

# The Protein Data Bank (in Europe)

An introduction to protein structure and the PDB

Genevieve L. Evans

Structural Biologist  
& Scientific Database  
Curator

 **PDB**e  
Protein Data Bank in Europe



[pdbhelp@ebi.ac.uk](mailto:pdbhelp@ebi.ac.uk)



[proteindatabank](https://www.facebook.com/proteindatabank)



[@PDBEurope](https://twitter.com/PDBEurope)



[proteindatabank](https://www.youtube.com/proteindatabank)



[pdbeurope](https://www.instagram.com/pdbeurope)



[pdbart](https://www.pinterest.com/pdbart)

# What is the PDB?

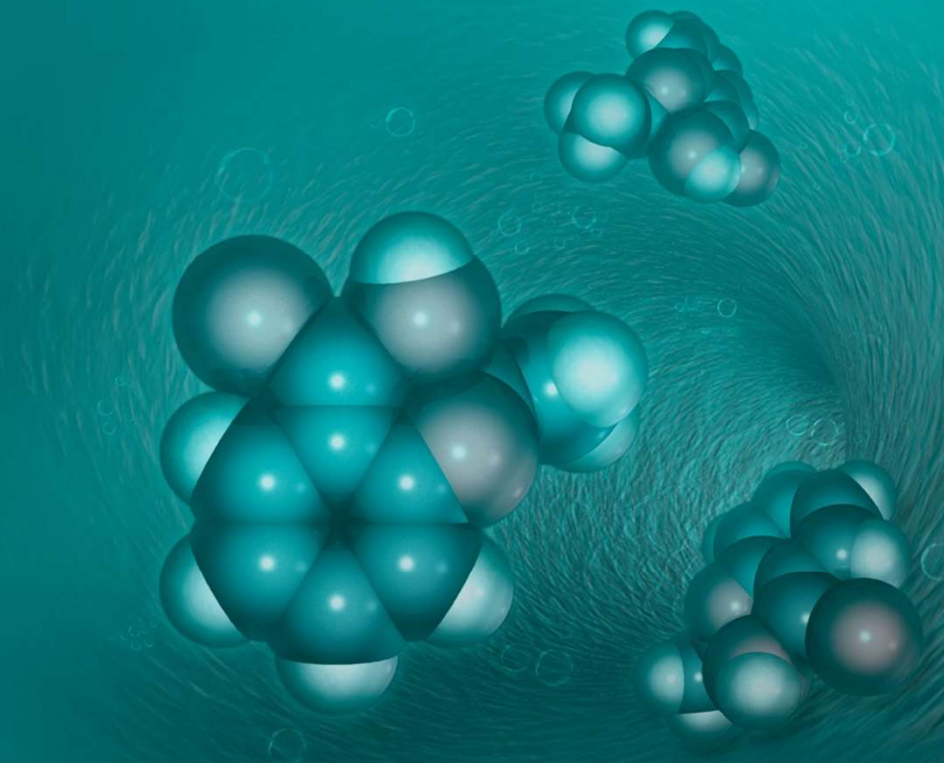
<https://www.menti.com/alpvu5wen5e4>

Join code: 8419 3777



 **PDB**e  
Protein Data Bank in Europe

EMBL-EBI 



# More about PDBe tools and services

**<https://www.menti.com/alpvu5wen5e4>**

Go to [PDBe.org/training](https://pdbe.org/training) > tutorials

**Join code: 8419 3777**

- Links to EBI train online tutorials for PDBe services
  - From basic introductory tutorials to advanced tools like PDBeFold
  - <https://www.ebi.ac.uk/pdbe/online-tutorials>
- Webinars available on our YouTube channel
  - Walk-throughs of specific features on PDBe and PDBe-KB (KnowledgeBase) webpages
  - <https://www.youtube.com/@ProteinDataBank>



# API - Application Programming Interface

<https://www.menti.com/alpvu5wen5e4>

Join code: 8419 3777

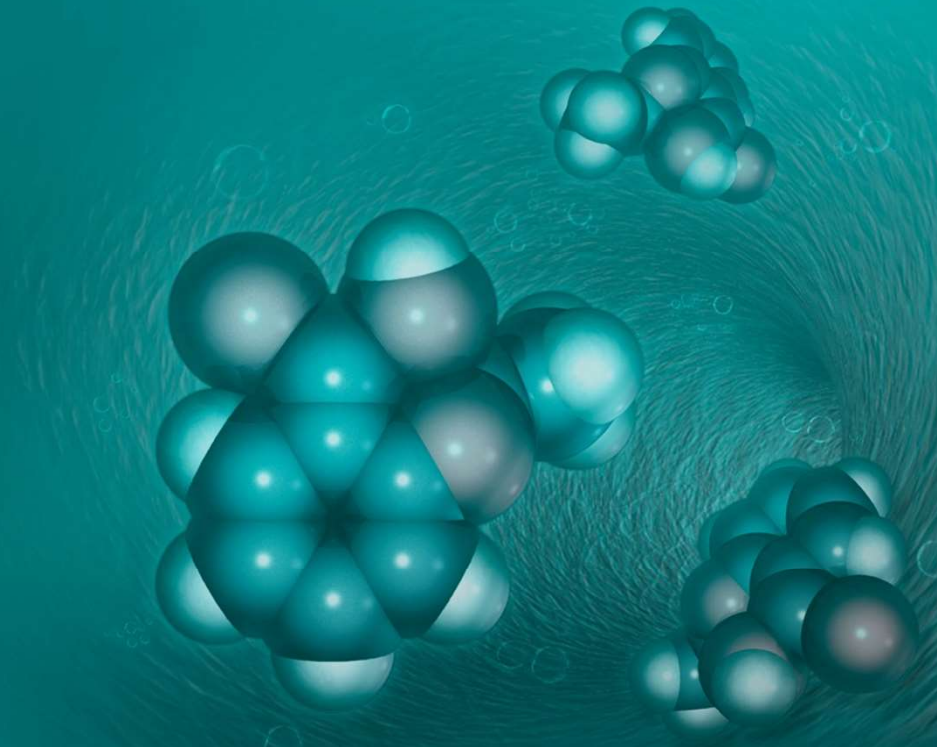
- Programmatic access webinar series
  - Six-part series ranging from basics in PDBe programmatic access to advanced visualisation
  - Visit [pdbeurope.github.io/api-webinars](https://pdbeurope.github.io/api-webinars) for videos and tutorials
- API 'start page'
  - Visit: <https://www.ebi.ac.uk/pdbe/pdbe-rest-api>



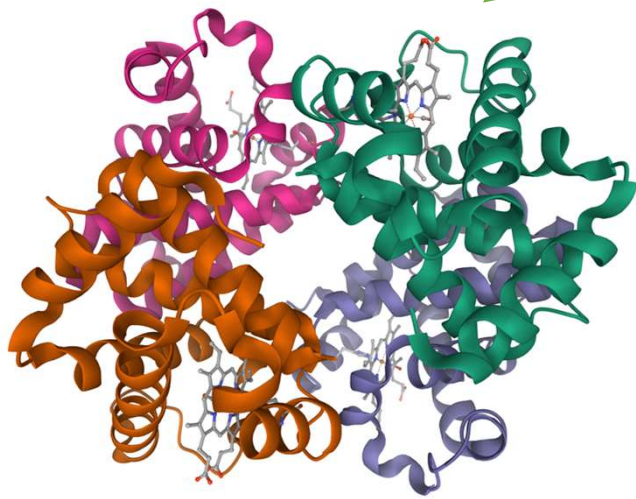
# What is the PDB?

 **PDB**e  
Protein Data Bank in Europe

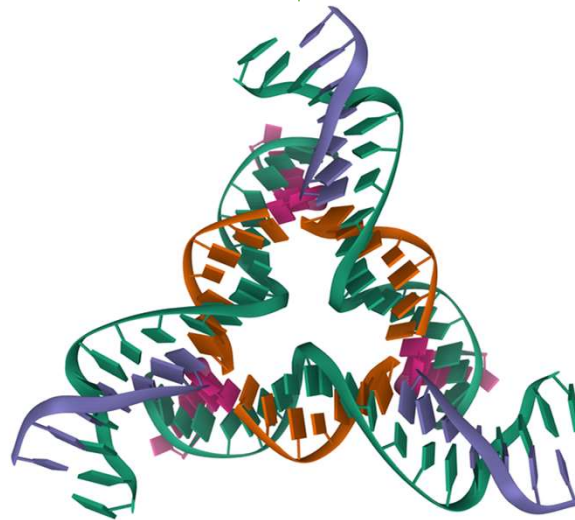
EMBL-EBI 



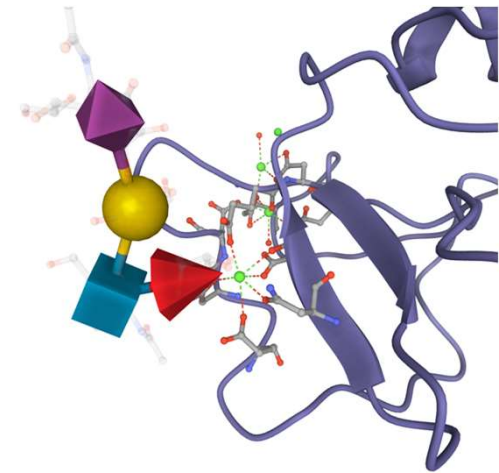
The Protein Data Bank (PDB) is an archive of experimentally determined 3-dimensional structures of biological macromolecules



Proteins

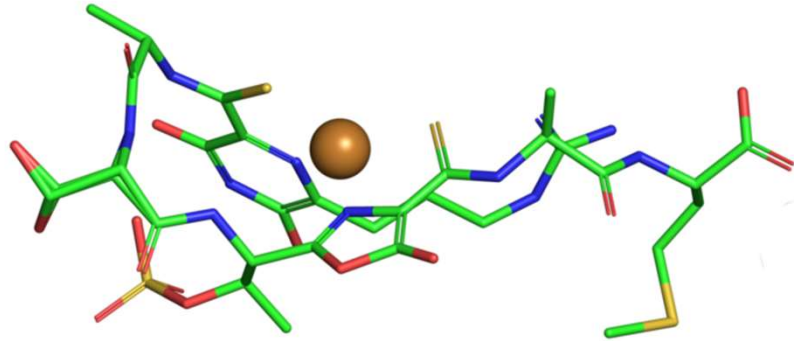


DNA/RNA

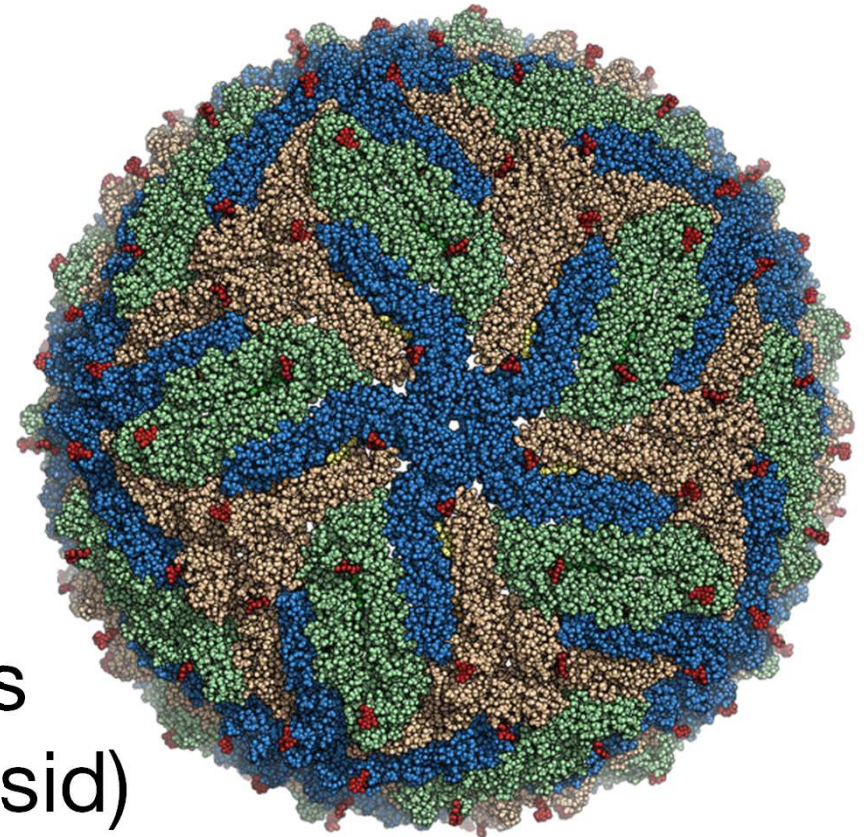


Bound ligands

Varying sizes of structure - can be very large!

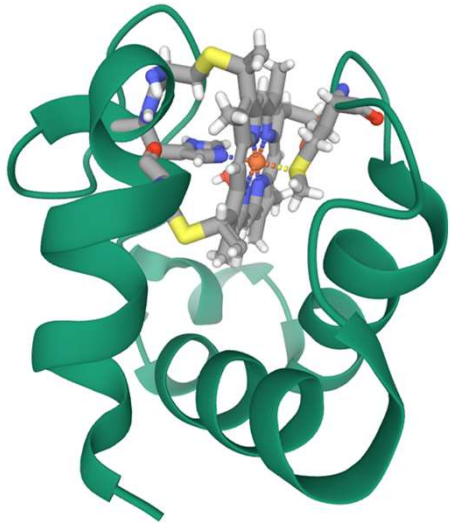


Copper scavenger  
<1kDa



Zika virus  
4 MDa (capsid)

# A “PDB code” refers to a structure



Deposit  
in the  
PDB ↓

**PDB code**

Described in a paper  
(or maybe not published) →



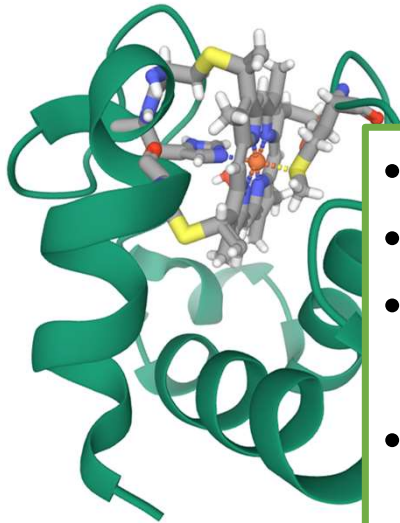
Referenced in the paper →



**nature** International weekly journal of science  
Menu Advanced search Search  
research > letters > article  
NATURE | LETTER  
Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB  
Manuel A. Ortega, Yue Hao, Qi Zhang, Mark C. Walker, Wilfred A. van der Donk & Satish K. Nair  
Primary accessions  
Protein Data Bank  
4WD9 4WD9



## A “PDB code” refers to a structure



Deposit  
in the  
PDB ↓

**PDB code**

- Unique code, currently 4 characters
- Identifies the data within the PDB archive
- Always starts with a number
- e.g. 2ins, 4xyz, 2f48  
→ pdb\_00002ins, pdb\_00004xyz,  
pdb\_00002f48

More information:

<https://www.wwpdb.org/documentation/new-format-for-pdb-ids>

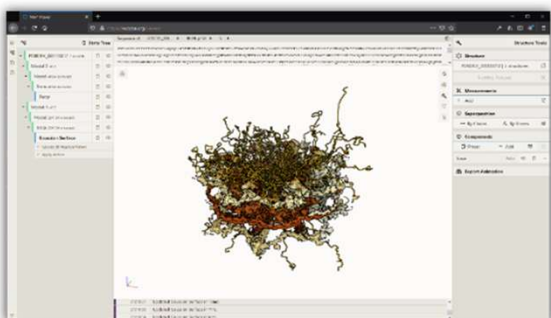


of the tRNA-dependent  
B  
alker, Wilfred A. van der Donk & Satish

# Mol\* website - molstar.org



Mol\* (*/ˈmolstar/*) is a modern web-based open-source toolkit for visualisation and analysis of large-scale molecular data



[Open Mol\\* Viewer](#)

[Viewer Documentation](#)

[Issues & Feedback](#)

High-performance graphics and data handling of the Mol\* Viewer allow users to simultaneously visualise up to hundreds of (superimposed) protein structures, play molecular dynamics trajectories, render cell-level models at atomic detail with tens of millions of atoms, or display huge models obtained by I/HM such as the Nuclear Pore Complex.



Select/Rotate



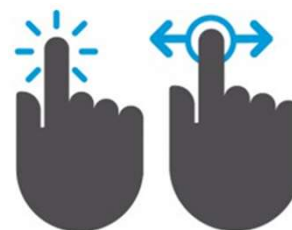
Move



Zoom



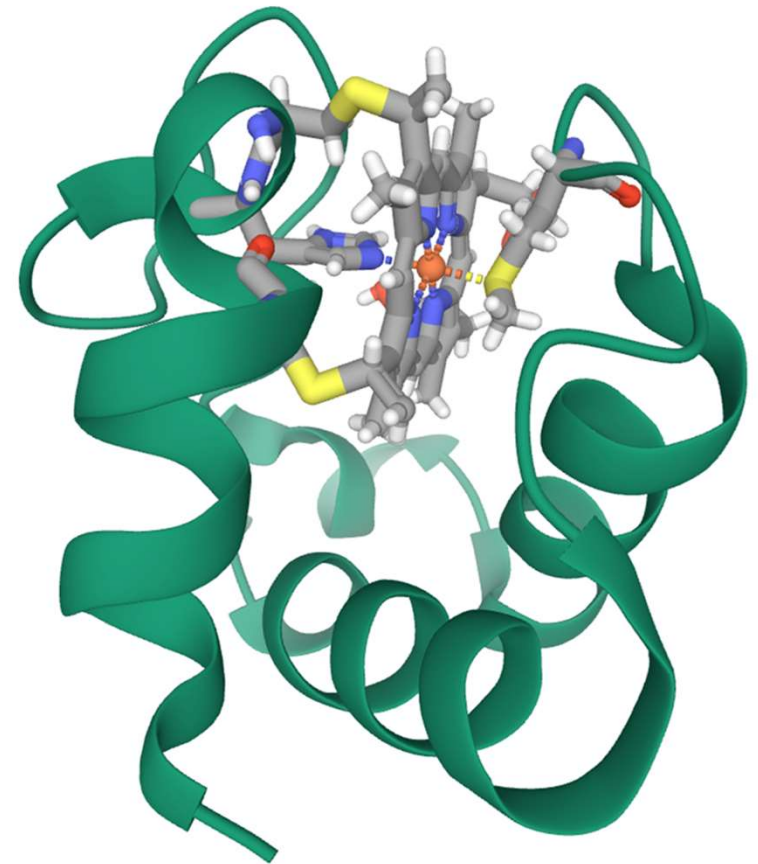
Slab



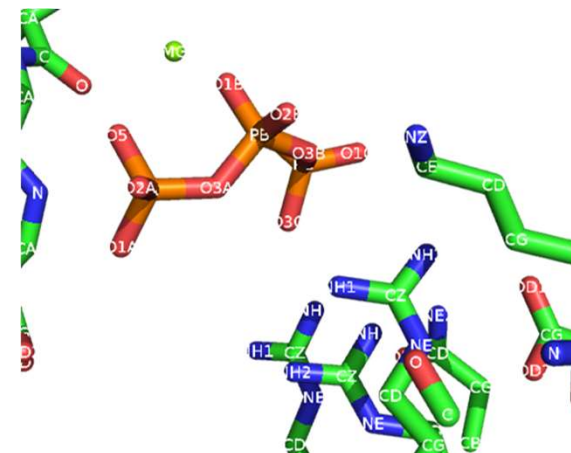
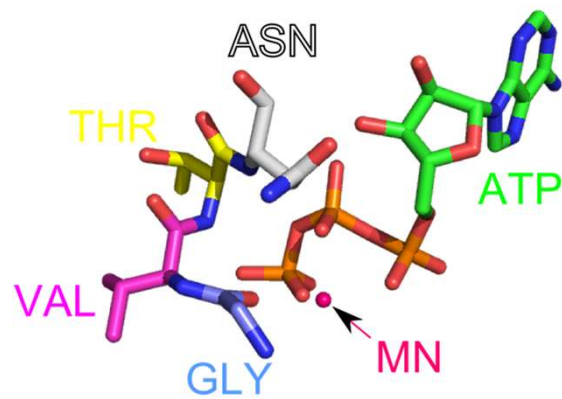
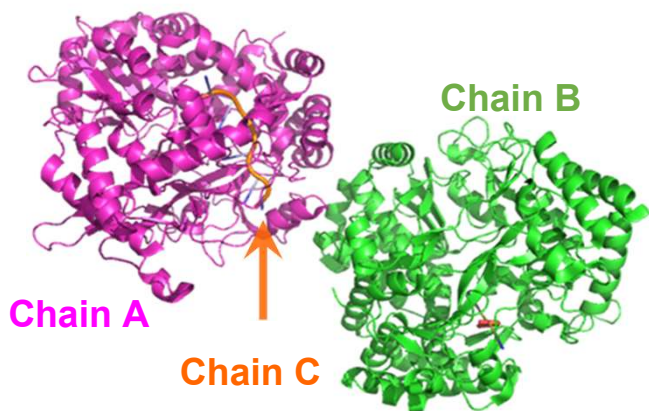
<https://molstar.org/viewer/>

# Ligands

- In the PDB:  
drugs, metal ions, cofactors, etc  
= ligands  
= small molecules binding to protein
- Each unique ligand has its own ID code
  - Currently 3 characters, *e.g.* AIN (aspirin)  
→ IDs with 5 characters will exist in future
  - also called: Chemical component (CC) id
  - Helps find it in the archive
- PDBeChem web pages contain ligand-specific info:  
<https://www.ebi.ac.uk/pdbe-srv/pdbechem/>



# A hierarchy in each PDB entry



## Chains

Macromolecules are referred to as 'chains'

## Residues

Chains have **residues** with three letter code for chemical type (but some are not 3 letters!) and unique number for where it is in the chain

## Atoms

Residues have **atoms**  
Each has defined position in 3D space

# What does a PDB file look like?

A text file with fixed column width - Card legacy

'Chain' name	ATOM	2365	O	GLU	S	271	-11.042	-31.638	22.562	1.00	13.19	O
	ATOM	2366	CB	GLU	S	271	-9.351	-31.199	25.481	1.00	12.72	C
	ATOM	2367	CG	GLU	S	271	-10.019	-32.565	25.731	1.00	14.90	C
	ATOM	2368	CD	GLU	S	271	-10.069	-32.942	27.205	1.00	15.70	C
	ATOM	2369	OE1	GLU	S	271	-10.068	-34.150	27.487	1.00	21.99	O
	ATOM	2370	OE2	GLU	S	271	-10.101	-32.059	28.084	1.00	19.89	O
	ATOM	2371	N	ALA	S	272	-11.559	-29.817	23.813	1.00	11.35	N
	ATOM	2372	CA	ALA	S	272	-12.918	-29.744	23.288	1.00	10.72	C
	ATOM	2373	C	ALA	S	272	-12.958	-29.381	21.778	1.00	10.89	C
Residue name	ATOM	2374	O	ALA	S	272	-13.709	-29.992	21.005	1.00	10.98	O
	ATOM	2375	CB	ALA	S	272	-13.814	-28.804	24.129	1.00	11.26	C
	ATOM	2376	N	ALA	S	273	-12.102	-28.459	21.336	1.00	9.29	N
	ATOM	2377	CA	ALA	S	273	-12.087	-28.103	19.894	1.00	10.76	C
	ATOM	2378	C	ALA	S	273	-11.709	-29.274	18.936	1.00	10.55	C
	ATOM	2379	O	ALA	S	273	-12.179	-29.288	17.806	1.00	11.31	O
	ATOM	2380	CB	ALA	S	273	-11.222	-26.891	19.632	1.00	9.88	C
Residue number	ATOM	2381	N	ALA	S	274	-10.961	-30.295	19.389	1.00	12.12	N
	ATOM	2382	CA	ALA	S	274	-10.493	-31.383	18.449	1.00	14.23	C
	ATOM	2383	C	ALA	S	274	-11.100	-32.890	18.262	1.00	14.94	C
	ATOM	2384	O	ALA	S	274	-10.339	-33.762	18.453	1.00	13.37	O
	ATOM	2385	CB	ALA	S	274	-8.960	-31.501	18.670	1.00	13.94	C
	ATOM	2386	N	GLN	S	275	-12.304	-33.243	17.708	1.00	19.11	N
	ATOM	2387	CA	GLN	S	275	-12.815	-34.714	17.812	1.00	16.43	C
Atom name	ATOM	2388	C	GLN	S	275	-13.255	-35.682	16.572	1.00	17.45	C
	ATOM	2389	O	GLN	S	275	-13.460	-36.964	16.661	1.00	4.38	O
	ATOM	2390	CB	GLN	S	275	-13.871	-34.737	18.905	1.00	19.77	C
	ATOM	2391	CG	GLN	S	275	-14.310	-36.127	19.356	1.00	23.54	C
	HETATM	2396	ZN	ZN	S	278	-11.252	-10.370	14.483	1.00	28.39	ZN
	HETATM	2397	ZN	ZN	S	279	-10.199	-35.599	11.656	1.00	18.26	ZN
	HETATM	2398	ZN	ZN	S	280	16.091	-23.317	24.137	1.00	37.16	ZN
	HETATM	2399	ZN	ZN	S	281	2.562	-32.376	26.687	1.00	26.51	ZN

```

loop_
  _atom_site.group_PDB
  _atom_site.id
  _atom_site.type_symbol
  _atom_site.label_atom_id
  _atom_site.label_alt_id
  _atom_site.label_comp_id
  _atom_site.label_asym_id
  _atom_site.label_entity_id
  _atom_site.label_seq_id
  _atom_site.pdbx_PDB_ins_code
  _atom_site.Cartn_x
  _atom_site.Cartn_y
  _atom_site.Cartn_z
  _atom_site.occupancy
  _atom_site.B_iso_or_equiv
  _atom_site.Cartn_x_esd
  _atom_site.Cartn_y_esd
  _atom_site.Cartn_z_esd
  _atom_site.occupancy_esd
  _atom_site.B_iso_or_equiv_esd
  _atom_site.pdbx_formal_charge
  _atom_site.auth_seq_id
  _atom_site.auth_comp_id
  _atom_site.auth_asym_id
  _atom_site.auth_atom_id
  _atom_site.pdbx_PDB_model_num

```

mmCIF is the ‘master format’

It’s still a text file!

More (modern) computer readable

The fastest open-source mmCIF parser:  
GEMMI (<https://gemmi.readthedocs.io/en/latest/>)

```

ATOM 1  N N  . MET A 1 1  ? 4.940  -11.534  16.748  1.00 0.00 ? ? ? ? ? ? ? 1 MET A N  1
ATOM 2  C CA . MET A 1 1  ? 4.627  -10.082  16.703  1.00 0.00 ? ? ? ? ? ? ? 1 MET A CA  1
ATOM 3  C C  . MET A 1 1  ? 4.375  -9.616  15.273  1.00 0.00 ? ? ? ? ? ? ? 1 MET A C  1
ATOM 4  O O  . MET A 1 1  ? 4.604  -10.360  14.318  1.00 0.00 ? ? ? ? ? ? ? 1 MET A O  1
ATOM 5  C CB . MET A 1 1  ? 5.800  -9.312  17.310  1.00 0.00 ? ? ? ? ? ? ? 1 MET A CB  1
ATOM 6  C CG . MET A 1 1  ? 7.097  -9.468  16.535  1.00 0.00 ? ? ? ? ? ? ? 1 MET A CG  1
ATOM 7  S SD . MET A 1 1  ? 8.437  -8.471  17.215  1.00 0.00 ? ? ? ? ? ? ? 1 MET A SD  1
ATOM 8  C CE . MET A 1 1  ? 8.819  -9.388  18.706  1.00 0.00 ? ? ? ? ? ? ? 1 MET A CE  1
ATOM 9  H H1 . MET A 1 1  ? 5.323  -11.746  17.691  1.00 0.00 ? ? ? ? ? ? ? 1 MET A H1  1
ATOM 10 H H2 . MET A 1 1  ? 5.641  -11.730  16.006  1.00 0.00 ? ? ? ? ? ? ? 1 MET A H2  1
ATOM 11 H H3 . MET A 1 1  ? 4.056  -12.055  16.578  1.00 0.00 ? ? ? ? ? ? ? 1 MET A H3  1
ATOM 12 H HA . MET A 1 1  ? 3.740  -9.904  17.292  1.00 0.00 ? ? ? ? ? ? ? 1 MET A HA  1
ATOM 13 H HB2 . MET A 1 1  ? 5.548  -8.261  17.340  1.00 0.00 ? ? ? ? ? ? ? 1 MET A HB2  1

```

Coordinates of atom  
in space (Å)

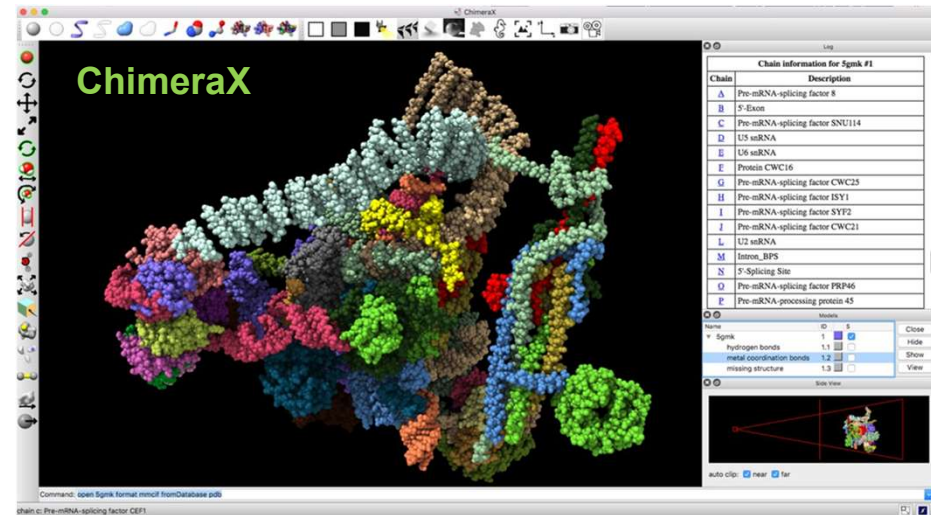
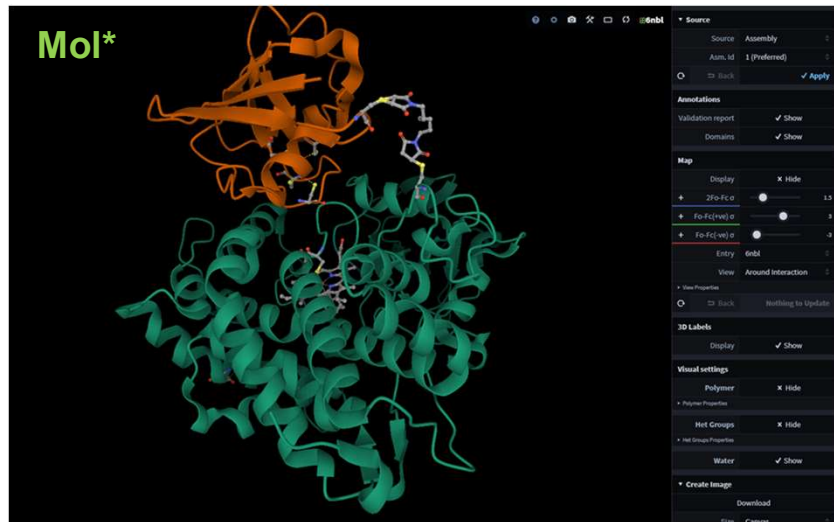
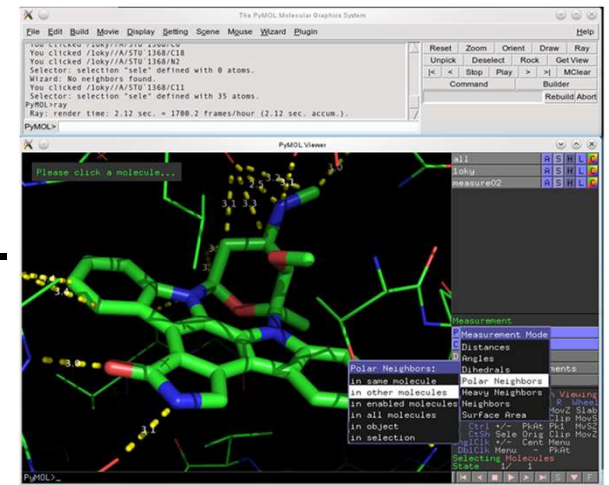
Residue  
name

“Occupancy”

# How to view mmCIF or PDB file in 3D

- Different molecular graphics programmes
- Display the atoms in 3D space and interpret the bonds.
- PyMol, ChimeraX, Mol\*

## PyMol



# Why have a central archive of structures?



## 1. Accessibility

One-stop shop  
Searchable

2.1Å  
PDB: 3L21

## 2. Context & connection

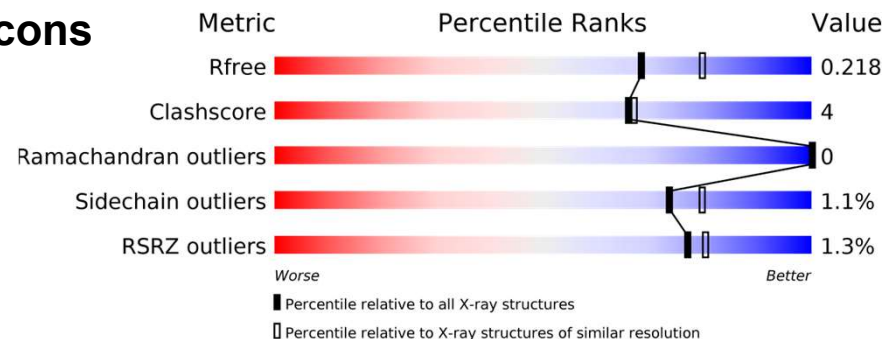
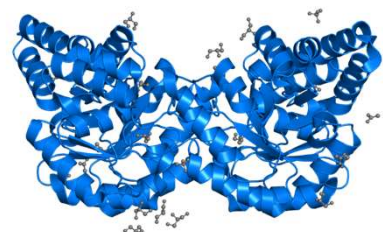
‘Structural bioinformatics’  
Integration with other resources



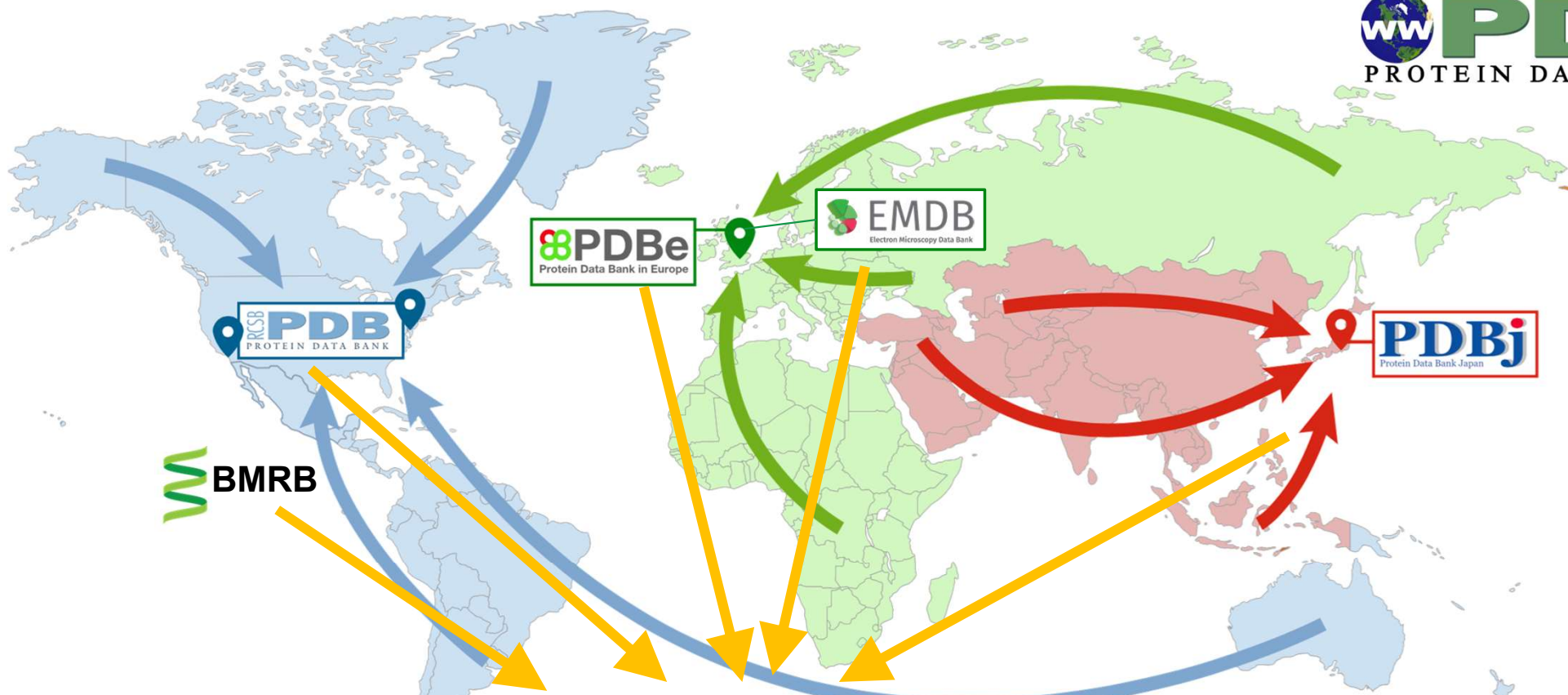
3D Beacons

## 3. Consistency


Comparison against all other entries  
→ Validation







**BMRB**

 **Data Archives**

300+  
BLAST, UniProt, PubChem,  
*Molprobitry, AlphaFold, etc*

Collect, curate, store and distribute experimentally determined 3D structures of biological macromolecules.

# The Worldwide Protein Data Bank (wwPDB)



RCSB PDB

1997-9 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

1P2B-101

Developers: Join the RCSB PDB Team

### 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 biochemistry and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data. 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.

**Join the RCSB PDB Team**

**COVID-19 CORONAVIRUS Resources**

**May Molecule of the Month**

Nicotine, Cancer, and Addiction

Latest Entries: As of Jan 10, 2022

Features & Highlights

News

PDBe Protein Data Bank in Europe

Advanced Search

Home | PDBe-KB | Deposit | Services | Training | PDB Act Project | Documentation | About | Feedback

PDBe is a founding member of the Worldwide Protein Data Bank (wwPDB) which collects, organises and disseminates data on biological macromolecular structures. wwPDB Partners: RCSB PDB, PDBE, EMDB. Read more about PDBe.

### Services

[PDBe-APP](#)
[Advanced Search](#)
[BFTS](#)
[PDBe-Kit](#)
[PDBe-ISA](#)
[PDBe-Chem](#)
[Download service](#)
[MMP](#)

### Latest archive statistics

Total PDB entries in the archive: 197822

Latest Data:

336 Latest Entries
80 Latest Chemistry
137 Latest Biology

### Featured

**A molecular switch for fluorescence**

Like the coral in our swimming pools, some colourful proteins can be switched on.

**Latest news**

**Compare experimental models from the PDB with AlphaFold models at PDBe-KB**

2022-10-20

You can now superpose AlphaFold models onto PDB structures at the touch of a button using the PDBe-KB.

**wwPDB Charter: Full and Associate Members**

2022-10-13

wwPDB is committed to responsible, international stewardship of public structure biology data.

**Secondary structure variance on PDBe-KB aggregated views of proteins**

EMDB Electron Microscopy Data Bank

EMBL-EBI | Services | Research | Training | About us | EMBL-EBI

Enter your search term(s) in the box below or build an advanced search query.

Search EMDs

Examples: 102I, Apoferritin, Tomography, Rossmann MD, SA1A

Home | Deposit | Documentation | Resources | FTP Archive | REST API | About | Feedback | Share

### EMDB News

- In February 2022, wwPDB switched to using version 3 of the EMDs data model and the old data model (v1.9.0) is no longer supported. Read the [wwPDB announcement](#) for more details.
- In July 2021, the EMDs website moved to a new location. [What do I need to know?](#)

As of 18 May 2022, EMDs contains 20102 entries (latest entries, trends).

### Quick Links

- EMDB Policies
- Talks & Tutorials
- Validation Analysis
- EMDB Citations
- EMPIAR
- PDBx
- Biomeq Archive
- EMDataResource
- EM Navigator
- 3D EM History

Recent Entries (Show all)

Explore EMDs

[Browse EMDs](#)
[EMDB statistics](#)
[SARS-CoV-2 entries](#)
[Deposit data](#)

PDBj Protein Data Bank Japan

English | 日本語 | 简体中文 | 繁體中文 | 한국어

Search pdbj.org

Home | PDB | PDBE | PDBx | PDB-ISA | About | Feedback

### About PDBj

**PDBj (Protein Data Bank Japan)** is a project team operating under the Joint Usage and Research activities of the [Institute for Protein Research](#), Osaka University. We maintain the single global PDB, BMRB, EMDs archives of macromolecular structures and provide integrated tools, in collaboration with RCSB PDB in the USA, PDBE in the EU, [EMDB](#) in the USA, and [EMBL](#) in the EU. PDBj is supported by [JST/JSPARC](#) and [EMBL-BMRB](#). PDBj's new logo design incorporates cyclochrom c (PDBID: 1JYC), which was the first structure determined in Japan.

### Find the service you need

Choose a keyword listed below or input keywords into the textbox at the right of the keyword list. The brief explanation of the matched services will be displayed.

- Click the "Show all services" button to display the explanation for all services.
- Input some keywords into the "Word Search" box to narrow down the search results.

PDB
  BMRB
  EMDs

search
  deposition
  viewer
  education/dictionary

BMRB
  electron microscopy
  secondary structure
  sequence

similarity
  function prediction
  chemical component
  structure prediction

changing site
  surface structure
  3D structure
  genome

iEDF
  SPARQL
  gene
  disease

Cling

Show all services

### Latest news

- 2022-05-18 [235 new PDB entries have been released on 2022-05-18.](#)
- 2022-05-04 [\[wwPDB\] Distributes PDBx/mmCIF-Formatted Assembly Files](#)
- 2022-04-11 [Protein Data Bank 50th Anniversary Virtual Issue: Featured structural articles from The Journal of Biochemistry were released](#)
- 2022-04-04 [\[wwPDB\] ModelCIF Extension for Computed Structure Models](#)

BMRB Biological Magnetic Resonance Data Bank

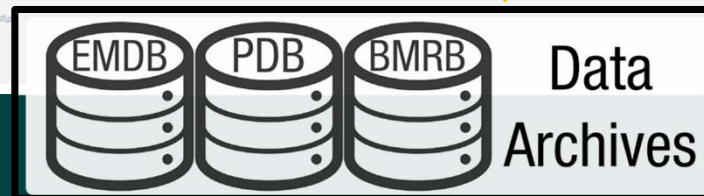
Statistics Calculated for All Chemical Shifts from Atoms in the 20 Common Amino Acids

The statistics presented in this table were calculated from the full BMRB database. This includes paramagnetic proteins, proteins with aromatic prosthetic groups, and entries where chemical shifts are reported relative to uncommon chemical shift reference compounds.

Jump to different statistics as CSV: Protein | Full

Download current statistics as CSV

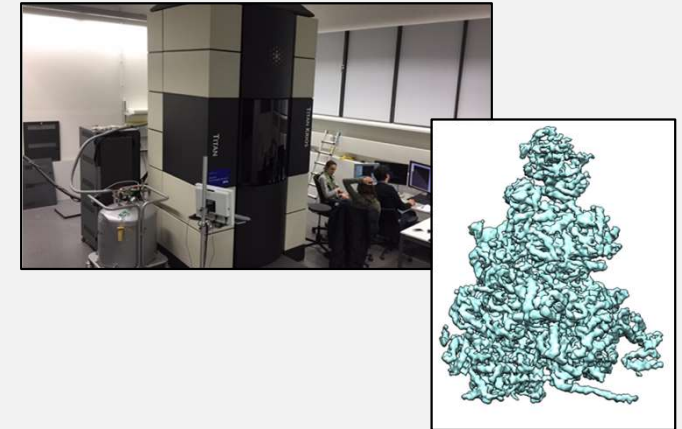
Compound ID	Amino Name	Number of shifts	Minimum Shift	Maximum Shift	Average Shift	Standard Deviation	Shift Outliers	Distribution Histogram
Ala	Ala	95419	-0.914	131.25	8.193	0.843	166	
Asp	Asp	62374	-2.52	17.870	4.244	0.441	1218	
Arg	Arg	60439	-14.040	5.48	1.353	0.278	1089	
Pro	Pro	61540	0.037	187.2	177.747	3.655	41	
Leu	Leu	84427	17.207	354.698	53.165	2.706	89	
Val	Val	80347	-40.993	318.868	19.043	2.979	205	
Ile	Ile	91438	0.049	706	123.349	5.867	105	
Met	Met	63333	0.011	178	8.240	1.019	43	
Thr	Thr	42986	1.212	12.97	4.289	0.467	910	
His	His	38833	-4.78	27.530	1.790	0.309	502	



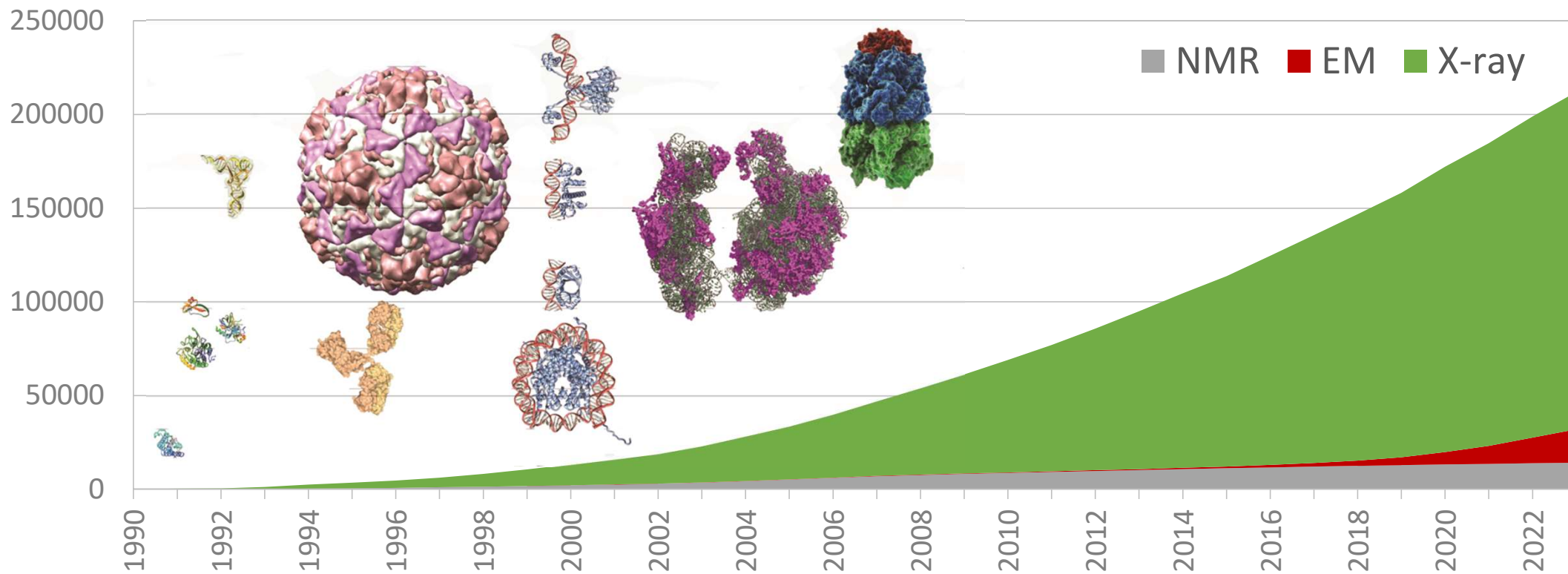
# What experimental methods are accepted?

Method	Total	2019	2020	2021	2022	2023
X-ray Crystallography	85%	84%	80%	74%	69%	66%
Solution NMR Spectroscopy	6.7%	3.3%	2.7%	2.9%	2.1%	2.0%
Electron Microscopy	8.4%	13%	17%	24%	29%	32%

~1%: Fibre Diffraction, Neutron Diffraction, Electron Crystallography, Solid state NMR...

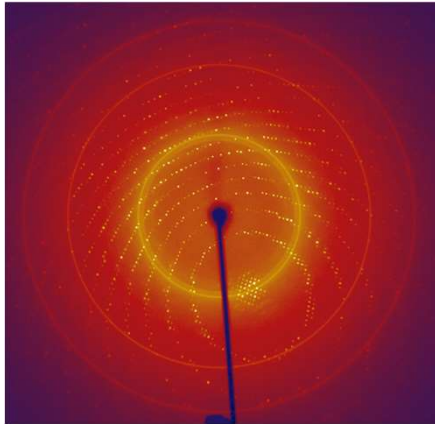


## Growth of data in the PDB



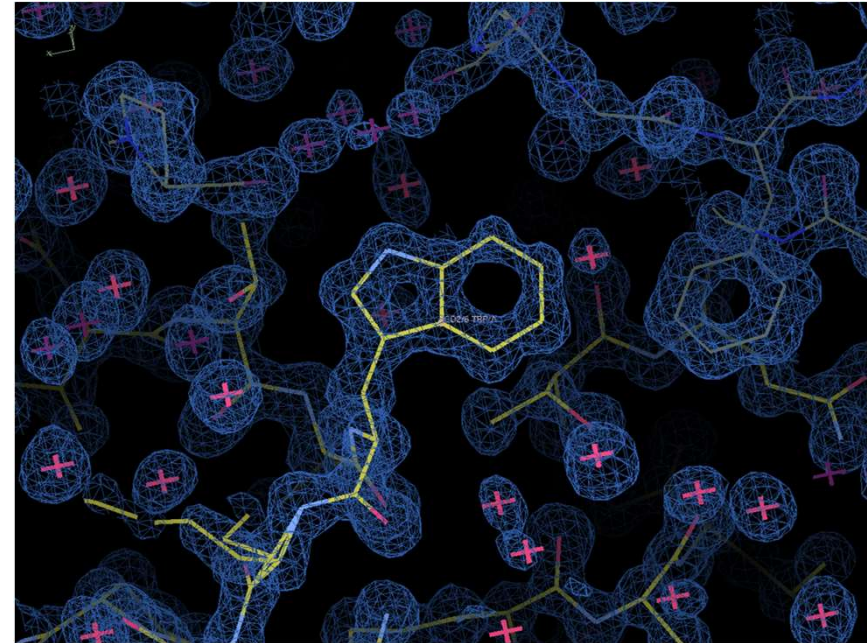
In January 2023, we passed the 200,000 milestone. As of today, ~215,000 structures

# X-ray crystallography data



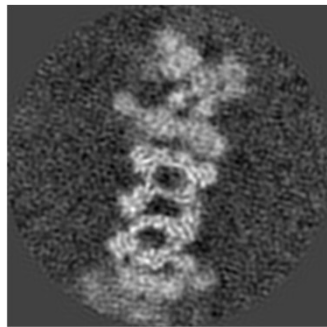
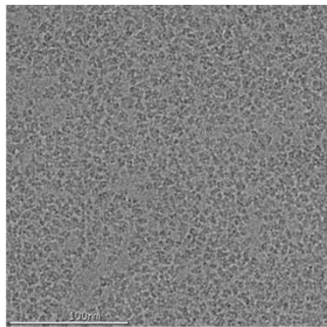
From the diffraction pattern, a map can be calculated.

This indicates the location of electrons (therefore atoms) in the crystal  
Hence - 'electron density map'

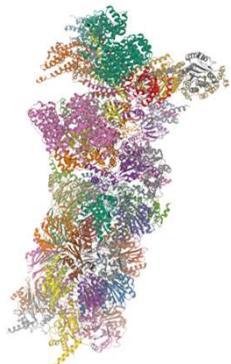


Model is built into this map in an iterative process

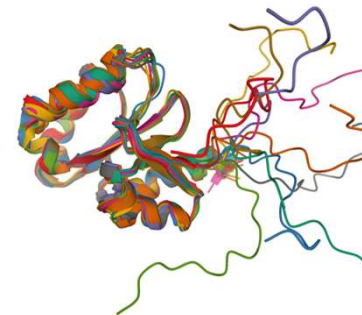
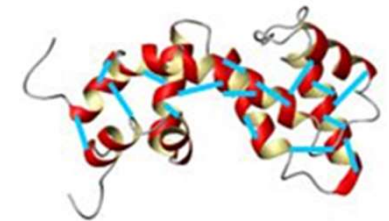
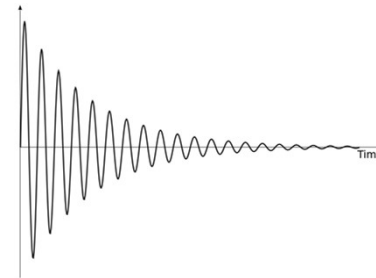
# Cryo-electron Microscopy (cryoEM)



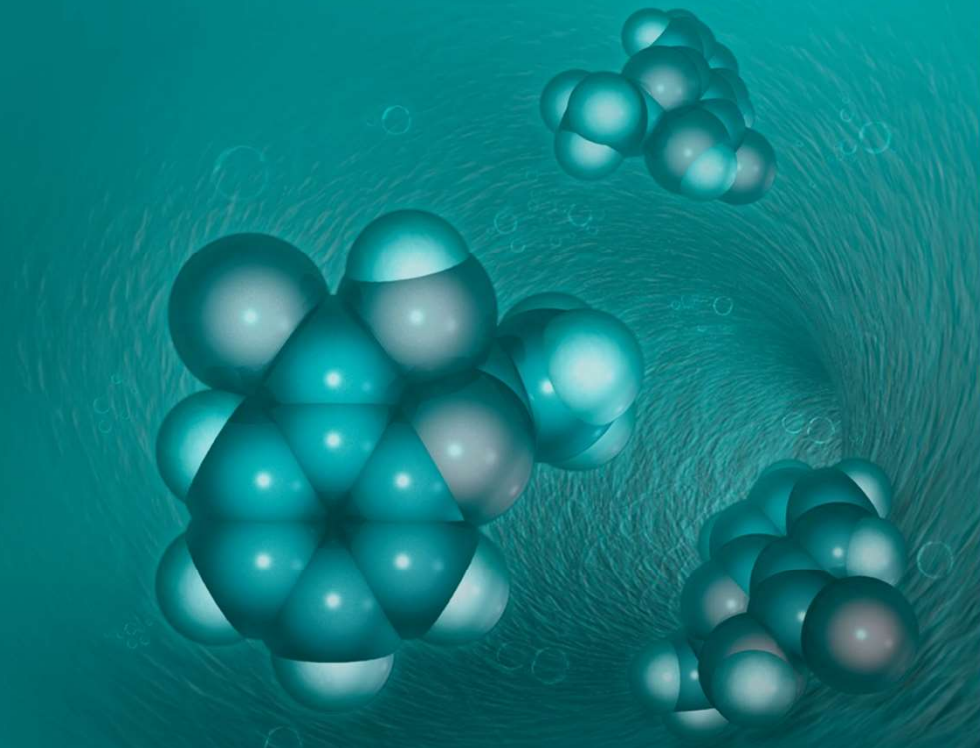
**EM volume map**



# Nuclear Magnetic Resonance (NMR)

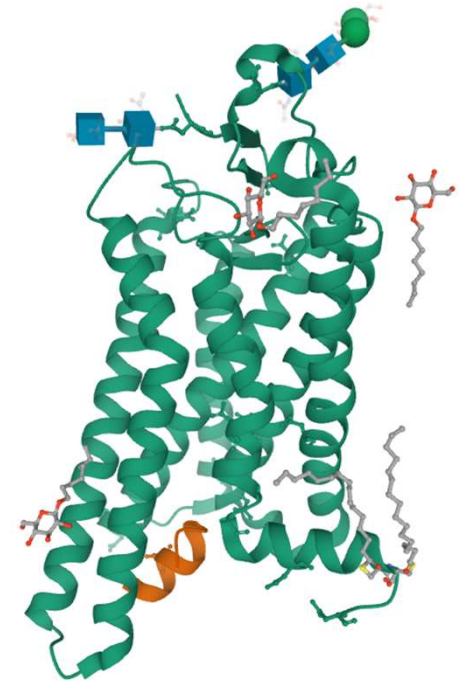
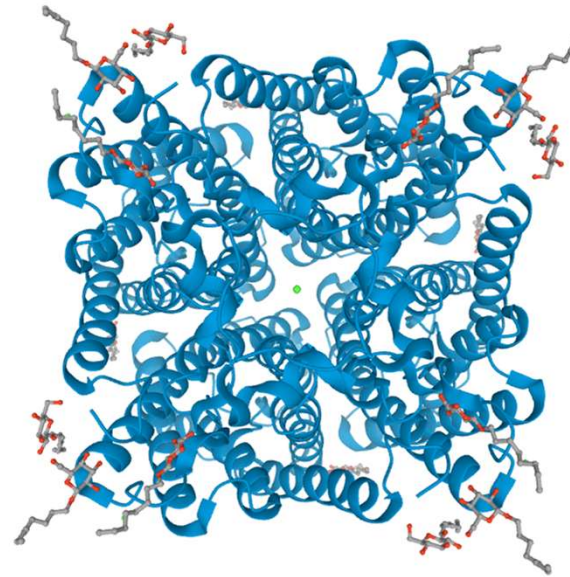
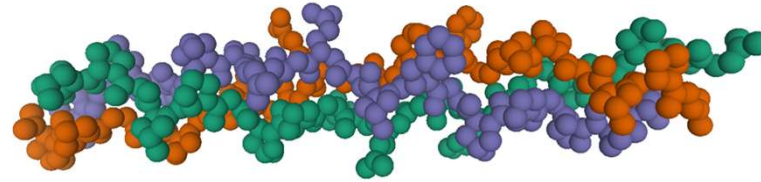


# Macromolecules and ligands



# Functions of proteins

- Structural
- Catalysis (i.e. enzymes)
- Receptors
- Channels
- Transport

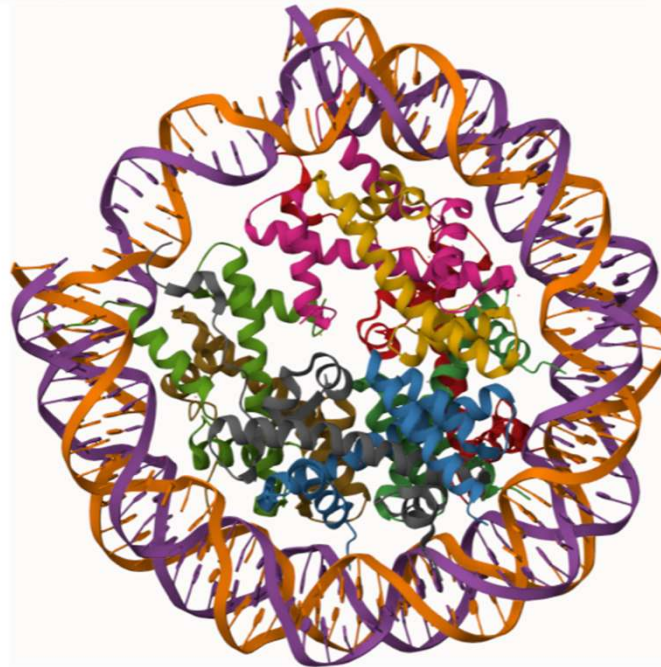
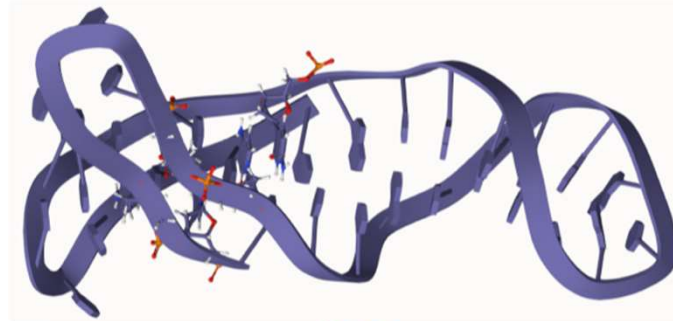


1Q7D, 3DQB, 2W1P



# Functions of nucleic acids

- Storage of genetic information
- Catalysis (i.e. ribozymes)
- Protein synthesis
- Regulation of gene expression
- Epigenetic regulation



3KWQ, 4PRF, 7O5E

## Secondary structure - nucleotides



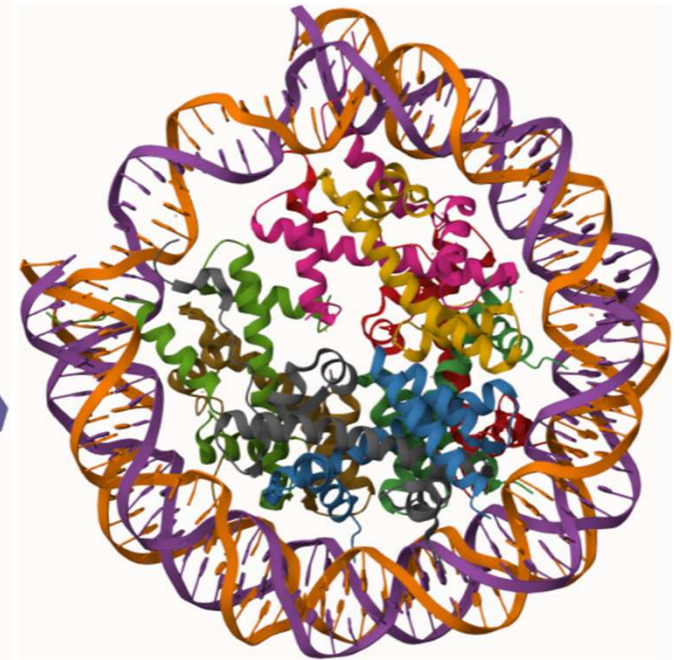
Single stranded  
DNA



Double stranded  
DNA



Quadruplex DNA

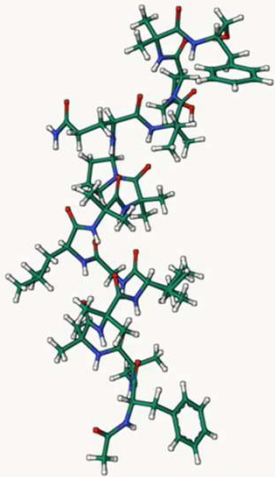


Nucleosome

3BBZ, 3E6Z

# Secondary structure - proteins

Peptide  
(Alpha helix)



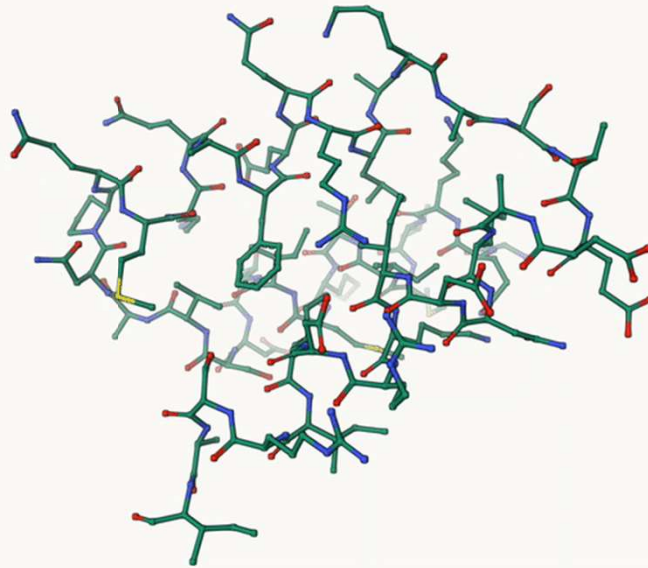
4G13, 3BBZ, 3E6Z

# Secondary structure - proteins

Peptide  
(Alpha helix)



Alpha helices



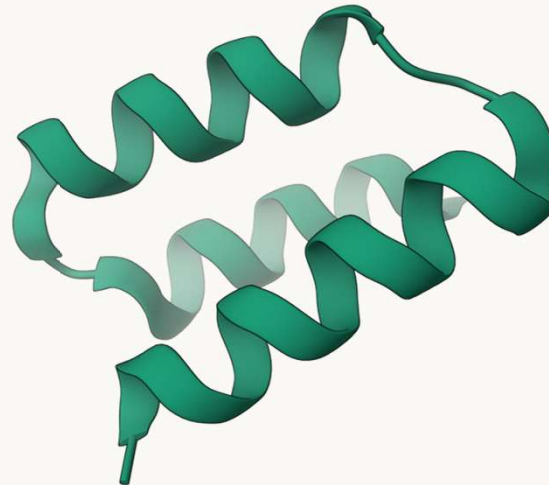
4G13, 3BBZ, 3E6Z

# Secondary structure - proteins

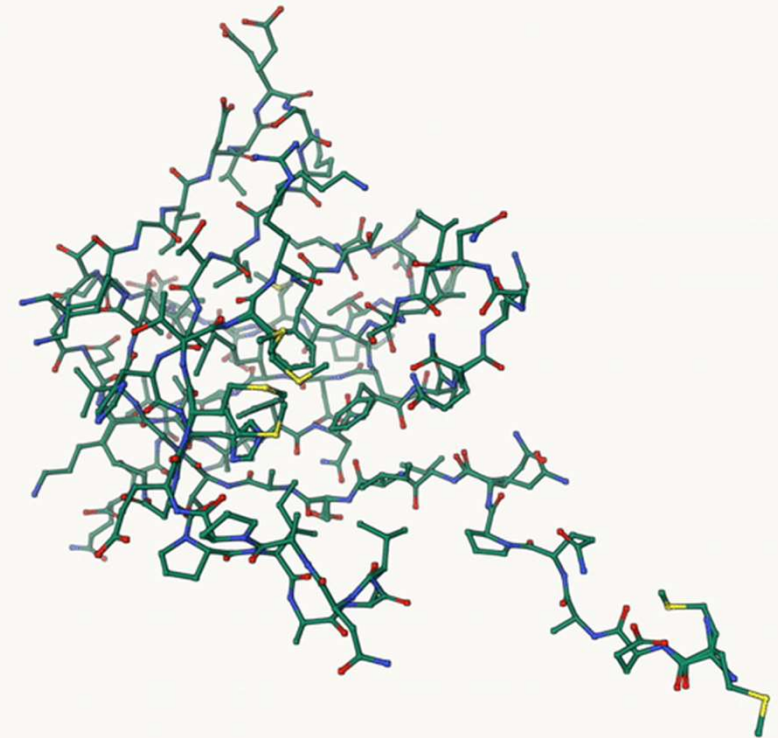
Peptide  
(Alpha helix)



Alpha helices



Beta strands



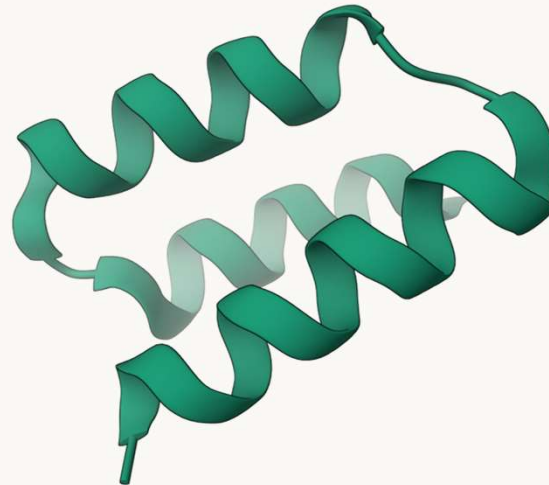
4G13, 3BBZ, 3E6Z

## Secondary structure - proteins

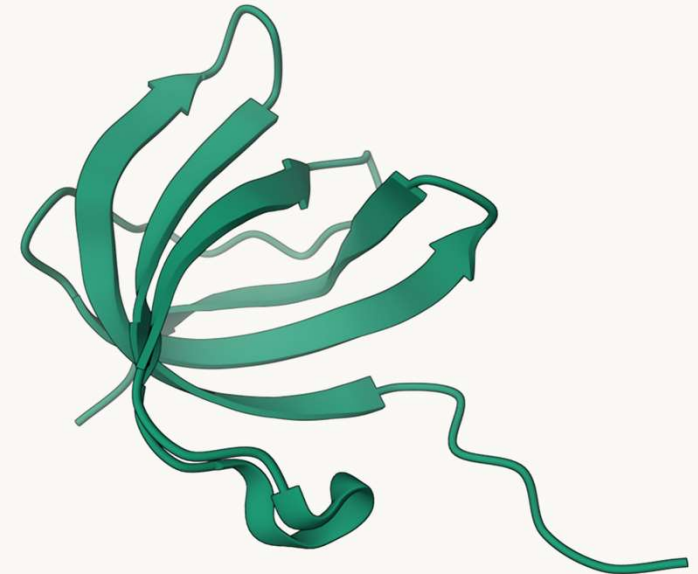
Peptide  
(Alpha helix)



Alpha helices



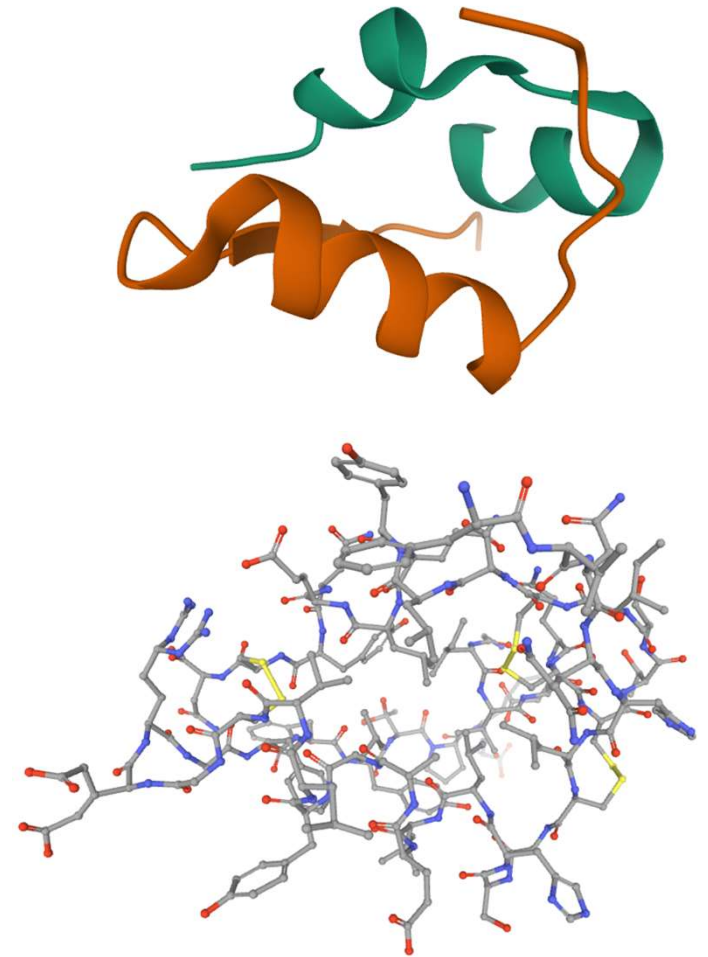
Beta strands



4G13, 3BBZ, 3E6Z

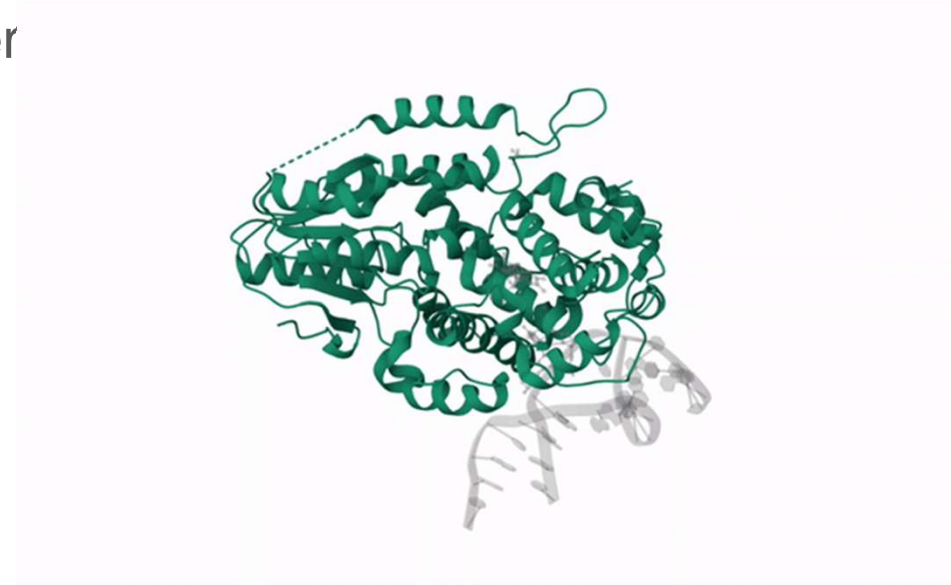
## Structural folds - PDBeFold

- Tertiary folds up into a specific shape
    - Driven by hydrophobic interactions
    - Also, more specific interactions:  
e.g. disulphide bonds, salt bridges
    - Puts amino acids in the right place for function
- **PDBeFold** – search for structural similarity  
(<https://www.ebi.ac.uk/msd-srv/ssm/>)
- Protein domains / protein families (InterPro)



# Assemblies - PDBePISA

- Multiple macromolecules interacting together
- Can be simple dimers or huge complexes
- Amino acids mediate interactions
  - Usually, hydrophobic patches
  - Can also be disulphide bridges



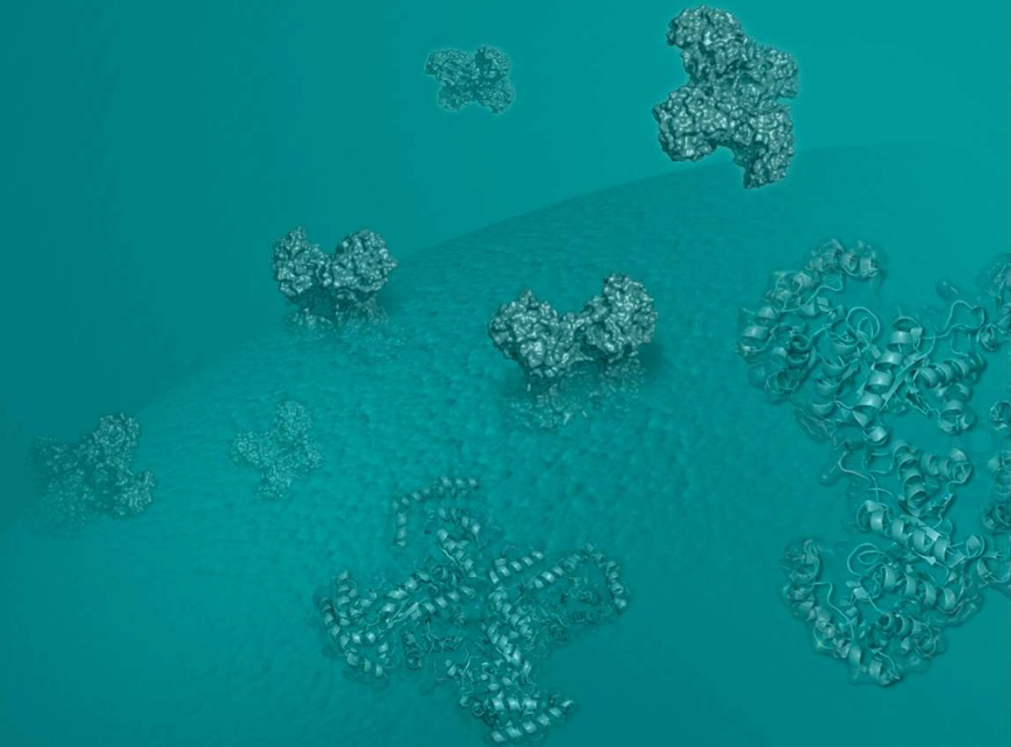
→ **PDBePISA** to analyse assemblies and interfaces  
(<https://www.ebi.ac.uk/pdbe/pisa/>)




# Searching the PDBe

 **PDBe**  
Protein Data Bank in Europe

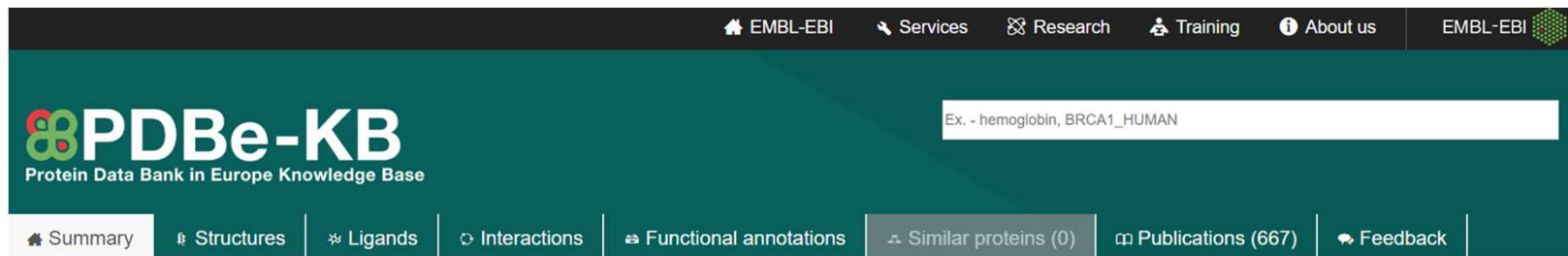
EMBL-EBI 



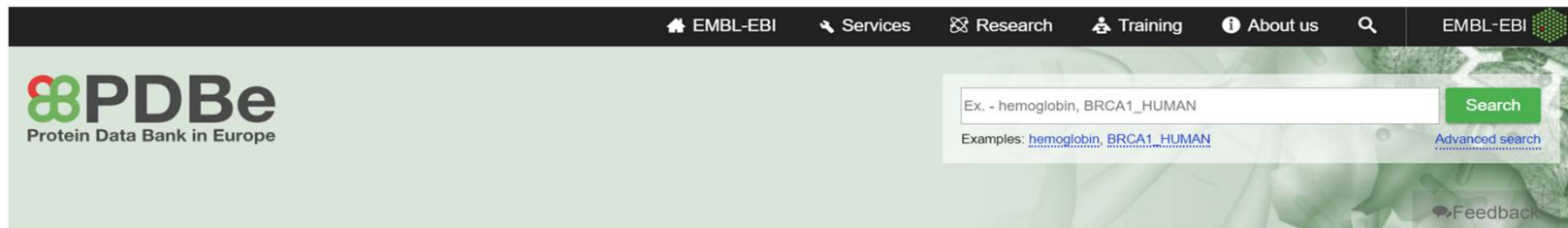
# At PDBe: look for the search bars



The screenshot shows the top navigation bar of the PDBe website. On the left is the PDBe logo (Protein Data Bank in Europe). To its right is a search bar containing the text "Ex. - hemoglobin, BRCA1\_HUMAN" and a magnifying glass icon. Below the search bar is a link for "Advanced Search". The main navigation menu includes: Home, PDBe-KB, Deposition, Services, Training, PDB Art Project, Documentation, About, and Feedback. On the far right are social media icons for a plus sign, Facebook, and Twitter.



The screenshot shows the top navigation bar of the PDBe-KB website. At the top right, there are links for EMBL-EBI, Services, Research, Training, About us, and EMBL-EBI with a logo. Below this is the PDBe-KB logo (Protein Data Bank in Europe Knowledge Base). To the right is a search bar with the text "Ex. - hemoglobin, BRCA1\_HUMAN". Below the search bar is a horizontal menu with the following items: Summary, Structures, Ligands, Interactions, Functional annotations, Similar proteins (0), Publications (667), and Feedback.



The screenshot shows a close-up of the search bar on the PDBe website. The search bar contains the text "Ex. - hemoglobin, BRCA1\_HUMAN" and a green "Search" button. Below the search bar, there are "Examples: hemoglobin, BRCA1\_HUMAN" and a link for "Advanced search". A "Feedback" link is visible at the bottom right of the search area.

Search bars → multi-purpose tool → ‘Advanced search’

The screenshot shows the top navigation bar of the PDBe website. On the left is the PDBe logo (Protein Data Bank in Europe). In the center is a search input field containing the text "Ex. - hemoglobin, BRCA1\_HUMAN". Below the input field, the text "Advanced Search" is highlighted with a red rectangular box. To the right of the input field is a magnifying glass icon. Below the search bar is a horizontal menu with items: Home, PDBe-KB, Deposition, Services, Training, PDB Art Project, Documentation, About, and Feedback. On the far right are social media icons for a plus sign, Facebook, and Twitter.

The screenshot shows the top navigation bar of the PDBe-KB website. On the left is the PDBe-KB logo (Protein Data Bank in Europe Knowledge Base). In the center is a search input field containing the text "Ex. - hemoglobin, BRCA1\_HUMAN". Below the input field is a horizontal menu with items: Summary, Structures, Ligands, Interactions, Functional annotations, Similar proteins (0), Publications (667), and Feedback. On the far right is the EMBL-EBI logo.

The screenshot shows the top navigation bar of the PDBe website. On the left is the PDBe logo (Protein Data Bank in Europe). In the center is a search input field containing the text "Ex. - hemoglobin, BRCA1\_HUMAN". Below the input field, the text "Advanced search" is highlighted with a red rectangular box. To the right of the input field is a green "Search" button. Below the input field is a horizontal menu with items: Examples: hemoglobin, BRCA1\_HUMAN, and Feedback. On the far right is the EMBL-EBI logo.

# 'Advanced search' - Protein domains

The screenshot displays the InterPro website interface. At the top left, the InterPro logo and tagline "Classification of protein families" are visible. A search bar at the top right contains the text "Ex. - hemoglobin, BRCA1\_HUMAN" and a magnifying glass icon. Below the search bar, a red box highlights the "Advanced Search" button. The main content area is divided into two sections. On the left, a sidebar shows the "Overview" for the domain "IPR001487 Bromodomain", listing statistics for Proteins (100k), Domain Architectures (2k), Taxonomy (9k), Proteomes (2k), Structures (2k), and AlphaFold (43k). On the right, the "Advanced search form" is open, showing a search condition set to "Contains" with the example "ipr013783 / immunoglobulin-like fold" and the input field containing "IPR001487". Below the input field, there is a "Select to add a search field" dropdown menu and "Submit" and "Cancel" buttons.

Search bars → multi-purpose tool → ‘autocomplete options’

The screenshot shows the PDBe (Protein Data Bank Europe) search interface. At the top left is the PDBe logo and the text 'Protein Data Bank Europe'. A search bar at the top right contains the text 'bromod' and a search icon. Below the search bar, three columns of results are displayed: 'Ligand', 'Molecule name', and 'Sequence family'. The 'Ligand' column lists chemical entities like '12-bromododecan-1-ol' and '12-bromododecanoic acid'. The 'Molecule name' column lists protein families like 'Bromodomain-containing protein 4'. The 'Sequence family' column lists InterPro identifiers like 'IPR001487 : Bromodomain'. A tooltip is visible over the 'IPR001487 : Bromodomain' entry in the 'Sequence family' column. On the left side of the page, there are navigation links for 'Home' and 'PDBe-KB', and a 'Services' section with a diagram of a protein structure.

Ligand	Molecule name	Sequence family
BDD : 12-bromododecan-1-ol (2)	Bromodomain-containing protein 4 (552)	IPR001487 : Bromodomain
BRC : 12-bromododecanoic acid (1)	Bromodomain and PHD finger-containing protein 1 (317)	IPR036427 : Bromodomain-like superfamily
U33 : Bromodeoxyuridine (1)	Bromodomain-containing protein 1 (317)	PF00439 : Bromodomain
BRC : 12-BROMODODECANOIC ACID (1)	Bromodomain adjacent to zinc finger domain ... (264)	IPR018359 : Bromodomain, conserved
	Bromodomain-containing protein 2 (164)	IPR043508 : Brdt, bromodomain, repeated
	Bromodomain and PHD finger-containing protein ... (63)	IPR037374 : BAZ2A/BAZ2B, bromodomain
	Bromodomain adjacent to zinc finger domain ... (61)	IPR043509 : Brdt, bromodomain, repeated
	Bromodomain-containing protein 9 (48)	IPR030411 : Nuclear body protein Sp1
	Bromodomain and PHD finger-containing protein ... (43)	IPR006565 : Bromodomain associated

# API tutorials

<https://github.com/glevans/7ADD-workshop-2024>

1\_API\_basics.ipynb

2\_GET\_vs\_POST.ipynb

<https://bit.ly/3SBNjNx>

More advanced:

<https://github.com/PDBEurope/pdbe-api-training/>  
→ subfolder api\_tutorials

# How do I pick the best structure for me? (part1)

# How good is a structure?

- Maps\*

\* - *X-ray crystallography & cryo electron microscopy*

- Resolution\*

- Geometry

- Validation



# How good is a structure?

- Maps\* are the real data

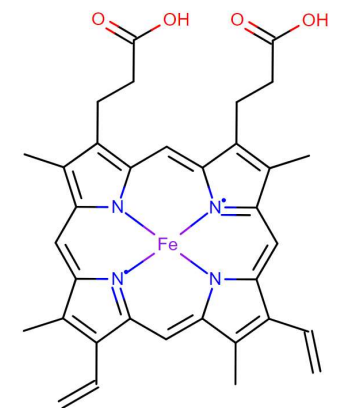
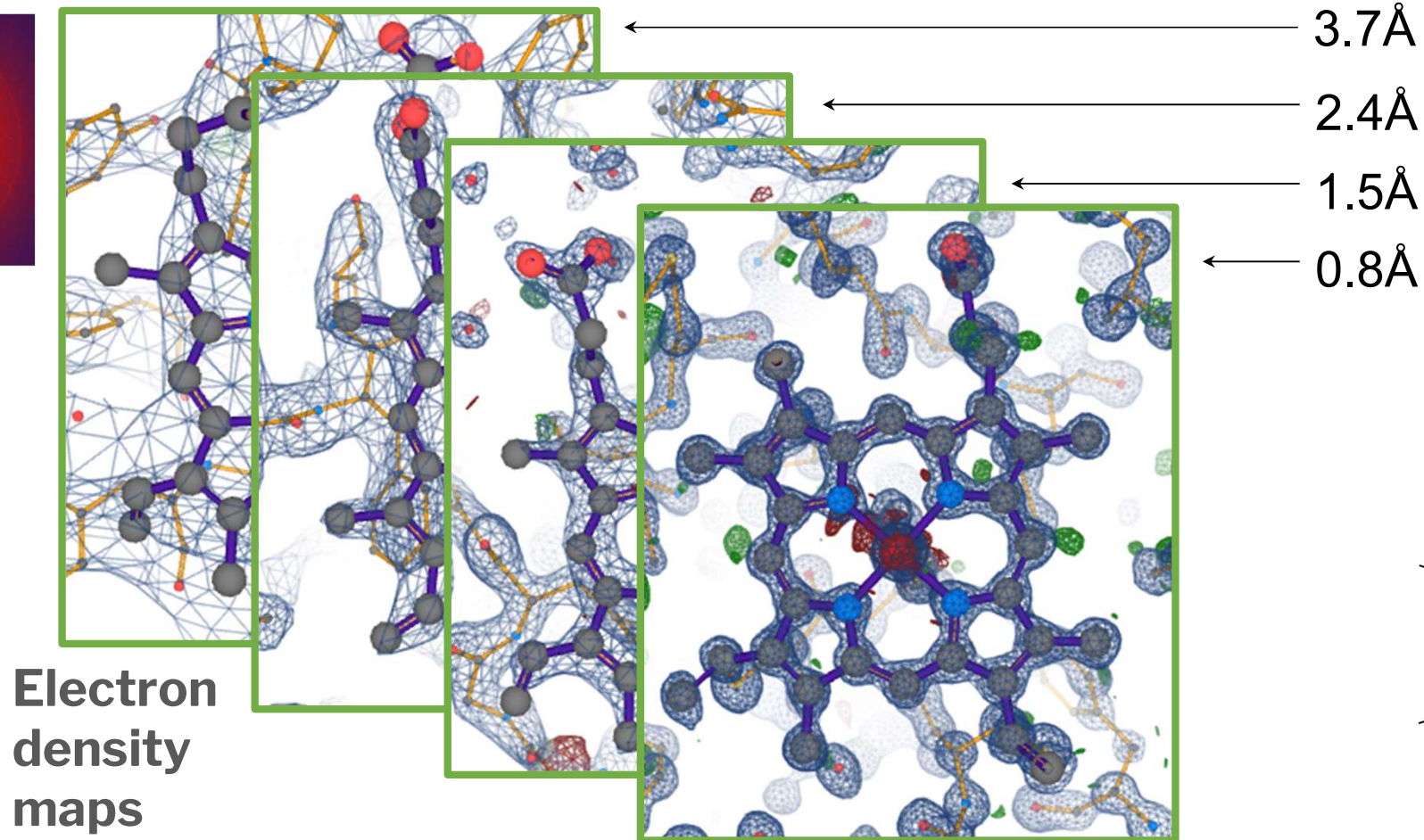
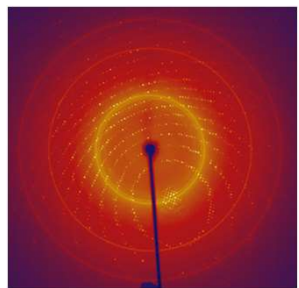
*\* - X-ray crystallography & cryo electron microscopy*

- Resolution\* matters, but isn't everything

- Geometry

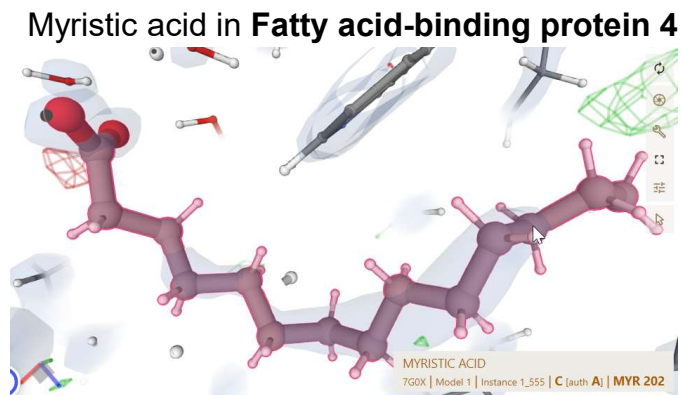
- Validation

# Maps are the real data - X-ray crystallography density maps

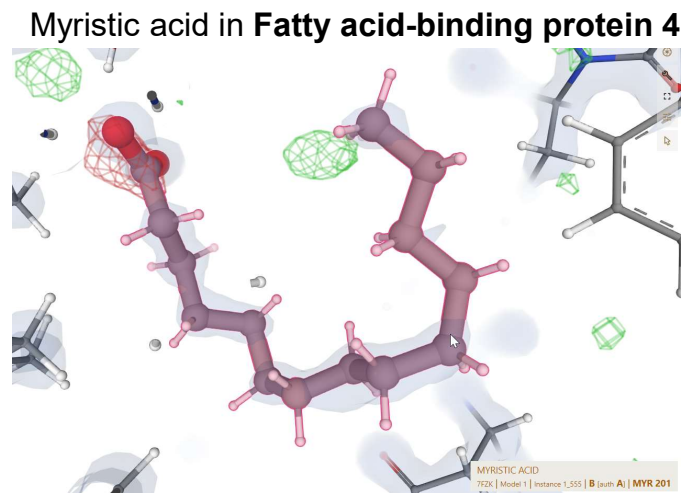


# Viewing electron density in Mol\*

- To initiate electron maps display: clicking on ligand or protein amino acid  
→ should surround atoms
- Regular map (blue) '2Fo-Fc' electron density map  
→ should surround atoms
- **Negative** and **positive** density  
→ highlights extra and missing atoms, respectively



PDB ID: 7G0X



PDB ID: 7FZK

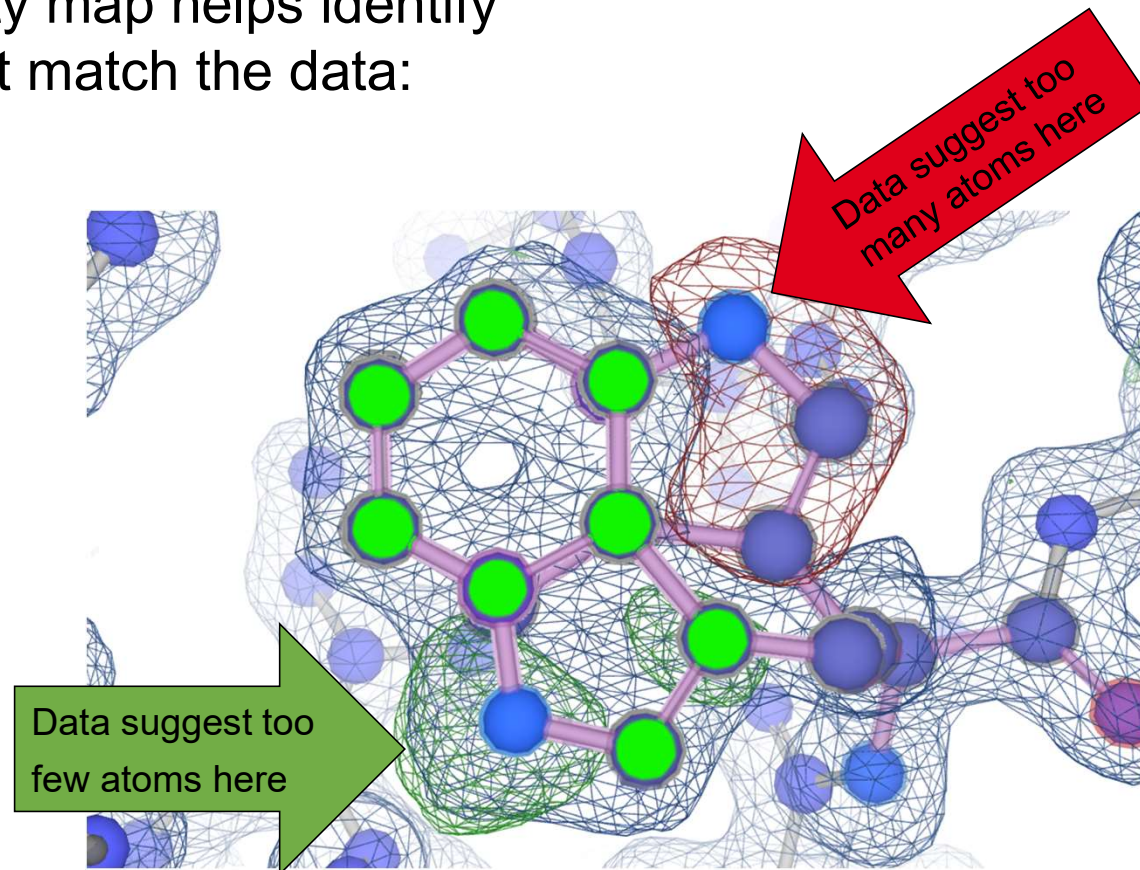
Volume Streaming		7G0X
2Fo-Fc $\sigma$	1.5	
Color		
Wireframe	<input type="checkbox"/> Off	
Opacity	0.15	
Fo-Fc(+ve) $\sigma$	3	
Color		
Wireframe	<input checked="" type="checkbox"/> On	
Opacity	0.3	
Fo-Fc(-ve) $\sigma$	-3	
Color		
Wireframe	<input checked="" type="checkbox"/> On	
Opacity	0.3	
Entry	7g0x	
View	Around Focus	

# Maps are the real data - X-ray crystallography density maps

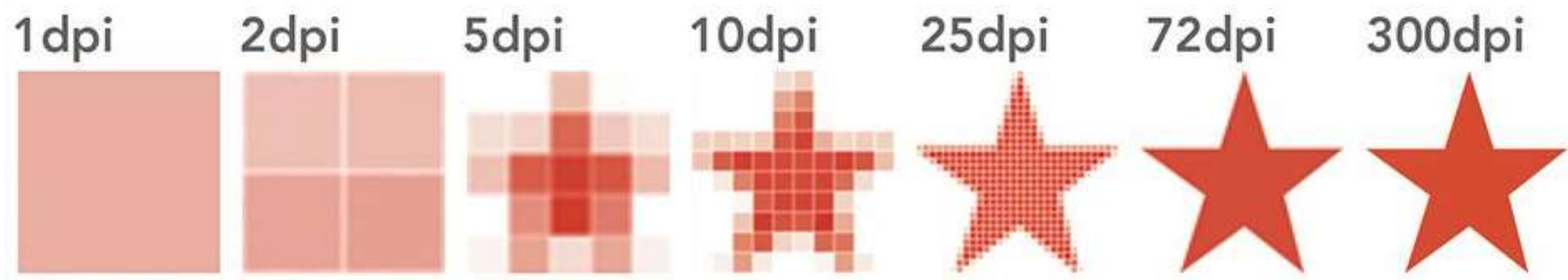
- Difference electron density map helps identify where the model does not match the data:

Areas where the model has too many atoms for the data

Areas where the model has too few atoms for the data

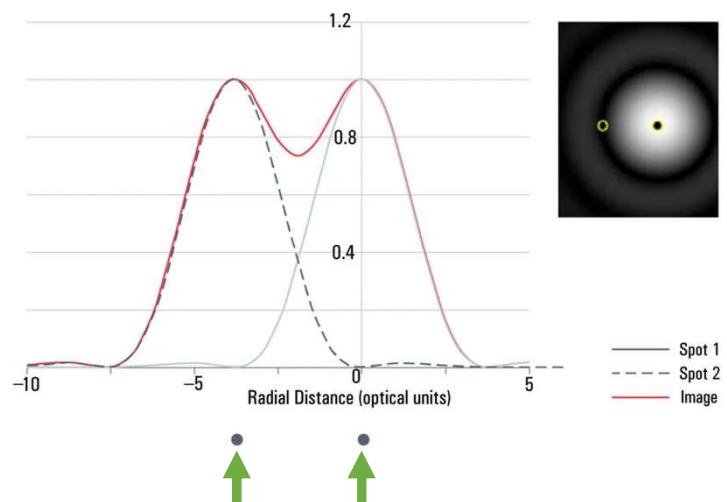


# Resolution



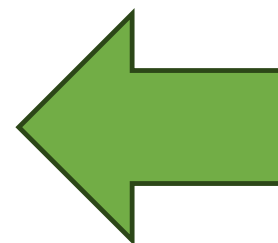
**Digital resolution**

**Larger number  
= better resolution**



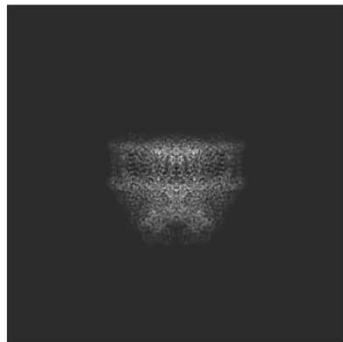
**Microscope resolution**

**Smaller number  
= better resolution**

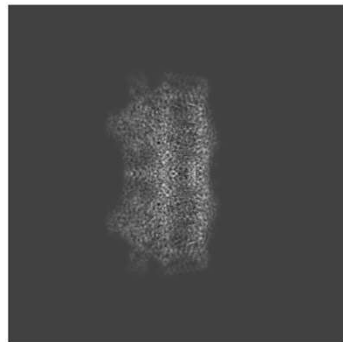


# Maps are the real data - Cryo-electron microscopy volumes

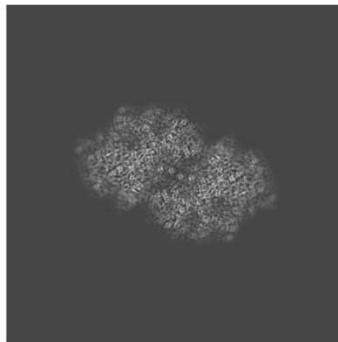
Photosystem II from cyanobacterium  
(Resolution: 1.9 Å)



X



Y



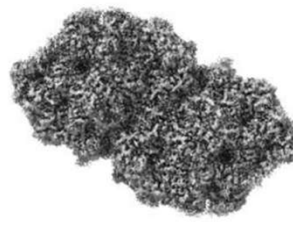
Z



X

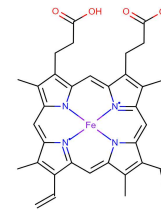


Y

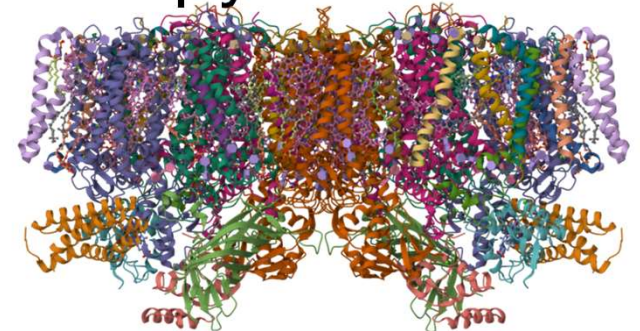


Z

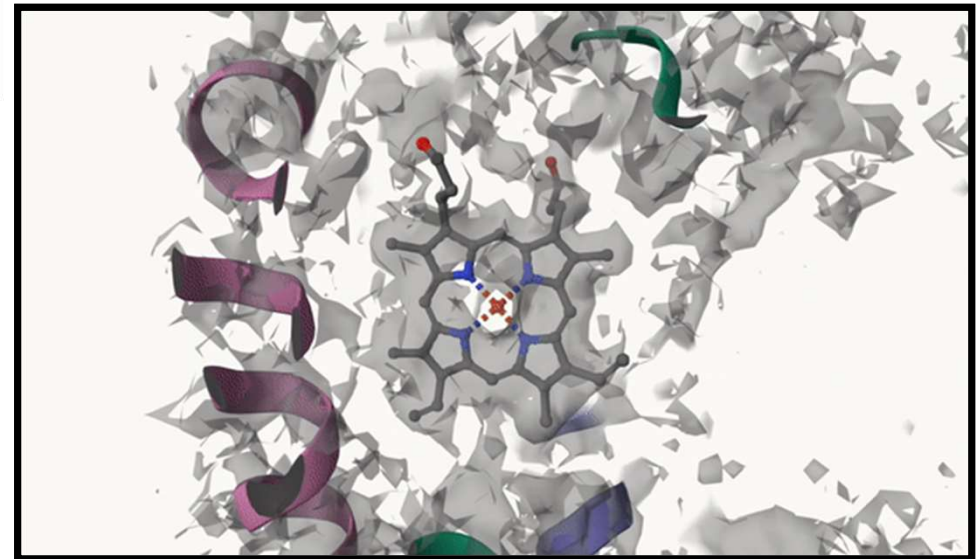
EM volume map



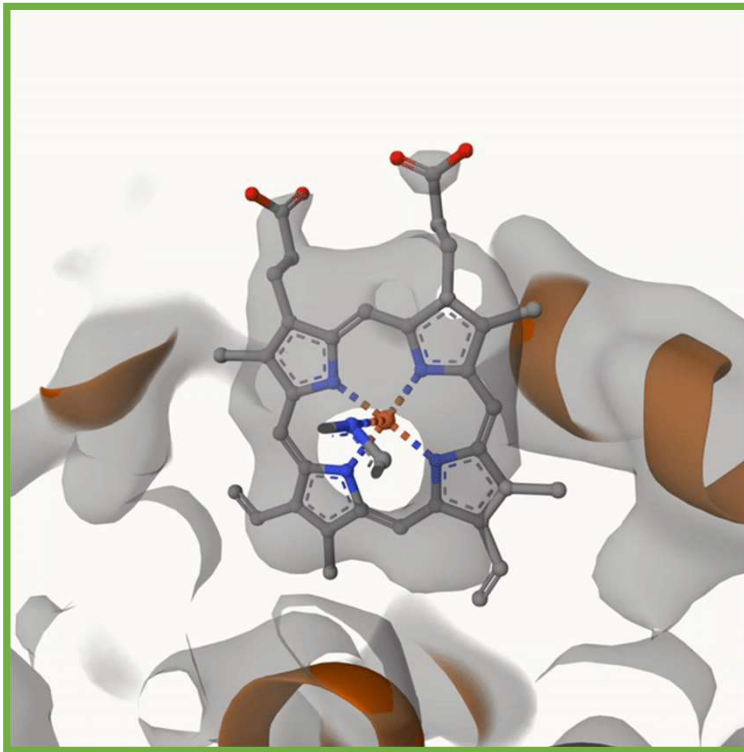
HEME



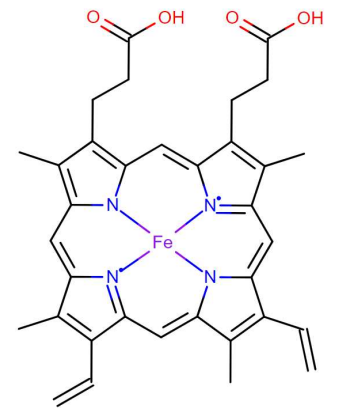
21 distinct protein chains



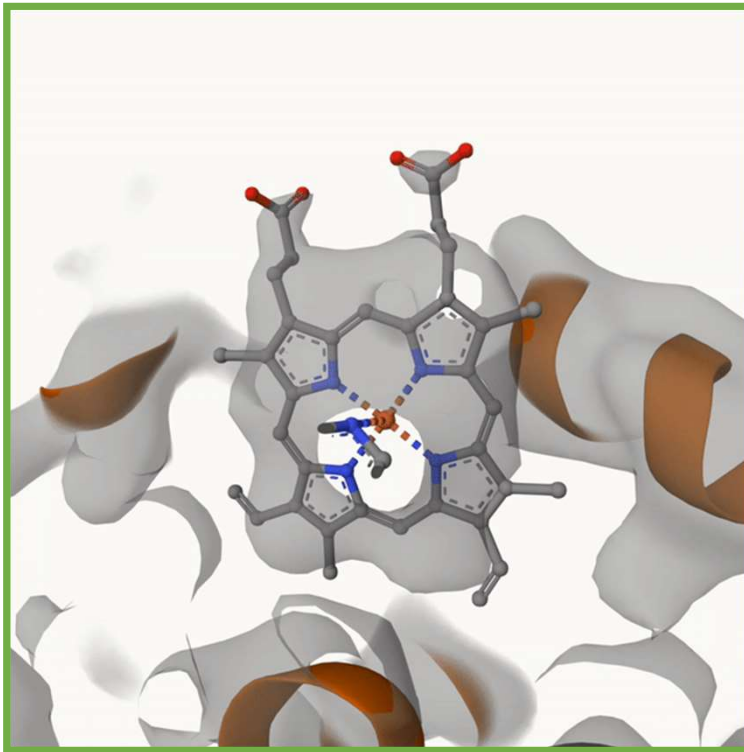
# Maps are the real data - Cryo-electron microscopy volumes



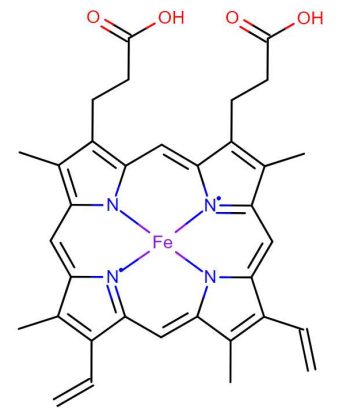
← 3.5Å



# Maps are the real data - Cryo-electron microscopy volumes

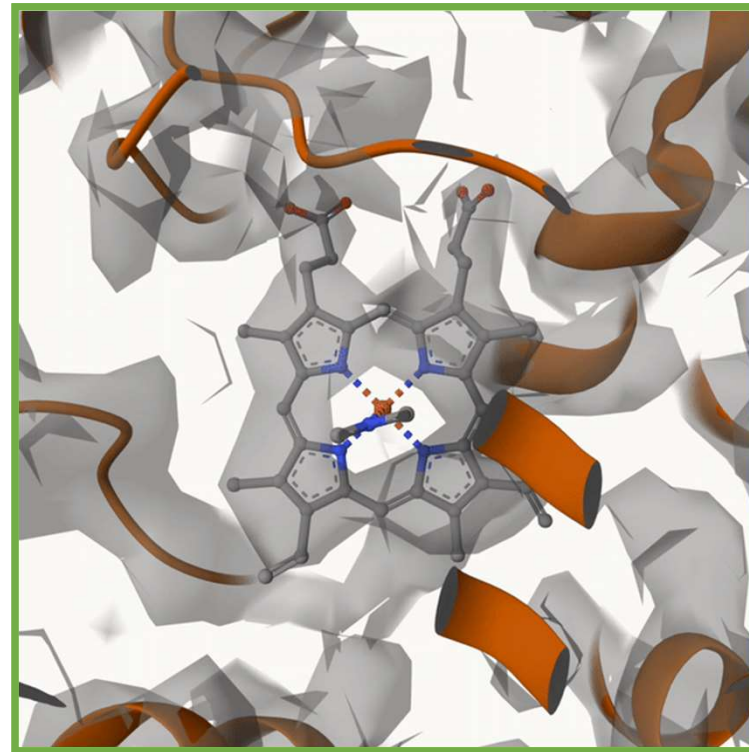


← 3.5Å

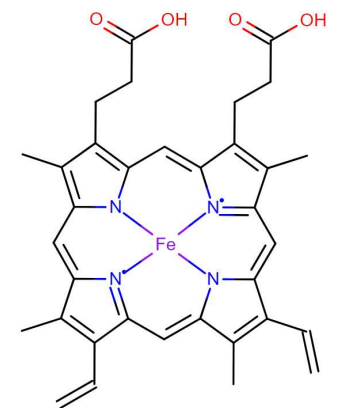




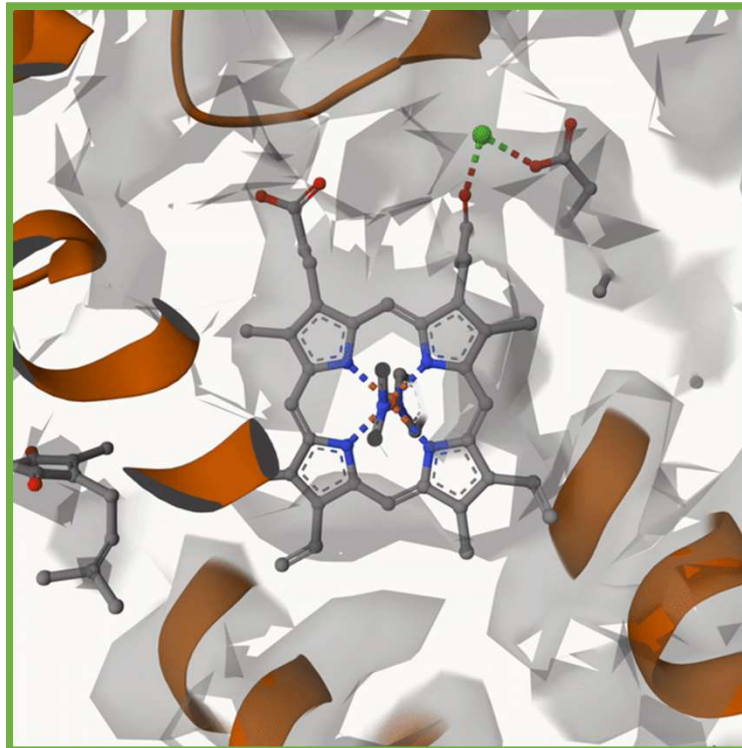
# Maps are the real data - Cryo-electron microscopy volumes



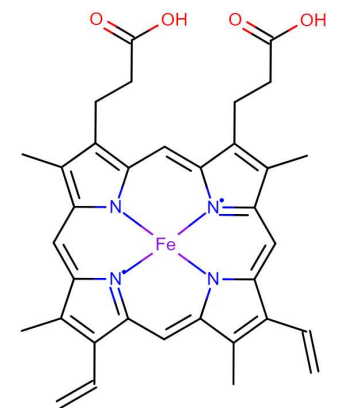
← 2.7Å



# Maps are the real data - Cryo-electron microscopy volumes

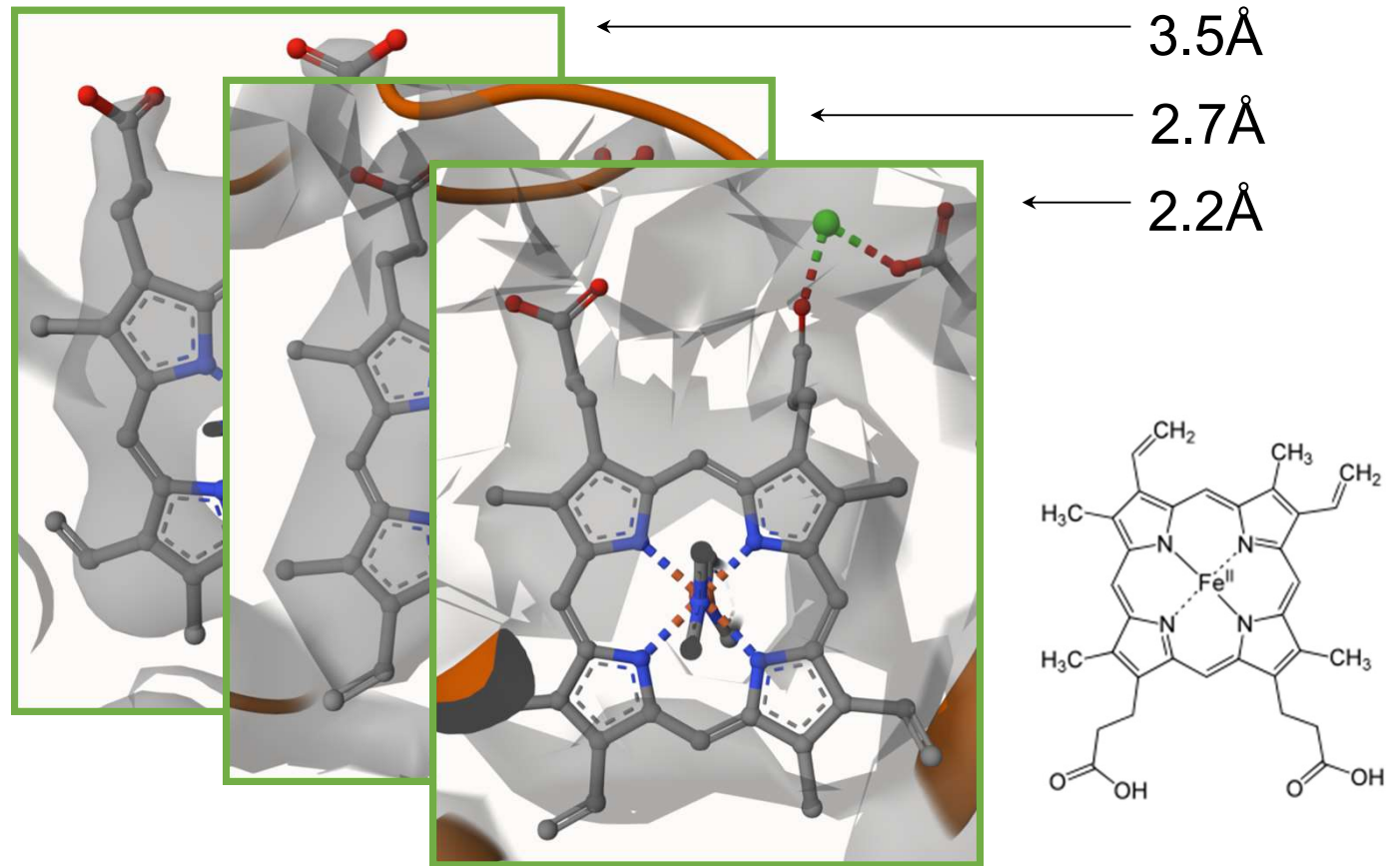


← 2.2Å



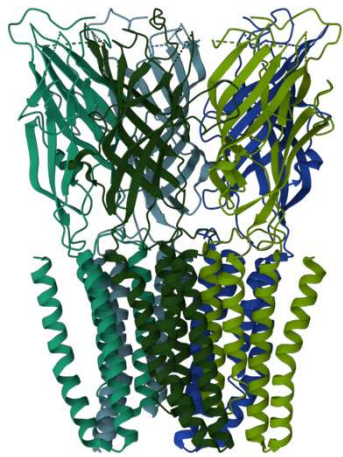
# Maps are the real data - Cryo-electron microscopy volumes

## Global resolution



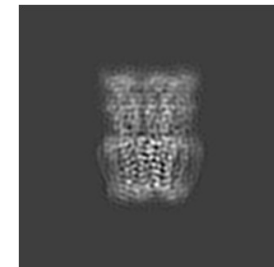
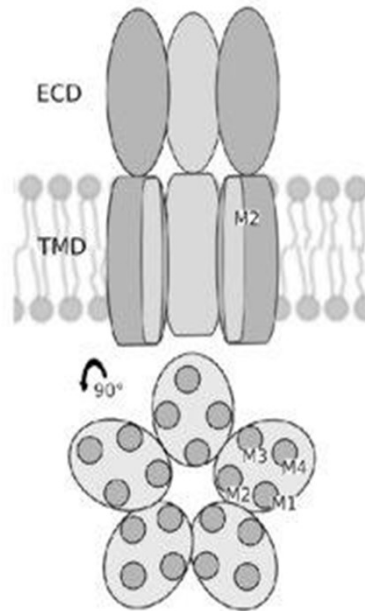
# Local resolution vs. Global resolution

## Example: Ion channel (ligand-gated)

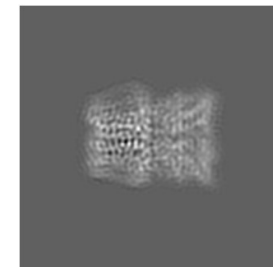


**ECD**  
= extra-cellular  
domain

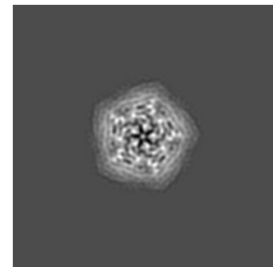
**TMD**  
= transmembrane  
domain



**X**



**Y**



**Z**

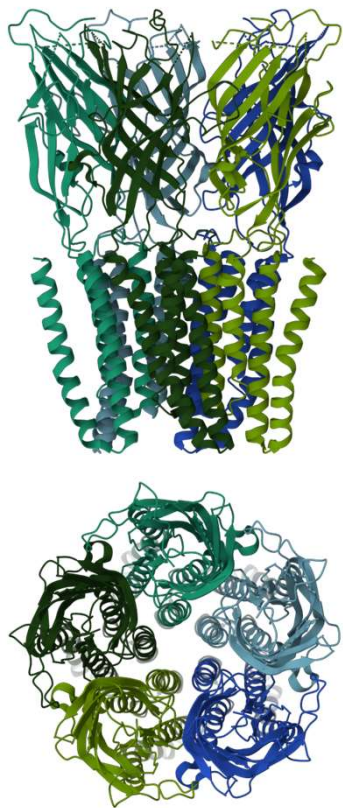


**EM volume map**

6ZGD, 6GZJ, 6ZGK

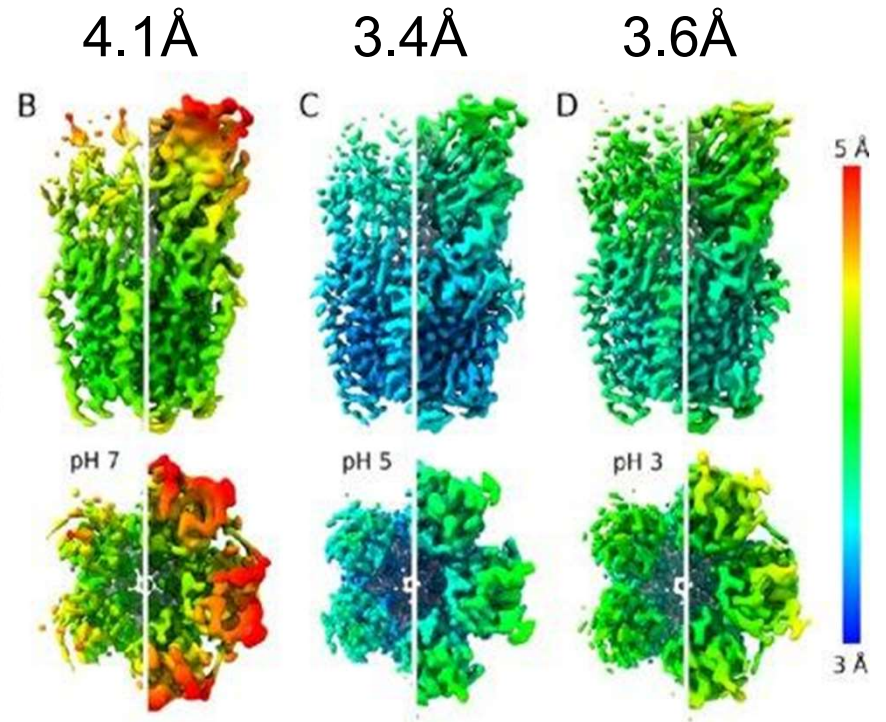
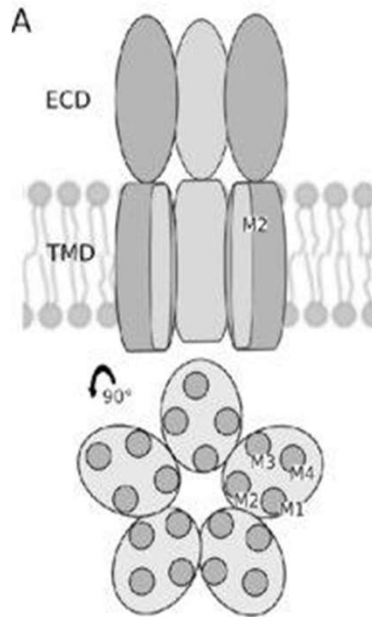
# Local resolution vs. Global resolution

## Example: Ion channel (ligand-gated)



**ECD**  
= extra-cellular  
domain

**TMD**  
= transmembrane  
domain



EM volume maps

6ZGD, 6GZJ, 6ZGK

# Molstar & ligand site tutorials

<https://github.com/glevans/7ADD-workshop-2024>

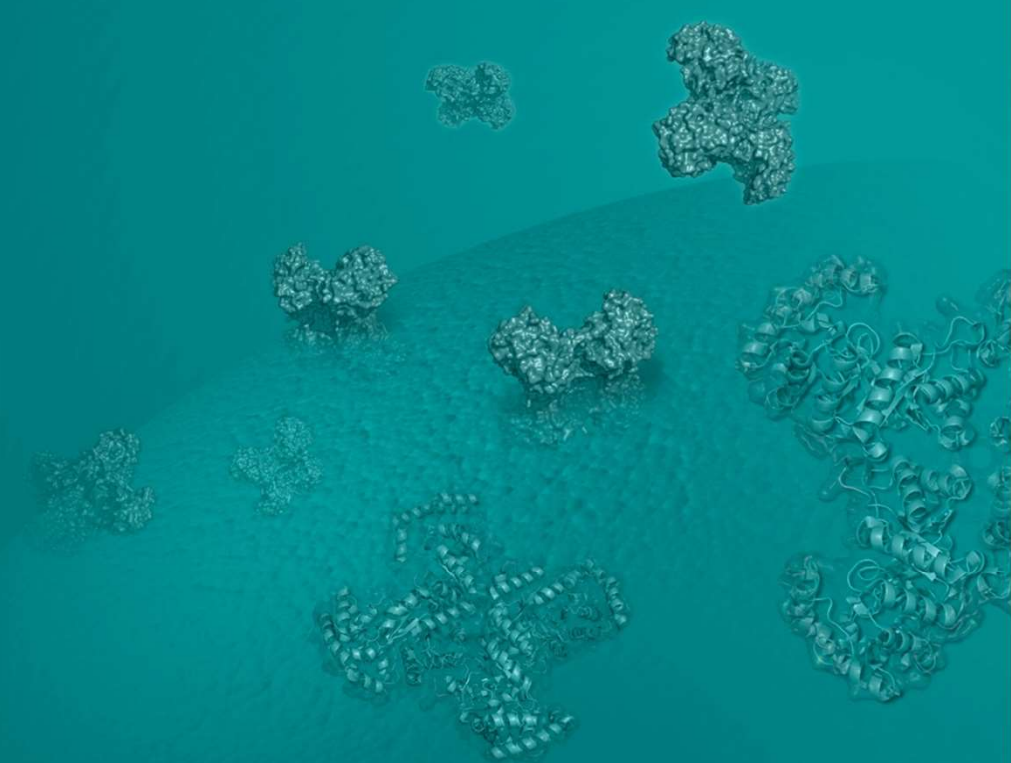
3\_PDBe\_PDBe-KB\_ligand\_view\_tutorial.docx

3\_PDBe\_PDBe-KB\_ligand\_view\_tutorial.pdf

<https://bit.ly/3SkNRpo>

 **PDBe**  
Protein Data Bank in Europe

EMBL-EBI 



# How do I pick the best structure for me? (part2)

# How good is a structure?

- Maps are the real data\*

\* - *X-ray crystallography & cryo electron microscopy*

- Resolution matters, but isn't everything
- Geometry can be evaluated
- Validation reports with 'sliders'



# View Validation information

- Validation information (geometry information etc) can be displayed on the structure



The image shows a screenshot of the PDBe structure viewer interface. On the left, a 3D ribbon representation of a protein structure is displayed, colored in shades of green and yellow. The structure is shown in a perspective view. On the right, the 'Structure Tools' panel is open, showing various settings for the structure. The 'Annotations' section is highlighted with a red box, and the 'Validation Report' option is selected, indicated by a red arrow pointing to it. The 'Validation Report' option has an eye icon and a three-dot menu icon next to it. The 'Structure Tools' panel also shows the structure ID '1OJ6', the type 'Assembly', and the assembly ID '3: Author And Softwar...'. The 'Dynamic Bonds' option is set to 'Off'. The 'SER 51 | C' residue is selected, and the 'Annotations' section is expanded to show 'Validation Report', 'Domains', and 'Quick Styles'. The 'Quick Styles' section has three options: 'Default', 'Stylized', and 'Illustrative'.

# View Validation information

- Validation information (geometry information etc) can be displayed on the structure



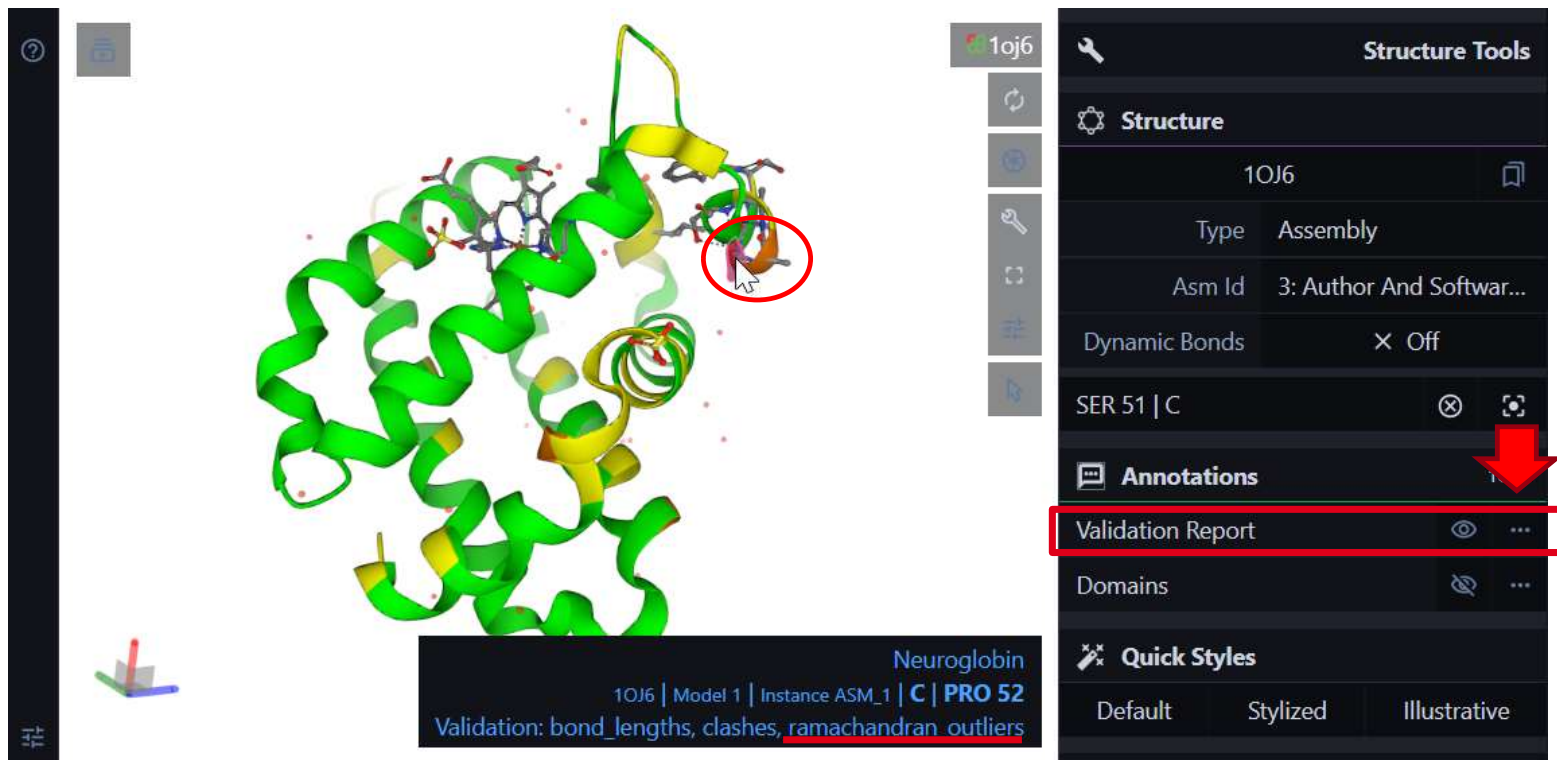
The image shows a 3D ribbon model of the protein Neuroglobin (PDB ID: 1OJ6). The protein is primarily colored green, with some yellow and blue regions. A red circle highlights a specific side chain outlier. The interface includes a top navigation bar, a left sidebar with a search icon, and a right sidebar with a 'Structure Tools' menu. The 'Structure Tools' menu is open, showing options like 'Structure', 'Annotations', and 'Validation Report'. The 'Validation Report' option is highlighted with a red box. Below the model, a tooltip displays the following information: Neuroglobin, 1OJ6 | Model 1 | Instance ASM\_1 | C | ARG 18, and Validation: sidechain\_outliers.

Structure Tools	
<b>Structure</b>	
1OJ6	
Type	Assembly
Asm Id	3: Author And Softwar...
Dynamic Bonds	× Off
SER 51   C	
<b>Annotations</b>	1OJ6
Validation Report	👁 ...
Domains	👁 ...
<b>Quick Styles</b>	
Default	Stylized Illustrative

Neuroglobin  
1OJ6 | Model 1 | Instance ASM\_1 | C | ARG 18  
Validation: sidechain\_outliers

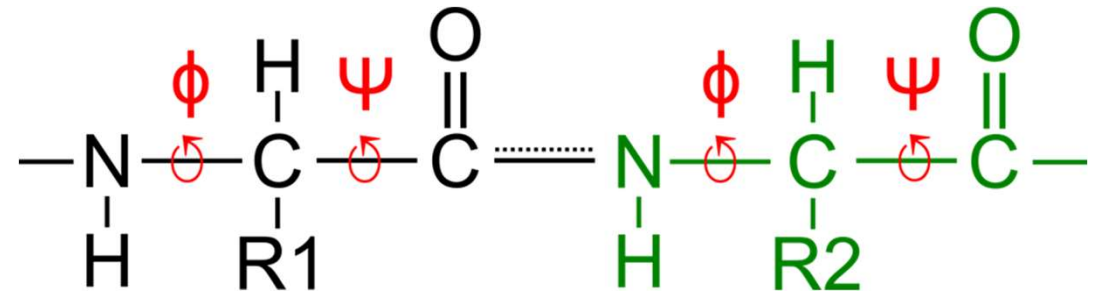
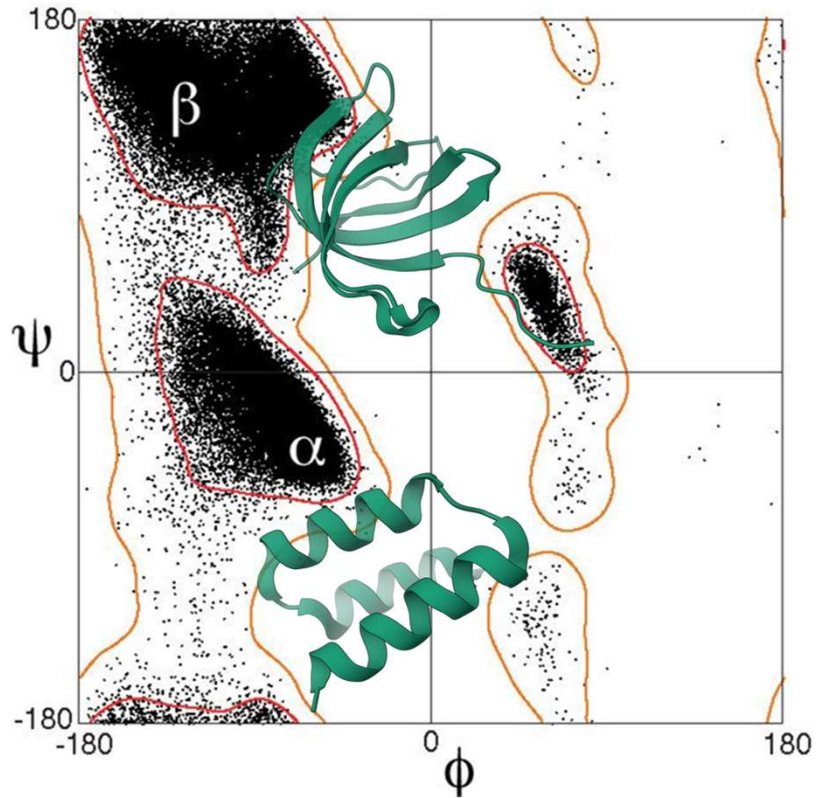
# View Validation information

- Validation information (geometry information etc) can be displayed on the structure



The screenshot displays the PDBe interface for the protein structure 1OJ6. The main view shows a ribbon representation of the protein in green and yellow, with a red circle highlighting a specific region. A sidebar on the right contains the 'Structure Tools' menu, which is open to the 'Annotations' section. The 'Validation Report' option is highlighted with a red box and a red arrow pointing to it. Below the structure, a text box provides details: 'Neuroglobin 1OJ6 | Model 1 | Instance ASM\_1 | C | PRO 52' and 'Validation: bond\_lengths, clashes, ramachandran\_outliers'. The bottom of the page features the URL 'pdbe.org/1OJ6' and the EMBL-EBI logo.

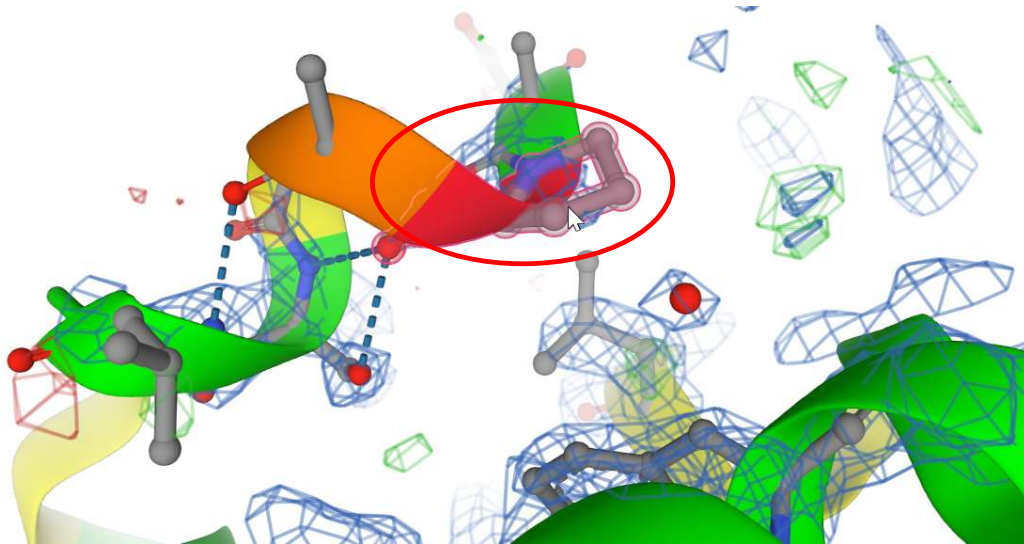
# How we judge geometry: Ramachandran Plot / Ramachandran Outliers



# Geometry and fit to data can go hand-in-hand

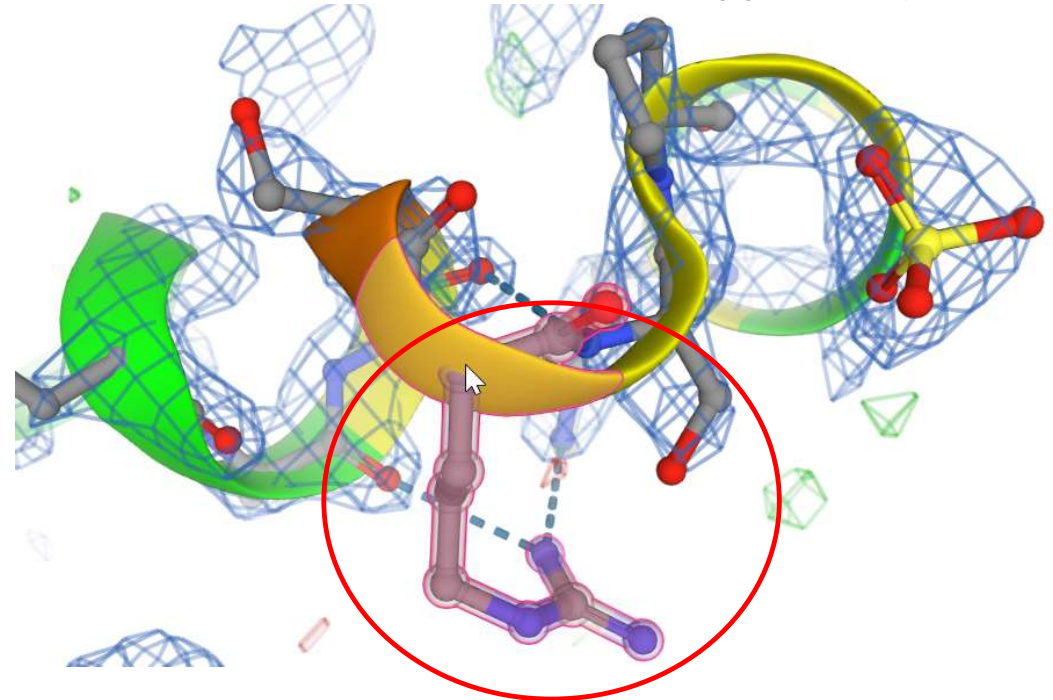
Pro 52 in entry 1oj6 is a Ramachandran outlier  
(and has unusual bond lengths etc)

The strained conformation is NOT supported by data



Arg 18 in entry 1oj6 is a Sidechain outlier

The unusual conformation is NOT supported by data



# Not all structures are created equal!

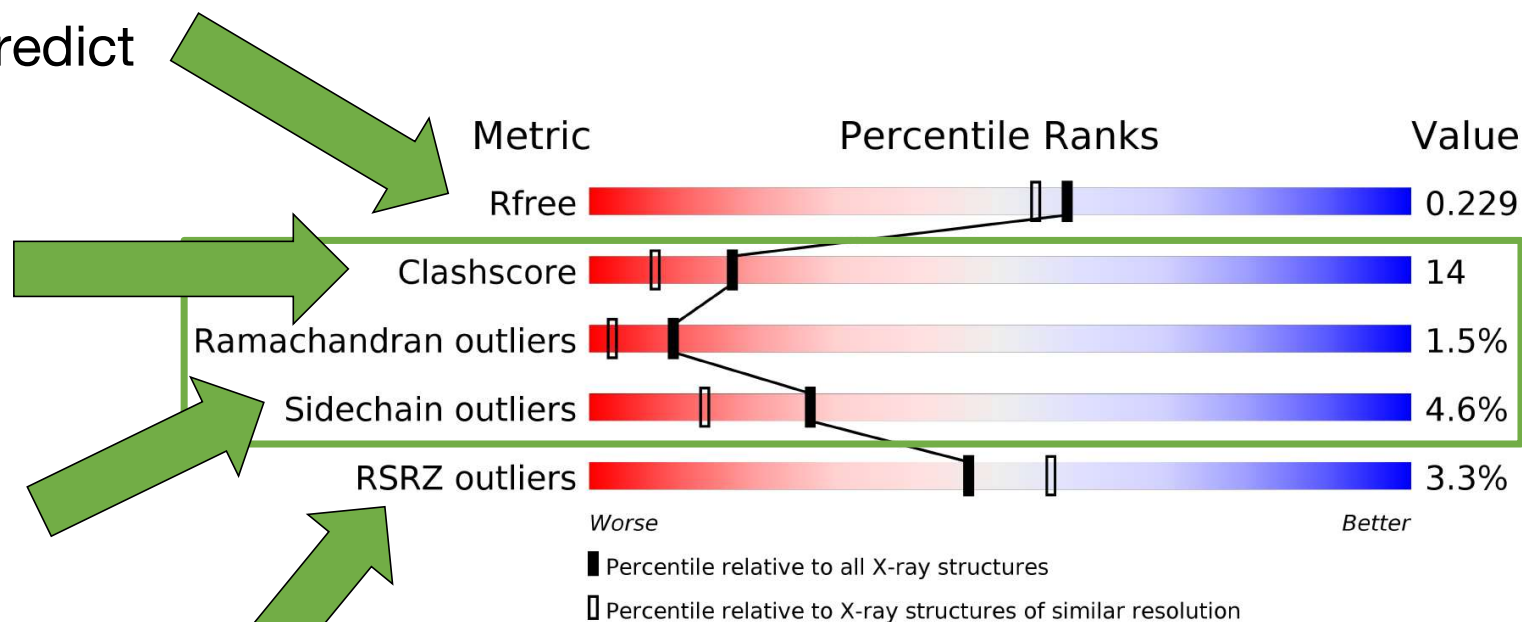
## *Validation sliders are here to help*

How well does the model back-predict the data?

Atoms bumping into each other

Surprising bond angles

Atoms not in electron density



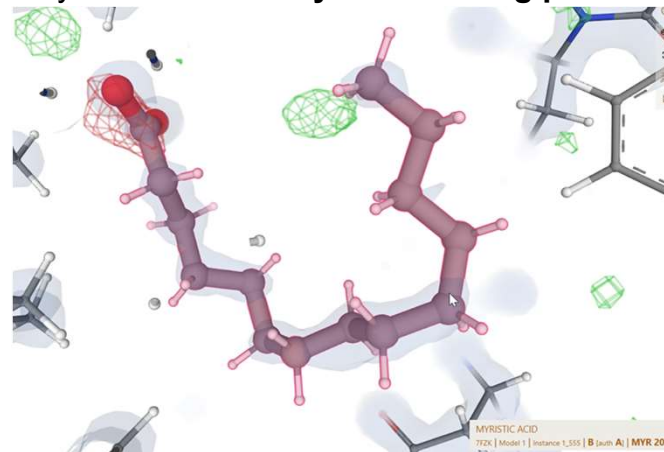
# Viewing maps / electron density in Mol\*

- To initiate electron maps display: clicking on ligand or protein amino acid
- Regular map (blue) '2Fo-Fc' electron density map *should surround atoms*
- **Negative** and **positive** density *highlights extra and missing atoms, respectively*



PDB ID: 7G0X

Myristic acid in Fatty acid-binding protein

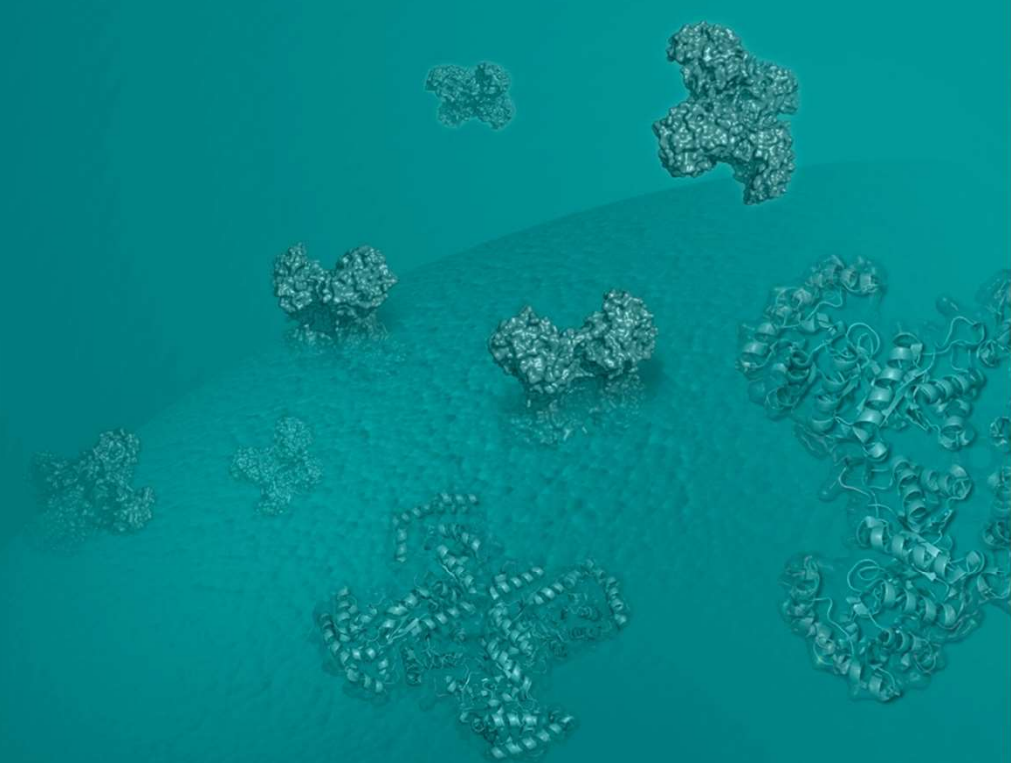


PDB ID: 7FZK

Volume Streaming		7G0X
2Fo-Fc $\sigma$	<input type="range" value="1.5"/>	1.5
Color	<input type="color" value="#0000FF"/>	
Wireframe	<input type="checkbox"/> Off	
Opacity	<input type="range" value="0.15"/>	0.15
Fo-Fc(+ve) $\sigma$	<input type="range" value="3"/>	3
Color	<input type="color" value="#00FF00"/>	
Wireframe	<input checked="" type="checkbox"/> On	
Opacity	<input type="range" value="0.3"/>	0.3
Fo-Fc(-ve) $\sigma$	<input type="range" value="-3"/>	-3
Color	<input type="color" value="#FF0000"/>	
Wireframe	<input checked="" type="checkbox"/> On	
Opacity	<input type="range" value="0.3"/>	0.3
Entry	7g0x	
View	Around Focus	

# PDBe-KB aggregated views of proteins

Genevieve Evans





In January 2023, we passed the 200,000 milestone. As of today, ~214,000 structures

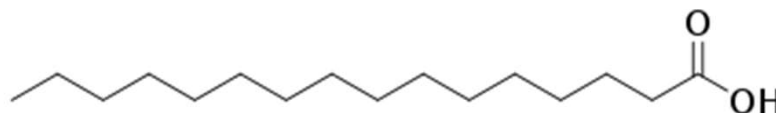
## Not all ~215,000 structures are unique molecules

There are ~60k structures of human proteins

But only ~15k different human proteins

(~4K different protein families)

Why?

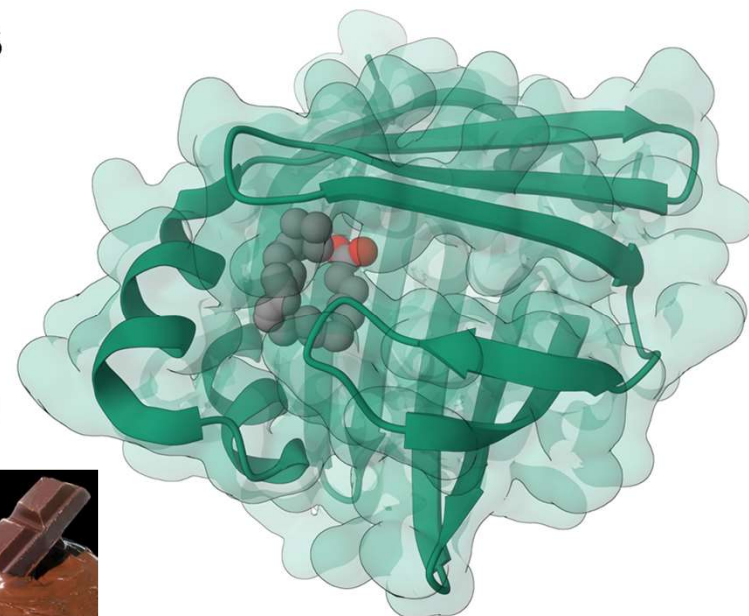


Solved by different experimental methods

With different compounds bound

By *different people*

In *different conditions*



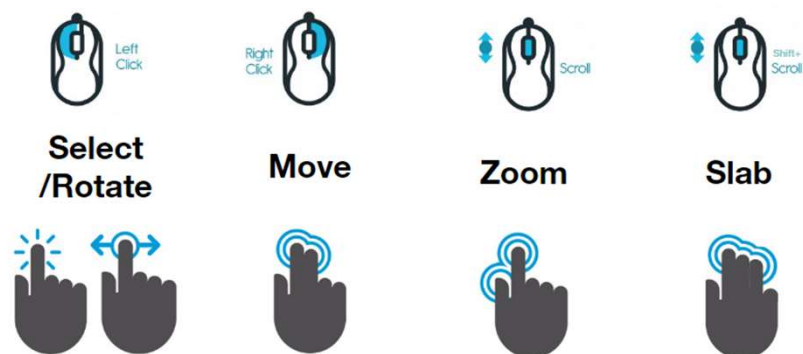
Fatty acid-binding protein: 242 structures,  
227 different small molecules

# Visualization in 3D with Mol\*

- Browser-based

- Interactive & many features
- Generate images / animations in browser window
- Relatively easy to navigate

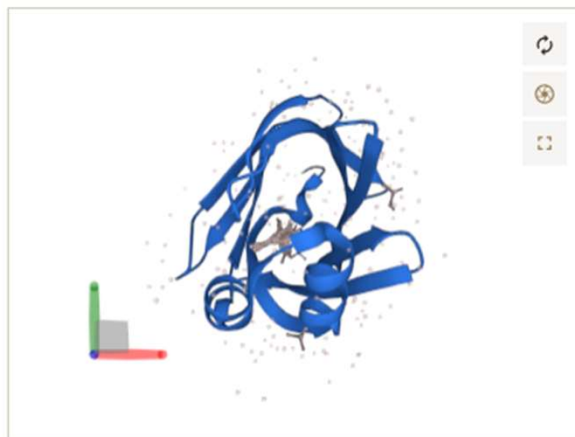
## Mol\* controls



# Protein KB pages

Representative structure for UniProt P15090

PDB chain with highest data quality, coverage and best resolution ⓘ



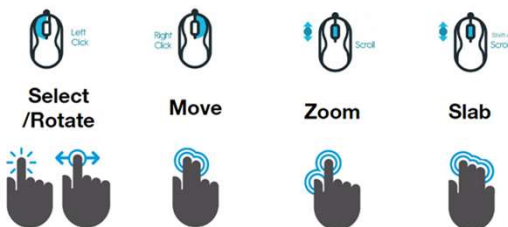
PDB chain shown: **7fwa A** [go to PDBe](#)

UniProt residues 1 - 132

Coverage: 100%

SOURCE: <https://www.ebi.ac.uk/pdbe/entry/pdb/7fwa/protein/1>

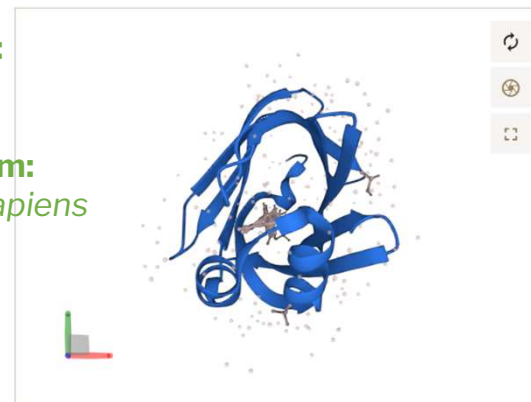
Mol\* controls



## Fatty acid-binding protein

UniProt:  
P15090

Organism:  
*Homo sapiens*



242  
Structures



227  
Ligands



0  
Interactions



Annotations




17  
Similarity



51  
Publications



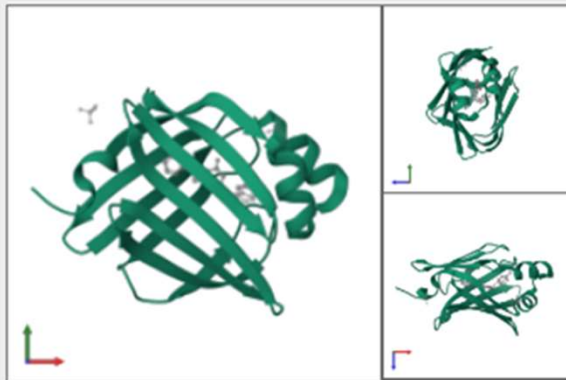
# Protein Entry pages

 Protein Data Bank in Europe  
Bringing Structure to  
Biology

PROTEIN DATA BANK

Search

**7fwa** › Fatty acid-binding protein, adipocyte



Chain: A

Length: 135 amino acids

Theoretical weight: 15.02 KDa

Source organism: *Homo sapiens*

Expression system: *Escherichia coli* BL21(DE3)

UniProt:

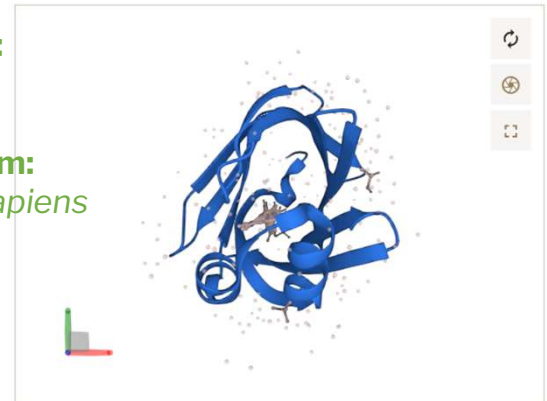
◦ Canonical:  **P15090**  (Residues: 1-132; Coverage: 100%)

**Assembly:  
monomer**

## Fatty acid-binding protein

UniProt:  
P15090

Organism:  
*Homo sapiens*



242  
Structures



227  
Ligands



0  
Interactions



Annotations



17  
Similarity



51  
Publications

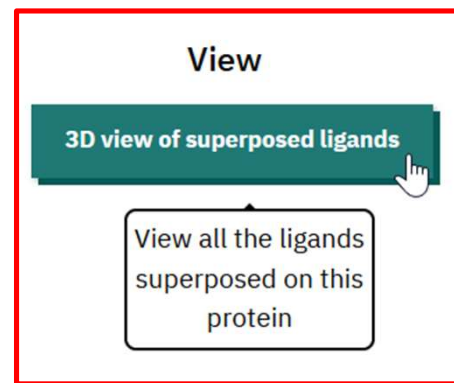


# Protein KB pages

- Structural context for a given protein

UniProt

P15090



SOURCE: <https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P15090/ligands>

A screenshot of the PDBeKB.org page for the protein 'Fatty acid-binding protein' (UniProt: P15090). The page features a 3D ribbon model of the protein in blue, with a small bar chart below it. To the left of the model, the UniProt ID 'P15090' and the organism 'Homo sapiens' are listed. Below the model, there are six summary cards: '242 Structures', '227 Ligands', '0 Interactions', 'Annotations', '17 Similarity', and '51 Publications'. Each card contains an icon representing its category.



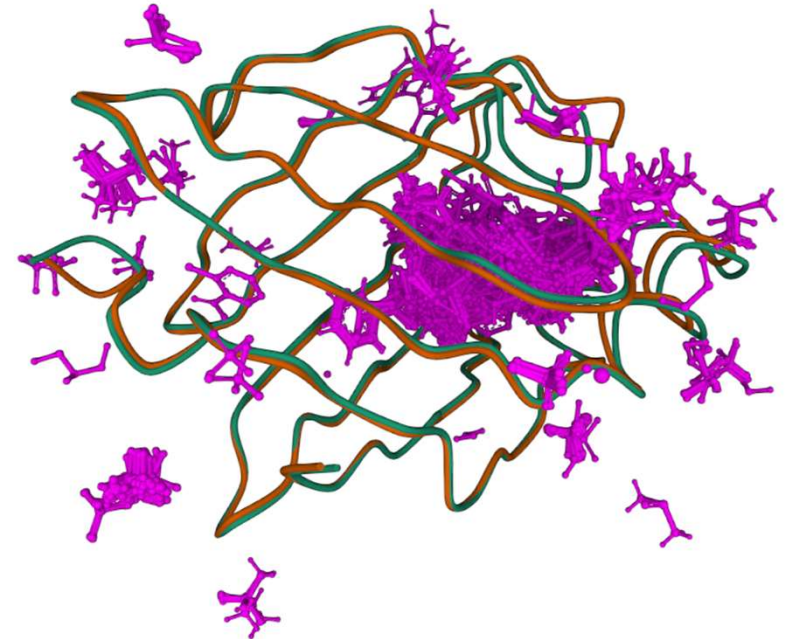
# Protein KB pages -- superposition

- We have a process that calculates superposition for individual proteins in the whole PDB archive

→ YouTube tutorial:

<https://www.youtube.com/watch?v=fqP50iFZHtY&t=17s>

- These data are displayed in: 'superposed clusters' (orange vs green)
- But:  
The clusters do not necessarily correspond to biological functions
- Allows you to view common interaction sites and important regions of the protein



Fatty acid-binding protein: 242 structures,  
227 different small molecules

SOURCE: <https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P15090>

# Protein KB pages -- superposition

- We have a process that calculates superposition for individual proteins in the whole PDB archive

→ YouTube tutorial:

<https://www.youtube.com/watch?v=fqP50iFZHtY&t=17s>

- These data are displayed in: 'superposed clusters' (orange vs green)
- But:  
The clusters do not necessarily correspond to biological functions
- Allows you to view common interaction sites and important regions of the protein



Salicylate Synthase: 14 structures,  
28 different small molecules

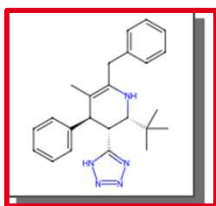
SOURCE: <https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P15090>

More information: <https://pubs.acs.org/doi/10.1021/bi3002067>

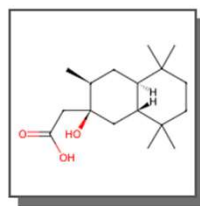
# Protein KB pages -- ligands

## Overview of ligands and binding residues

- Legends:
- Annotated small molecules
  - Other small molecules
  - Not interacting small molecules



**UUX**   
Found in **2** PDB entries [↗](#)

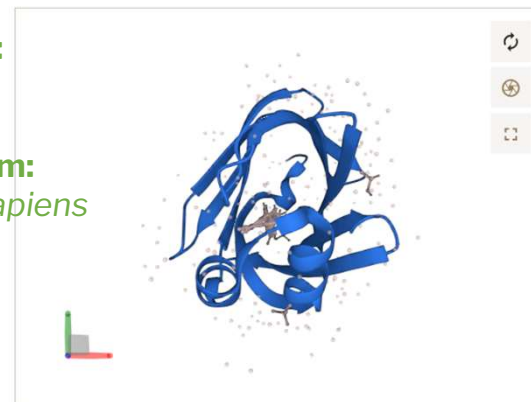


**WOO**   
Found in **1** PDB entry [↗](#)

# Fatty acid-binding protein

UniProt:  
P15090

Organism:  
*Homo sapiens*

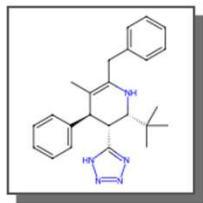




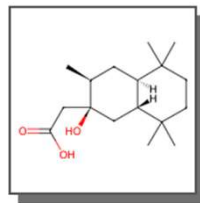
# Protein KB pages -- ligands

## Overview of ligands and binding residues

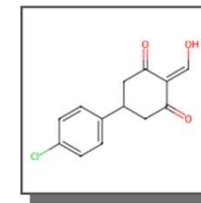
- Legends:
- Annotated small molecules
  - Other small molecules
  - Not interacting small molecules



**UUX**   
Found in [2 PDB entries](#)



**W00**   
Found in [1 PDB entry](#)



**M69**   
Found in [1 PDB entry](#)



# Protein KB pages

- Structural context

UniProt ID

= specific protein / protein component  
from a specific organism

- Information clustered in different sections

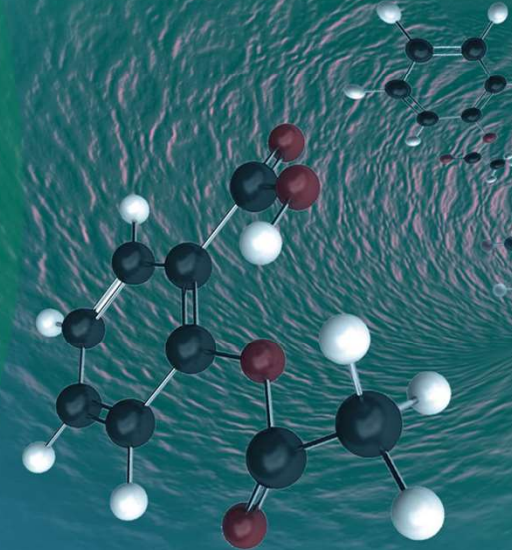
- Structure coverage
- Ligands and binding sites
- Protein interactions
- Additional annotations
- Proteins with similar sequence
- Publications

**We've used APIs  
and our analysis  
pipeline  
to pre-cluster  
& analysis  
multiple structures  
for you 😊**





# Acknowledgements



**Genevieve Evans**

PDBe

-  [pdhelp@ebi.ac.uk](mailto:pdhelp@ebi.ac.uk)
-  [proteindatabank](https://www.facebook.com/proteindatabank)
-  [@PDBeurope](https://twitter.com/PDBeurope)
-  [proteindatabank](https://www.youtube.com/proteindatabank)
-  [pdbeurope](https://www.instagram.com/pdbeurope)
-  [pdbart](https://www.pinterest.com/pdbart)



# Feedback Welcome Thank you!

PDBe.org



[pdbhelp@ebi.ac.uk](mailto:pdbhelp@ebi.ac.uk)



[proteindatabank](https://www.facebook.com/proteindatabank)



[@PDBEurope](https://twitter.com/PDBEurope)



[proteindatabank](https://www.youtube.com/proteindatabank)



[pdbeurope](https://www.instagram.com/pdbeurope)



[pdbart](https://www.pinterest.com/pdbart)

