

# PDBe resources to help with starting model selection for molecular dynamics simulations

**Sudakshina Ganguly**

Scientific Database Curator

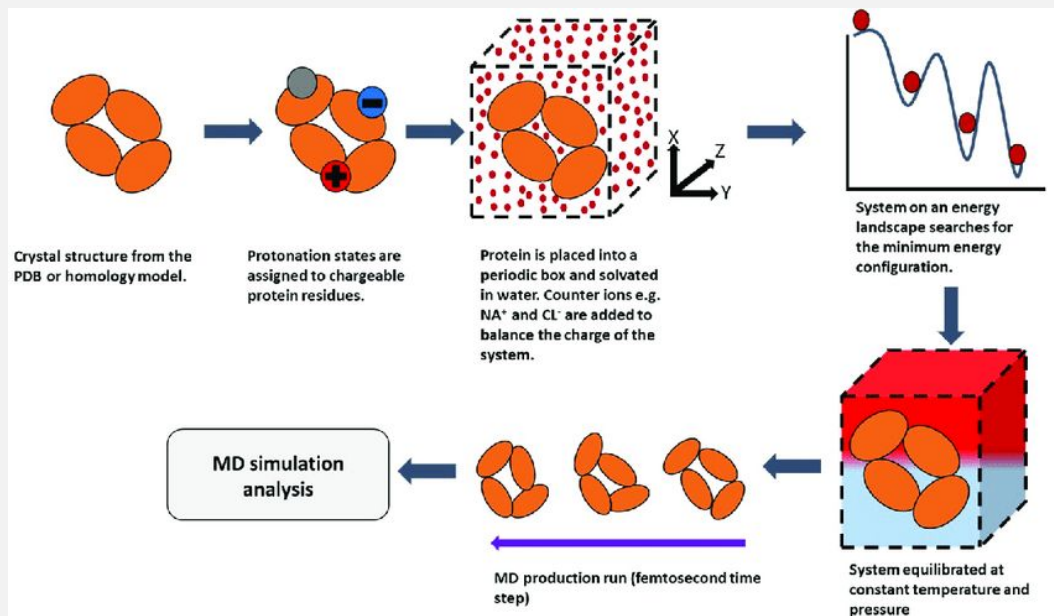


Protein Data Bank in Europe

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



# Molecular Dynamics Simulations



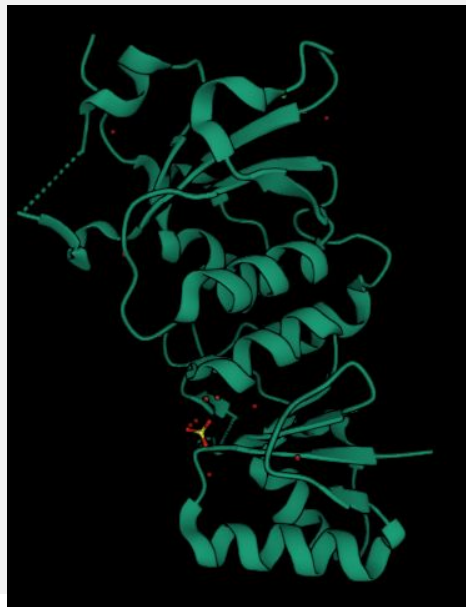
# Starting model

- Why is a starting model important?
- An example where different models led to different results?

**PDB id 3pxc**

BRCA BRCT  
protein

UniProt:  
P38398



**PDB id 3pxe**

BRCA BRCT  
protein

UniProt:  
P38398



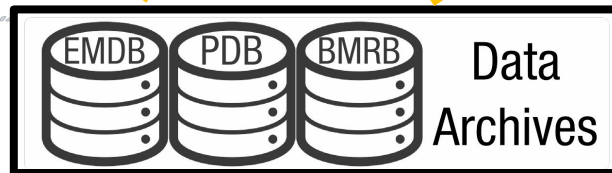
**Relevant difference in clashes**

# Objectives of this seminar

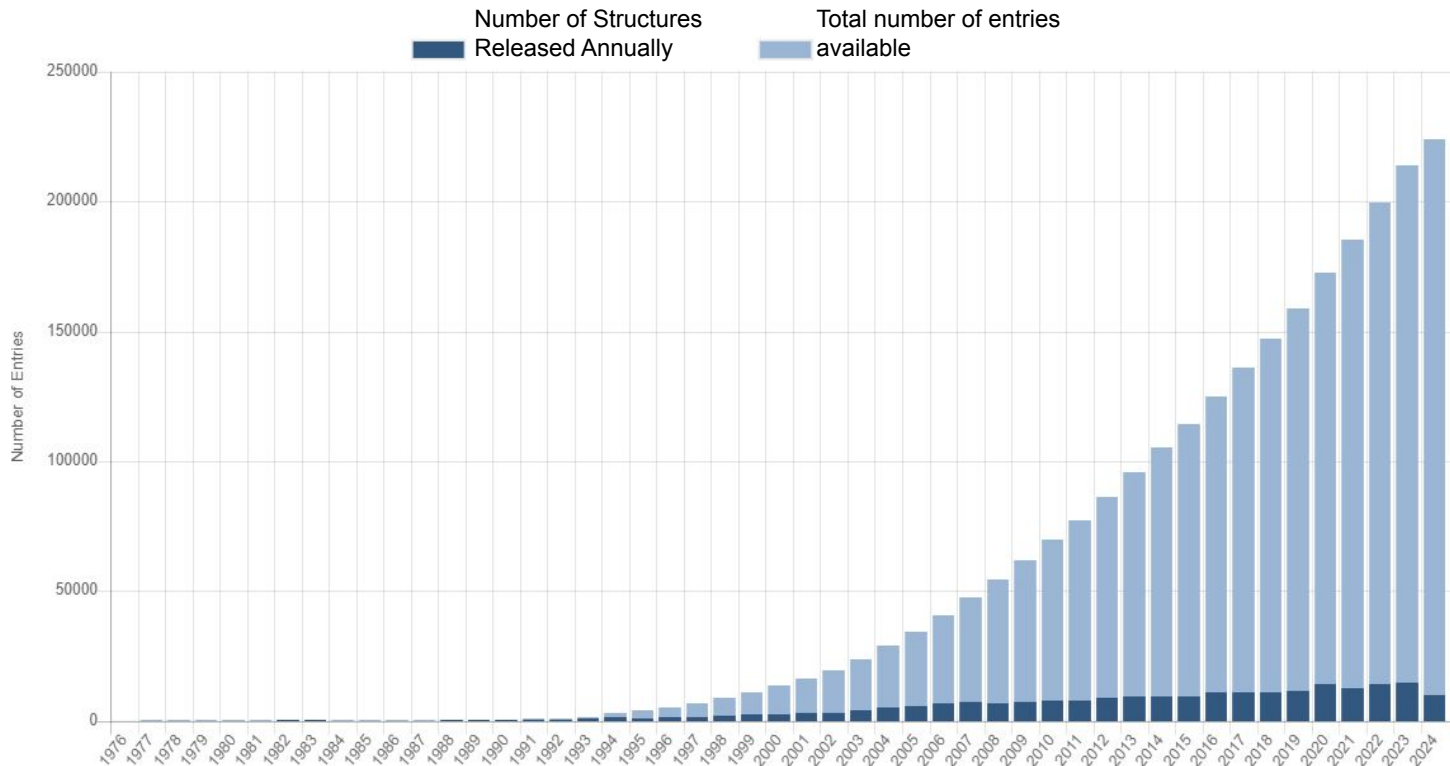
- Learn where and how to search for a structure
- Understanding the validation metrics
- Assimilate methods to visualise and compare structures
- Use an existing MD frame as an initial structure

# The Worldwide Protein Data Bank (wwPDB)

Residue ID	Atom Name	Number of Shifts	Minimum Shift	Maximum Shift	Average Shift	Standard Deviation	Shift Outliers
H	95419	-0.914	131.25	8.193	0.843	166	
HA	62374	-2.52	17.870	4.244	0.441	1218	
MB	65439	-14.040	5.48	1.553	0.278	1089	
C	61640	0.037	187.2	177.747	3.055	41	
CA	84427	17.027	354.698	53.165	2.706	89	
CB	80347	-45.993	318.868	19.043	2.979	205	
N	91488	0.849	766	123.549	5.867	105	
H	63333	0.011	178	8.240	1.019	43	
HA	42696	1.212	12.57	4.288	0.467	510	
HB	39833	-4.78	27.330	1.790	0.309	502	



# How many structures in the PDB?



**224572**  
structures  
as of today,  
10th Sep  
2024

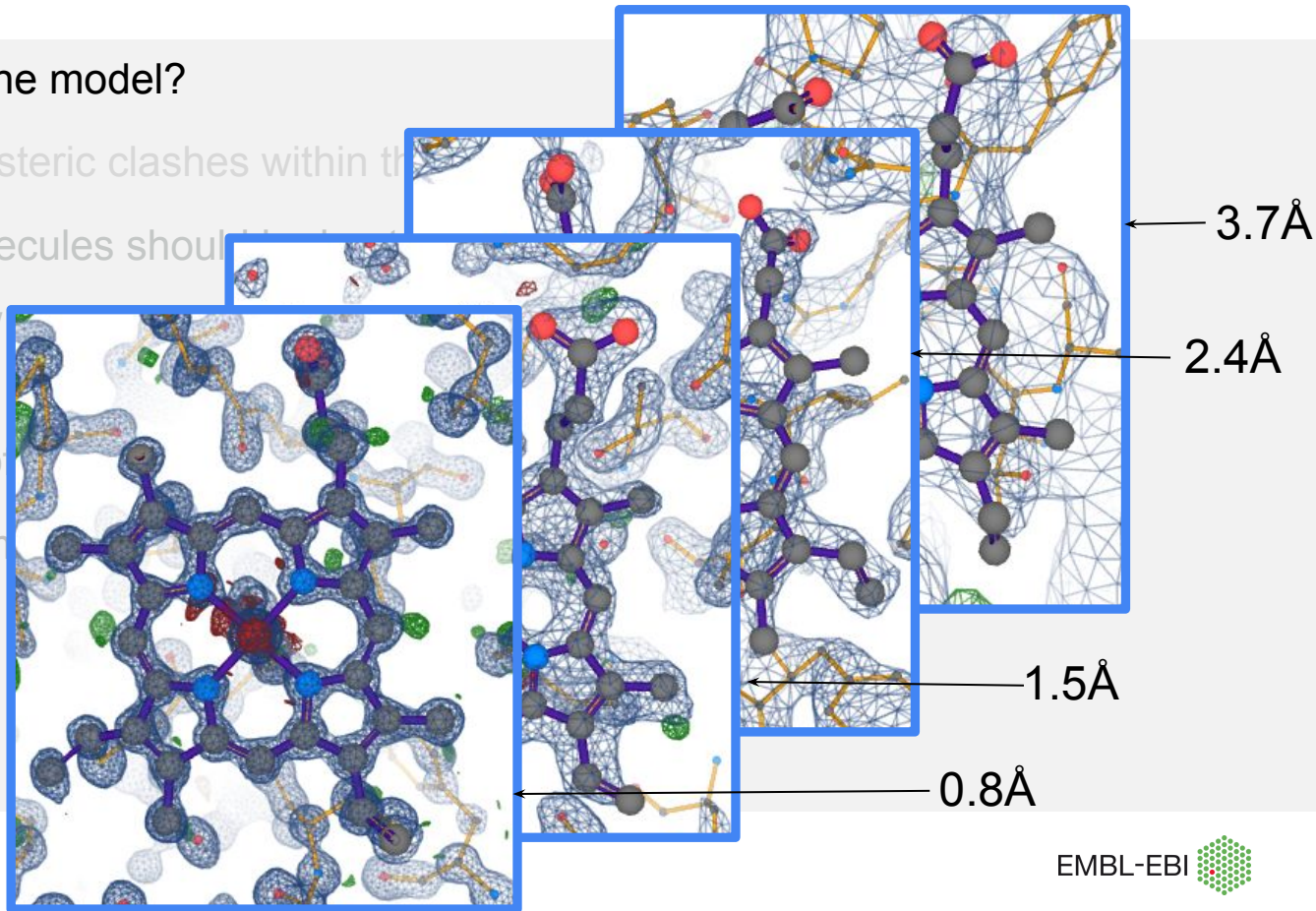
# Important points for selecting the starting model

- What is the resolution for the model?
- Are there any outstanding steric clashes within the structure?
- Assess whether water molecules should be kept in the structure or stripped.
- Are there missing regions / loops of the protein and do they have to be modelled using another tool?
- What is the conformation of the protein?
- What is the experimental method used to derive the structure?



# Important points for selecting the starting model

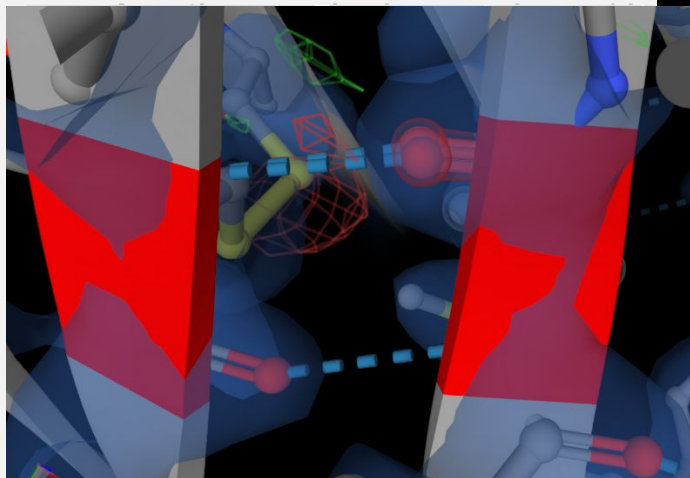
- What is the resolution for the model?
- Are there any outstanding steric clashes within the model?
- Assess whether water molecules should be included or not?
- Are there missing regions / missing atoms? Should you use another tool?
- What is the conformation of the side chains?
- What is the experimental map?





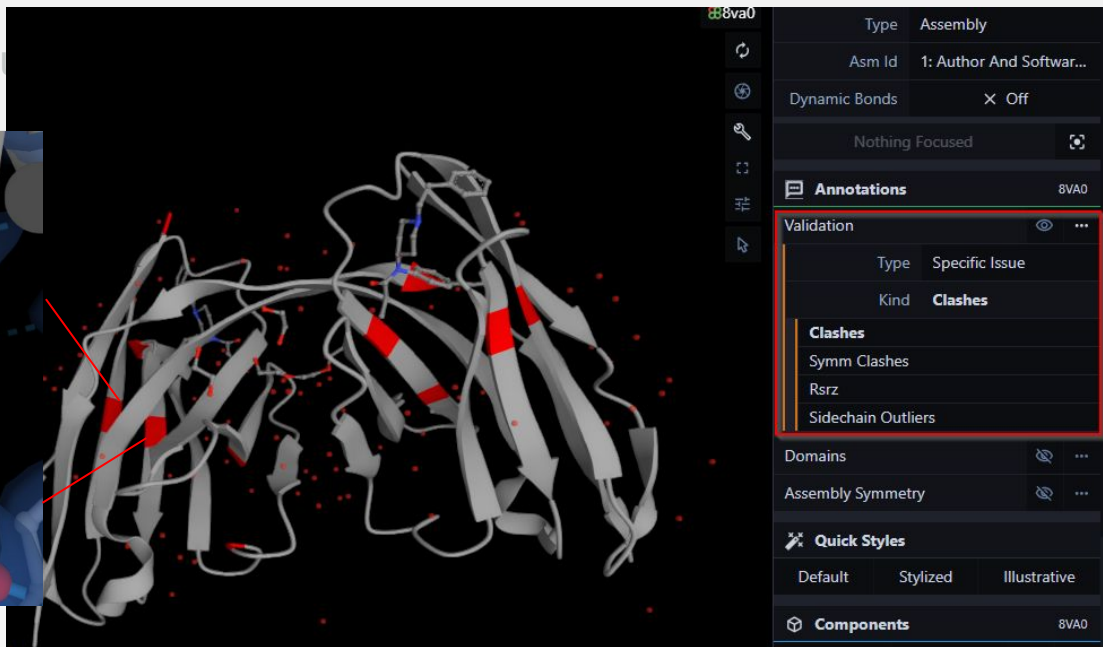
# Important points for selecting the starting model

- What is the resolution for the model?
- Are there any outstanding steric clashes within the structure?
- Assess whether water molecules



+ Assembly composition (9)

+ Assembly polymer count (42)



A screenshot of a software interface, likely a molecular visualization tool, showing a protein structure with a red mesh overlay. The interface includes a sidebar menu on the right with various options. The 'Annotations' section is expanded, showing a 'Validation' panel with a table of issues. The table has columns for 'Type' and 'Specific Issue'. The 'Kind' is 'Clashes'. The 'Clashes' section is further expanded, showing 'Symm Clashes', 'Rsrz', and 'Sidechain Outliers'. The 'Quick Styles' section is also visible, with options for 'Default', 'Stylized', and 'Illustrative'. The 'Components' section is at the bottom, showing '8VAO'.

Type	Specific Issue
Kind	Clashes
<b>Clashes</b>	
	Symm Clashes
	Rsrz
	Sidechain Outliers

Domains [icon] ...

Assembly Symmetry [icon] ...

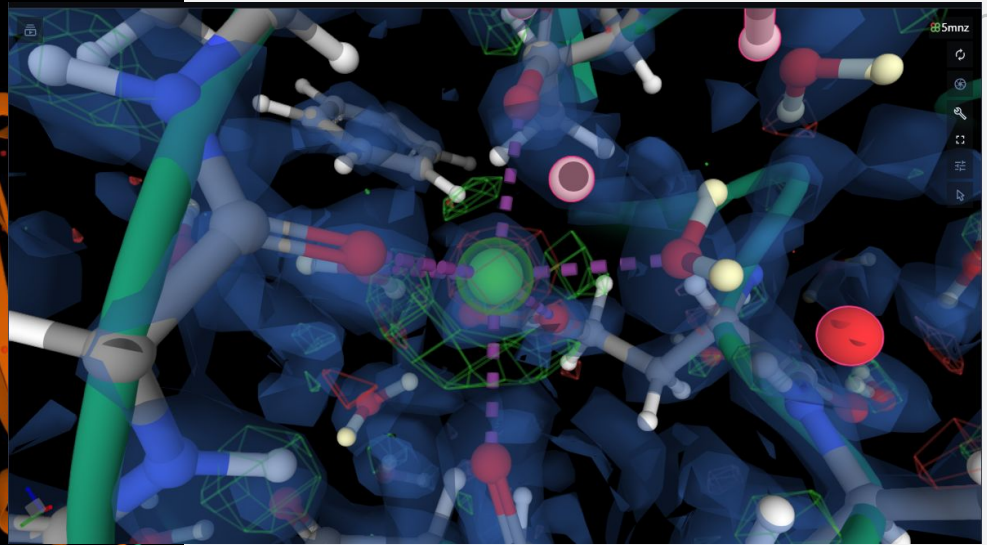
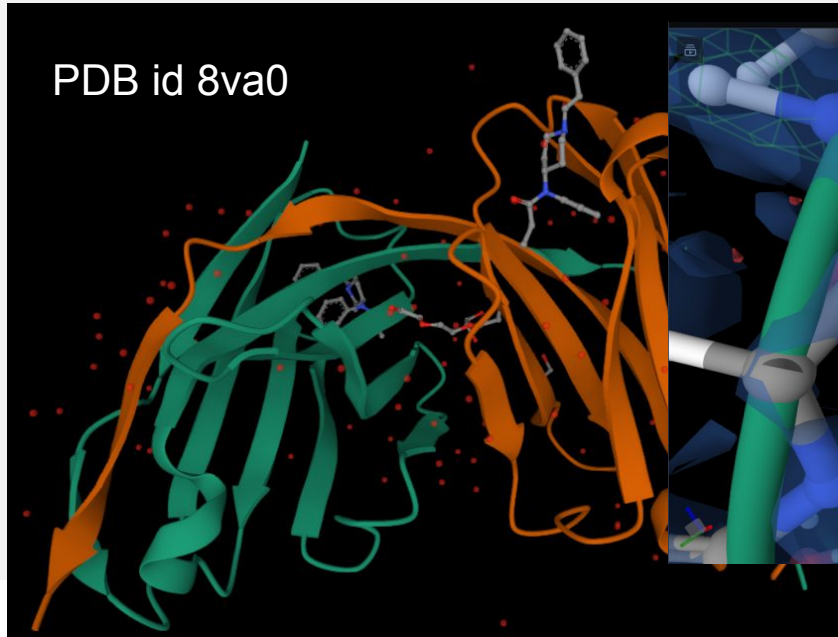
**Quick Styles**

Default Stylized Illustrative

**Components** 8VAO

# Important points for selecting the starting model

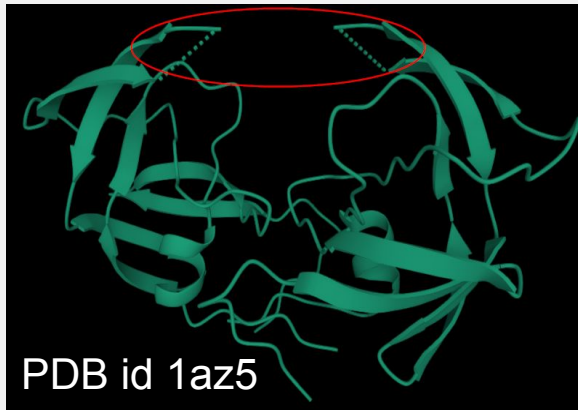
- What is the resolution for the model?
- Are there any outstanding steric clashes within the structure?
- Assess whether water molecules should be kept in the structure or stripped.



PDB id 5mnz

# Important points for selecting the starting model

- What is the resolution for the model?
- Are there any outstanding steric clashes within the structure?
- Assess whether water molecules should be kept in the structure or stripped.
- Are there missing regions / loops of the protein and do they have to be modelled using another tool?

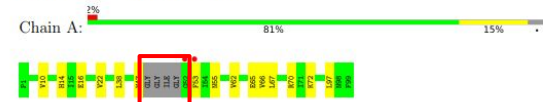


```
loop_
_pdbx_unobs_or_zero_occ_residues.id
_pdbx_unobs_or_zero_occ_residues.PDB_model_num
_pdbx_unobs_or_zero_occ_residues.polymer_flag
_pdbx_unobs_or_zero_occ_residues.occupancy_flag
_pdbx_unobs_or_zero_occ_residues.auth_asym_id
_pdbx_unobs_or_zero_occ_residues.auth_comp_id
_pdbx_unobs_or_zero_occ_residues.auth_seq_id
_pdbx_unobs_or_zero_occ_residues.PDB_ins_code
_pdbx_unobs_or_zero_occ_residues.label_asym_id
_pdbx_unobs_or_zero_occ_residues.label_comp_id
_pdbx_unobs_or_zero_occ_residues.label_seq_id
1 1 Y 1 A GLY 48 ? A GLY 48
2 1 Y 1 A GLY 49 ? A GLY 49
3 1 Y 1 A ILE 50 ? A ILE 50
4 1 Y 1 A GLY 51 ? A GLY 51
```

## 3 Residue-property plots [i](#)

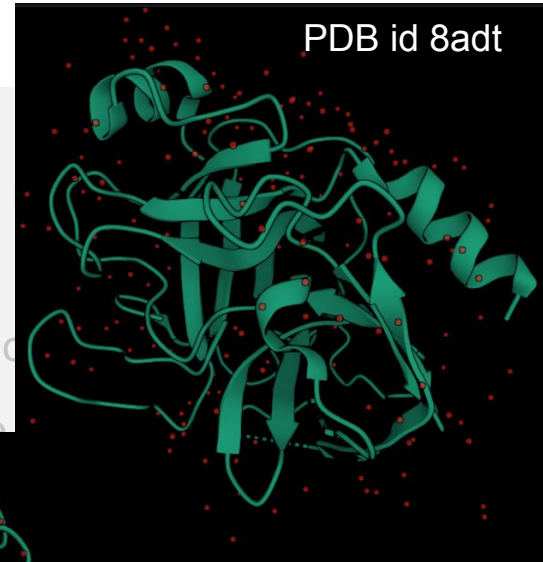
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: SIV PROTEASE



# Important points for selecting the starting model

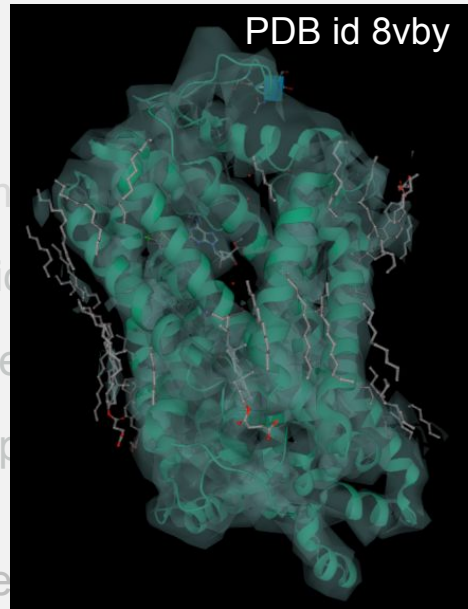
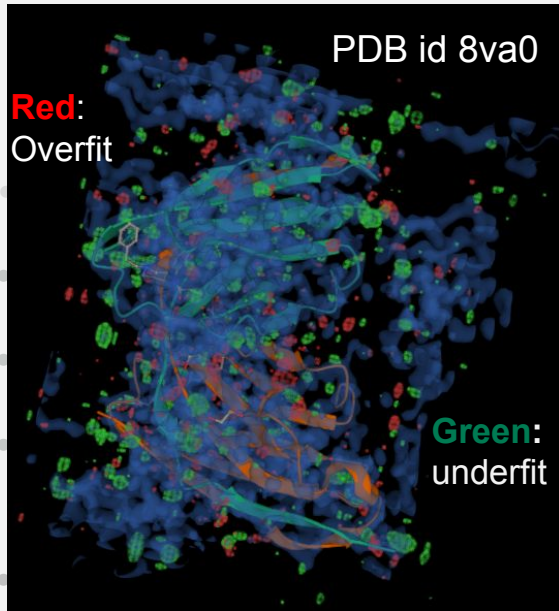
- What is the resolution for the model?
- Are there any outstanding steric clashes within the structure?
- Assess whether water molecules should be kept in the structure or not
- Are there missing regions / loops of the protein and do they have a good alternative in another tool?
- What is the conformation of the protein?
- What is the experimental method used to determine the structure?



Serine protease 1  
UniProt: **P00760**



# Important points for selecting the starting model



- What is the experimental method used to derive the structure?

# Validation sliders: what you need to know

## PDBe > 8xi5

Structure of Eastern Equine Encephalitis VLP in complex with the receptor VLDLR LA3-5

### Source organisms:

- Eastern equine encephalitis virus
- Homo sapiens

### Primary publication:

The receptor VLDLR binds Eastern Equine Encephalitis virus through multiple distinct modes.

Cao D, Ma B, Cao Z, Xu X, Zhang X, Xiang Y

Nat Commun 15 6866 (2024)

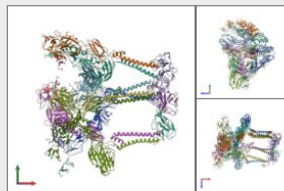
PMID: 39127734

Related structures: [EMD-38371](#)

## Electron Microscopy 3.4Å resolution

Released: 28 Aug 2024

DOI: [10.2210/pdb8xi5/pdb](#)



## Quick links

### 8xi5 overview

- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

View

Downloads

Archive mmCIF file

Updated mmCIF file

PDB file

PDB header

PDB file (gz)

PDBML

PDBML (ATOM lines)

PDBML (no atoms)

Assembly composition XML

Assembly 1 (mmCIF; gz)

Assembly 1 (atom only; mmCIF)

FASTA (Entry)

SIFTS XML file with residue-level mappings

Summary report (PDF)

Full report (PDF)

Percentile plot (PNG)

Percentile plot (SVG)

Validation data (XML)

PDB format is not being supported any longer

mmCIF format is the new 'master' format!

More info on the mmCIF format -

<https://mmcif.wwpdb.org/docs/user-guide/guide.html>

## Function and Biology

Details

### Reaction catalysed:

Autocatalytic release of the core protein from the N-terminus of the togavirus structural polyprotein by hydrolysis of a -Trp|-Ser- bond.

Biochemical function: not assigned

Biological process: not assigned

Cellular component: not assigned

## Ligands and Environments

### 1 bound ligand:

Ca<sup>2+</sup>

8 x CA

No modified residues

## Structure analysis

Details

Assembly composition: hetero icosamer (preferred)

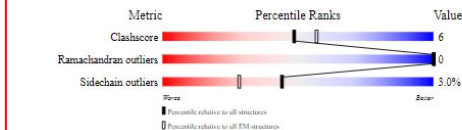
Assembly name: Very low-density lipoprotein receptor and Spike glycoprotein E1 (preferred)

PDBe Complex ID: PDB-CPX-250800 (preferred)

Entry contents: 5 distinct polypeptide molecules

## Experiments and Validation

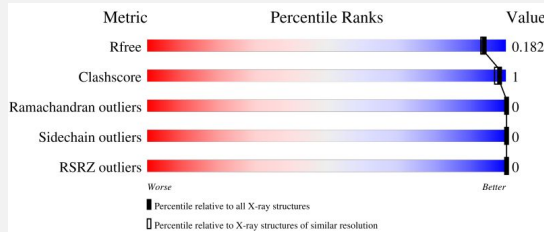
Details



# Validation sliders: what you need to know

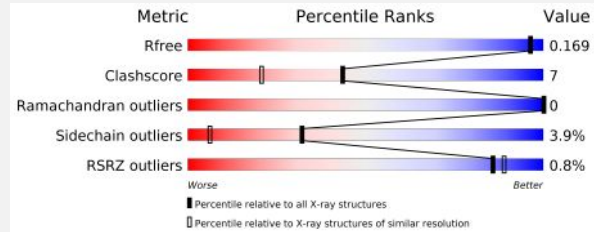
- Overall quality of a specific structure can be assessed
- Overall assessment of the quality of the modelling of the structure compared to a number of key validation metrics
- Includes both geometric validation and fit of the model to the experimental data

**PDB id 2xhh**



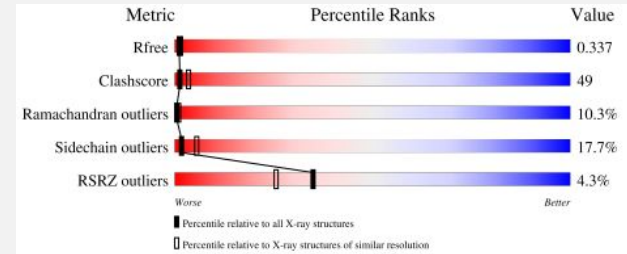
**High quality**

**PDB id 5ovo**



**Medium quality**

**PDB id 3Inn**

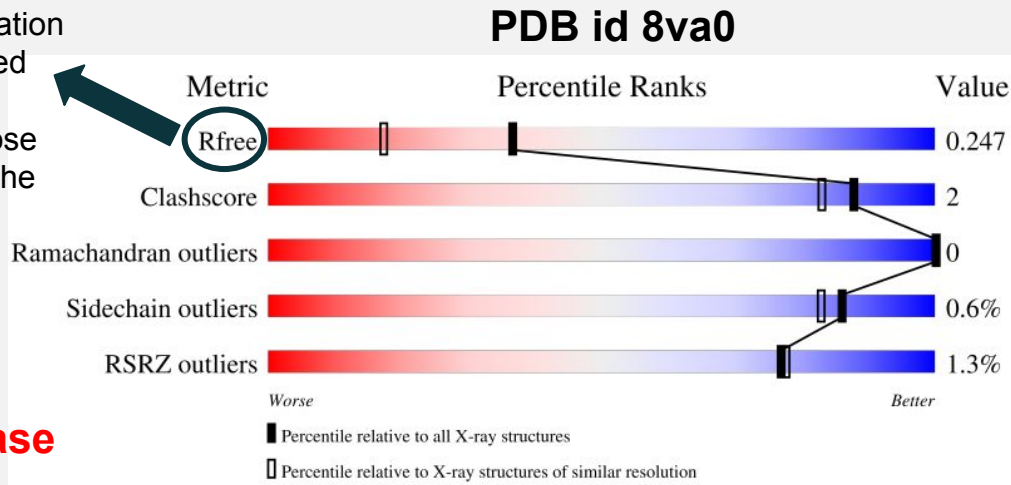


**Low quality**



# Validation sliders: what you need to know

measures correlation between observed structure factor amplitudes & those calculated from the model



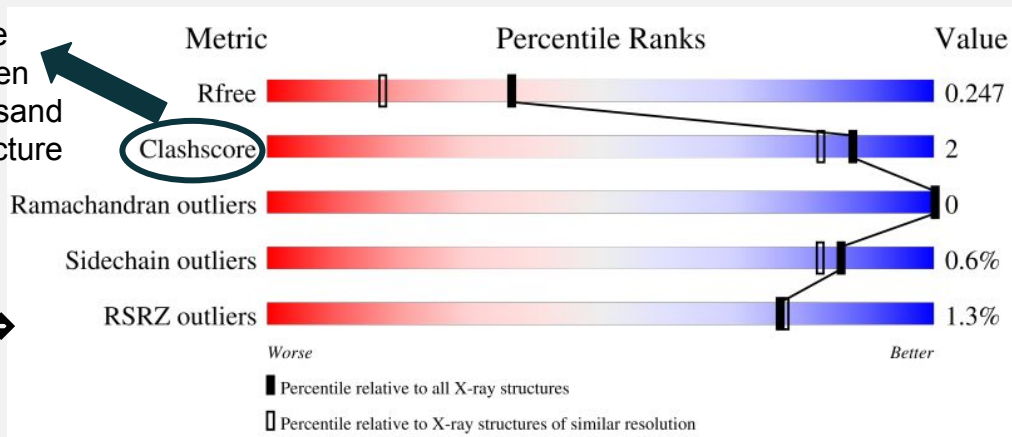
**Rfree increase**  
⇒ quality  
**decrease**

**Rfree is specific for X-ray diffraction entries**

# Validation sliders: what you need to know

quantifies the number of close contacts between atoms per thousand atoms of a structure

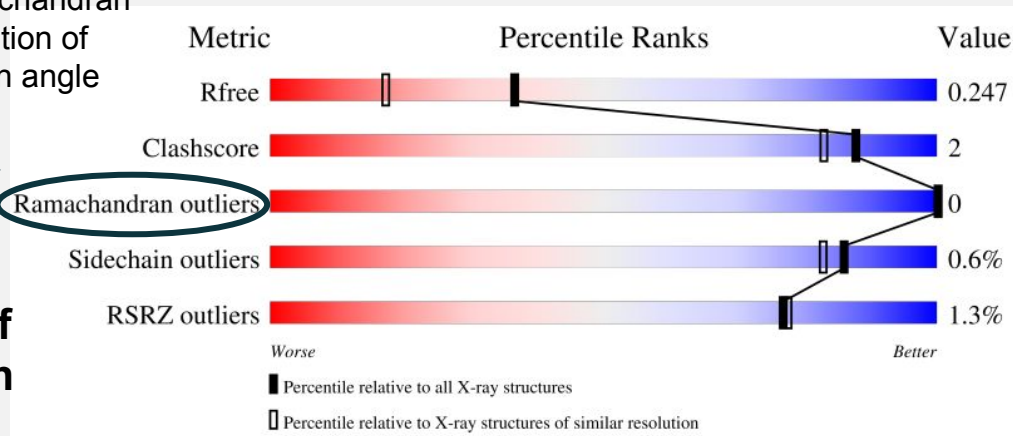
**Clashcore**  
**increases** ⇒  
**quality**  
**decreases**



# Validation sliders: what you need to know

A standard residue is identified as a Ramachandran outlier if the combination of backbone  $\phi$ - $\psi$  torsion angle values is unusual

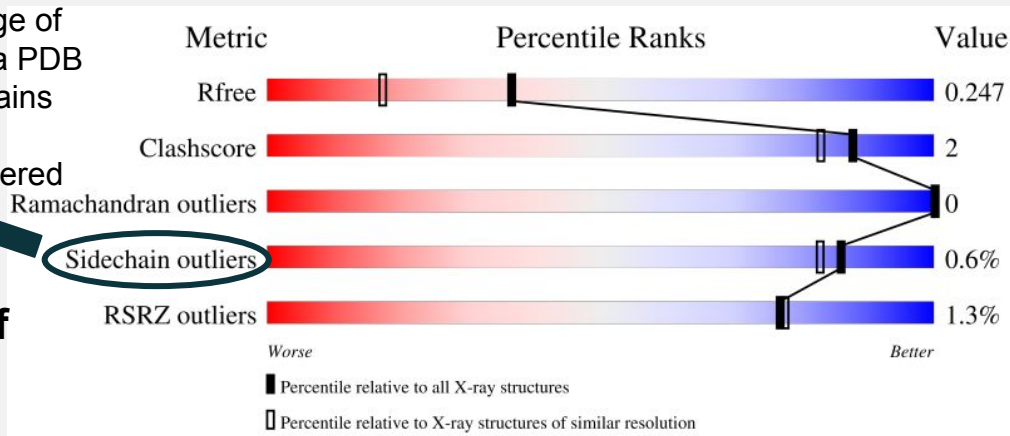
**Low** number of Ramachandran outliers  $\Rightarrow$  **higher** quality



# Validation sliders: what you need to know

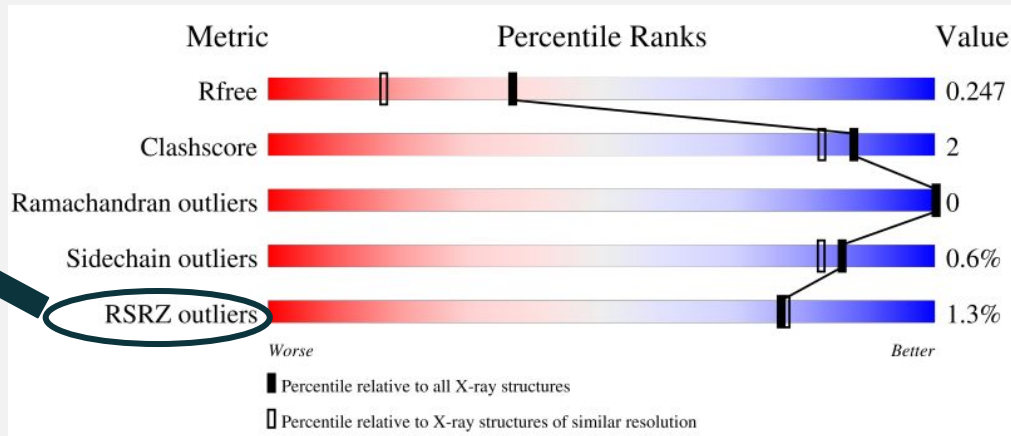
Represents percentage of standard residues in a PDB structure with side chains whose torsion angle combination is considered to be an outlier

**Low** number of sidechain outliers ⇒ **higher** quality



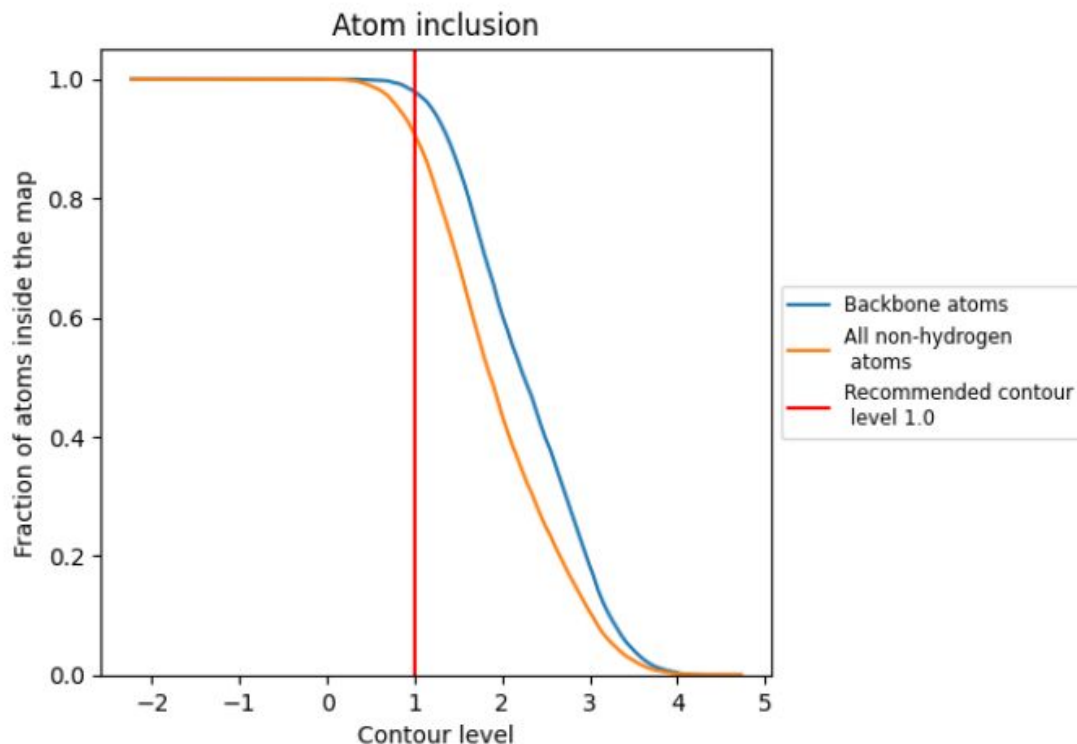
# Validation sliders: what you need to know

Quantify deviation of bond angles and bond lengths in standard residues of a PDB structure



**Low** number of  
**RSRZ outliers** ⇒  
**higher** quality

# Data quality for EM structures

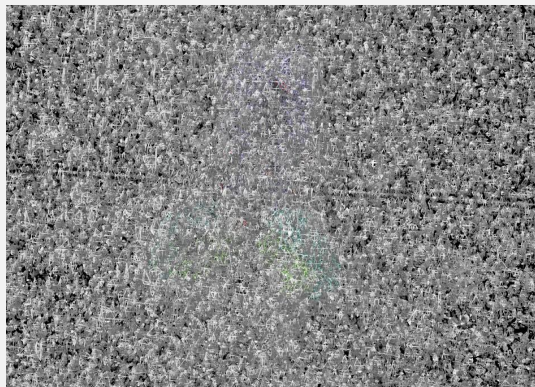


At the recommended contour level, 98% of all backbone atoms, 91% of all non-hydrogen atoms, are inside the map.

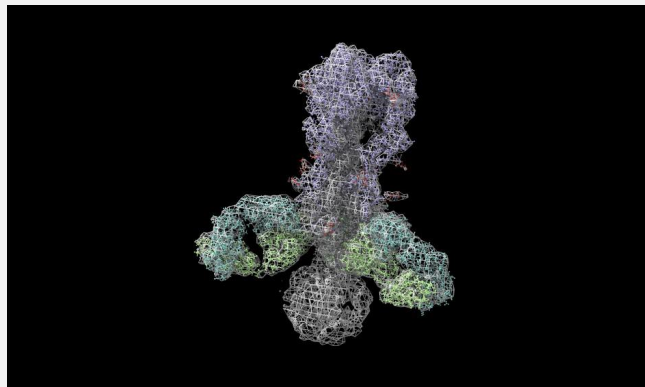
- Displays graphically the fraction of atoms that fit within the map at different map contour levels
- Also displays the recommended contour level for this map, as provided by the depositor at deposition.

# Data quality for EM structures

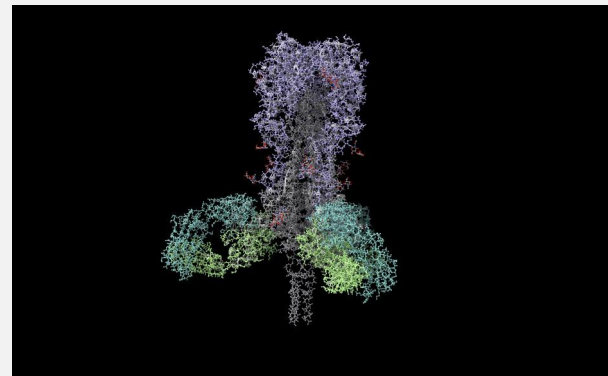
PDB id 6hjq



Contour level 0.2



Contour level 1



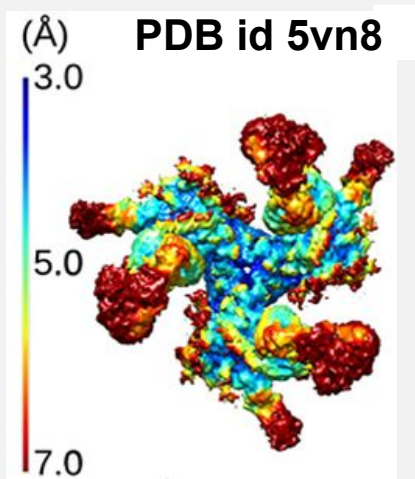
Contour level 2

Comparison of the map model fit at these different contour levels gives an indication of how changes in the contour level affect the apparent fit of atoms within the map.



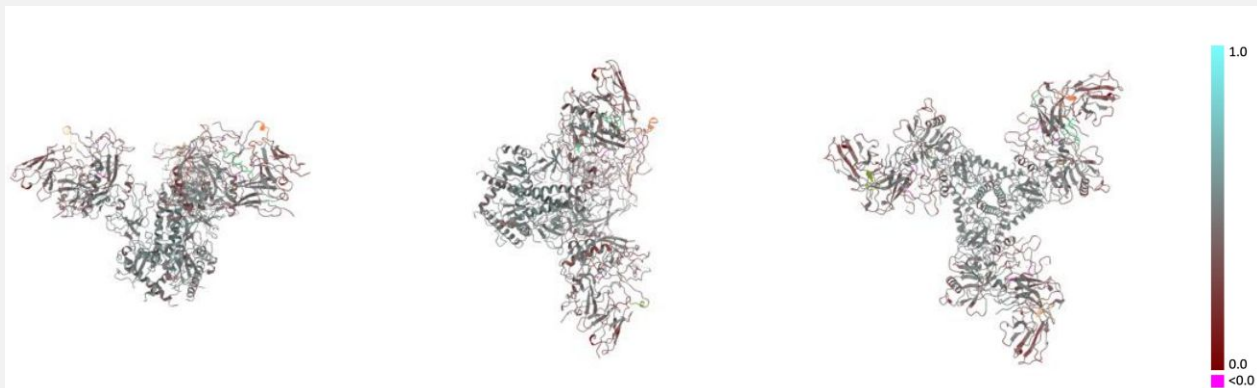
# Data quality for EM structures

It is important to understand the concept of **global** vs **local** resolution for cryo-EM structures



**Q-score**: calculated directly from map values around an atom's position

**PDB id 5vn8**

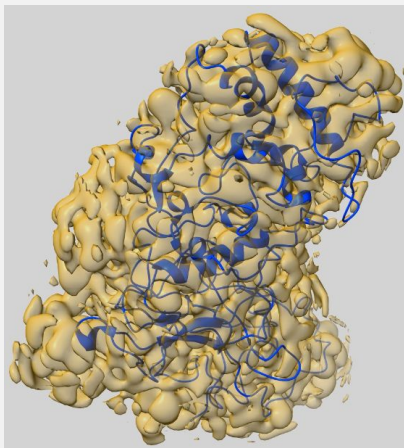


Vilas, J.L. et al. Local resolution estimates of cryoEM reconstructions. *Curr Opin Struct Biol.* 2020, 64:74-78

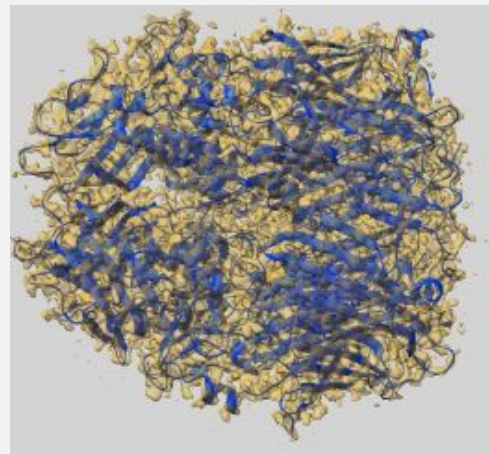
# Data quality for EM structures


Higher Q-score values reflect better resolvability


PDB id 8fd5



PDB id 9ikc



Model	Metric	Evaluation bar	Average value
8fd5	Q-score		0.006
		◆ : Percentile relative to EM structures of $\pm 1$ Å    ◇ : Percentile relative to all EM structures	

Model	Metric	Evaluation bar	Average value
9ikc	Q-score		0.763
		◆ : Percentile relative to EM structures of $\pm 1$ Å    ◇ : Percentile relative to all EM structures	

# PDBe-KB: Comparing structures in the PDB

8vih EgtB-IV from *Crocospaera subtropica*, an ergothioneine-biosynthetic type IV sulfoxide synthase in complex with N,N-dimethyl-histidine

Ireland KA, Davis KM

Structure (2024)

Source organism: *Crocospaera subtropica* ATCC 51142

Assembly composition: protein only structure

Bound ligands: FE AVI CL EDO NA

Assembly name: Sulfatase-modifying factor enzyme domain-containing protein (Preferred) [search this complex](#)

PDBe complex ID: PDB-CPX-247958 (Preferred) [search this ID](#)

**PDBe-KB: B1WTS6**

[3D Visualisation](#) [Download files](#)

Search in PDBe

Summary Structures Ligands Interactions Annotations Similarity Publications Feedback

### Sulfatase-modifying factor enzyme domain-containing protein

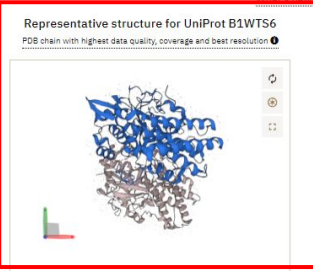
Organism: *Crocospaera subtropica* (strain ATCC 51142 / BH68) (*Cyanothece* sp. (strain ATCC 51142))

Uniprot: B1WTS6 [go to UniProt](#)

Biological function: Catalytic activity: undefined [go to UniProt](#)

Representative structure for UniProt B1WTS6

PDB chain with highest data quality, coverage and best resolution









PDB chain shown: 8vih B [go to PDBe](#)

UniProt residues 1 - 448

Coverage: 97%

3D view of superposed structures for region 1

Click on the icons below to view the relevant page:

 4 Structures <a href="#">Download</a> 3D view of superposed structures	 8 Ligands <a href="#">Download</a> 3D view of superposed ligands	 0 Interactions	 Annotations	 0 Similarity	 2 Publications
---	---	---	--	---	---

PDBe-KB selects the best representative structure based on the PDB chains with highest data quality, coverage and resolution

# PDBe-KB: Comparing structures in the PDB

Summary Structures Ligands Interactions Annotations Similarity Publications Feedback of

Structures and Domains Download View

Structures 3D view of superposed structures

Annotations

**Serine/threonine-protein kinase**

The visualisation below shows information on protein structures covering various regions of the sequence, domains (Pfam, CATH, SCOP and InterPro), known secondary structural elements content and predicted intrinsic flexibility of the protein. It also shows all the theoretical structures available for this protein.

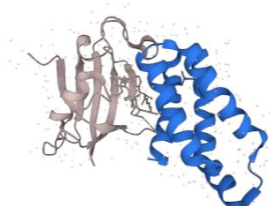
Gene:

Organism:

Synonyms:

Uniprot:

Biological function:



PDB Chain shown: 5gpg B go to PDBe of

Help: Click on any PDB and Other Structures segment (i.e. coloured box) on the ProtVista sequence feature viewer below to display a different structure. The visualisation is using UniProt numbering for residues, not PDB numbering.

Click on the icons below to

50 Structures

Download

3D view of superposed structures

PDB ID	Resolution (Å)	Download
5gpg	1.67 Å	Download
4dri	1.45 Å	Download
8xi9	1.85 Å	Download
6m4u	2.2 Å	Download
5wbh	1.75 Å	Download

All the PDB entries with this protein

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

# PDBe-KB: Comparing structures in the PDB

Structure clusters P62942 (1 - 108)

Select Segment 1 (1 - 108)

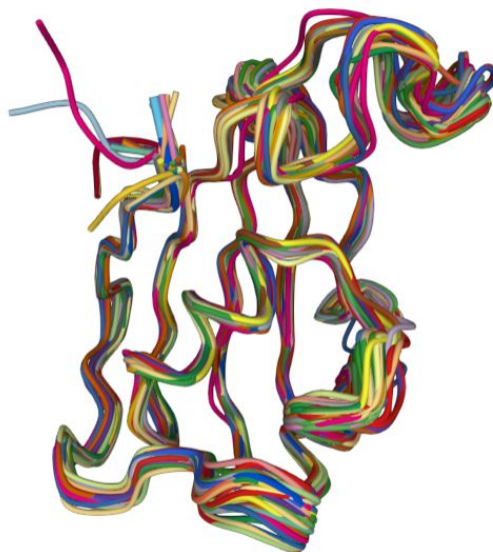
Cluster 1 104 chains ✓ All ✕ None

Search PDB ID Enter PDB ID. ✕

- 1fkh chain A
- 1d7h chain B
- 8er7 chain E
- 1b14 chain A
- 2ppo chain A
- 1o6o chain A
- 2ppn chain A
- 7u0t chain B
- 6y10 chain A
- 1bkf chain A

Cluster 2 3 chains ✓ All ✕ None

- 1f1t chain A (Representative)
- 1f1r chain A
- 1f1s chain A



Structure Tools

Components

- Chain
- AlphaFold Structure

AlphaFold Structure Opacity

pLDDT less than

Opacity

AlphaFold PAE

AlphaFold Superposition

Entry	RMSD (Å)
8x6p chain A	0.37
1f1t chain A	1.31

Measurements

+ Add

Export Models

Structural clusters of the protein

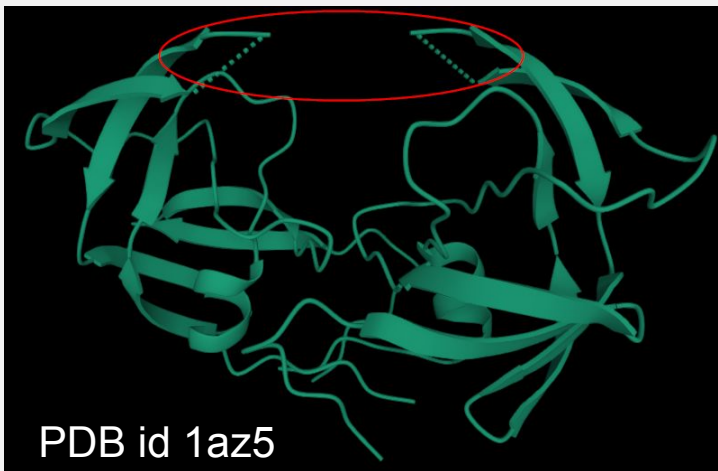
Download Download Download

3D view of superposed structures

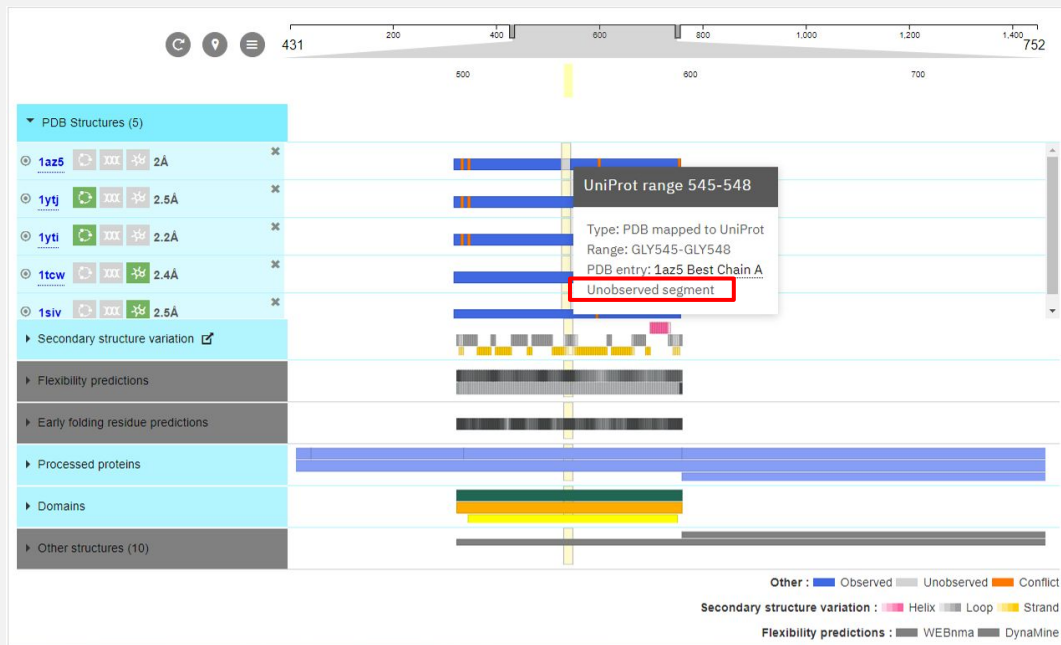
3D view of superposed ligands

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

# PDBe-KB: Comparing structures in the PDB



1az5





## Ligands and Environments

## Directly Interacting Ligands (23)

## All Ligands (28)

This section, by default, shows ligands observed directly bound to the protein of interest, if such ligands are available. Click on the checkbox below to see every ligand from all PDB entries (some may not directly interact with the protein). If there are no directly interacting ligands, all ligands will be shown by default. Click on the images to see the related PDB entries. For ligand binding residues, see the sequence viewer at the bottom.

Filter the ligands: 

Filter by molecule name, code or PDB id.

Show all ligands from PDB entries containing

Help: Checking this box will show ligands which may not

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

IHP

Found in 10 PDB entries of



RAP

drug-like

Found in 9 PDB entries of



ARD

Found in 2 PDB entries of



XYU

Found in 2 PDB entries of

Show all

all the ligands observed in the same PDB entries as this protein

Click on the icons below to view the relevant page:

50  
Structures

28  
Ligands

15  
Interactions

Annotations

113  
Similarity

722  
Publications

Download

Download

Download

3D view of superposed structures

3D view of superposed ligands

Download

Structures with Ligands

View

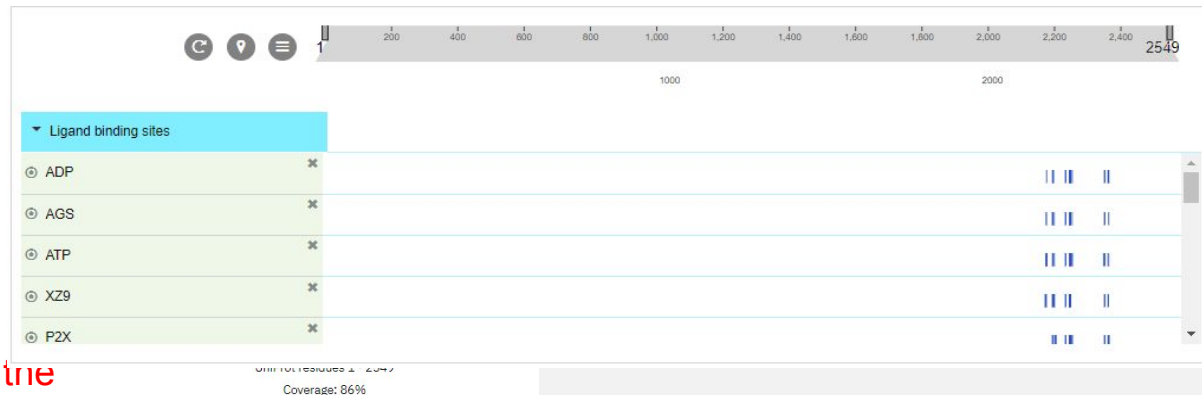
3D view of superposed ligands

Ligand Annotations

## structures in the PDB

## Ligand-binding Residues

The visualisation is using UniProt numbering for residues, not PDB numbering.



3D view of superposed structures for region 1

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



# PDBe-KB: Comparing structures in the PDB

Structure clusters P42345 (1 - 2114)

Select Segment 1 (1 - 2549)

- Cluster 1 1 chain
  - 6bcu A (Representative)
- Cluster 2 2 chains
  - 5h64 D (Representative)
  - 5h64 A No Ligand found!
- Cluster 3 21 chains
  - 7pe8 A (Representative)
  - 7tzo A No Ligand found!
  - 4t6 C
  - 4j5v C
  - 6s02 A No Ligand found!
  - 5fic F No Ligand found!
  - 5wby A No Ligand found!
  - 7owg A No Ligand found!
  - 5wbu A No Ligand found!
  - 5ccs A No Ligand found!

Download

Download

Download

3D view of superposed structures

3D view of superposed ligands

AlphaFold Structure Opacity

pLDDT less than 70

Opacity 0.2

AlphaFold PAE

Expected position error (Ångströms)

AlphaFold Superposition

Entry	RMSD (Å)
6bcu chain A	2.10
7pe8 chain A	6.56
5h64 chain D	12.24

Measurements

+ Add

all the ligands superposed  
on this protein

# PDBe-KB: Comparing structures in the PDB

## Serine/threonine-protein kinase mTOR (P42345)

### Macromolecular Interactions

#### Interacting Partners (15)

This section shows macromolecules observed together with the protein of interest in PDB entries. Click on the images to see the related PDB entries. The interaction partner is colored blue.

Filter the molecules:

Filter by molecule name, code or PDB id.

In the gallery below, the structure of the protein of interest (i.e. the protein this page focuses on) is colored grey, while the interaction partner (i.e. the macromolecule the gallery item focuses on) is colored blue.

<p><b>P42345 (self)</b> ↓ Found in 10 PDB entries of</p> <p><a href="#">View protein page</a></p>	<p><b>Q9BVC4</b> ↓ Found in 26 PDB entries of</p> <p><a href="#">View protein page</a></p>
<p><b>Q8N122</b> ↓ Found in 10 PDB entries of</p> <p><a href="#">View protein page</a></p>	<p><b>Q8TB45</b> ↓ Found in 7 PDB entries of</p> <p><a href="#">View protein page</a></p>

[Show all](#)

Download [What's new?](#)

[Structures of Complexes](#) niProt P42345 and best resolution

[Interface Annotations](#)

### Interface Residues

The visualisation is using UniProt numbering for residues, not PDB numbering.



<p>50 Structures</p> <p><a href="#">Download</a></p> <p>3D view of superposed structures</p>	<p>28 Ligands</p> <p><a href="#">Download</a></p> <p>3D view of superposed ligands</p>	<p>15 Interactions</p> <p><a href="#">Download</a></p>	<p>Annotations</p>	<p>113 Similarity</p>	<p>722 Publications</p>
--	--	--	--------------------	-----------------------	-------------------------

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

all the macromolecular interaction partners observed in the same PDB entries as this protein

# PDBe-KB: Comparing structures in the PDB

Summary Structures Ligands Interactions Annotations Similarity Publications Feedback

## Serine/threonine-protein kinase mTOR

Gene: MTOR [Enzyme: EC 2.7.11.1](#) [Disease](#)

Organism: *Homo sapiens (Human)*

Synonyms: FRAP, FRAP1, FRAP2, RAFT1, RAPT1

Uniprot: P42345 [go to UniProt](#)

Biological function: Serine/threonine protein kinase which is a central regulator of cellular metabolism, growth and survival in response to hormones, growth factors, nutrients, energy and stress signals ([PubMed:12087098](#), [PubMed:12150925](#), [PubMed:12150926](#), [PubMed:12231510](#), [PubMed:12718876](#), [PubMed:14651849](#), [PubMed:15268862](#), [PubMed:15467718](#), [+ \[show more\]](#) [go to UniProt](#))







Representative structure for UniProt P42345  
PDB chain with highest data quality, coverage and best resolution



PDB chain shown: 7pe7 A [go to PDBe](#)  
UniProt residues 1 - 2549  
Coverage: 86%

3D view of superposed structures for region 1

Click on the icons below to view the relevant page:

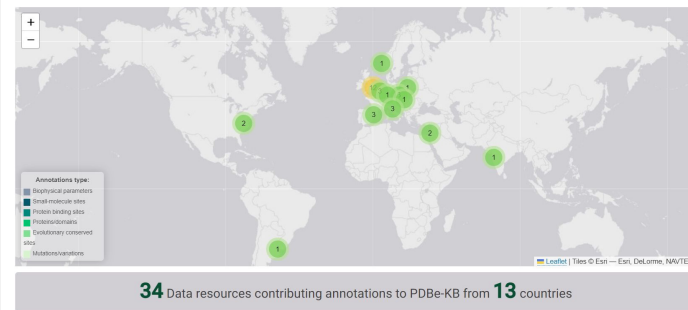
 50 Structures	 28 Ligands	 15 Interactions	 <b>Annotations</b>	 113 Similarity	 722 Publications
<a href="#">Download</a>	<a href="#">Download</a>	<a href="#">Download</a>			

3D view of superposed structures

3D view of superposed ligands

additional annotations derived from PDBe-KB partners ([pdbekb.org/partners](http://pdbekb.org/partners))

<https://www.ebi.ac.uk/pdbe/pdbe-kb/partners>



# PDBe-KB: Comparing structures in the PDB

Summary Structures Ligands Interactions Annotations Similarity Publications Feedback

[What's new?](#)

## Serine/threonine-protein kinase mTOR

Gene: MTOR [Enzyme: EC 2.7.11.1](#) [Disease](#)

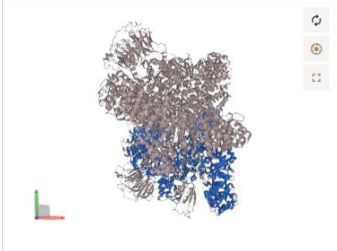
Organism: *Homo sapiens (Human)*

Synonyms: FRAP, FRAP1, FRAP2, RAFT1, RAPT1

Uniprot: P42345 [go to UniProt](#)

Biological function: Serine/threonine protein kinase which is a central regulator of cellular metabolism, growth and survival in response to hormones, growth factors, nutrients, energy and stress signals (PubMed:12087098 , PubMed:12150925 , PubMed:12150926 , PubMed:12231510 , PubMed:12718876 , PubMed:14651849 , PubMed:15268862 , PubMed:15467718 , [+ \[show more\]](#) [go to UniProt](#))

Representative structure for UniProt P42345  
PDB chain with highest data quality, coverage and best resolution



PDB chain shown: 7pe7 A [go to PDBe](#)  
UniProt residues 1 - 2549  
Coverage: 86%

3D view of superposed structures for region 1

Click on the icons below to view the relevant page:

 50 Structures	 28 Ligands	 15 Interactions	 Annotations	 113 Similarity	 722 Publications
<a href="#">Download</a>	<a href="#">Download</a>	<a href="#">Download</a>			
<a href="#">3D view of superposed structures</a>	<a href="#">3D view of superposed ligands</a>				

proteins that have associated PDBs with sequence identities of 90%+ or belong to the same UniRef90 cluster

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

# PDBe-KB: Comparing structures in the PDB

Summary Structures Ligands Interactions Annotations Similarity Publications Feedback ↗

What's new?

## Serine/threonine-protein kinase mTOR

Gene: **mTOR** [Enzyme: EC 2.7.11.1 ↗](#) [Disease ↗](#)

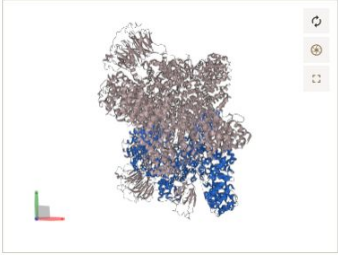
Organism: *Homo sapiens (Human)*

Synonyms: FRAP, FRAP1, FRAP2, RAFT1, RAPT1

Uniprot: P42345 [go to UniProt ↗](#)

Biological function: Serine/threonine protein kinase which is a central regulator of cellular metabolism, growth and survival in response to hormones, growth factors, nutrients, energy and stress signals ([PubMed:12087098 ↗](#), [PubMed:12150925 ↗](#), [PubMed:12150926 ↗](#), [PubMed:12231510 ↗](#), [PubMed:12718876 ↗](#), [PubMed:14651849 ↗](#), [PubMed:15268862 ↗](#), [PubMed:15467718 ↗](#), [+ \[show more\] go to UniProt ↗](#))







Representative structure for UniProt P42345  
PDB chain with highest data quality, coverage and best resolution ⓘ



PDB chain shown: 7pe7 A [go to PDBe ↗](#)  
UniProt residues 1 - 2549  
Coverage: 86%

3D view of superposed structures for region 1

Click on the icons below to view the relevant page:

 50 Structures	 28 Ligands	 15 Interactions	 Annotations	 113 Similarity	 722 Publications
<a href="#">Download</a>	<a href="#">Download</a>	<a href="#">Download</a>			
<a href="#">3D view of superposed structures</a>	<a href="#">3D view of superposed ligands</a>				

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

all the primary PDB publications, publications from UniProt and other reviews

# 3D-Beacons: Finding computational models

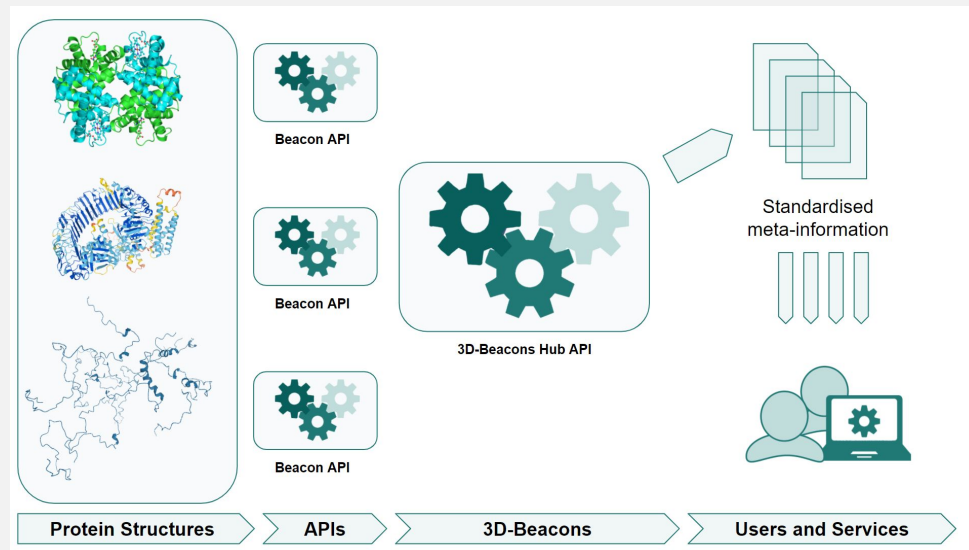
- The 3D-Beacons Network<sup>[5]</sup> provides standardised access to both experimentally determined and predicted protein structures
- It is an open collaboration between many model providers:
  - PDBe
  - SWISS-MODEL
  - AlphaFold DB
  - Genome3D
  - SASBDB
  - AlphaFill
  - ModelArchive
  - Protein Ensemble Database

<https://3d-beacons.org>



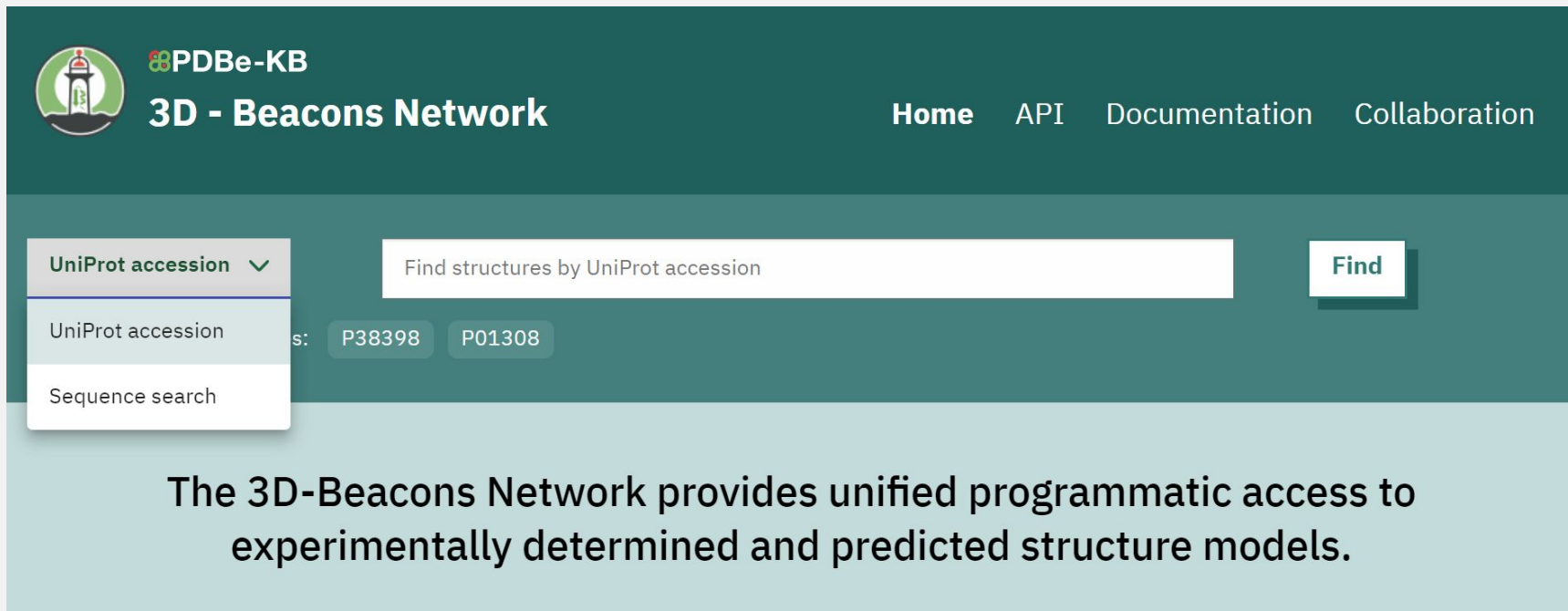
# 3D-Beacons: Finding computational models


- You can access data and their descriptions in a standardised format
  - Where are the models (i.e. URLs)?
  - What is the overall quality of a model?
  - What is the context (i.e. metadata)?
    - Species?
    - Gene?
    - Sequence identifier?



# 3D-Beacons: Finding computational models

Search by UniProt or sequence



 PDBe-KB  
**3D - Beacons Network**

Home API Documentation Collaboration

UniProt accession ▾  
UniProt accession  
Sequence search

Find structures by UniProt accession

Find

UniProt accession: P38398 P01308

The 3D-Beacons Network provides unified programmatic access to experimentally determined and predicted structure models.

<https://3d-beacons.org>



# 3D-Beacons: Finding computational models

P38398 (BRCA1\_HUMAN) - 64 Structures available

## Information

Protein Breast cancer type 1 susceptibility protein [Go to UniProt](#)

Gene BRCA1

Source organism *Homo sapiens*

Biological function E3 ubiquitin-protein ligase that specifically mediates the formation of polyubiquitin chains and plays a central role in DNA repair by fa responses to DNA ... [Show more](#)



31

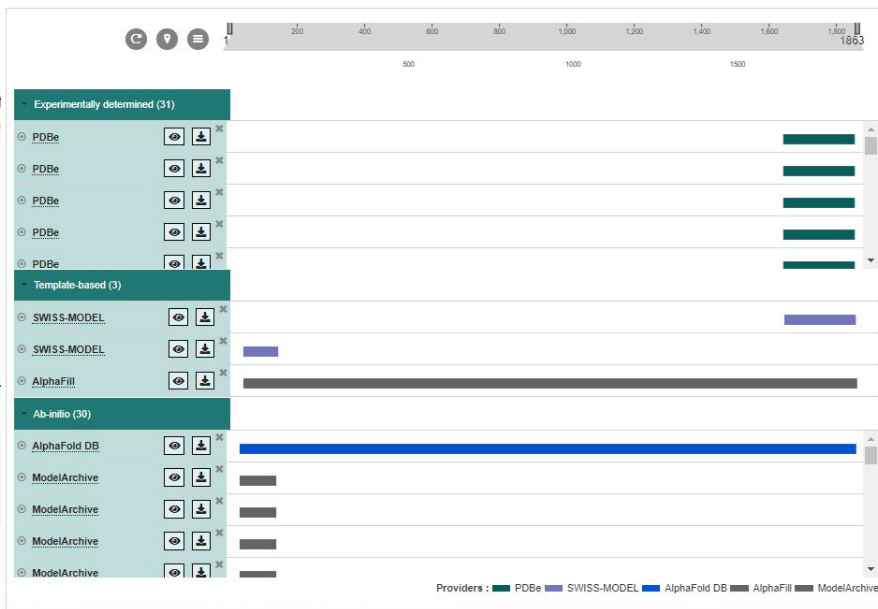
Experimentally Determined Structures



0

Conformational Ensembles

Ter



<https://3d-beacons.org>

# Starting model from a MD simulation: MDposit

MDposit

HOME BROWSE SEARCH REST API META-ANALYSIS HELP CONTACT

- open platform designed to provide web access to atomistic molecular dynamics (MD) simulations
- Possible to search simulations by title, authors or group names
- Download trajectory files and extract structures where the protein is solvated, energy minimized and ions added & continue the MD.

<https://mdposit-dev.mddbr.eu/>

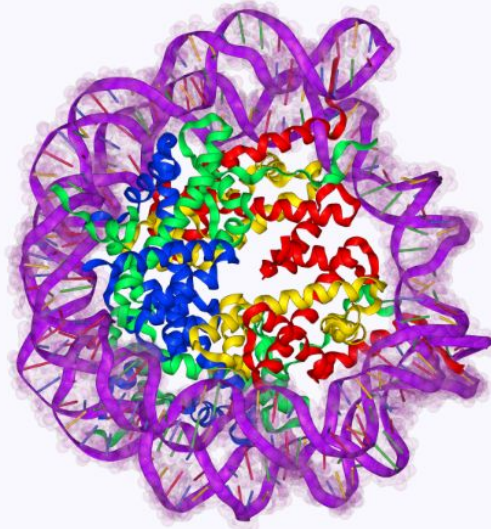
# Starting model from a MD simulation: MDDDB

## The MDDDB Project

Your gateway to comprehensive molecular dynamics data

At MDDDB we are harnessing decades of cutting-edge computational resources to build a **unified database** that compiles and organises all data generated by **molecular dynamics simulations**. By making these data **accessible to a wider scientific community**, we hope to drive **new research and discoveries** in fields such as biochemistry, pharmacology, and personalized medicine.

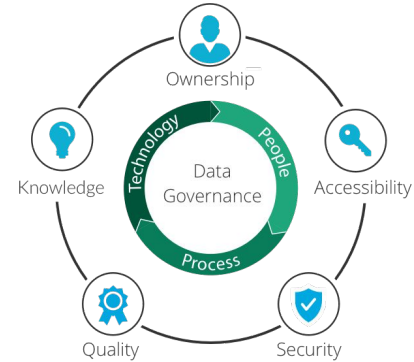
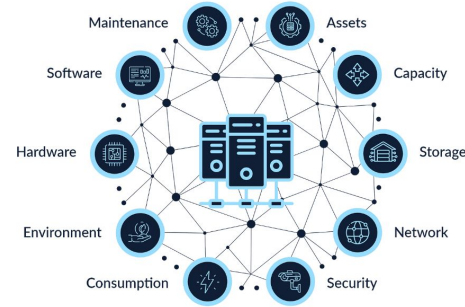
ABOUT



- unified database for MD simulations
- multidisciplinary consortium for validated MD simulations
- EU horizon project including 7 partners and lead by IRB Barcelona

<https://mddbr.eu/>

# MDDDB Objectives



Establish FAIR principles for MD data

Define and promote good practices for the generation and analysis of trajectories

Design the technical infrastructure for a general repository for simulation data

Design a sustainability and governance model for MDDDB



Funded by

the European Union

MDDDB received funding from EU's Horizon Europe programme under grant agreement 101094651

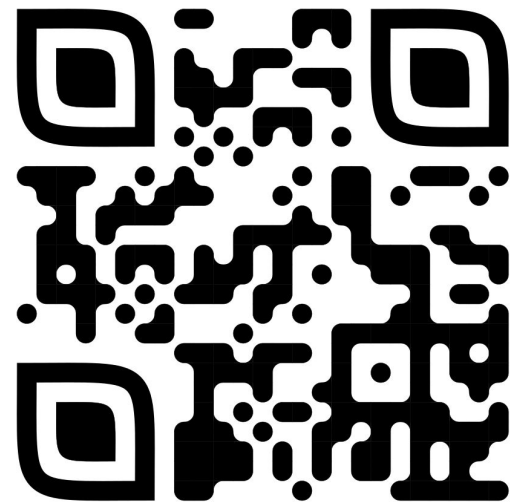
# Conclusions

- Selection of the correct starting model is extremely important for a successful MD simulation
- You should be careful of several factors, such as the protein conformation, missing loops, presence of ligands, etc.
- It is also important to be aware of the resolution and experimental technique used to derive the structure
- PDBe-KB & 3D Beacons are useful resources to compare protein structures across experimental and computational models.



# Feedback

<https://tinyurl.com/PDBe-bioexcel>



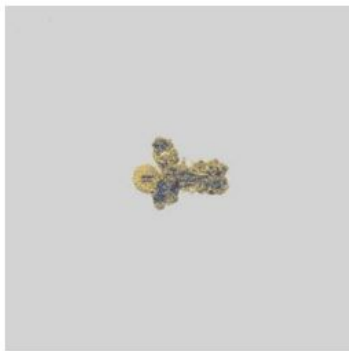


# Data quality for EM structures

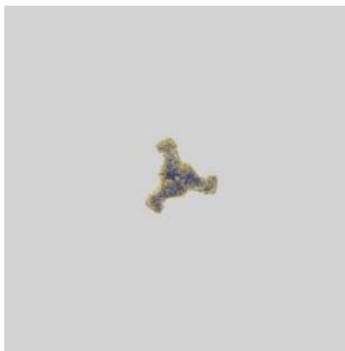
## 9.1 Map-model overlay [i](#)



X



Y



Z

- Displays graphically the fraction of atoms that fit within the map at different map contour levels
- Also displays the recommended contour level for this map, as provided by the depositor at deposition.