Using the PDBe
The Protein Data Bank (PDB) is an archive of experimentally determined 3-dimensional structures of biological macromolecules.
A “PDB code” refers to a structure

- A unique 4 character code
- Identifies the data within the PDB archive
- Always starts with a number
  - eg 2nuu, 4xyz, 2f48
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
The PDB
FTP Archive

Value added data

PDB
Protein Data Bank

PDBj
Protein Data Bank Japan

PDBe
Protein Data Bank in Europe

BMRB

wwPDB.org
What can we learn from structures?
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 Å
  - Medium = 1.8 - 3.0 Å
  - High = 1.0 - 1.8 Å
  - Atomic = < 1.0 Å
- 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

Data suggest too many atoms here

Data suggest too few atoms here
Ramachandran Plot

\[ \begin{array}{c}
\text{N} & \text{C} & \text{C} \\
\phi & \text{H} & \psi \\
\text{H} & \text{R1} & \\
\end{array} \quad \begin{array}{c}
\text{N} & \text{C} & \text{C} \\
\phi & \text{H} & \psi \\
\text{H} & \text{R2} & \\
\end{array} \]

The Ramachandran Plot

- Beta sheet
- Left handed alpha-helix
- Right handed alpha-helix

\[ \begin{array}{c|c}
\text{phi} & \text{psi} \\
\hline
-180 & +180 \\
0 & 0 \\
-psi & +psi \\
\end{array} \]
Irregular bonding at N1 and NS 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?
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
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 2yf7 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.
Structural characterization of 5-Aryl-4-(5-substituted-2-4-dihydroxyphenyl)-1,2,3-thiadiazole Hsp90 inhibitors.

Source organism: Homo sapiens

Primary publication:

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

Sharp SY, Roe SM, Kuzalskas E, Cilkotiené I, Workman P, Matulis D, Prodromou C

PLoS ONE 7 e44642 (2012)

PMID: 22984537

Function and Biology

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:

- BZ8
- Mg²⁺

No modified residues

Structure analysis

Assembly composition: homo dimer (preferred)

Entry contents: 1 distinct polypeptide molecule

Macromolecule:

- Heat shock protein HSP 90-alpha

Experiments and Validation

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

Quick links

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

Additional links:

- 3D Visualisation
Searching the PDB
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
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...

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 -
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>GO mapping</th>
<th>Journal</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosidase, i.e. enzymes hydrolyzing polysaccharides</td>
<td>GO:0018788: hydrolase activity, acid activity</td>
<td></td>
<td>PEG: dimethyl sulfoxide (2501)</td>
</tr>
<tr>
<td>Hydro-halides</td>
<td>GO:004553: hydrolase activity, acid activity</td>
<td></td>
<td>PEG: 2-(2-hydroxyethyl)ethanol (2501)</td>
</tr>
<tr>
<td>Phosphoric monoester hydrolases</td>
<td>GO:000305: nucleic acid phosphohydrolase</td>
<td></td>
<td>MAN: (2S,3S,4S,5R)-5-(hydroxyethyl)imidazole (2087)</td>
</tr>
<tr>
<td>Carboxylic ester hydrolases</td>
<td>GO:000502: RNA phosphodiesterase</td>
<td></td>
<td>BMA: (2R,3S,4S,5R)-5-(hydroxyethyl)imidazole (2012)</td>
</tr>
<tr>
<td>Carboxylic acid hydrolase (carbonic anhydrase)</td>
<td>GO:0042542: response to hydrophobic stimulus</td>
<td></td>
<td>FAD: [(2R,3S,4R,5R)-5-(6-aminopurin-6-yl)] (1890)</td>
</tr>
<tr>
<td>Beta lactam hydrolase</td>
<td>GO:000501: RNA phosphodiesterase</td>
<td></td>
<td>ADP: (2R,3S,4R,5R)-5-(6-aminopurin-6-yl) (1892)</td>
</tr>
<tr>
<td>Phosphoric diester hydrolases</td>
<td>GO:0042754: hydrogen peroxide catabolism</td>
<td></td>
<td>NAD: [(2R,3S,4R,5R)-5-(6-aminopurin-6-yl)] (1239)</td>
</tr>
<tr>
<td>Hydrolyzing N-glycosyl compounds</td>
<td>GO:0017078: hydrolase activity</td>
<td></td>
<td>NAP: [(2R,3S,4R,5R)-5-(3-aminocyclopentane-1-carboxamide) (1181)</td>
</tr>
<tr>
<td>3.4.19.12: Ubiquitinyl hydrolase 1</td>
<td>GO:0052089: carboxylic ester hydrolase</td>
<td></td>
<td>TRS: 2-AMINO-2-HYDROXYMETHYLCYCLOPENTANE (1155)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecule name</th>
<th>Organism</th>
<th>Sequence family</th>
<th>Structure domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soluble epoxide hydrolase</td>
<td>Hydrozoa</td>
<td>IPR027417: P-loop containing nucleotide-binding domain</td>
<td>Methane Monoxygenase Hydroxylase</td>
</tr>
<tr>
<td>Bifunctional epoxide hydrolase 2</td>
<td>Arthrobacter Hydrocarboxyglutamicus</td>
<td>IPR017853: Glycoside hydrolase superfamily</td>
<td>P-loop containing nucleotide triphosphate</td>
</tr>
<tr>
<td>Cytosolic epoxide hydrolase 2</td>
<td>Aeromonas hydrophila</td>
<td>CL0058: Glyco_hydrolase</td>
<td>alpha/beta-Hydrolases</td>
</tr>
<tr>
<td>Pyrophosphate phospho-hydrolase</td>
<td>Carboxythermus hydrogenoformans</td>
<td>IPR029058: Alpha/Beta hydrolase family</td>
<td>Phosphatase:hydrolase-like</td>
</tr>
<tr>
<td>4-hydroxy-4-oxo-5-carboxypicolinate synthase</td>
<td>Aeromonas Hydrophila</td>
<td>CL0028: AB_hydrolase</td>
<td>Glycosyl hydrolase domain</td>
</tr>
<tr>
<td>Organism</td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrozoa</td>
<td>365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthrobacter Hydrocarboglutamicus</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromonas hydrophila</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carboxydothermus hydrogenoform...</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromonas Hydrophila</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus Hydrophilus Fuscus</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterium Hydrophilum</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteus Hydrophilus</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudomonas Hydrophila</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromonas Hydrophila Proteolytica</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More...
<table>
<thead>
<tr>
<th>Enzyme Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrolases</td>
<td>21099</td>
</tr>
<tr>
<td>Glycosidases, i.e. enzymes hydrolyzing sugars</td>
<td>3634</td>
</tr>
<tr>
<td>Hydro-lyases</td>
<td>1683</td>
</tr>
<tr>
<td>Phosphoric monoester hydrolases</td>
<td>1543</td>
</tr>
<tr>
<td>Carboxylic ester hydrolases</td>
<td>1038</td>
</tr>
<tr>
<td>Carbonic acid hydro-lyase (carbonic acid dioxygenase)</td>
<td>845</td>
</tr>
<tr>
<td>Beta-lactam hydrolase</td>
<td>537</td>
</tr>
<tr>
<td>Phosphoric diester hydrolases</td>
<td>499</td>
</tr>
<tr>
<td>Hydrolyzing N-glycosyl compounds</td>
<td>465</td>
</tr>
<tr>
<td>3.4.19.12 : Ubiquitinyl hydrolase 1</td>
<td>450</td>
</tr>
</tbody>
</table>

More...
Query – “Enzyme: Hydrolases”
Drill down with facets

- **Organism superkingdom (4)**
  - **Organism name (921)**
    - Homo sapiens (6129)
    - Bos taurus (1115)
    - Escherichia coli (782)
    - Gallus gallus (760)
    - Escherichia virus T4 (701)
    - Human immunodeficiency virus 1 (539)
Drill down with facets
Query – “Enzyme: Hydrolases”
Introducing PDBe-KB

pdbe-kb.org
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

Molecular & Cellular Structures
- Protein Data Bank in Europe
- Electron Microscopy Data Bank

Chemical Biology
- ChEMBL
- MetaboLights
- ChEBI

Molecular Systems
- BioModels
- Reactome
- IntAct
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]

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

**PDB chain shown:** 3k05 B

Uniprot residues 1891 - 2089

Coverage: 10%
Protein pages - summary

- Icons highlight structural information related to specific protein

PDBe-KB > Mediator of DNA damage checkpoint protein 1

Gene: MDC1
Organism: Homo sapiens (Human)
Synonyms: KIAA0170, NFBD1
Uniprot: Q14676 [go to UniProt]

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

PDBe-KB › Mediator of DNA damage checkpoint protein 1
Gene: MDC1
Organism: Homo sapiens (Human)
Synonyms: KIAA0170, NFBD1
Uniprot: Q14676

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]

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: ☑

- COH: co-factor
  - 3D view
  - Found in 7 entries

- BOG
  - 3D view
  - Found in 5 entries

- MAN
  - 3D view
  - Found in 7 entries

- NAG
  - 3D view
  - Found in 7 entries

- RCX
  - 3D view
  - Found in 1 entry

- FLF
  - 3D view
  - Found in 1 entry
Protein pages - ligands

Overview of ligands and binding residues in the PDB

Show all molecules from PDB entries containing this protein: 

- COH: co-factor
  - Found in 7 entries

- MAN: Found in 7 entries

- NAG: Found in 7 entries

- Cofactors highlighted preferentially where present
Protein pages - ligands

Overview of ligands and binding residues in the PDB

- COH
  - co-factor
  - Found in 7 entries
- MAN
  - Found in 7 entries
- NAG
  - Found in 7 entries

- Other bound ligands listed by number of entries present
Protein pages - ligands

- ProtVista shows location of all ligand binding sites
  - Begin to see trend of important binding residues
Overview of interaction partners and interacting residues in the PDB
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.

- 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

<table>
<thead>
<tr>
<th>Predicted functional sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATH-FunSites</td>
</tr>
<tr>
<td>Predicted PTM sites</td>
</tr>
<tr>
<td>Predicted ligand binding sites</td>
</tr>
<tr>
<td>Domains</td>
</tr>
<tr>
<td>Ligand binding sites</td>
</tr>
<tr>
<td>Interaction interfaces</td>
</tr>
<tr>
<td>Variants</td>
</tr>
</tbody>
</table>
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

Crystal structure of Turkey (Meleagris gallopavo) hemoglobin at 2.3 Angstrom

Ramesh P, Sundaresan SS, Ponnuwamy MN

To be published

Source organism: Meleagris gallopavo

Assembly composition: protein/protein complex

Bound ligands: HEM

X-ray diffraction
2.3A resolution
Released: 22 Dec 2009
Model geometry
Fit model/data

3D Visualisation  Download files