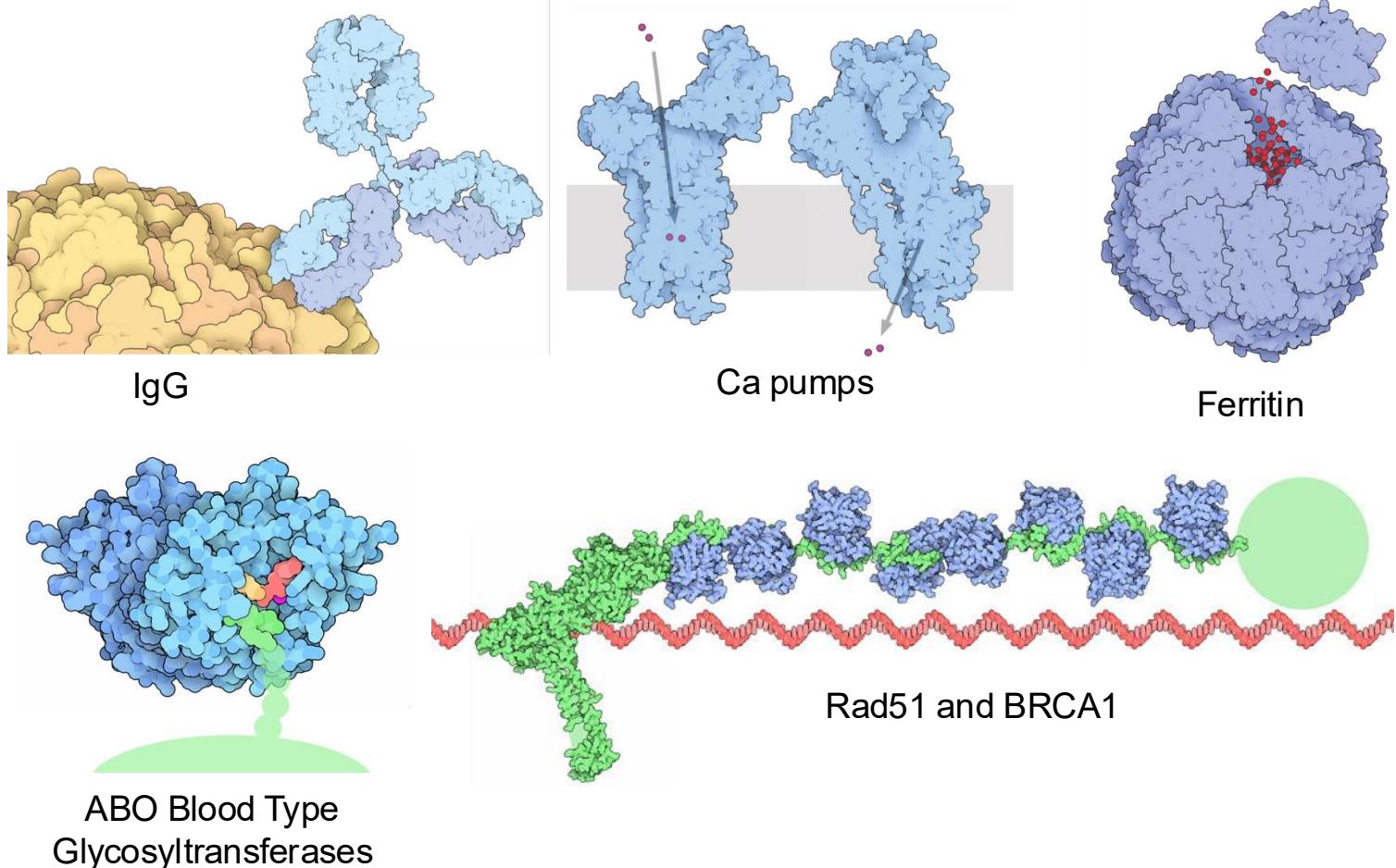


Visualizing macromolecular structures with PyMOL

Ha An Nguyen
Lima Lab

Structure informs function of proteins



- Amino acid sequence to full functional complex
 - Mutations
 - Drug design

How Do DRUGS Work?

Examples from the PDB archive

PROTEINS are tiny molecular machines that perform most of the tasks needed to keep cells alive. These machines are far too small to see, so you might imagine that it is impossible to affect their action. However, drugs can be used to turn proteins on or off. DRUGS are small molecules that bind to proteins and modify their action. Some very powerful drugs, such as antibiotics or anticancer drugs, are used to completely disable a critical molecular machine. These drugs can kill a bacterial or cancer cell. Other molecules, such as aspirin, gently block less-critical proteins for a few hours. With the use of these drugs, we can make changes inside our own cells, such as the blocking of pain signals. Many structures of drugs that bind to proteins have been determined by scientists. These atomic structures allow us to see how drugs work, and perhaps how to modify them to improve their action. A few examples are shown here. Some of these drugs, like penicillin, were discovered in nature. Other drugs, such as HIV protease inhibitors, were created by using the target protein structure to design new drug molecules. These structures of proteins and drugs, along with many others, can be explored at the RCSB Protein Data Bank (PDB).

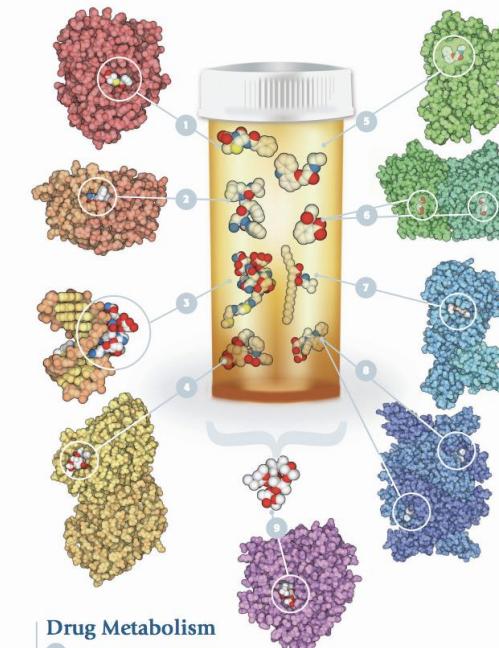
Antibiotics & Antivirals

1

2

Antibiotics and antiviral drugs are specific poisons. They need to kill pathogenic organisms like bacteria and viruses without poisoning the patient at the same time. Often, these drugs attack proteins that are unique to the targeted bacterium or virus and which are crucial for their survival or multiplication. For instance, penicillin attacks the enzyme that builds bacterial cell walls, and HIV protease inhibitors like saquinavir attack an enzyme that is needed for HIV maturation.

1. D-alanyl-D-alanine carboxypeptidase with penicillin (1pvc)
2. HIV protease with saquinavir (1hd)



Anticancer Chemotherapy

3

4

Cancer cells grow and multiply without control. Since these cells are still similar to normal cells, it is difficult to kill them selectively with drugs that can distinguish between the two. Many drugs currently used for cancer chemotherapy attack all growing cells, including cancer cells and normal cells. This causes many side effects of cancer chemotherapy, because the drugs attack rapidly-growing cells in hair follicles and the stomach. Two examples are shown here. Paclitaxel attacks DNA in actively growing cells, often cleaving the DNA chain and killing the cell. Paclitaxel (Taxol) binds to the microtubules, preventing the action of microtubules during cell division.

3. DNA with bleomycin (1msk)
4. Tubulin with taxol (1pft)

Drug Metabolism

5

You have probably noticed that when you take drugs, the effects gradually wear off in a few hours. Enzymes like cytochrome P450 continually search for drugs and destroy them. This is important because it protects us from poisonous molecules in our diet and in the environment, but it means that we have to take multiple doses of drugs when being treated for a disease.

5. Cytochrome P450 3A4 with erythromycin (2jod)

Molecular Mimics

6

7

Most drugs mimic the molecules that are normally processed by an enzyme or receptor protein. They bind tightly to the protein and block the site that usually performs a specific function. The protein usually binds to a protein chain, like the one shown at the left, and clips it into two pieces. Drugs used to treat HIV infection, like saquinavir shown here, are smaller than the protein chain but chemically similar. The drug binds in a similar position as the peptide, completely blocking the active site so the enzyme is unable to cleave the protein chain. (Image created with the Python Molecular Viewer—mgltools.scripps.edu)

Peptide bound (2nd) and drug bound (1hd) structures of HIV protease.

Suicide Inhibitors

8

9

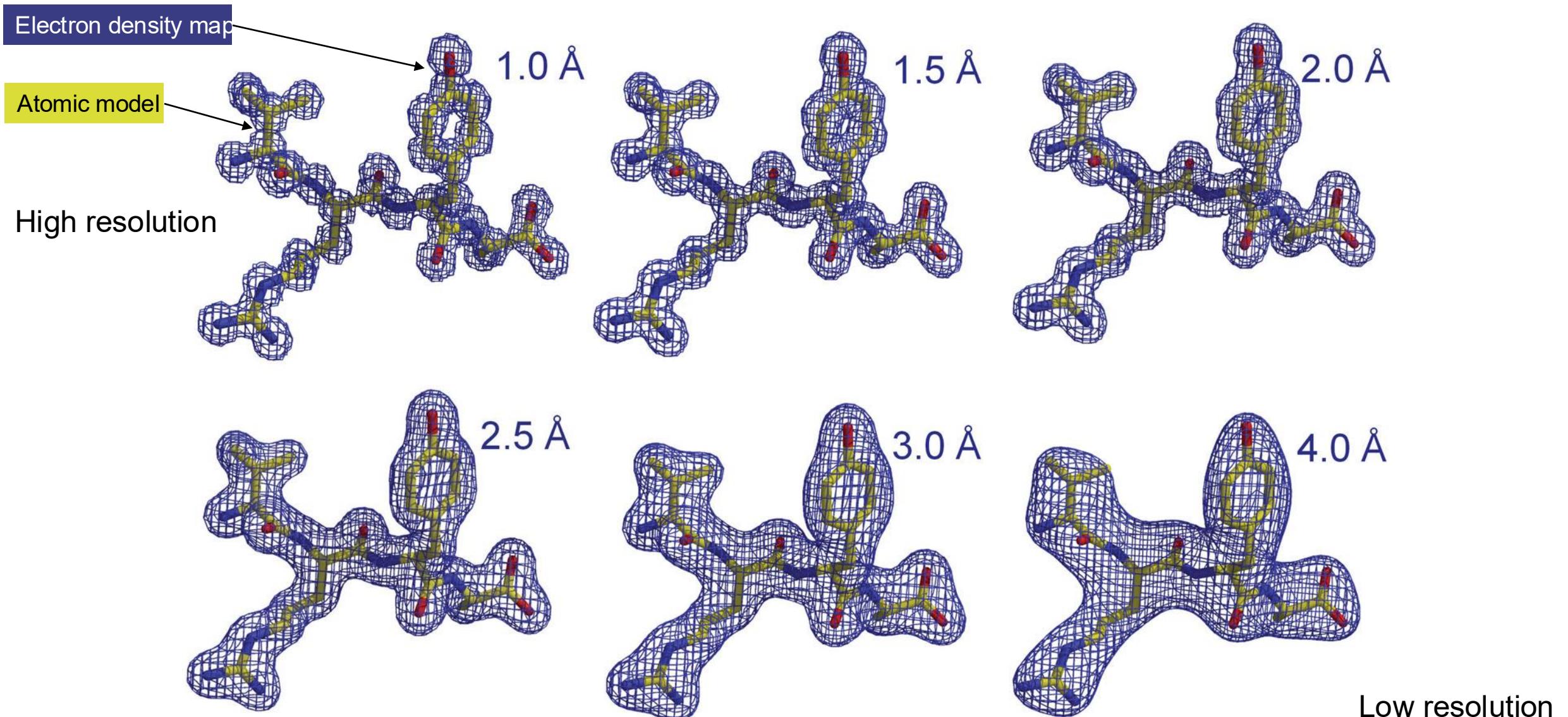
Some drugs are particularly effective because they form a chemical bond to the protein target (shown in turquoise), totally disabling it in that location. For instance, penicillin (shown with atomic colors) reacts with a serine amino acid in the bacterial enzyme, forming a new covalent bond to the enzyme. This completely blocks the enzyme, so the enzyme is unable to perform its role in cell wall synthesis. Another suicide inhibitor, aspirin (shown in #6), attaches an acetyl group to its target which blocks an inflammation pathway.

Penicillin bound structure of D-alanyl-D-alanine carboxypeptidase (PDB entry 1pvc)

How do we determine protein structures?

Method	Sample	Instrument	Data	Advantages	Limitations
X-ray crystallography	Crystals	Synchrotron	Diffraction pattern	<ul style="list-style-type: none"> • High-res data • Very common • No size restriction 	<ul style="list-style-type: none"> • Dependent on crystallization • Limited conformations
Cryo-electron microscopy	Sample vitrified on grids	Electron microscope	Micrographs	<ul style="list-style-type: none"> • Can capture states • Small amount of sample required • Fast sample prep 	<ul style="list-style-type: none"> • Size limit (smaller = harder) • Preferred orientation • Computationally intensive
Nuclear Magnetic Resonance Spectroscopy	Sample in solvent	NMR Spectrometer	NMR spectra	Only method for real-time, in-solution motion detection	<ul style="list-style-type: none"> • Size limit (larger = harder) • Need a lot of very pure sample

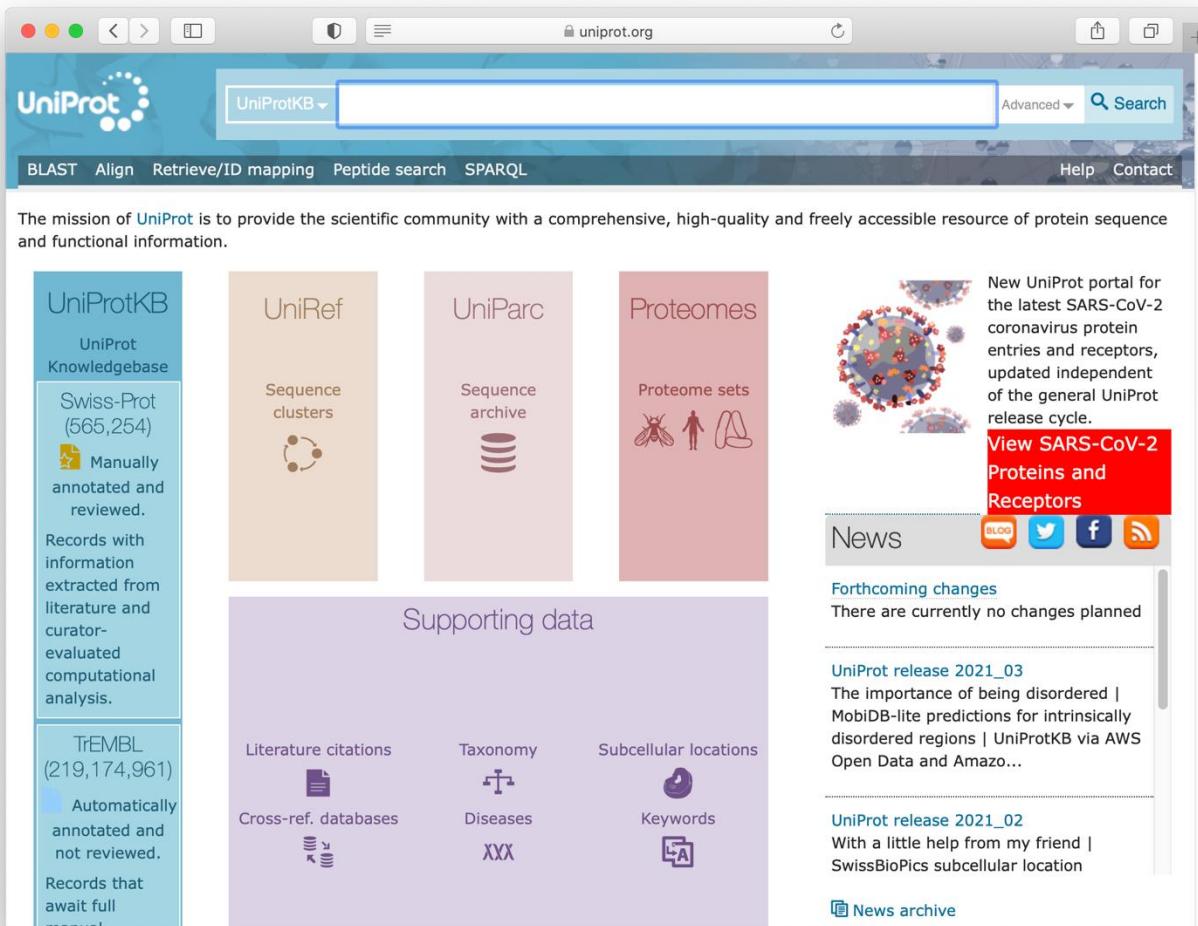
Resolution and information level



Where to find protein information

UniProt: Universal Protein Resource

www.uniprot.org



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

- UniProt Knowledgebase
- Swiss-Prot (565,254)
Manually annotated and reviewed.
- Records with information extracted from literature and curator-evaluated computational analysis.
- TrEMBL (219,174,961)
Automatically annotated and not reviewed.
- Records that await full manual

UniRef

- Sequence clusters

UniParc

- Sequence archive

Proteomes

- Proteome sets

Supporting data

- Literature citations
- Taxonomy
- Subcellular locations
- Cross-ref. databases
- Diseases
- Keywords

News

New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.

[View SARS-CoV-2 Proteins and Receptors](#)

Forthcoming changes

There are currently no changes planned

UniProt release 2021_03

The importance of being disordered | MobiDB-lite predictions for intrinsically disordered regions | UniProtKB via AWS Open Data and Amazo...

UniProt release 2021_02

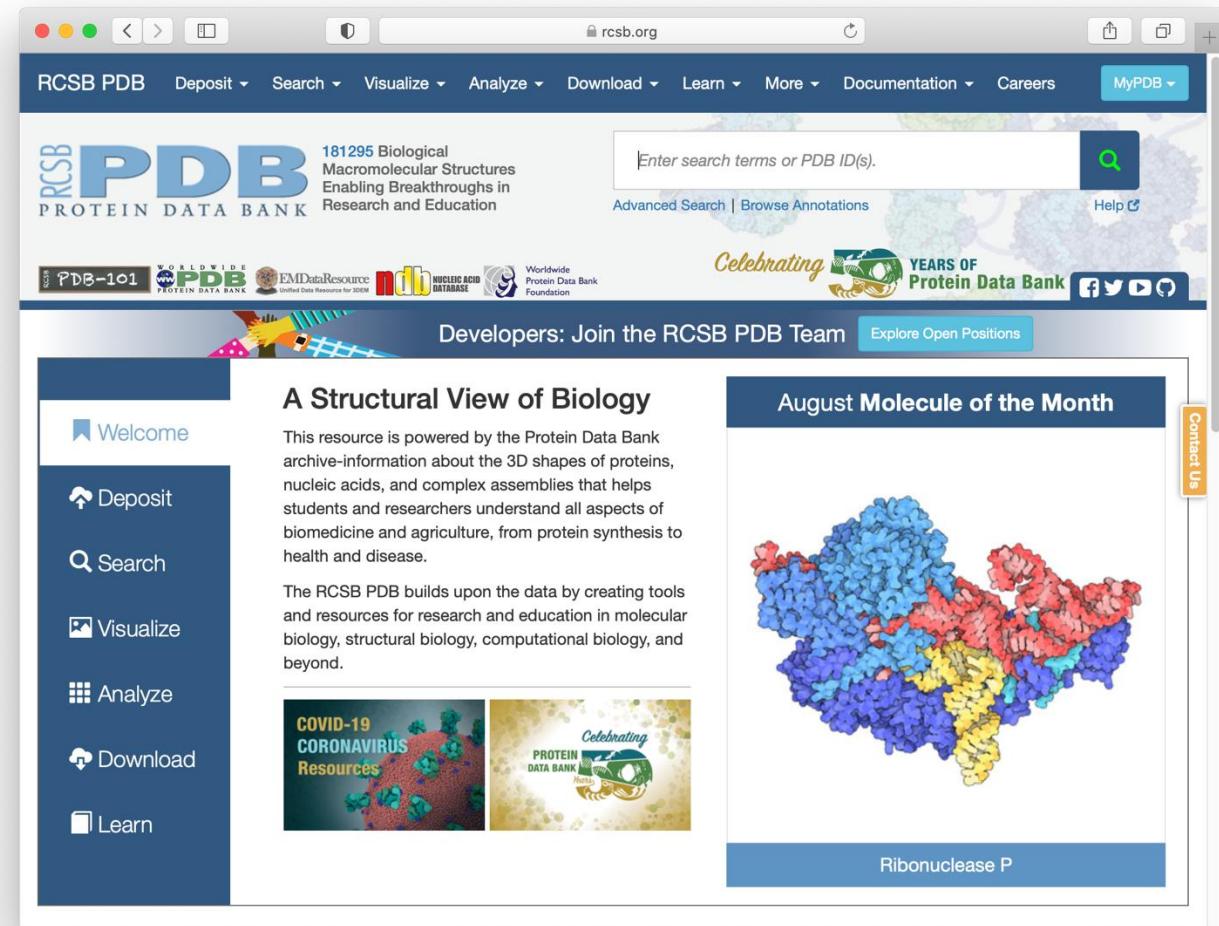
With a little help from my friend | SwissBioPics subcellular location

[News archive](#)

protein sequence and functional information

RCSB PDB: Protein DataBank

www.rscb.org



RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB

181295 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search terms or PDB ID(s).

Advanced Search | Browse Annotations Help

PDB PROTEIN DATA BANK

PDB-101 WORLDWIDE PROTEIN DATA BANK EMDDataResource NDB NUCLEIC ACID DATABASE Worldwide Protein Data Bank Foundation

Developers: Join the RCSB PDB Team Explore Open Positions

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

August Molecule of the Month

Ribonuclease P

central repository of biomolecular structures

AlphaFold

Database: <https://alphafold.ebi.ac.uk/>

models for almost every protein in Uniprot

Revolutionary tool for structural biology, but proteins are chaotic
post-translational modifications (PTMs), environment, ligands, oligomerization etc...

The image shows the homepage of the AlphaFold Protein Structure Database. The background is a dark blue gradient with a faint, stylized 3D protein structure model. The title 'AlphaFold Protein Structure Database' is prominently displayed in large, white, sans-serif font in the center. Below the title, a subtitle 'Developed by DeepMind and EMBL-EBI' is shown in a smaller, white, sans-serif font. At the top, a navigation bar includes links for 'EMBL-EBI home', 'Services', 'Research', 'Training', 'About us', and 'EMBL-EBI'. Below the navigation bar, a secondary navigation bar features links for 'Home', 'About', 'FAQs', 'Downloads', and 'API'. A search bar at the bottom left contains the placeholder text 'Search for protein, gene, UniProt accession or organism' and includes a 'BETA' button and a blue 'Search' button. Below the search bar, a 'Examples:' section lists several search terms: 'Free fatty acid receptor 2', 'At1g58602', 'Q5VSL9', and 'E. coli'. To the right of these examples is a link 'See search help' with a small arrow icon.

3D viewer ⓘ

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Sequence of AF-Q8N3Z6-F1 ⓘ Chain ⓘ 1: Zinc finger ... ⓘ A ⓘ

Residue	Sequence	Start	End
1	MMFGGYETIEAYEDDLYRDESSSEL	1	21
21	SVDEVEFQQLYSQIHYAQD	21	31
31	DDVIREEEHEEKNSGNSESSSS	31	41
41	KPNQQKLIVLS	41	51
51	DSEVIQLSDGSEVITLSD	51	61
61	EDSIYRCKGK	61	71
71	NVRVQAQENAHGLSS	71	81
81	LSQSNELVDKRC	81	91
91	KSDIEKPKSEERSGVIRE	91	101
101	MIE	101	111
111	VSSSEEEESTISEGDNV	111	121
121	ESWMLLGCEV	121	131
131	DDDKDILNLVGC	131	141
141	ENSVT	141	151
151	EGEDGINWSISDKDIE	151	161
161	QAQIANNRTPGRWTQRY	161	171
171	VSS	171	181
181	SEESTISEGDNV	181	191
191	ESWMLLGCEV	191	201
201	DDDKDILNLVGC	201	211
211	ENSVT	211	221
221	EGEDGINWSISDKDIE	221	231
231	QAQIANNRTPGRWTQRY	231	241
241	VSS	241	251
251	SEESTISEGDNV	251	261
261	ESWMLLGCEV	261	271
271	DDDKDILNLVGC	271	281
281	ENSVT	281	291
291	EGEDGINWSISDKDIE	291	301
301	QAQIANNRTPGRWTQRY	301	311
311	VSS	311	321
321	SEESTISEGDNV	321	331
331	ESWMLLGCEV	331	341
341	DDDKDILNLVGC	341	351
351	ENSVT	351	361
361	EGEDGINWSISDKDIE	361	371
371	QAQIANNRTPGRWTQRY	371	381
381	VSS	381	391
391	SEESTISEGDNV	391	401
401	ESWMLLGCEV	401	411
411	DDDKDILNLVGC	411	421
421	ENSVT	421	431
431	EGEDGINWSISDKDIE	431	441
441	QAQIANNRTPGRWTQRY	441	451
451	VSS	451	461
461	SEESTISEGDNV	461	471
471	ESWMLLGCEV	471	481
481	DDDKDILNLVGC	481	491
491	ENSVT	491	501
501	EGEDGINWSISDKDIE	501	511
511	QAQIANNRTPGRWTQRY	511	521
521	VSS	521	531
531	SEESTISEGDNV	531	541
541	ESWMLLGCEV	541	551
551	DDDKDILNLVGC	551	561
561	ENSVT	561	571
571	EGEDGINWSISDKDIE	571	581
581	QAQIANNRTPGRWTQRY	581	591
591	VSS	591	601
601	SEESTISEGDNV	601	611
611	ESWMLLGCEV	611	621
621	DDDKDILNLVGC	621	631
631	ENSVT	631	641
641	EGEDGINWSISDKDIE	641	651
651	QAQIANNRTPGRWTQRY	651	661
661	VSS	661	671
671	SEESTISEGDNV	671	681
681	ESWMLLGCEV	681	691
691	DDDKDILNLVGC	691	701
701	ENSVT	701	711
711	EGEDGINWSISDKDIE	711	721
721	QAQIANNRTPGRWTQRY	721	731
731	VSS	731	741
741	SEESTISEGDNV	741	751
751	ESWMLLGCEV	751	761
761	DDDKDILNLVGC	761	771
771	ENSVT	771	781
781	EGEDGINWSISDKDIE	781	791
791	QAQIANNRTPGRWTQRY	791	801
801	VSS	801	811
811	SEESTISEGDNV	811	821
821	ESWMLLGCEV	821	831
831	DDDKDILNLVGC	831	841
841	ENSVT	841	851
851	EGEDGINWSISDKDIE	851	861
861	QAQIANNRTPGRWTQRY	861	871
871	VSS	871	881
881	SEESTISEGDNV	881	891
891	ESWMLLGCEV	891	901
901	DDDKDILNLVGC	901	911
911	ENSVT	911	921
921	EGEDGINWSISDKDIE	921	931
931	QAQIANNRTPGRWTQRY	931	941
941	VSS	941	951
951	SEESTISEGDNV	951	961
961	ESWMLLGCEV	961	971
971	DDDKDILNLVGC	971	981
981	ENSVT	981	991
991	EGEDGINWSISDKDIE	991	1001
1001	QAQIANNRTPGRWTQRY	1001	1011
1011	VSS	1011	1021
1021	SEESTISEGDNV	1021	1031
1031	ESWMLLGCEV	1031	1041
1041	DDDKDILNLVGC	1041	1051
1051	ENSVT	1051	1061
1061	EGEDGINWSISDKDIE	1061	1071
1071	QAQIANNRTPGRWTQRY	1071	1081
1081	VSS	1081	1091
1091	SEESTISEGDNV	1091	1101
1101	ESWMLLGCEV	1101	1111
1111	DDDKDILNLVGC	1111	1121
1121	ENSVT	1121	1131
1131	EGEDGINWSISDKDIE	1131	1141
1141	QAQIANNRTPGRWTQRY	1141	1151
1151	VSS	1151	1161
1161	SEESTISEGDNV	1161	1171
1171	ESWMLLGCEV	1171	1181
1181	DDDKDILNLVGC	1181	1191
1191	ENSVT	1191	1201
1201	EGEDGINWSISDKDIE	1201	1211
1211	QAQIANNRTPGRWTQRY	1211	1221
1221	VSS	1221	1231
1231	SEESTISEGDNV	1231	1241
1241	ESWMLLGCEV	1241	1251
1251	DDDKDILNLVGC	1251	1261
1261	ENSVT	1261	1271
1271	EGEDGINWSISDKDIE	1271	1281
1281	QAQIANNRTPGRWTQRY	1281	1291
1291	VSS	1291	1301
1301	SEESTISEGDNV	1301	1311
1311	ESWMLLGCEV	1311	1321
1321	DDDKDILNLVGC	1321	1331
1331	ENSVT	1331	1341
1341	EGEDGINWSISDKDIE	1341	1351
1351	QAQIANNRTPGRWTQRY	1351	1361
1361	VSS	1361	1371
1371	SEESTISEGDNV	1371	1381
1381	ESWMLLGCEV	1381	1391
1391	DDDKDILNLVGC	1391	1401
1401	ENSVT	1401	1411
1411	EGEDGINWSISDKDIE	1411	1421
1421	QAQIANNRTPGRWTQRY	1421	1431
1431	VSS	1431	1441
1441	SEESTISEGDNV	1441	1451
1451	ESWMLLGCEV	1451	1461
1461	DDDKDILNLVGC	1461	1471
1471	ENSVT	1471	1481
1481	EGEDGINWSISDKDIE	1481	1491
1491	QAQIANNRTPGRWTQRY	1491	1501
1501	VSS	1501	1511
1511	SEESTISEGDNV	1511	1521
1521	ESWMLLGCEV	1521	1531
1531	DDDKDILNLVGC	1531	1541
1541	ENSVT	1541	1551
1551	EGEDGINWSISDKDIE	1551	1561
1561	QAQIANNRTPGRWTQRY	1561	1571
1571	VSS	1571	1581
1581	SEESTISEGDNV	1581	1591
1591	ESWMLLGCEV	1591	1601
1601	DDDKDILNLVGC	1601	1611
1611	ENSVT	1611	1621
1621	EGEDGINWSISDKDIE	1621	1631
1631	QAQIANNRTPGRWTQRY	1631	1641
1641	VSS	1641	1651
1651	SEESTISEGDNV	1651	1661
1661	ESWMLLGCEV	1661	1671
1671	DDDKDILNLVGC	1671	1681
1681	ENSVT	1681	1691
1691	EGEDGINWSISDKDIE	1691	1701
1701	QAQIANNRTPGRWTQRY	1701	1711
1711	VSS	1711	1721
1721	SEESTISEGDNV	1721	1731
1731	ESWMLLGCEV	1731	1741
1741	DDDKDILNLVGC	1741	1751
1751	ENSVT	1751	1761
1761	EGEDGINWSISDKDIE	1761	1771
1771	QAQIANNRTPGRWTQRY	1771	1781
1781	VSS	1781	1791
1791	SEESTISEGDNV	1791	1801
1801	ESWMLLGCEV	1801	1811
1811	DDDKDILNLVGC	1811	1821
1821	ENSVT	1821	1831
1831	EGEDGINWSISDKDIE	1831	1841
1841	QAQIANNRTPGRWTQRY	1841	1851
1851	VSS	1851	1861
1861	SEESTISEGDNV	1861	1871
1871	ESWMLLGCEV	1871	1881
1881	DDDKDILNLVGC	1881	1891
1891	ENSVT	1891	1901
1901	EGEDGINWSISDKDIE	1901	1911
1911	QAQIANNRTPGRWTQRY	1911	1921
1921	VSS	1921	1931
1931	SEESTISEGDNV	1931	1941
1941	ESWMLLGCEV	1941	1951
1951	DDDKDILNLVGC	1951	1961
1961	ENSVT	1961	1971
1971	EGEDGINWSISDKDIE	1971	1981
1981	QAQIANNRTPGRWTQRY	1981	1991
1991	VSS	1991	2001
2001	SEESTISEGDNV	2001	2011
2011	ESWMLLGCEV	2011	2021
2021	DDDKDILNLVGC	2021	2031
2031	ENSVT	2031	2041
2041	EGEDGINWSISDKDIE	2041	2051
2051	QAQIANNRTPGRWTQRY	2051	2061
2061	VSS	2061	2071
2071	SEESTISEGDNV	2071	2081
2081	ESWMLLGCEV	2081	2091
2091	DDDKDILNLVGC	2091	2101
2101	ENSVT	2101	2111
2111	EGEDGINWSISDKDIE	2111	2121
2121	QAQIANNRTPGRWTQRY	2121	2131
2131	VSS	2131	2141
2141	SEESTISEGDNV	2141	2151
2151	ESWMLLGCEV	2151	2161
2161	DDDKDILNLVGC	2161	2171
2171	ENSVT	2171	2181
2181	EGEDGINWSISDKDIE	2181	2191
2191	QAQIANNRTPGRWTQRY	2191	2201
2201	VSS	2201	2211
2211	SEESTISEGDNV	2211	2221
2221	ESWMLLGCEV	2221	2231
2231	DDDKDILNLVGC	2231	2241
2241	ENSVT	2241	2251
2251	EGEDGINWSISDKDIE	2251	2261
2261	QAQIANNRTPGRWTQRY	2261	2271
2271	VSS	2271	2281
2281	SEESTISEGDNV	2281	2291
2291	ESWMLLGCEV	2291	2301
2301	DDDKDILNLVGC	2301	2311
2311	ENSVT	2311	2321
2321	EGEDGINWSISDKDIE	2321	2331
2331	QAQIANNRTPGRWTQRY	2331	2341
2341	VSS	2341	2351
2351	SEESTISEGDNV	2351	2361
2361	ESWMLLGCEV	2361	2371
2371	DDDKDILNLVGC	2371	2381
2381	ENSVT	2381	2391
2391	EGEDGINWSISDKDIE	2391	2401
2401	QAQIANNRTPGRWTQRY	2401	2411
2411	VSS	2411	2421
2421	SEESTISEGDNV	2421	2431
2431	ESWMLLGCEV	2431	2441
2441	DDDKDILNLVGC	2441	2451
2451	ENSVT	2451	2461
2461	EGEDGINWSISDKDIE	2461	2471
2471	QAQIANNRTPGRWTQRY	2471	2481
2481	VSS	2481	2491
2491	SEESTISEGDNV	2491	2501
2501	ESWMLLGCEV	2501	2511
2511	DDDKDILNLVGC	2511	2521
2521	ENSVT	2521	2531
2531	EGEDGINWSISDKDIE	2531	2541
2541	QAQIANNRTPGRWTQRY	2541	2551
2551	VSS	2551	2561
2561	SEESTISEGDNV	2561	2571
2571	ESWMLLGCEV	2571	2581
2581	DDDKDILNLVGC	2581	2591
2591	ENSVT	2591	2601
2601	EGEDGINWSISDKDIE	2601	2611
2611	QAQIANNRTPGRWTQRY	2611	2621
2621	VSS	2621	2631
2631	SEESTISEGDNV	2631	2641
2641	ESWMLLGCEV	2641	2651
2651	DDDKDILNLVGC	2651	2661
2661	ENSVT	2661	2671
2671	EGEDGINWSISDKDIE	2671	2681
2681	QAQIANNRTPGRWTQRY	2681	2691
2691	VSS	2691	2701
2701	SEESTISEGDNV	2701	2711
2711	ESWMLLGCEV	2711	2721
2721	DDDKDILNLVGC	2721	2731
2731	ENSVT	2731	2741
2741	EGEDGINWSISDKDIE	2741	2751
2751	QAQIANNRTPGRWTQRY	2751	2761
2761	VSS	2761	2771
2771	SEESTISEGDNV	2771	2781
2781	ESWMLLGCEV	2781	2791
2791	DDDKDILNLVGC	2791	2801
2801	ENSVT	2801	2811
2811	EGEDGINWSISDKDIE	2811	2821
2821	QAQIANNRTPGRWTQRY	2821	2831
2831	VSS	2831	2841
2841	SEESTISEGDNV	2841	2851
2851	ESWMLLGCEV	2851	2861
2861	DDDKDILNLVGC	2861	2871
2871	ENSVT	2871	2881
2881	EGEDGINWSISDKDIE	2881	2891
2891	QAQIANNRTPGRWTQRY	2891	2901
2901	VSS	2901	2911
2911	SEESTISEGDNV	2911	2921
2921	ESWMLLGCEV	2921	2931
2931	DDDKDILNLVGC	2931	2941
2941	ENSVT	2941	2951
2951	EGEDGINWSISDKDIE	2951	2961
2961	QAQIANNRTPGRWTQRY	2961	2971
2971	VSS	2971	2981
2981	SEESTISEGDNV	2981	2991</

<https://alphafold.ebi.ac.uk/entry/Q8N3Z6>

Useful resources

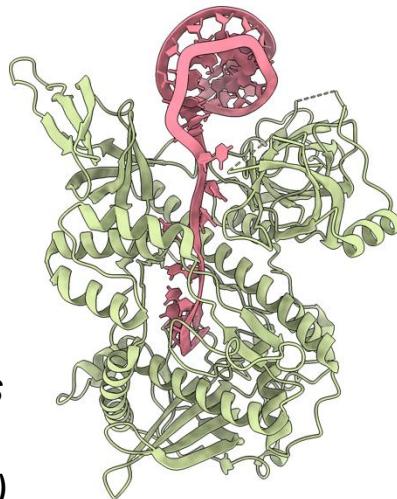
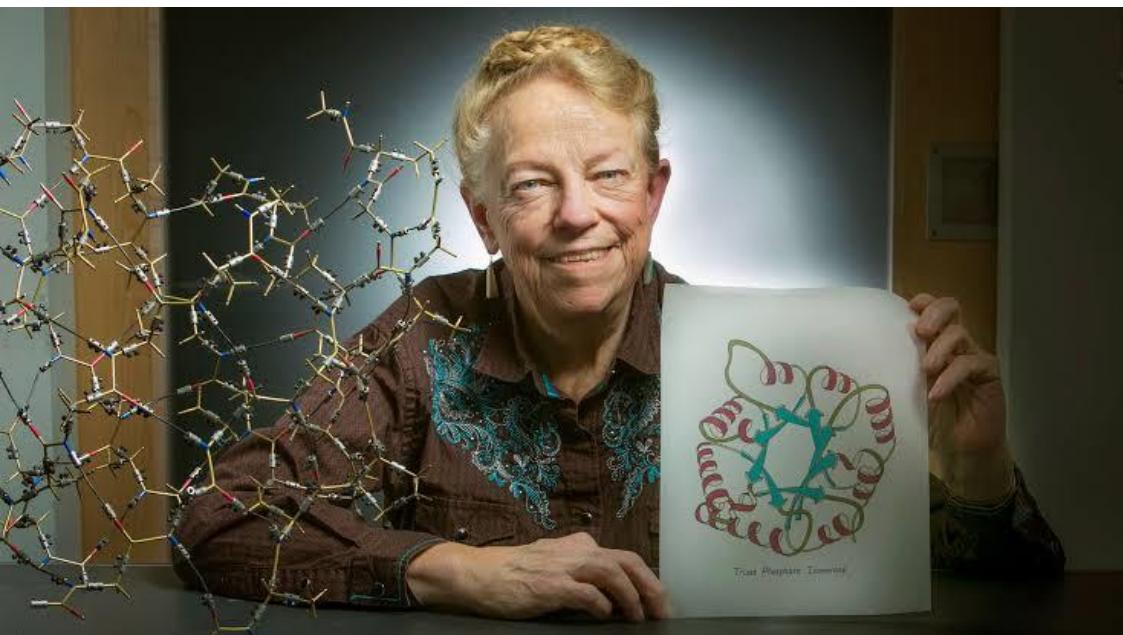
Name	Link	Purpose
UniProt	https://www.uniprot.org/	Protein information database
Protein Data Bank (PDB)	https://www.rcsb.org/	Protein structure repository
PyMOL	https://pymol.org/	Download PyMOL
PyMOL wiki	www.pymolwiki.org	Information on PyMOL such as examples of settings
ConSurf	http://consurf.tau.ac.il	Evolutionary conservation of amino/nucleic acid sequences
PISA	www.ebi.ac.uk/pdbe/pisa/	Protein interface and assemblies
Dali	http://ekhidna2.biocenter.helsinki.fi/dali/	Search PDB for similar structures
EMDB	https://www.ebi.ac.uk/emdb/	Repository for EM/ET data
AlphaFold database	https://alphafold.ebi.ac.uk/	Database of AlphaFold predictions
ChimeraX	https://www.cgl.ucsf.edu/chimerax/	Another visualization software, works well with EM maps

overview of structure files

model

- the coordinates of the atoms
- file types: .pdb, .cif

Note: Cartoon (ribbon) representation is an artistic interpretation of structures invented by Dr. Jane Richardson (Duke)



map

- the experimental data with which you can place atoms into position (x-ray or EM)
- can also be called structure factors or map coefficients
- file types: .map, .mrc, .ccp4, .mtz, .cif



emd_27829.map	
1	2c01 0000 2c01 0000 2c01 0000 0200 0000
2	0000 0000 0000 0000 0000 0000 2c01 0000
3	2c01 0000 2c01 0000 0000 4943 0000 4943
4	0000 4943 0000 b442 0000 b442 0000 b442
5	0100 0000 0200 0000 0300 0000 9a89 00bf
6	ecc7 3c3f 34c1 c1b8 0100 0000 0000 0000
7	0000 0000 0000 0000 0000 0000 0000 0000
8	0000 0000 0000 0000 0000 0000 0000 0000
9	0000 0000 0000 0000 0000 0000 0000 0000
10	0000 0000 0000 0000 0000 0000 0000 0000
11	0000 0000 0000 0000 0000 0000 0000 0000
12	0000 0000 0000 0000 0000 0000 0000 0000
13	0000 0000 0000 0000 0000 0000 0000 0000
14	4d41 5020 4441 0000 e0eb 913c 0100 0000
15	3a3a 3a3a 454d 4441 5441 4241 4e4b 2e6f
16	7267 3a3a 3a3a 454d 442d 3237 3832 393a
17	3a3a 3a20 2020 2020 2020 2020 2020 2020
18	2020 2020 2020 2020 2020 2020 2020 2020
19	2020 2020 2020 2020 2020 2020 2020 2020
20	2020 2020 2020 2020 2020 2020 2020 2020
21	2020 2020 2020 2020 2020 2020 2020 2020
22	2020 2020 2020 2020 2020 2020 2020 2020
23	2020 2020 2020 2020 2020 2020 2020 2020
24	2020 2020 2020 2020 2020 2020 2020 2020

The primary information stored in the PDB archive consists of coordinate files that list of atoms and their 3D location in space stored as plain text (ASCII) files
Either as .pdb or mmcif (.cif) format

1f88.pdb

ATOM	16	ND2	ASN	A	2	46.634	-7.296	-23.187	1.00	59.21	N
ATOM	17	N	GLY	A	3	44.982	-2.466	-23.598	1.00	49.38	N
ATOM	18	CA	GLY	A	3	45.060	-1.050	-23.332	1.00	48.15	C
ATOM	19	C	GLY	A	3	44.975	-0.193	-24.576	1.00	47.25	C
ATOM	20	O	GLY	A	3	45.481	-0.545	-25.645	1.00	48.00	O
ATOM	21	N	THR	A	4	44.321	0.946	-24.419	1.00	45.08	N
ATOM	22	CA	THR	A	4	44.165	1.892	-25.494	1.00	44.41	C
ATOM	23	C	THR	A	4	45.409	2.775	-25.508	1.00	44.61	C
ATOM	24	O	THR	A	4	45.662	3.492	-24.545	1.00	44.85	O
ATOM	25	CB	THR	A	4	42.913	2.754	-25.261	1.00	43.60	C
ATOM	26	OG1	THR	A	4	41.781	1.901	-25.083	1.00	44.84	O
ATOM	27	CG2	THR	A	4	42.649	3.650	-26.435	1.00	43.69	C
ATOM	28	N	GLU	A	5	46.225	2.647	-26.555	1.00	45.88	N
ATOM	29	CA	GLU	A	5	47.438	3.458	-26.710	1.00	47.26	C
ATOM	30	C	GLU	A	5	47.146	4.651	-27.612	1.00	48.31	C
ATOM	31	O	GLU	A	5	46.687	4.480	-28.742	1.00	49.86	O
ATOM	32	CB	GLU	A	5	48.599	2.656	-27.325	1.00	46.89	C
ATOM	33	CG	GLU	A	5	49.822	3.538	-27.693	1.00	49.26	C
ATOM	34	CD	GLU	A	5	51.044	2.764	-28.230	1.00	52.49	C
ATOM	35	OE1	GLU	A	5	51.016	1.511	-28.240	1.00	55.51	O
ATOM	36	OE2	GLU	A	5	52.047	3.411	-28.638	1.00	51.62	O
ATOM	37	N	GLY	A	6	47.387	5.855	-27.102	1.00	48.23	N

Header

Atom coordinates

8e29.pdb x

```

1 HEADER RNA BINDING PROTEIN/RNA 14-AUG-22 8E29
2 TITLE HUMAN DIS3L2 IN COMPLEX WITH HAIRPIN C-U12
3 COMPD MOL_ID: 1;
4 COMPD 2 MOLECULE: DIS3-LIKE EXONUCLEASE 2;
5 COMPD 3 CHAIN: A;
6 COMPD 4 SYNONYM: HDIS3L2;
7 COMPD 5 EC: 3.1.13.-;
8 COMPD 6 ENGINEERED: YES;
9 COMPD 7 MOL_ID: 2;
10 COMPD 8 MOLECULE: RNA HAIRPIN C-U12;
11 COMPD 9 CHAIN: B;
12 COMPD 10 ENGINEERED: YES
13 SOURCE MOL_ID: 1;
14 SOURCE 2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
15 SOURCE 3 ORGANISM_COMMON: HUMAN;
16 SOURCE 4 ORGANISM_TAXID: 9606;
17 SOURCE 5 GENE: DIS3L2, FAM6A;
18 SOURCE 6 EXPRESSION_SYSTEM: SPODOPTERA FRUGIPERDA;
19 SOURCE 7 EXPRESSION_SYSTEM_TAXID: 7108;
20 SOURCE 8 MOL_ID: 2;
21 SOURCE 9 SYNTHETIC: YES;
22 SOURCE 10 ORGANISM_SCIENTIFIC: SYNTHETIC CONSTRUCT;
23 SOURCE 11 ORGANISM_TAXID: 32630
24 KEYWDS 3'-5' EXONUCLEASE, DS-RNA BOUND EXONUCLEASE, HUMAN EXONUCLEASE, RNA
25 KEYWDS 2 BINDING PROTEIN-RNA COMPLEX
26 EXPTA ELECTRON MICROSCOPY

```

8e29.pdb x

```

427 SEQRES 61 A 885 ALA PHE ASP VAL LEU VAL LEU ARG TYR GLY VAL GLN LYS
428 SEQRES 62 A 885 ARG ILE TYR CYS ASN ALA LEU ALA ARG SER HIS HIS
429 SEQRES 63 A 885 PHE GLN LYS VAL GLY LYS LYS PRO GLU LEU THR LEU VAL
430 SEQRES 64 A 885 TRP GLU PRO GLU ASP MET GLU GLN GLU PRO ALA GLN GLN
431 SEQRES 65 A 885 VAL ILE THR ILE PHE SER LEU VAL GLU VAL VAL LEU GLN
432 SEQRES 66 A 885 ALA GLU SER THR ALA LEU LYS TYR SER ALA ILE LEU LYS
433 SEQRES 67 A 885 ARG PRO GLY THR GLN GLY HIS LEU GLY PRO GLU LYS GLU
434 SEQRES 68 A 885 GLU GLU GLU SER ASP GLY GLU PRO GLU ASP SER SER THR
435 SEQRES 69 A 885 SER
436 SEQRES 1 B 33 U C G G C G C C U U U C G
437 SEQRES 2 B 33 A G G C G C C G U U U U U
438 SEQRES 3 B 33 U U U U U U U
439 HELIX 1 AA1 SER A 55 GLY A 66 1 12
440 HELIX 2 AA2 ILE A 329 GLY A 341 1 13
441 HELIX 3 AA3 SER A 348 GLU A 353 1 6
442 HELIX 4 AA4 PRO A 364 PHE A 368 5 5
443 HELIX 5 AA5 SER A 421 ARG A 430 1 10
444 HELIX 6 AA6 PRO A 445 GLU A 451 1 7
445 HELIX 7 AA7 GLU A 495 GLU A 502 1 8
446 HELIX 8 AA8 PRO A 509 LEU A 513 5 5
447 HELIX 9 AA9 SER A 521 GLY A 547 1 27
448 HELIX 10 AB1 GLU A 577 PHE A 601 1 25
449 HELIX 11 AB2 GLN A 614 GLN A 627 1 14
450 HELIX 12 AB3 ALA A 637 GLY A 649 1 13
451 HELIX 13 AB4 ASP A 651 SER A 666 1 16
452 HELIX 14 AB5 PHE A 706 LEU A 719 1 14
453 HELIX 15 AB6 ALA A 728 SER A 765 1 38
454 SHEET 1 AA1 6 ILE A 69 VAL A 72 0
455 SHEET 2 AA1 6 LEU A 110 LEU A 115 -1 0 VAL A 113 N ILE A 69
456 SHEET 3 AA1 6 ARG A 231 GLU A 240 -1 0 TYR A 237 N VAL A 112
457 SHEET 4 AA1 6 ILE A 94 ASP A 97 1 N PHE A 95 0 ARG A 231
458 SHEET 5 AA1 6 GLU A 82 ILE A 85 -1 N ILE A 85 0 ILE A 94
459 SHEET 6 AA1 6 ARG A 74 ILE A 75 -1 N ARG A 74 0 PHE A 84
460 SHEET 1 AA2 6 ALA A 246 LYS A 252 0
461 SHEET 2 AA2 6 TYR A 265 PRO A 270 -1 0 LEU A 267 N LYS A 252
462 SHEET 3 AA2 6 ILE A 278 PRO A 281 -1 0 VAL A 280 N ALA A 266
463 SHEET 4 AA2 6 ALA A 315 GLY A 324 1 0 ALA A 315 N TYR A 279
464 SHEET 5 AA2 6 LEU A 300 ASP A 307 -1 N ILE A 302 0 ALA A 320
465 SHEET 6 AA2 6 ALA A 246 LYS A 252 -1 N ALA A 247 0 CYS A 303

```

what's in a pdb file

atom #, atom name, residue name, chain id, residue number, x/y/z
coordinates of the atom, occupancy, bfactor, atom element

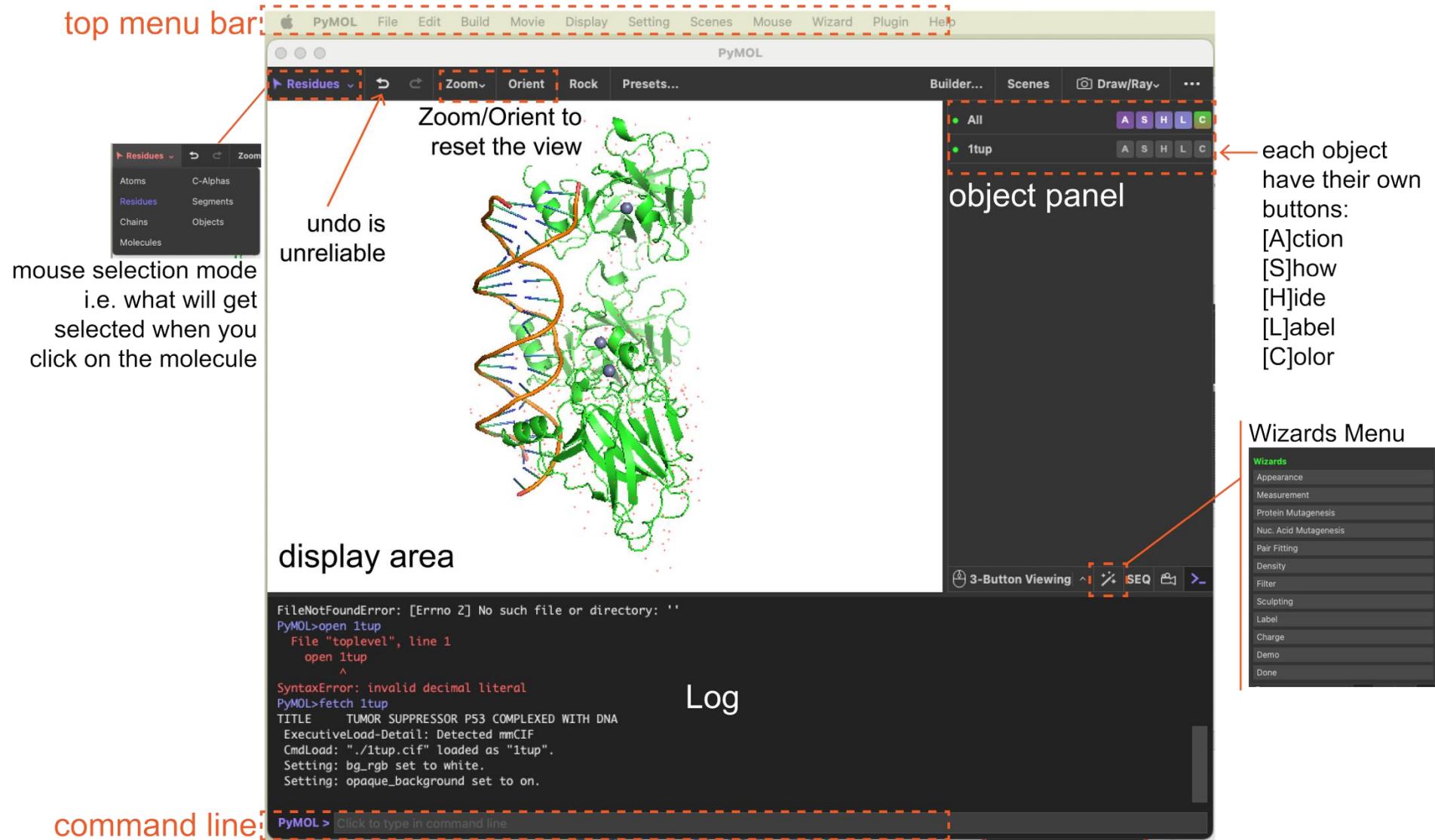
513	ATOM	20	N	GLU	A	51	87.388	80.367	121.779	1.00	85.42	N
514	ATOM	21	CA	GLU	A	51	86.228	79.551	121.454	1.00	85.42	C
515	ATOM	22	C	GLU	A	51	86.500	78.738	120.196	1.00	85.42	C
516	ATOM	23	O	GLU	A	51	87.163	79.215	119.271	1.00	85.42	O
517	ATOM	24	CB	GLU	A	51	84.981	80.405	121.263	1.00	85.42	C
518	ATOM	25	CG	GLU	A	51	83.734	79.732	121.773	1.00	85.42	C
519	ATOM	26	CD	GLU	A	51	82.477	80.461	121.378	1.00	85.42	C
520	ATOM	27	OE1	GLU	A	51	82.293	80.702	120.169	1.00	85.42	O
521	ATOM	28	OE2	GLU	A	51	81.670	80.783	122.274	1.00	85.42	O
522	ATOM	29	N	THR	A	52	85.998	77.507	120.177	1.00	86.60	N
523	ATOM	30	CA	THR	A	52	86.234	76.597	119.068	1.00	86.60	C
524	ATOM	31	C	THR	A	52	85.477	77.029	117.819	1.00	86.60	C
525	ATOM	32	O	THR	A	52	84.455	77.710	117.882	1.00	86.60	O
526	ATOM	33	CB	THR	A	52	85.797	75.186	119.437	1.00	86.60	C
527	ATOM	34	OG1	THR	A	52	84.375	75.100	119.320	1.00	86.60	O
528	ATOM	35	CG2	THR	A	52	86.189	74.865	120.861	1.00	86.60	C
529	ATOM	36	N	TYR	A	53	85.988	76.602	116.670	1.00	86.91	N
530	ATOM	37	CA	TYR	A	53	85.354	76.896	115.395	1.00	86.91	C
531	ATOM	38	C	TYR	A	53	84.110	76.027	115.212	1.00	86.91	C
532	ATOM	39	O	TYR	A	53	83.850	75.102	115.982	1.00	86.91	O
533	ATOM	40	CB	TYR	A	53	86.329	76.677	114.245	1.00	86.91	C
534	ATOM	41	CG	TYR	A	53	87.400	77.727	114.118	1.00	86.91	C
535	ATOM	42	CD1	TYR	A	53	88.429	77.574	113.214	1.00	86.91	C
536	ATOM	43	CD2	TYR	A	53	87.373	78.876	114.883	1.00	86.91	C
537	ATOM	44	CE1	TYR	A	53	89.401	78.522	113.088	1.00	86.91	C
538	ATOM	45	CE2	TYR	A	53	88.344	79.831	114.759	1.00	86.91	C
539	ATOM	46	CZ	TYR	A	53	89.350	79.648	113.862	1.00	86.91	C
540	ATOM	47	OH	TYR	A	53	90.316	80.599	113.736	1.00	86.91	O
541	ATOM	48	N	MET	A	54	83.333	76.325	114.178	1.00	98.72	N
542	ATOM	49	CA	MET	A	54	82.160	75.527	113.852	1.00	98.72	C
543	ATOM	50	C	MET	A	54	82.382	74.709	112.587	1.00	98.72	C
544	ATOM	51	O	MET	A	54	83.355	74.893	111.854	1.00	98.72	O
545	ATOM	52	CB	MET	A	54	80.920	76.405	113.682	1.00	98.72	C
546	ATOM	53	CG	MET	A	54	80.567	77.235	114.879	1.00	98.72	C
547	ATOM	54	SD	MET	A	54	78.941	77.984	114.674	1.00	98.72	S
548	ATOM	55	CE	MET	A	54	79.049	78.572	112.987	1.00	98.72	C
549	ATOM	56	M	SER	A	55	81.452	72.786	112.252	1.00	105.56	N

Why and how to visualize protein structures?

- To observe and analyze a structure of interest.
- To make images or movies to illustrate certain properties of the structure of interest.
 - There are numerous programs and websites to visualize macromolecular structures. PyMOL and UCSF Chimera / ChimeraX are most used.

PyMOL tutorial

PyMOL interface



In this ppt,
[square brackets]
indicates these
buttons

1. Loading structures

File > Get PDB > 1tup

OR

1. Go to PDB website: www.rcsb.org

2. Find molecule of interest using search bar

Search for: **p53 complexed with dna**

3. Download Files > PDB Format > 1tup.pdb file

and open it in PyMOL

using File>Open

OR

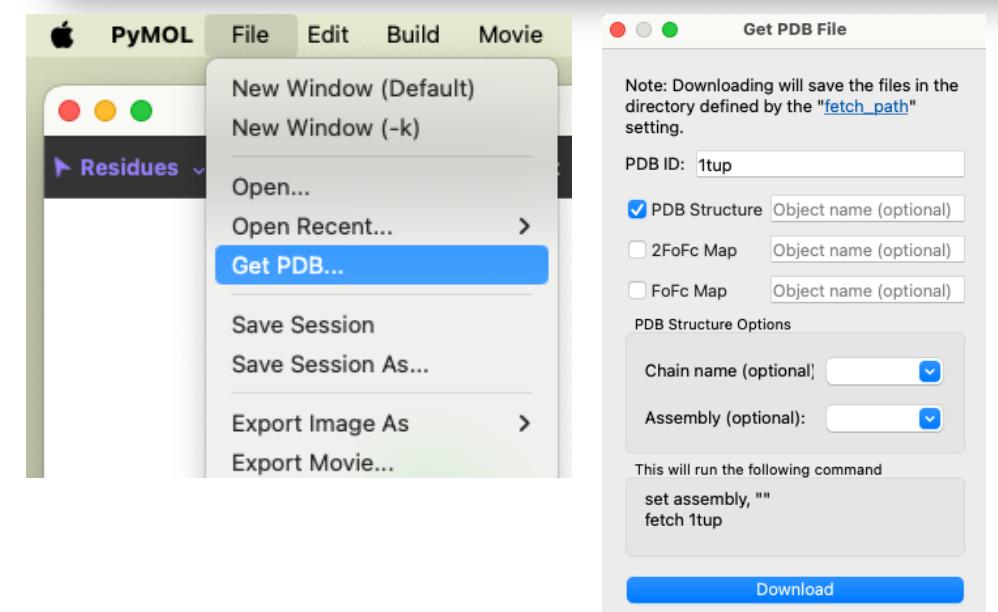
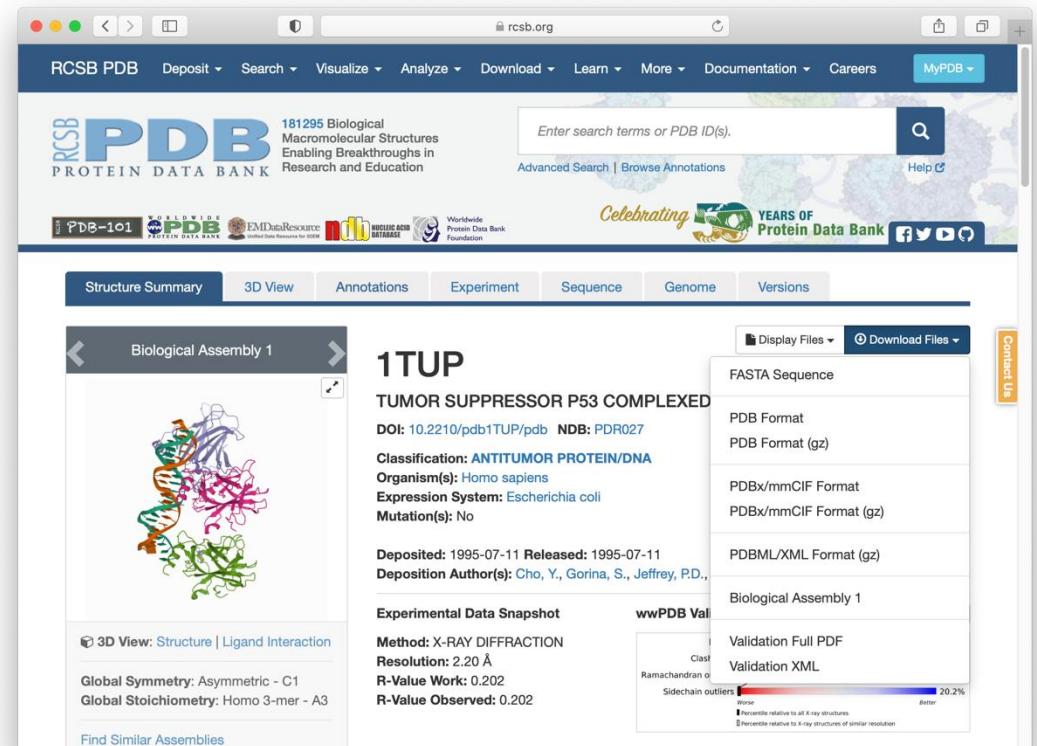
Fetch .pdb file directly to PyMOL using

File>Get PDB... and providing 4 letter code

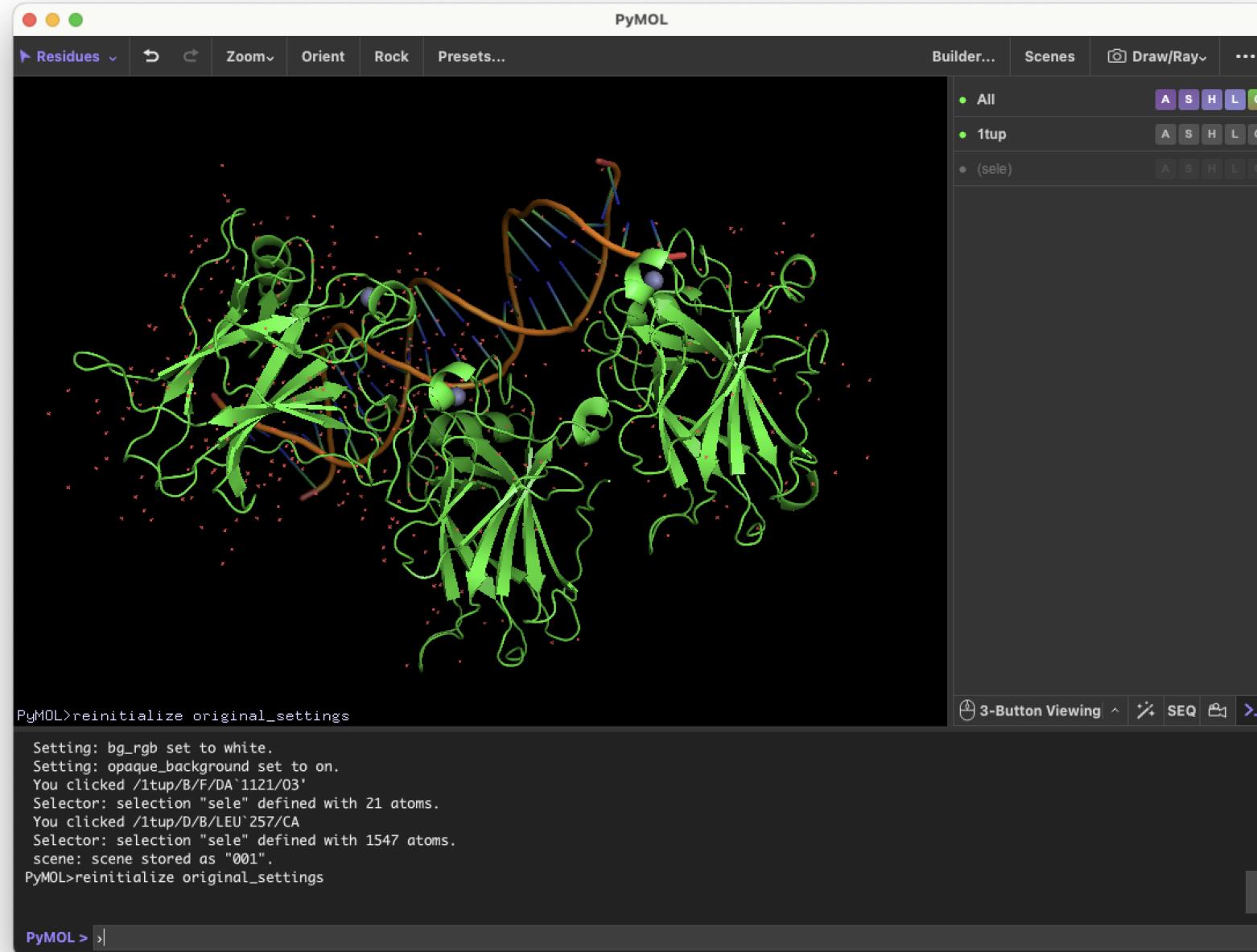
from PDB entry: **1TUP**

OR

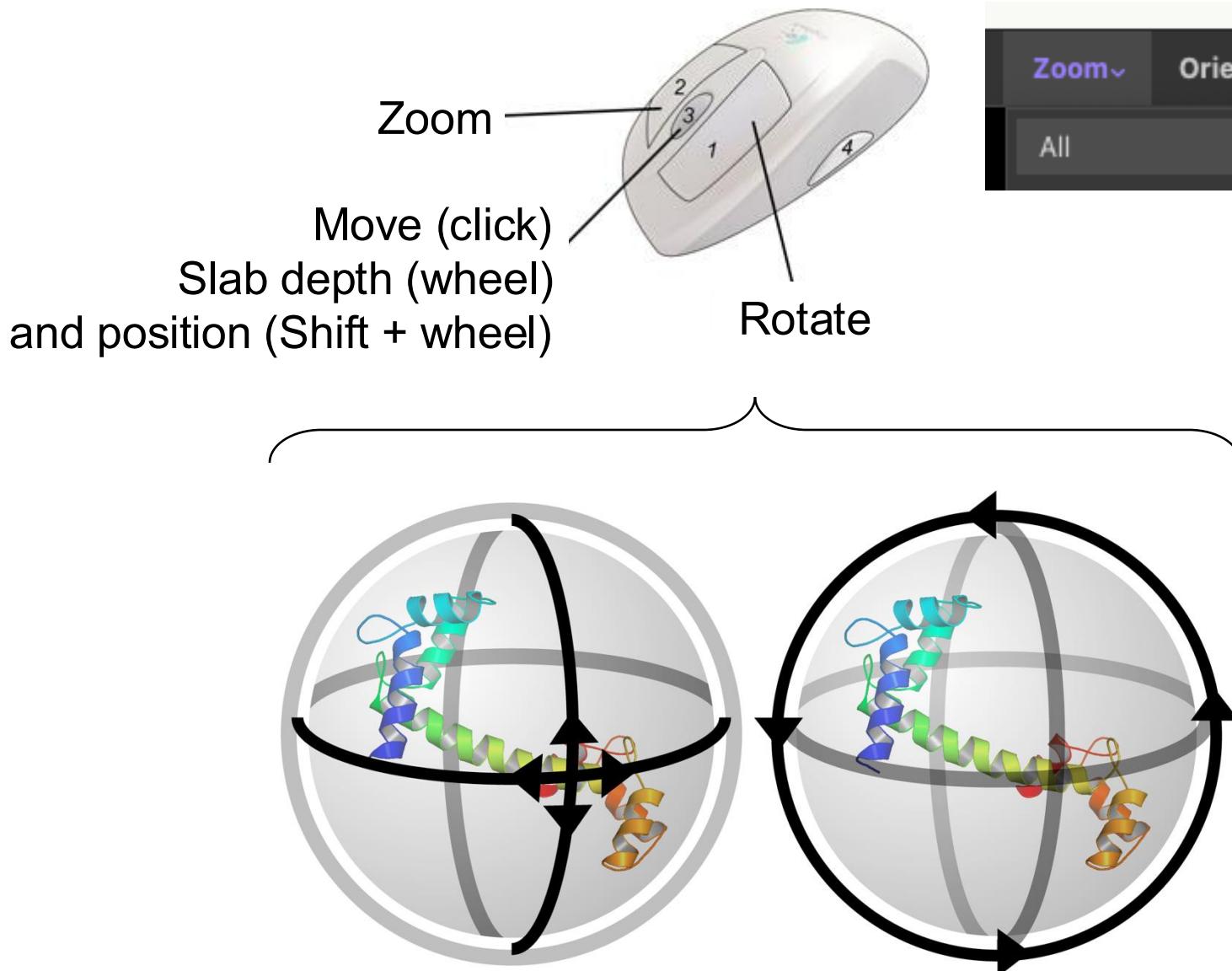
Command line: `fetch 1tup`



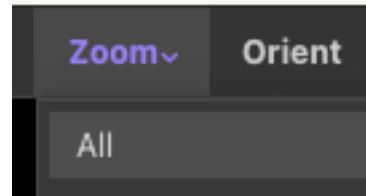
1tup.pdb loaded in PyMOL



2. Mouse controls



If you get lost:
Zoom>All or Orient



Mouse functions + modifier keys:

3-Button Viewing Mouse Controls				
Buttons & Keys	L	M	R	Wheel
	Rota	Move	MovZ	Slab
Shift	+Box	-Box	Clip	MovS
Ctrl	Move	PkAt	Pk1	MvSZ
Ctrl+Shift	Sele	Orig	Clip	MovZ
Single Click	+/-	Cent	Menu	
Double Click	Menu	-	PkAt	

Mouse modes

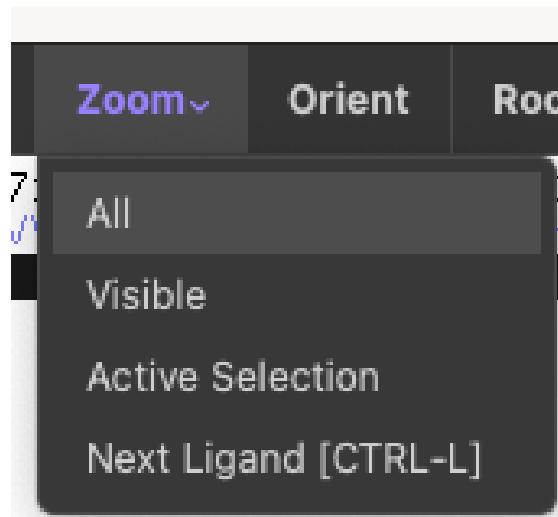


Viewing Mode: view, rotate, translate, and change the representations of objects (you will use this mostly)

Editing Mode: rotate bonds, replace atoms, physically move atoms and residues, etc.

General notes on PyMOL

- Save sessions (**File->Save session...**) from time to time
- Use the **Zoom** or **Orient** buttons on the top panel if you get lost
- To change display quality, select **Display > Quality > Maximum quality** (or other)



Exercise 1: Adjusting view and slab

1. Reset the view.

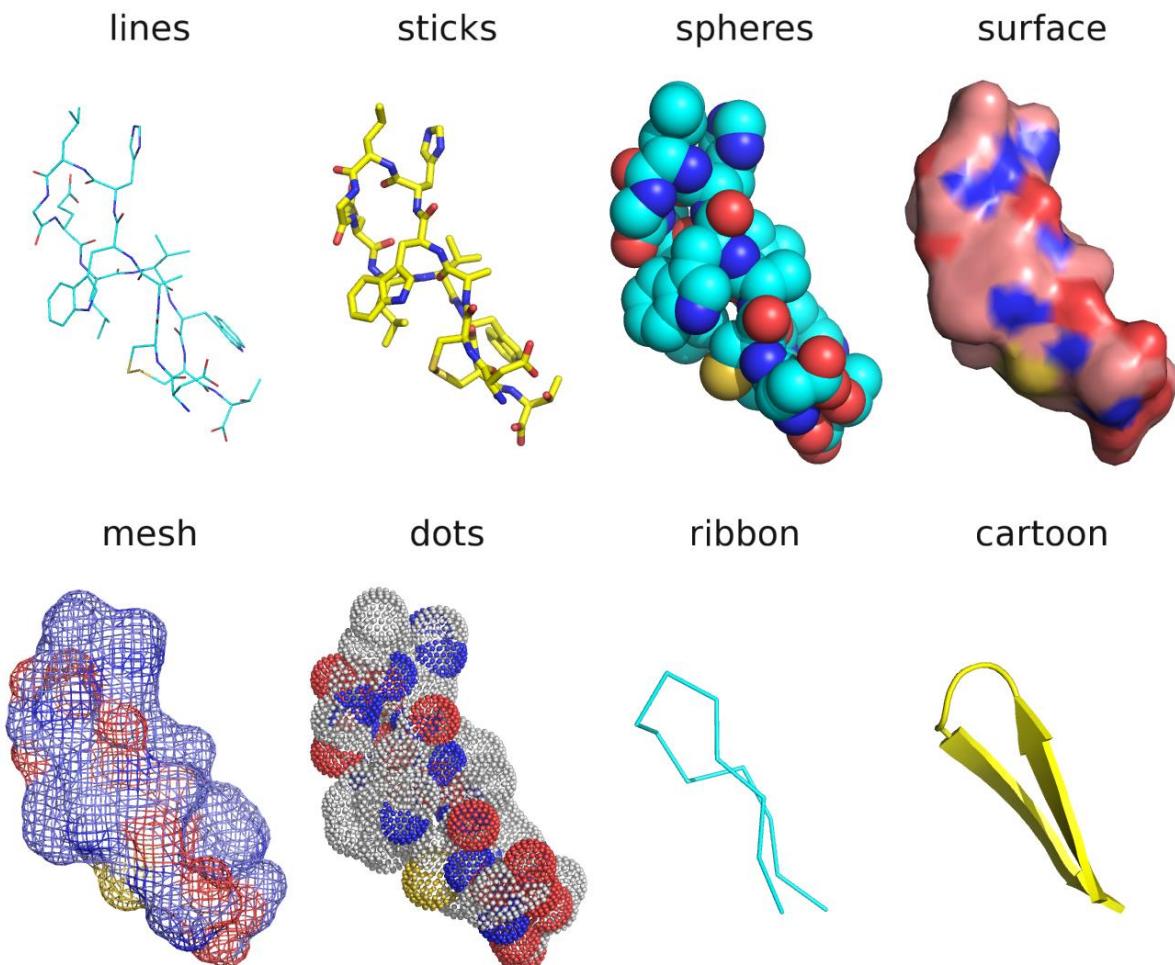
- Locate DNA in the complex, adjust view and slab so you can see single base pair and move the slab so you can see single base-pair as you move the slab along Z-axis.

2. Reset, look around

- How many protein chains are bound to this dsDNA?
- Do all the protein chains bind DNA similarly?



3. Representations



Different representations convey different information

Object and Selections panel – used to change representations and much more



Selections and objects play slightly different roles in PyMOL.

- An **Object** can be created from selections
 - Can be moved and aligned!
- A **Selection** belongs to a specific object
 - Selections are used to modify representation of parts of an object
 - **You cannot display a selection without displaying its original object source**

4. Selections – how to make them

There are multiple ways to perform selections:

By clicking on the molecule:

- Single left mouse click by default selects single residue
(can also be used to select atom/chain/molecule etc. depending by setting in Mouse->Selection Mode menu)
- Shift + Left Mouse button selects a box

By clicking on the sequence:

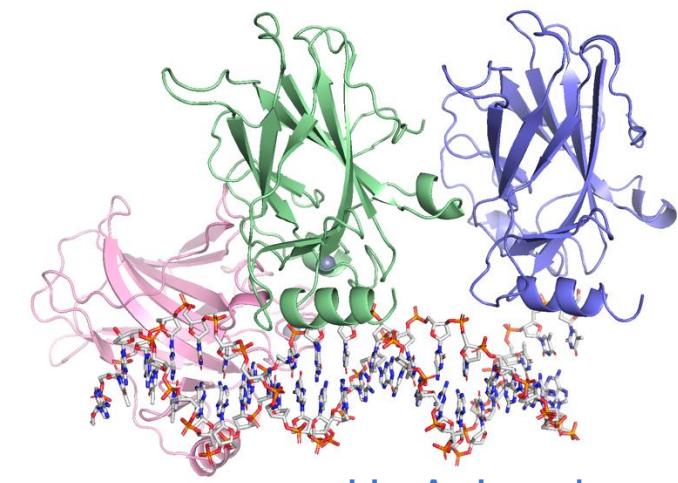
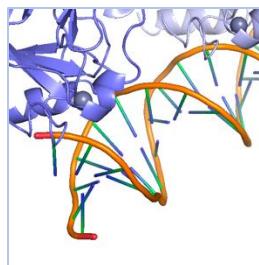
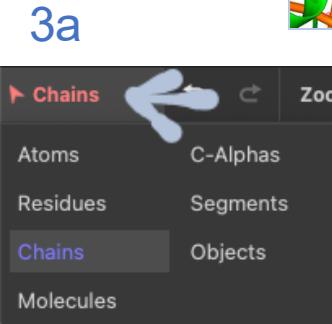
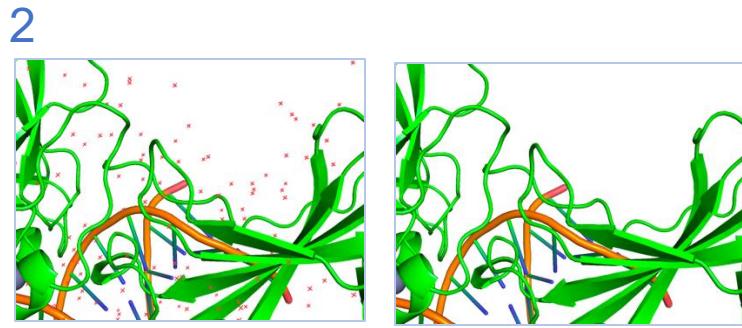
- Select Sequence On in the Display Menu to show sequence; click on the residues

OR

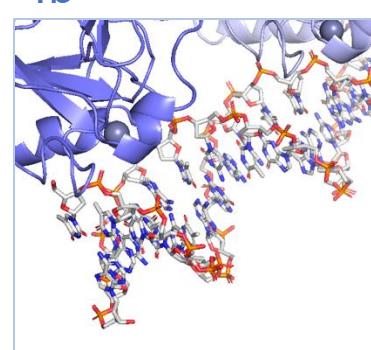
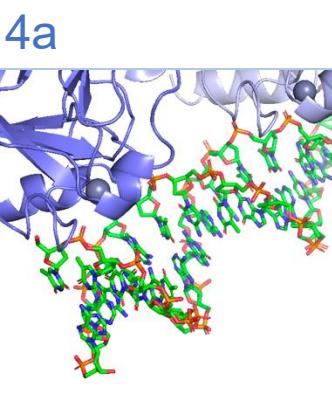
command line: https://pymolwiki.org/index.php/Selection_Algebra

Exercise 2: Changing representations

1. Reset the view.
2. Hide waters:
[H] > waters
3. Select individual protein chains and change their colors:
 - a. Change the mouse selection mode to chains.
 - b. Click on a chain. In the Object Panel, click on [C] button of <sele>, choose a color.
4. Set the DNA representation to licorice sticks and change color by element:
Make sure your mouse selection mode is still chains.
 - a. Click the 2 DNA strands. [S] > as > licorice – sticks.
 - b. [C] > by element > choose a color palette.



Ha An's colors:
pink, palegreen, slate



Making Selections using Sequence Display

-Select Display->Sequence On to show sequence; click on the residues



-Colors of sequence corresponds to colors in structure

-0 = water

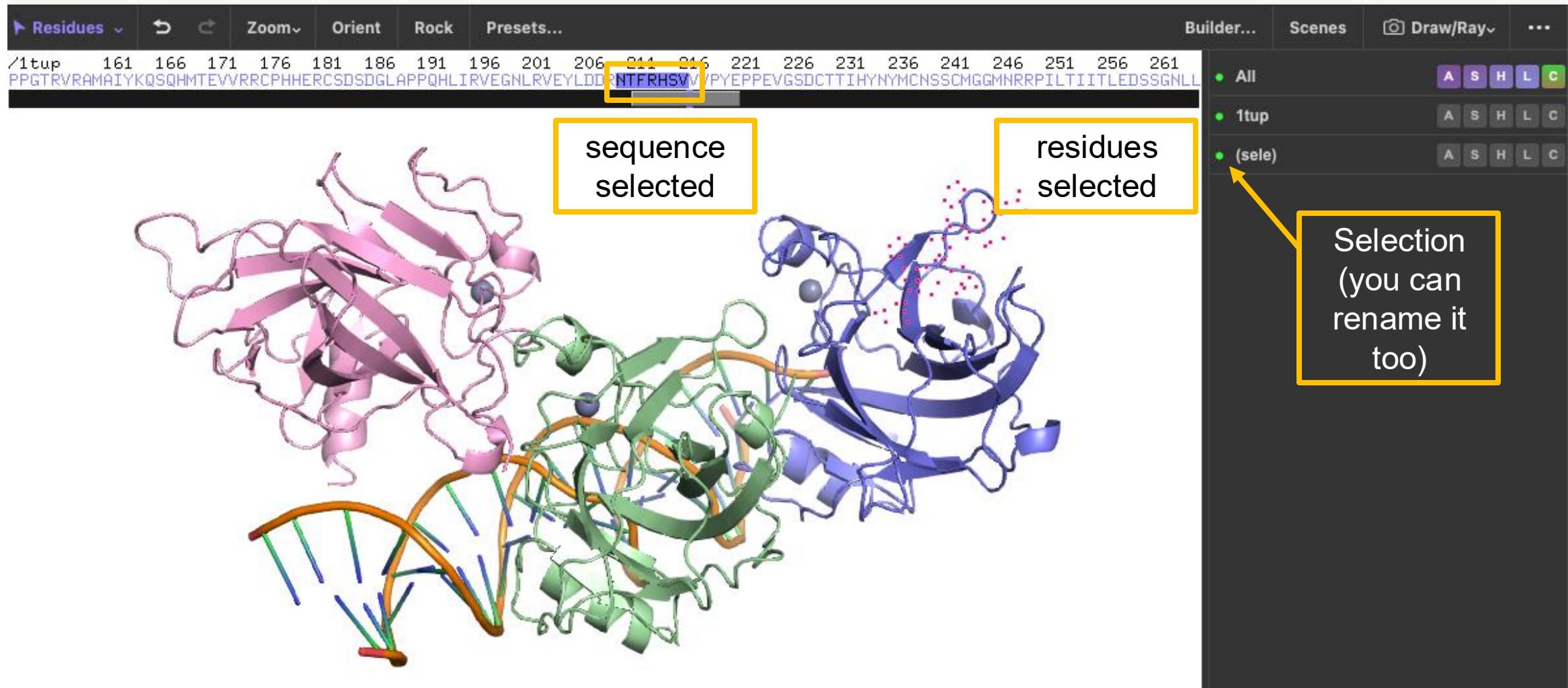
-Residue number corresponds to the residue directly below the first digit

-Display->Sequence Mode->Residue Names changes Sequence display to three-letter amino acid code



if you don't see the numbering, might need to change the background color to white (Display > Background > White)

How selections are indicated



Selections – command line:

Command line offers powerful way of making selections

Hierarchy: model → chain → resn (residue name) → resi (residue number) → name (atom name)

Examples:

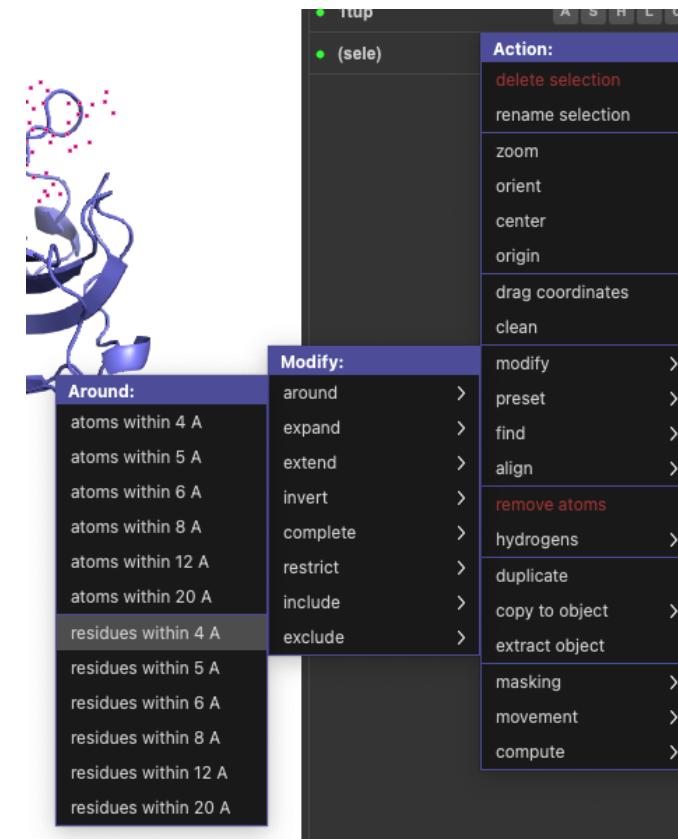
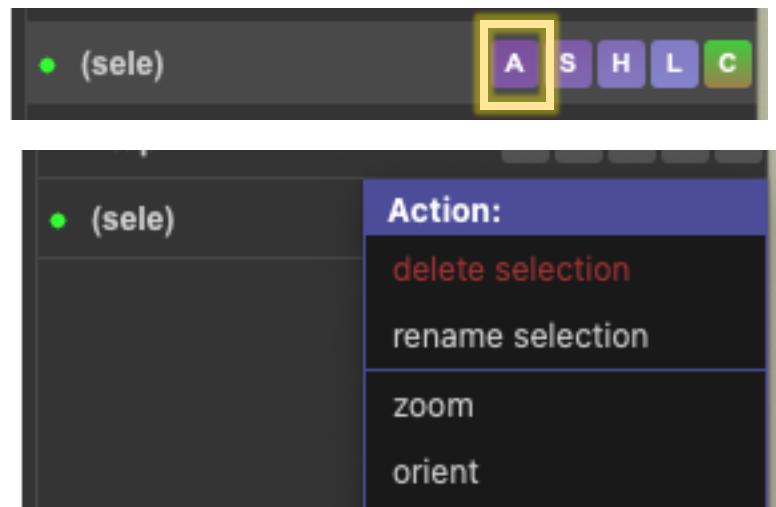
- select chain A **– select chain A**
- select resn lys **– select all lysines**
- select resi 175 **– select residue 175 in all chains**
- select (resi 175 or resi 245) and chain B
 - select residues 175 and 245 in chain B**
- select interface, (chain B within 4 of chain E:F) or (chain E:F within 4 of chain B) **– make selection named interface with all atoms located less than 4A apart between chain B and chain E:F**

Type: help select to see help

Full explanation: https://pymolwiki.org/index.php/Selection_Algebra

Selection and Objects

- rename selection: <sele> [A] > rename selection
- copy selection to new object: <sele> [A] > copy to object
- select all residues within 4Å of selection: <sele> [A] > modify > around > residues within 4 A (this selects the atoms; you can show sidechains by going to [S] > sidechain > stick)



Exercise 3: Working with sequences

Certain residues are most frequently mutated in cancer:

R175, G245, R248, R249, R273, R282.

Select and show the side chains for these residues

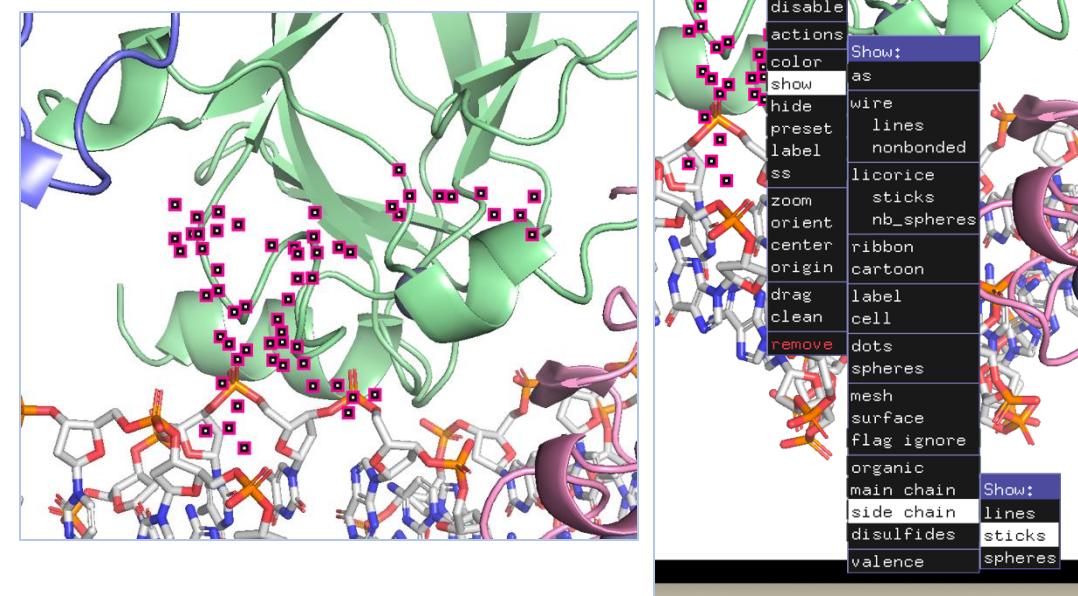
in chain B (middle protein molecule) and look at them.

Q: Where are these residues located?

Q: What could be the role of these residues?

- Display > Sequence
- Make sure your selection mode is 'Residues'

Selecting Residues

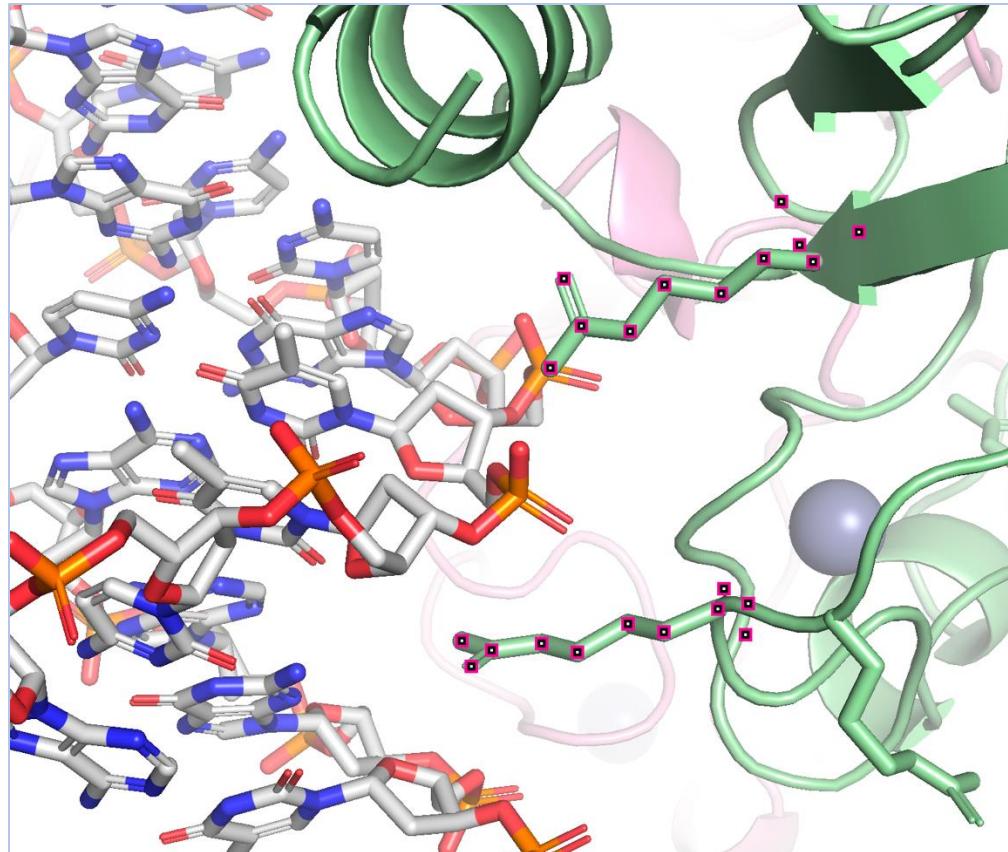


/1tup 141 146 151 156 161 166 171 176 181 186 191 196 201 206 211 216 221 226 231 236 241 246 251 256 261 266 271 276 281 286 291 296 301
FCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVRRCPHHERCSDSDGLAPPQHLIRVEGLDRNTFRHSVWPYEPPEVGSDCTTIHYNYMCNSCMGGMNRPILTITLEDSSGNLLGRNSFEVRVACPGDRDRTEEEENLRKKGEPEHLLPPG

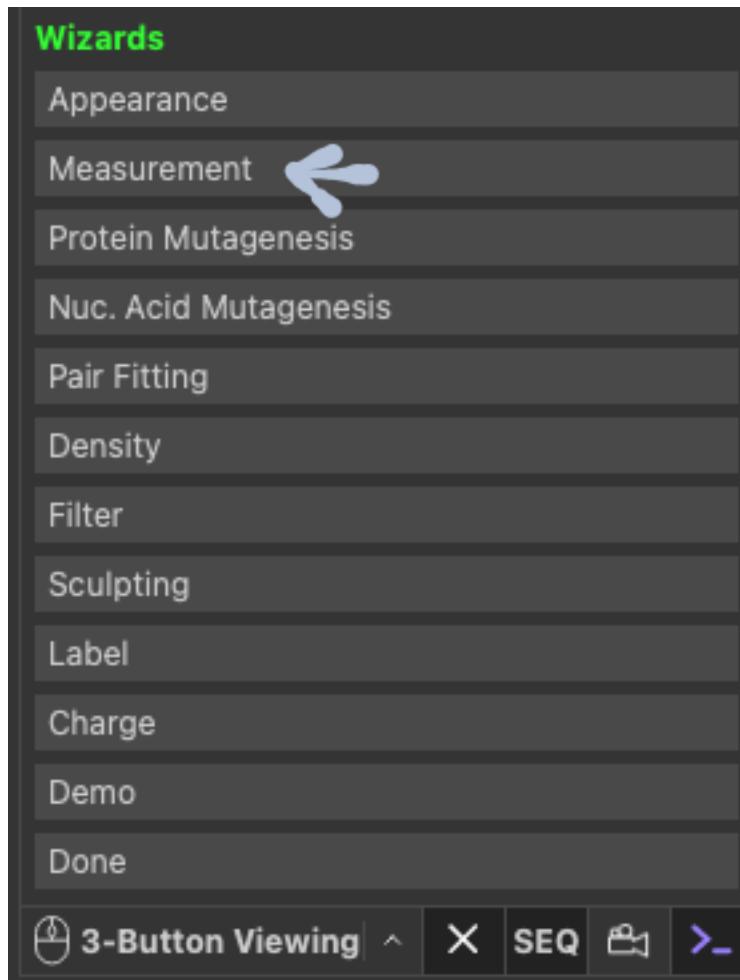
Exercise 3: Working with sequences

Q: Where are these residues located?

Q: What could be the role of these residues?

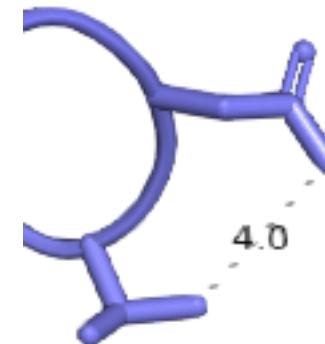
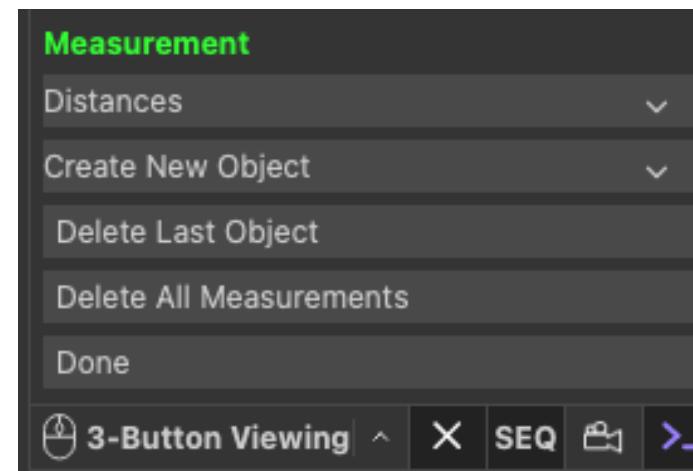


5. Making measurements



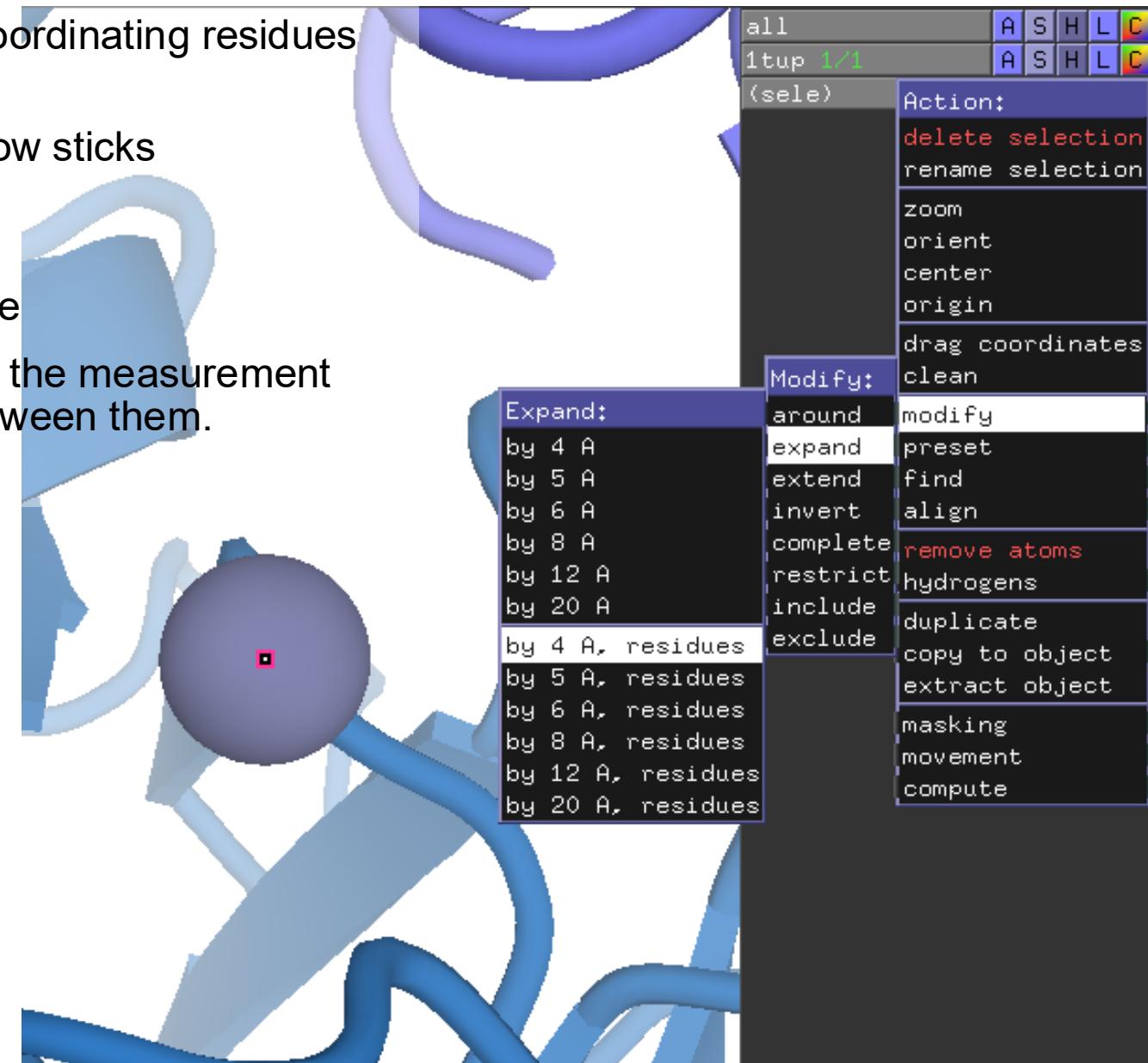
Please click on the first atom...

click the first atom, then
click the second atom



Exercise 4: Measure the coordinating distances of the Zn atom

- Select a Zn atom, expand the selection to select coordinating residues
<sele> [A] > modify > expand > by 4 Å, residues
- change the representation of the side chains to show sticks
<sele> [S] > sidechain > sticks
- you can also change the color of this selection:
<sele> [C] > by element > choose your color palette
- In the menu bar, Wizard > Measurement to turn on the measurement option click two atoms to measure the distance between them.



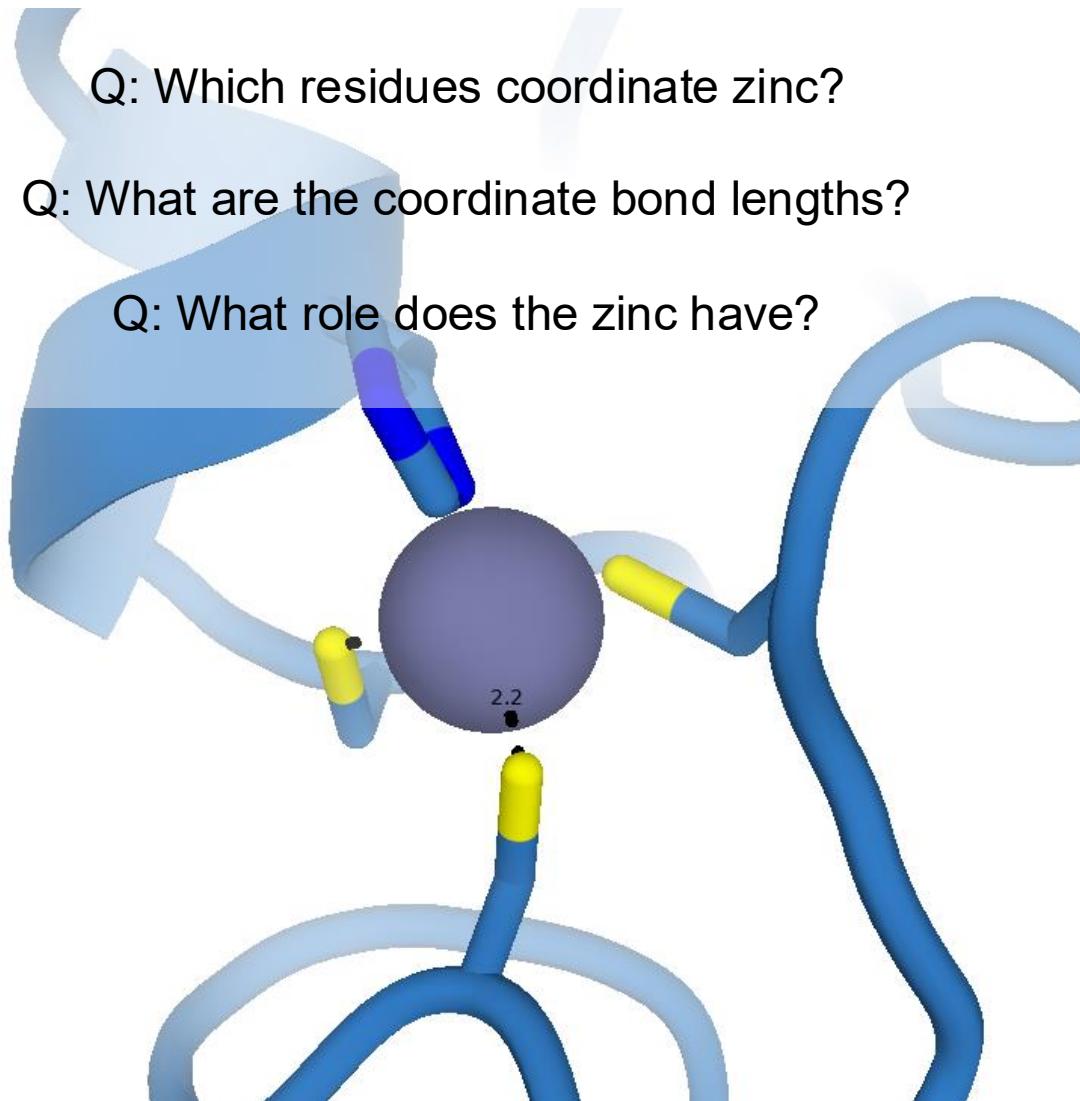
Exercise 4: Measure the coordinating distances of the Zn atom

The zinc ion has been shown to be important in DNA binding.

Q: Which residues coordinate zinc?

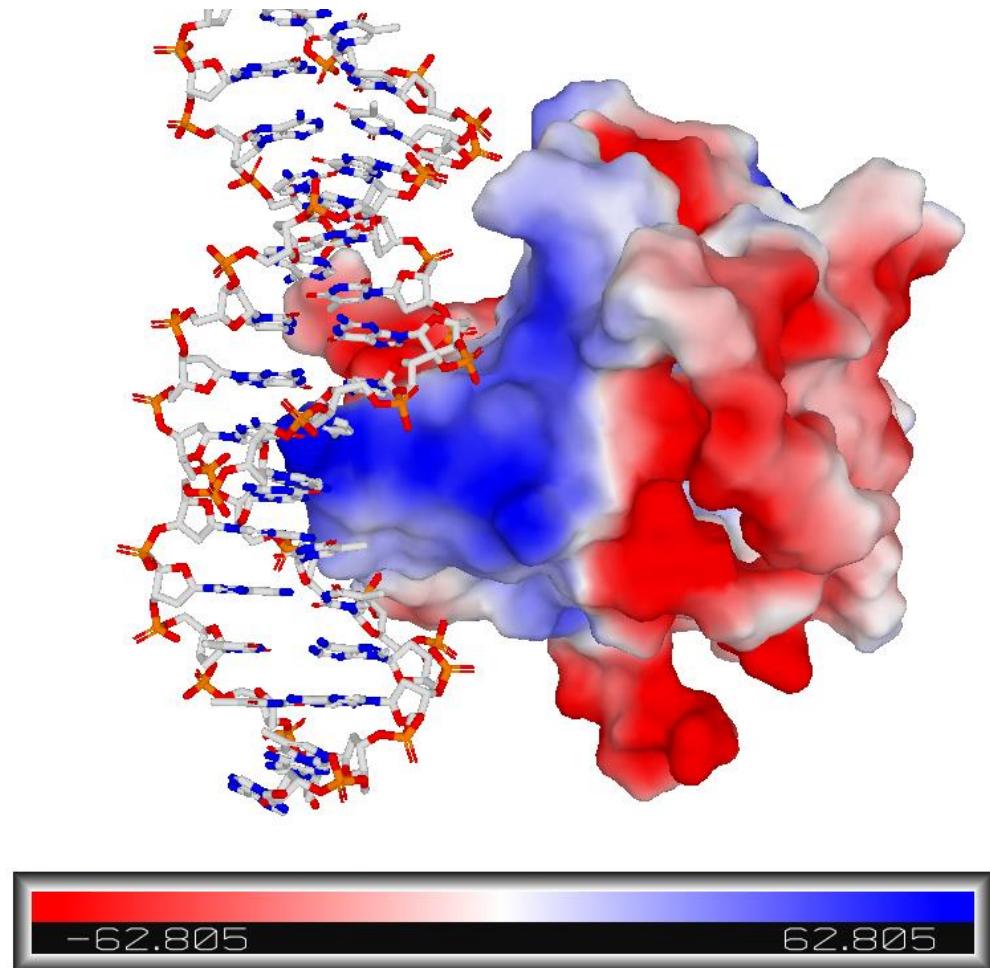
Q: What are the coordinate bond lengths?

Q: What role does the zinc have?



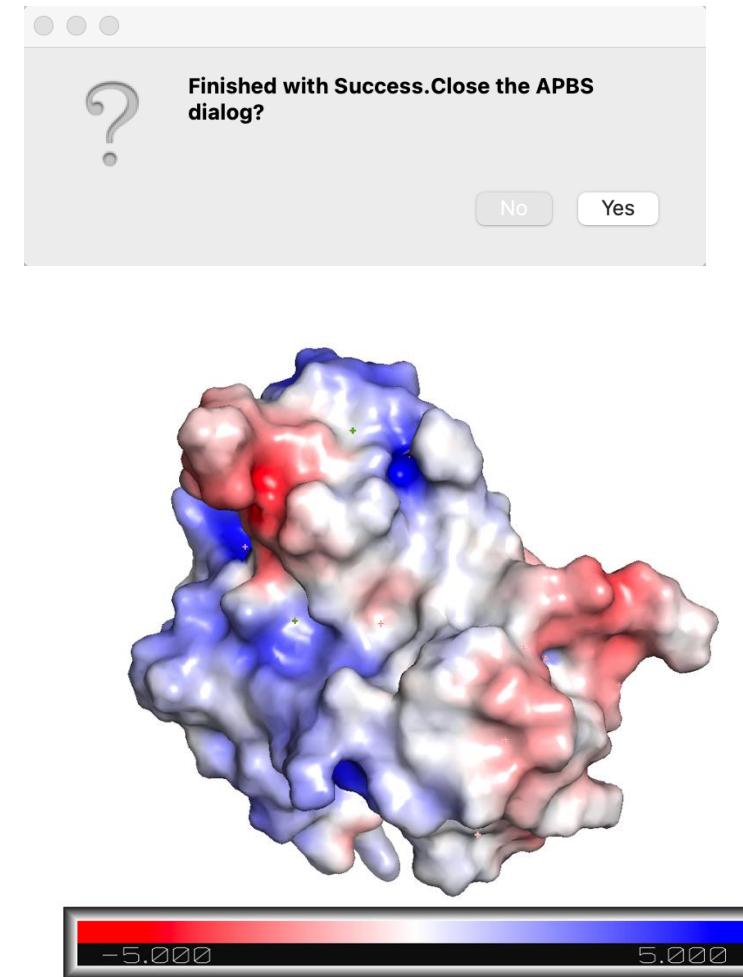
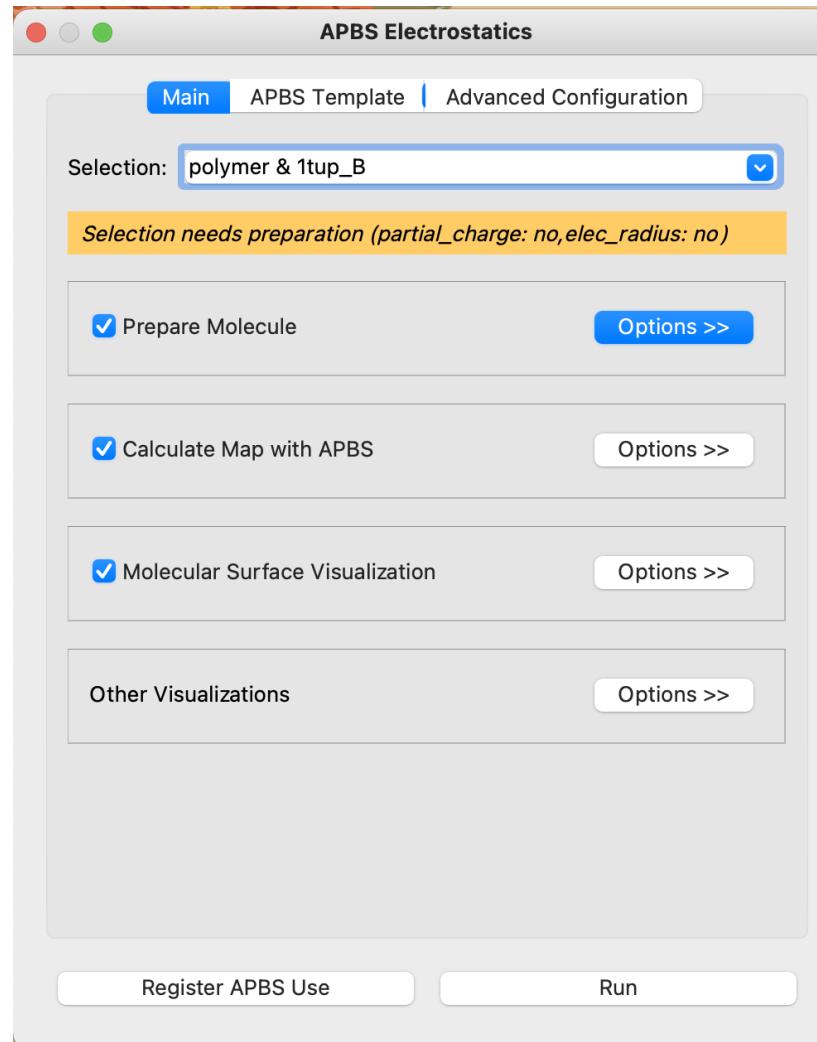
Exercise 5: Showing electrostatics

- Create new individual objects that contain the protein chains A, B and C individually
(use the command `split_chains 1tup`)
- Generate electrostatic potential surfaces quickly:
[A] > generate > vacuum electrostatics > protein contact potential
- Do you see any electronegative or electropositive patches?
Where are they located?



Better (more proper) way to generate electrostatics by APBS:

- Plugin > APBS electrostatics > choose your selection > Run (takes some time)
wait till “Finished with Success” popup



If this errors out for you, try going into Prepare Molecule > Options > Method: change to “use formal_charge and vdw”

Exercise 6: Align the three protein chains

- you should have the 3 protein chains in separate objects
- align chains A and C to chain B
[A] > align > to molecule > chain B
- Are the proteins adopting different conformations?

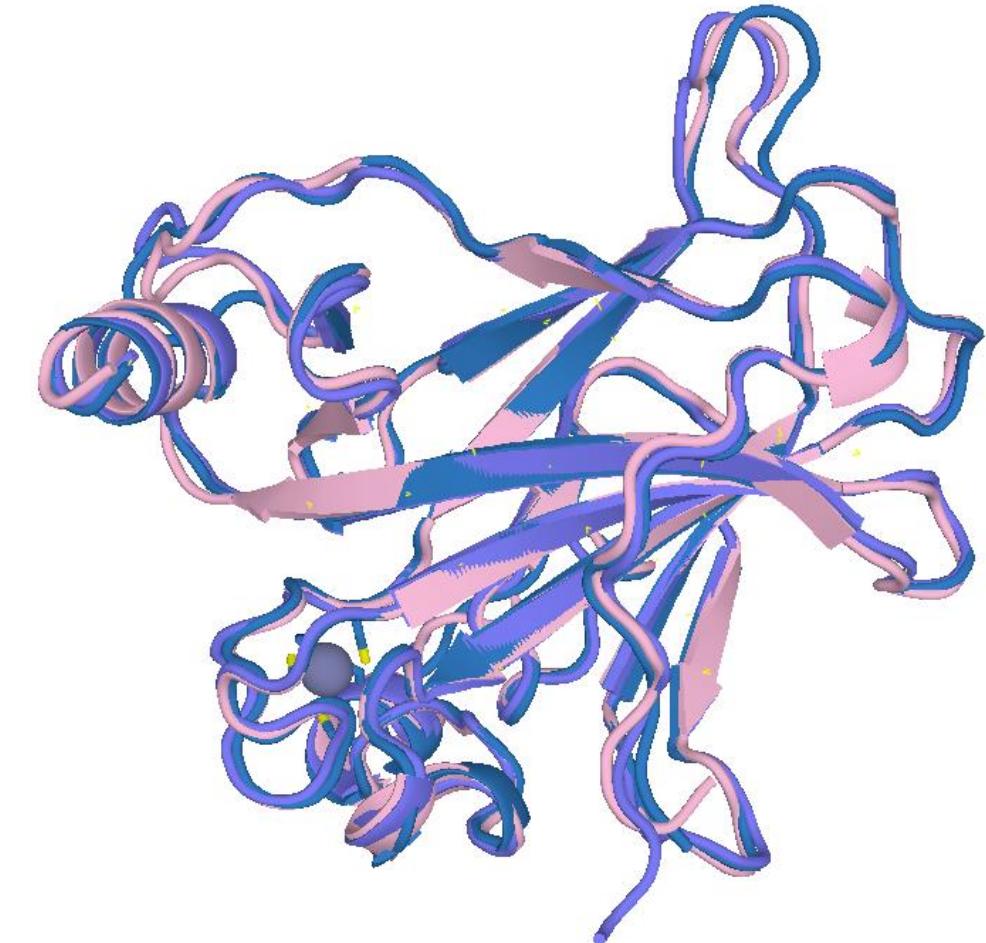
You can align all proteins to a currently selected one
A→Action→Align→All to this

or align proteins one by one
A→Action→Align→To Molecule

Exercise 6: Align the three protein chains

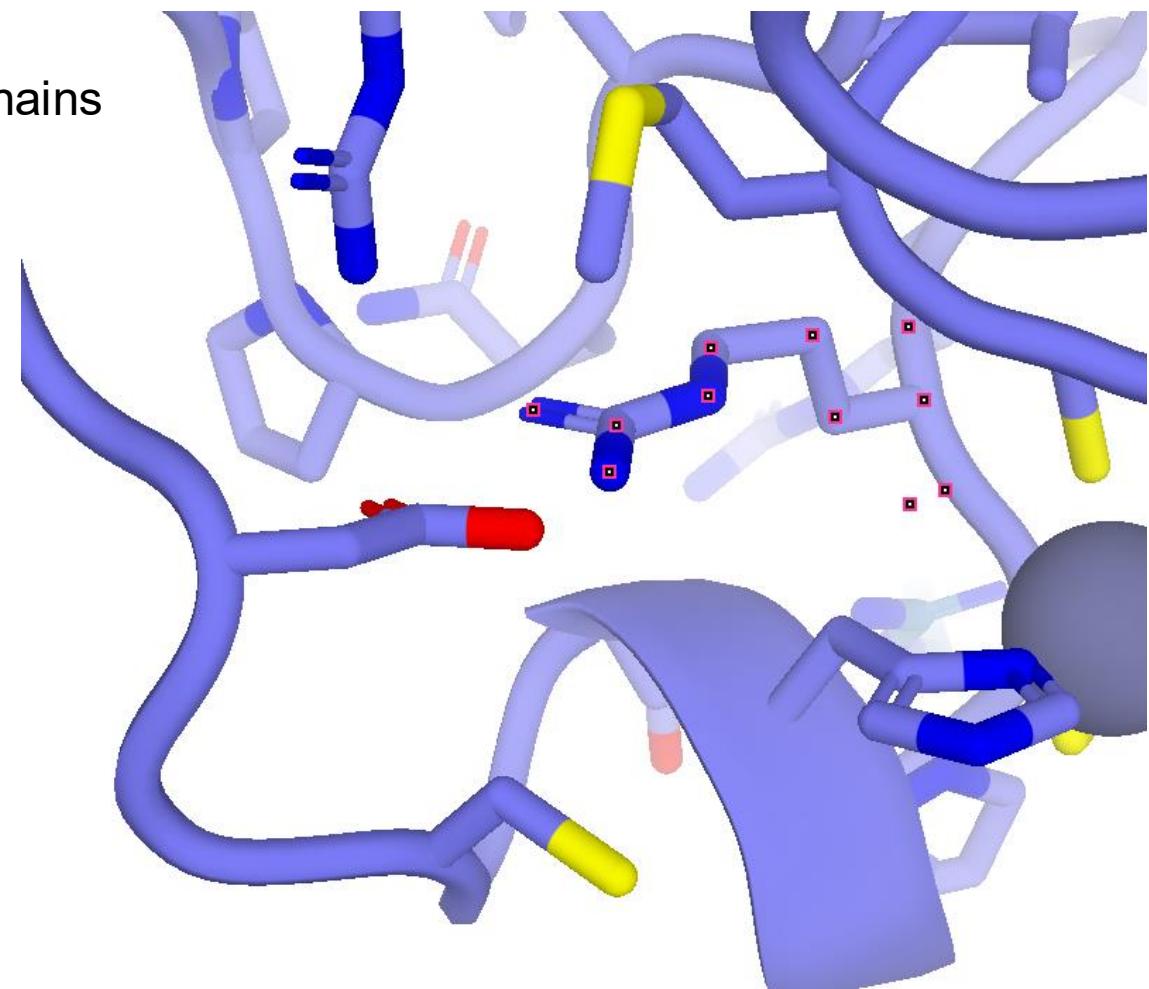
- you should have the 3 protein chains in separate objects
- align chains A and C to chain B
[A] > align > to molecule > chain B
- Are the proteins adopting different conformations?

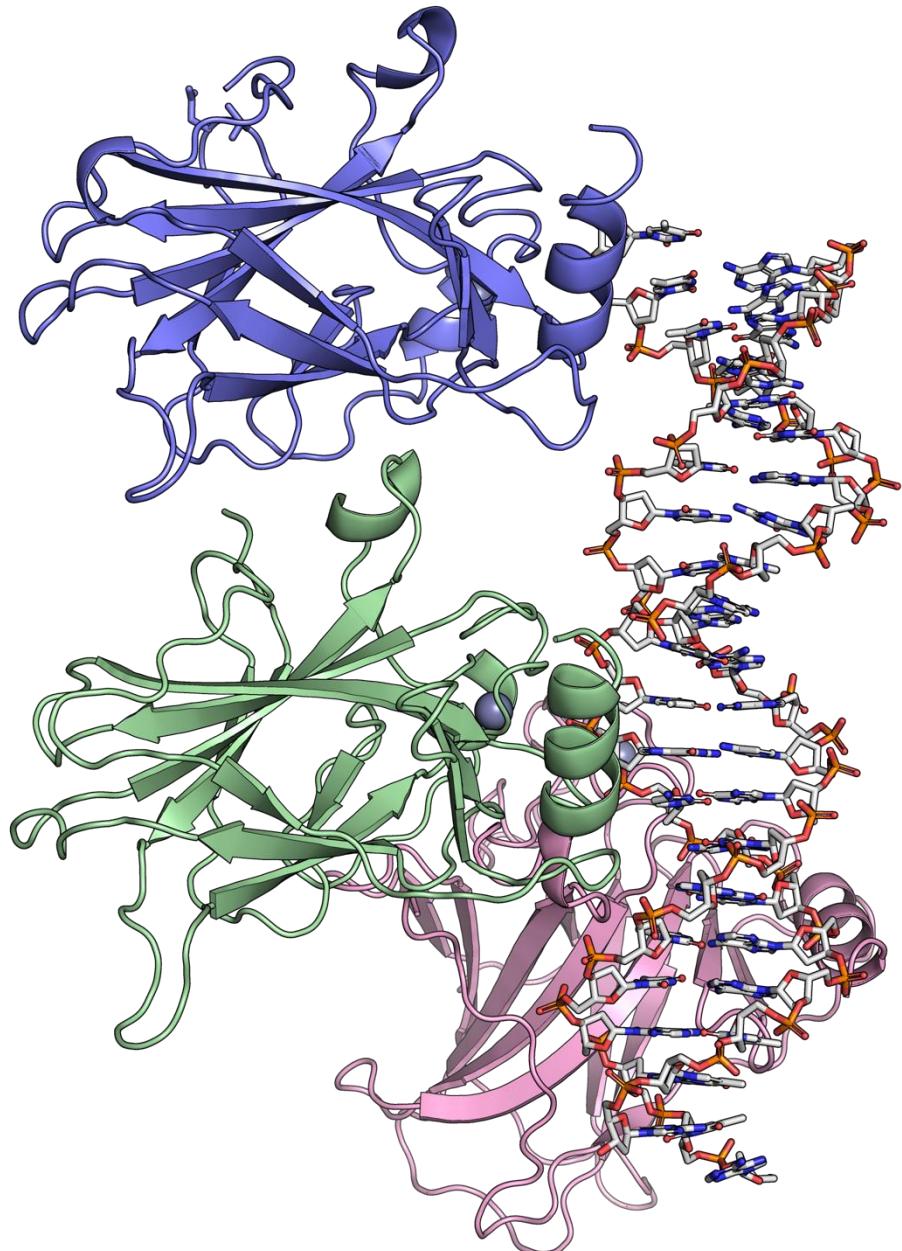
```
Executive: RMS: 2 atoms rejected during cycle 5 (RMSD=0.44).  
Executive: RMSD = 0.429 (169 to 169 atoms)  
Executive: object "aln_1tup_C_to_1tup_B" created.  
PyMOL> set transparency, 0.5  
Setting: transparency set to 0.50000.  
Executive: RMSD = 0.429 (169 to 169 atoms)
```



Exercise 7: perform an *in silico* alanine mutation

- Select R175 in chain A, show and color its sidechain.
- expand the selection to residues within 5 Å and show sidechains
- Go to menu bar Wizard > Mutagenesis
- In the panel on the right, click on 'No Mutation', select ALA.
- Click on the R175 that you want to mutate
- Then click Apply
- What could be an effect of this mutation?





10. Make a figure

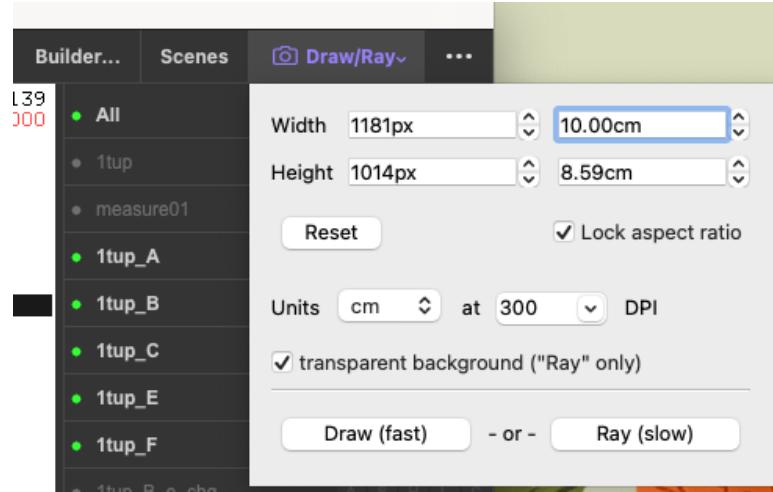
Reset everything (can just open a new window)

1tup

Recolor each protein chain a different color

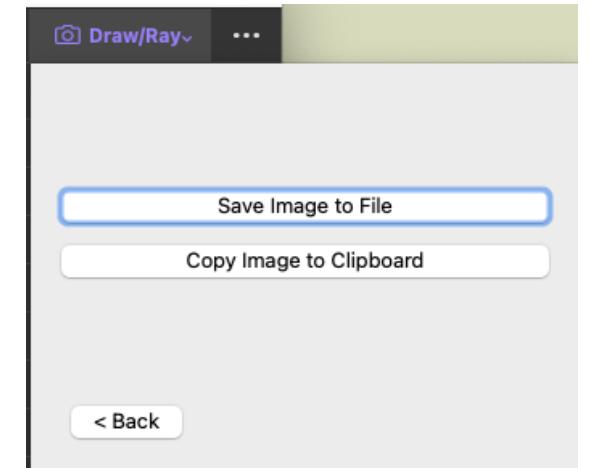
Set DNA chains to be sticks, color by element

Draw/Ray: Ray(slow) > save image as > png to somewhere



Ha An's settings to
make it look pretty:
set spec_count, 0
set spec_reflect, 1

set ray_shadows, off
set ray_trace_mode, 1

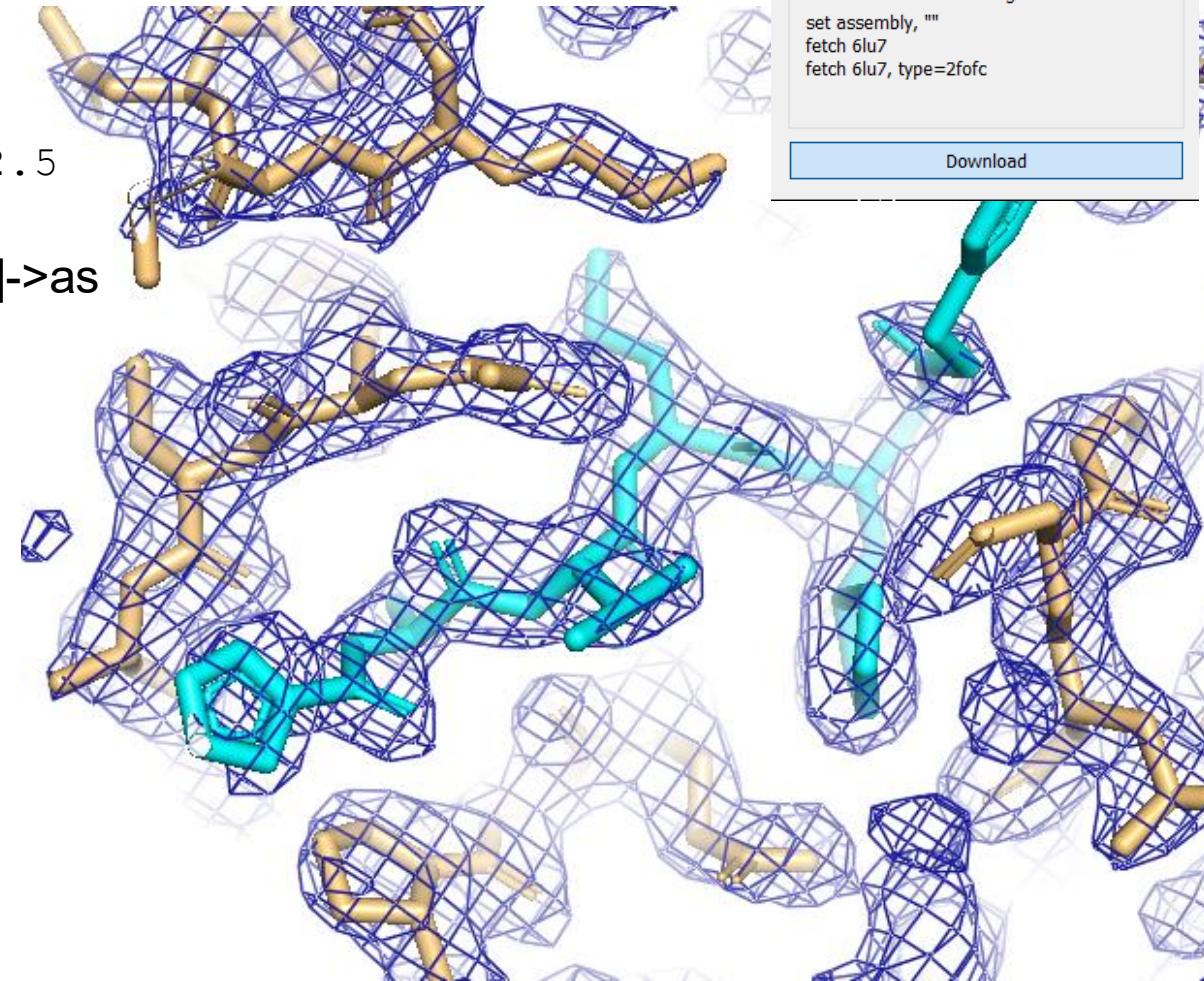


Bonus: look at electron density of an inhibitor

Explore electron density in the crystal structure of COVID-19 main protease in complex with an inhibitor N3 (pdb: 6lu7)

Get pdb... **6lu7**: PDB and 2FoFc map

- Hide waters
- Create a mesh representing the map:
 - Action > mesh > @level 1.0
 - isomesh map, 6lu7_2fofc, 1, 3bep, carve = 2.5
- Color the mesh (e.g. blue > density)
- Color protein chain and inhibitor ([C]->color by chain, [S]->as licorice (sticks)
- Explore electron density for the inhibitor



Note: Downloading will save the files in the directory defined by the "fetch_path" setting.

PDB ID: **6lu7**

PDB Structure Object name (optional)

2FoFc Map Object name (optional)

FoFc Map Object name (optional)

PDB Structure Options

Chain name (optional):

Assembly (optional):

This will run the following command
 set assembly, ""
 fetch 6lu7
 fetch 6lu7, type=2fofc

Download