Recent updates to the PDBe Knowledge Base

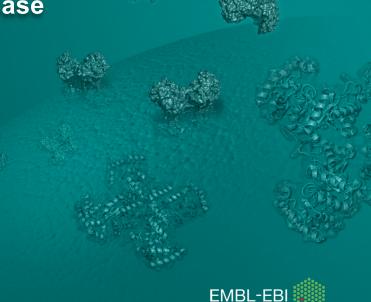
(PDBe-KB)

Protein Data Bank in Europe - Knowledge Base

Lukáš Pravda Bioinformatician

CCP4 Virtual Study Weekend 7th January 2021





Overview

- Brief introduction to PDBe-KB
- PDBe-KB aggregated views of proteins
- Major new features in 2020
 - Covid-19 data portal (https://www.ebi.ac.uk/pdbe/covid-19)
 - Structures superposition
 - Batch download
 - "Processed proteins"
 - Similar proteins

A macromolecular structure is but one piece of the puzzle

"Coordinates by themselves just specify shape and are not necessarily of intrinsic biological value, unless they can be related to other information"

Integrative database analysis in structural genomics, Mark Gerstein, Nature Structural Biology 7, 960, 2000

Protein Data Bank in Europe Knowledge Base (https://pdbe-kb.org)

Placing macromolecular structure data in their biological context by establishing a community-driven, integrated resource for structural annotations to promote basic and applied research

- Create data standards
- Create data access mechanisms
- Reduce fragmentation

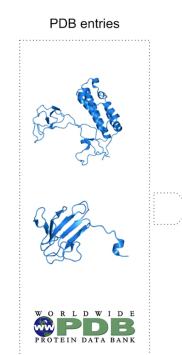
PDBe-KB: a community-driven resource for structural and functional annotations, Nucl.Acids Res. 2019 Oct https://doi.org/10.1093/nar/gkz853

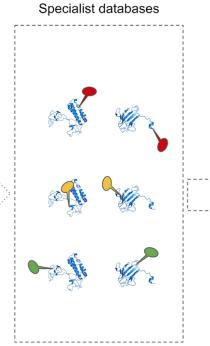




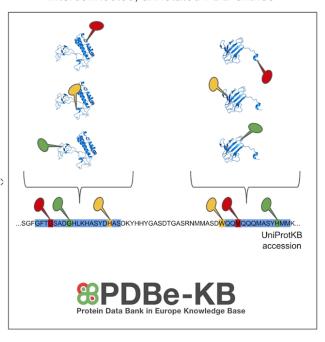
8PDBe-KB

Integrating annotations from specialist data resources





Interconnected, annotated PDB entries



PDBe-KB integrates data on:

- Domain annotations
- Rfam classifications
- Ligand-binding sites
- Macromolecular interaction interfaces
- Variant annotations
- Sequence conservation
- Post-translational modifications
- Backbone flexibility and intrinsic disorder
- Residue accessibility and depth





























Rfam









EC-PDE



























14-3-3-Pred











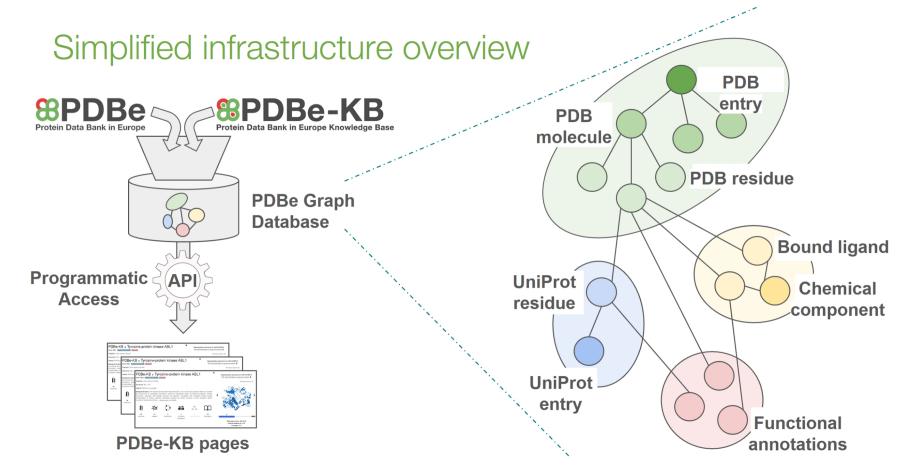


3DLigandSite **ProKinO**











Aggregated view of proteins





Accessing the aggregated views of proteins

The protein-centric PDBe-KB pages can be accessed by:

- 1. From the PDBe-KB protein aggregated views landing page https://pdbekb.org/proteins/
- 2. Using a PDB or UniProt identifier
 - e.g. https://pdbekb.org/protein/2etx
 - e.g. https://pdbekb.org/protein/Q92793
- 3. From PDBe entry pages / search
- 4. UniProt entry pages





Live demo

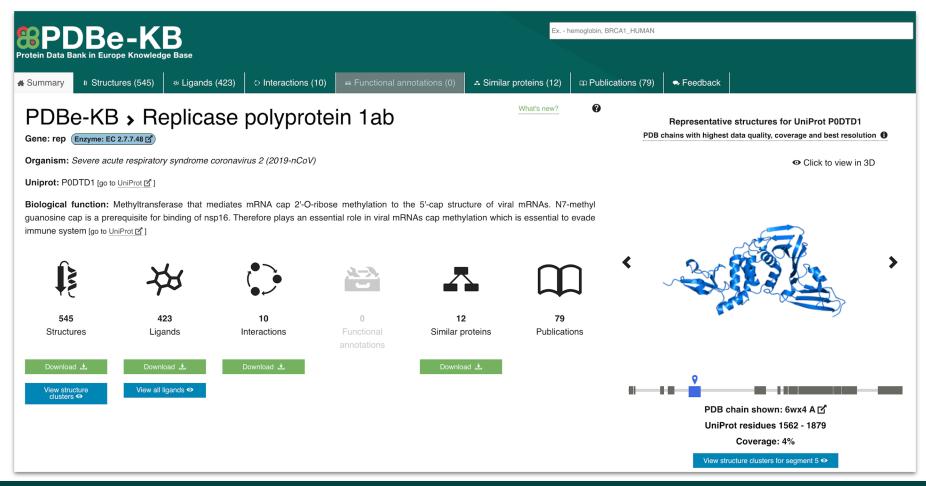




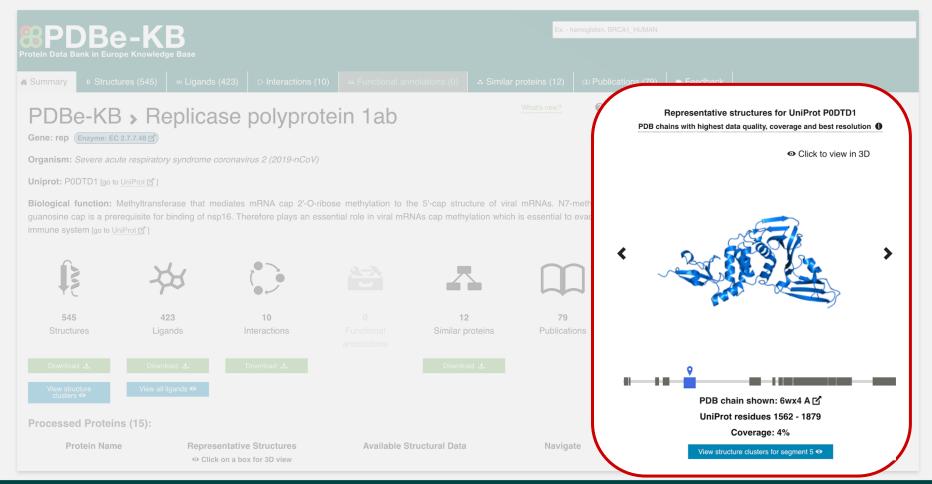
Summary







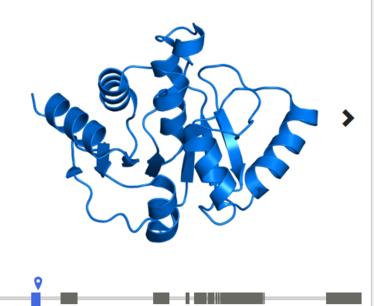




Representative structures for UniProt P0DTD1

PDB chains with highest coverage and resolution 1

Click to view in 3D



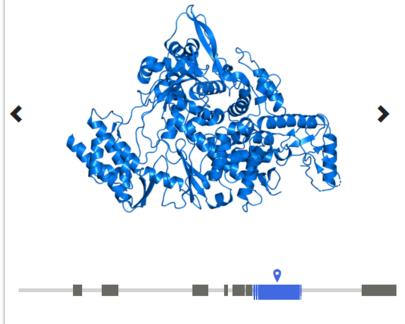
PDB chain shown: 6wen A ☐ UniProt residues 1024 - 1192

Coverage: 2%

Representative structures for UniProt P0DTD1

PDB chains with highest coverage and resolution 19

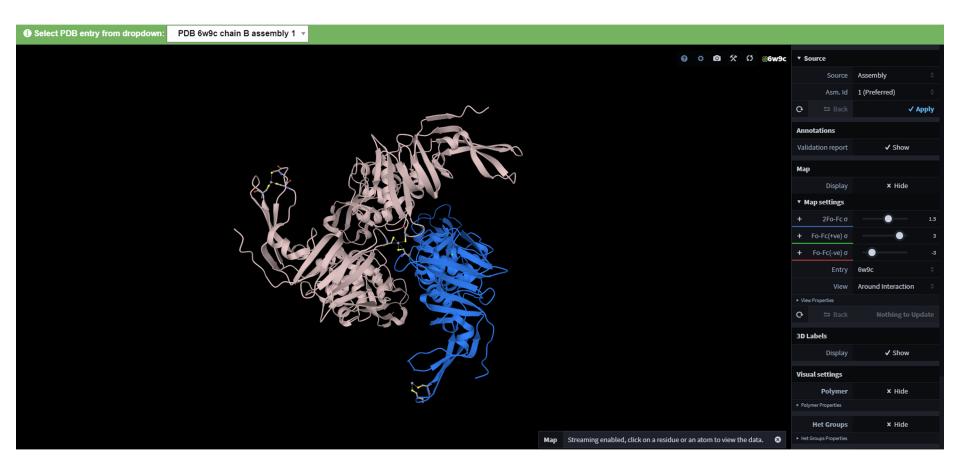
Click to view in 3D



PDB chain shown: 6m71 A ☑ UniProt residues 4393 - 5324
Coverage: 12%









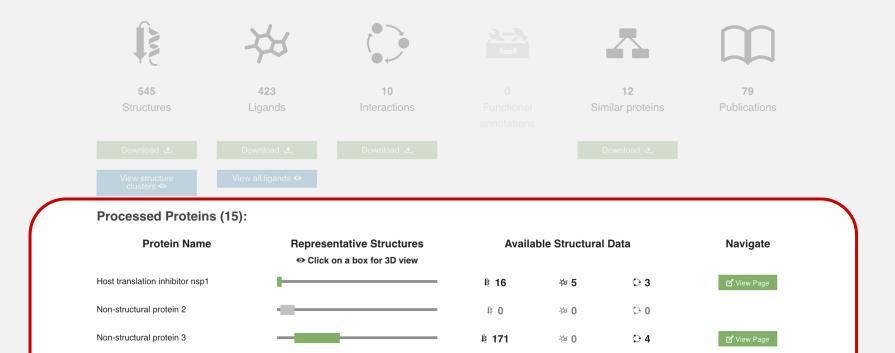




Processed proteins







Non-structural protein 4

+ Show more items



₿ 1

 \bigcirc 0

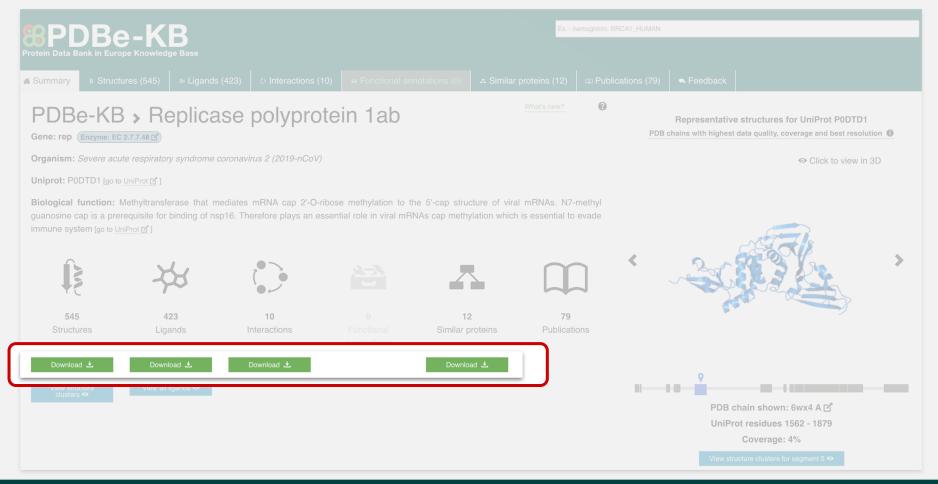
₩ 0



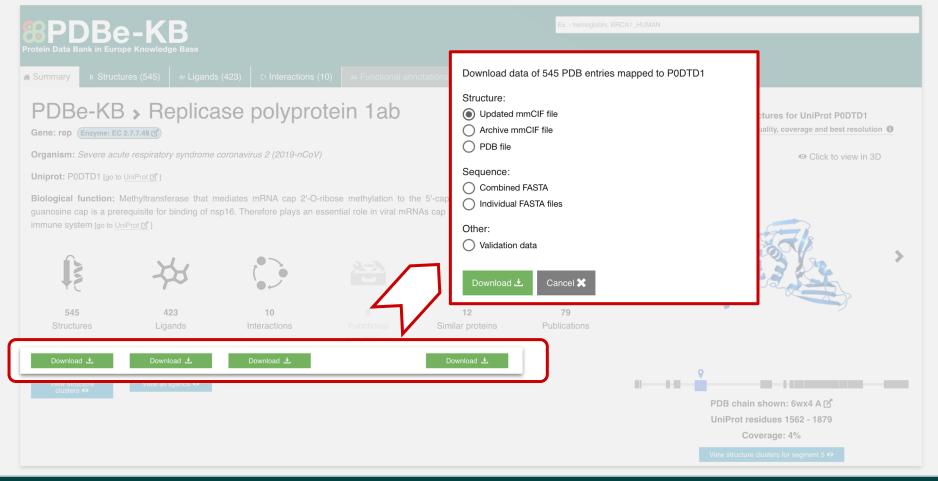
Structure download















PDBe File Download service 12 OAS3

/pdbe/download/openapi.json

PDBe download service API entry point

This API allows for a batch download of arbitrary subsets of the PDB data.

Servers		
/pdbe/download	~	

Macromolecular data



POST /pdb/entry/updated Updated mmCIF files download

GET /pdb/entry/archive Archive coordinate files download





Structures superposition

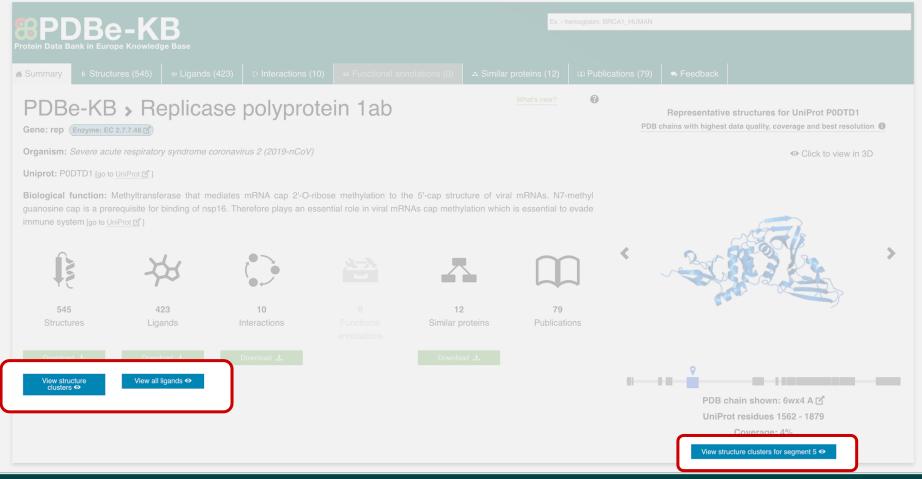




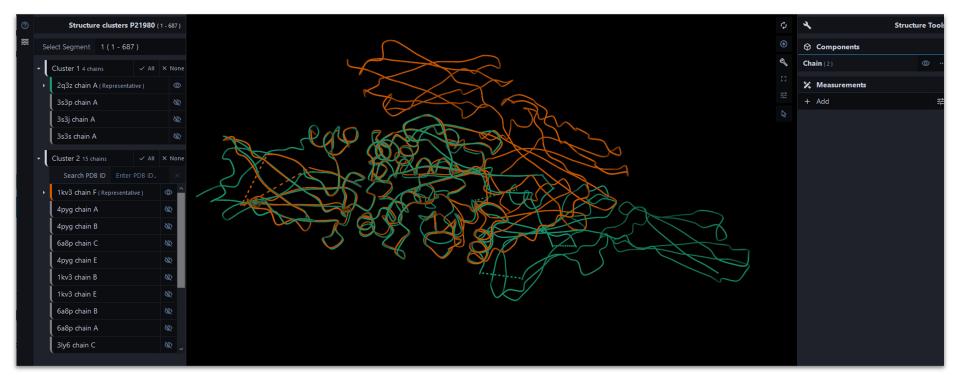
Superposition

- We have implemented a new process that calculates superposition matrices for individual proteins in the whole PDB archive
- The process uses GESAMT to align PDB chains mapped to UniProt segments
- Clusters are generated based on the alignments
- The matrix files and the superposition clusters are accessible via API
- The clusters do not necessarily correspond to biological functions.





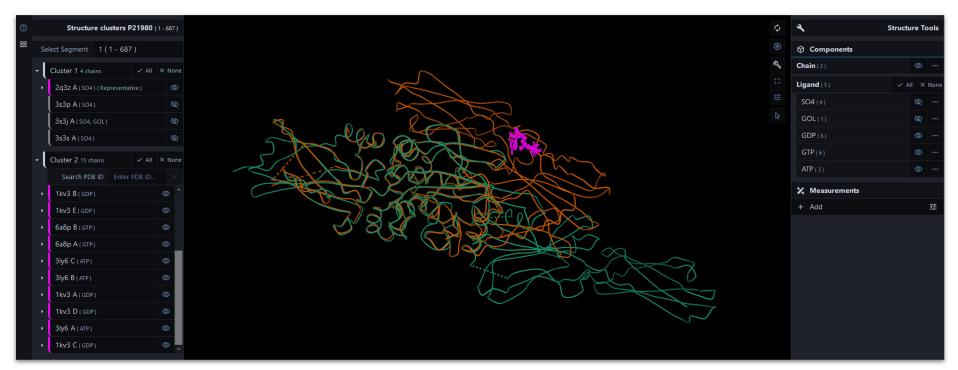
Protein-glutamine gamma-glutamyltransferase 2





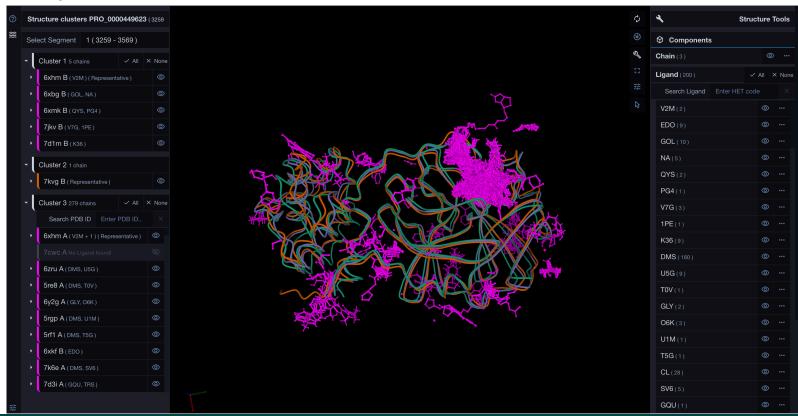


Protein-glutamine gamma-glutamyltransferase 2





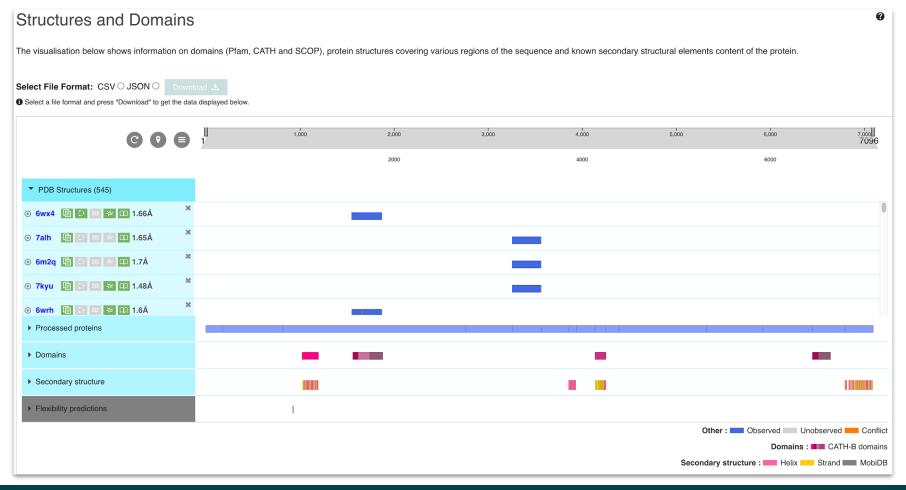
3C-like proteinase



Structures and domains









Ligands



All Ligands (430)

This section, by default, shows ligands observed directly bound to the protein of interest, if such ligands are available. Click on the checkbox 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 section below.

Q Search: e.g. S-ADENOSYLMETHIC

1 Search by molecule name, code or PDB id.



SAM 6

cofactor-like

Q Found in 13 entries

Q Found in 1 entry



S8B 6

 \pm



SAH 6

cofactor-like

Q Found in 6 entries



• 3D view

3D view

SFG 6

cofactor-like

Q Found in 2 entries







3D view

UOM 6

Q Found in 1 entry



X1Y 🚯

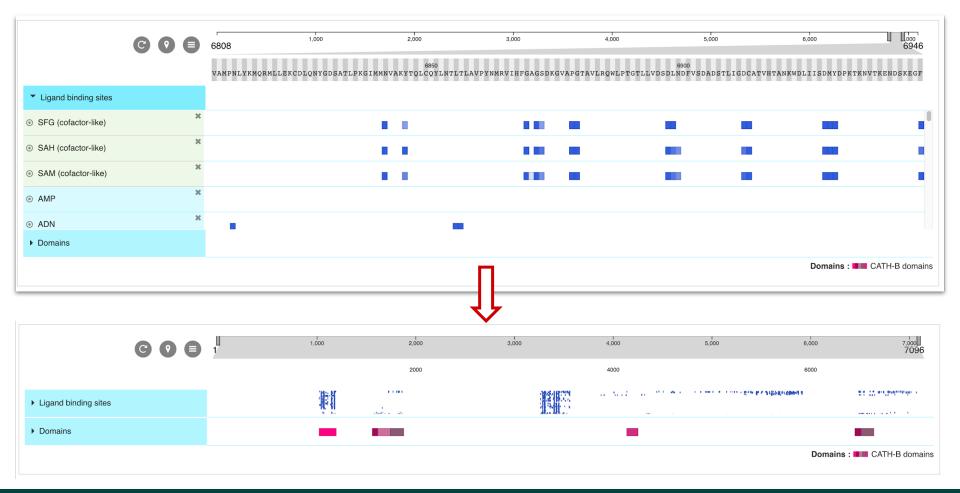


Q Found in 1 entry



+ Show all items

3D view

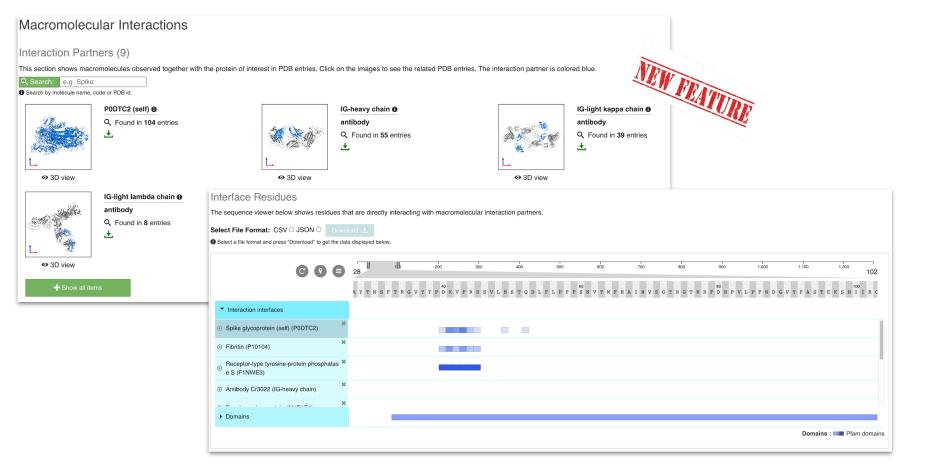




Macromolecules







Additional annotations



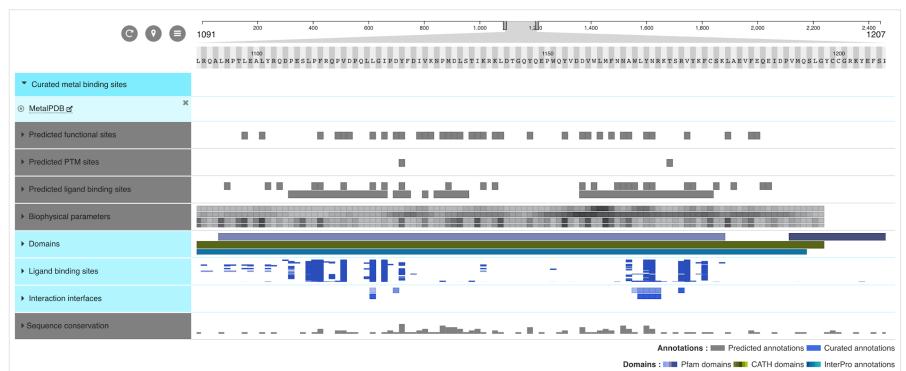
Additional Functional Annotations

Additional residue-level functional annotations available for the PDB entries related to this protein are displayed below. The majority of the annotations are contributed by collaborators of PDBe-KB. (Find out more about PDBe-KB partners 🗹).

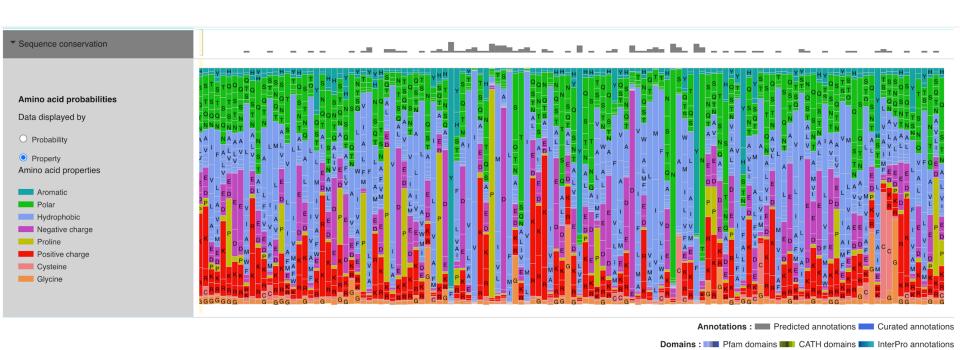
Select File Format: CSV ○ JSON ○

Download 🕹

1 Select a file format and press "Download" to get the data displayed below.













Similar proteins





Similar Proteins (2128)

Proteins with 90% or greater sequence identity to P0DTC2

• Search: e.g. Severe acute respiratc

• Search by species, taxonomy id or protein name.

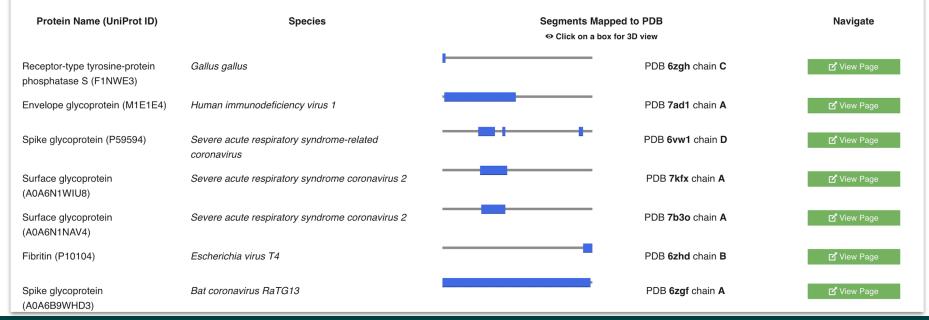
Select File Format: CSV ○ JSON ○

Download 🕹

1 Select a file format and press "Download" to get the data displayed below.

Similar Proteins with highly identical PDB sequences to P0DTC2 (20)

PDB entries with more than 90% sequence identity to the protein of interest.



Publications





Publications

a

Primary PDB publications, reviews associated with PDB entries and UniProt publications are listed below. Click on the plus signs to expand the publication lists.

Q Search: E.g. Mutation

1 Search by keyword in the title or by PDB id or PubMed id.

PDB publications (77)

Go to page:

Select File Format: CSV ○ BibTeX ○

Select a file format and press "Download" to get the data displayed below.

D614G mutation alters SARS-CoV-2 spike conformational dynamics and protease cleavage susceptibility at the S1/S2 junction.

Gobeil S. Janowska K. McDowell S. Mansouri K. Parks R et al.

bioRxiv (2020)

PMID: 33052347 F*

doi> 10.1101/2020.10.11.335299 🗹

Related PDB entries: 7kdl 7kdj 7kdk 7kdh 7kdi 7kea 7kdg 7kec 7keb 7ke9

7ke8 7ke4 7ke7 7ke6

A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2.

Chi X, Yan R, Zhang J, Zhang G, Zhang Y et al.

Science (2020)

PMID: 32571838

doi> 10.1126/science.abc6952 <a>С

Related PDB entries: 7c2l

A pH-dependent switch mediates conformational masking of SARS-CoV-2 spike.

PDB-related reviews (182)

Go to page:







Select File Format: CSV ○ BibTeX ○

• Select a file format and press "Download" to get the data displayed below.

Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks.

PMID: 32143502 14

Related PDB entries: 6vsb

Should COVID-19 Concern Nephrologists? Why and to What Extent? The Emerging Impasse of Angiotensin Blockade.

PMID: 32203970 F*

Related PDB entries: 6m17

One size does not fit all - Patterns of vulnerability and resilience in the COVID-19 pandemic and why heterogeneity of disease matters.

PMID: 32205119 1

Related PDB entries: 6vsb

Elevated Plasmin(ogen) as a Common Risk Factor for COVID-19 Susceptibility.

PMID: 32216698

Related PDB entries: 6vxx 6vsb 6vyb

UniProt publications (20)

Go to page:







Select File Format: CSV ○ BibTeX ○

1 Select a file format and press "Download" to get the data displayed below.

A new coronavirus associated with human respiratory disease in China.

Wu F., Zhao S., Yu B., Chen Y.M., Wang W. et al.

Nature 265-269 (2020)

PMID: 32015508

SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor.

Hoffmann M., Kleine-Weber H., Schroeder S., Krueger N., Herrler T. et al. Cell 1-10 (2020)

PMID: 32142651

Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2.

Yan R., Zhang Y., Li Y., Xia L., Guo Y. et al.

Science - (2020)

PMID: 32132184

A multibasic cleavage site in the Spike protein of SARS-CoV-2 is essential for infection of human lung cells





Data availability

- Aggregated views: https://pdbekb.org/proteins
- Graph database: https://pdbekb.org/graph-download
- REST API: http://pdbe.org/aggregated-api
- Download service API: https://www.ebi.ac.uk/pdbe/download/docs
- Visualization components: https://www.ebi.ac.uk/p1dbe/pdb-component-library
- API webinars: https://bit.ly/PDBe_API_webinars



Protein Data Bank in Europe - Knowledge Base

COVID-19 Data Portal

PDBe.org/covid19



PDBe-KB COVID-19 Data Portal

An unprecedented number of scientific efforts are taking place worldwide in order to help combat the new coronavirus epidemic (COVID-19). One of the biggest challenges in this fast-moving situation is to share data and findings in a coordinated way, in order to understand the disease and to develop treatments and vaccines.

To support research efforts to understand more about the SARS-CoV-2 virus and the structures of its proteins, we have created dedicated PDBe-KB pages to highlight important structural features from released PDB entries. These pages include all **observed ligand binding sites and protein-protein interaction residues**, to help researchers easily identify important structural features to support the development of treatments and vaccines.

The pages available are listed below, with links to the relevant PDBe-KB protein pages.

PDBe-KB page **Data Summary** PODTD1 - Replicase polyprotein 1ab The orf1ab polyprotein is a multifunctional protein involved in the transcription and replication of viral RNAs. It contains the proteinases 545 423 10 12 responsible for the cleavages of the polyprotein. Structures Ligands Interactions Functional Similar Annotations Proteins PRO_0000449619 - Host translation Inhibitor nsp1 (nsp1) Inhibits host translation by interacting with the 40S ribosomal subunit. The nsp1-40S ribosome complex further induces an endonucleolytic cleavage near the 5'UTR of host mRNAs, targeting them for Interactions Functional Similar Structures Ligands degradation. Viral mRNAs are not susceptible to nsp1-mediated **Annotations** Proteins endonucleolytic RNA cleavage thanks to the presence of a 5'-end leader sequence and are therefore protected from degradation. By suppressing host gene expression, nsp1 facilitates efficient viral gene expression in infected cells and evasion from host immune response.

Please provide your feedback!





Questions and discussion





pdbekb_help@ebi.ac.uk



f proteindatabank



@PDBeurope pdbeurope





pdbart

Funding















