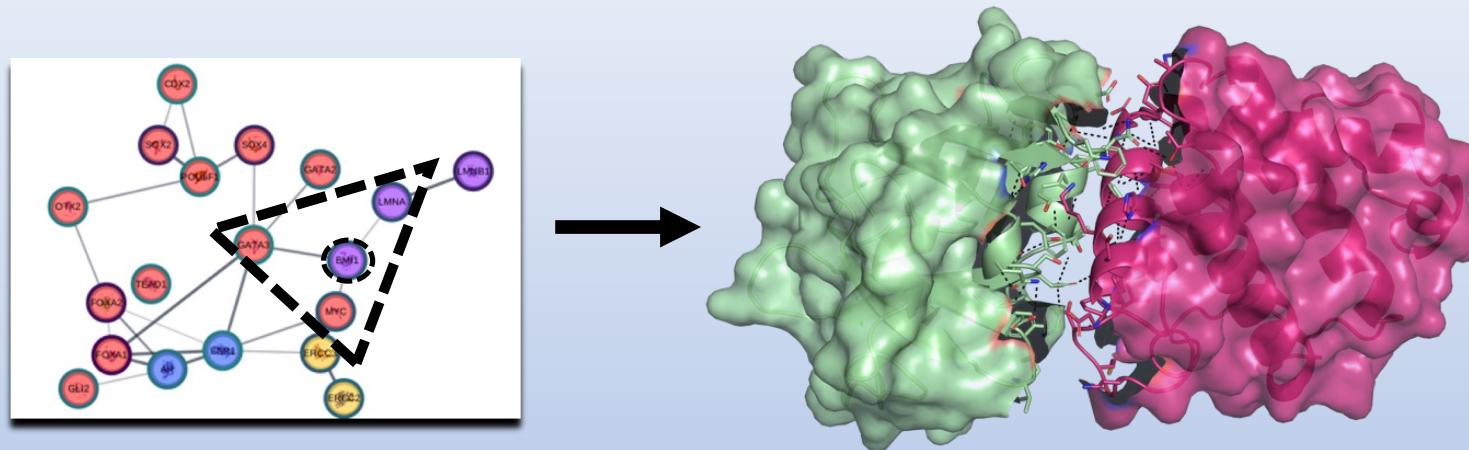


# Day4: Protein Structures



## Visualization of Protein Interactions

Prof Tiina A. Salminen, Dr Mia Åstrand, MSc Marion Alix



Structural Bioinformatics Laboratory

InFLAMES Research Flagship Center

Biochemistry, Faculty of Science and Engineering

Tiina.Salminen@abo.fi

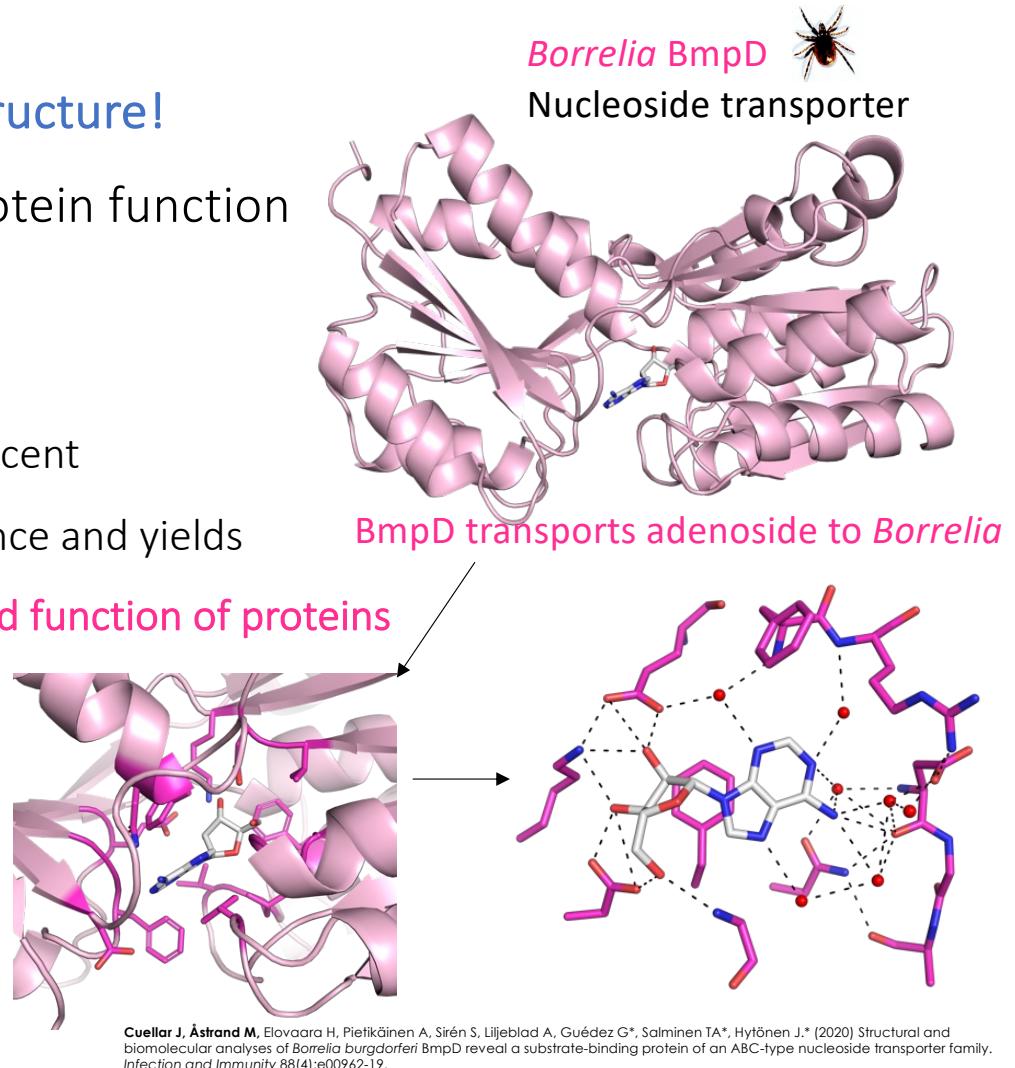
1. Introduction
  - 5 min break
2. Pymol Exercise - everyone
3. Group meeting & work
  - "group" breaks

# Why do we want to study protein structures?

- The function of a protein is determined by its structure!
- If we know the structure, we can understand protein function
- Examples:
  - Drug design → therapeutic agents can be designed
  - Chemical industry → enzymes can be made more efficient
  - Agriculture → modified proteins may increase tolerance and yields
  - In your project → analyze interactions and understand function of proteins

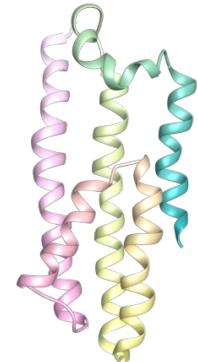
Typically, ~10% of the residues are functionally important while 90% create the 3D structure essential for function (structurally important)

Åstrand M, Cuellar J, Hytönen J, Salminen TA (2019) Predicting the ligand-binding properties of *Borrelia burgdorferi* s.s. Bmp proteins in light of the conserved features of related *Borrelia* proteins. *Journal of Theoretical Biology* 462:97-108.



# Tertiary and quaternary structure

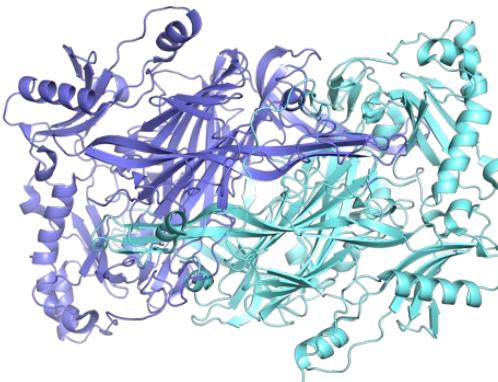
- The **tertiary** structure is the three-dimensional structure of an entire polypeptide chain (folded protein)
- The **quaternary** structure consists of more than one polypeptide chain



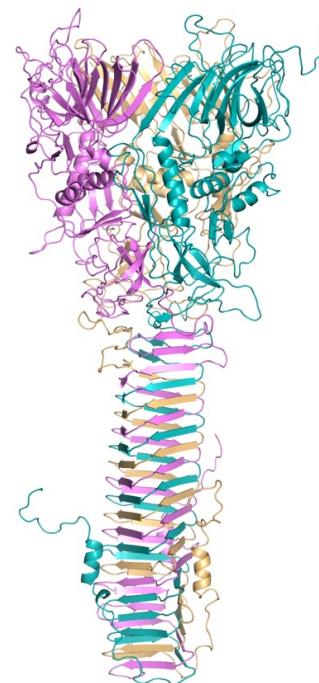
**Monomer** =  
only one chain

**Multimers/oligomers** = more than one chain

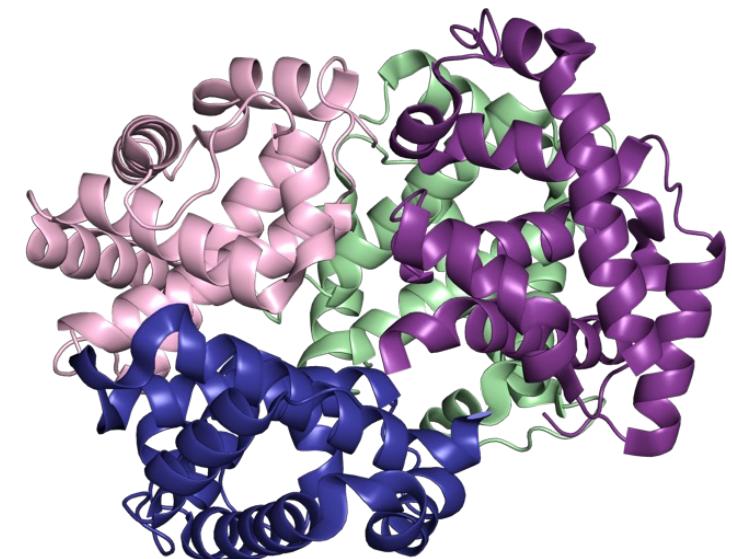
- **Homomer** = Identical chains
- **Heteromer** = Different chains



**Homodimer** = two identical chains colored by chain



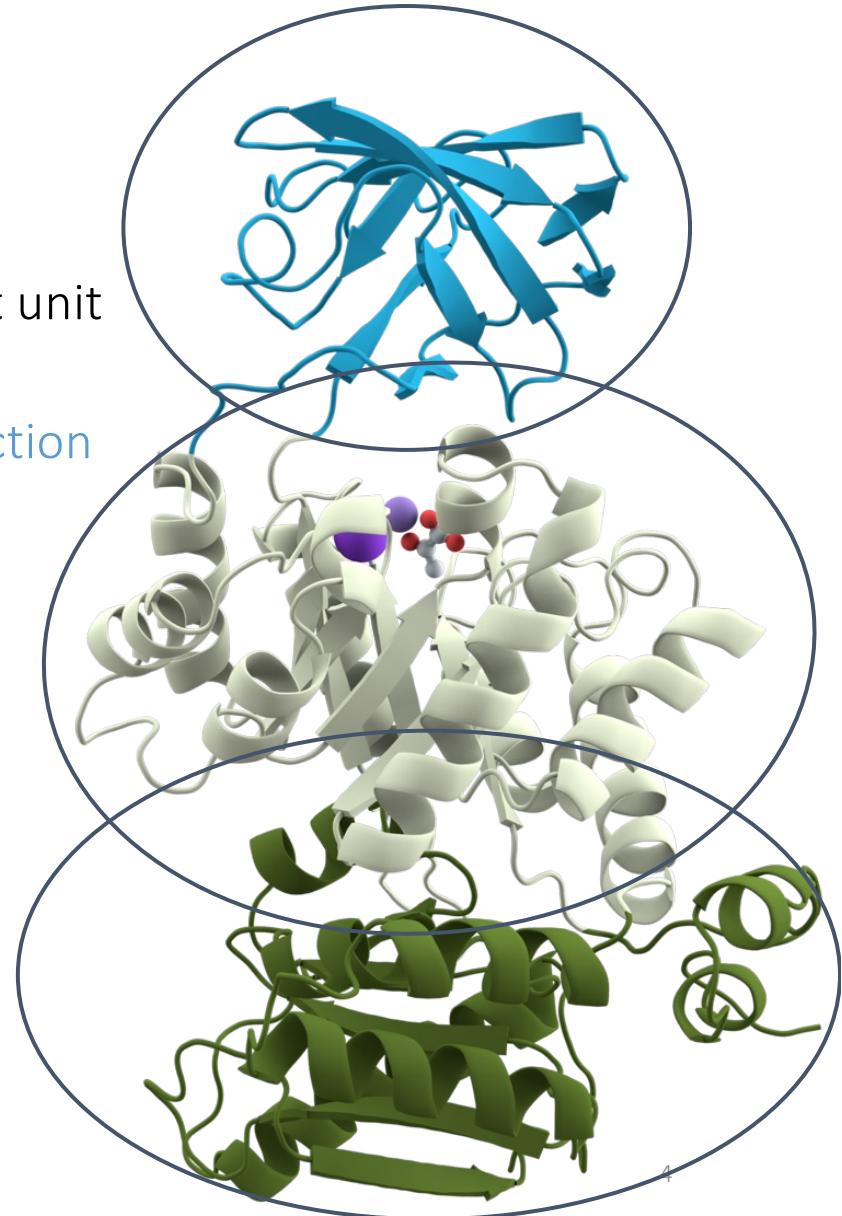
**Homotrimer** = three identical chains



**Heterotetramer** = different chains

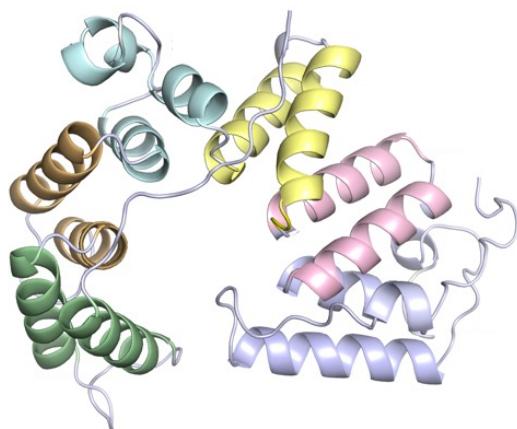
# Protein domains

- A **domain** is a functional, compact and semi-independent unit of the 3D structure
- Usually **folds independently** and often has **a specific function**
- Many proteins consist of several domains

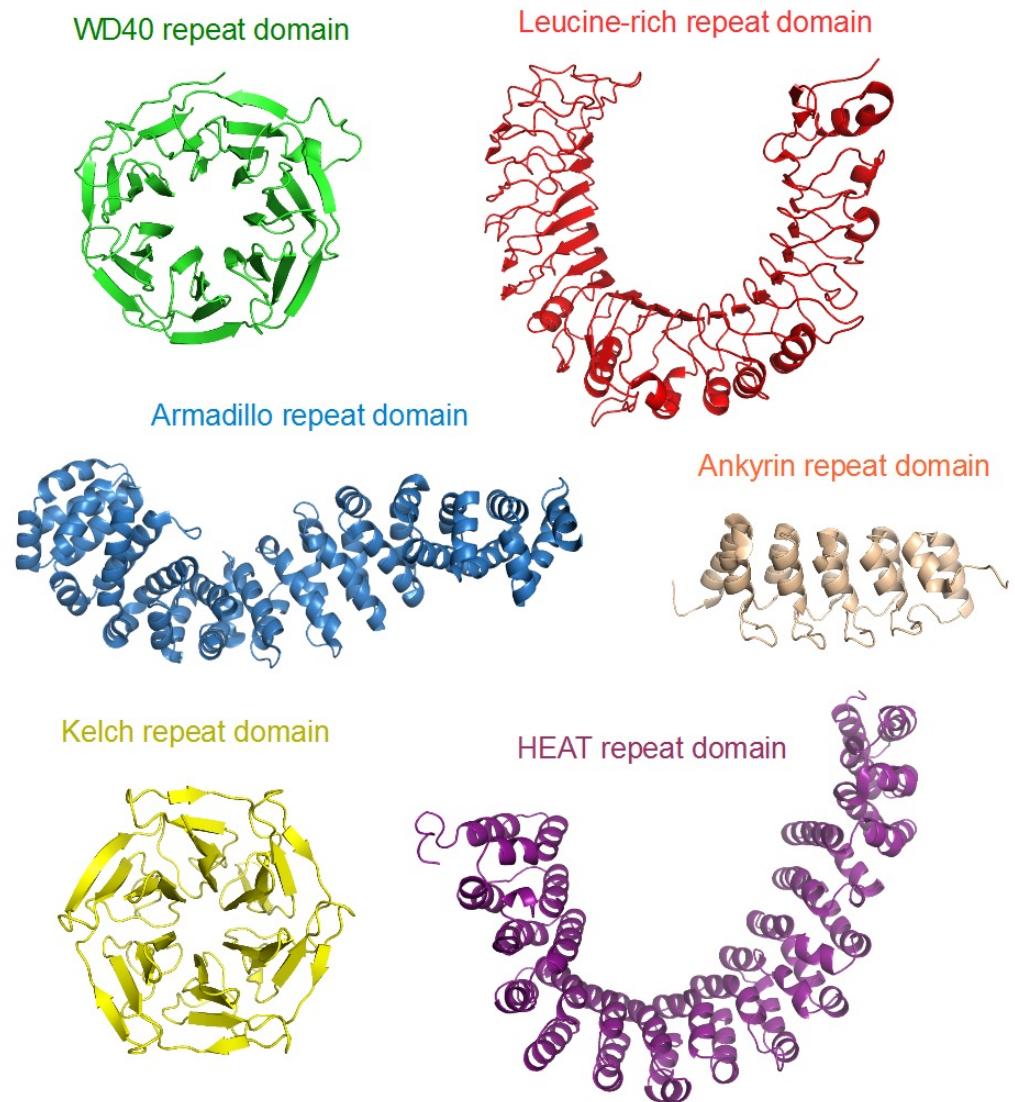


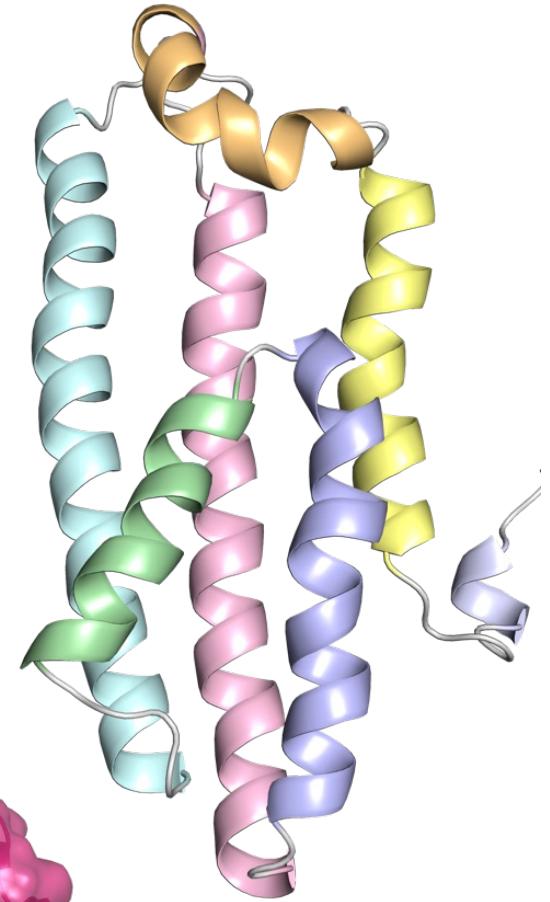
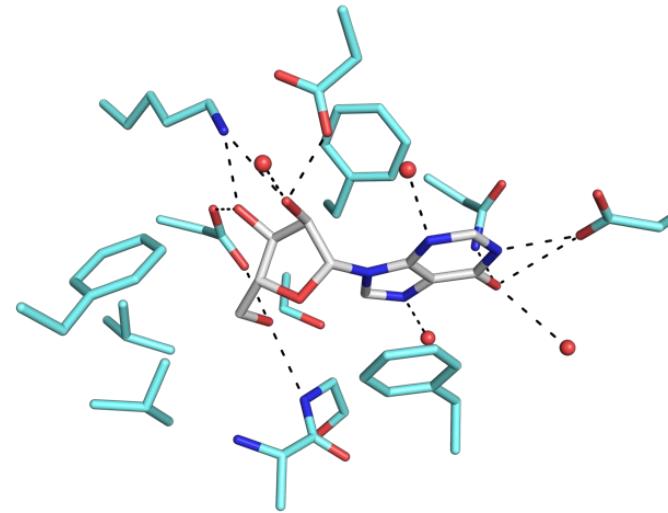
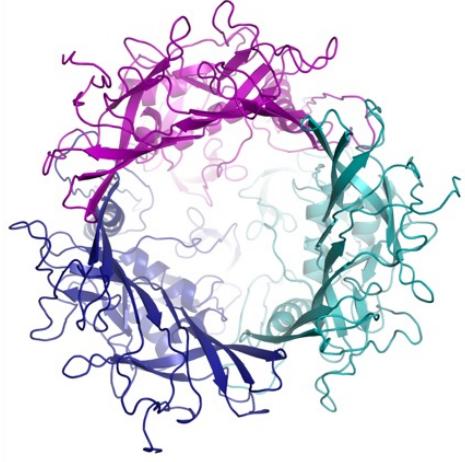
# Repeat domains

- Tandem repeat domains are very versatile
- Often involved in protein-protein interactions

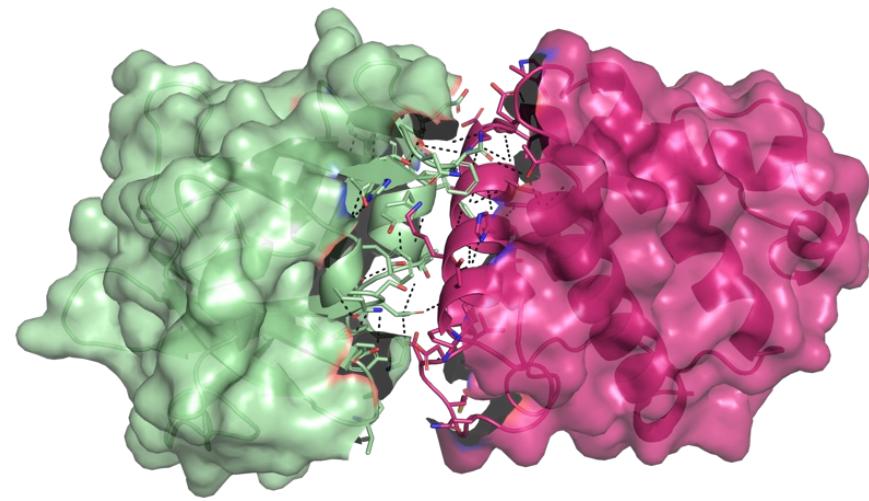
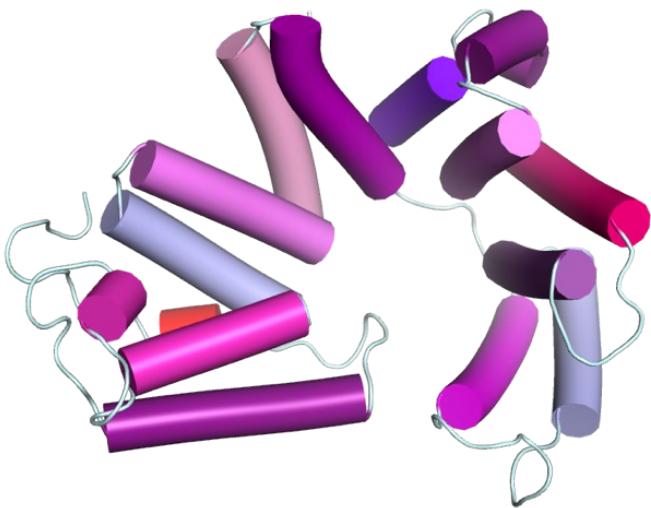


TPR repeat





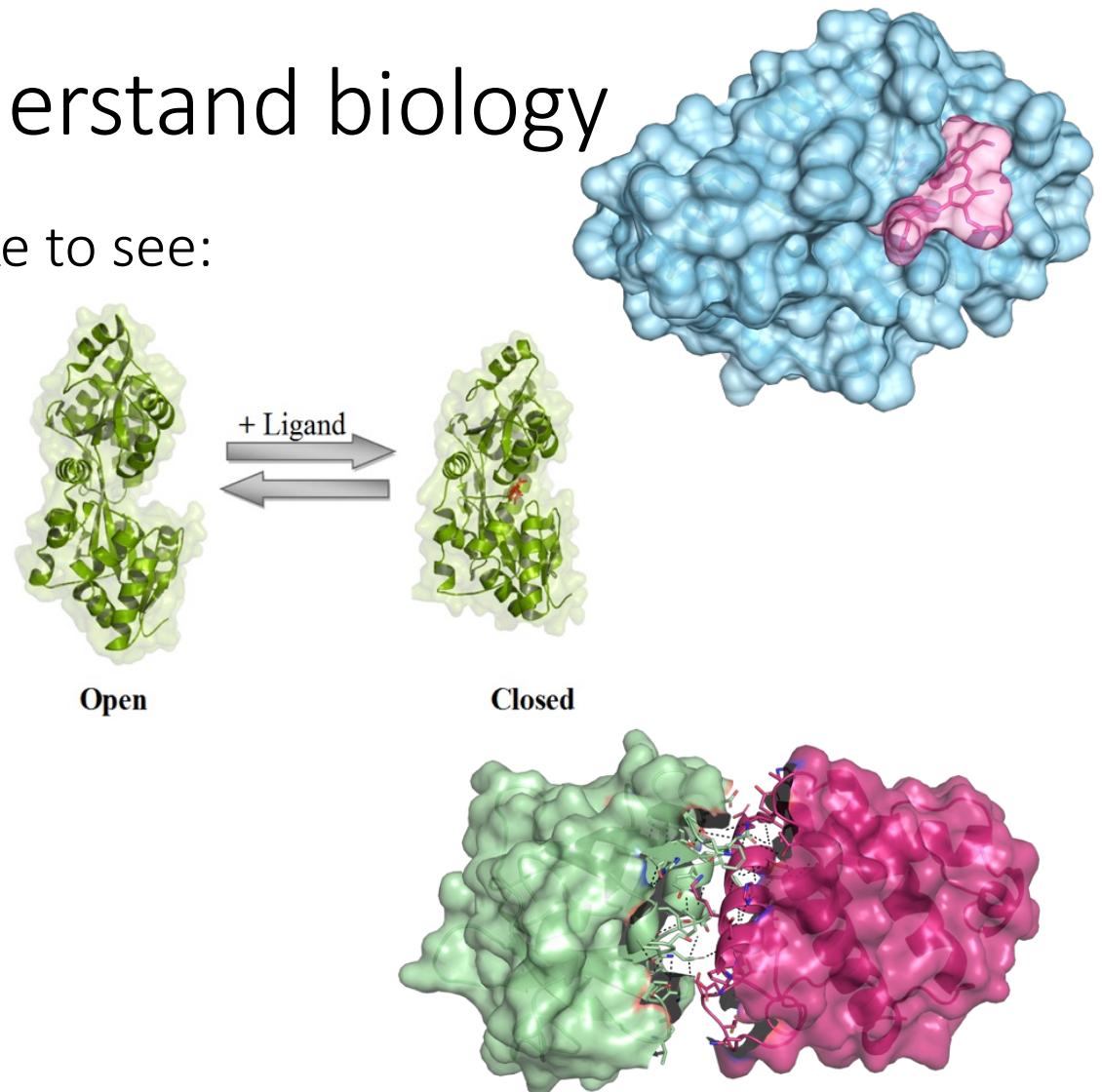
## Visualization of 3D structures



# Structure is a tool to understand biology

In addition to the protein, we would like to see:

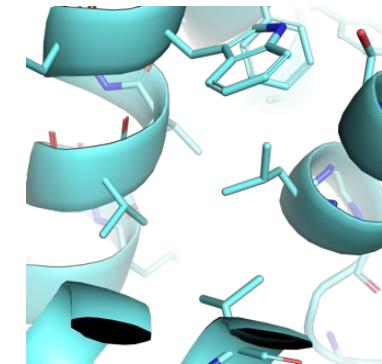
- Ligands, cofactors
- Functional conformations, protein dynamics
- Molecular interactions
  - Protein-Small molecule
  - Protein-Protein (PPI)
  - Protein-DNA
- Complex structures



# Molecular interactions

## 1. Hydrophobic interactions

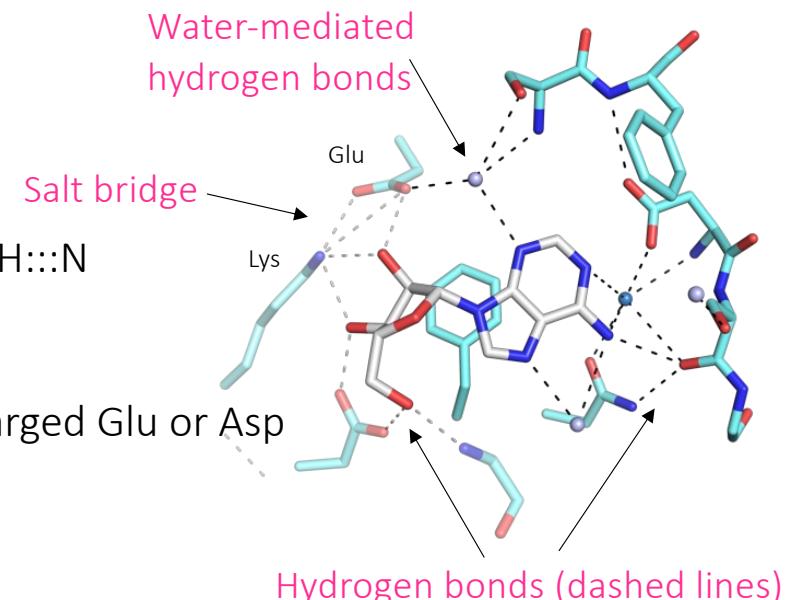
- Non-polar, hydrophobic residues pack together because they cannot participate in hydrogen bonding with water
  - Inside protein
  - Protein-protein interfaces



Hydrophobic interactions

## 2. Polar interactions

- Hydrogen bonds
  - Two polar groups interact: O-H:::O; N-H:::O; N-H:::N; O-H:::N
- Ionic interactions (Salt bridges)
  - Positively charged Arg or Lys (or His) with negatively charged Glu or Asp



## 3. Covalent interactions

- Disulphide bonds (formed between two cysteine residues)
- Metal ions e.g.  $Mg^{2+}$ ,  $Mn^{2+}$ ,  $Cu^{2+}$ ,  $Zn^{2+}$

# 3D structures from PDB

- Includes 3D structures from macromolecules
  - Proteins, DNA, RNA
- Established in 1971 at Brookhaven National Laboratory
  - Contained 7 structures!
  - Helped to store and distribute structural data
- Structures have been determined using
  - X-ray crystallography
  - Nuclear Magnetic Resonance
  - Electron microscopy (cryo-EM)



CRYSTALLOGRAPHY

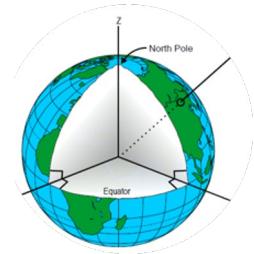
## Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There

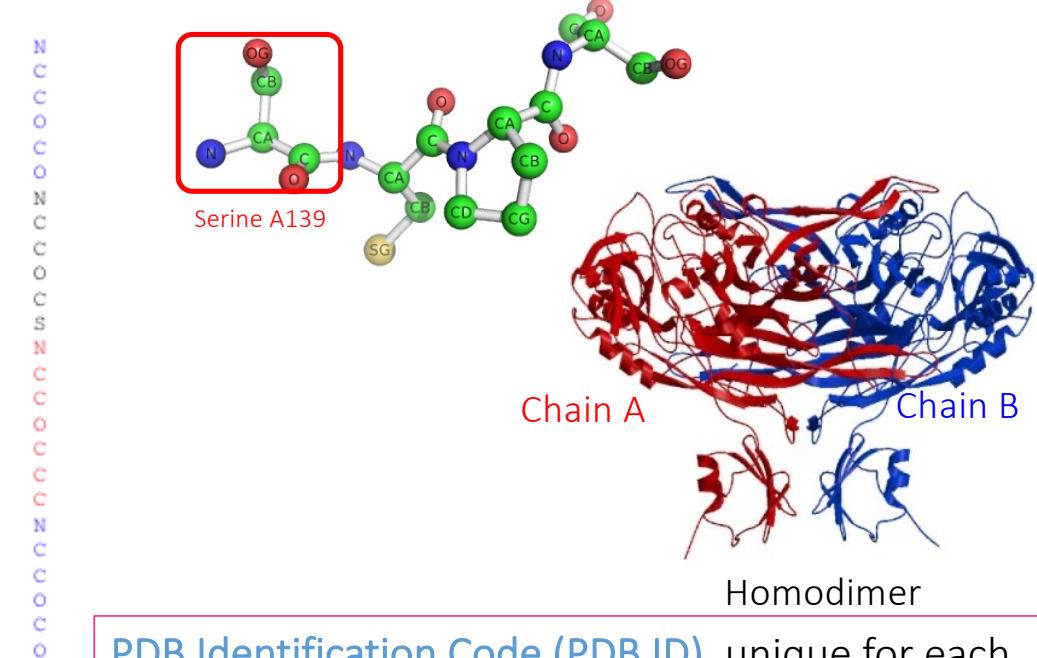
Crystallography: Protein Data Bank. *Nature New Biology* **233**, 223 (1971).  
<https://doi.org/10.1038/newbio233223b0>

# The PDB File

Atomic coordinates describe the location in space for each atom in the protein



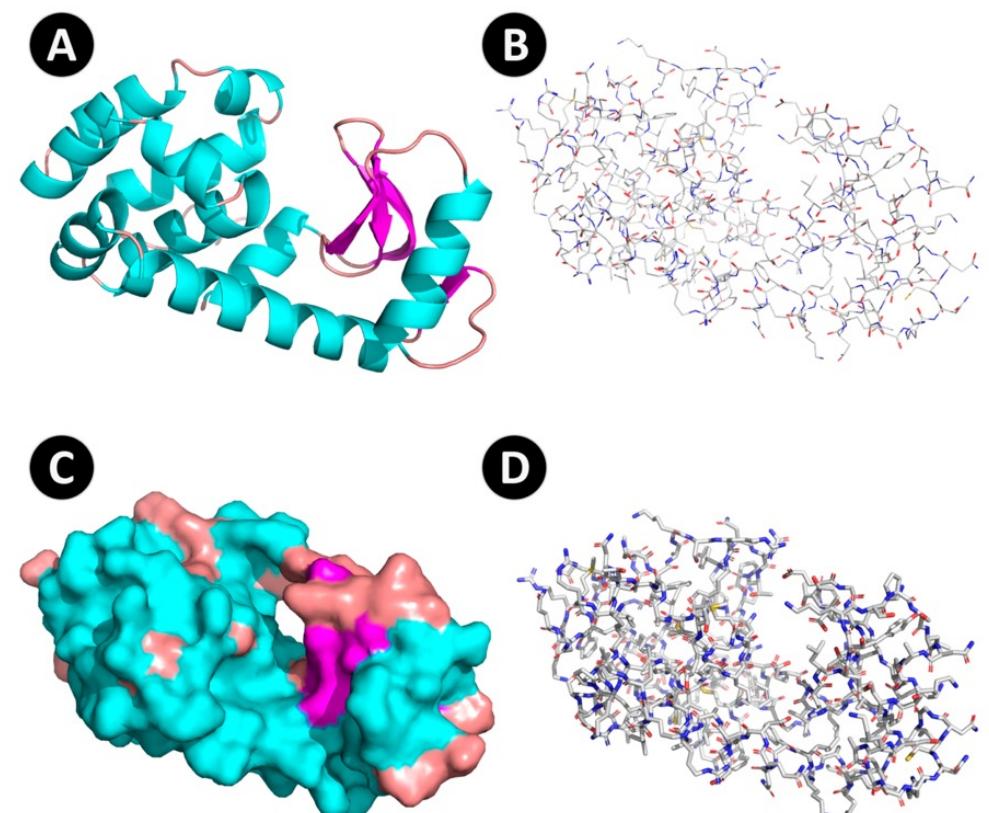
Atom	Amino acid	Chain identifier		X	Y	Z	Occupancy	B-factor
		Residue number						
ATOM	1	N	SER A 139	22.479	-4.111	18.945	1.00	34.34
ATOM	2	CA	SER A 139	22.123	-2.660	18.772	1.00	36.15
ATOM	3	C	SER A 139	21.917	-2.021	20.148	1.00	35.23
ATOM	4	O	SER A 139	21.829	-2.724	21.148	1.00	37.66
ATOM	5	CB	SER A 139	20.891	-2.485	17.863	1.00	36.40
ATOM	6	OG	SER A 139	19.742	-3.154	18.362	1.00	43.71
ATOM	7	N	CYS A 140	21.832	-0.699	20.188	1.00	34.49
ATOM	8	CA	CYS A 140	21.716	0.023	21.450	1.00	36.17
ATOM	9	C	CYS A 140	20.371	0.642	21.738	1.00	37.47
ATOM	10	O	CYS A 140	19.731	1.189	20.844	1.00	36.62
ATOM	11	CB	CYS A 140	22.749	1.139	21.490	1.00	35.87
ATOM	12	SG	CYS A 140	24.467	0.525	21.122	1.00	36.23
ATOM	13	N	PRO A 141	20.006	0.713	23.028	1.00	38.50
ATOM	14	CA	PRO A 141	18.735	1.279	23.481	1.00	38.39
ATOM	15	C	PRO A 141	18.517	2.743	23.086	1.00	39.20
ATOM	16	O	PRO A 141	19.407	3.590	23.232	1.00	38.41
ATOM	17	CB	PRO A 141	18.809	1.093	25.001	1.00	39.58
ATOM	18	CG	PRO A 141	20.284	1.177	25.286	1.00	38.94
ATOM	19	CD	PRO A 141	20.849	0.333	24.181	1.00	39.16
ATOM	20	N	SER A 142	17.342	3.006	22.515	1.00	37.85
ATOM	21	CA	SER A 142	16.939	4.352	22.097	1.00	39.22
ATOM	22	C	SER A 142	17.444	4.856	20.744	1.00	34.22
ATOM	23	O	SER A 142	17.045	5.945	20.316	1.00	31.72
ATOM	24	CB	SER A 142	17.288	5.385	23.185	1.00	40.67
ATOM	25	OG	SER A 142	16.498	5.192	24.346	1.00	48.68



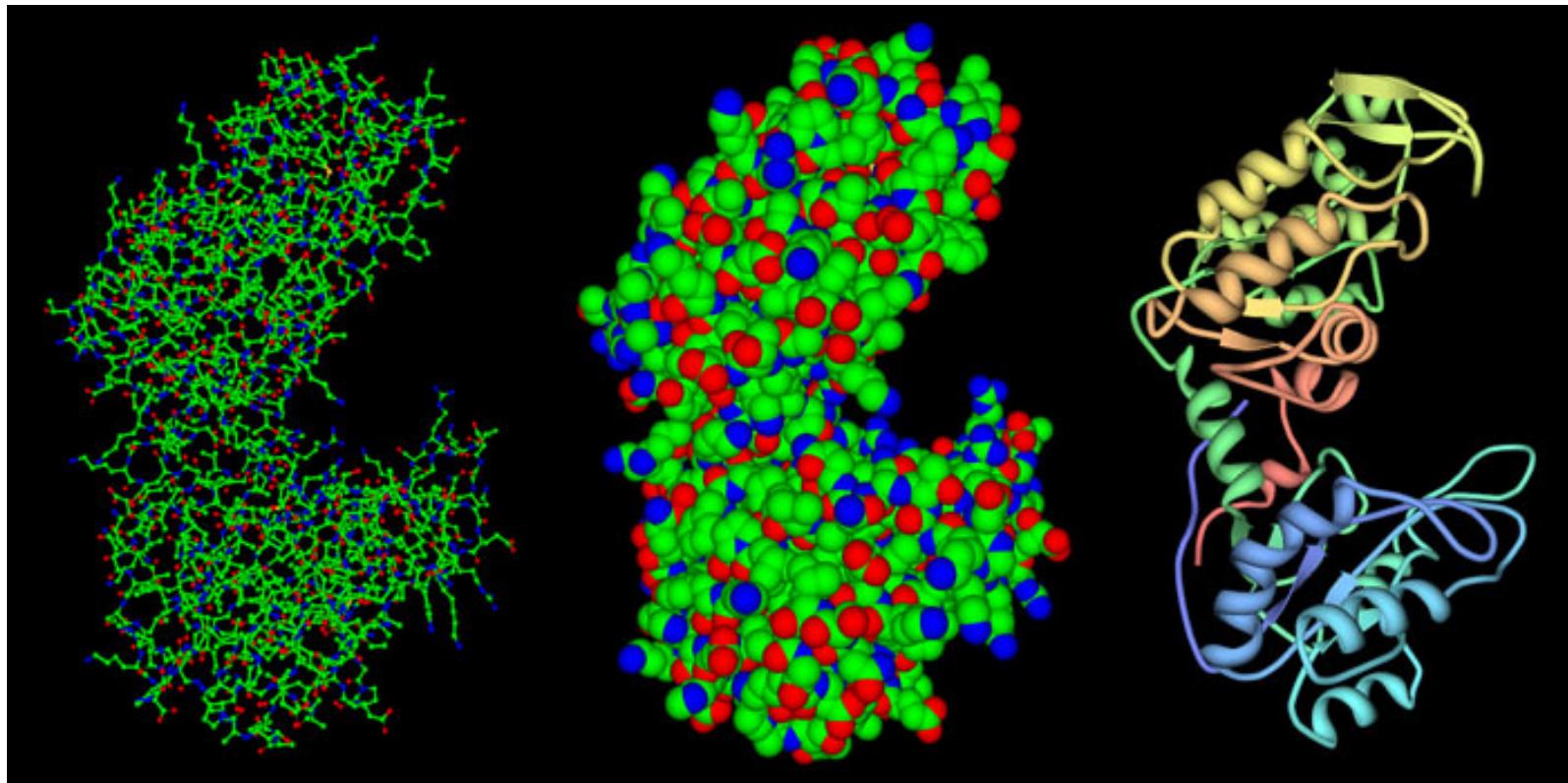
**PDB Identification Code (PDB ID)**, unique for each structure, is a 4-character alphanumeric identifier  
e.g. [1AOX](#)

# Why do we need visualization?

- We depend on visual information!
- Visualization programs enable us to
  - Understand structures in 3D
  - Study protein interactions in detail
  - Understand the function of a protein
  - Prepare descriptive figures



# Visualization styles



Wireframe  
Ball and Stick

Space filling  
or  
Surface

Cartoon  
Ribbon diagram

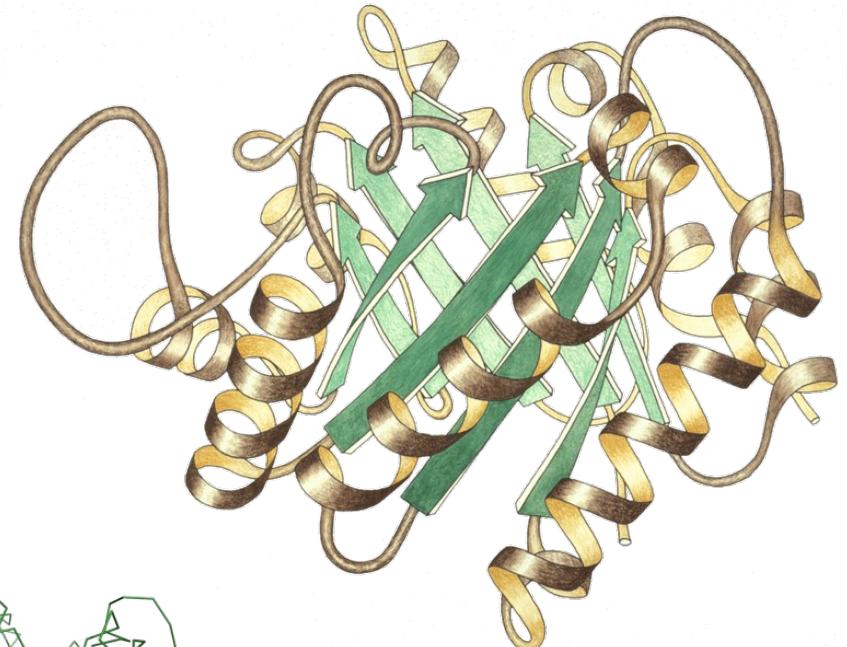
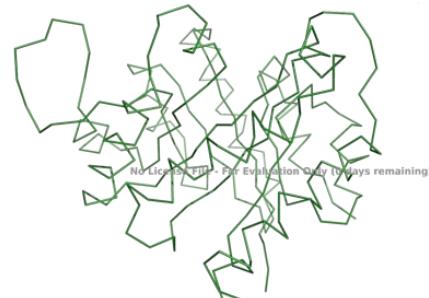
Pdb-101

# Visualization styles: Cartoon or ribbon

- Highlights secondary structures and folds
- **Cartoon** draws secondary structures as arrows/helices
- **Ribbon** is a CA trace, which draws a line from the C $\alpha$  atom on each amino acid to the next → traces the fold



PDB ID 1TIM in ribbon

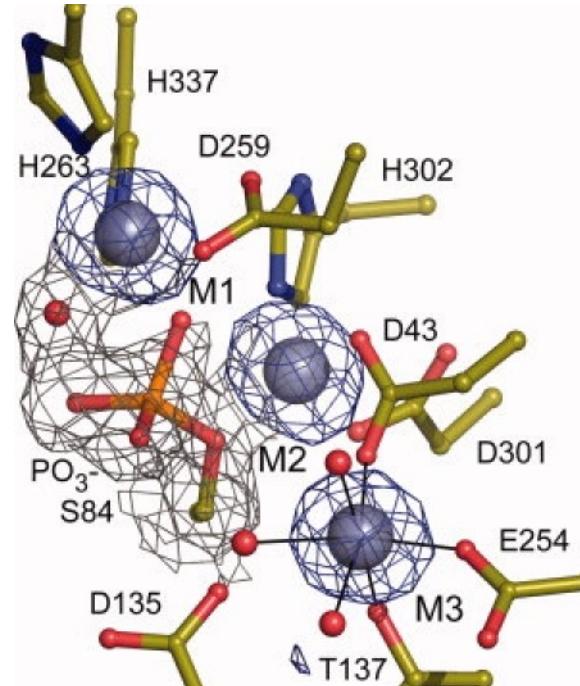


TIM-barrel  
Jane Richardson  
Wikimedia Commons

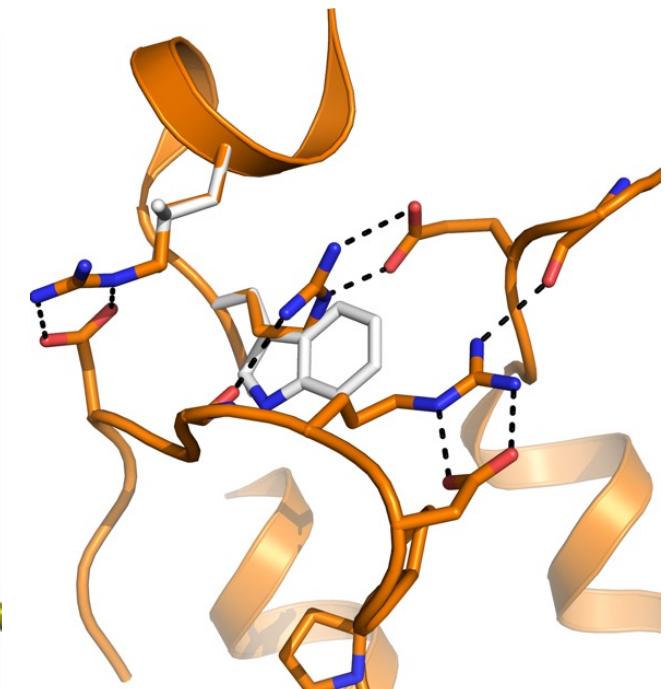
# Visualization styles: Wireframe – Ball and Stick

Lines, sticks and ball-and-sticks

- For displaying details
  - Amino acids
  - Bonds (interactions)
  - All atoms shown in the same size!
- Waters and other small molecules as spheres relative to their atom size



Zn<sup>2+</sup> coordination and electron density map at alkaline phosphatase active site  
Koutsilis 2010, doi: 10.1002/pro.284.

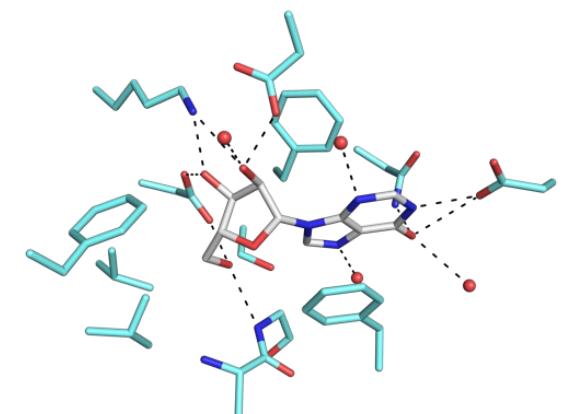
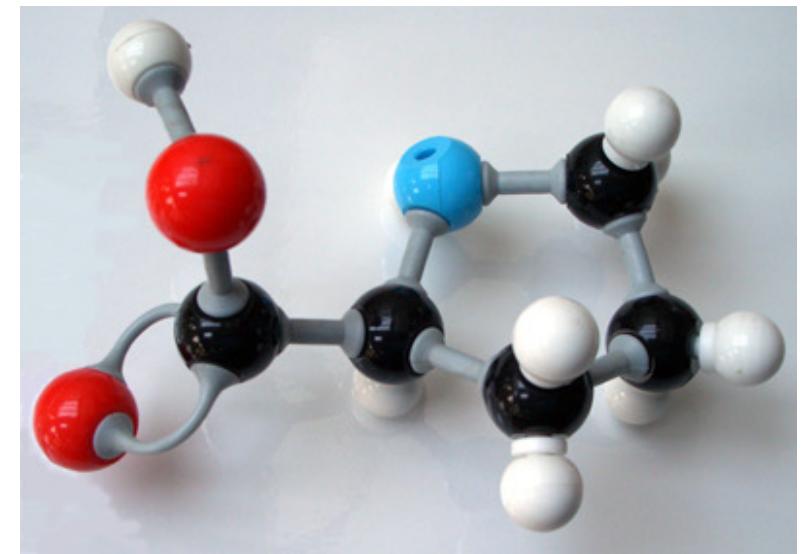


Inherited α-mannosidase mutations. Change of amino acid from R to W destroys interactions in human lysosomal α-mannosidase.  
Heikinheimo, unpublished

# Visualization styles: Colouring

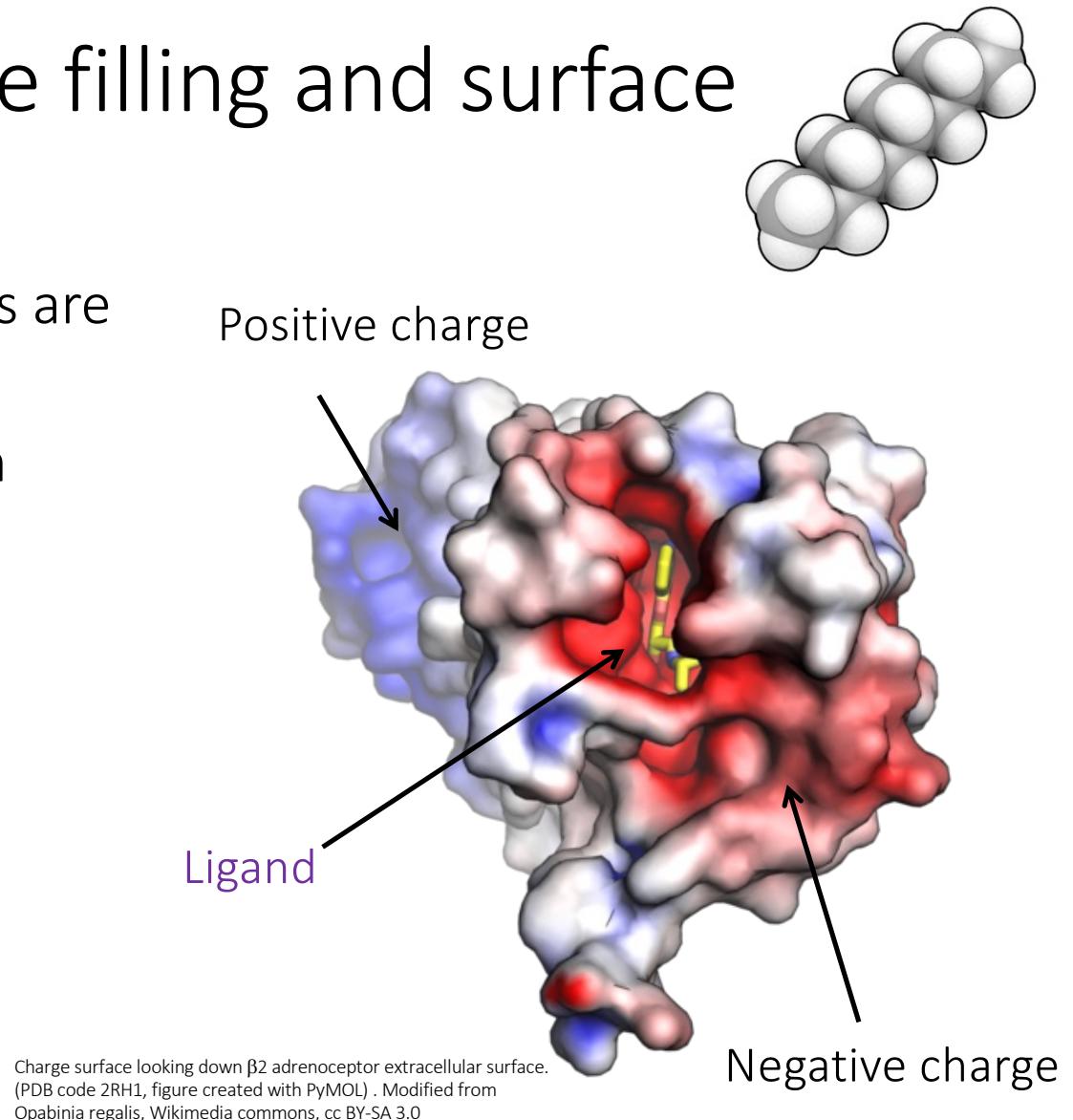
- Typical colour convention for most common atoms
  - Black for carbon
  - Blue for nitrogen
  - Red for oxygen
- White for hydrogen
- Deep yellow for sulphur
- Purple for phosphorus
- Light, medium, medium dark, and dark green for the halogens (F, Cl, Br, I)
- Silver for metals (Co, Fe, Ni, Cu)

A plastic ball-and-stick model of proline



# Visualization styles: Space filling and surface

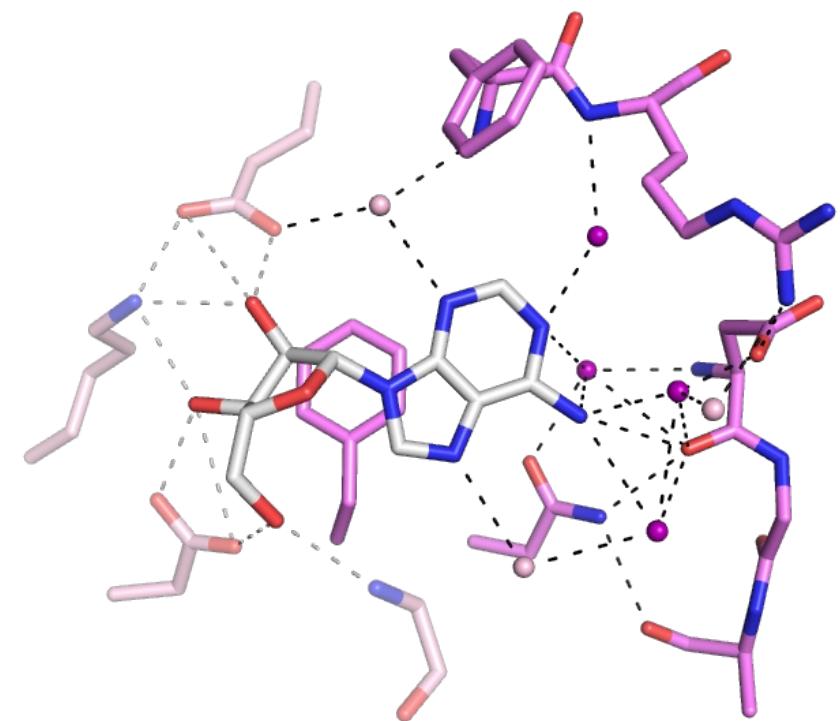
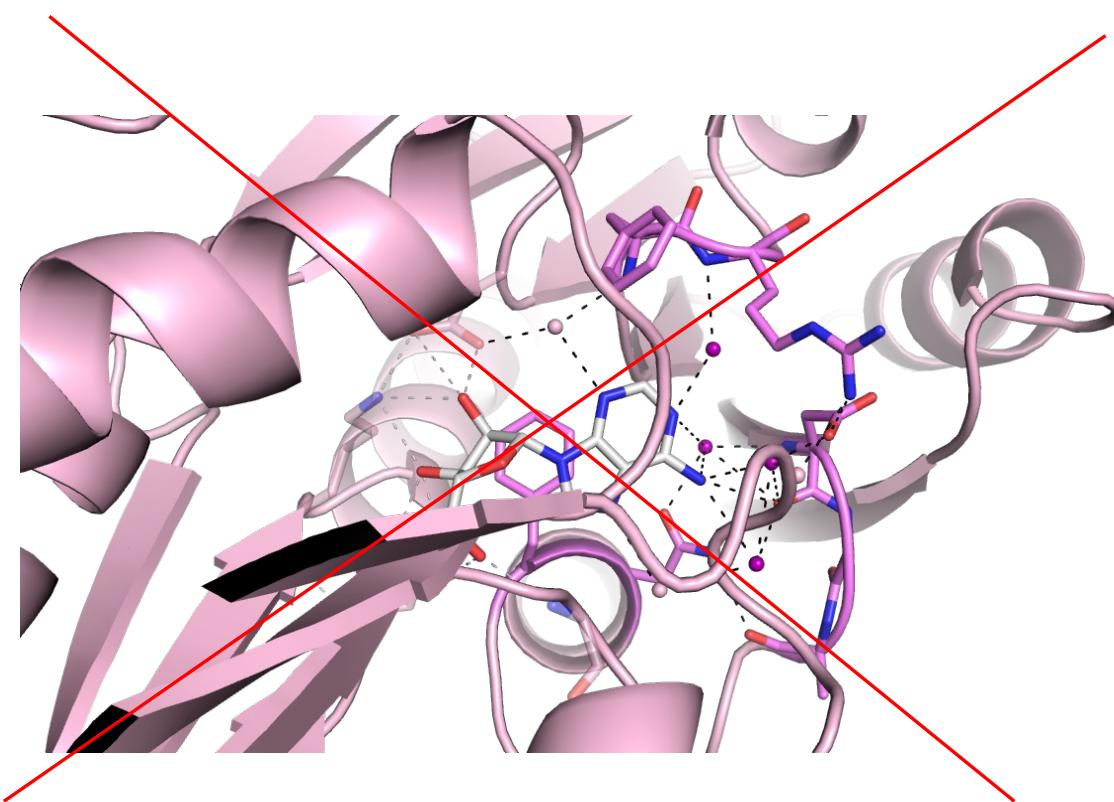
- In space-filling representation, atoms are shown as spheres
  - Radius is *proportional* to atom radius
- Surfaces
  - Surface charges
  - Good for displaying volume, cavities, shape etc.



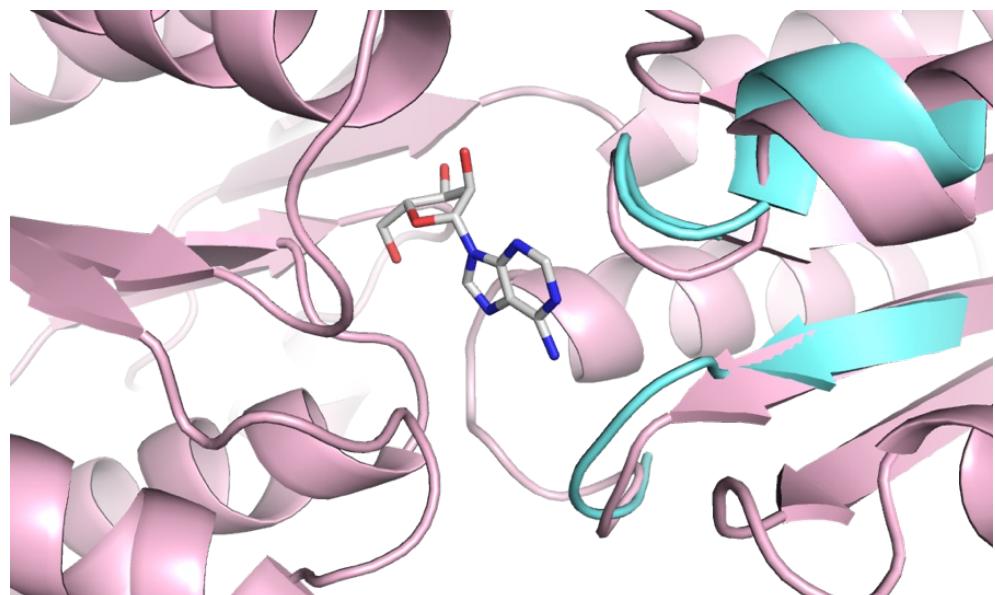
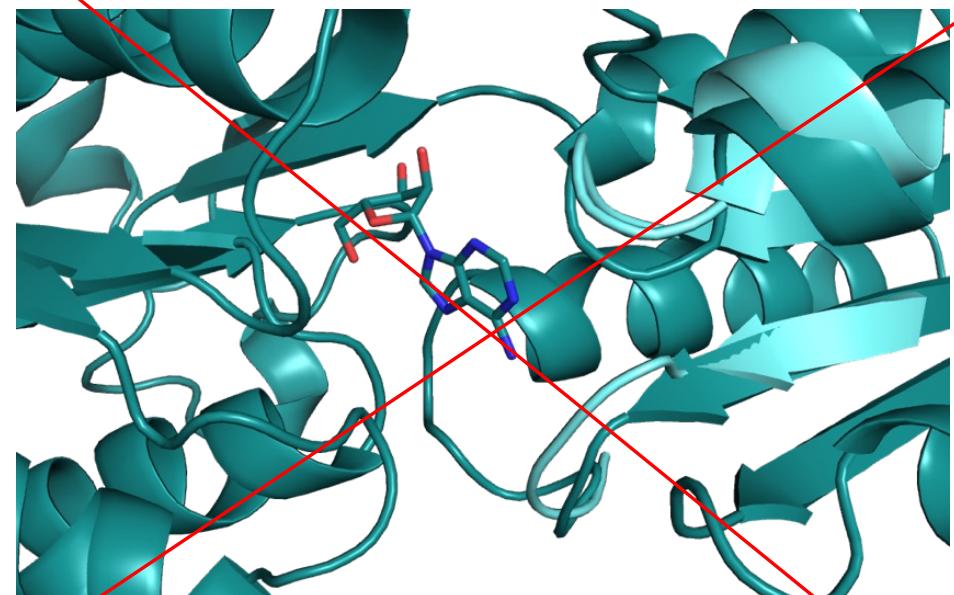
# How to make a good protein figure:

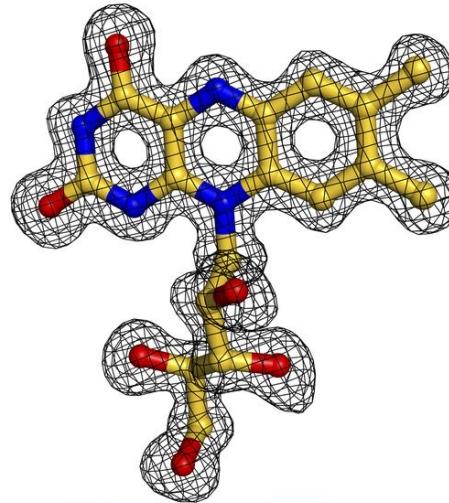
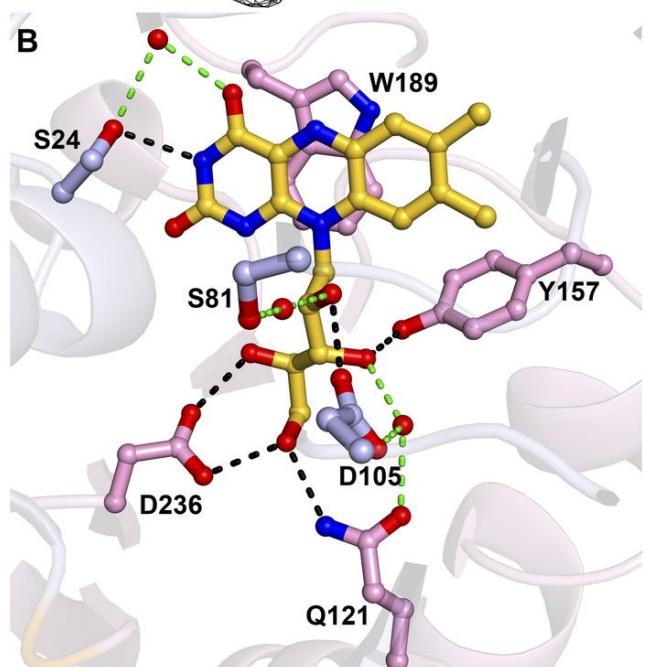
1. Make it **clear and simple!**
  - Highlight the most important things and **remove distractions**
2. Use colors, labels, arrows etc. wisely
  - Select **contrasting colors** to highlight differences
  - Label the most important things (e.g. catalytic residues, secondary structures involved in interactions)
3. Choose a format that is suitable for what you want to show
  - E.g. cartoon, sticks and surface representations can be used to highlight different aspects of a protein
4. Write a **clear and descriptive figure legend!**
  - All colors, signs etc. should be clearly explained
  - PDB structures used should be referenced

# Example: Showing ligand-binding interactions



# Example: Coloring



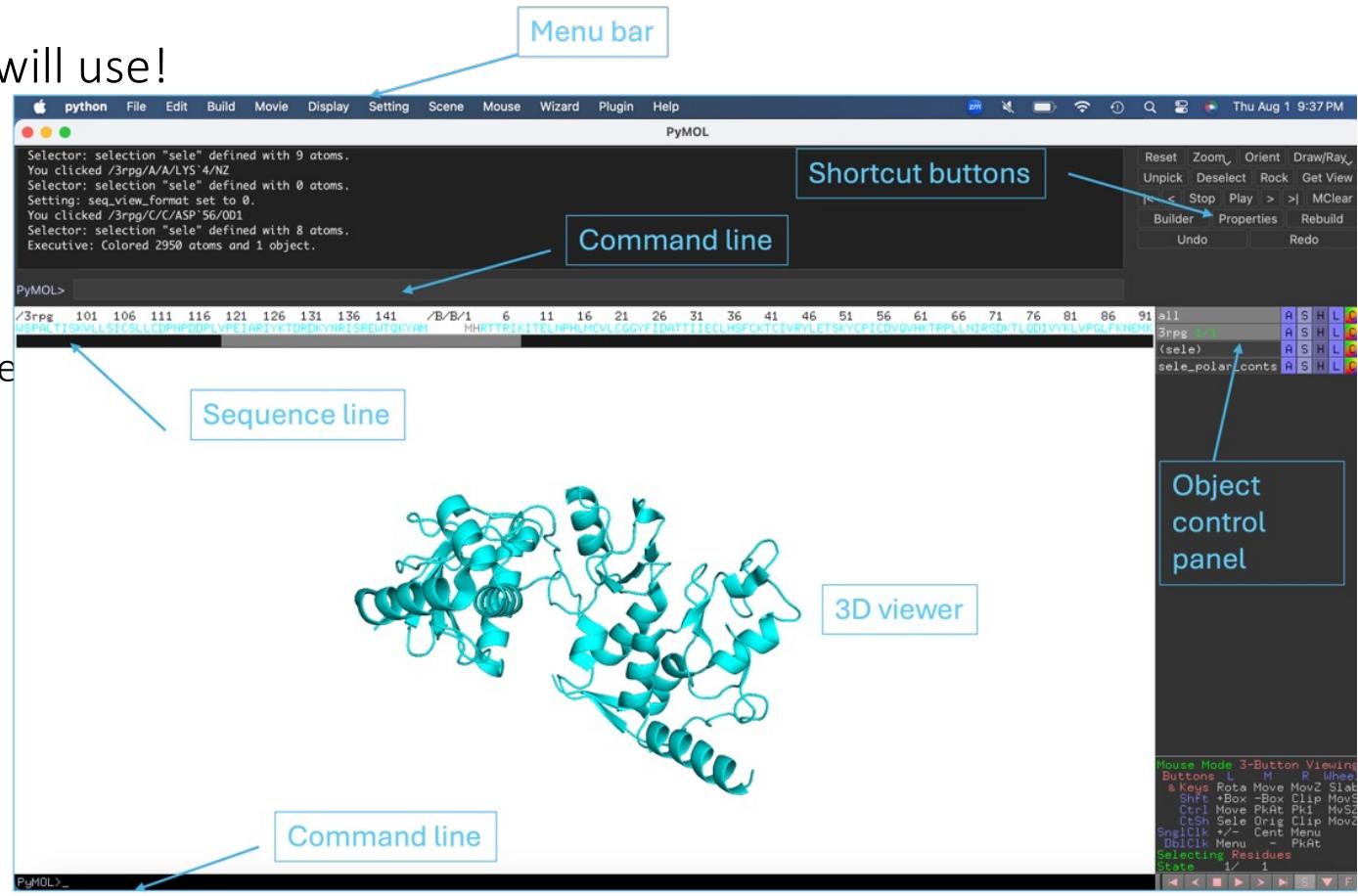
**A****B**

## Example: Figure legend

**FIG 6** Closeup views of the riboflavin bound to TP0298. (A) Electron density for the riboflavin. Shown is an  $mF_o - DF_c$  omit map superposed on the final refined coordinates of the TP0298 model. The map is contoured at the 3  $\sigma$  level. (B) Contacts between TP0298 and riboflavin. Apparent hydrogen bonds between the riboflavin and TP0298 are shown as black dashes, and apparent hydrogen bonds that are involved in water-mediated contacts between TP0298 and the ligand are shown as green dashes; all distances between atoms shown in this figure are less than 3.0 Å. Atoms are colored as described for riboflavin in Fig. 5, except carbon atoms from the N lobe are colored light blue, and those from the C lobe are pink. Protein secondary structure is shown semitransparently for clarity.

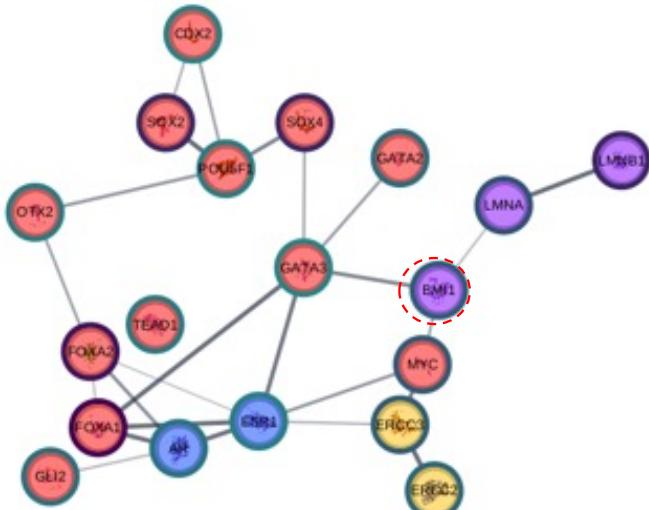
# PyMOL - The program we will use!

- Created by Warren DeLano
  - First release in **2000**
  - Python script, OpenGL, open source
  - Since 2010 supported and distributed by Schrödinger, Inc.
- **Possibility to save sessions – Do it!**
- High quality rendering, movies etc.
- Automatic retrieval of molecules from PDB
- Possibility to use **scripts** to do complicated tasks
- Allow user to fully control and adjust the items displayed



<https://en.wikipedia.org/wiki/PyMOL>

# Exercise – To get familiar with Pymol



## PyMOL exercise

### Bmi1/Ring1b-UbcH5c complex structure

You will look at the 3D structures of three interacting proteins including RING2 (chain C) and BMI1 (chain B), both being part of human polycomb repressive complex 1-like, which plays a role in histone and gene regulation as described by Wang et al. (2004) in “Role of histone H2A ubiquitination in polycomb silencing”. These two proteins interact as a heterodimer with the E2 enzyme UbcH5c (chain A) and the complex structure has been crystallized by Bentley et al. (2011) “Recognition of UbcH5c and the nucleosome by the BMI1/RING1b ubiquitin ligase complex” (PDB ID 3RPG). Now you will analyze how these proteins interact with each other.

Link to the article in PDB

## Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex

Matthew L Bentley<sup>1</sup>, Jacob E Corn<sup>1</sup>,  
Ken C Dong<sup>2</sup>, Qui Phung<sup>3</sup>,  
Tommy K Cheung<sup>3</sup> and Andrea G Cochran<sup>1,\*</sup>

<sup>1</sup>Department of Early Discovery Biochemistry, Genentech Research and Early Development, South San Francisco, CA, USA, <sup>2</sup>Department of Structural Biology, Genentech Research and Early Development, South San Francisco, CA, USA and <sup>3</sup>Department of Protein Chemistry, Genentech Research and Early Development, South San Francisco, CA, USA

The Polycomb repressive complex 1 (PRC1) mediates gene silencing, in part by monoubiquitination of histone H2A on lysine 119 (uH2A). Bmi1 and Ring1b are critical components of PRC1 that heterodimerize via their N-terminal RING domains to form an active E3 ubiquitin ligase. We have determined the crystal structure of a complex between the Bmi1/Ring1b RING–RING heterodimer and the E2 enzyme UbcH5c and find that UbcH5c interacts with Ring1b only, in a manner fairly typical of E2–E3 interactions. However, we further show that the Bmi1/Ring1b RING domains bind directly to duplex DNA through a basic surface patch unique to the Bmi1/Ring1b RING–RING dimer. Mutation of residues on this interaction surface leads to a loss of H2A ubiquitination activity. Computational modelling of the interface between Bmi1/Ring1b–UbcH5c and the nucleosome suggests that Bmi1/Ring1b interacts with both nucleosomal DNA and an acidic patch on histone H4 to achieve specific monoubiquitination of H2A. Our results point to a novel mechanism of substrate recognition, and control of product formation, by Bmi1/Ring1b.

# PDB Search for the structure

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

**PDB** 223,166 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures 3RPG Advanced Search | Browse Annotations Include CSM Help

PDB-101 wwpdb EMDDataResource NAKB wwPDB Foundation PDB-Dev

Structure Summary Structure Annotations Experiment Sequence Genome Versions Display Files Download Files Data API

### 3RPG

Bmi1/Ring1b-UbcH5c complex structure  
PDB DOI: <https://doi.org/10.2210/pdb3RPG/pdb>

Classification: LIGASE  
Organism(s): Homo sapiens  
Expression System: Escherichia coli  
Mutation(s): No

Deposited: 2011-04-26 Released: 2011-08-17  
Deposition Author(s): Bentley, M.L., Dong, K.C., Cochran, A.G.

Experimental Data Snapshot

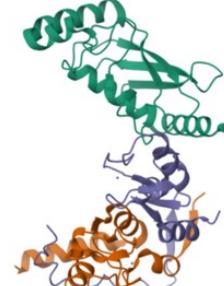
Method: X-RAY DIFFRACTION  
Resolution: 2.65 Å  
R-Value Free: 0.243  
R-Value Work: 0.217  
R-Value Observed: 0.219

Starting Models: experimental  
View more details

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree	11	0.237
Clashscore	10	11
Ramachandran outliers	0	0
Sidechain outliers	6%	0.6%
RSRZ outliers	9%	0.9%

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



Explore in 3D: Structure | Sequence Annotations | Electron Density | Validation Report | Ligand Interaction (ZN)

Global Symmetry: Asymmetric - C1  
Global Stoichiometry: Hetero 3-mer - A1B1C1



## 3 proteins in the complex structure: chain A, B and C

### Macromolecules

Find similar proteins by: Sequence (by identity cutoff) | 3D Structure

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
Ubiquitin-conjugating enzyme E2 D3	A	149	Homo sapiens	Mutation(s): 0 Gene Names: UBE2D3, UBC5C, UBC5 EC: 6.3.2.19	

### UniProt & NIH Common Fund Data Resources

Find proteins for P61077 (Homo sapiens)

Explore P61077

Go to UniProtKB: P61077

Entity ID: 2

Molecule	Chains	Sequence Length	Organism	Details	Image
Polycomb complex protein BMI-1	B	117	Homo sapiens	Mutation(s): 0 Gene Names: BMI1, PCGF4, RNF51	

### UniProt & NIH Common Fund Data Resources

Find proteins for P35226 (Homo sapiens)

Explore P35226

Go to UniProtKB: P35226

Entity ID: 3

Molecule	Chains	Sequence Length	Organism	Details	Image
E3 ubiquitin-protein ligase RING2	C	121	Homo sapiens	Mutation(s): 0 Gene Names: RNF2, BAP1, DING, HIP13, RING1B EC: 6.3.2	

### UniProt & NIH Common Fund Data Resources

Find proteins for Q99496 (Homo sapiens)

Explore Q99496

Go to UniProtKB: Q99496

Biological assembly 1 assigned by authors.

### Macromolecule Content

- Total Structure Weight: 44.26 kDa
- Atom Count: 2,912
- Modelled Residue Count: 349
- Deposited Residue Count: 387
- Unique protein chains: 3

### Literature

Download Primary Citation ▾

Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex.

Bentley, M.L., Corn, J.E., Dong, K.C., Phung, Q., Cheung, T.K., Cochran, A.G.

(2011) EMBO J 30: 3285-3297

PubMed: 21772249 Search on PubMed Search on PubMed Central

DOI: <https://doi.org/10.1038/emboj.2011.243>

Primary Citation of Related Structures:

3RPG

### PubMed Abstract:

The Polycomb repressive complex 1 (PRC1) mediates gene silencing, in part by monoubiquitination of histone H2A on lysine 119 (H2A). Bmi1 and Ring1b are critical components of PRC1 that heterodimerize via their N-terminal RING domains to form an active E3 ubiquitin ligase. We have determined the crystal structure of a complex...

View More

### Organizational Affiliation:

Department of Early Discovery Biochemistry, Genentech Research and Early Development, South San Francisco, CA, USA.

Small Molecules

Ligands

1 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
ZN Query on ZN	D [auth B], E [auth B], F [auth C], G [auth C]	ZINC ION Zn PTFCDOFLQPIGS-UHFFFAOYSA-N		Zn <sup>2+</sup>

Download Ideal Coordinates CCD File

Download Instance Coordinates

Interactions ▾

Interactions & Density ▾

# Pymol exercise and Group exercise

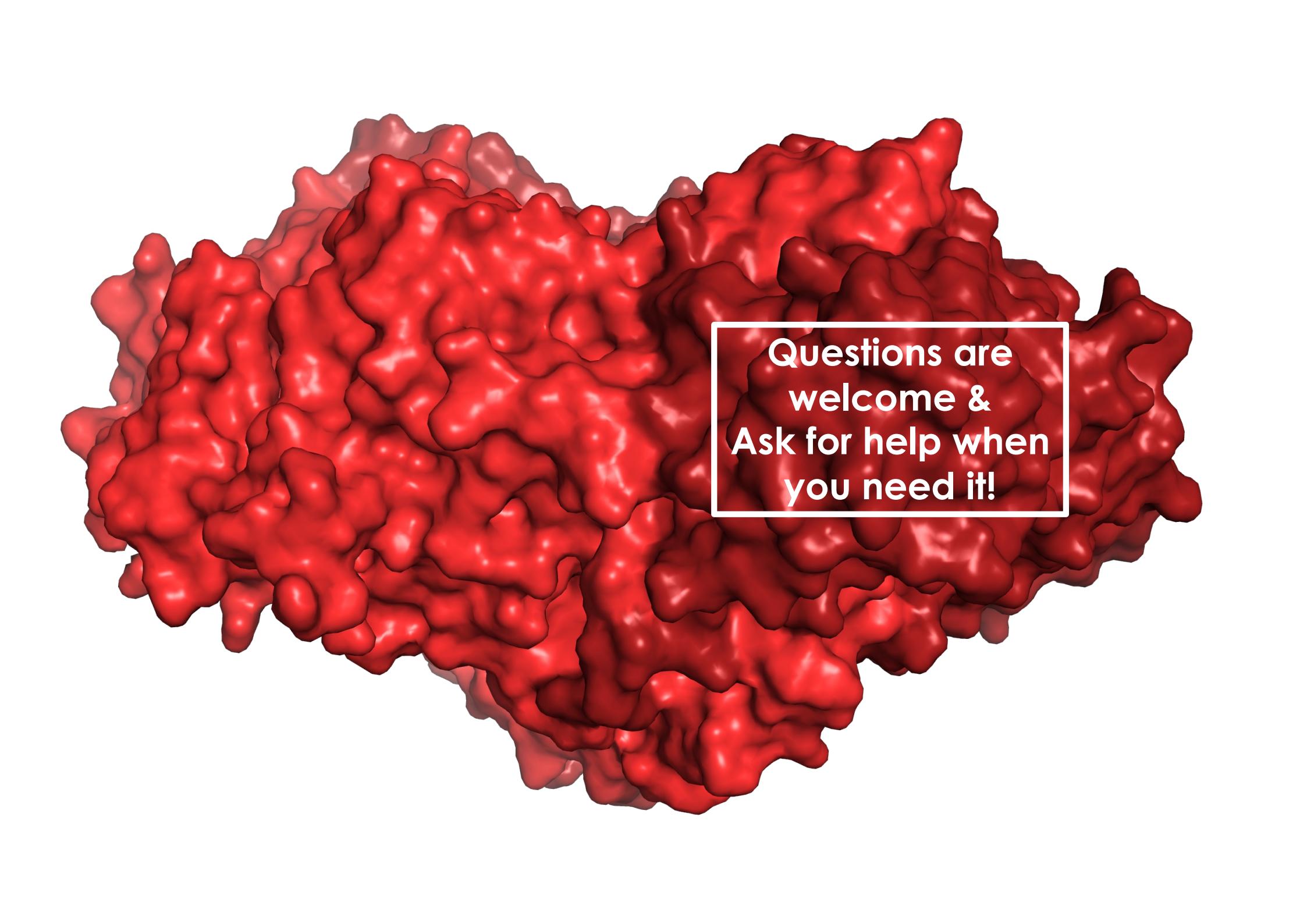
1. Individual Pymol exercise: Everyone in the group will do the exercise to learn how to use Pymol

- Have you installed PyMol on your computer?
- Save the license file from GitHub
- Open Pymol and read the license file in Pymol
- Follow the detailed information in the Pymol exercise instructions (GitHub)

2. Group exercise for the report:

In the group:

- Discuss and select one of these alternative group projects:
  1. Continue working with the same complex structure (Bentley et al. article in the GitHub)
  2. OR you may ask us for another protein complex to work with if you want to work e.g. DNA-binding protein
  3. OR work with the AlphaFold models you have earlier created
- Analyze the complex structure and get familiar with the proteins using the article linked in the PDB.
- Design what kind of figures and figure legends you will prepare to explain how the protein complex functions
- Submit figures and figure legends in the report



**Questions are  
welcome &  
Ask for help when  
you need it!**