Protein Data Bank in Europe

PDBe.org

pdbhelp@ebi.ac.uk

proteindatabank

@PDBeurope

pdbeurope

pdbart

PDBe

Protein Data Bank in Europe

EMBL-EBI

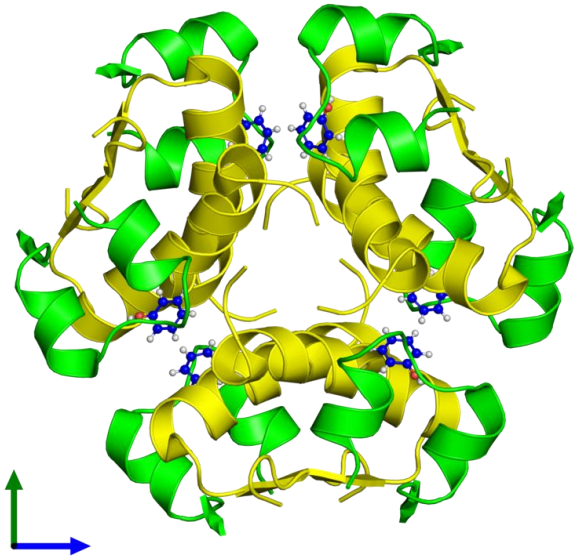
Using the PDBe

PDBe

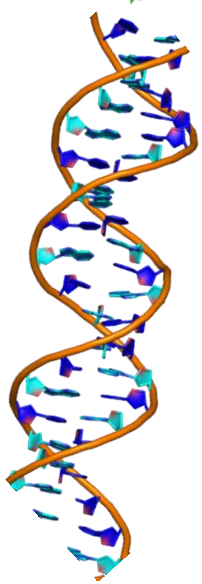
Protein Data Bank in Europe

EMBL-EBI

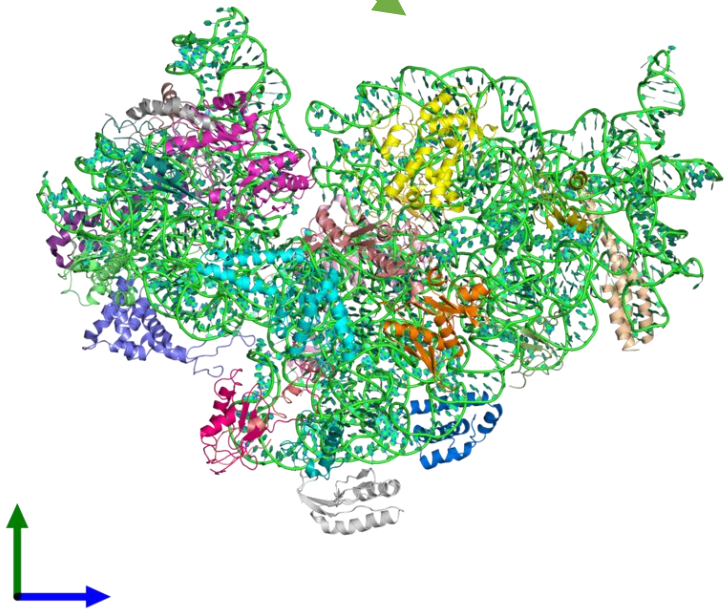
The Protein Data Bank (PDB) is an archive of ***experimentally determined*** 3-dimensional structures of biological macromolecules



Proteins



DNA/RNA



Complexes of Protein and NA

A “PDB code” refers to a structure

Described in a paper
(or maybe not p

- A unique 4 character code
- Identifies the data within the PDB archive
- Always starts with a number

- eg 2nuu, 4xyz, 2f48

Depos
in the
PDB

PDB code

Referenced in the paper

Methods of solving the structures

89% X-ray crystallography

9% NMR Spectroscopy

2% Electron microscopy

Releases in 2017…

91% X-ray crystallography

4% NMR Spectroscopy

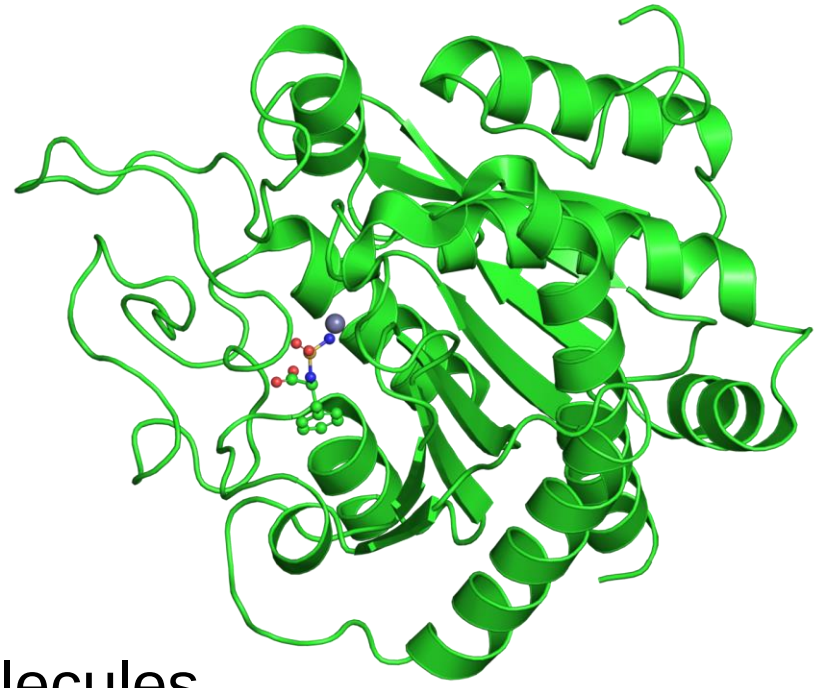
5% Electron microscopy

Not all 150 000 structures are unique molecules

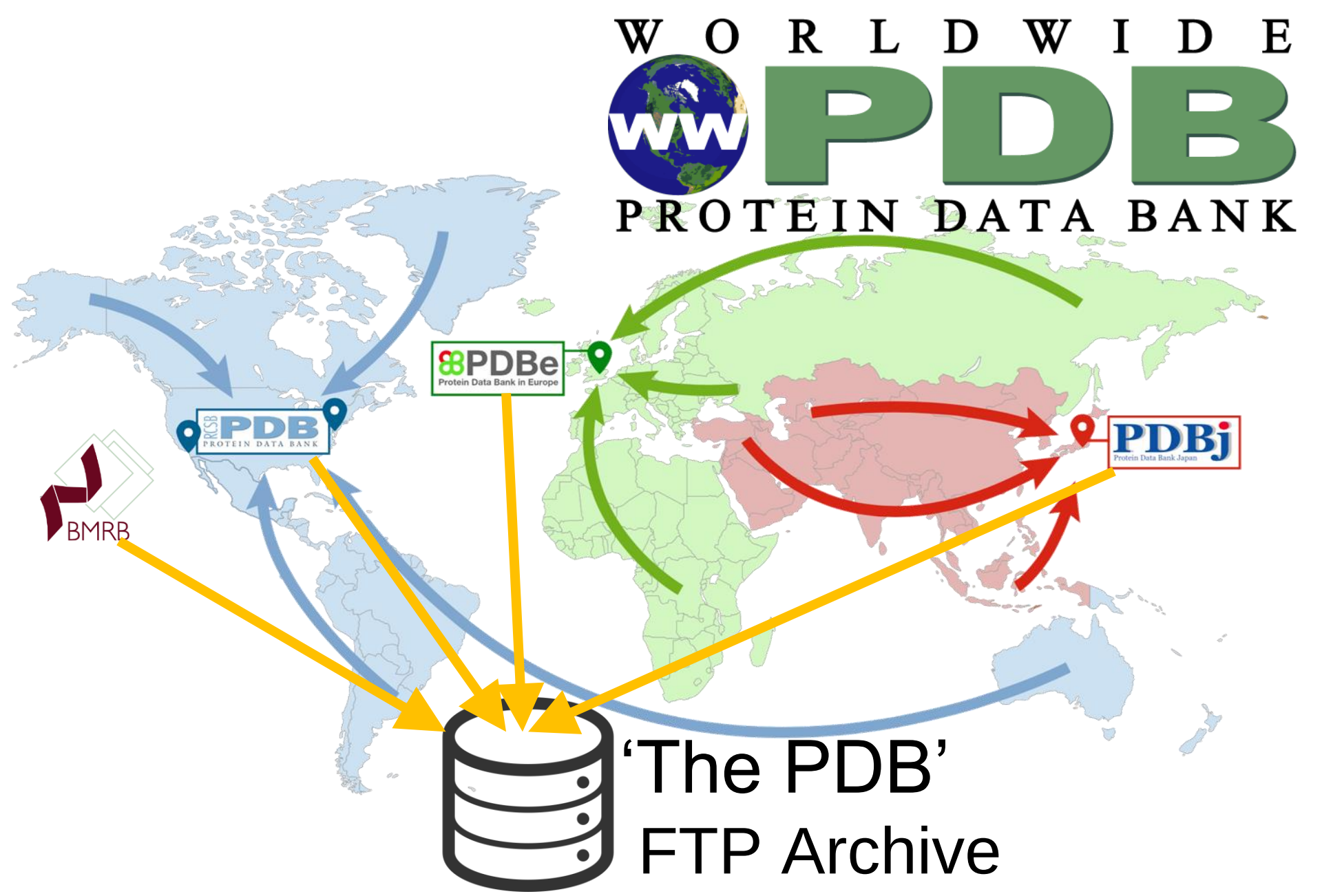
eg There are 39,775 structures of Human proteins
But only 9,640 different Human proteins

Why?

- Solved by different methods
- With different compounds bound
- By different people
- In different conditions



Carboxypeptidase A:
32 structures - 15 different small molecules


WORLDWIDE
PDB
PROTEIN DATA BANK
PDBe
Protein Data Bank in Europe
RCSB PDB
PROTEIN DATA BANK
PDBj
Protein Data Bank Japan
BMRB
'The PDB'
FTP Archive

WORLDWIDE PDB
PROTEIN DATA BANK

PDBe
Protein Data Bank in Europe

RCSB PDB
PROTEIN DATA BANK

PDBj
Protein Data Bank Japan

BMRB

Value added data

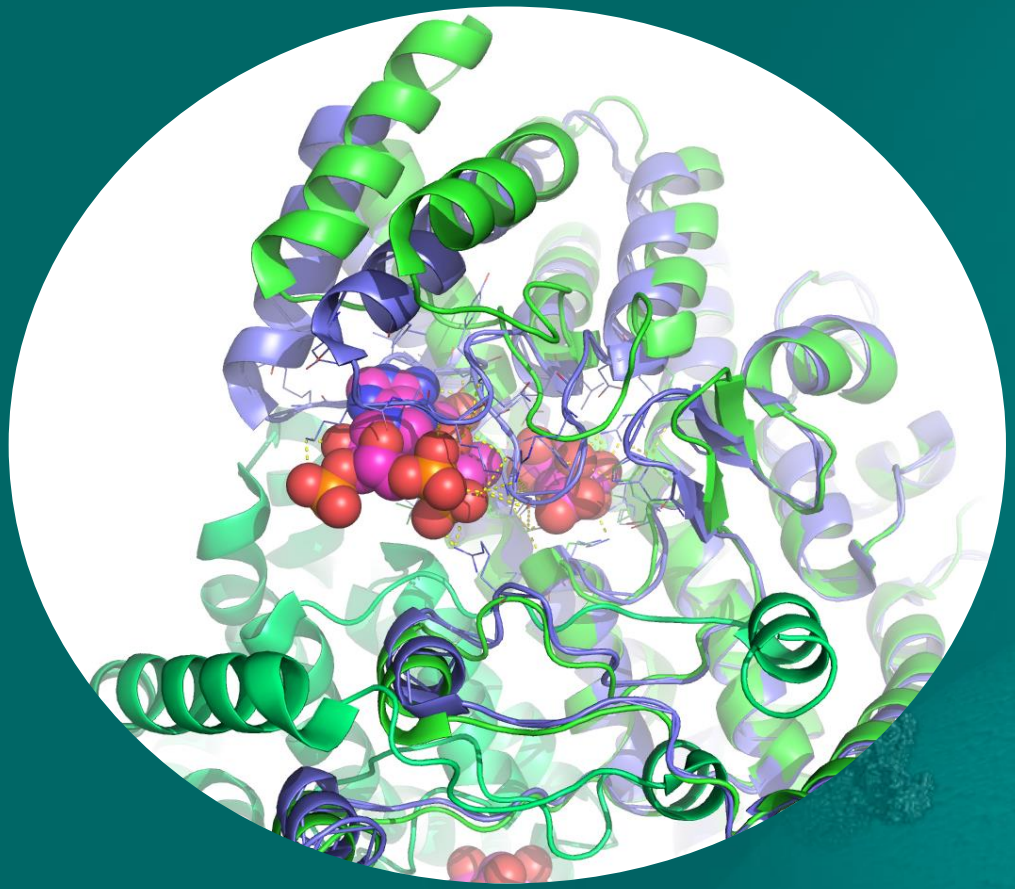
Value added data

Value added data

Value added data

‘The PDB’ FTP Archive

What can we learn from structures?



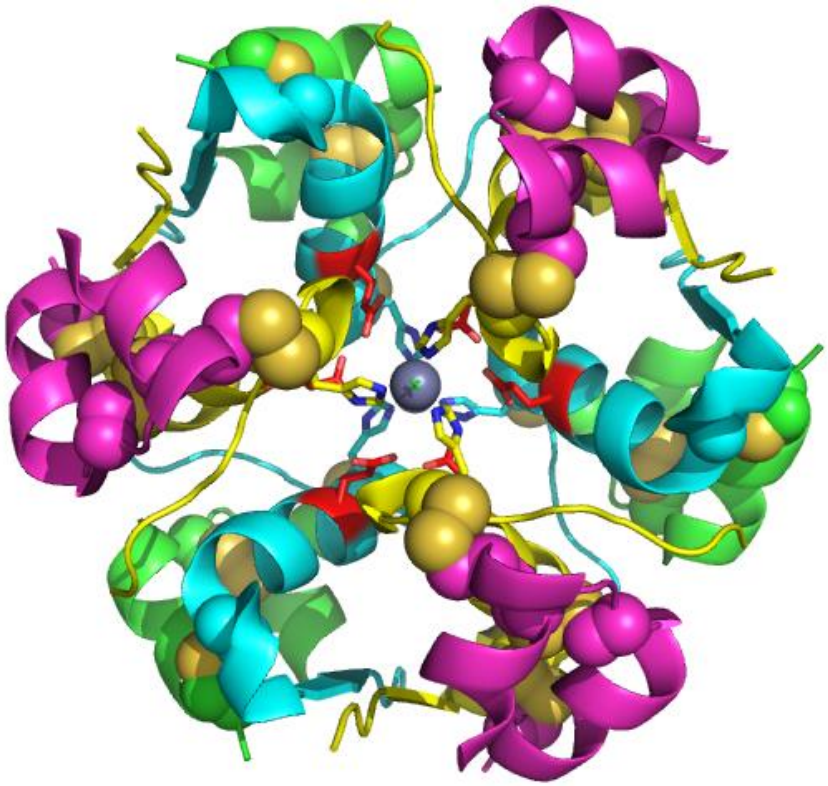
EMBL-EBI

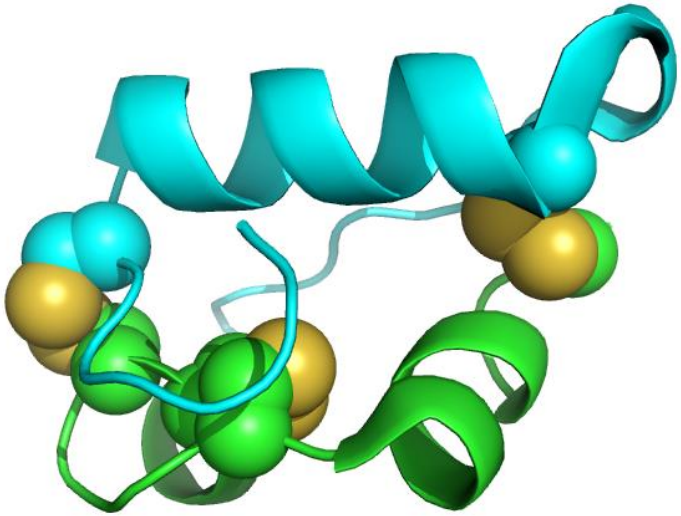
Why study protein structures?

- Sequence similarity and function not necessarily intrinsically linked
- Active site amino acids not always in close proximity in the protein sequence
- Modelling binding sites can help determine future drug targets
- Interfaces between macromolecules in structure can help identify likely biological assemblies.
- Is a proposed mutation likely to disrupt the structure significantly?

How multimeric state effects function, e.g. Insulin

- Inactive complex - 6 insulin units bind zinc
- pH shift from 6 to 7.4 - releases active monomer
  - de-protonation of Glutamate - breaks hydrogen bonds





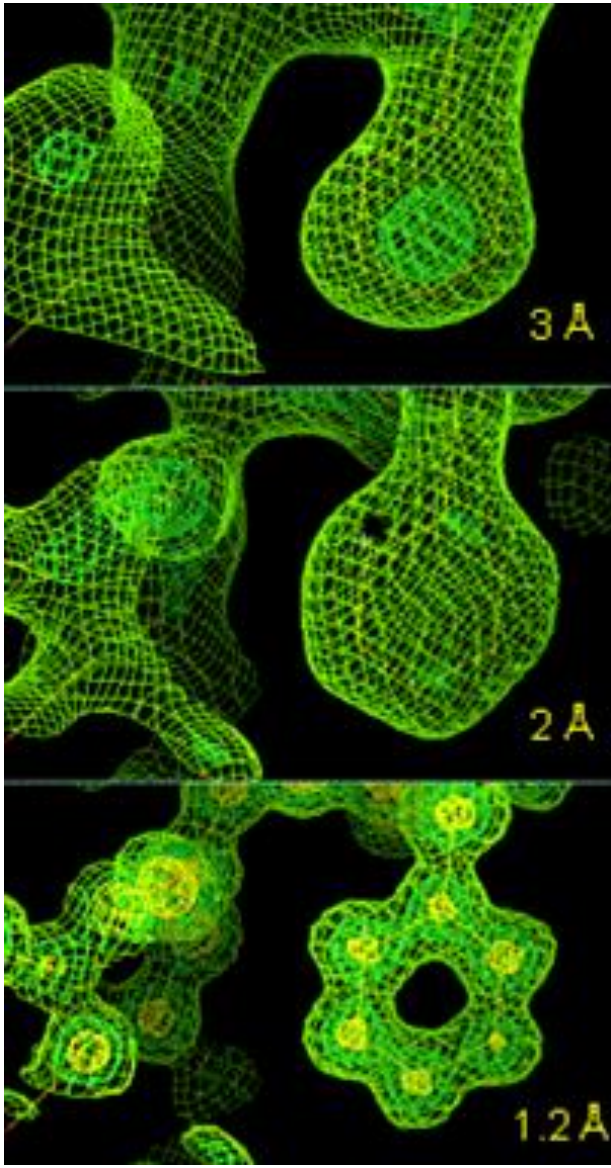
- Disulphide bonds identified
- Difficult to establish from sequence alone.

How good is a structure?

- Resolution
- Geometry
- Density maps are the real data
- Validation reports

Resolution

- Resolution indicates level of detail in the data
  - for structures determined by X-ray crystallography and electron microscopy
- Higher resolution (lower number) means more detail!
  - Low = > 3.0 A
  - Medium = 1.8 - 3.0 A
  - High = 1.0 - 1.8 A
  - Atomic = < 1.0 A
- But, not all parts of the structure are at the same resolution…

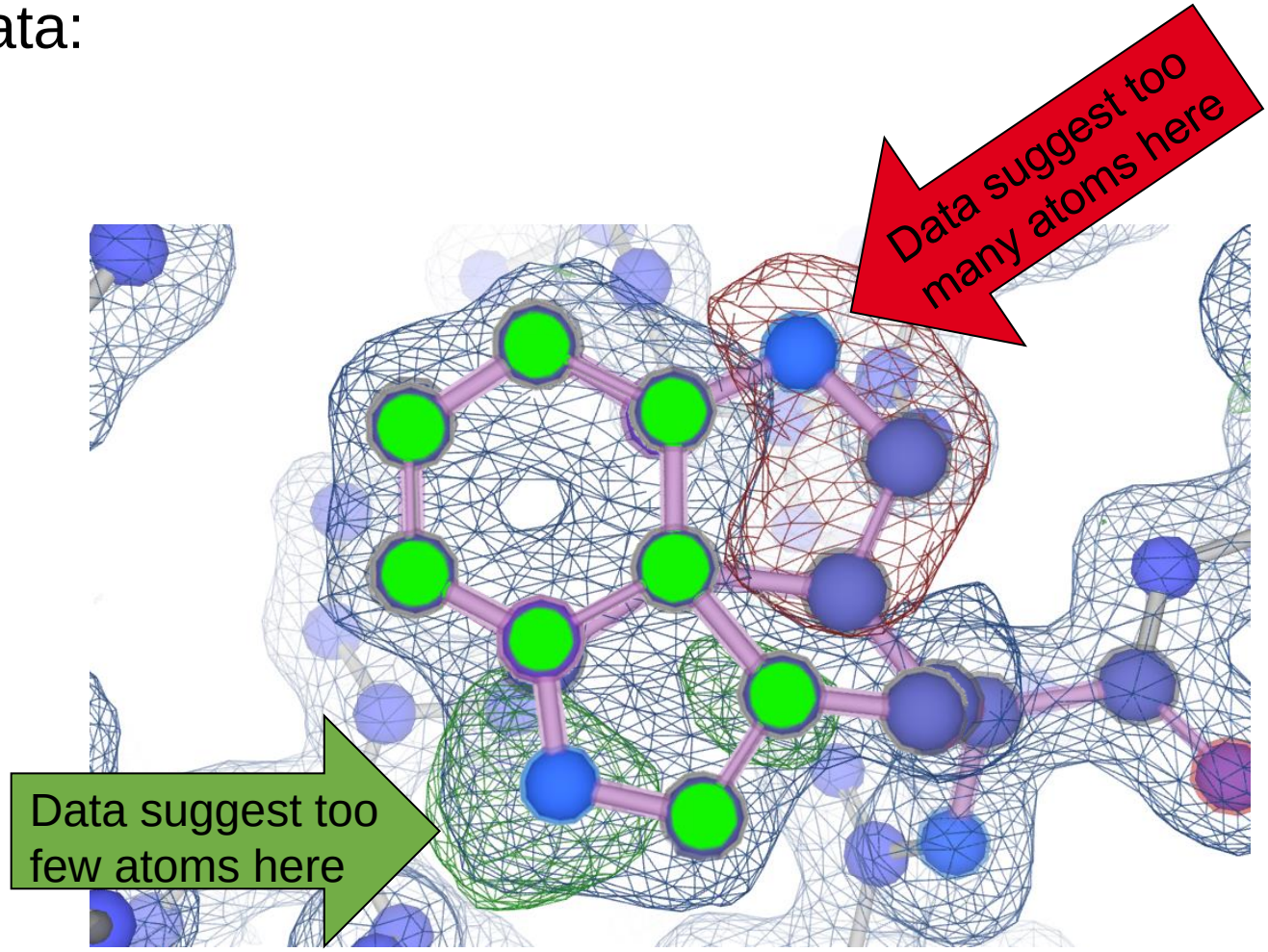


The PDB file is only an interpretation of the data

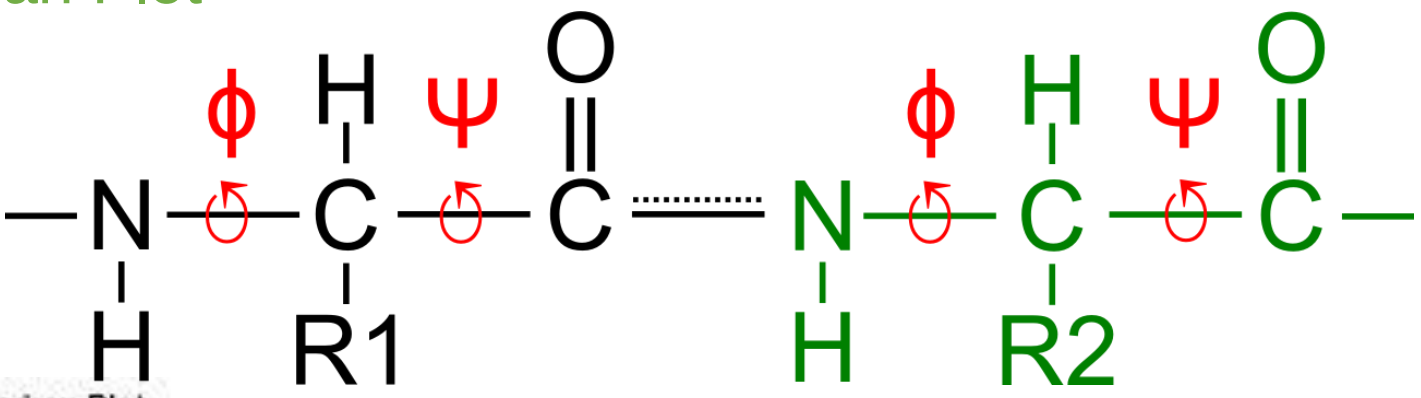
- Difference density helps identify where the model does not match the data:

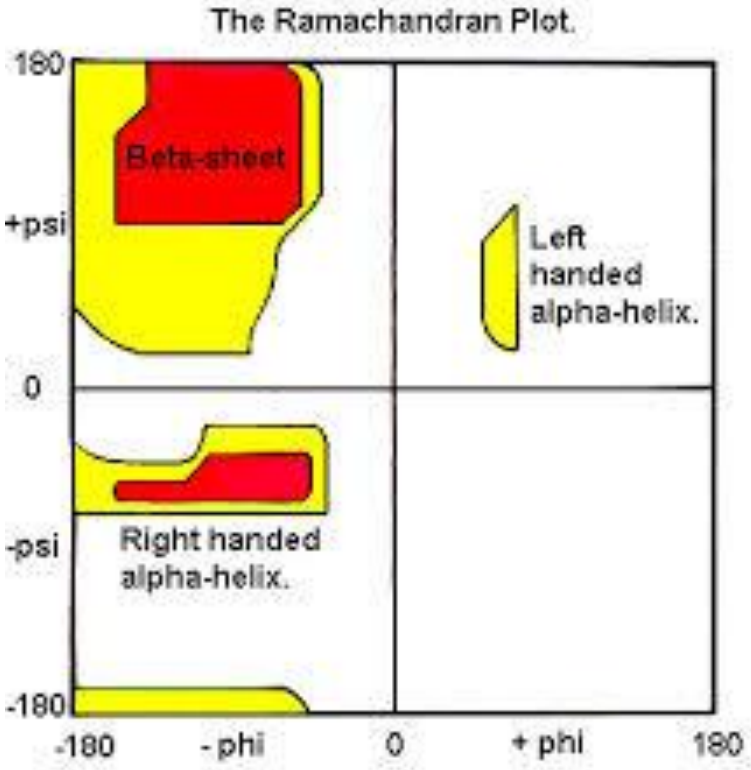
Areas where the model has too many atoms for the data

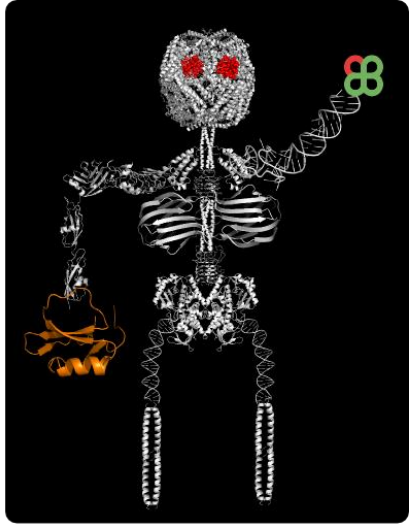
Areas where the model has too few atoms for the data

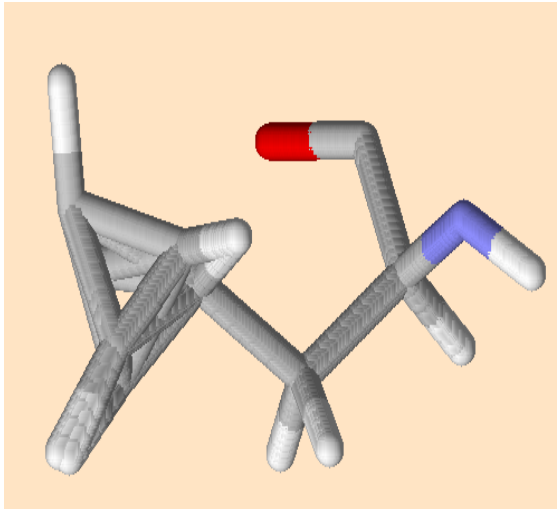


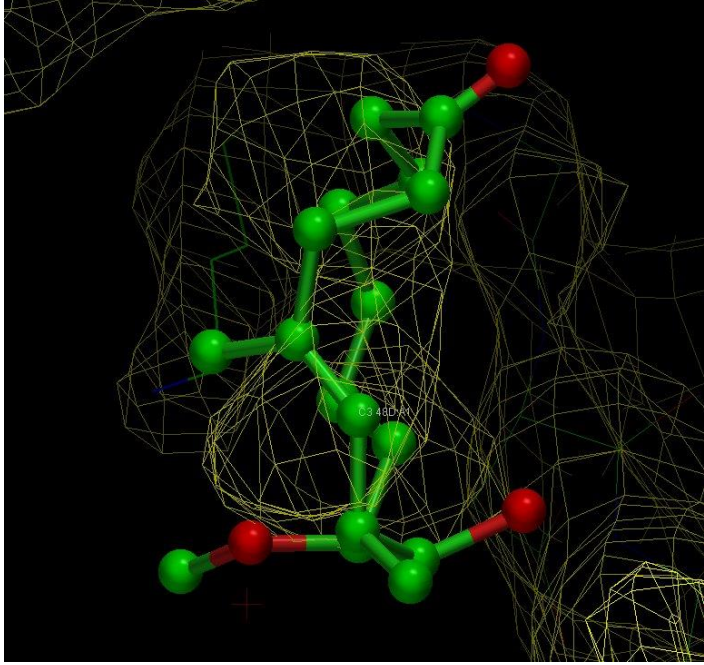
Ramachandran Plot

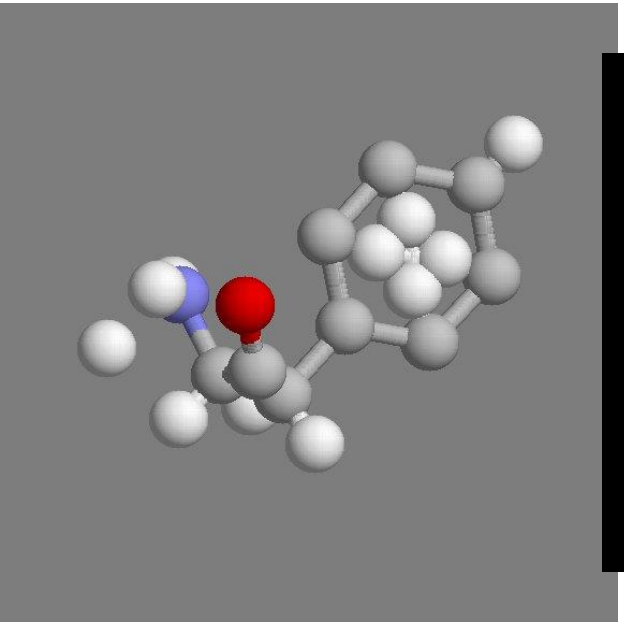


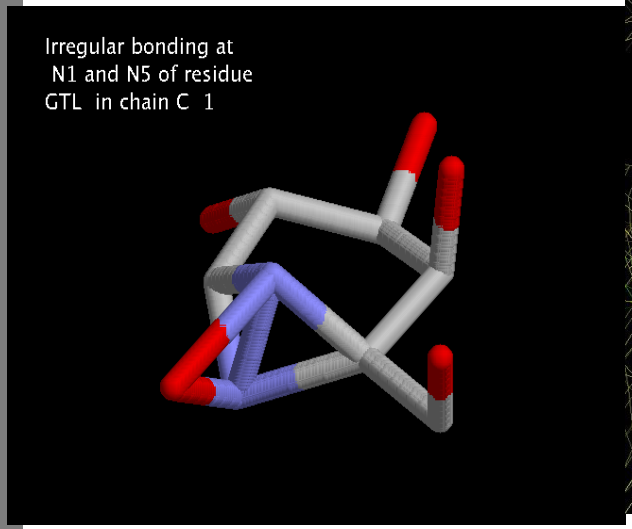


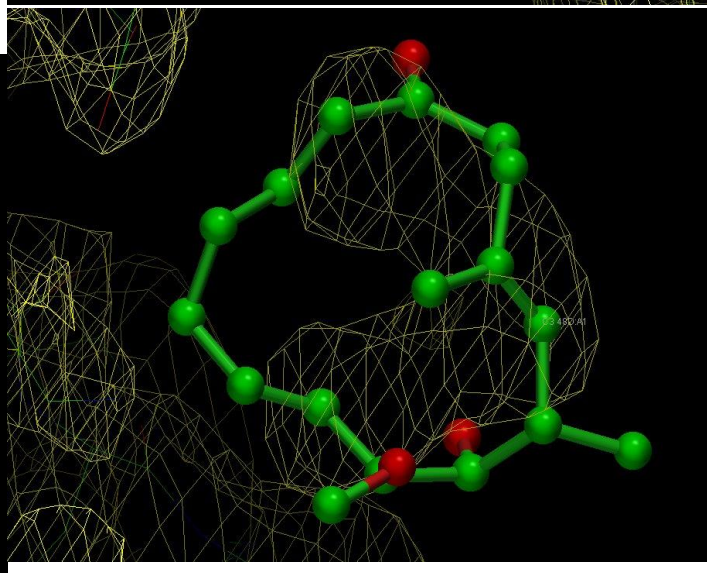






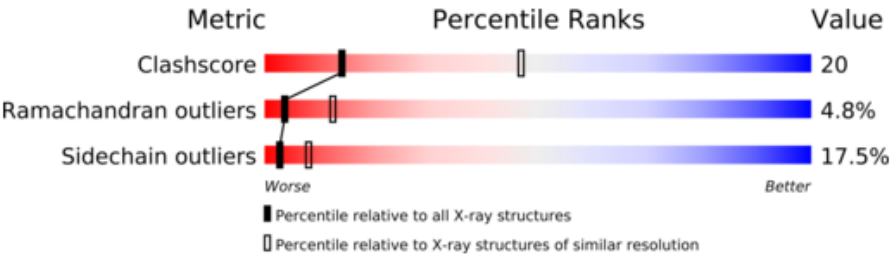


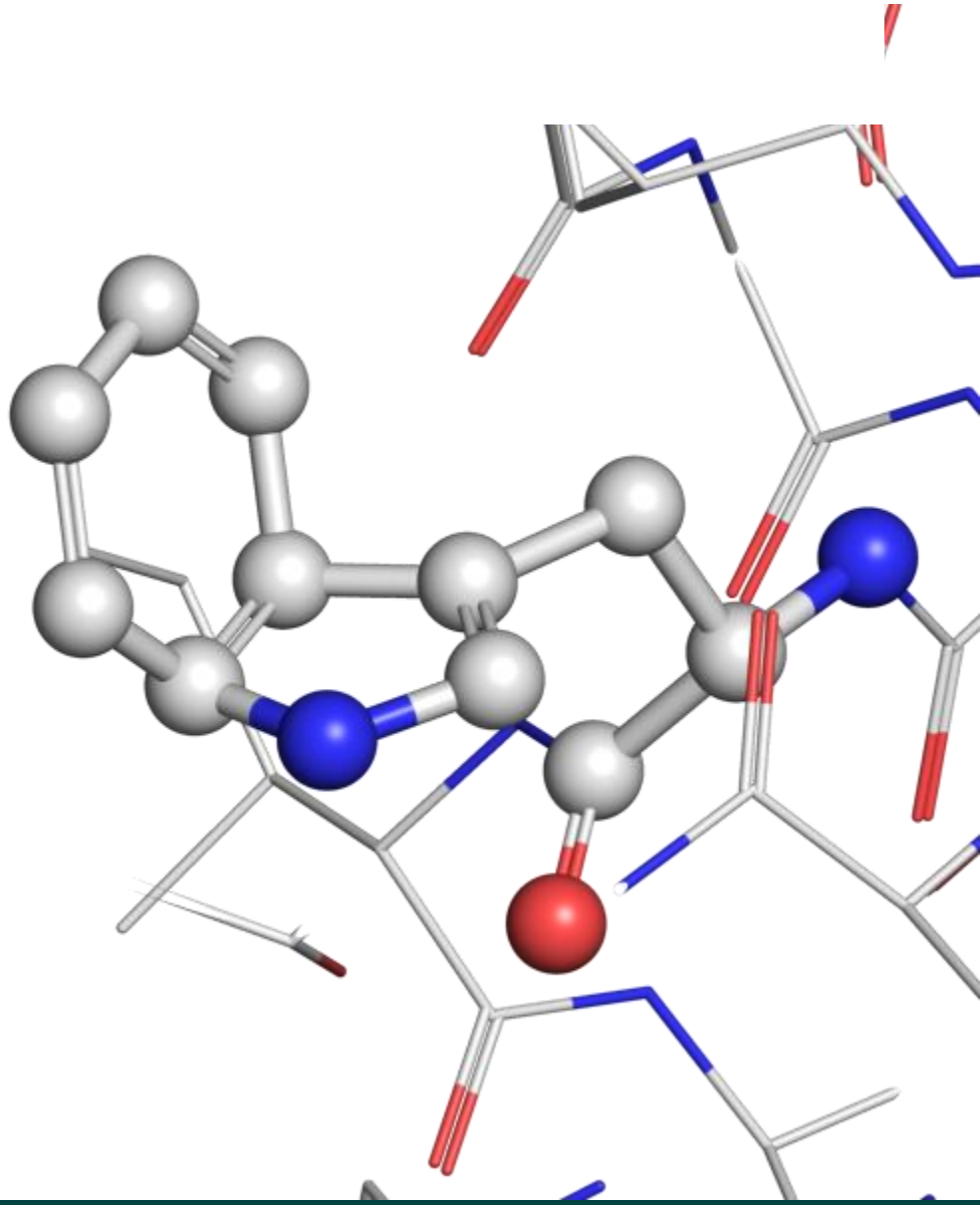

Irregular bonding at
N1 and N5 of residue
GTL in chain C 1



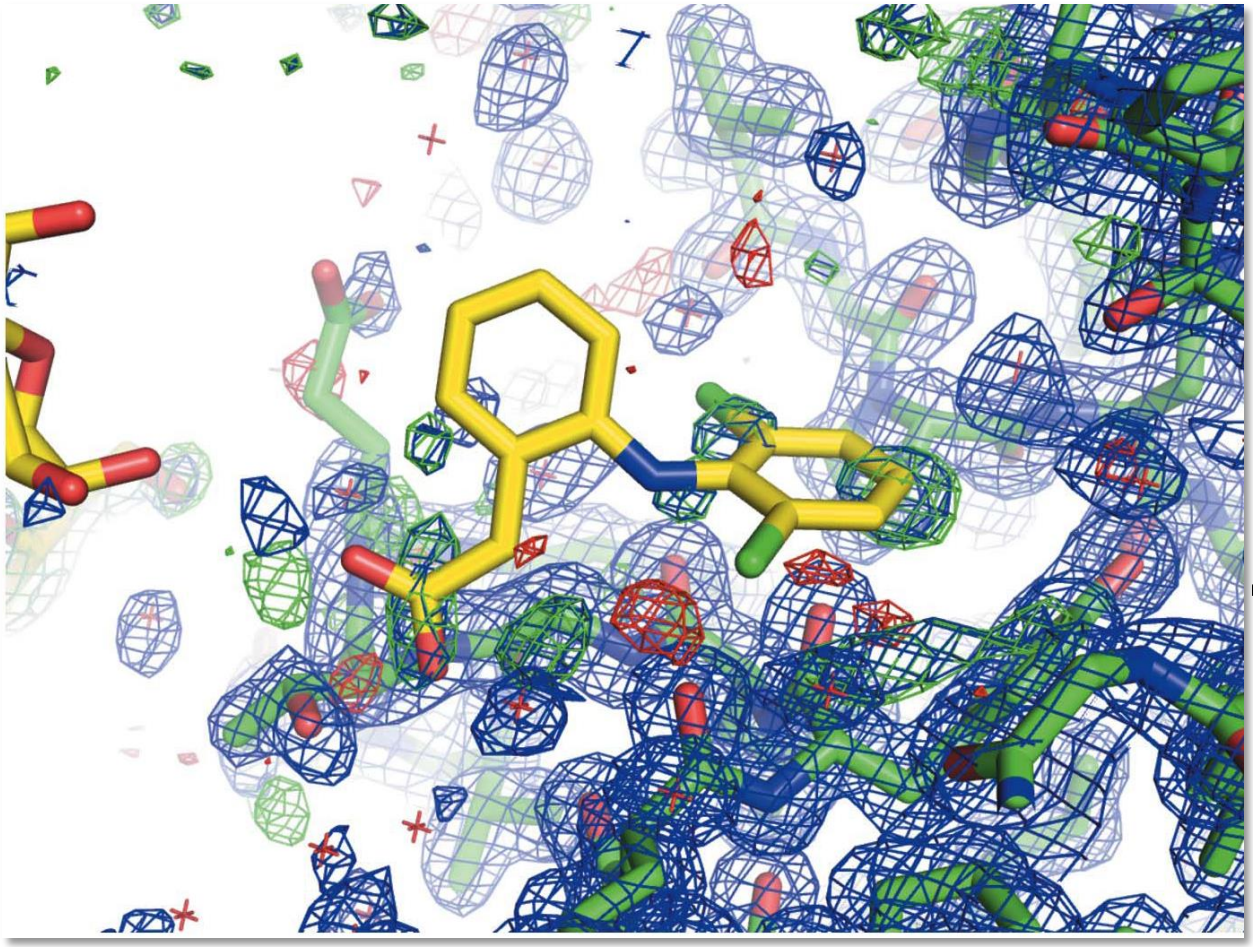
Interesting geometry

7GPB 2.9Å





Poor density for ligand

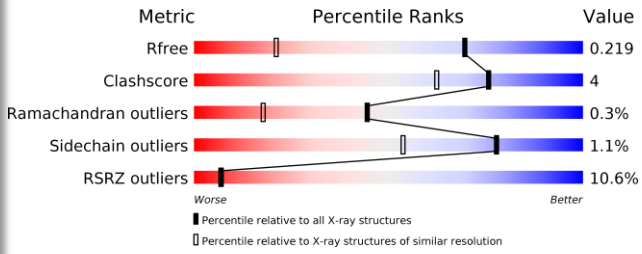


PDB: 3IB0 (1.4Å)

DIF A701

Ligand diclofenac not supported by electron density

https://www.ebi.ac.uk/pdbe/entry/pdb/3ib0/bound/DIF



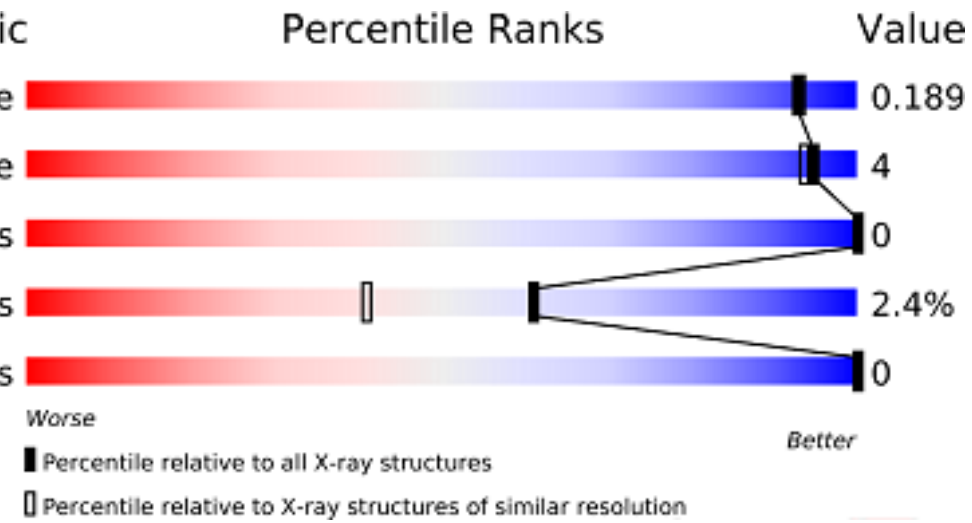
Not all structures are created equal!

How well does the model back-predict the data?

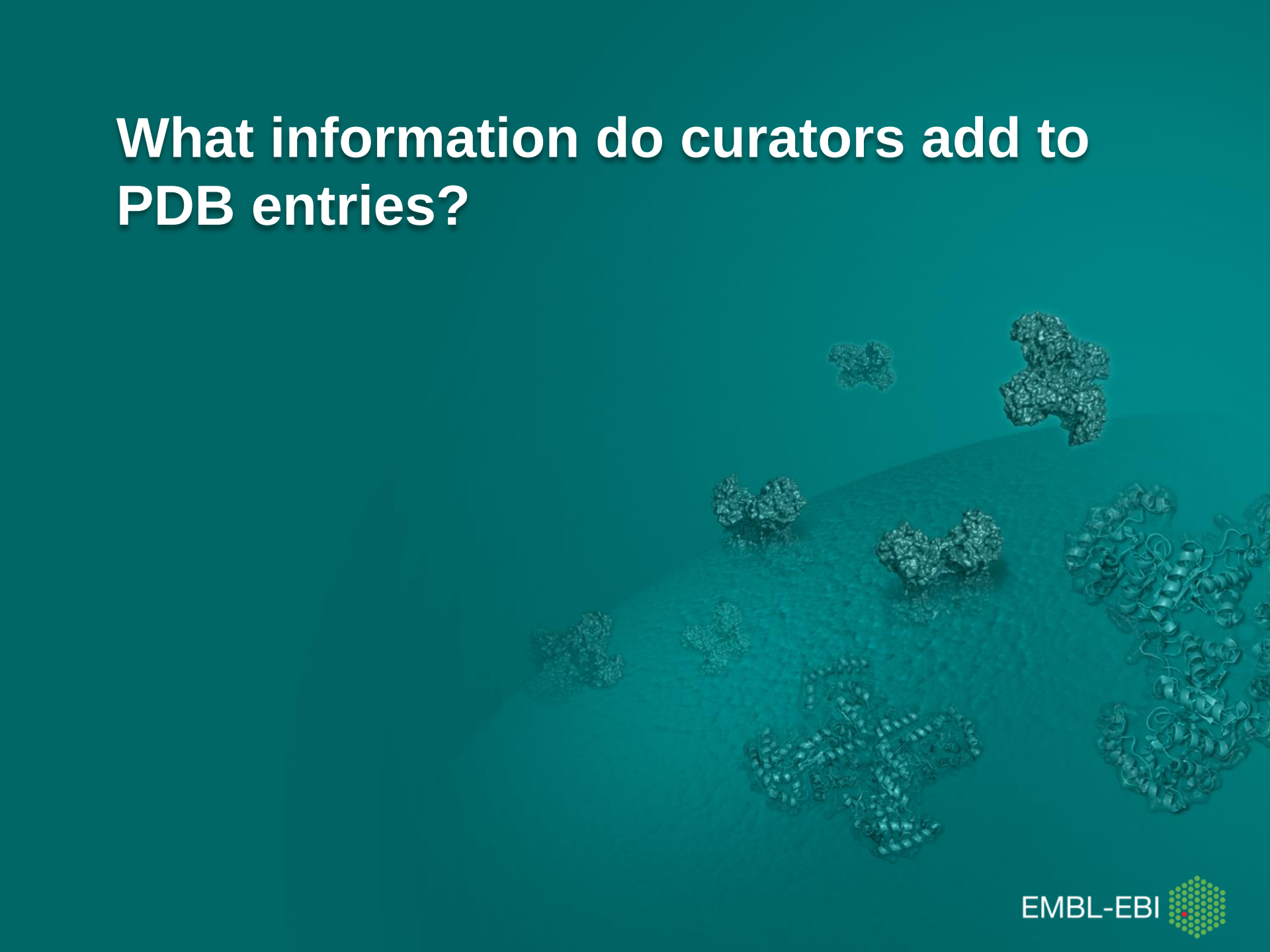
Atoms bumping into each other

Surprising bond angles

Atoms not in electron density



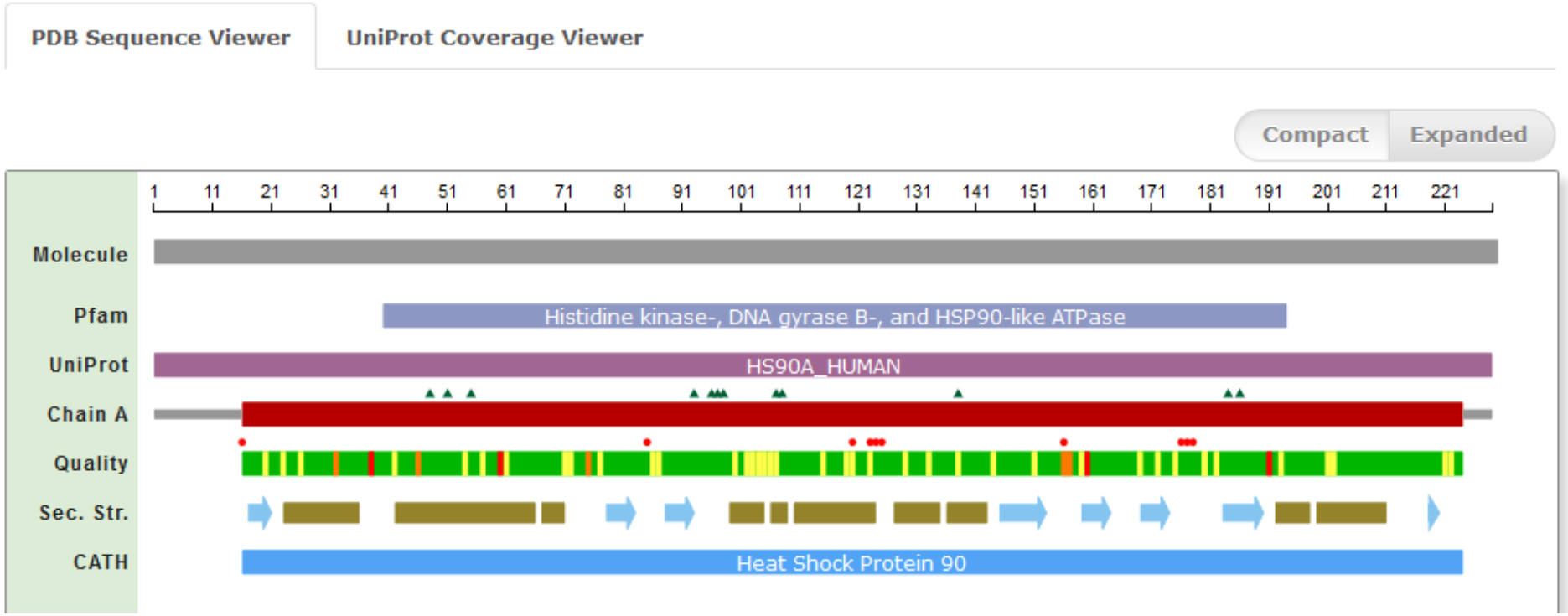
What information do curators add to PDB entries?



EMBL-EBI

Sequence cross references

- Protein sequences are cross referenced to UniProt

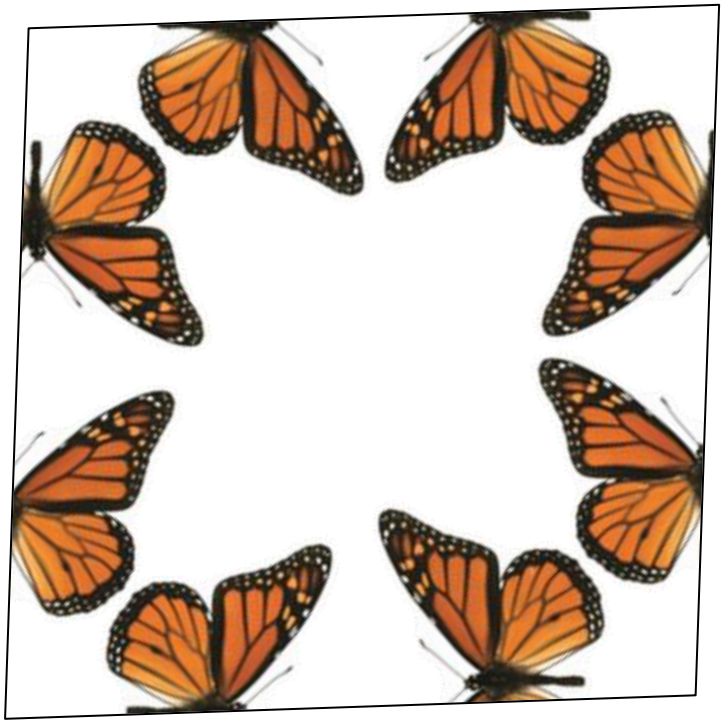


Assembly annotation

- How the molecule exists in ‘solution’

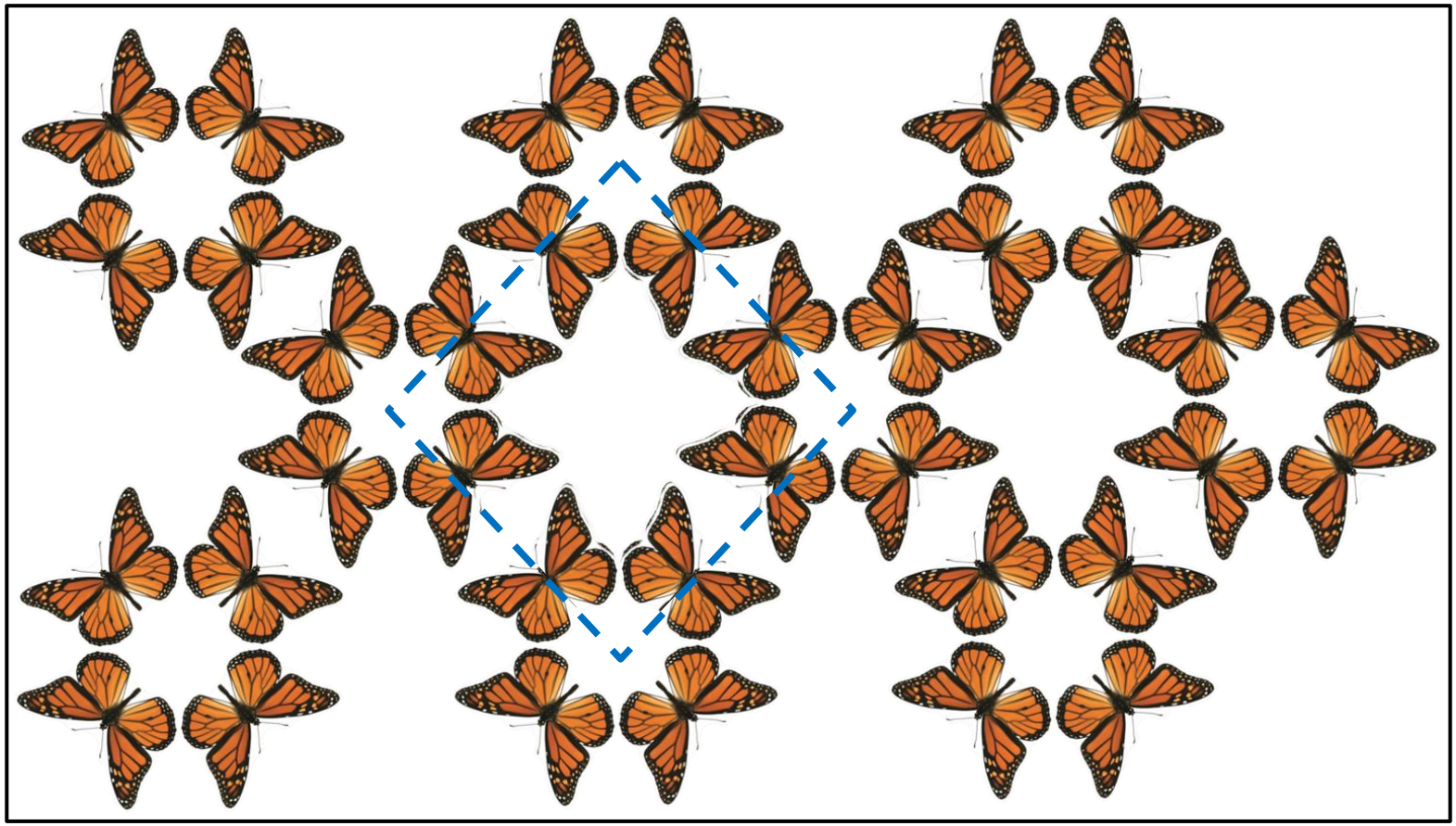
Crystal packing

- What we observe in a crystal unit cell



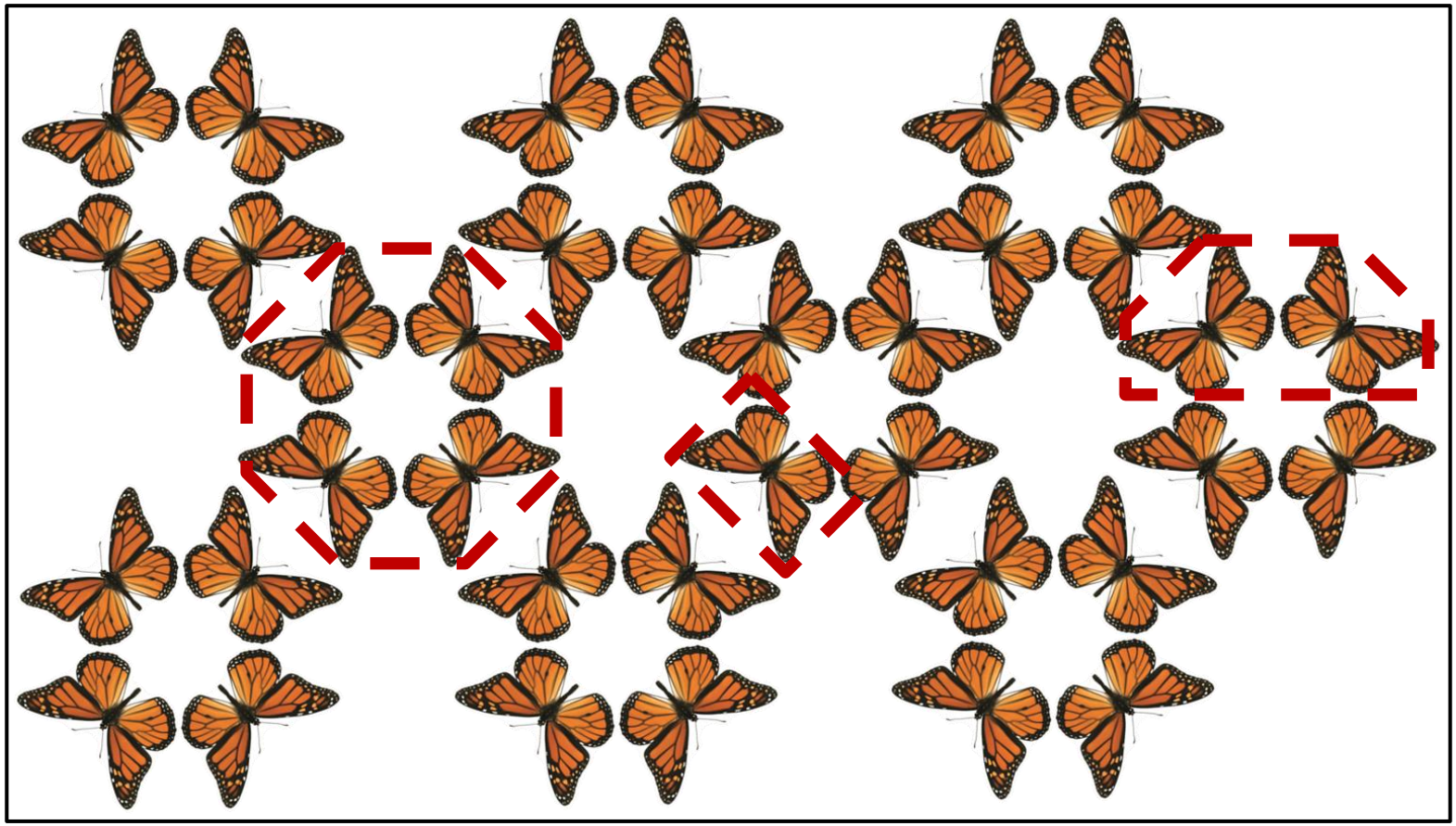
Crystal packing

- The rest of the crystal can be generated by symmetry



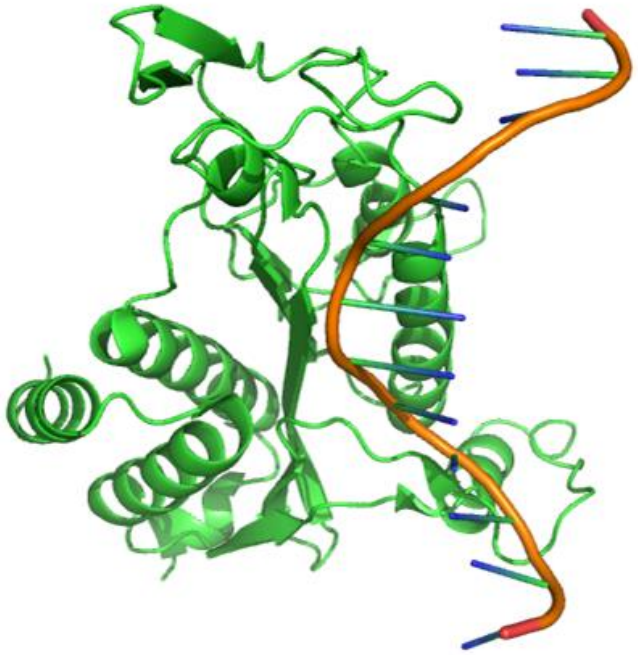
Crystal packing

- Can analyse which of the complexes are likely to be significant

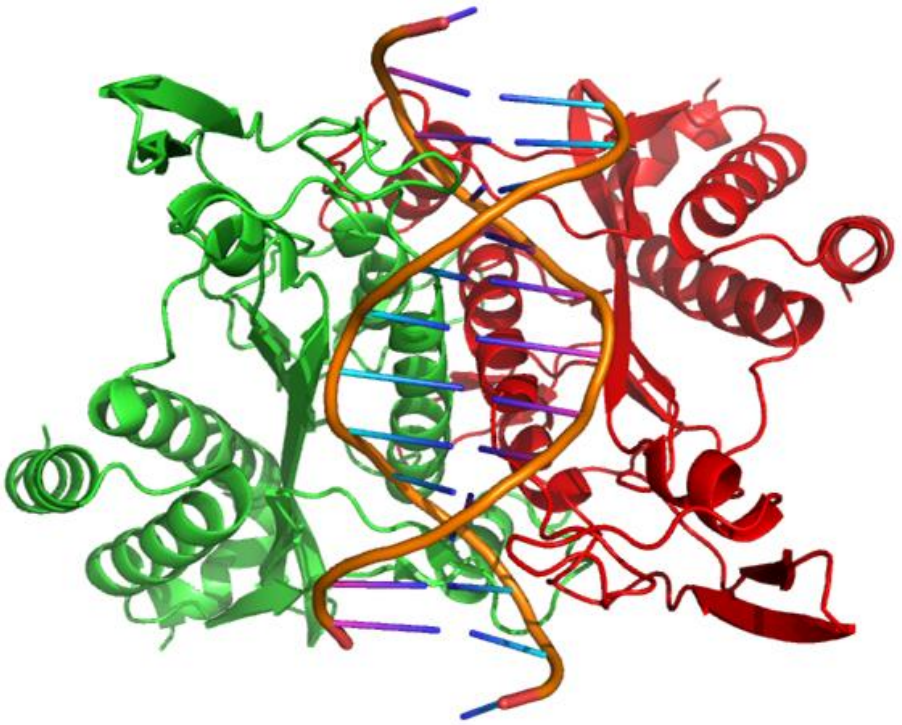


DNA-protein complex

A.S.U.



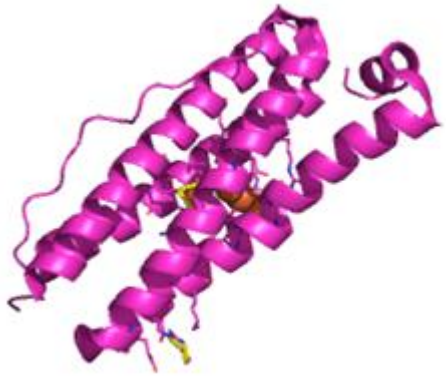
Assembly



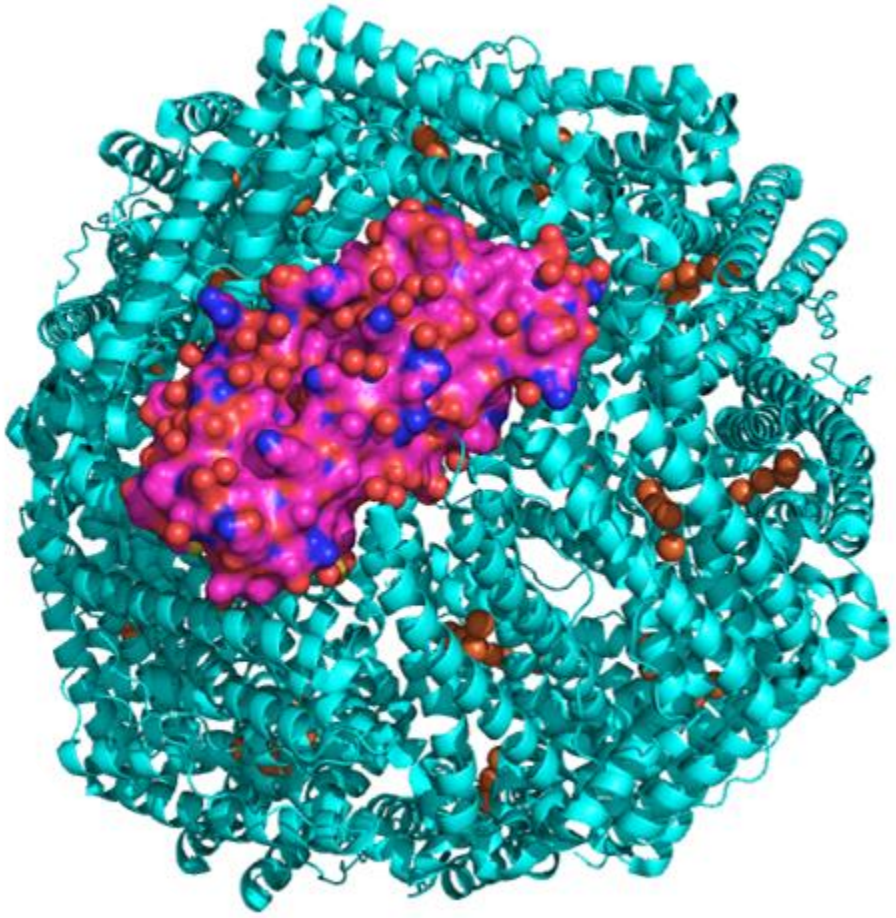
- Asymmetric unit contains protein and single DNA strand
- In assembly, two proteins bind double-stranded DNA.

The importance of assemblies

- Ferritin
- PDB entry 3r2s



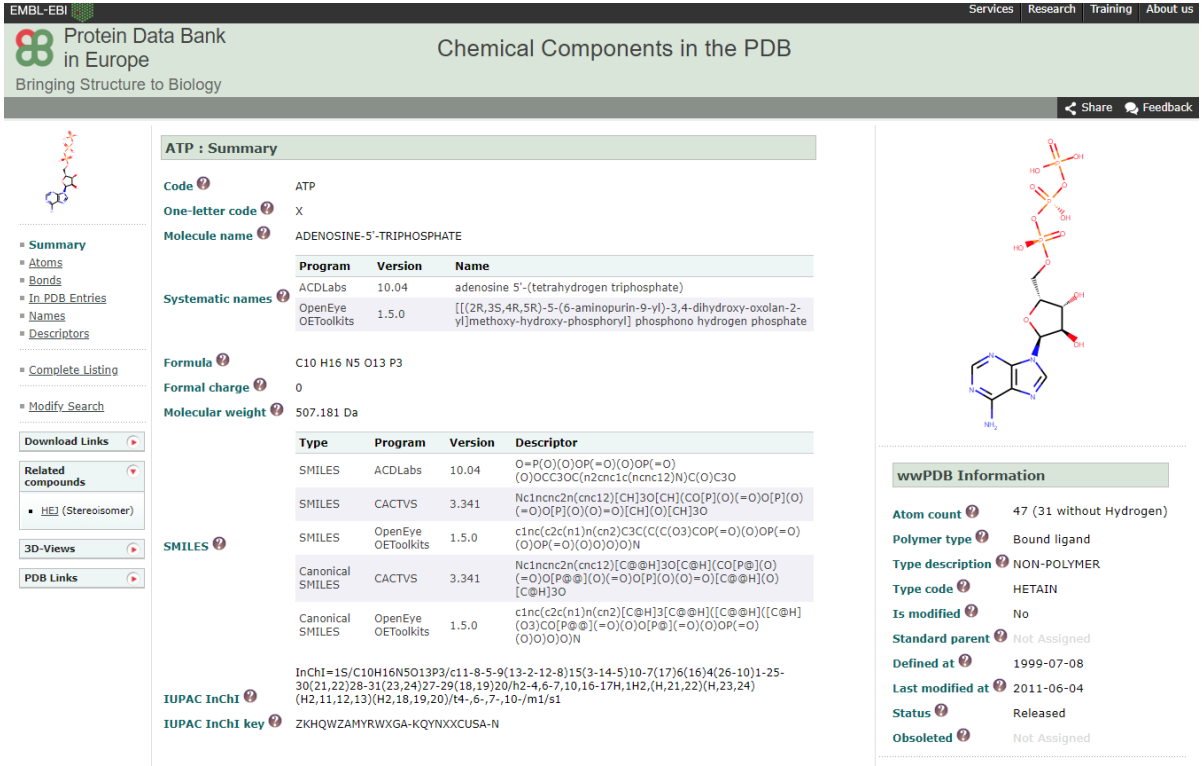
- Deposited coordinates



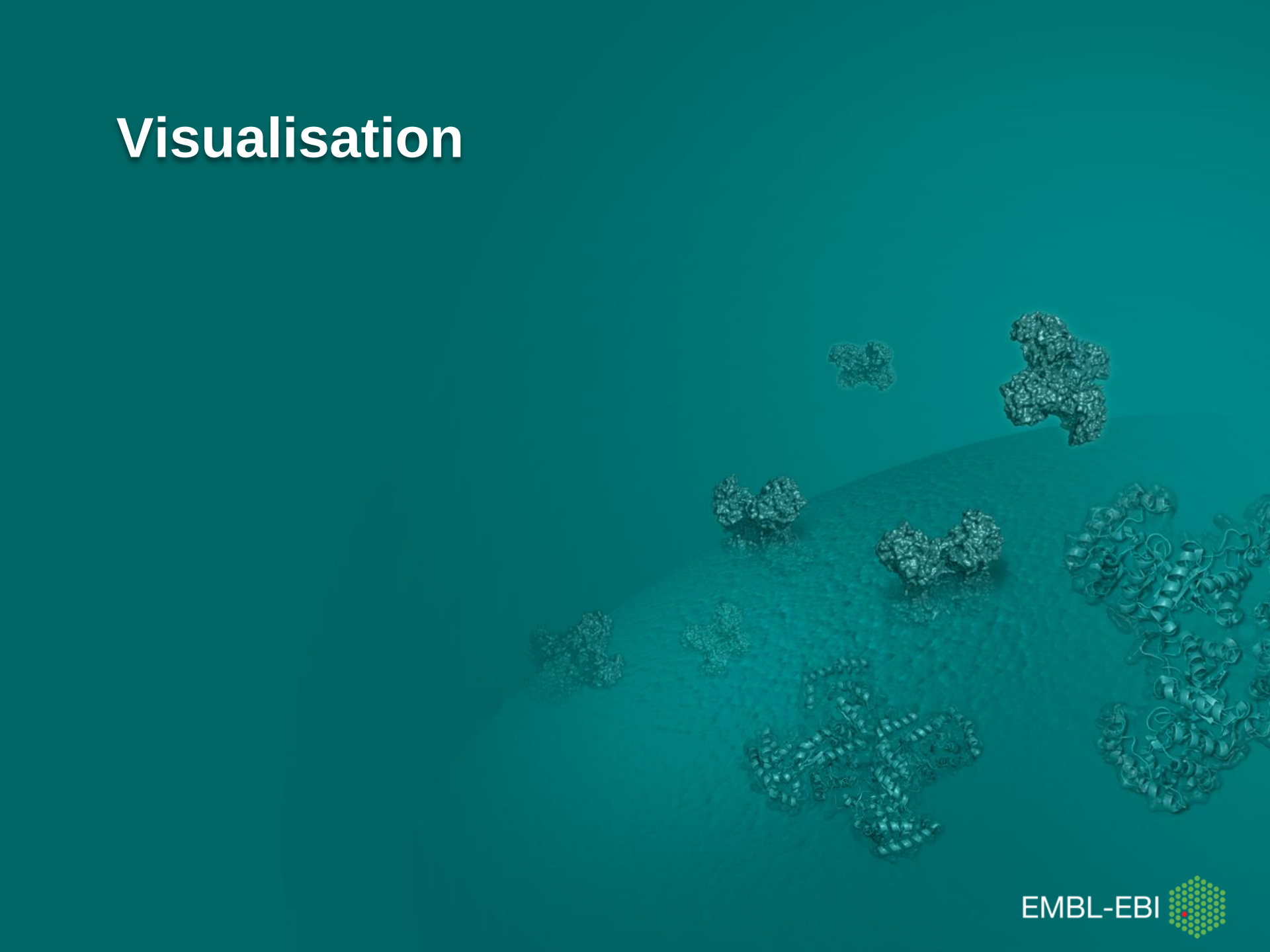
- Symmetry-related assembly

Small molecules

- All molecules mapped to PDB's Chemical Component Dictionary



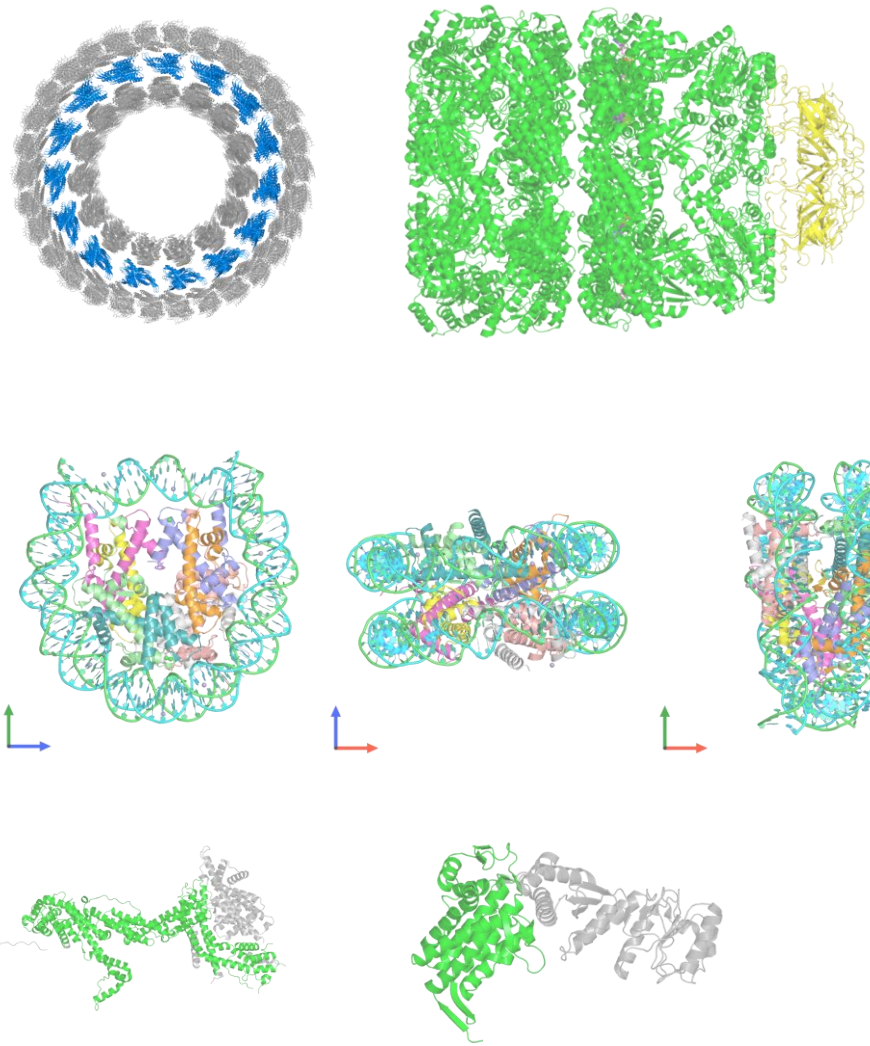
Visualisation



EMBL-EBI

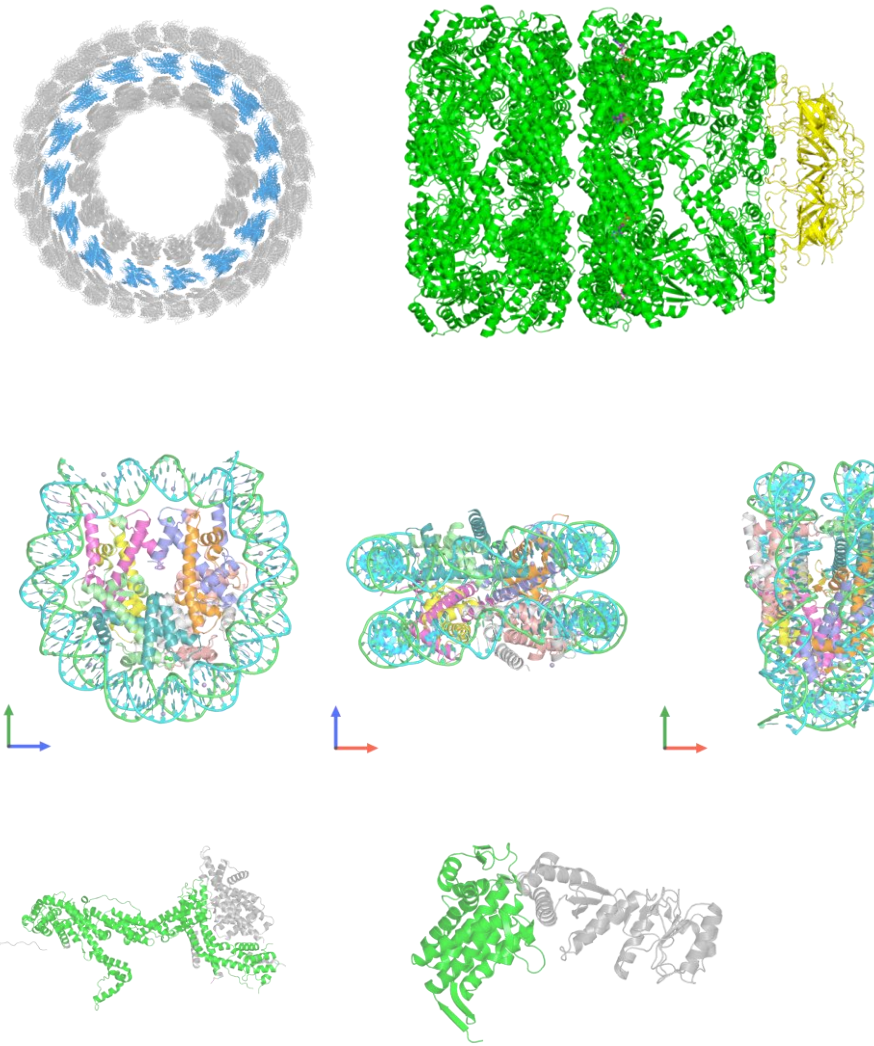
Using images to convey information

- Where is “my protein” in this assembly?
- How many unique macromolecules are there in this complex?
- What is the overall shape of the complex?
- Where is this sequence or structure domain in the 3D structure?



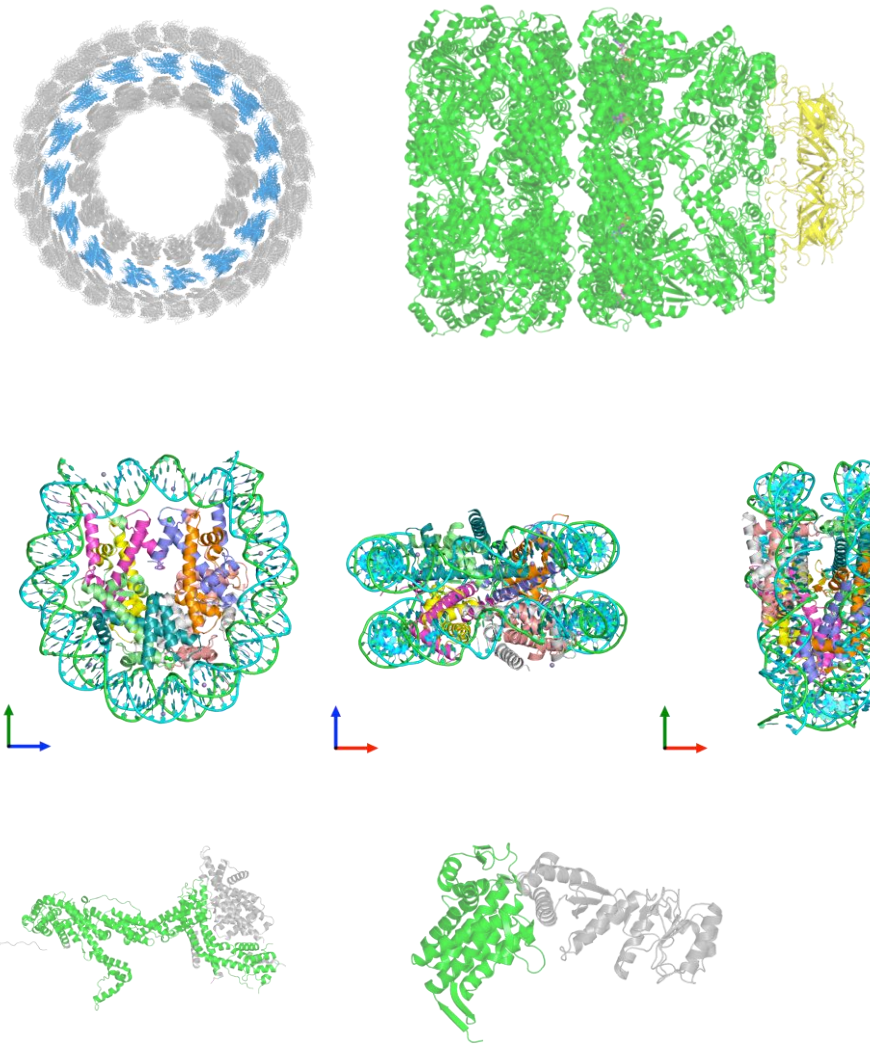
Using images to convey information

- Where is “my protein” in this assembly?
- How many unique macromolecules are there in this complex?
- What is the overall shape of the complex?
- Where is this sequence or structure domain in the 3D structure?



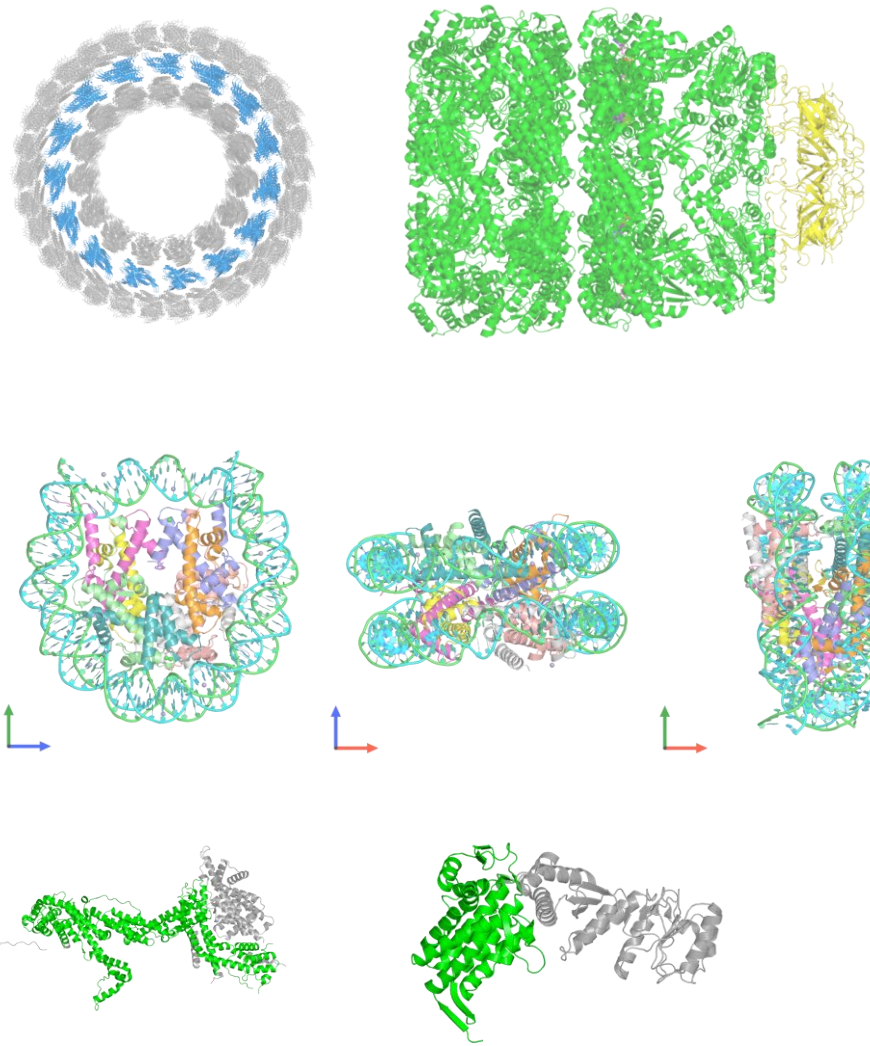
Using images to convey information

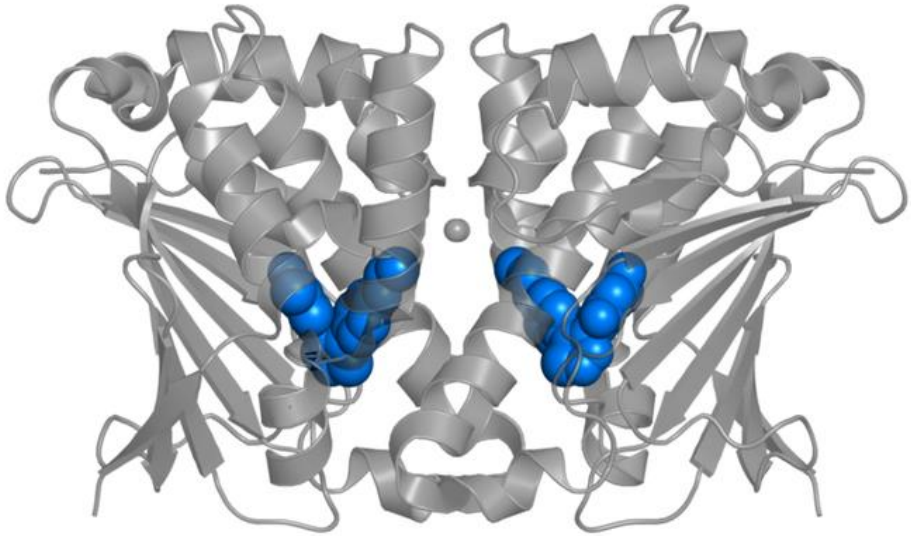
- Where is “my protein” in this assembly?
- How many unique macromolecules are there in this complex?
- What is the overall shape of the complex?
- Where is this sequence or structure domain in the 3D structure?



Using images to convey information

- Where is “my protein” in this assembly?
- How many unique macromolecules are there in this complex?
- What is the overall shape of the complex?
- Where is this sequence or structure domain in the 3D structure?





PDB 2yi7 contains 2 copies of 4-CHLORO-6-[5-(4-ETHOXYPHENYL)-1,2,3-THIADIAZOL-4-YL BENZENE-1,3-DIOL in assembly 1. This small molecule is highlighted and viewed from the front.

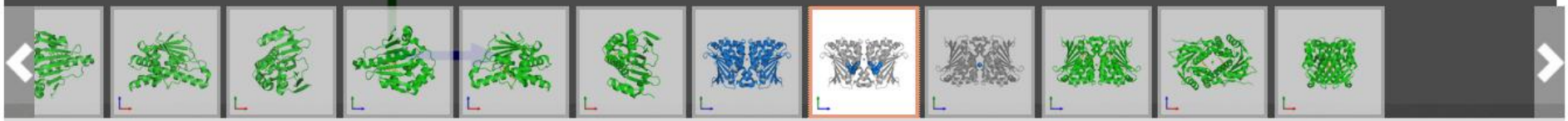


Image 8 of 12

Show captions Share

Protein Data Bank in Europe
Bringing Structure to Biology

PDBe › 2yi7

X-ray diffraction
1.4Å resolution

Structural characterization of 5-Aryl-4-(5-substituted-2-4-dihydroxyphenyl)-1,2,3-thiadiazole Hsp90 inhibitors.

Source organism: *Homo sapiens*

Primary publication:

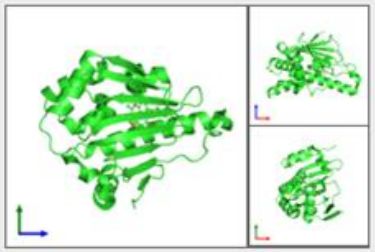
Co-crystalization and in vitro biological characterization of 5-aryl-4-(5-substituted-2-4-dihydroxyphenyl)-1,2,3-thiadiazole hsp90 inhibitors.

Sharp SY, Roe SM, Kazlauskas E, Cikotienė I, Workman P, Matulis D, Prodromou C

PLoS ONE **7** e44642 (2012)
PMID: 22984537

Released: 16 May 2012

Model geometry
Fit model/data



Quick links
2yi7 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

View
3D Visualisation

Function and Biology

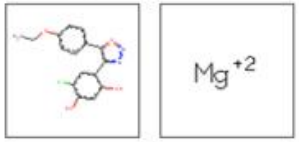
Details

Biochemical function: ATP binding
Biological process: response to stress
Cellular component: not assigned

Sequence domains:
- Heat shock protein Hsp90, N-terminal
- Histidine kinase-like ATPase, C-terminal domain
- Heat shock protein Hsp90 family
- Heat shock protein Hsp90, conserved site

Ligands and Environments

2 bound ligands:



1 x BZ8 1 x MG

No modified residues

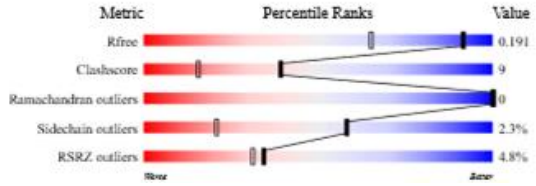
Structure analysis

Details

Assembly composition: homo dimer (preferred)
Entry contents: 1 distinct polypeptide molecule
Macromolecule:
Heat shock protein HSP 90-alpha
Chain: A
Molecule details ›

Experiments and Validation

Details



Citations

2 citation in other articles

Volume of Hsp90 ligand binding and the unfolding phase diagram as a function of pressure and temperature.
Petrauskas et al. (2013)

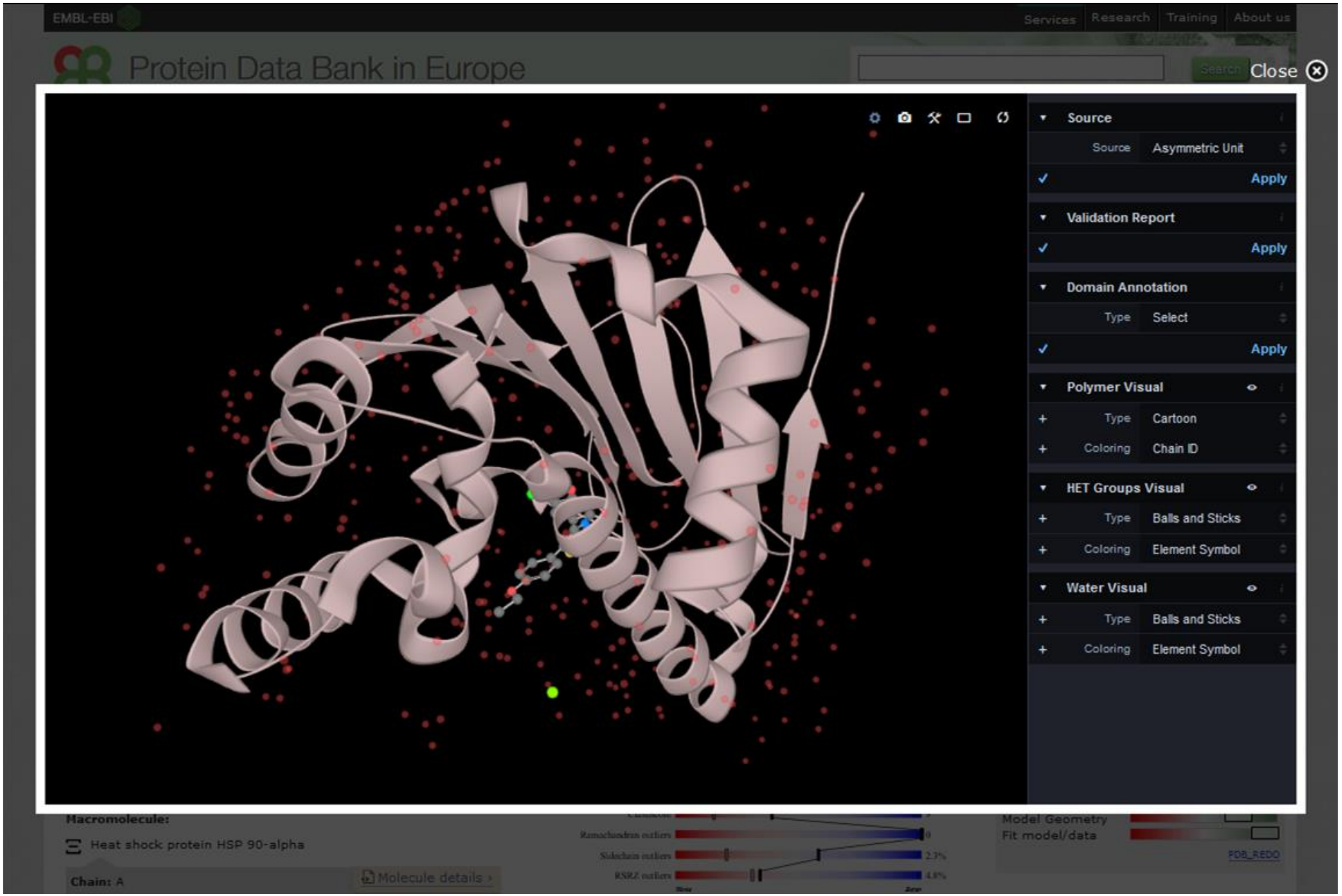
1 more

PDB_REDO

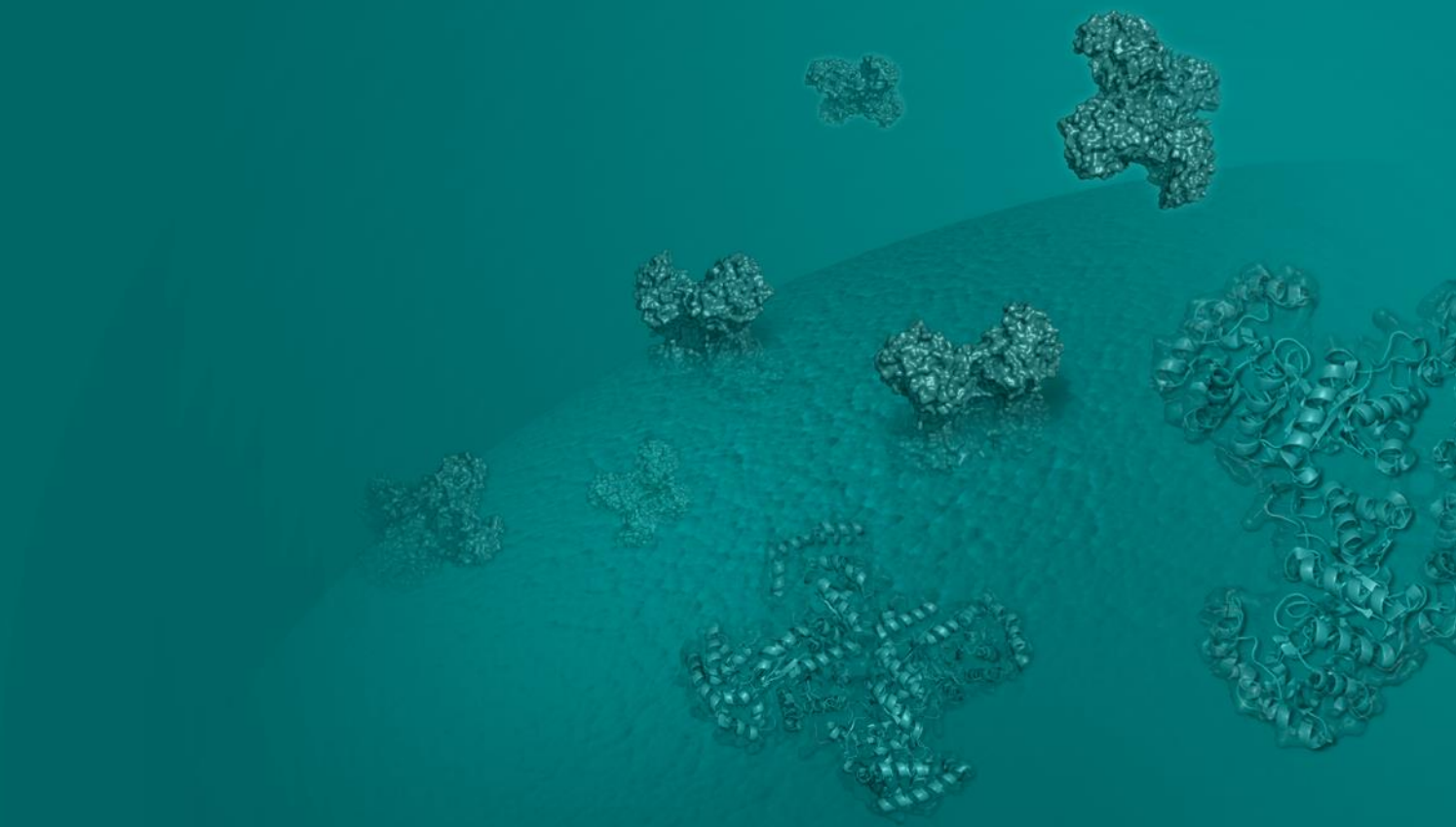
The sliders below show the change in model quality between original PDB entry and the PDB_REDO entry

Model Geometry
Fit model/data

PDB_REDO


EMBL-EBI
Services
Research
Training
About us
Protein Data Bank in Europe
Search
Close
Source
Source
Asymmetric Unit
Apply
Validation Report
Apply
Domain Annotation
Type
Select
Apply
Polymer Visual
Type
Cartoon
Coloring
Chain ID
HET Groups Visual
Type
Balls and Sticks
Coloring
Element Symbol
Water Visual
Type
Balls and Sticks
Coloring
Element Symbol
Macromolecule:
Heat shock protein HSP 90-alpha
Chain: A
Molecule details
Ramachandran outliers
Sidechain outliers
RSRZ outliers
Model Geometry
Fit model/data
PDB_REDO

Searching the PDB



EMBL-EBI

Data out

The problem with the PDB is…

- I can't find what I want
- Too many false hits when I search
  - CaM, CaM-Kinase, CaM binding protein, Cam-like…
- Results are too complicated
  - Which lysozyme to use? 47 species!
    - Which is best?/what is best? >250 ligands
- Redundancy
  - How many unique human protein structures?

Searching the PDB made simple

At PDBe we've implemented:

- Auto-complete suggestions
- Facets to narrow down search results
- Quality presented on search results
- Four views of results
  - Moving away from entry-centricity

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

Protein Data Bank in Europe

Bringing Structure to Biology

Search

Examples: hemoglobin, BRCA1_HUMAN

EMsearch

PDBe home | Deposition | PDBe services | PDBe training | Documentation | About PDBe | Share | Feedback

PDBe is the European resource for the collection, organisation and dissemination of data on biological macromolecular structures. Read more about PDBe.

Featured structure

Stop motion: The muscular system

1st May 2018

The image for May in our 2018 calendar captures a molecular snapshot of one of our primary groups of organs – the muscular system. Here we discuss the individual molecules responsible for the large motions of these muscles.

Read more...

Previous featured structures

News

David Blow Poster Prize awarded at BCA

13 April, 2018

PDBe Explores Art

29 March, 2018

Events

University of Cambridge Protein Structure Analysis course

University of Cambridge, Cambridge, UK
24 May 2018

Three Dimensional Electron Microscopy -

Popular

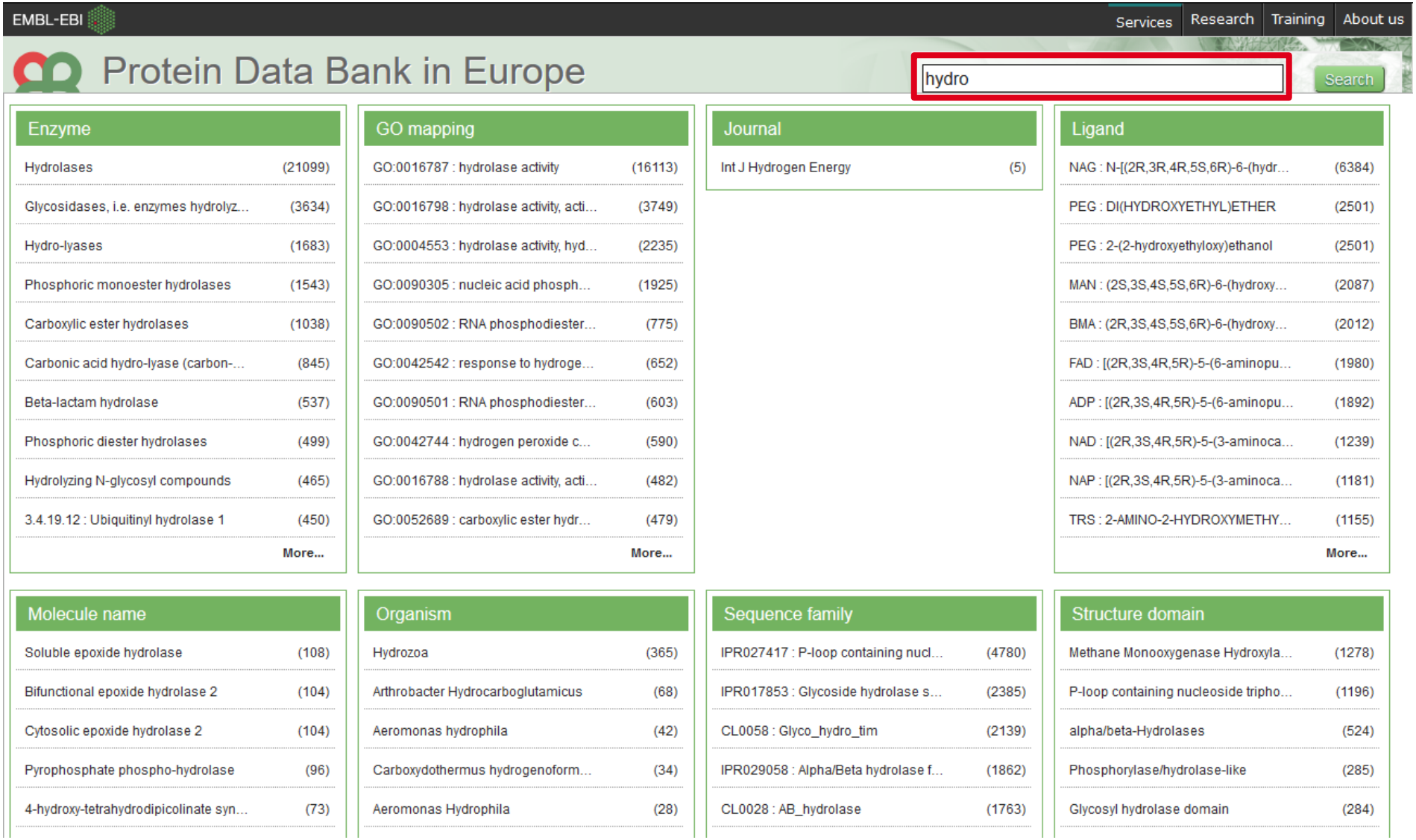
- EMsearch
- PDBeFold
- PDBePISA
- PDBeChem
- Sequence search
- PDBe REST API
- EM resources
- NMR resources
- EMPIAR
- Coordinate Server
- PDB Component Library
- News
- Events
- Training
- Contact us

Latest archive statistics

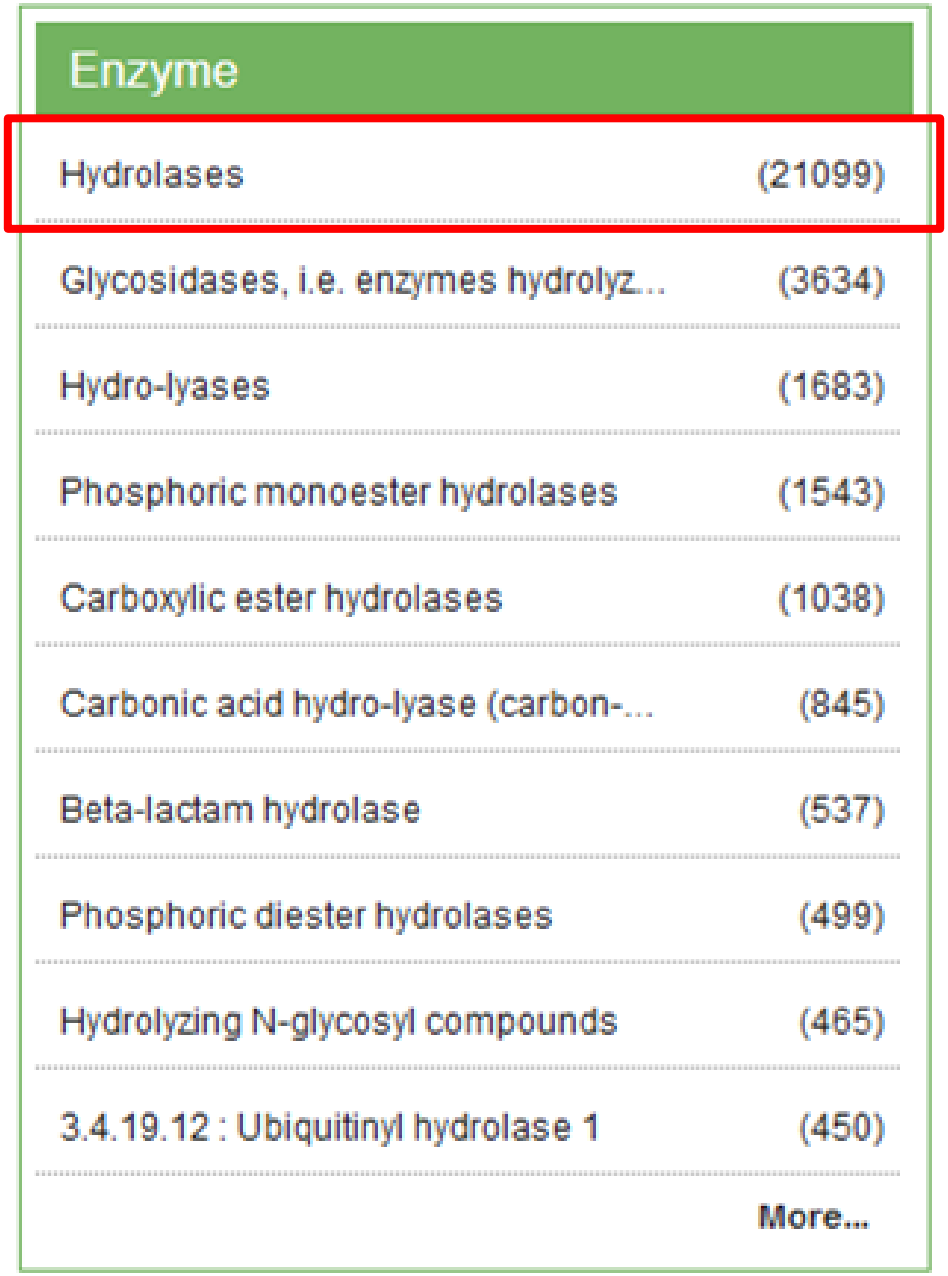
As of 23 May 2018 the PDB contains 140591 entries (latest PDB entries, chemistry, biology) and EMDB contains 6215 entries (latest map releases, latest header releases, latest updates).

Tweets by @PDBeurope

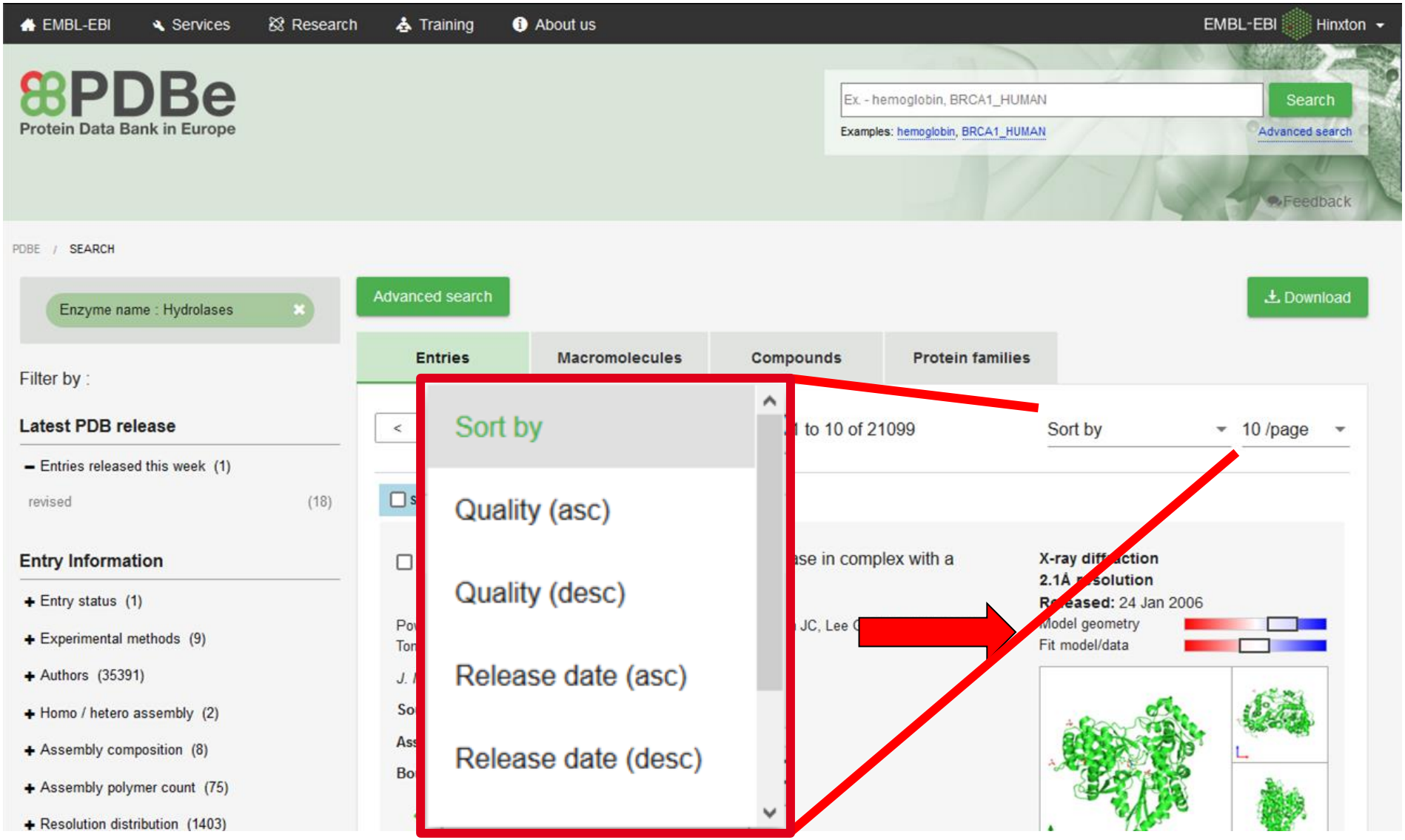
Protein Data Bank
@PDBeurope
How snail toxins could treat pain. This structure,


EMBL-EBI
Services
Research
Training
About us
Protein Data Bank in Europe
hydro
Search
Enzyme
Hydrolases (21099)
Glycosidases, i.e. enzymes hydrolyz... (3634)
Hydro-lyases (1683)
Phosphoric monoester hydrolases (1543)
Carboxylic ester hydrolases (1038)
Carbonic acid hydro-lyase (carbon-... (845)
Beta-lactam hydrolase (537)
Phosphoric diester hydrolases (499)
Hydrolyzing N-glycosyl compounds (465)
3.4.19.12 : Ubiquitinyl hydrolase 1 (450)
More...
GO mapping
GO:0016787 : hydrolase activity (16113)
GO:0016798 : hydrolase activity, acti... (3749)
GO:0004553 : hydrolase activity, hyd... (2235)
GO:0090305 : nucleic acid phosph... (1925)
GO:0090502 : RNA phosphodiester... (775)
GO:0042542 : response to hydroge... (652)
GO:0090501 : RNA phosphodiester... (603)
GO:0042744 : hydrogen peroxide c... (590)
GO:0016788 : hydrolase activity, acti... (482)
GO:0052689 : carboxylic ester hydr... (479)
More...
Journal
Int J Hydrogen Energy (5)
Ligand
NAG : N-[(2R,3R,4R,5S,6R)-6-(hydr... (6384)
PEG : DI(HYDROXYETHYL)ETHER (2501)
PEG : 2-(2-hydroxyethyloxy)ethanol (2501)
MAN : (2S,3S,4S,5S,6R)-6-(hydroxy... (2087)
BMA : (2R,3S,4S,5S,6R)-6-(hydroxy... (2012)
FAD : [(2R,3S,4R,5R)-5-(6-aminopu... (1980)
ADP : [(2R,3S,4R,5R)-5-(6-aminopu... (1892)
NAD : [(2R,3S,4R,5R)-5-(3-aminoca... (1239)
NAP : [(2R,3S,4R,5R)-5-(3-aminoca... (1181)
TRS : 2-AMINO-2-HYDROXYMETHY... (1155)
More...
Molecule name
Soluble epoxide hydrolase (108)
Bifunctional epoxide hydrolase 2 (104)
Cytosolic epoxide hydrolase 2 (104)
Pyrophosphate phospho-hydrolase (96)
4-hydroxy-tetrahydrodipicolinate syn... (73)
Organism
Hydrozoa (365)
Arthrobacter Hydrocarboglutamicus (68)
Aeromonas hydrophila (42)
Carboxydothermus hydrogenoform... (34)
Aeromonas Hydrophila (28)
Sequence family
IPR027417 : P-loop containing nucl... (4780)
IPR017853 : Glycoside hydrolase s... (2385)
CL0058 : Glyco_hydro_tim (2139)
IPR029058 : Alpha/Beta hydrolase f... (1862)
CL0028 : AB_hydrolase (1763)
Structure domain
Methane Monooxygenase Hydroxyla... (1278)
P-loop containing nucleoside tripho... (1196)
alpha/beta-Hydrolases (524)
Phosphorylase/hydrolase-like (285)
Glycosyl hydrolase domain (284)

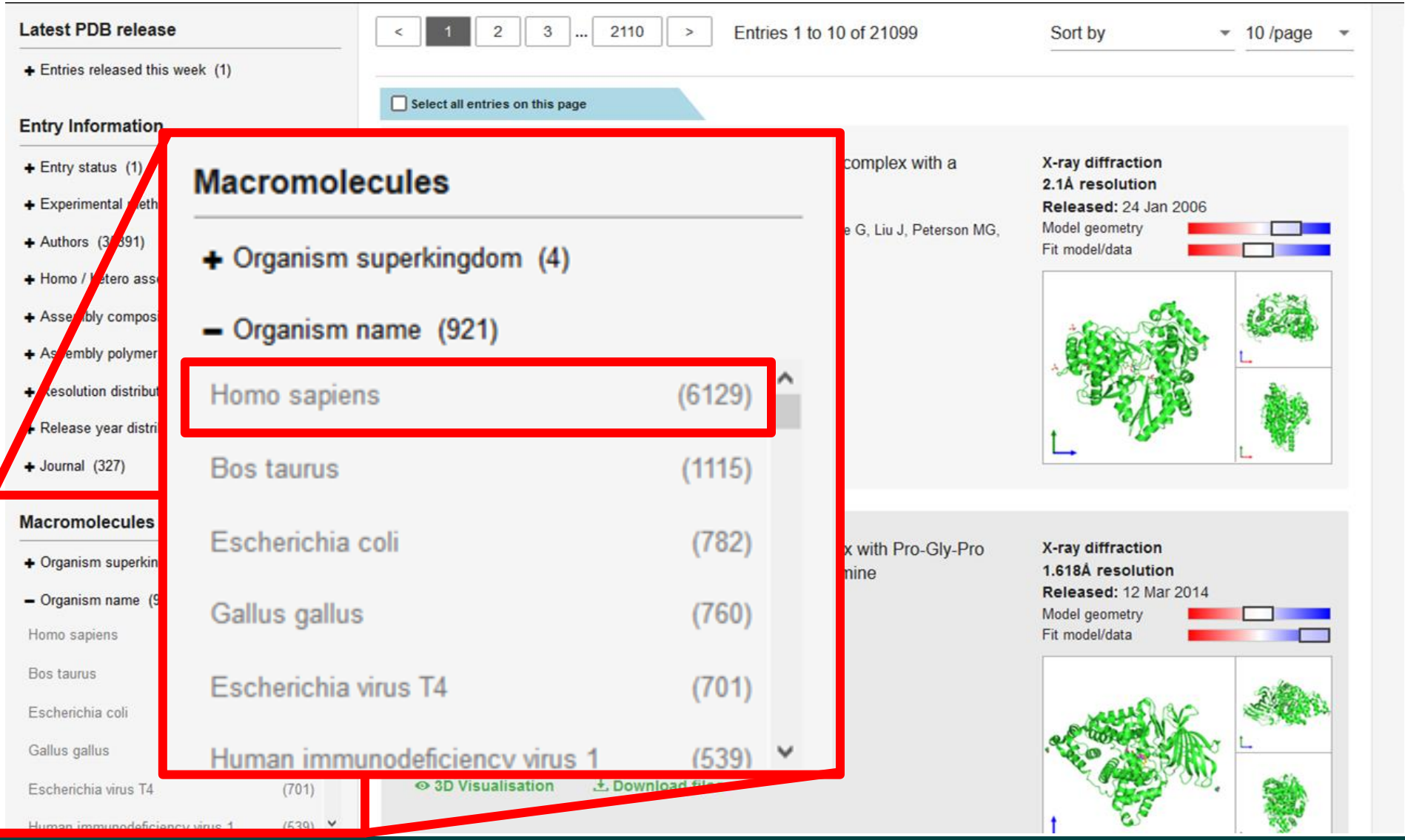

Organism
Hydrozoa (365)
Arthrobacter Hydrocarboglutamicus (68)
Aeromonas hydrophila (42)
Carboxydothermus hydrogenoform... (34)
Aeromonas Hydrophila (28)
Bacillus Hydrophilus Fuscus (28)
Bacterium Hydrophilum (28)
Proteus Hydrophilus (28)
Pseudomonas Hydrophila (28)
Aeromonas Hydrophila Proteolytica (25)
More...


Enzyme
Hydrolases (21099)
Glycosidases, i.e. enzymes hydrolyz... (3634)
Hydro-lyases (1683)
Phosphoric monoester hydrolases (1543)
Carboxylic ester hydrolases (1038)
Carbonic acid hydro-lyase (carbon-... (845)
Beta-lactam hydrolase (537)
Phosphoric diester hydrolases (499)
Hydrolyzing N-glycosyl compounds (465)
3.4.19.12 : Ubiquitinyl hydrolase 1 (450)
More...

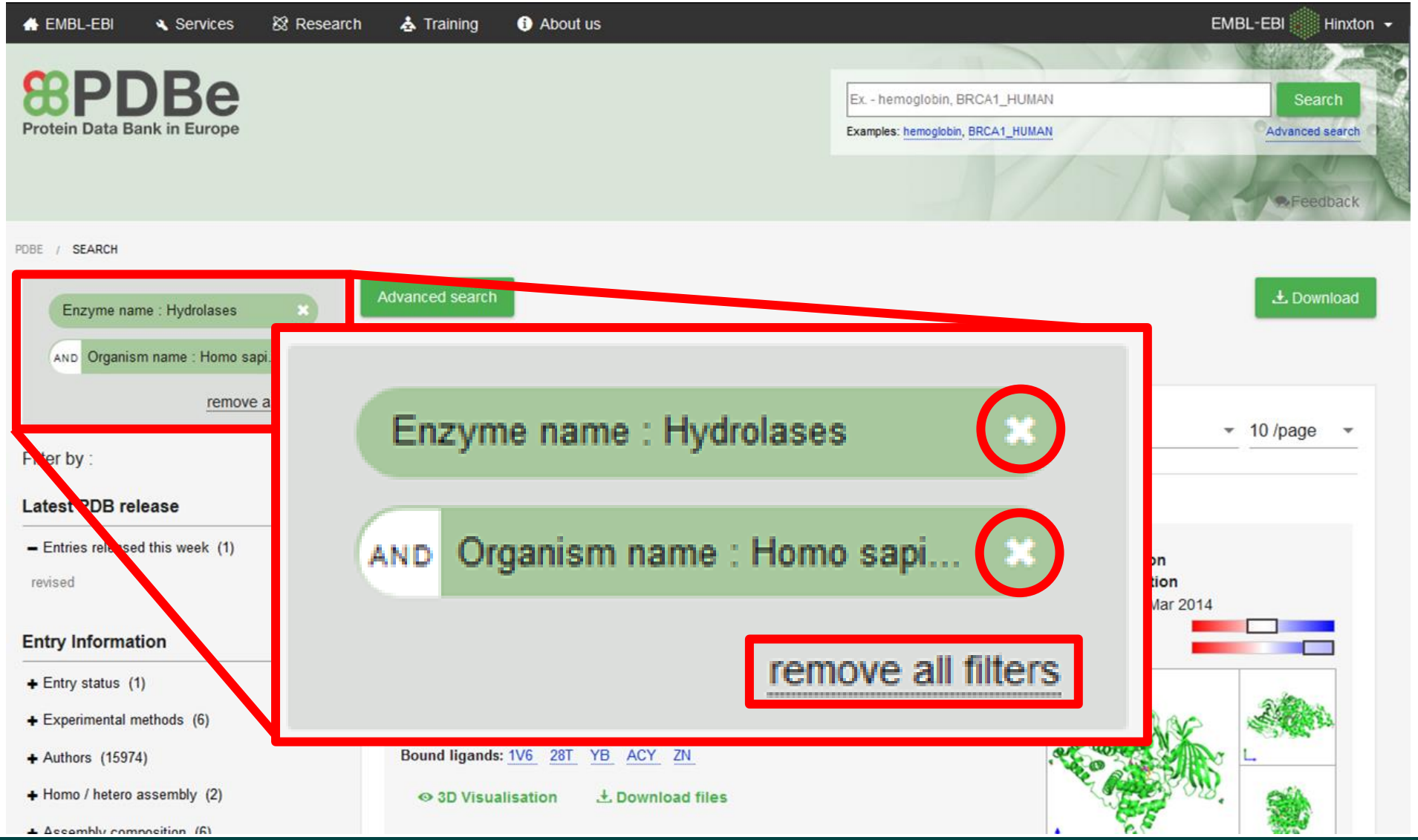
Query – “Enzyme: Hydrolases”



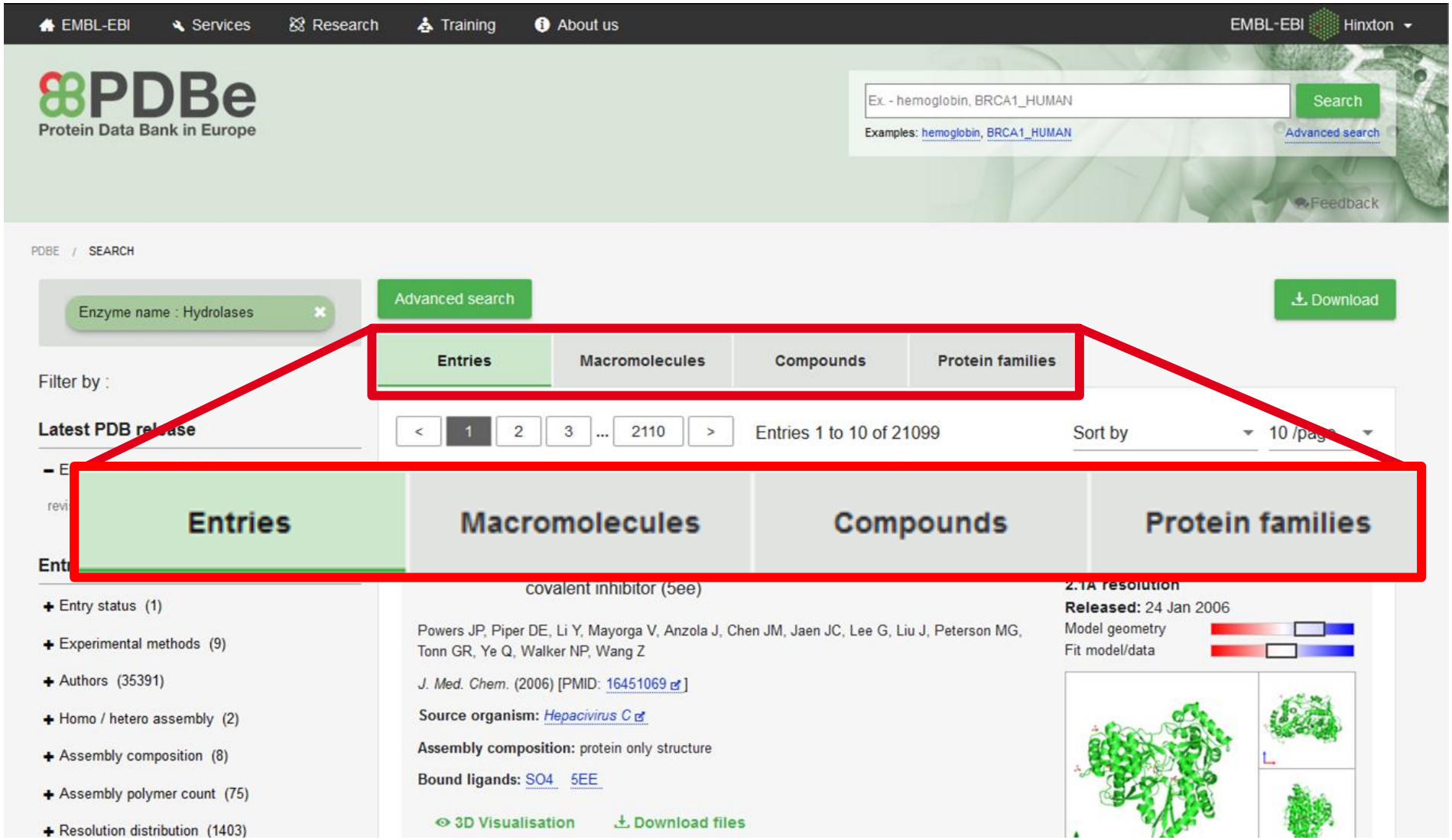
Drill down with facets



Drill down with facets

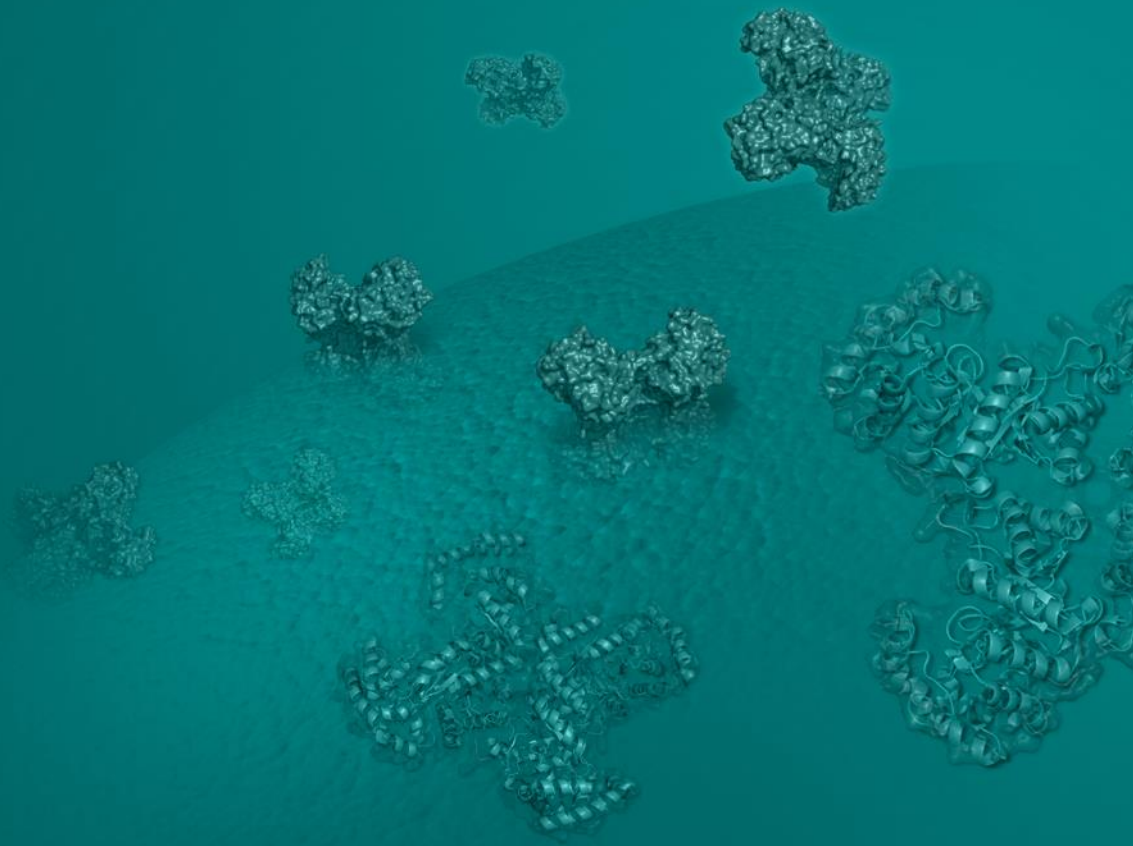


Query – “Enzyme: Hydrolases”



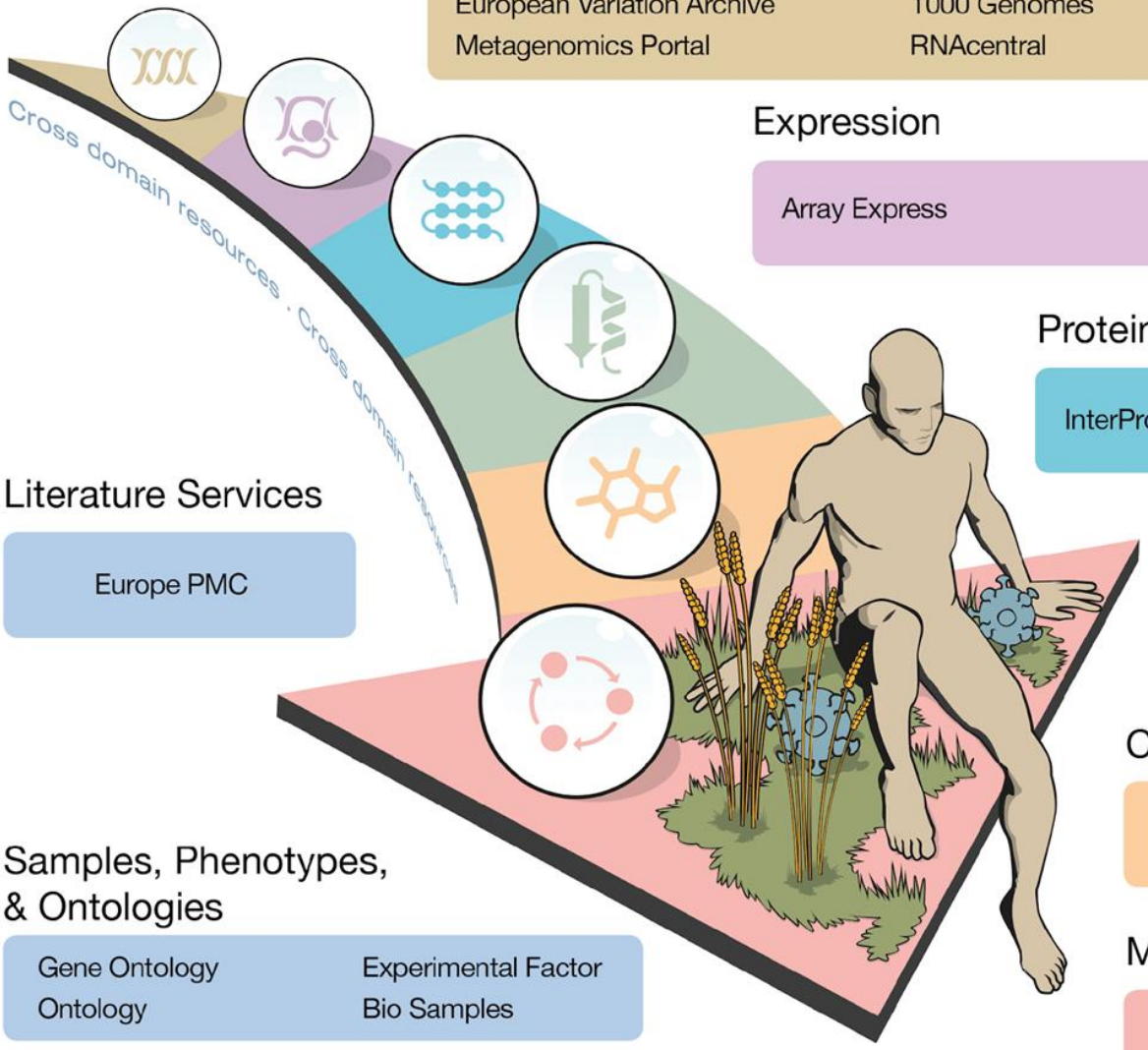
Introducing PDBe-KB

pdbe-kb.org



EMBL-EBI

At the heart of EMBL-EBI resources



Genes, Genomes & Variation

European Nucleotide Archive
European Variation Archive
Metagenomics Portal
Ensembl
1000 Genomes
RNAcentral
European Genome-phenome Archive
Ensembl Genomes

Expression

Array Express
Expression Atlas
PRIDE

Proteins & Protein Families

InterPro
Pfam
UniProt

Literature Services

Europe PMC

Molecular & Cellular Structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical Biology

ChEMBL
ChEBI
MetaboLights

Samples, Phenotypes, & Ontologies

Gene Ontology
Experimental Factor Ontology
Bio Samples

Molecular Systems

BioModels
IntAct
Reactome

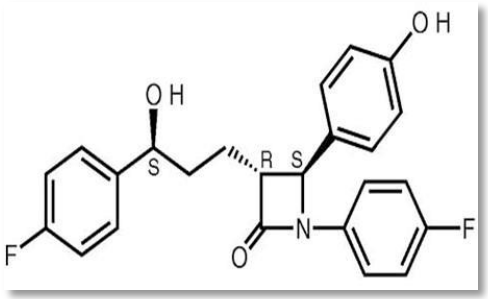
Integrating structure and other biological data

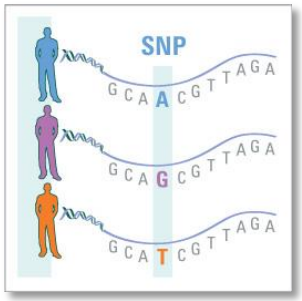
- Annotating and linking structure through biological information

Chemistry | Sequences | Variation | Interactions | Pathways

Information / Resources

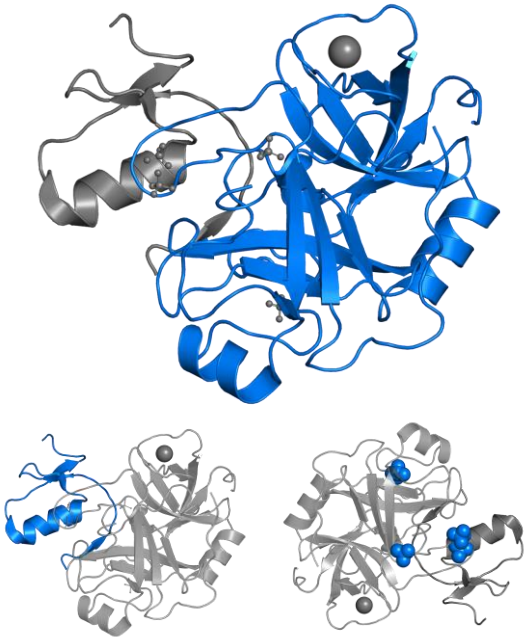
ChEBI | ENA/UniProt | Ensembl | ChEMBL/IntAct | Reactome



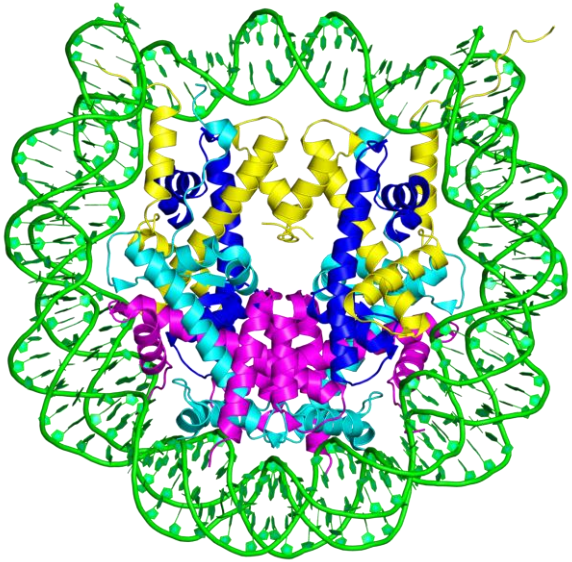


Change the context of structural data

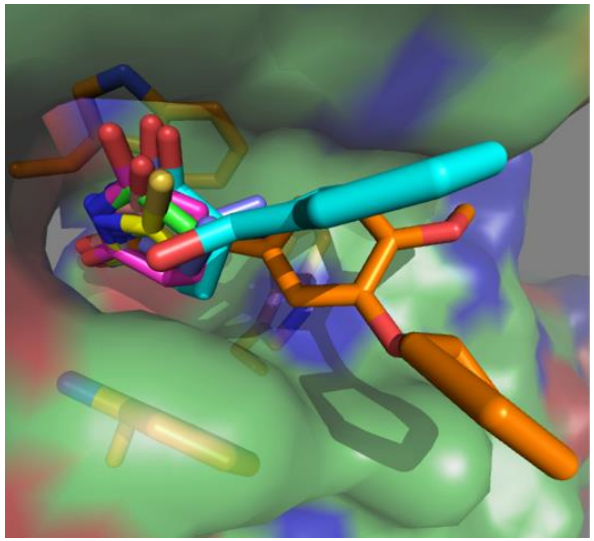
Proteins



Complexes

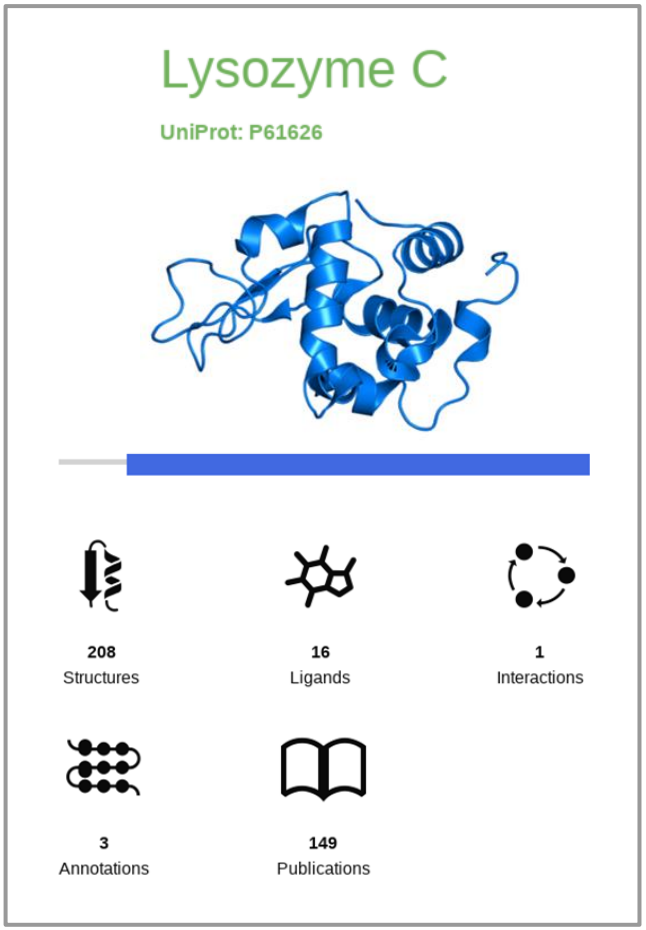


Binding sites



Protein pages

- Structural context for a given protein
  - Initially based on Uniprot ID
  - Serving data through a new API
- Data clustered by different sections
  - Structure coverage
  - Ligands and binding sites
  - Protein interactions
  - Additional annotations
  - Publications



Protein pages - summary

PDBe-KB › Mediator of DNA damage checkpoint protein 1

Gene: MDC1

Organism: *Homo sapiens (Human)*

Synonyms: KIAA0170, NFBD1

Uniprot: Q14676 [go to UniProt]

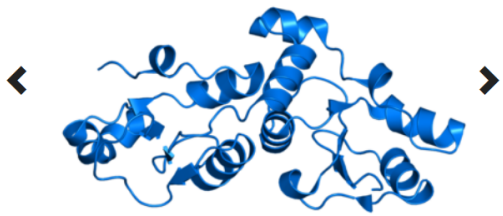
Biological function: Required for checkpoint mediated cell cycle arrest in response to DNA damage within both the S phase and G2/M phases of the cell cycle. May serve as a scaffold for the recruitment of DNA repair and signal transduction proteins to discrete foci of DNA damage marked by 'Ser-139' phosphorylation of histone H2AFX. Also required for downstream events subsequent to the ... [show more] [go to UniProt]

10	2	4	2	0	250
Structures	Ligands	Interactions	Functional annotations	Similar proteins	Publications

Representative structures for UniProt Q14676

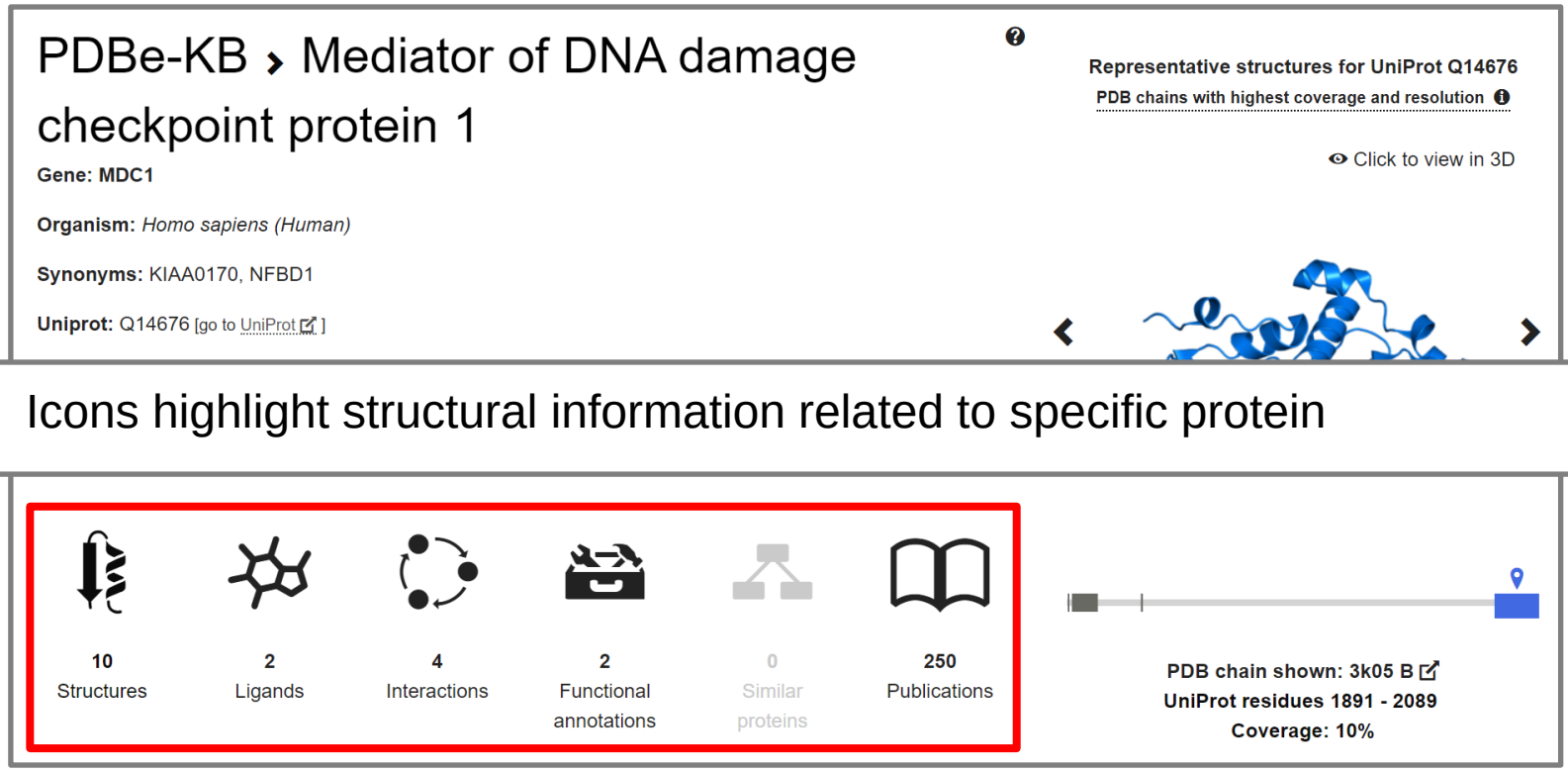
PDB chains with highest coverage and resolution

Click to view in 3D



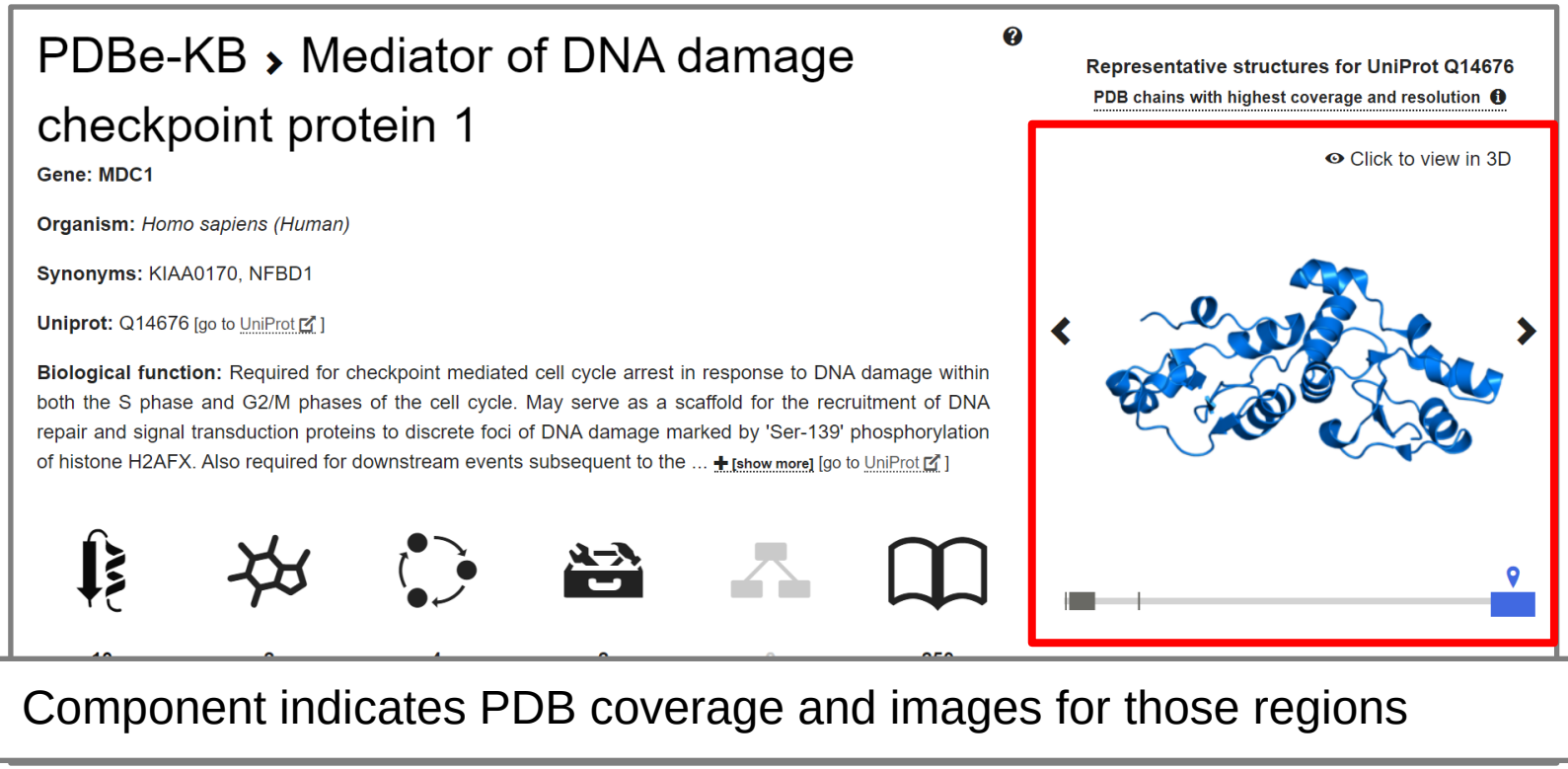
PDB chain shown: 3k05 B
UniProt residues 1891 - 2089
Coverage: 10%

Protein pages - summary



- Icons highlight structural information related to specific protein

Protein pages - summary



- Component indicates PDB coverage and images for those regions

Protein pages - structures

Overview of structures and domains in the PDB using ProtVista viewer

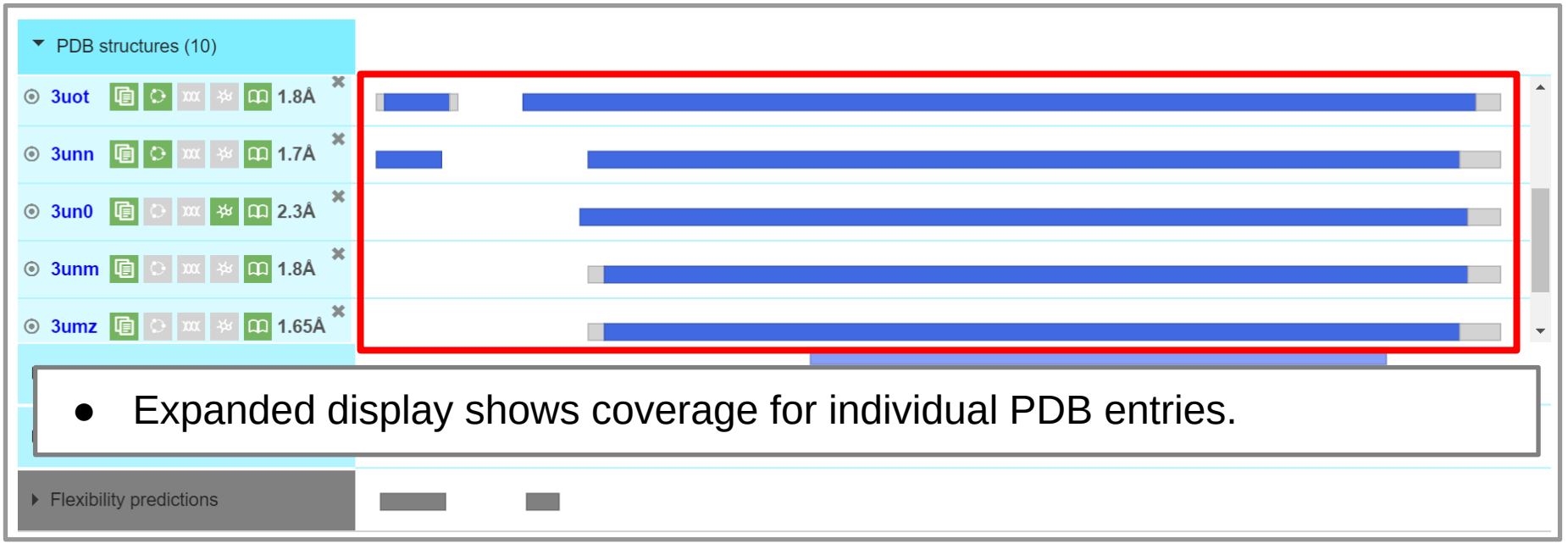
Protein pages - structures

Overview of structures and domains in the PDB using ProtVista viewer

- Protein coverage displayed for all PDB structures

Protein pages - structures

Overview of structures and domains in the PDB using ProtVista viewer



- Expanded display shows coverage for individual PDB entries.

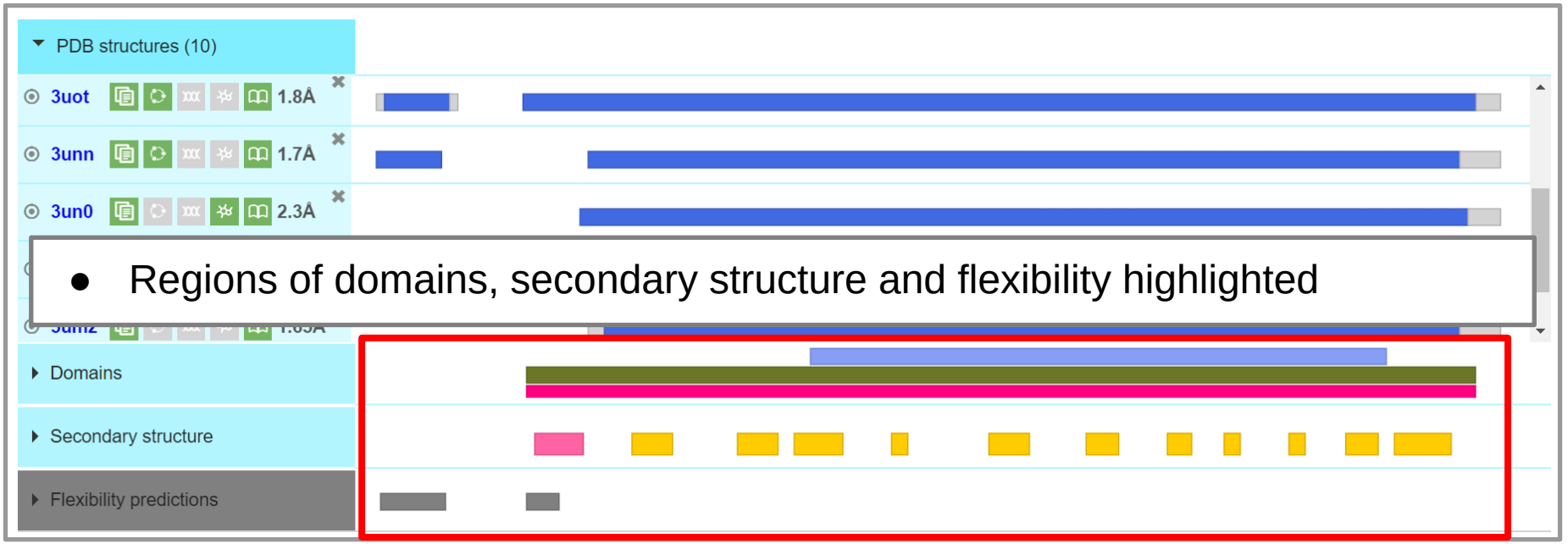
Protein pages - structures

Overview of structures and domains in the PDB using ProtVista viewer

- Clicking on the range allows you to go to that specific PDB entry

Protein pages - structures

Overview of structures and domains in the PDB using ProtVista viewer



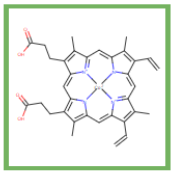
- Regions of domains, secondary structure and flexibility highlighted

Protein pages - ligands

Overview of ligands and binding residues in the PDB

Show all molecules from PDB entries containing this protein:

Checking this box will show molecules which may not directly interact with the protein.

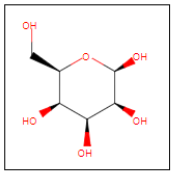


COH

co-factor

Found in **7** entries

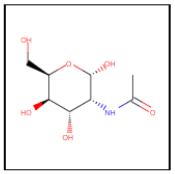
3D view



MAN

Found in **7** entries

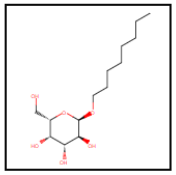
3D view



NAG

Found in **7** entries

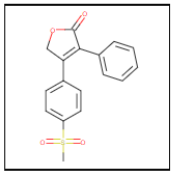
3D view



BOG

Found in **5** entries

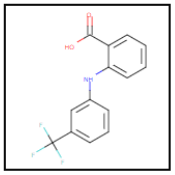
3D view



RCX

Found in **1** entry

3D view



FLF

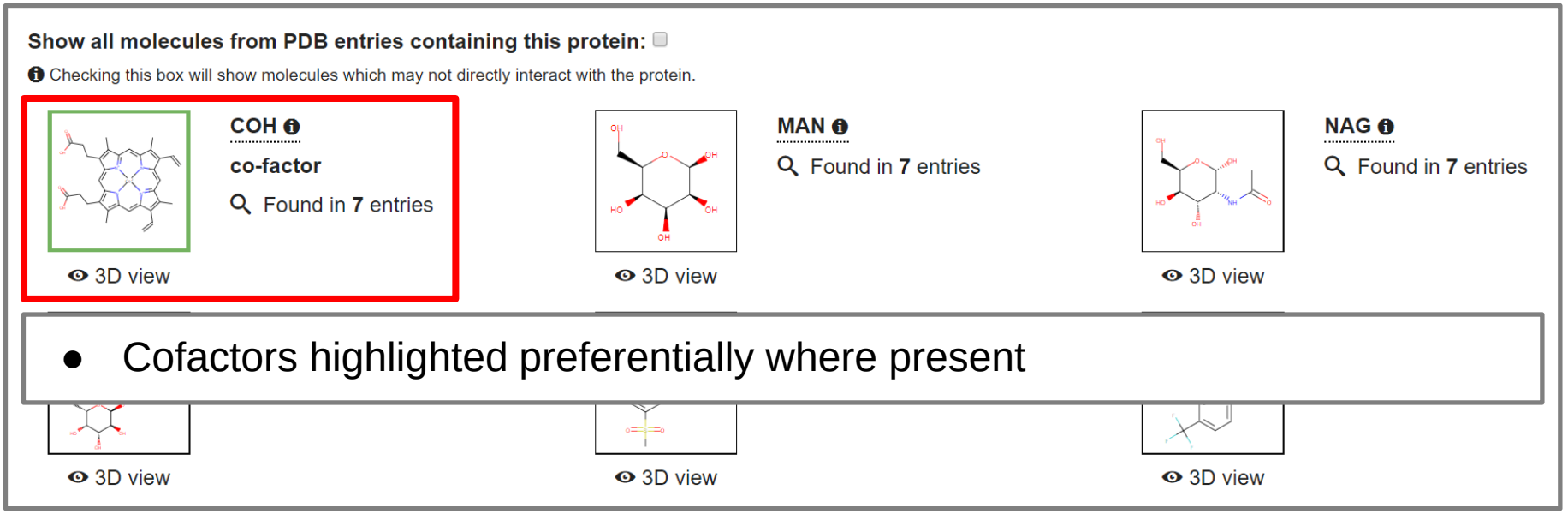
Found in **1** entry

3D view

EMBL-EBI

Protein pages - ligands

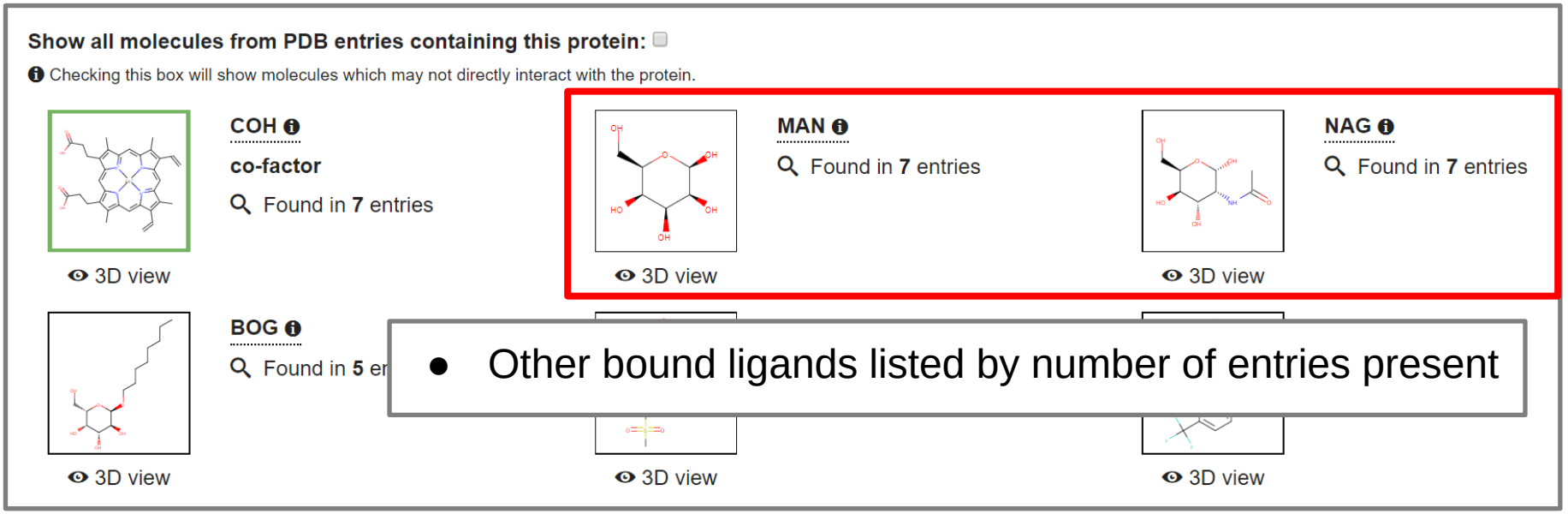
Overview of ligands and binding residues in the PDB



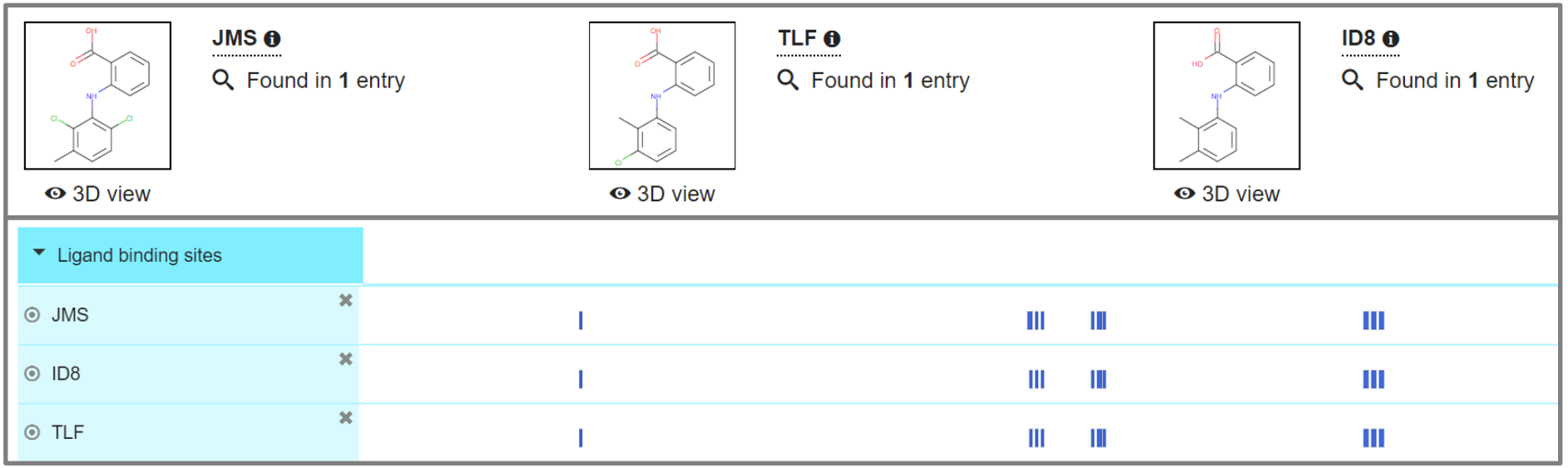
- Cofactors highlighted preferentially where present

Protein pages - ligands

Overview of ligands and binding residues in the PDB



Protein pages - ligands



- ProtVista shows location of all ligand binding sites
  - Begin to see trend of important binding residues

Protein pages - interactions

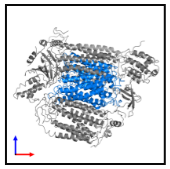
Overview of interaction partners and interacting residues in the PDB

Interaction Partners (21)

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.

Search: e.g. Cytochrome

Search by molecule name, code or PDB id.

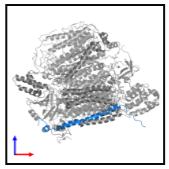


P00396

Found in **50** entries

View page

3D view

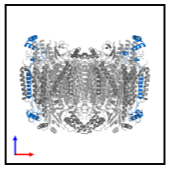


P04038

Found in **50** entries

View page

3D view

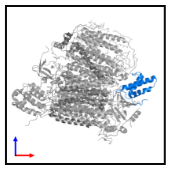


P00423

Found in **50** entries

View page

3D view

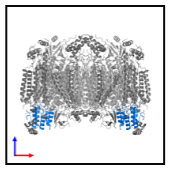


P00429

Found in **50** entries

View page

3D view

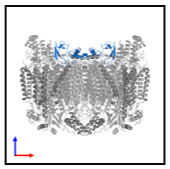


P00426

Found in **50** entries

View page

3D view



P00428

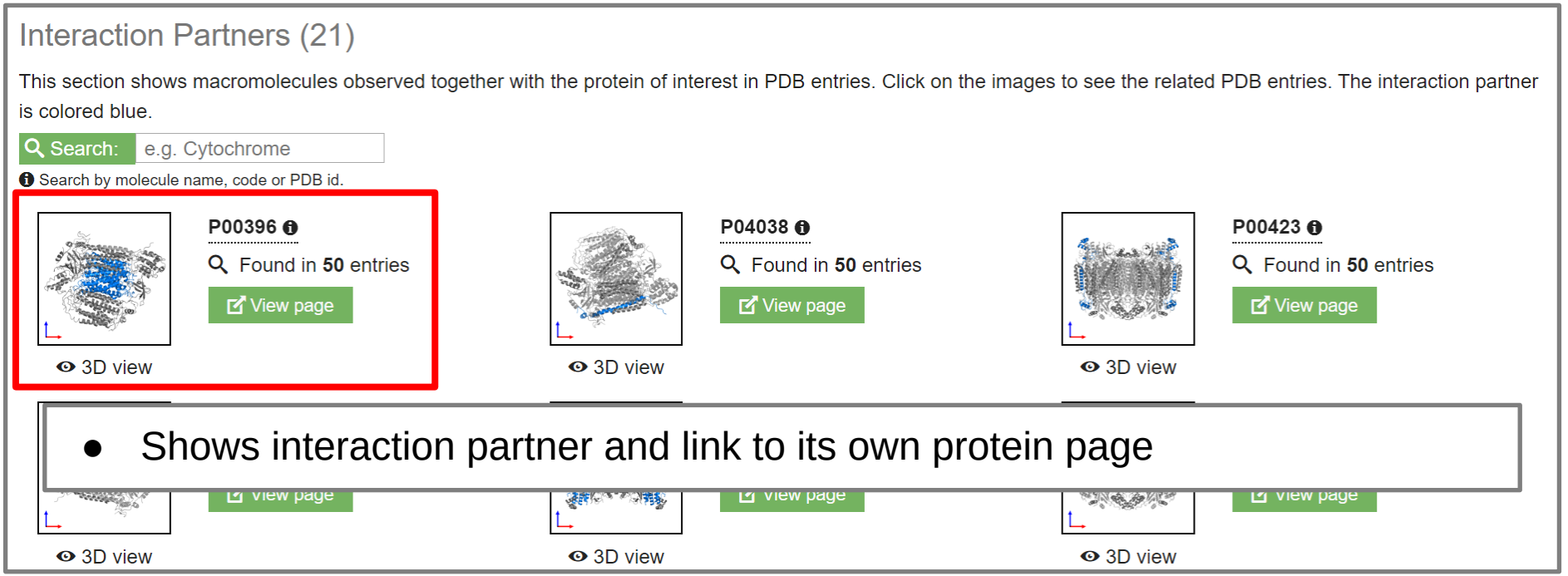
Found in **39** entries

View page

3D view

Protein pages - interactions

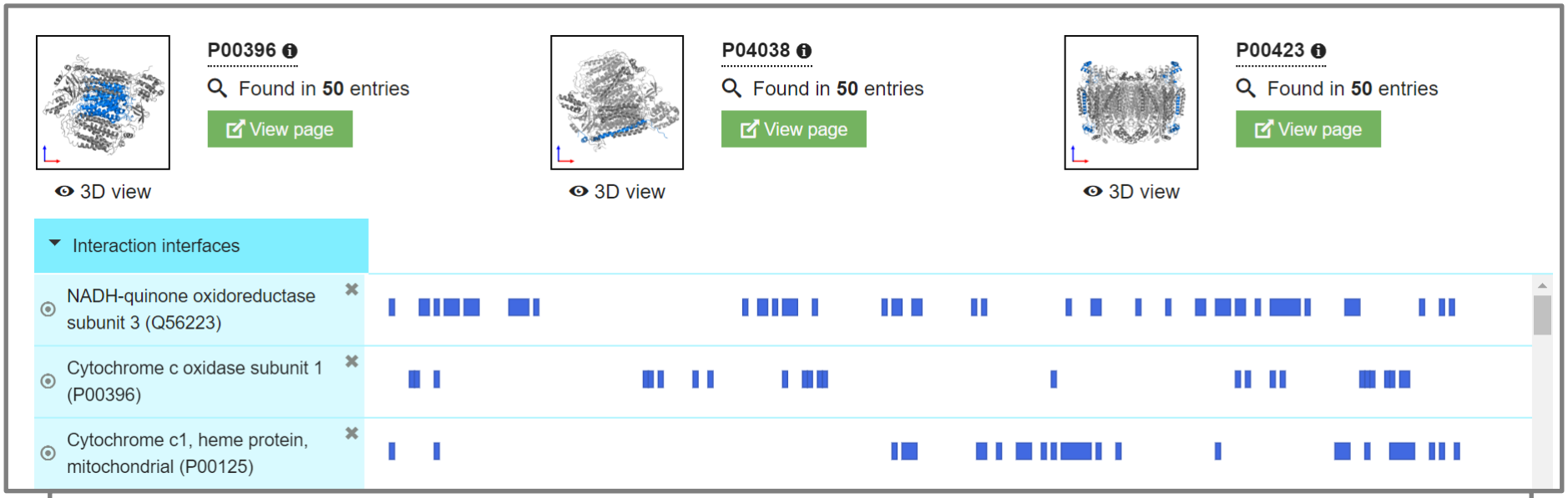
Overview of interaction partners and interacting residues in the PDB



- Shows interaction partner and link to its own protein page

Protein pages - interactions

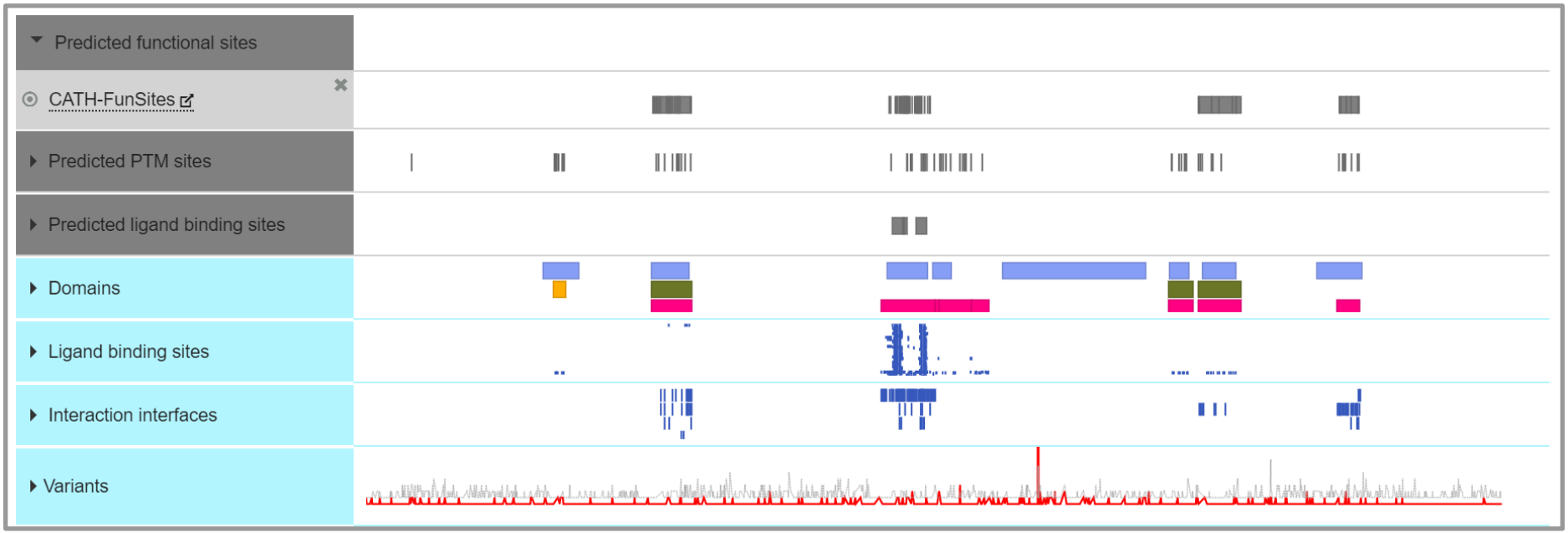
Overview of interaction partners and interacting residues in the PDB



- Shows location of all interaction sites

Protein pages - additional annotations

Adding data from related databases through PDBe-KB collaborations



Protein pages - try them yourself!

Directly accessible from **PDBe-KB.org/proteins**

- Try the examples or input your own **Uniprot ID** or **PDB entry**

Link to PDBe-KB pages from the **search at PDBe.org**