



Exploring 3D Molecular Structures with iCn3D

Alexa M. Salsbury, Ph.D.

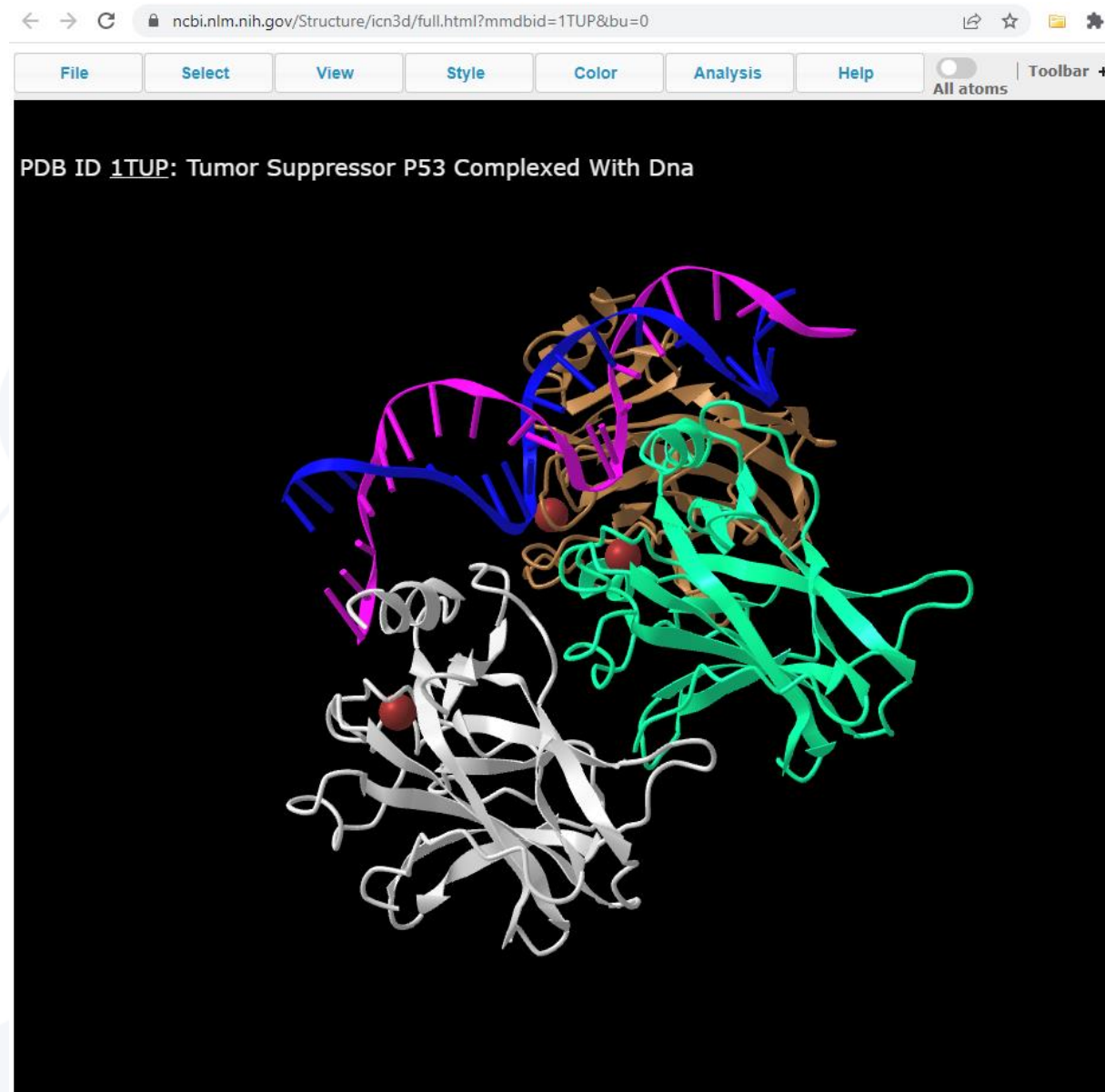


National Library of Medicine
National Center for Biotechnology Information



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.

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

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

Experimental techniques

	Advantages	Disadvantages
X-ray crystallography	<ul style="list-style-type: none">• Well developed• High resolution• Broad molecular weight range	<ul style="list-style-type: none">• Difficult sample prep• Static crystalline state
NMR	<ul style="list-style-type: none">• High resolution• 3D structure in solution• Good for dynamic study	<ul style="list-style-type: none">• Difficult sample prep• High sample purity needed• Static crystalline state captured
Cryo-EM	<ul style="list-style-type: none">• Simple sample prep• Structure in native state• Small sample size needed	<ul style="list-style-type: none">• 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?

The screenshot shows the RCSB PDB website homepage. The header includes navigation menus for Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. A search bar is prominently displayed with a dropdown menu for 'PDB Archive'. Below the search bar, there are logos for PDB-101, PDB, EMDatabank, Nucleic Acid Database, and Worldwide Protein Data Bank Foundation. A banner for 'Developers: Join the RCSB PDB Team' is visible. The main content area features a 'Welcome' section, a 'March Molecule of the Month' section with a 3D protein structure of Vascular Endothelial Growth Factor (VegF) and Angiogenesis, and a 'COVID-19 CORONAVIRUS Resources' section.

RCSB Protein Data Bank

The screenshot shows the NCBI Structure Database website. The header includes navigation menus for Resources and How To, and a 'Sign in to NCBI' button. A search bar is present with a dropdown menu for 'Structure' and an 'Advanced' search option. Below the search bar, there is a 3D protein structure visualization and a text box explaining that three-dimensional structures provide information on biological function and evolutionary history. A table of links is provided below:

Using Structure	Structure Tools	More Resources
Search	Macromolecular Resources Overview	PDB
How to (Quick Start) Guides	iCn3D (web-based 3D viewer)	Protein
Help	Cn3D (3D viewer application)	CDD
News	IBIS	PubChem
FTP	VAST	NCBI Structure Group Resources & Research
Publications	VAST+	
Discover		

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.

6LU7

The crystal structure of COVID-19 main protease in complex with an inhibitor N3

DOI: [10.2210/pdb6LU7/pdb](https://doi.org/10.2210/pdb6LU7/pdb)

Classification: VIRAL PROTEIN

Organism(s): Severe acute respiratory syndrome coronavirus 2, synthetic construct

Expression System: Escherichia coli BL21(DE3)

Mutation(s): No

Deposited: 2020-01-26 **Released:** 2020-02-05

Deposition Author(s): Liu, X., Zhang, B., Jin, Z., Yang, H., Rao, Z.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.16 Å

R-Value Free: 0.235

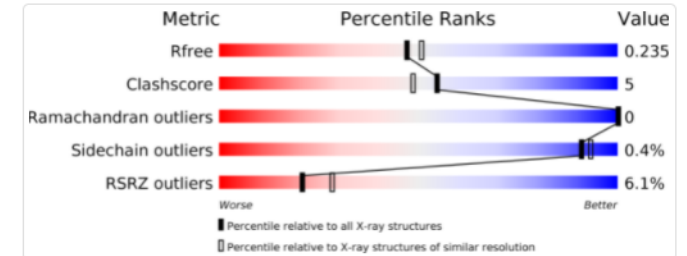
R-Value Work: 0.202

R-Value Observed: 0.204

wwPDB Validation

[3D Report](#)

[Full Report](#)



Literature

[Download Primary Citation](#)

Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors.

[Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., Yang, H.](#)

(2020) Nature **582**: 289-293

PubMed: [32272481](#) [Search on PubMed](#)

DOI: [10.1038/s41586-020-2223-y](https://doi.org/10.1038/s41586-020-2223-y)

Primary Citation of Related Structures:

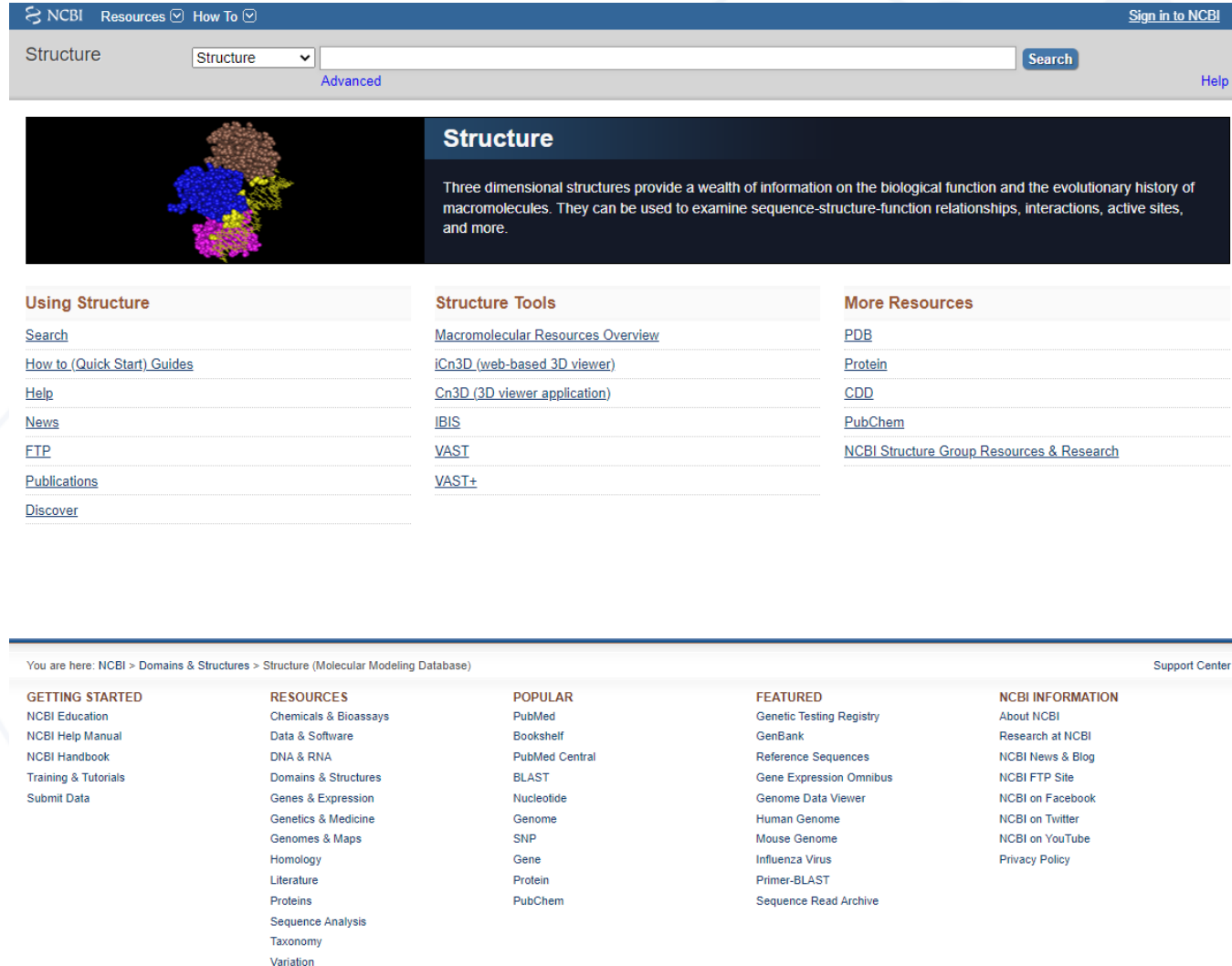
[7BQY](#), [6LU7](#)

PubMed Abstract:

A new coronavirus, known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the aetiological agent responsible for the 2019-2020 viral pneumonia outbreak of coronavirus disease 2019 (COVID-19)¹⁻⁴. Currently, there are no targeted therapeutic agents for the treatment of this disease, and effective treatment options remain very limited ...

NCBI's Structure Database

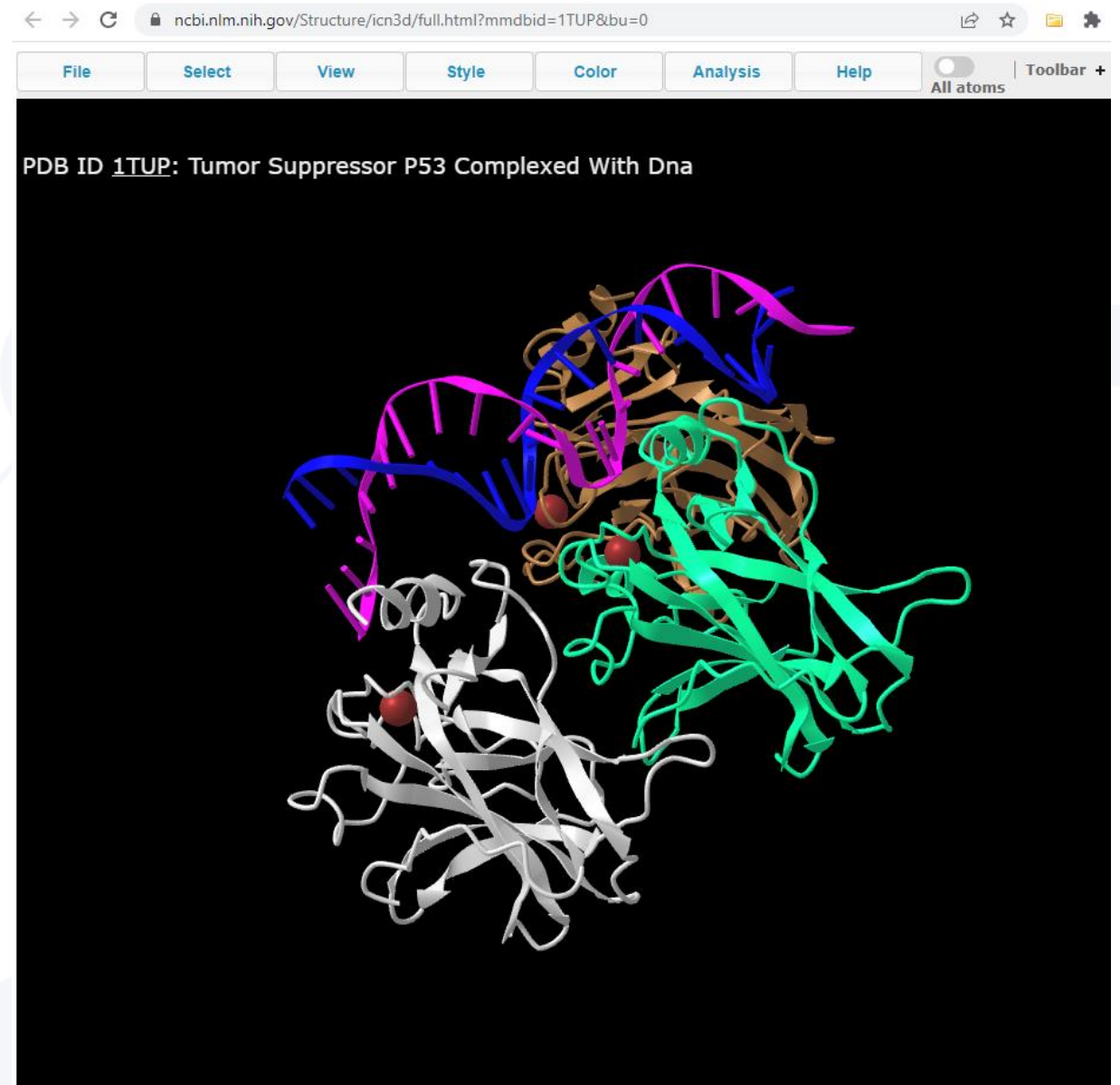
- Updated monthly
- Derived from PDB records
- Additional information added, including:
 - Explicit chemical graph information
 - Validation (secondary structure elements)
 - Includes taxonomy
- Connects 3D to associated literature, molecular data, chemical data, and other NCBI tools



The screenshot shows the NCBI Structure Database homepage. At the top, there is a navigation bar with 'NCBI', 'Resources', and 'How To' menus, and a 'Sign in to NCBI' link. Below this is a search bar with a dropdown menu set to 'Structure' and a 'Search' button. A 'Help' link is also present. The main content area features a large image of a 3D molecular structure on the left and a text box on the right stating: 'Structure Three dimensional structures provide a wealth of information on the biological function and the evolutionary history of macromolecules. They can be used to examine sequence-structure-function relationships, interactions, active sites, and more.' Below this, there are three columns of links: 'Using Structure' (Search, How to (Quick Start) Guides, Help, News, FTP, Publications, Discover), 'Structure Tools' (Macromolecular Resources Overview, iCn3D (web-based 3D viewer), Cn3D (3D viewer application), IBIS, VAST, VAST+), and 'More Resources' (PDB, Protein, CDD, PubChem, NCBI Structure Group Resources & Research). At the bottom, there is a breadcrumb trail: 'You are here: NCBI > Domains & Structures > Structure (Molecular Modeling Database)' and a 'Support Center' link. The footer contains five columns of links: 'GETTING STARTED' (NCBI Education, NCBI Help Manual, NCBI Handbook, Training & Tutorials, Submit Data), 'RESOURCES' (Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Variation), 'POPULAR' (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem), 'FEATURED' (Genetic Testing Registry, GenBank, Reference Sequences, Gene Expression Omnibus, Genome Data Viewer, Human Genome, Mouse Genome, Influenza Virus, Primer-BLAST, Sequence Read Archive), and 'NCBI INFORMATION' (About NCBI, Research at NCBI, NCBI News & Blog, NCBI FTP Site, NCBI on Facebook, NCBI on Twitter, NCBI on YouTube, Privacy Policy).

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](#):

The screenshot displays the iCn3D web interface. At the top, the NIH logo and the text "U.S. National Library of Medicine" and "NCBI National Center for Biotechnology Information" are visible. The main heading is "iCn3D" followed by "AlphaFold-related gallery with live examples". A "Menu" dropdown is open, listing options: "About iCn3D", "Live Gallery", "Tutorial >", "Search Structure", "Citing iCn3D", "Source Code >", "Develop >", and "Help Doc".

Two protein structure visualizations are shown side-by-side. The left one is for UniProt ID A0A044R7Z7, labeled "ALPHAFOLD MONOMER V2". It features a blue ribbon structure with a red helix and green arrows. The right one is for UniProt ID Q08426, labeled "ALPHAFOLD MONOMER V1". It features a blue ribbon structure with yellow and orange highlights. A legend for the right structure indicates: "Very high (pLDDT > 90)", "Confident (90 > pLDDT > 70)", "Low (70 > pLDDT > 50)", and "Very low (pLDDT < 50)".

Below each structure is a "Sequences and Annotations" panel. The left panel shows annotations for A0A044R7Z7, including "Conserved Domains" and "3D Domains". The right panel shows annotations for Q08426, including "Conserved Domains", "3D Domains", "Disulfide Bonds", and "Cross-Linkages".

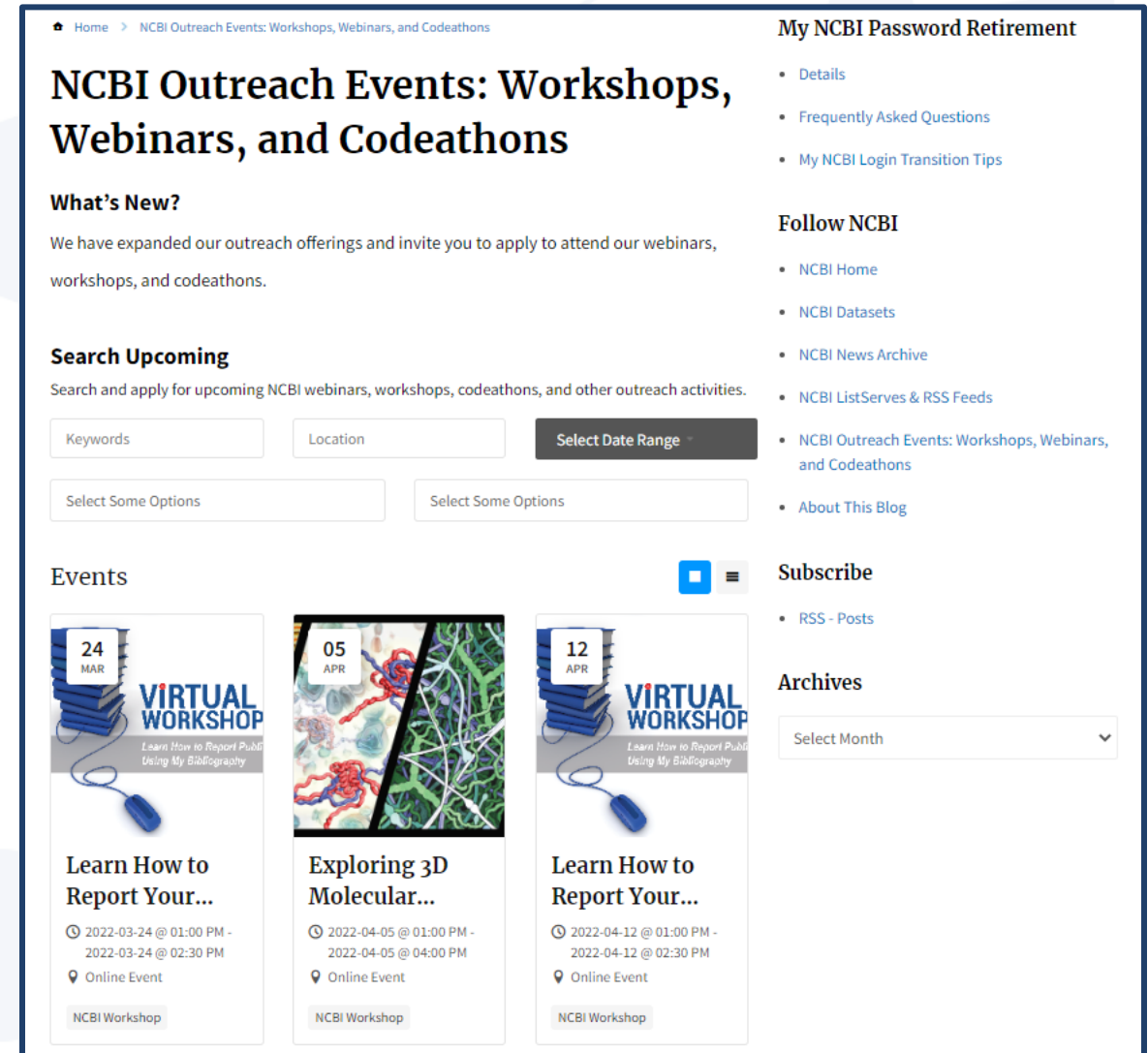
Below the left structure is the caption: "AlphaFold structures with conserved domain and 3D domain annotations (Uniprot ID A0A044R7Z7)". Below the right structure is the caption: "AlphaFold structures with SNP and ClinVar annotations (Uniprot ID Q08426)".

Continue learning about NCBI Resources

- Join us for workshops, webinars, or codeathons!

[NCBI Insights Blog](#)

- Follow us on social media:



The screenshot shows the NCBI Outreach Events page. At the top, there is a breadcrumb trail: Home > NCBI Outreach Events: Workshops, Webinars, and Codeathons. The main heading is "NCBI Outreach Events: Workshops, Webinars, and Codeathons". Below this, there is a "What's New?" section with a paragraph: "We have expanded our outreach offerings and invite you to apply to attend our webinars, workshops, and codeathons." There is also a "Search Upcoming" section with a search bar containing "Keywords", "Location", and "Select Date Range" buttons. Below the search bar are two "Select Some Options" dropdown menus. The "Events" section features three event cards. The first card is for a "VIRTUAL WORKSHOP" on 24 MAR, titled "Learn How to Report Your..." with a time slot of 2022-03-24 @ 01:00 PM - 2022-03-24 @ 02:30 PM. The second card is for a "VIRTUAL WORKSHOP" on 05 APR, titled "Exploring 3D Molecular..." with a time slot of 2022-04-05 @ 01:00 PM - 2022-04-05 @ 04:00 PM. The third card is for a "VIRTUAL WORKSHOP" on 12 APR, titled "Learn How to Report Your..." with a time slot of 2022-04-12 @ 01:00 PM - 2022-04-12 @ 02:30 PM. On the right side of the page, there are sections for "My NCBI Password Retirement" (with links for Details, Frequently Asked Questions, and My NCBI Login Transition Tips), "Follow NCBI" (with links for NCBI Home, NCBI Datasets, NCBI News Archive, NCBI ListServes & RSS Feeds, NCBI Outreach Events: Workshops, Webinars, and Codeathons, and About This Blog), "Subscribe" (with a link for RSS - Posts), and "Archives" (with a "Select Month" dropdown menu).

Questions & Discussion



Exploring 3D Molecular Structures with iCn3D Supplemental Learning Materials

Alexa M. Salsbury, Ph.D.



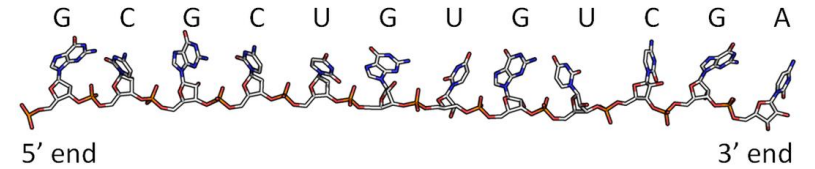
National Library of Medicine
National Center for Biotechnology Information



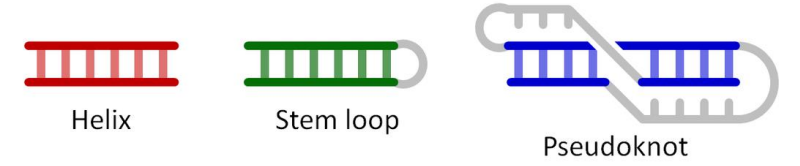
Nucleic Acid Structure

- **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)

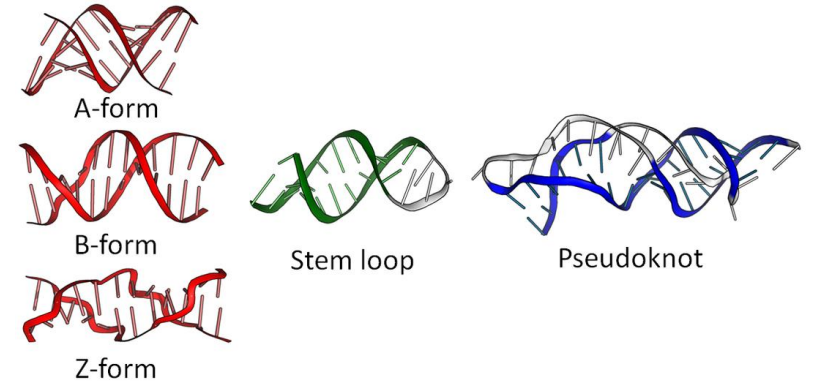
Primary



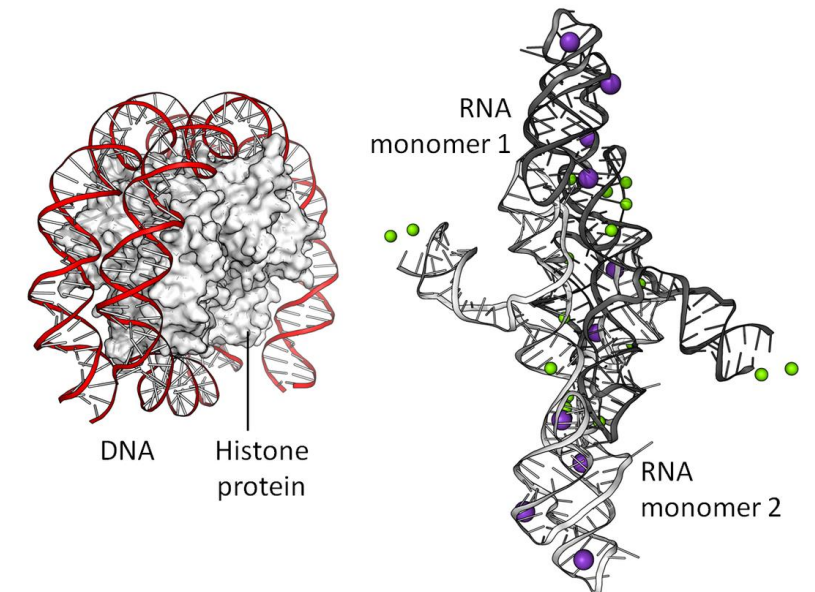
Secondary



Tertiary



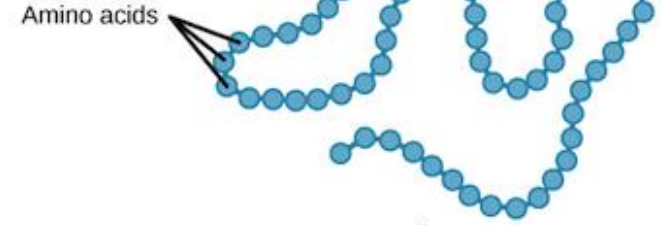
Quaternary



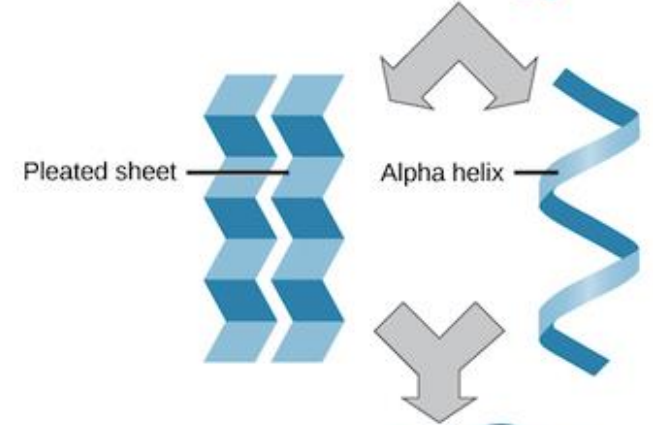
Protein Structure

- **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

Primary



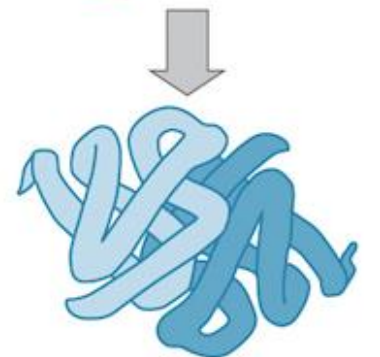
Secondary



Tertiary



Quaternary



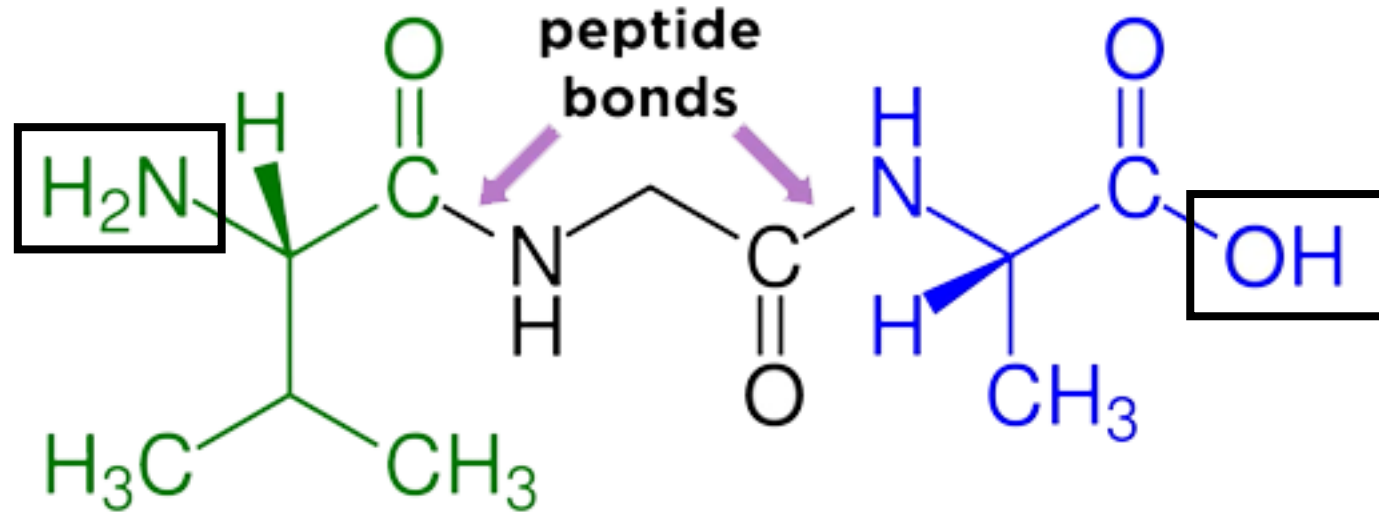
Terminology

N-terminus

(ends in amino group)

C-terminus

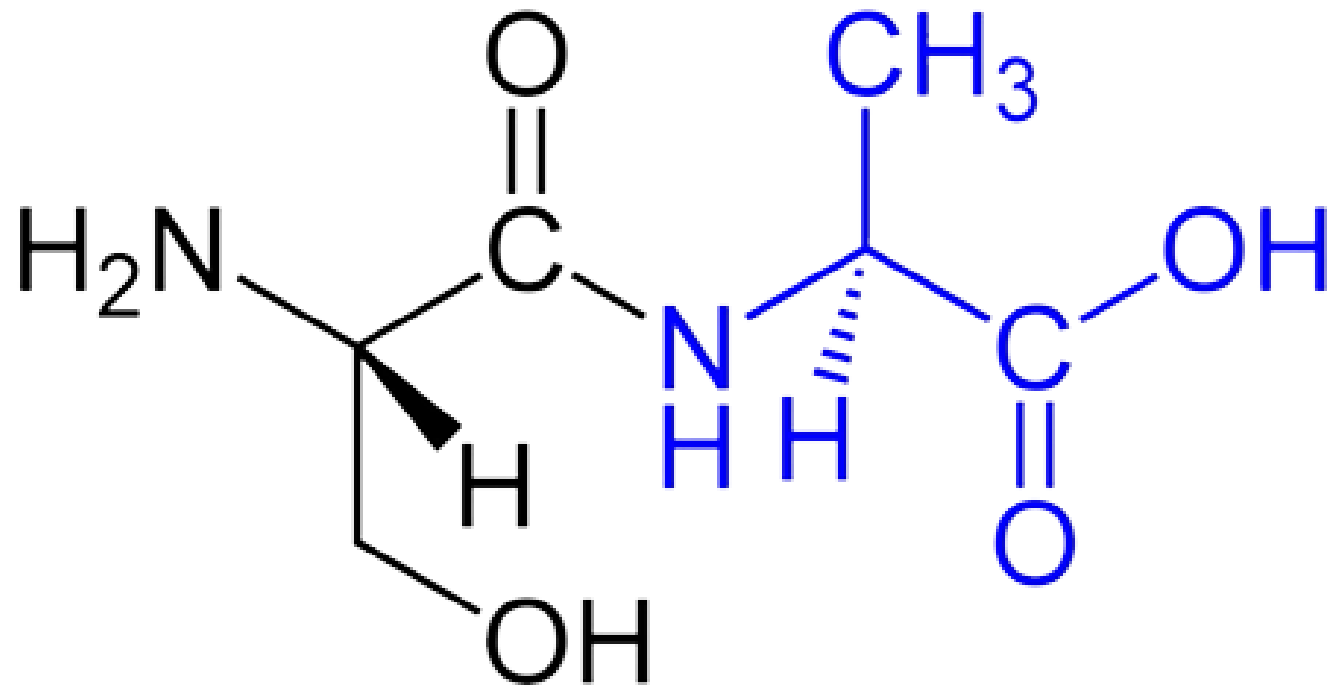
(ends in carboxyl group)



valine-glycine-alanine

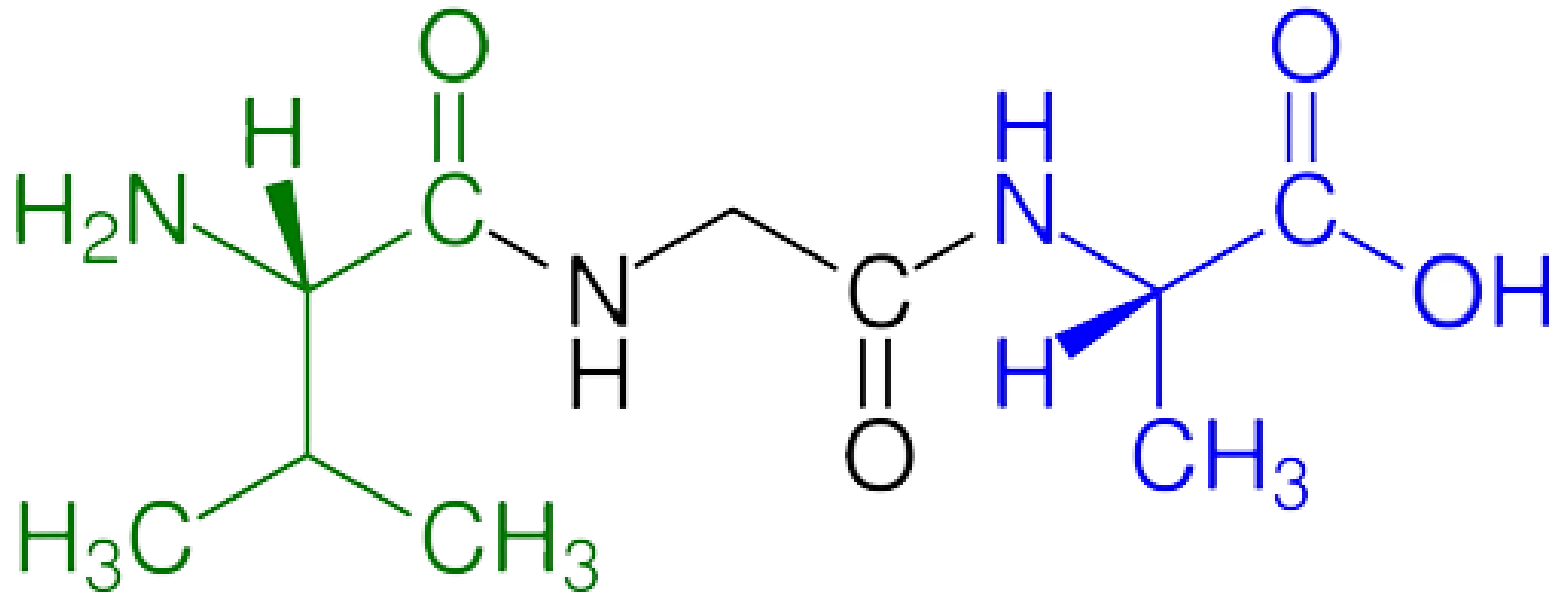
Terminology

dipeptide (2 amino acids)



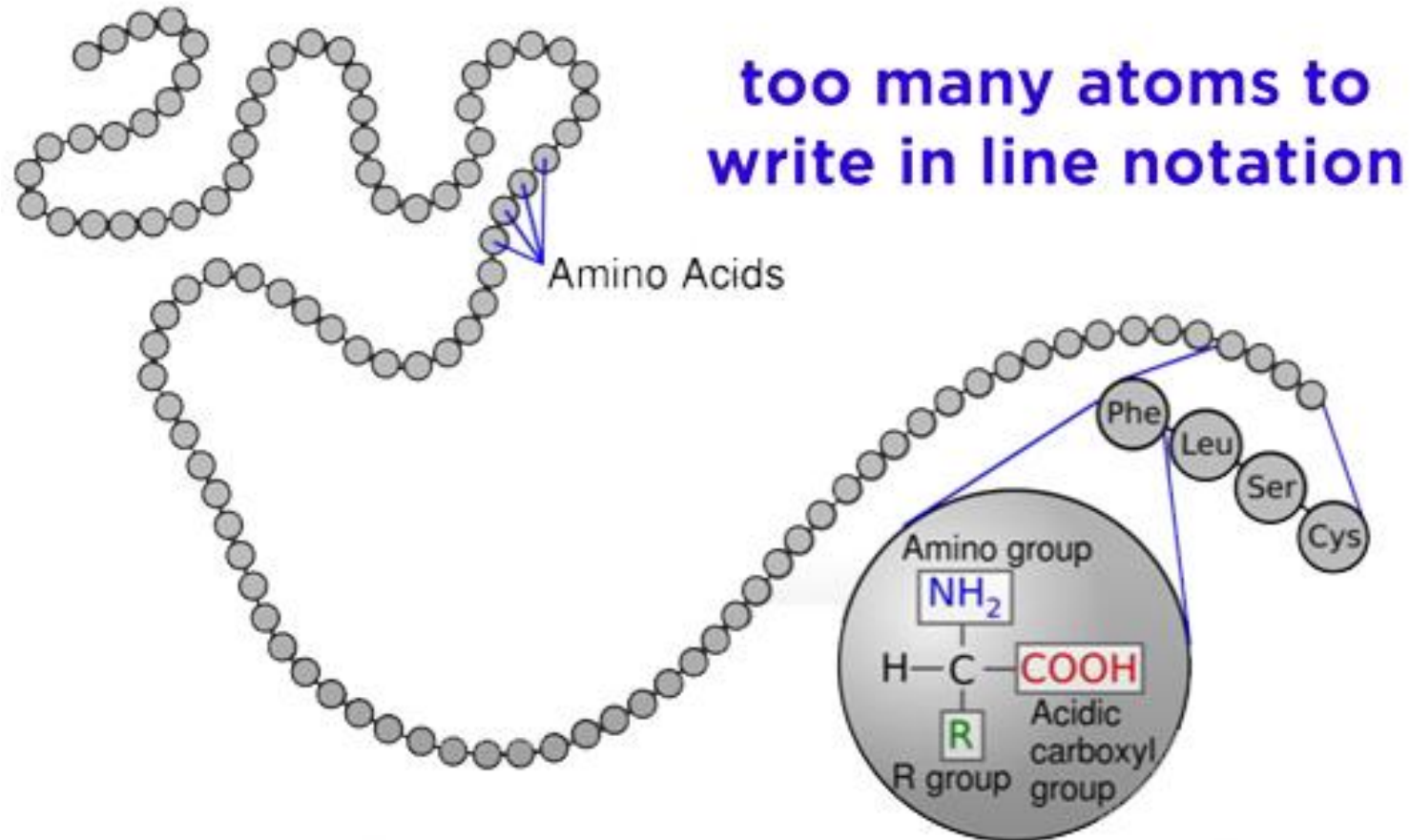
Terminology

oligopeptide (3-10 amino acids)



Terminology

polypeptide (>10 amino acids)

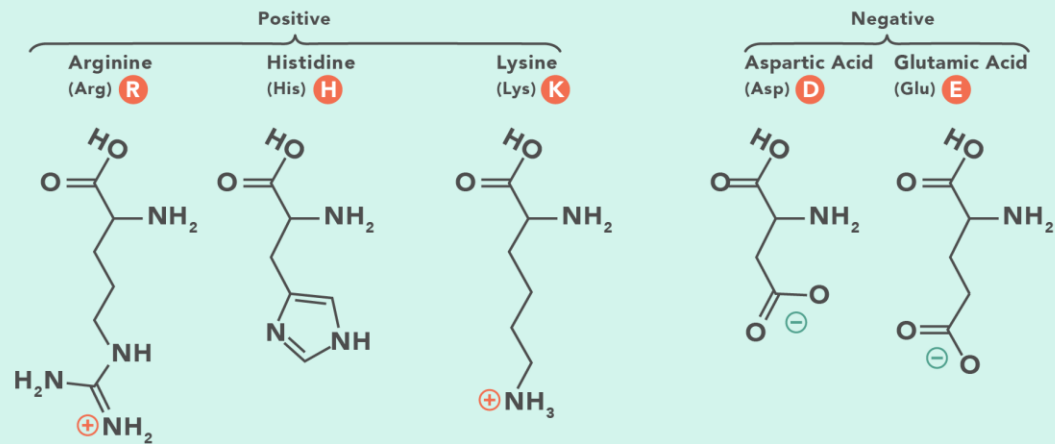


Terminology

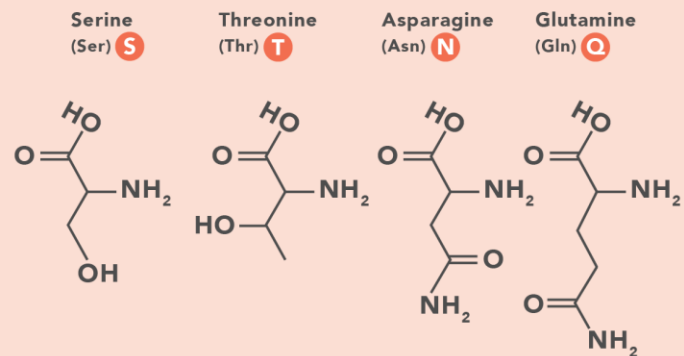
protein (generally 300-1000 amino acids)



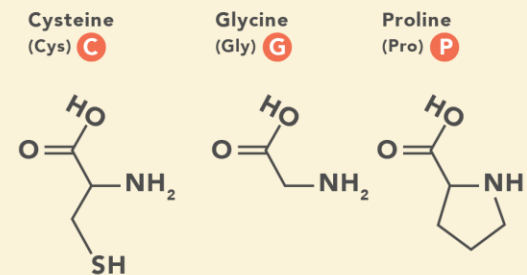
A. Amino Acids with Electrically Charged Side Chains



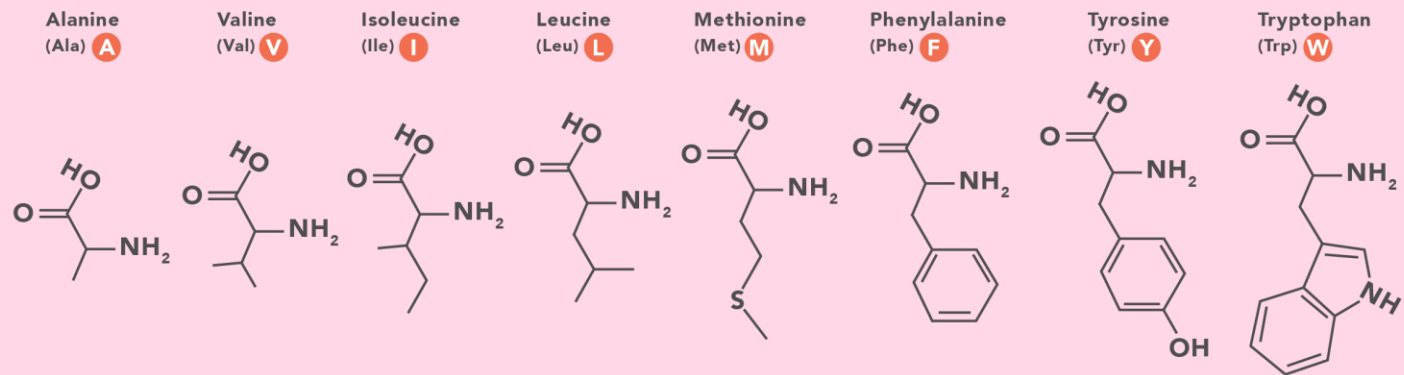
B. Amino Acids with Polar Uncharged Side Chains



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)

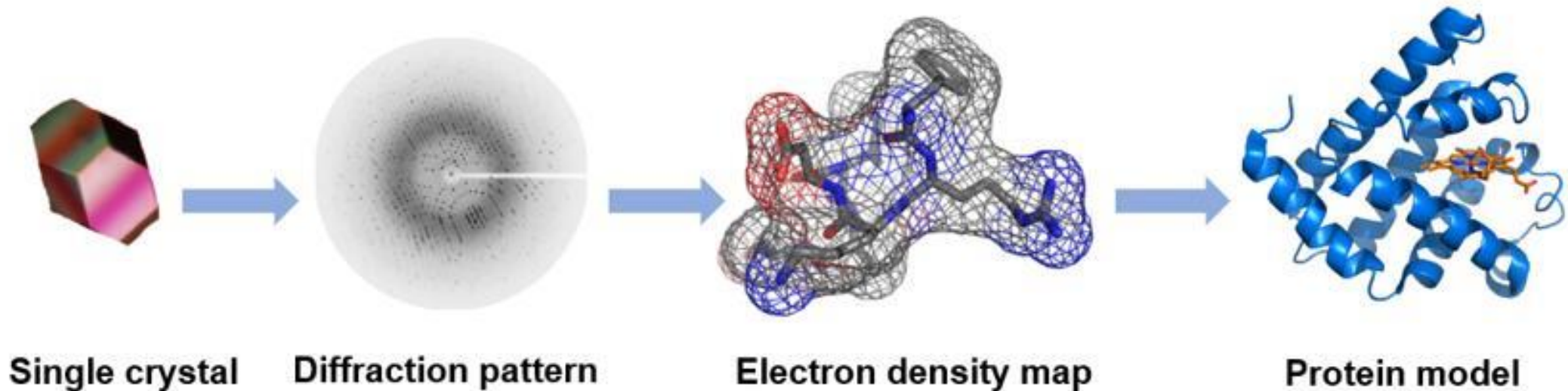
■ Nuclear magnetic resonance (NMR)

■ Cryo-electron microscopy (Cryo-EM)

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

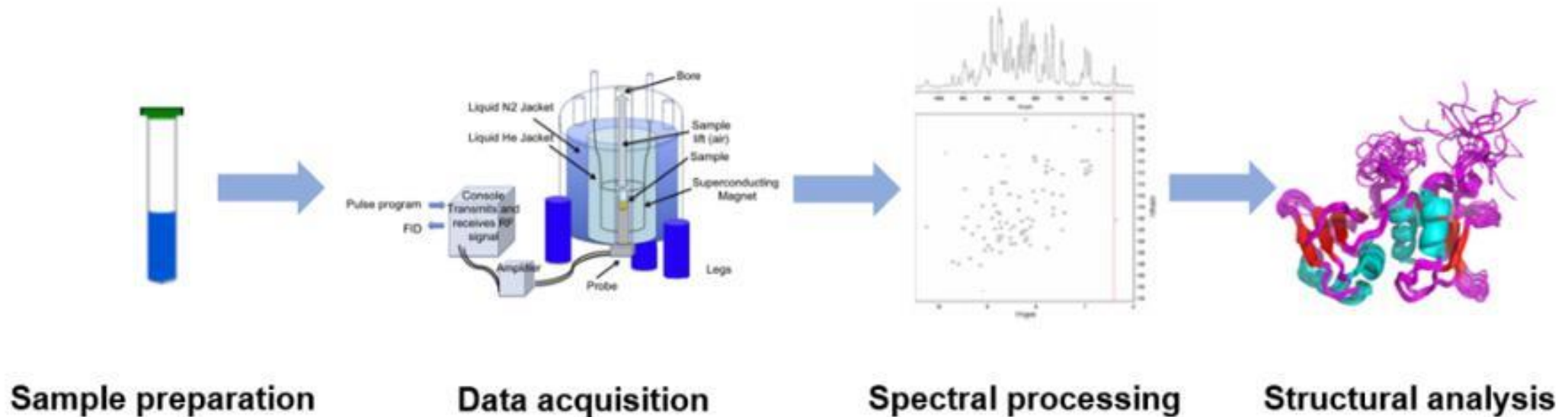
X-ray Crystallography

- 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



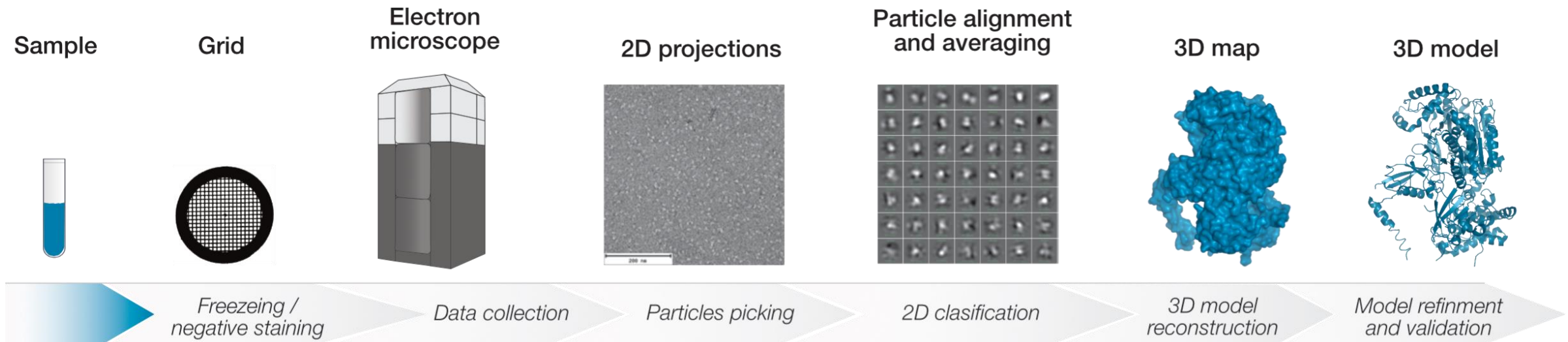
Nuclear Magnetic Resonance

- 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

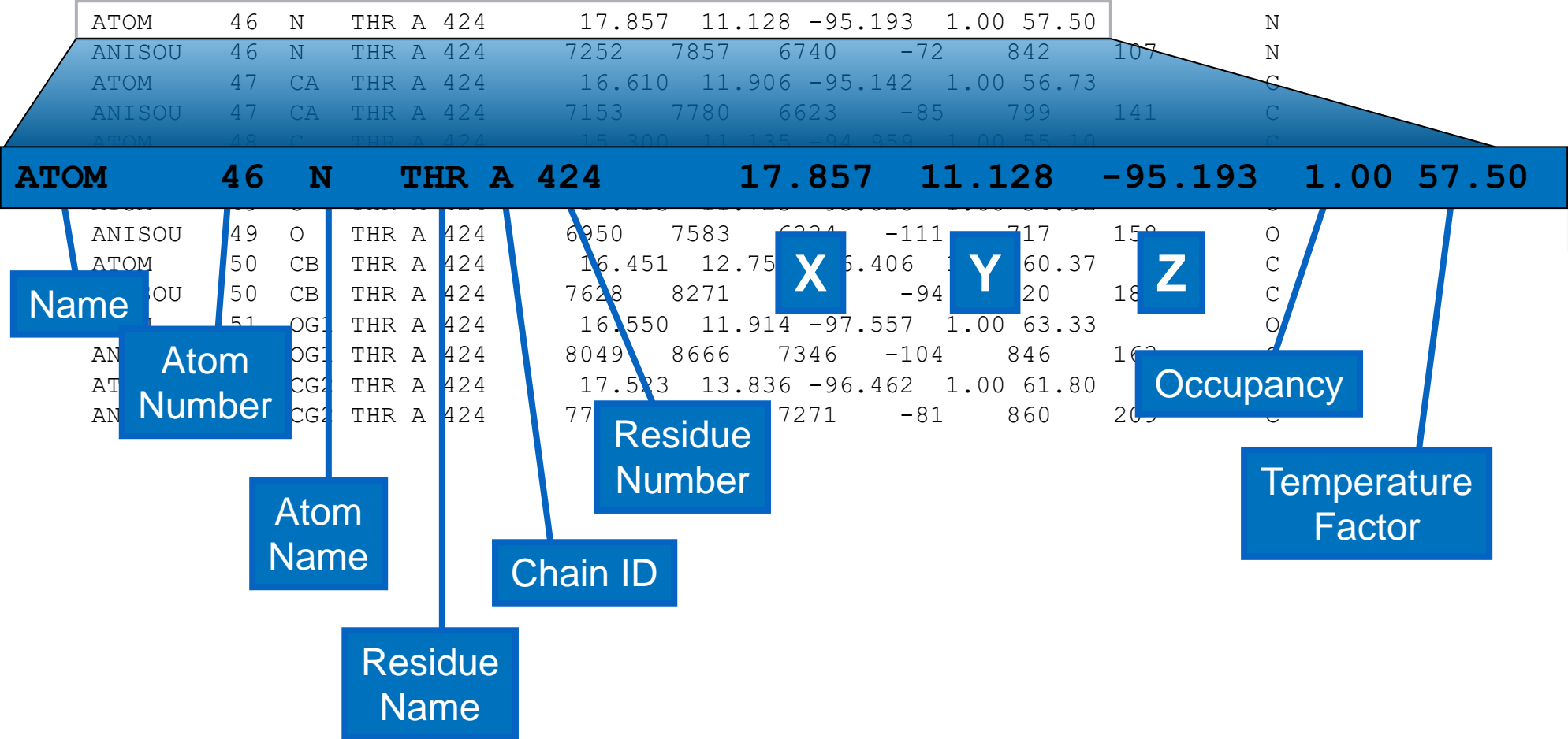
```
1TUP
```

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 ORGANISM_SCIENTIFIC: SACCHAROMYCES CEREVISIAE;
SOURCE      5 GENE: TOPOISOMERASE II;
SOURCE      6 EXPRESSION_SYSTEM: BACTERIAL EXPRESSION SYSTEM;
SOURCE      7 EXPRESSION_SYSTEM: BACTERIAL EXPRESSION SYSTEM;
SOURCE      8 EXPRESSION_SYSTEM: BACTERIAL EXPRESSION SYSTEM;
SOURCE      9 EXPRESSION_SYSTEM: BACTERIAL EXPRESSION SYSTEM;
SOURCE      10 EXPRESSION_SYSTEM: BACTERIAL EXPRESSION SYSTEM;
SOURCE      11 EXPRESSION_SYSTEM: BACTERIAL EXPRESSION SYSTEM;
SOURCE      12 MOL_ID: 2;
SOURCE      13 SYNTHETIC: YES;
SOURCE      14 MOL_ID: 3;
SOURCE      15 SYNTHETIC: YES;
```

```
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



Computational Structural Biology

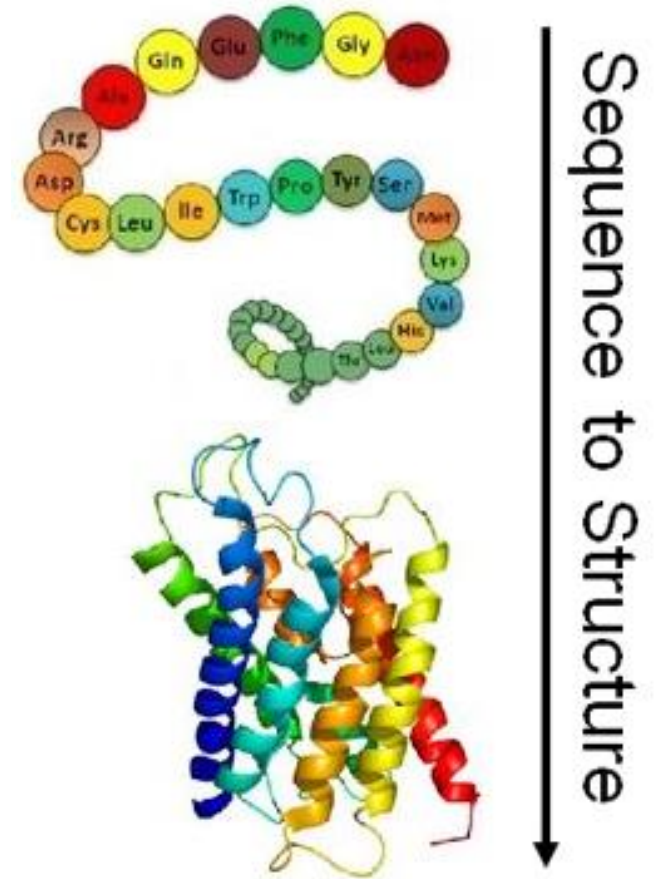
- Structure Prediction- inference of 3D structure from sequence data
- Molecular Docking- predicts the orientation of one molecule to another
- Molecular Dynamics Simulations- analyzes physical movements of atoms and molecules over time

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

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)

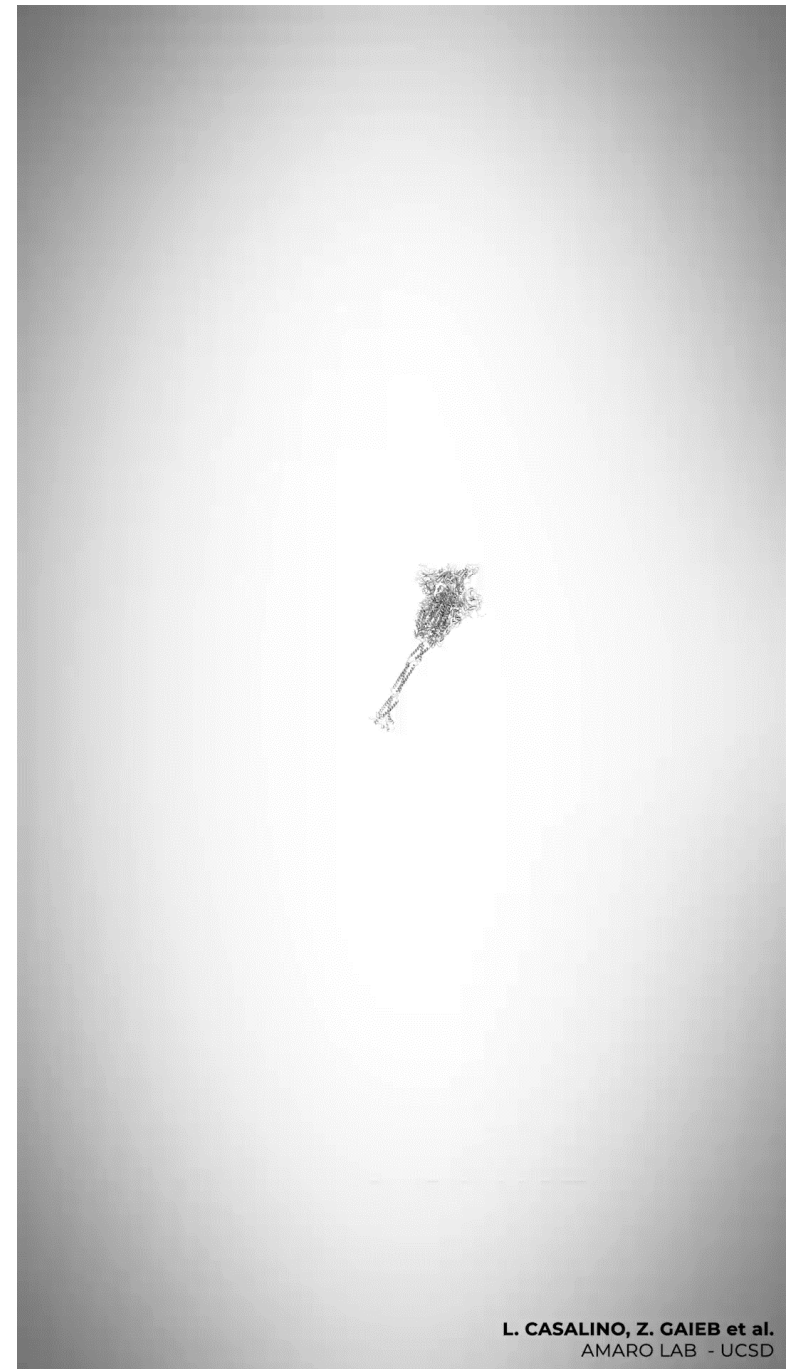


Structure Prediction Example

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 ($>4\text{\AA}$ RMSD)	Fairly accurate ($<3\text{\AA}$ 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](#)