

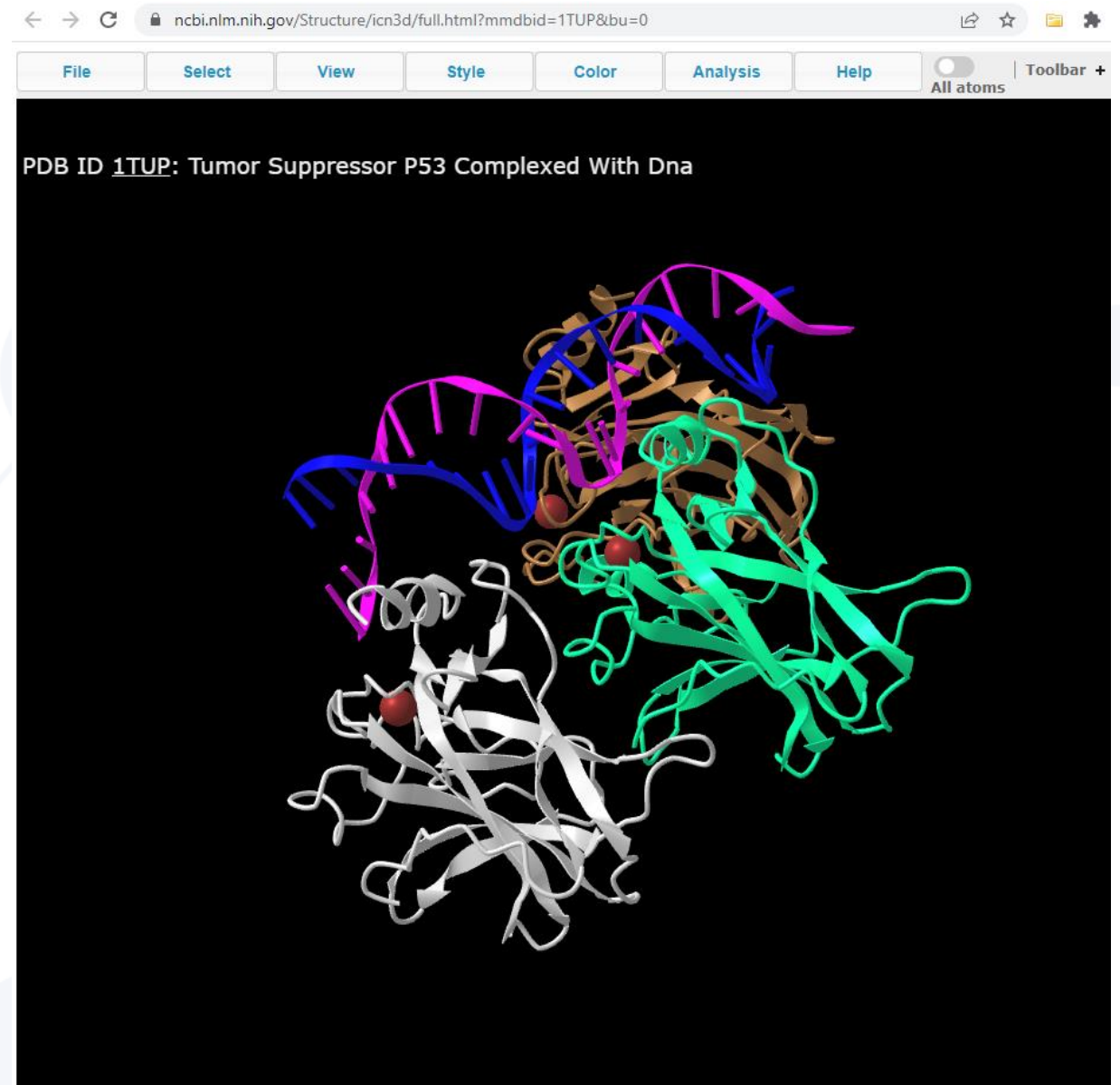


Exploring 3D Molecular Structures with iCn3D ProteinCodeathon@ISMB2022

Alexa M. Salsbury, Ph.D.

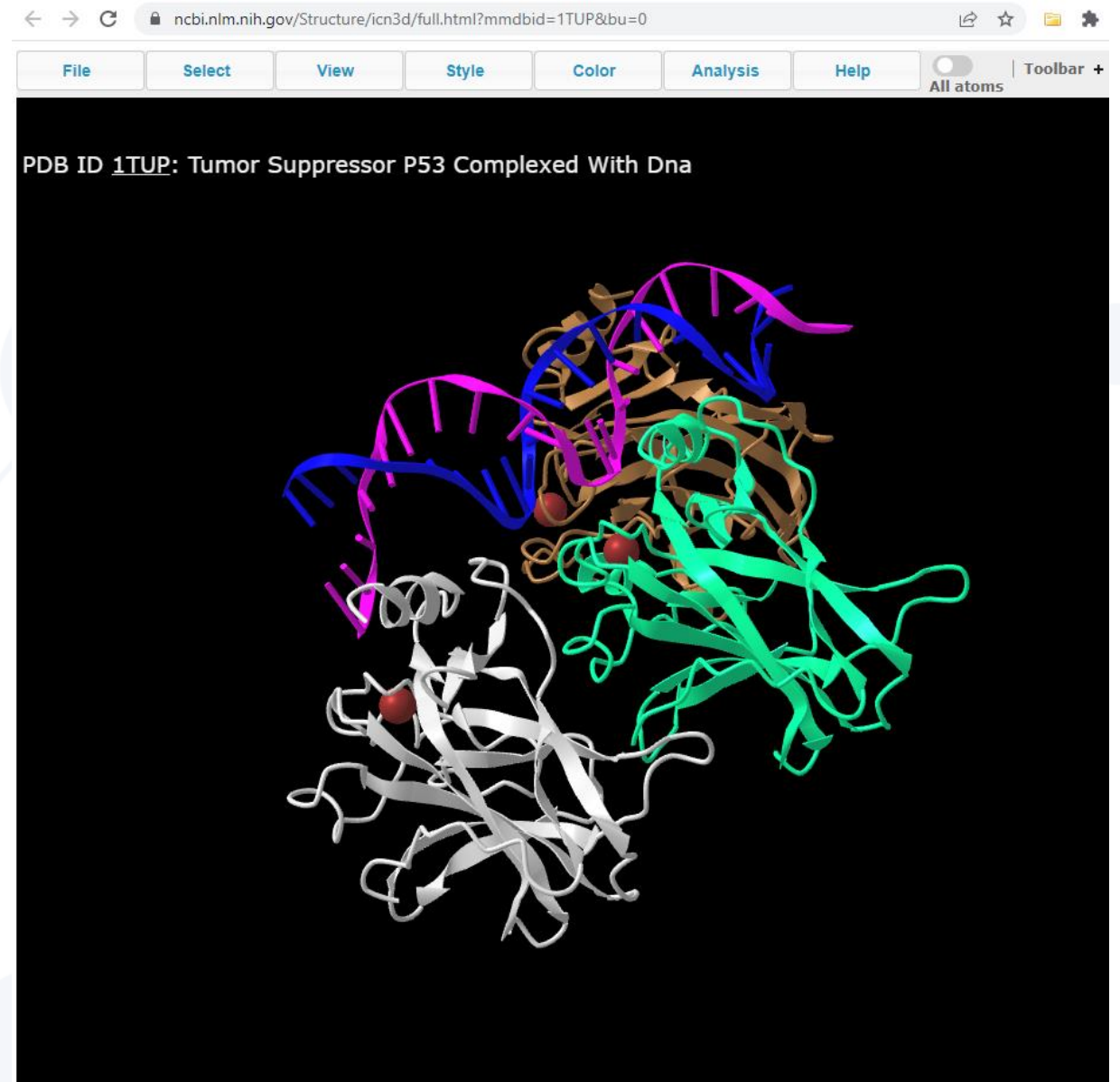
Overview

- Introduction
- iCn3D Orientation
 - Selection
 - Coloring
 - Style
- Interaction networks
- Mutation Analysis
- Continued learning materials



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
 - Save/share links of their customized display



iCn3D Features of Interest

- Use iCn3D in Jupyter Notebook: 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

New Features!

- Virtual reality view
- Batch analysis

NCBI at BOSC 2022

Jiyao Wang, Ph.D.

Poster

From web-based 3D viewer
to structural analysis tool
in batch mode

July 13

12:30 - 14:30 CDT



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

NCBI Resources How To Sign in to NCBI

Structure Structure Search Help

Advanced

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.

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+](#)

More Resources

- [PDB](#)
- [Protein](#)
- [CDD](#)
- [PubChem](#)
- [NCBI Structure Group Resources & Research](#)

You are here: NCBI > Domains & Structures > Structure (Molecular Modeling Database) Support Center

| | | | | |
|---|---|---|--|--|
| GETTING STARTED <ul style="list-style-type: none">NCBI EducationNCBI Help ManualNCBI HandbookTraining & TutorialsSubmit Data | RESOURCES <ul style="list-style-type: none">Chemicals & BioassaysData & SoftwareDNA & RNADomains & StructuresGenes & ExpressionGenetics & MedicineGenomes & MapsHomologyLiteratureProteinsSequence AnalysisTaxonomyVariation | POPULAR <ul style="list-style-type: none">PubMedBookshelfPubMed CentralBLASTNucleotideGenomeSNPGeneProteinPubChem | FEATURED <ul style="list-style-type: none">Genetic Testing RegistryGenBankReference SequencesGene Expression OmnibusGenome Data ViewerHuman GenomeMouse GenomeInfluenza VirusPrimer-BLASTSequence Read Archive | NCBI INFORMATION <ul style="list-style-type: none">About NCBIResearch at NCBINCBI News & BlogNCBI FTP SiteNCBI on FacebookNCBI on TwitterNCBI on YouTubePrivacy Policy |
|---|---|---|--|--|

National Center for Biotechnology Information, U.S. National Library of Medicine

iCn3D Shortcuts

Rotate

- **Left mouse button** can be used to rotate the structure
- **Key L** - left
- **Key J** - right
- **Key I** - up
- **Key M** - down

Zoom

- **Middle mouse button** can be used to zoom
- **Left Mouse + Shift** can be used as an alternative to the middle mouse button
- **Key Z** - zoom in
- **Key X** - zoom out

Translate

- **Right mouse button** - can be used to translate (slide) the structure to a different location within the 3D window
- **Left Mouse + Ctrl** can be used as an alternative to the right mouse button
- **Arrow Left** - left
- **Arrow Right** - right
- **Arrow Up** - up
- **Arrow Down** - down

Select

- **Alt + Click (PC)** or **Option + Click (Mac)**- can be used to select atom/residue/strand , hold **Ctrl + Click** to add another

iCn3D Demo (1/5)

- Click **full-feature 3D viewer** on the Molecular Graphic
- Or go to <https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>
 - Input **1TUP > Load Biological Unit**
- Orient yourself (see iCn3D Shortcuts for help)
- Hover over structure with your mouse to view residues
- **Select > Select on 3D > Atom** to see atomistic details
- You can revert to selecting by residues **Select > Select on 3D > Residue**

iCn3D Demo (2/5)

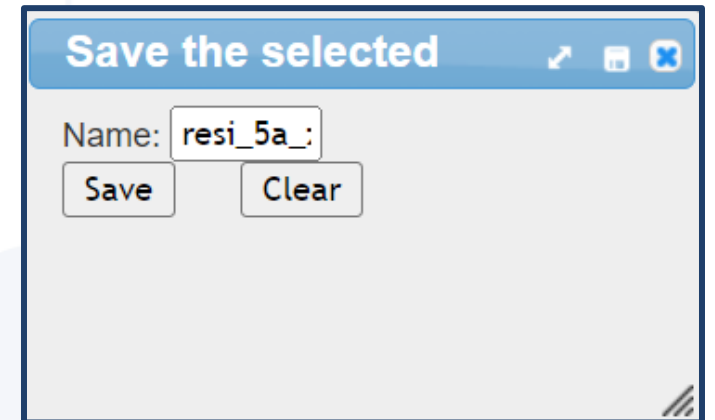
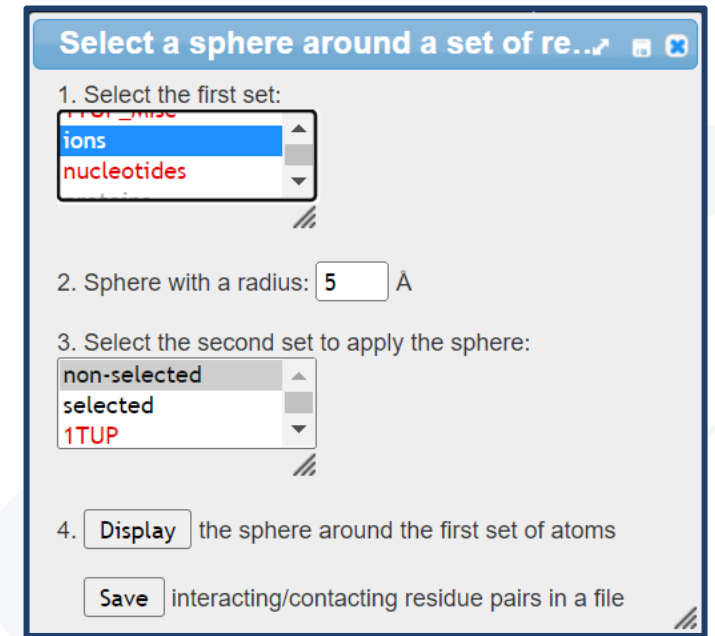
- Change styling with **Style > Sidechains > Lines**
- Explore different **Style** options
- Explore different **Color** options
 - **Chain** – default, colors structural components differently
 - **Rainbow** – N-term or 5' end is red and flows to blue for C-term or 3' end
 - **Charge** – colors positively charged as blue, negatively charged as red, and neutral as gray
 - **Atom** – colors C gray, O red, N blue, S yellow
 - **Secondary**, **Hydrophobicity**, and **Solvent Accessibility** options are useful for more in-depth analysis of structure

iCn3D Demo (3/5)

- Make specific selections with **Analysis > Defined Sets**
- Select **nucleotides** and change color with **Color > Rainbow**
- Select residues by sequence with **Analysis > Seq. & Annotations**
 - Uncheck annotations and click **Details**
 - Highlight residues from the sequence to select
- Close or minimize **Seq. & Annotations** when not using

iCn3D Demo (4/5)

- Explore ion interactions with **Select > By Distance**
 - Choose **ions**
 - Set sphere radius to **5 Å**
 - Choose **non-selected**
- Change the style of these residues with **Style > Proteins > Stick**
- Save the selection by **Select > Save Selection** and give name like **resi_5a_zn**



iCn3D Demo (5/5)

- View the ion interactions **Analysis > Defined Sets > ions** and your newly named selection **resi_5a_zn** and **View > View Selection**
- Change your background color with **Style > Background**
- Saving your files
 - As a PNG with **File > Save Files > iCn3D PNG image**
 - As an interactive link with **File > Share Link** and copy
 - For 3D printing with **File > 3D Printing**

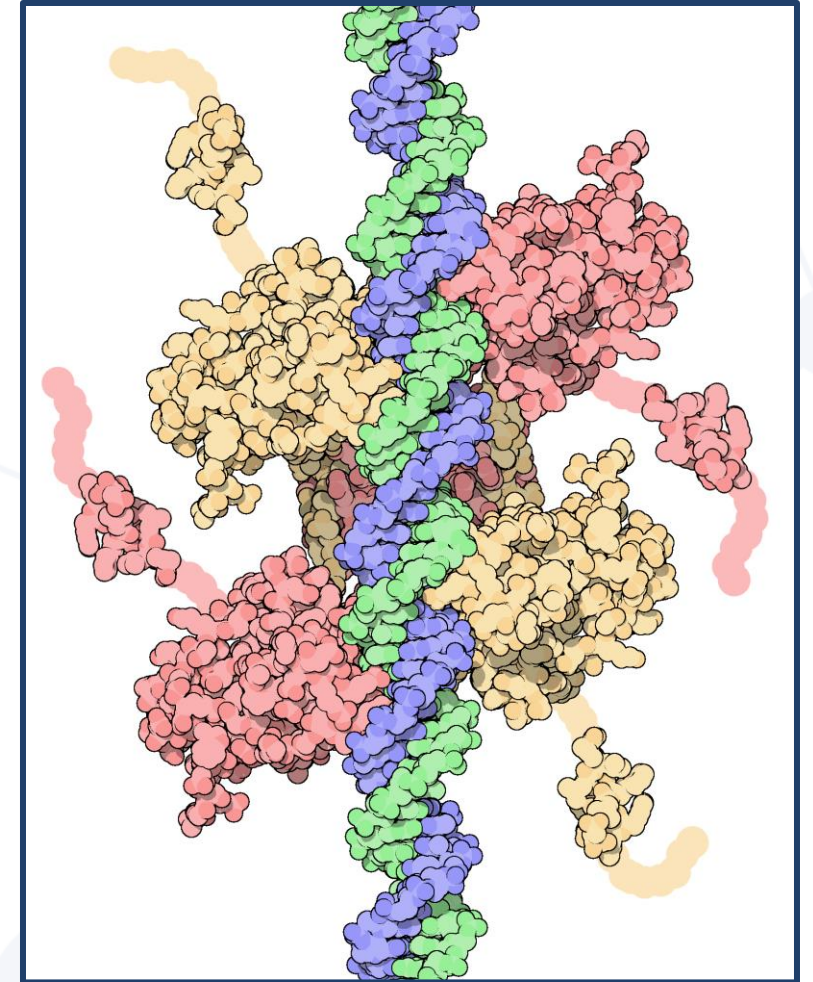
iCn3D Exploration

5-minute exercise!

- If you select something accidentally **Select > Clear Selection**
- If you need to undo **View > Undo**
- Get additional help by:
 - **Show Help > Help Docs**
 - **Help > Selection Hints**
- Like what you've rendered? Share your interactive link in the chat!
- Ask us!

P53 DNA-binding

- P53 binds to regulatory sites in the genome and:
 - Initiates protein production that stops cell division until damage is repaired
 - Initiates apoptosis
- Rich in + charged amino acids (Arg, His, Lys)
 - + charged amino acids commonly interact with negatively charged nucleic acid backbones
 - Usually interact at the major groove



P53 DNA- binding (from PDB)

Mutation Example 1

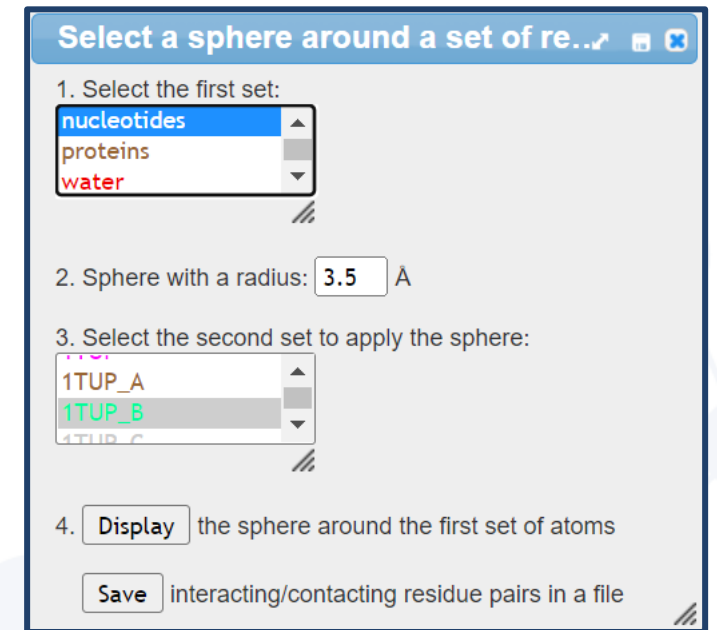
- Go to back to [iCn3D](#)
- What residues seem important to P53-binding domain?

Select > by Distance

- Take a closer look at these residues

Style > Protein > Stick

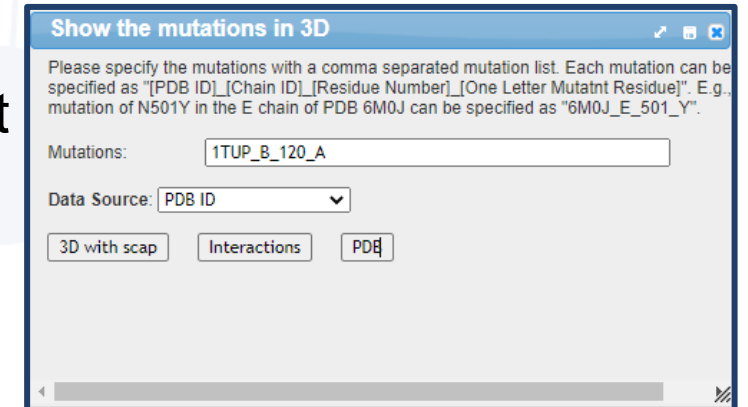
- Lys120 points into the major groove to make base-specific contact. How might a mutation affect interactions?
 - Analysis > Mutation > 1TUP_B_120_A and select Interactions



Lys120 (Lys117 in mouse) at the loop's tip points into the major groove to make a base-specific DNA contact. In contrast, here, Lys117 and the L1 loop in each subunit has moved nearly 15 Å away from the DNA and adopts some α -helical structure.

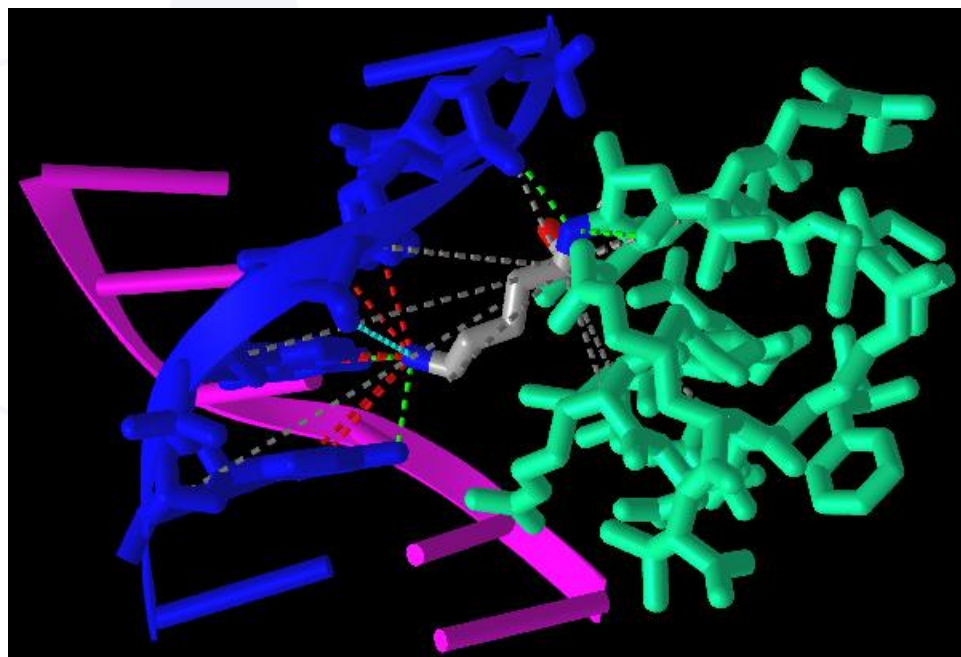
The DNA contacts made by each subunit are essentially as reported by Cho *et al.* (1994). Three minor variations, which are also observed in the other p53DBD/DNA structures, are seen in each subunit.

Oncogene

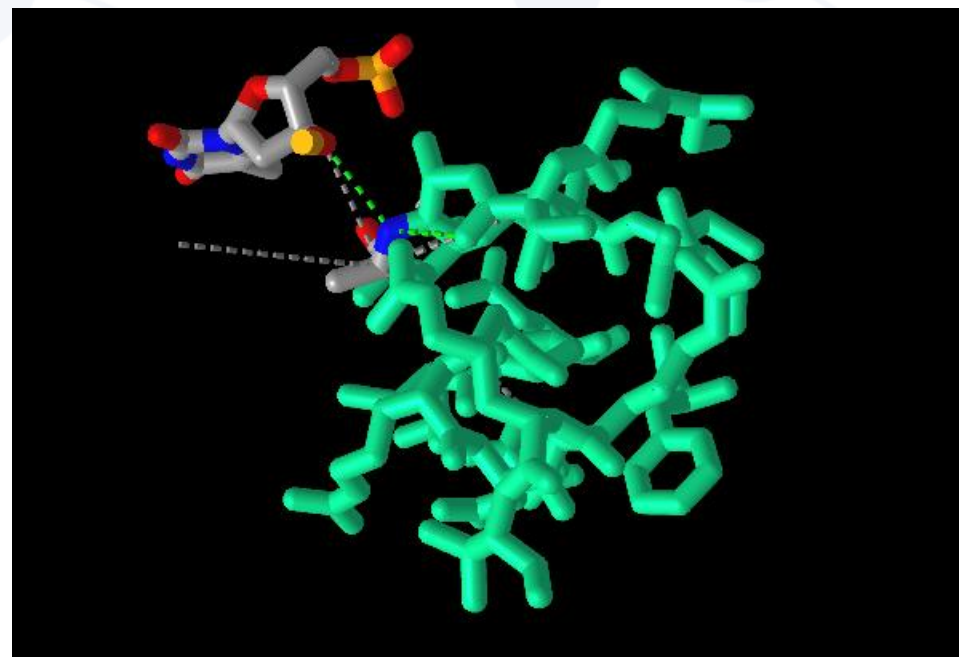


Mutation Discussion 1

- [K120A mutation](#) results in loss of interaction with the DNA groove



Wild type K120



Mutant A120

Mutation Example 2

- Literature shows that Arg248 and Arg273 are common P53 mutations implicated in disease
- Use the Mutation analysis to understand how these mutations may affect interactions

1TUP_B_248_W

1TUP_B_273_H

5-minute exercise!

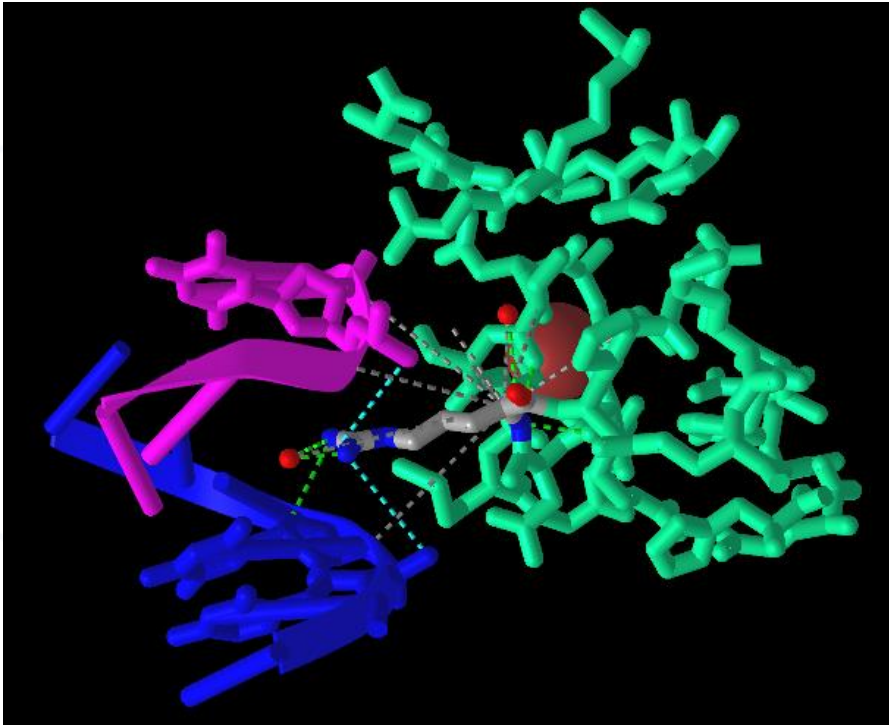
Distinct pattern of p53 phosphorylation in human tumors

Phosphorylation of mutant p53 in tumor-derived cell cultures

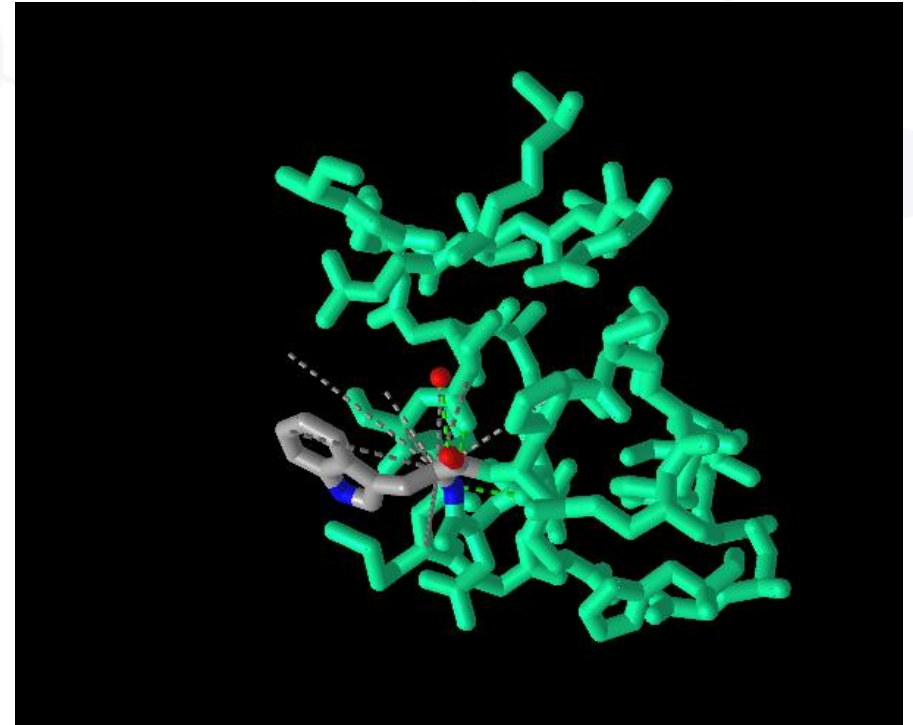
To determine the pattern of mutant p53 phosphorylation and acetylation in tumor-derived cell lines under normal growth conditions we analysed 18 cell lines with defined p53 mutations. Cell lines used for phospho-analysis were derived from seven tumor types and included a total of nine different mutations. Several tumor-derived cell lines with the same hot spot mutation (R248W or R273H) were included in this analysis to enable comparison of the phosphorylation pattern among different tumors that have the same mutation. As controls for this analysis, we used two non-transformed fibroblast cell lines (GM00038 and TIG) known to harbor wild type p53. Additional analysis has been carried out in parallel on tumor-derived tissues and cell lines that harbor wild type p53, thus allowing comparison of the phosphorylation pattern of wild type and mutant forms of p53 within the tumor environment.

Mutation Discussion 2

- [R248W mutation](#) results in loss of hydrogen and salt bridge interactions



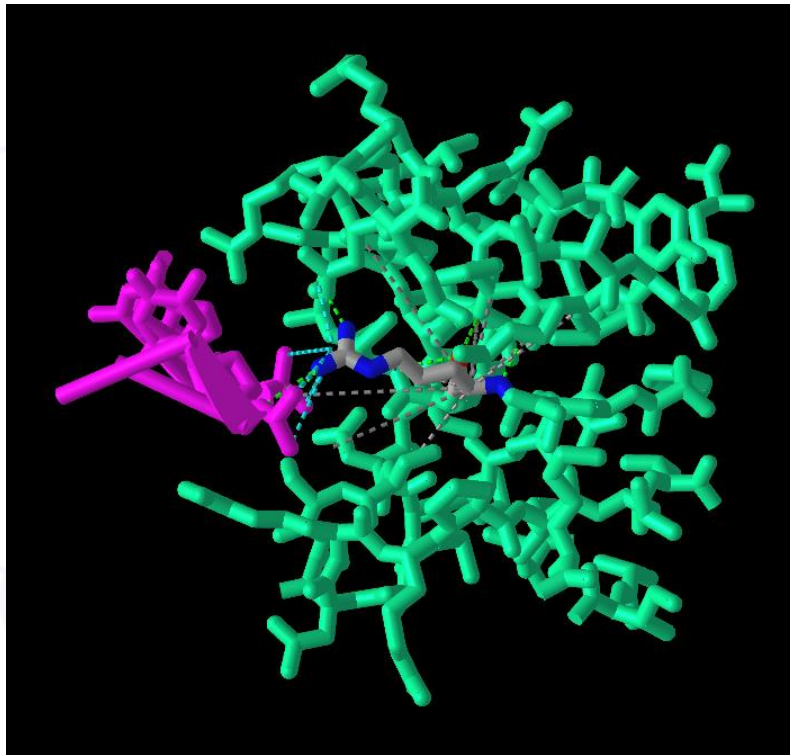
Wild type R248



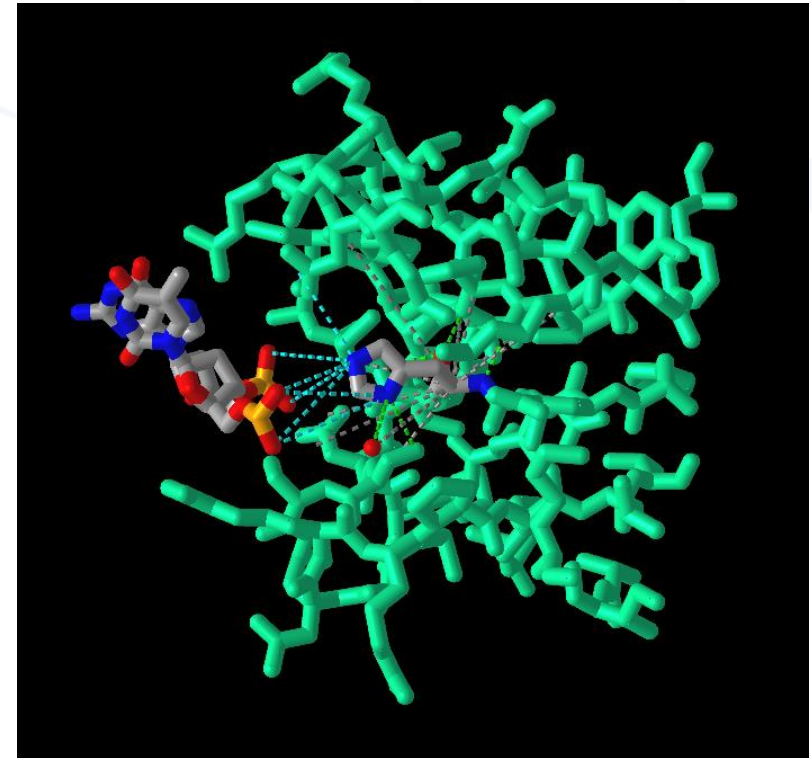
Mutant W248

Mutation Discussion 3

- [R273H mutation](#) results in loss of hydrogen bond interactions



Wild type R273



Mutant H273

Continue learning about iCn3D

Tutorials and help documents are available [here](#).

The screenshot displays the iCn3D web interface. At the top, it shows the NIH logo and the text "U.S. National Library of Medicine" and "NCBI National Center for Biotechnology Information". 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 V". 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 for keywords, a location dropdown, and a date range selector. Below the search bar are two "Select Some Options" dropdowns. 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).



Exploring 3D Molecular Structures with iCn3D Supplemental Learning Materials

Alexa M. Salsbury, Ph.D.



National Library of Medicine
National Center for Biotechnology Information



Structural Biology

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

Now- over 175,000 structures are publicly available!

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

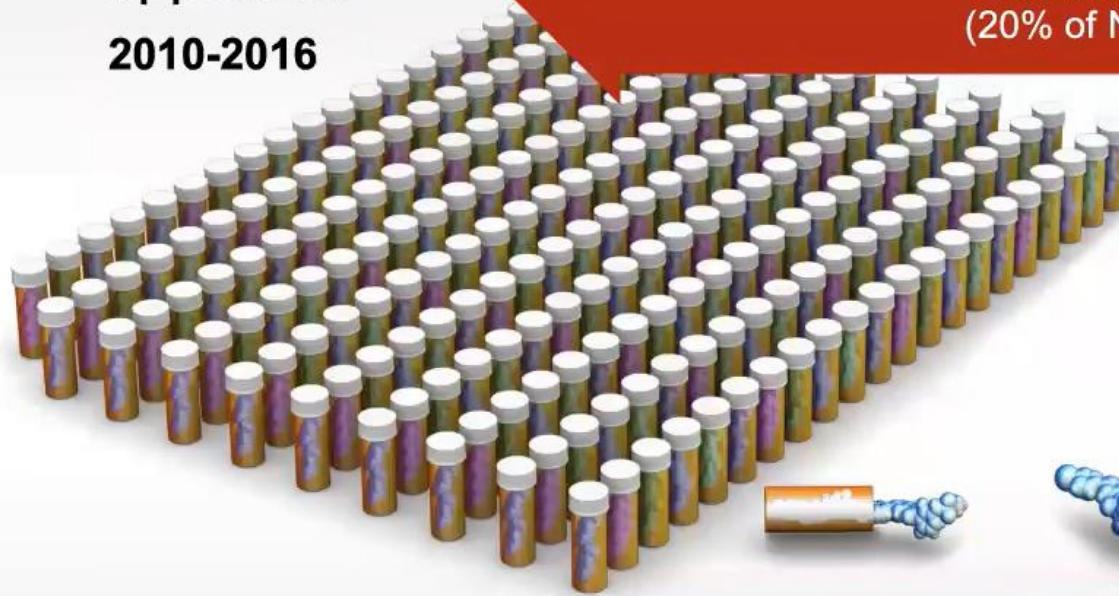
Impact of Structures on New Drug Approvals

210 NEW DRUGS
approved
2010-2016

>\$100 BILLION

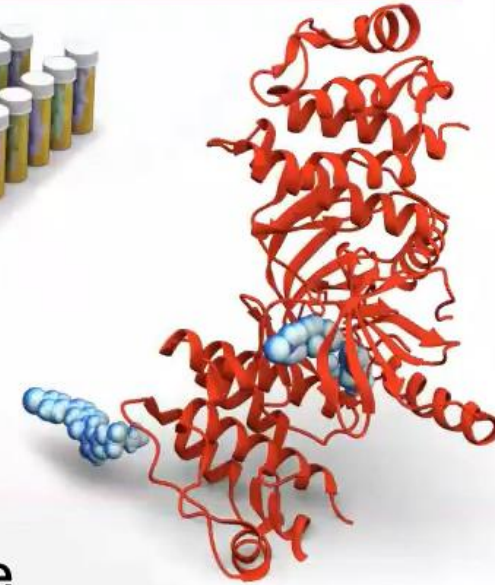
of NIH funding
contributed to these approvals
(20% of NIH Budget)²

2000-2016



5,913 PDB Structures
contributed to

184 of these
drug approvals



*B-Raf Kinase
complex with
Vemurafenib
PDB ID 3og7*

Impact of Structures on Anti-Cancer Drug Approvals

79 NEW ANTI-CANCER DRUGS Approved 2010-2018



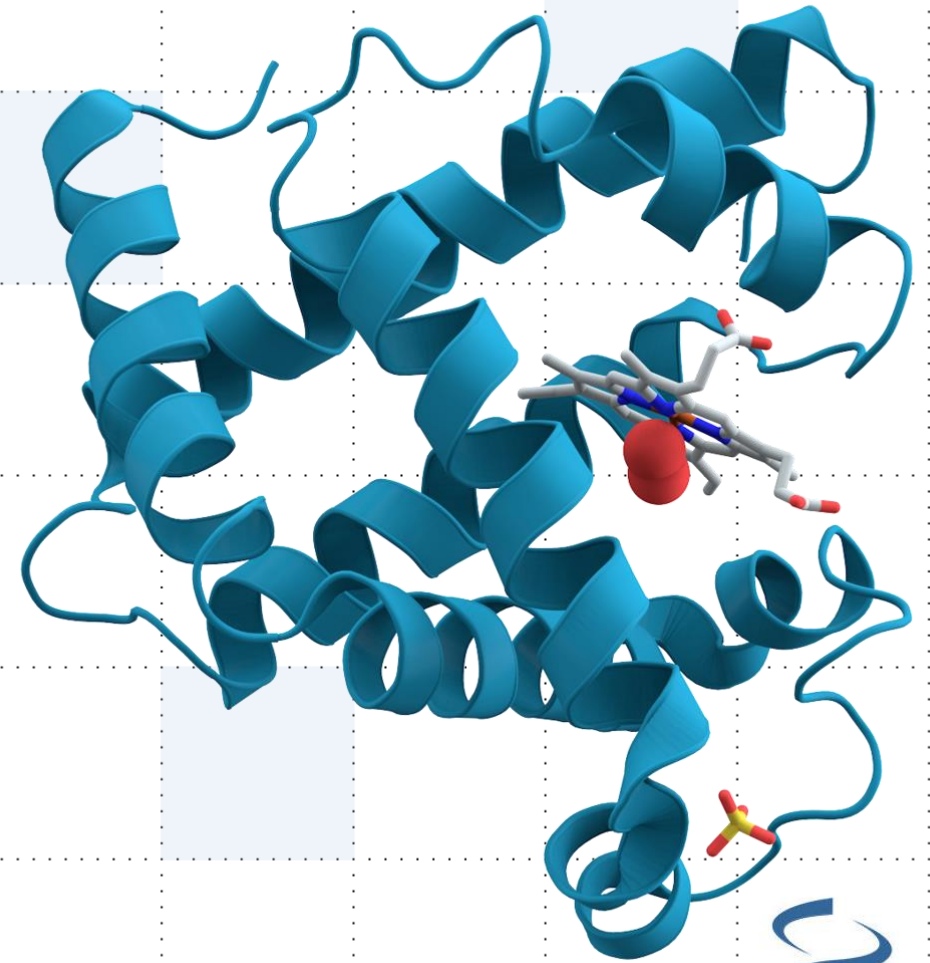
74 of these new drugs

had a total of **2412** unique structures in the PDB explaining target biology and facilitating discovery/development

Structure-guided drug discovery → >70% of small-molecule drugs

To keep you alive, your body must:

- Convert oxygen to usable energy
- Digest food and recover nutrients
- Build and maintain body tissue
- Fight off invasions and detoxify poisons

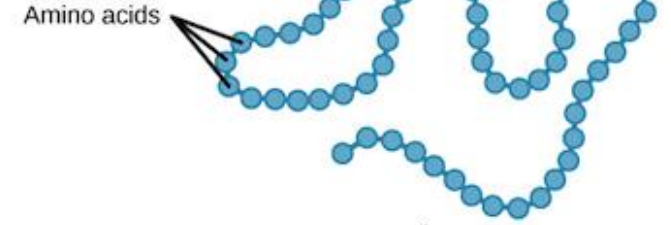


Proteins do all these things!

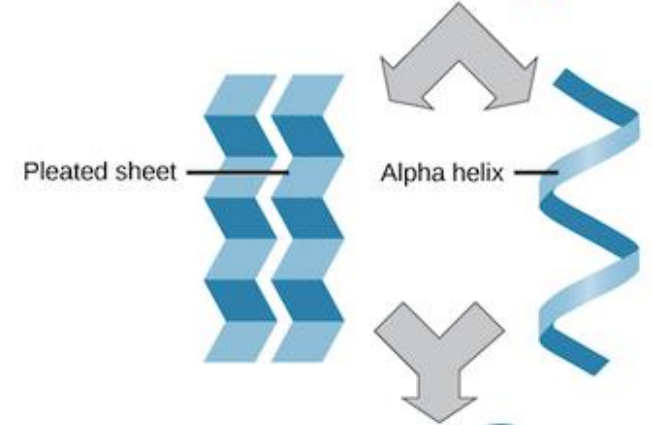
Protein Structure

- **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



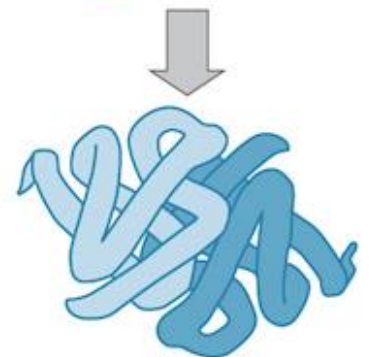
Secondary



Tertiary



Quaternary



Classification of Proteins – Size

- Most peptides and proteins have 2-2000 amino acids
- Assuming average MW per amino acid of 110 daltons, MW range for peptides and proteins in range 220-220,000 daltons
- Some proteins consist of single polypeptide chain and are monomeric
- Others consist of multiple polypeptide chains (sometimes identical, sometimes not) and are oligomeric

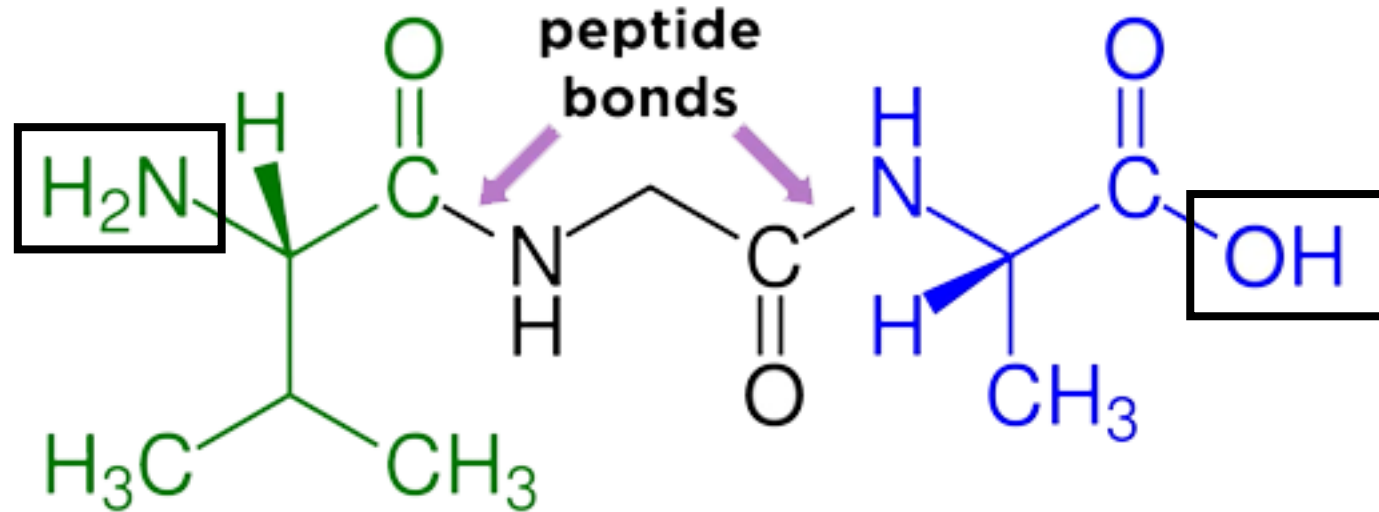
Terminology

N-terminus

(ends in amino group)

C-terminus

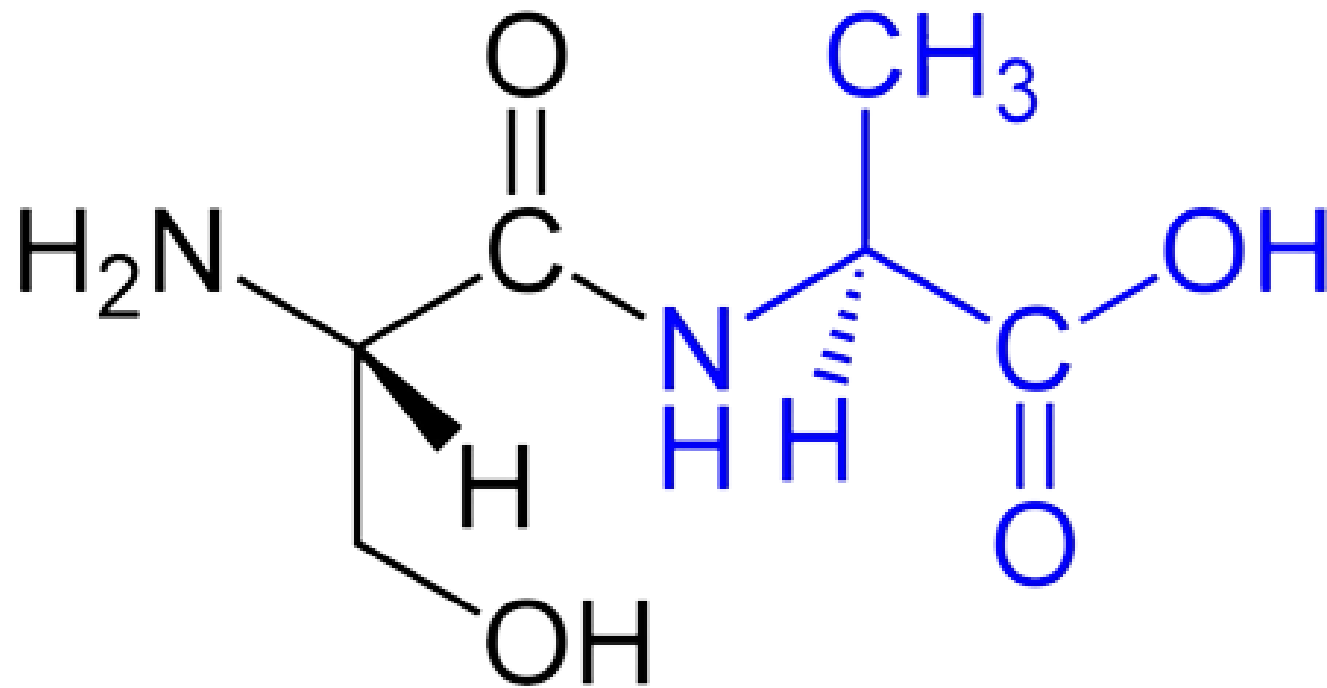
(ends in carboxyl group)



valine-glycine-alanine

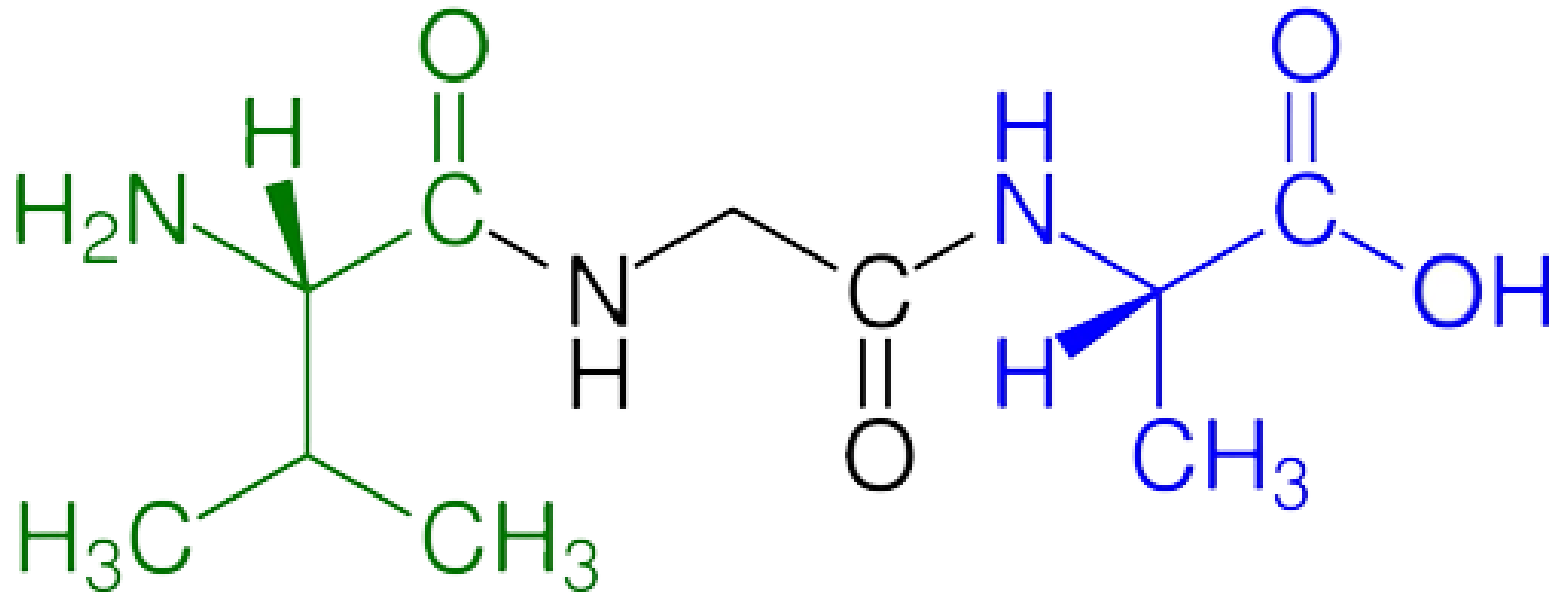
Terminology

dipeptide (2 amino acids)



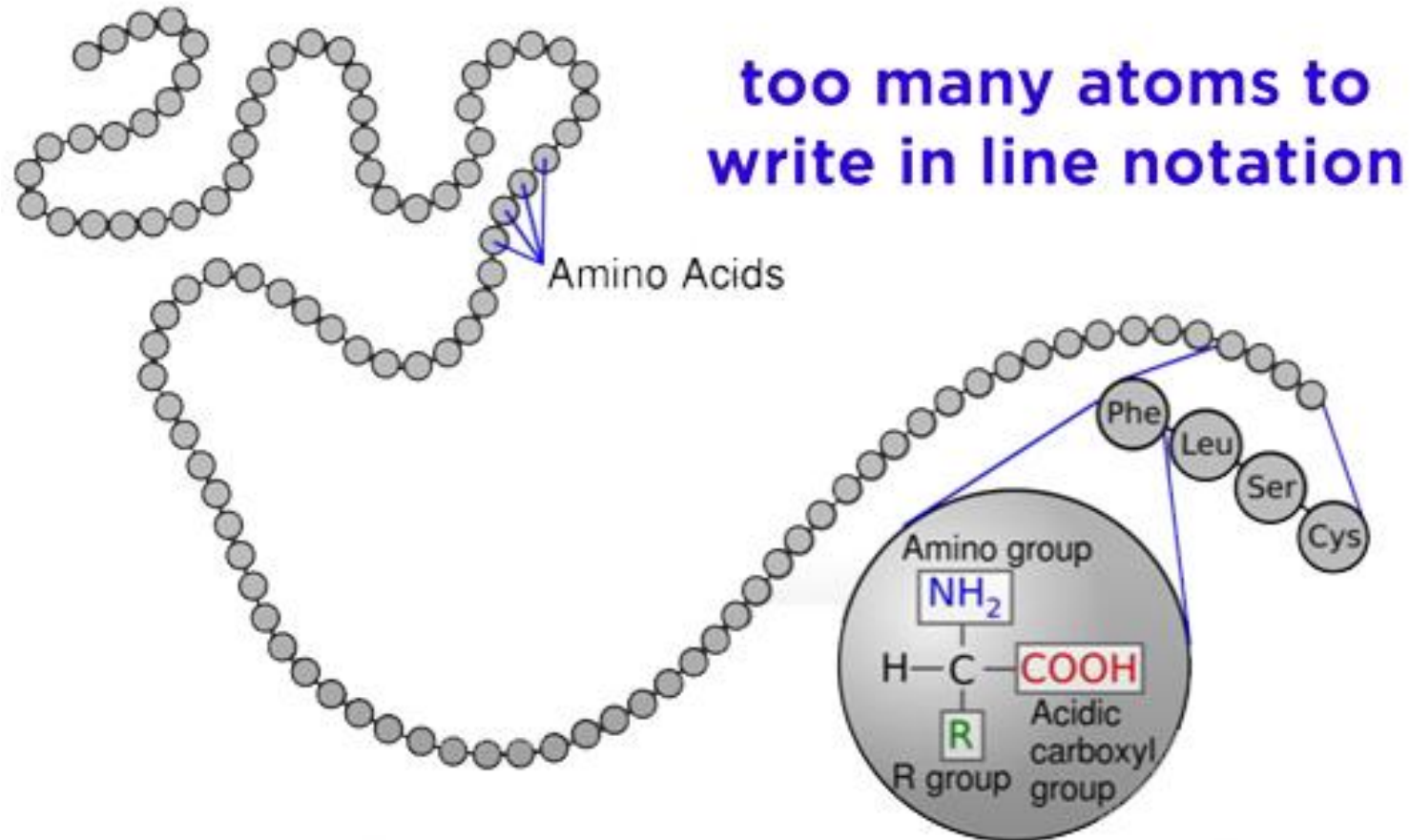
Terminology

oligopeptide (3-10 amino acids)



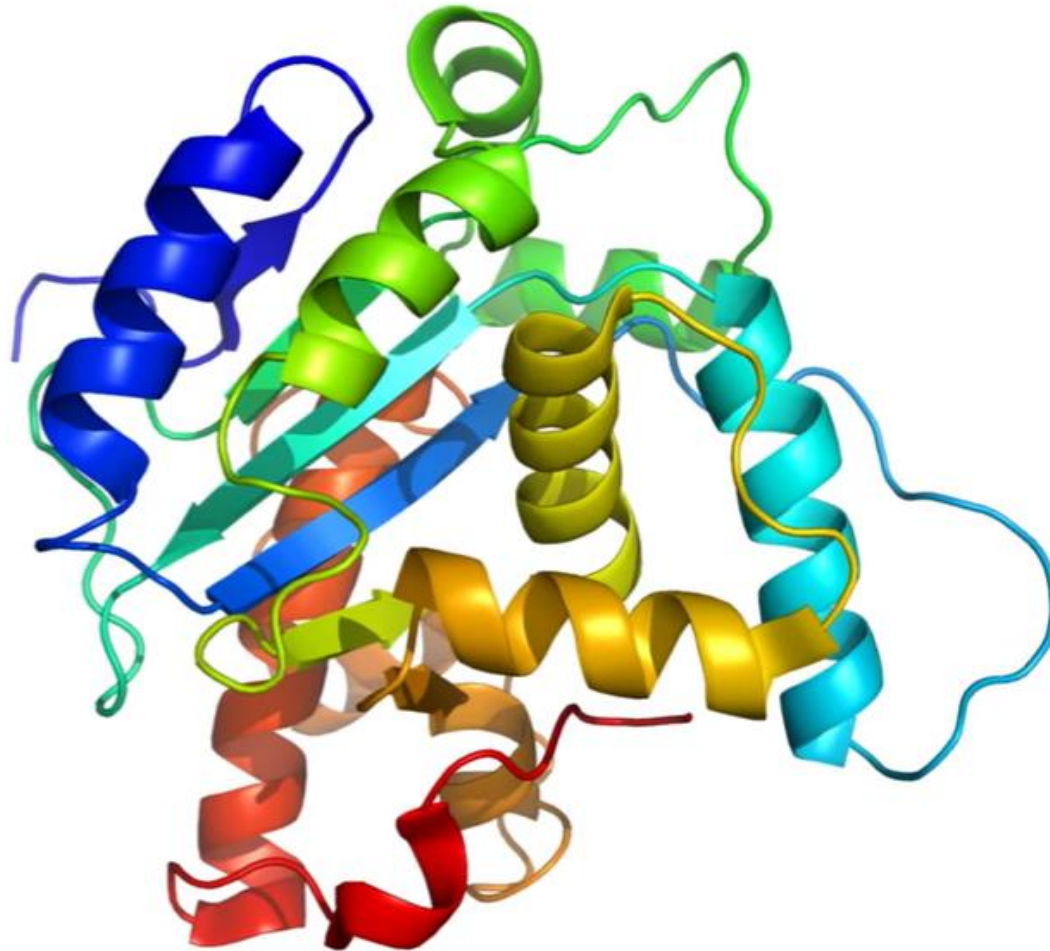
Terminology

polypeptide (>10 amino acids)

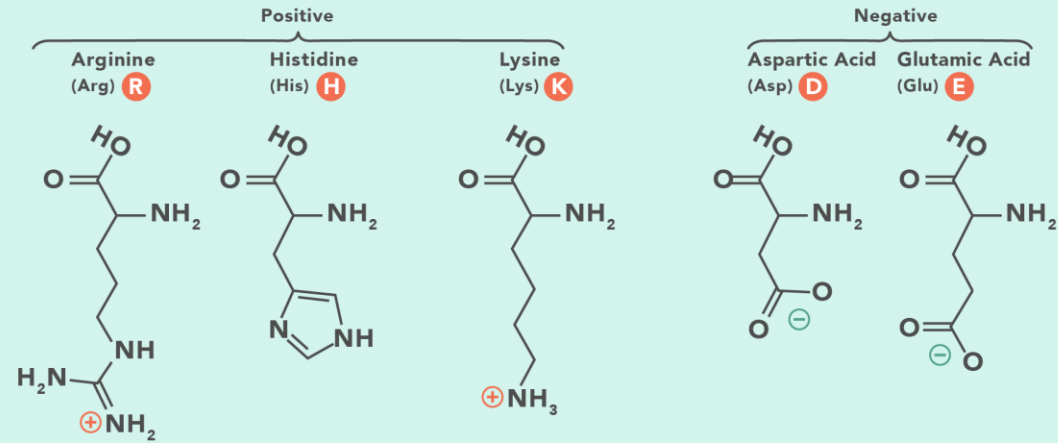


Terminology

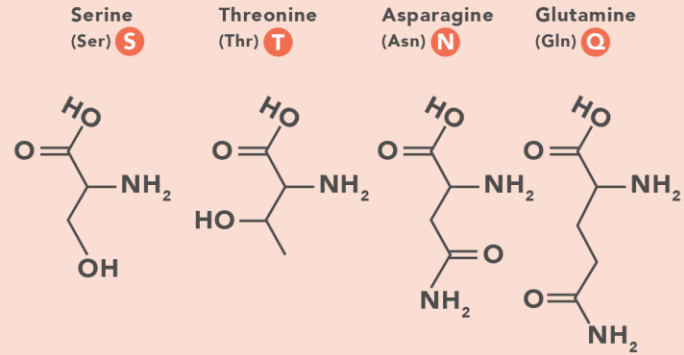
protein (generally 300-1000 amino acids)



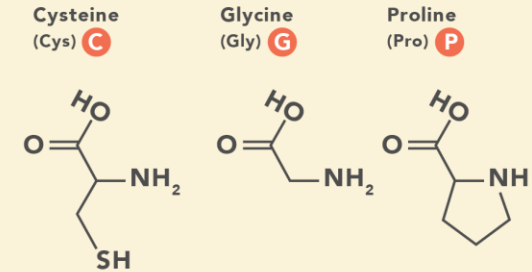
A. Amino Acids with Electrically Charged Side Chains



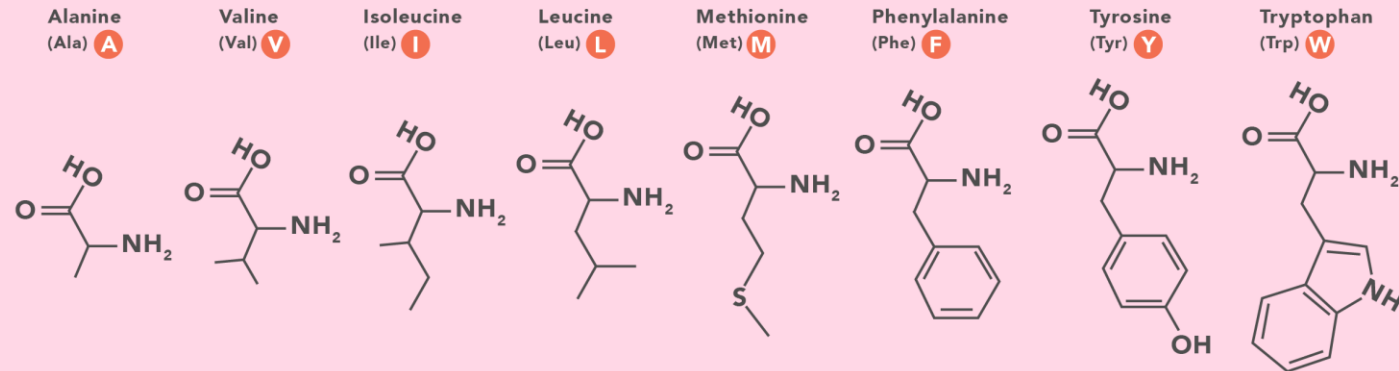
B. Amino Acids with Polar Uncharged Side Chains



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

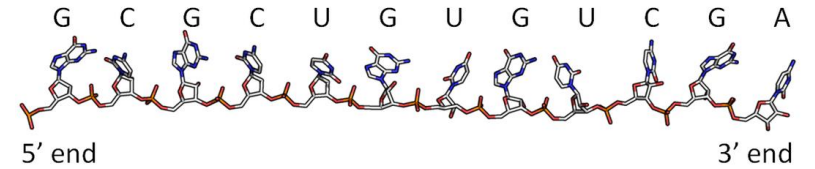
Protein Domains

- Structural Domain
 - Discrete independently folding unit of a protein
- Conserved Domain (sequence-based)
 - Protein region with recognizable position-specific pattern of sequence conservation
- Sequence-based domains often roughly correspond to structural domains
- Domains often have distinct, identifiable functions

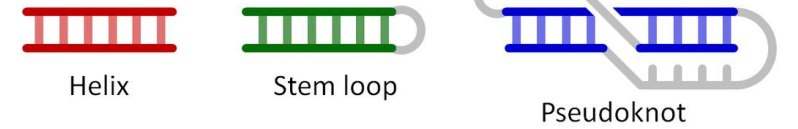
Nucleic Acid Structure

- **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)

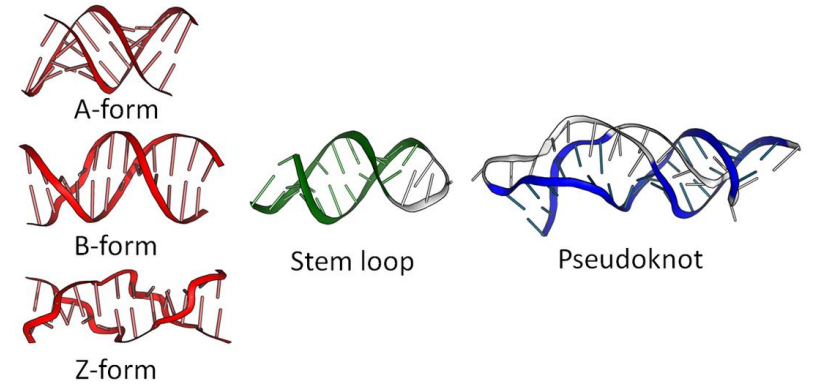
Primary



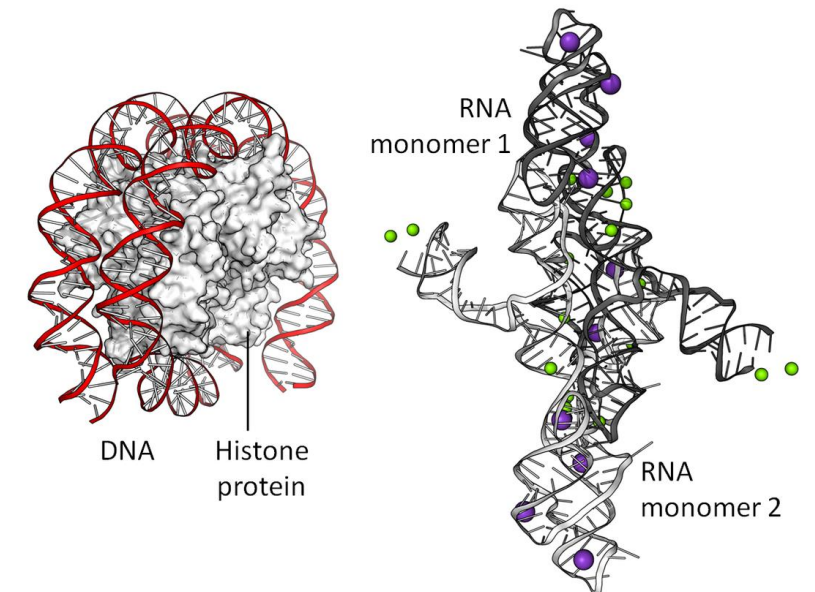
Secondary



Tertiary



Quaternary



Experimental techniques



■ Single crystal X-ray diffraction (SC-XRD)

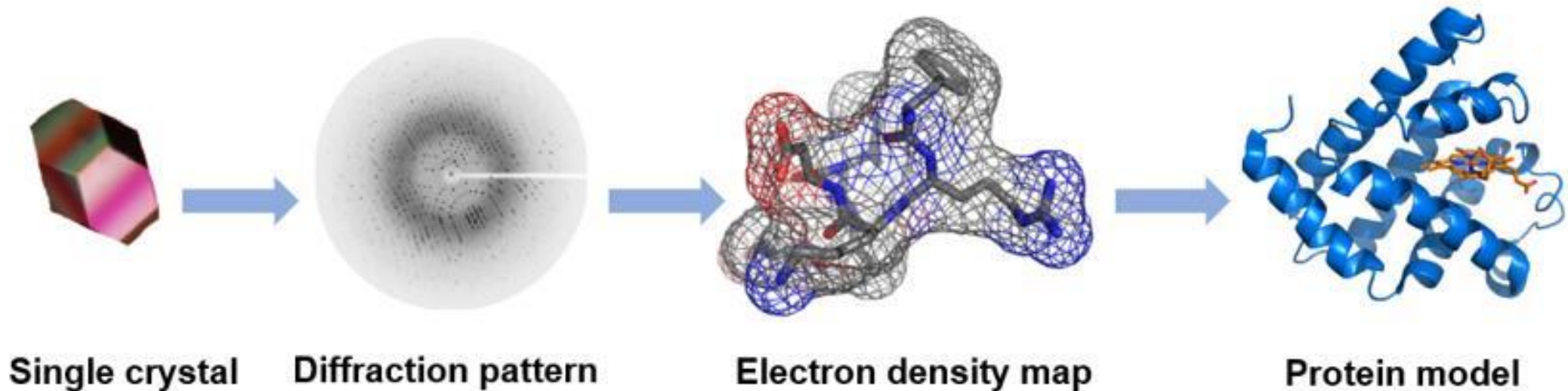
■ Nuclear magnetic resonance (NMR)

■ Cryo-electron microscopy (Cryo-EM)

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

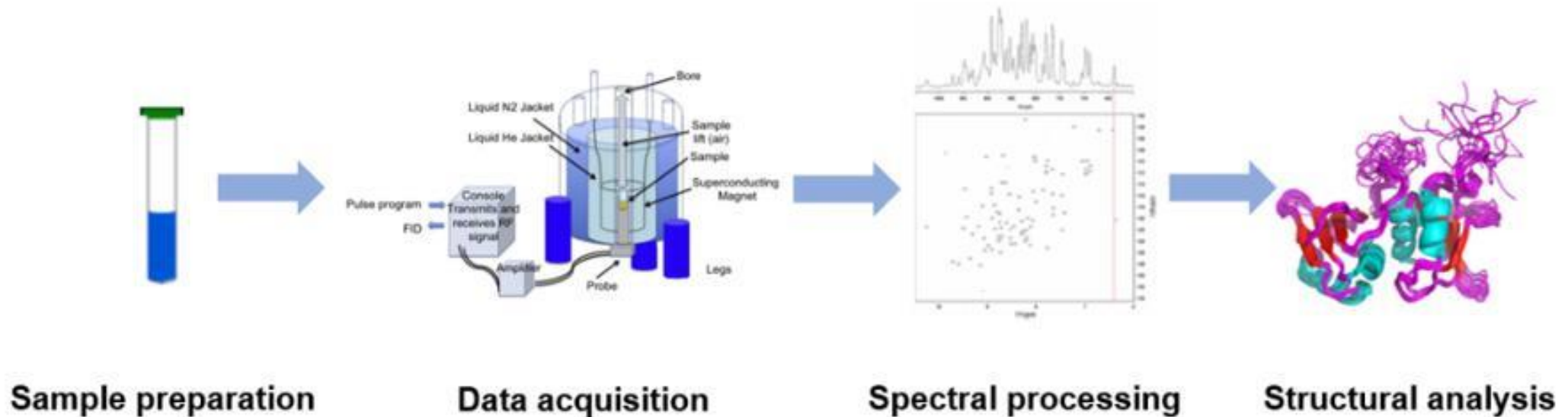
X-ray Crystallography

- 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



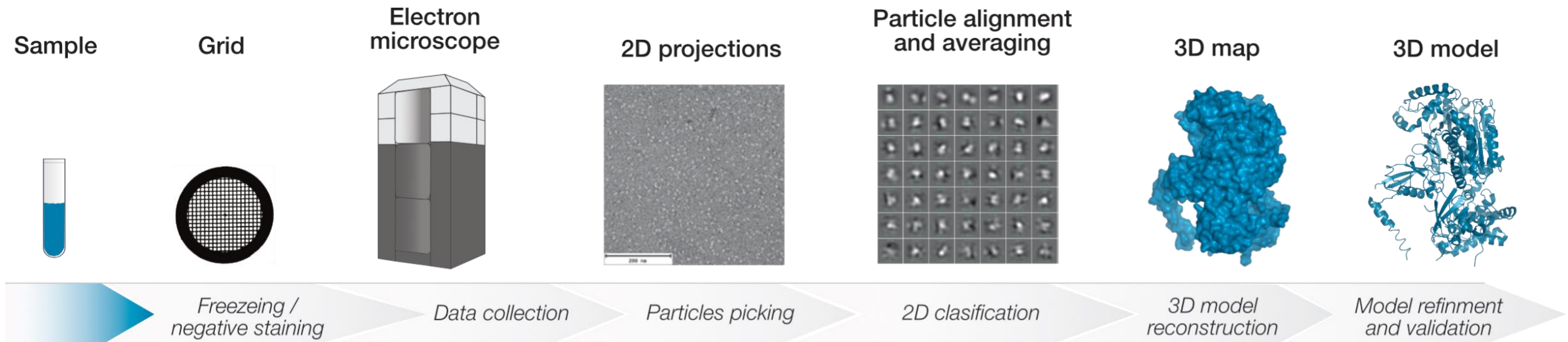
Nuclear Magnetic Resonance

- 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

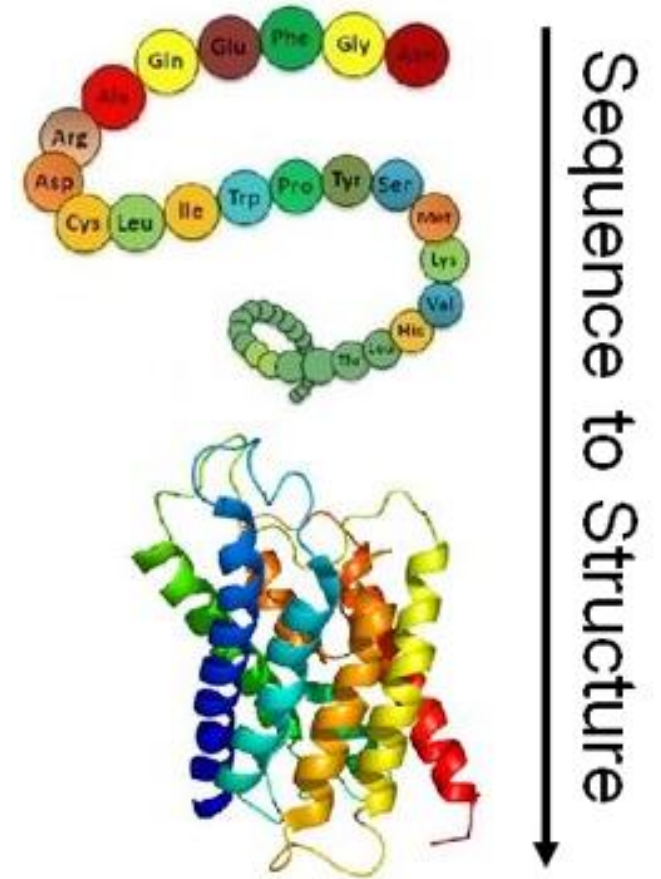


Experimental techniques

| | 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• EM equipment is costly |

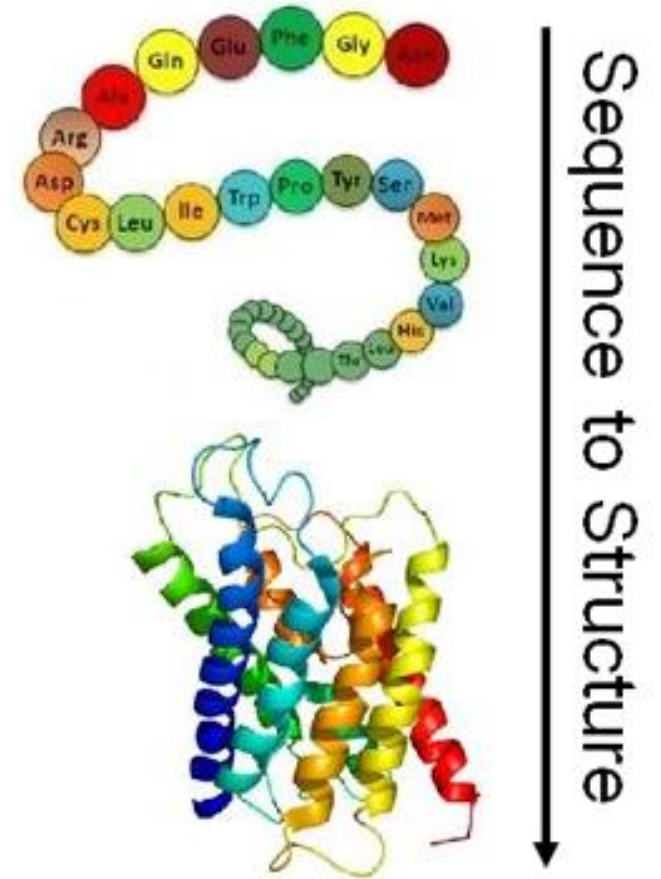
Structure Prediction Overview

- When no or incomplete structure is available
- Prominent research focus of bioinformatics and theoretical chemists
 - Drug discovery
 - Biotechnology/bioengineering
- Predicted structures aren't housed in most structural databases
 - Structure prediction websites, such as [AlphaFold](#) Protein Structure Database, are housed separately



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)

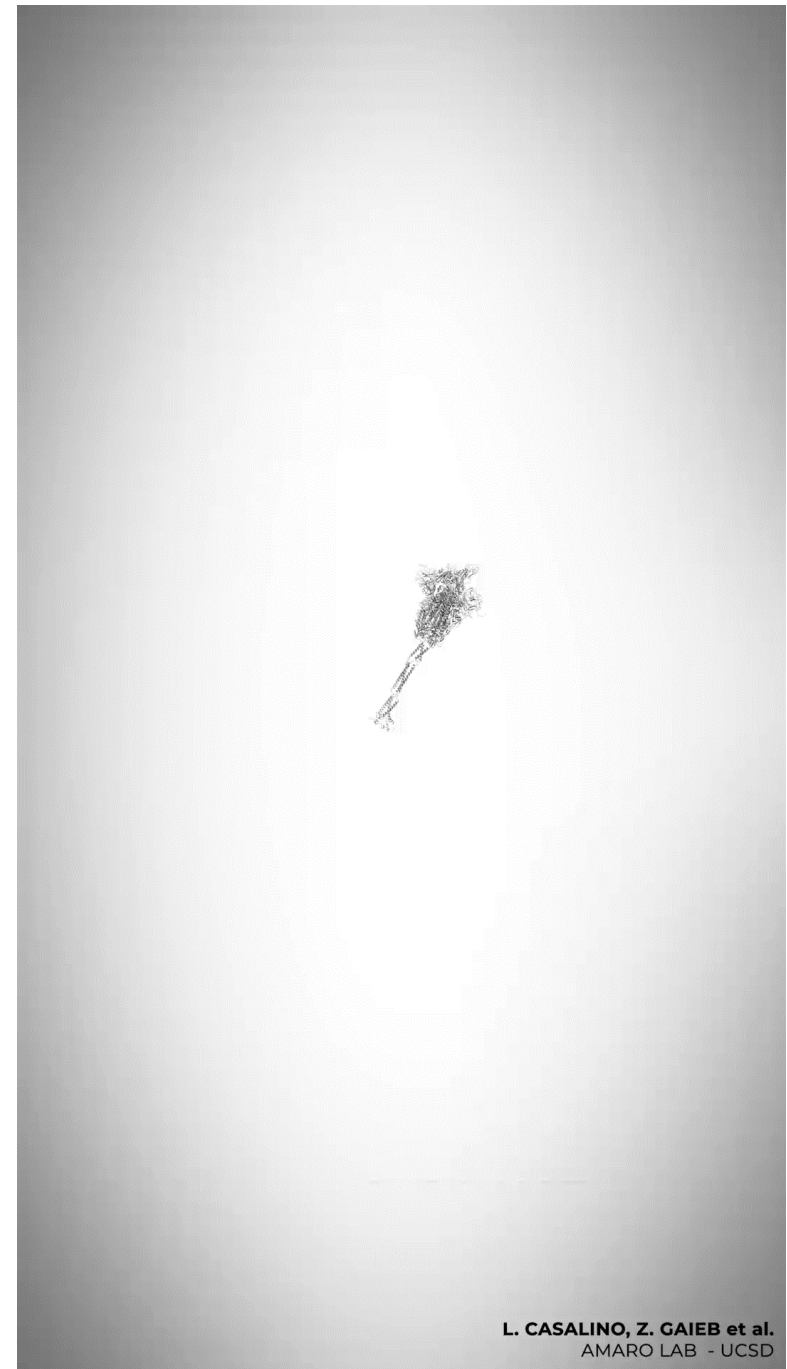


Structure Prediction Example

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



Where do I find 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. The main content area features a 'Welcome' sidebar, a 'A Structural View of Biology' section, and a 'March Molecule of the Month' section featuring a 3D model of Vascular Endothelial Growth Factor (VegF) and Angiogenesis. A 'Join the RCSB PDB Team' banner is also visible.

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. The main content area features a 3D model of a protein structure and a text box explaining that three-dimensional structures provide information on biological function and evolutionary history. Below this, there is a table with three columns: 'Using Structure', 'Structure Tools', and 'More Resources'. The 'Using Structure' column lists links for Search, How to (Quick Start) Guides, Help, News, FTP, Publications, and Discover. The 'Structure Tools' column lists links for Macromolecular Resources Overview, iCn3D (web-based 3D viewer), Cn3D (3D viewer application), IBIS, VAST, and VAST+. The 'More Resources' column lists links for PDB, Protein, CDD, PubChem, and NCBI Structure Group Resources & Research.

| 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)

~20 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

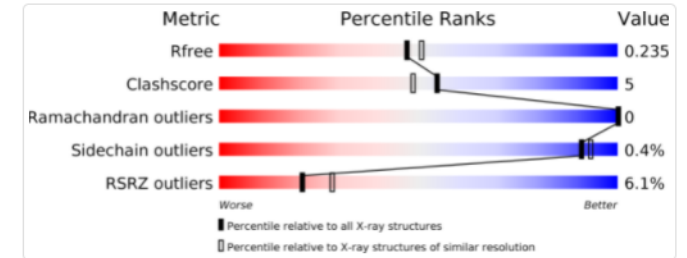
R-Value Work: 0.202

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:

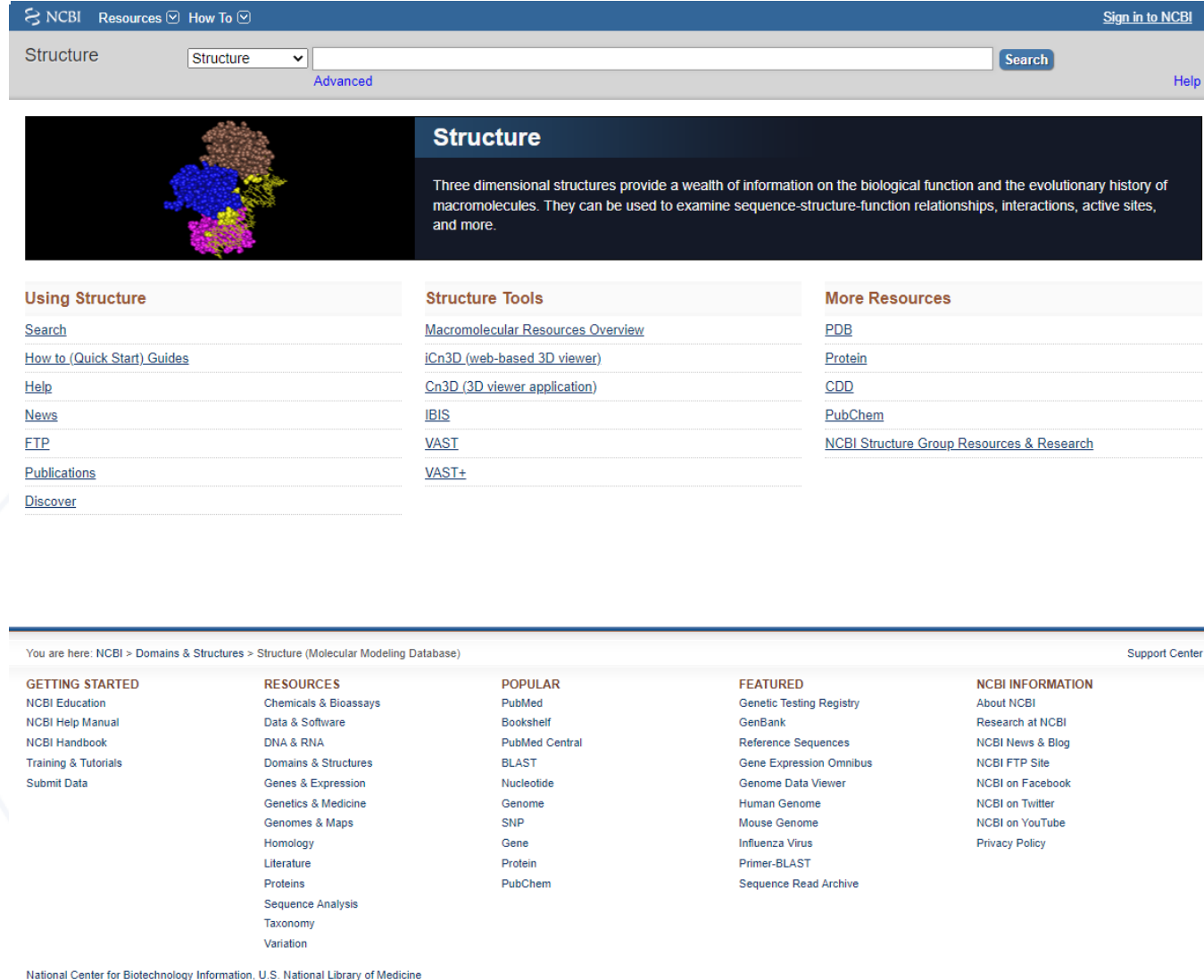
[7BQY](#), [6LU7](#)

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 ...

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



NCBI Resources How To Sign in to NCBI

Structure Structure Search Help

Advanced

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.

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+](#)

More Resources

- [PDB](#)
- [Protein](#)
- [CDD](#)
- [PubChem](#)
- [NCBI Structure Group Resources & Research](#)

You are here: NCBI > Domains & Structures > Structure (Molecular Modeling Database) Support Center

| | | | | |
|---|---|---|--|--|
| GETTING STARTED <ul style="list-style-type: none">NCBI EducationNCBI Help ManualNCBI HandbookTraining & TutorialsSubmit Data | RESOURCES <ul style="list-style-type: none">Chemicals & BioassaysData & SoftwareDNA & RNADomains & StructuresGenes & ExpressionGenetics & MedicineGenomes & MapsHomologyLiteratureProteinsSequence AnalysisTaxonomyVariation | POPULAR <ul style="list-style-type: none">PubMedBookshelfPubMed CentralBLASTNucleotideGenomeSNPGeneProteinPubChem | FEATURED <ul style="list-style-type: none">Genetic Testing RegistryGenBankReference SequencesGene Expression OmnibusGenome Data ViewerHuman GenomeMouse GenomeInfluenza VirusPrimer-BLASTSequence Read Archive | NCBI INFORMATION <ul style="list-style-type: none">About NCBIResearch at NCBINCBI News & BlogNCBI FTP SiteNCBI on FacebookNCBI on TwitterNCBI on YouTubePrivacy Policy |
|---|---|---|--|--|

National Center for Biotechnology Information, U.S. National Library of Medicine

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

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

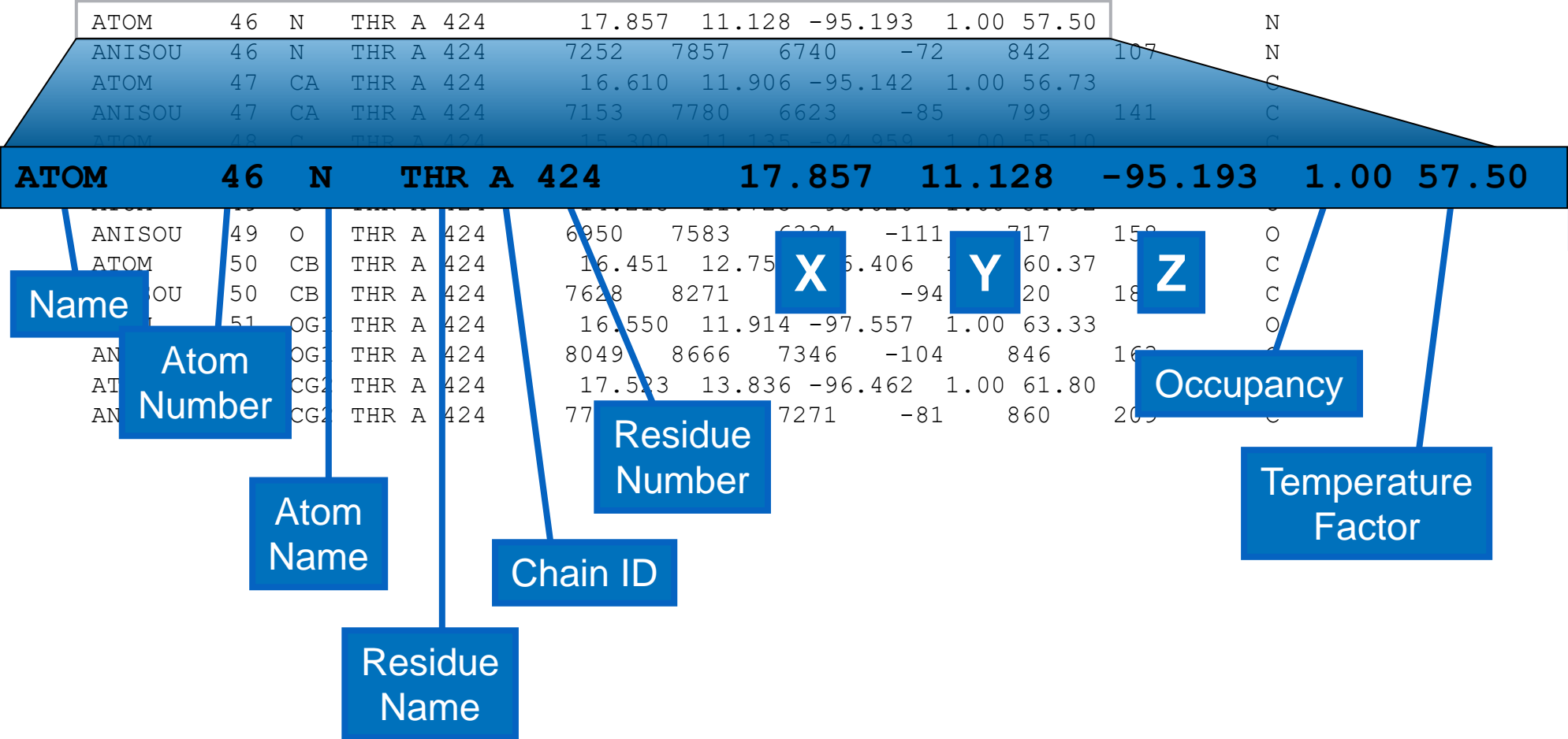
Items: 1 to 20 of 288

```
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 EXPRES: TOPOISOMERASE II;
SOURCE      7 EXPRES: TOPOISOMERASE II;
SOURCE      8 EXPRES: TOPOISOMERASE II;
SOURCE      9 EXPRES: TOPOISOMERASE II;
SOURCE     10 EXPRES: TOPOISOMERASE II;
SOURCE     11 EXPRES: TOPOISOMERASE II;
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

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 |