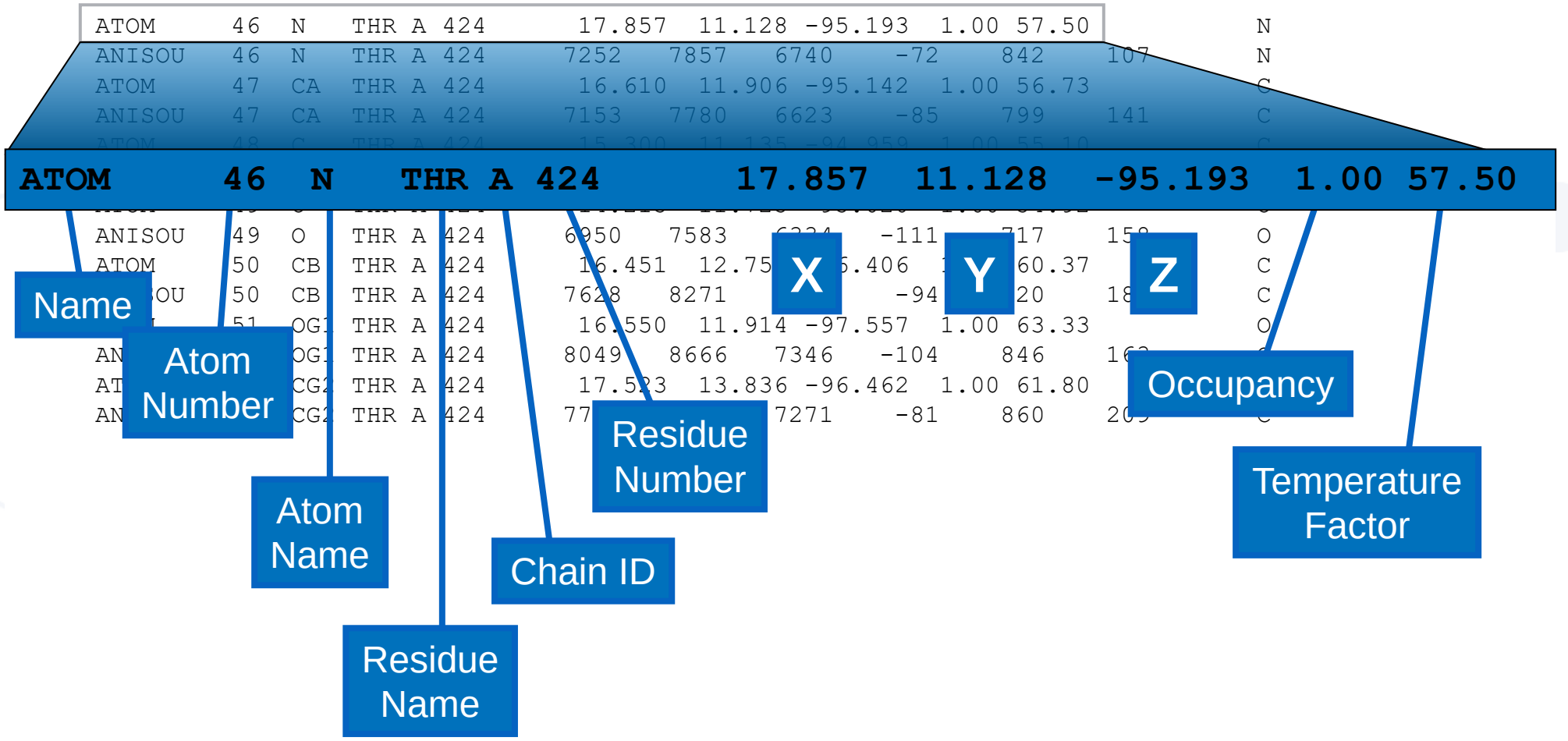
```
1TUP
```

PDB File

```
HEADER    ISOMERASE/DNA                           04-OCT-07   2RGR
TITLE     TOPOISOMERASE IIA BOUND TO G-SEGMENT DNA
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: DNA TOPOISOMERASE 2;
COMPND   3 CHAIN: A;
COMPND   4 FRAGMENT: DNA BINDING AND CLEAVAGE DOMAIN (RESIDUES 419-
COMPND   5 1177);
COMPND   6 SYNONYM: DNA TOPOISOMERASE II;
COMPND   7 EC: 5.99.1.3;
COMPND   8 ENGINEERED: YES;
COMPND   9 MOL_ID: 2;
COMPND  10 MOLECULE: DNA;
COMPND  11 CHAIN: C;
COMPND  12 ENGINEERED: YES;
COMPND  13 MOL_ID: 3;
COMPND  14 MOLECULE: DNA;
COMPND  15 CHAIN: D;
COMPND  16 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: SACCHAROMYCES CEREVISIAE;
SOURCE   3 ORGANISM_COMMON: BAKER'S YEAST;
SOURCE   4 ORGANI
SOURCE   5 GENE:
SOURCE   6 EXPRES
SOURCE   7 EXPRES
SOURCE   8 EXPRES
SOURCE   9 EXPRES
SOURCE  10 EXPRES
SOURCE  11 EXPRES
SOURCE  12 MOL_ID
SOURCE  13 SYNTHE
SOURCE  14 MOL_ID
SOURCE  15 SYNTHE
```

```
REMARK   2
REMARK   2 RESOLUTION.    3.00 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : PHENIX
…
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS   (%): 59.90
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 3.07
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: 12-20% PEG 1000, 100-250 MM MGCL2,
REMARK 280  100 MM SODIUM CACODYLATE, PH 7.0, VAPOR DIFFUSION, HANGING
REMARK 280  DROP, TEMPERATURE 277K
REMARK 290
```

PDB File: Data



More about PDB Data

Computational Structural Biology

- Structure Prediction- inference of 3D structure from sequence data
- Molecular Docking- predicts the orientation of one molecule to another
- Molecular Dynamics Simulations- analyzes physical movements of atoms and molecules over time

NCBI

Computational Structural Biology

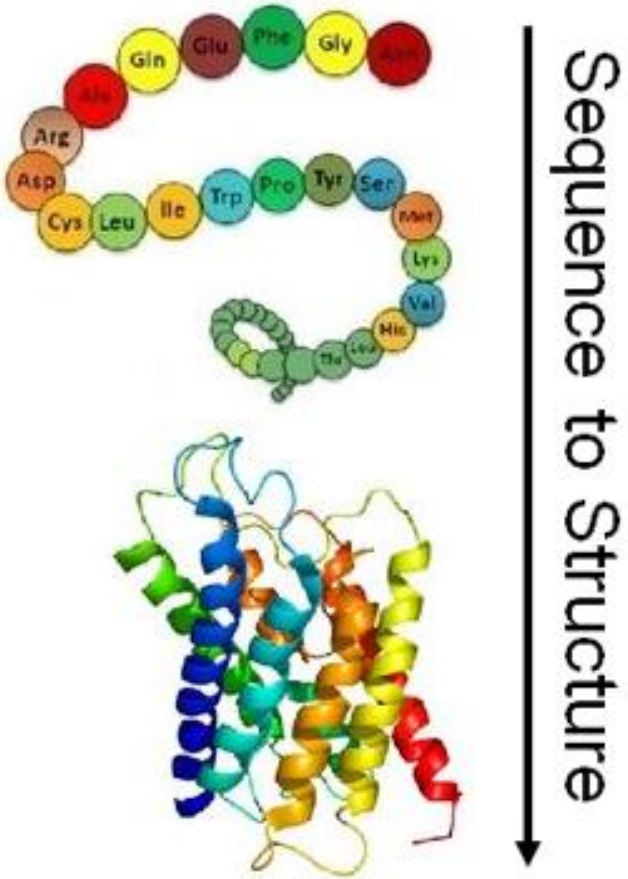
- Structure Prediction- inference of 3D structure from sequence data
- Molecular Docking- predicts the orientation of one molecule to another
- Molecular Dynamics Simulations- analyzes physical movements of atoms and molecules over time

- Rely on experimental information from public databases
  - NCBI Databases and RCSB Protein Data Bank

NCBI

Structure Prediction Methods

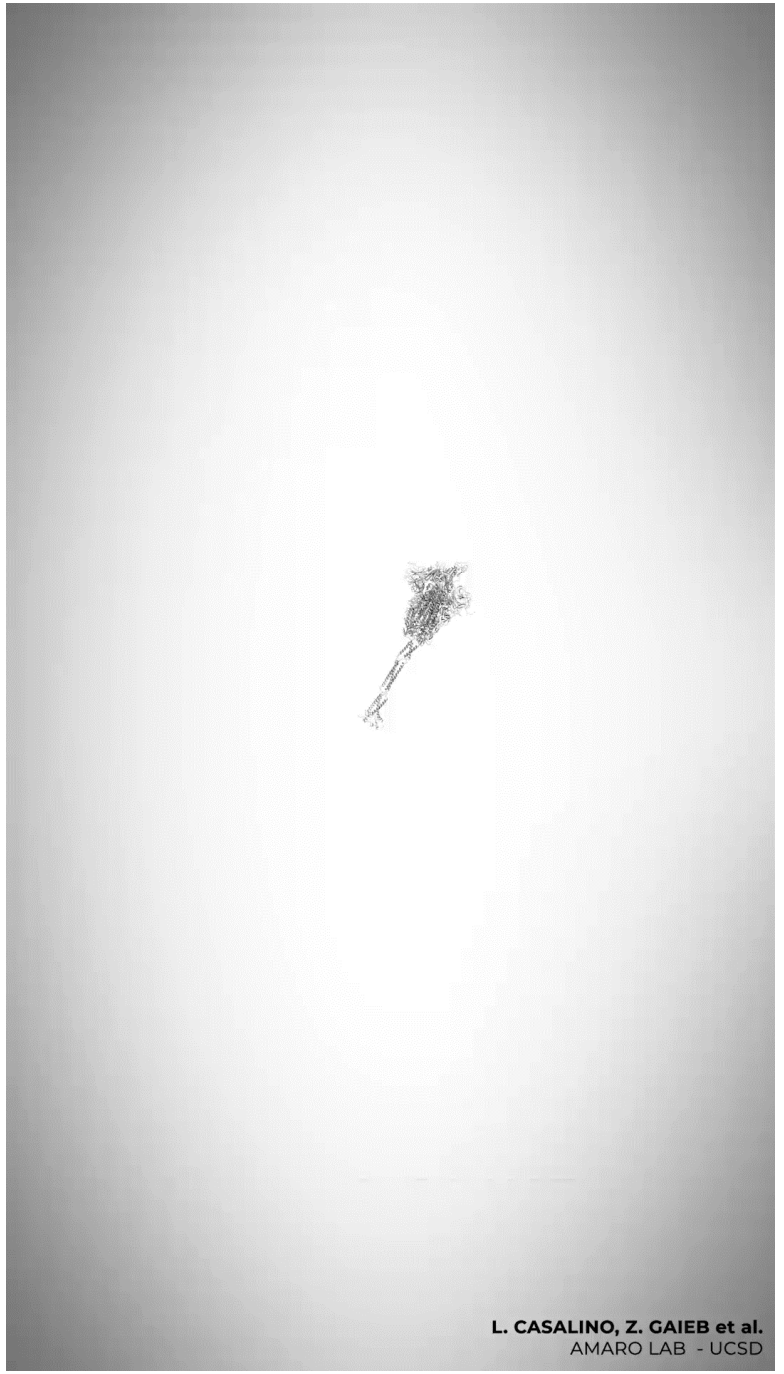
- Comparative Modeling
  - Prediction is based on amino acid sequence and structures of similar molecules available
- Fold recognition
  - Predicts folded structure by aligning a protein of unknown structure and a protein of known structure for low levels of sequence identity (<25%)
- Ab initio
  - Predicts the structure of proteins from the sequence and using molecular energy calculations (Schrodinger equation)



Structure Prediction Example

Impact on COVID-19 research

- Researchers have provided key insights into the SARS-CoV-2 proteins through structure prediction
  - Identified critical residues
  - Contextualized variant perturbations
  - Improved understanding of molecular recognition
- Spike fusion glycoprotein example
  - Challenging to characterize experimentally
  - Modeling + molecular dynamics helped researchers understand the roles of glycans on the dynamics of the protein



Casalino et al, *Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein*, PMID:33140034

Homology Modeling *vs Ab initio* Prediction

Ab initio Prediction	**Comparative Modeling**
Applicable to any sequence	Applicable to only those sequences with recognizable similarity to a template structure
Not very accurate (>4Å RMSD)	Fairly accurate (<3Å RMSD), similar to low resolution X-ray structure
Attempted for proteins of <100 residues	Not limited by size
Accuracy and applicability are limited by our understanding of the protein folding problem	Accuracy and applicability are limited by the number of known folds

Quick Background to AlphaFold

Learn more about AlphaFold here