
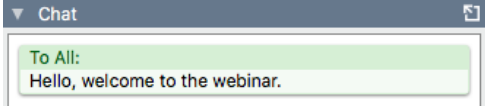


Welcome - webinar instructions

- GoToTraining works best in **Chrome** or on Linux, **Firefox**
- To access the full features of GoToTraining, use the desktop version by clicking “**switch to desktop version**”
- All **microphones will be muted** whilst the trainer is speaking 
- If you have a question please use the **chat box** at the bottom of the GoToTraining box 
- Please complete the **feedback survey** which will launch at the end of the webinar.
- This presentation is available in the materials section

PDBe API webinar series:

2) Searching with the PDBe API

[PDBe.org/API](https://pdbe.org/API)



John Berrisford



pdhelp@ebi.ac.uk



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



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



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



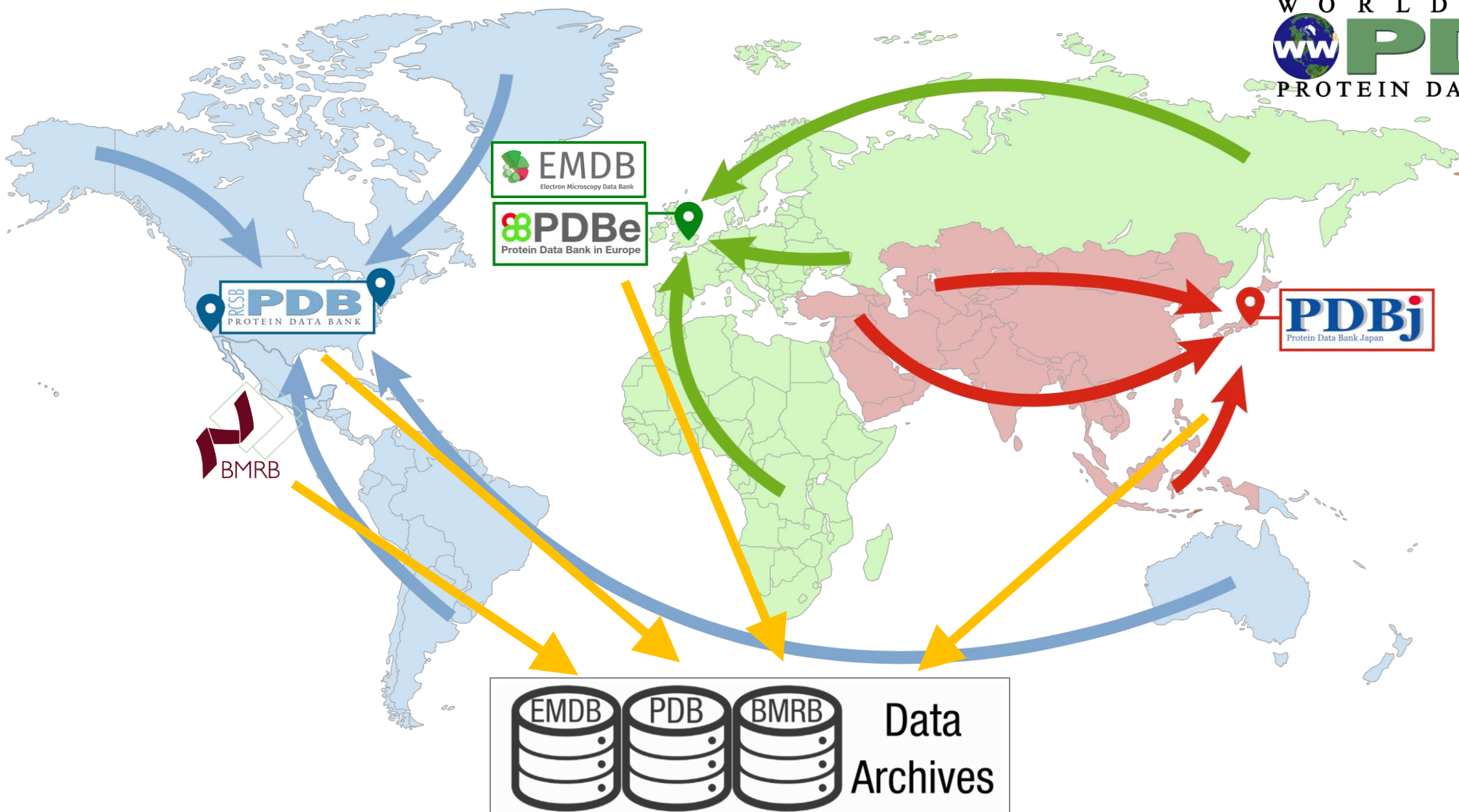
[pdbart](https://www.pinterest.com/pdbart)


PDBe API webinar series

- Sep 15th Introduction to PDBe programmatic access
- Sep 22nd **Searching with the PDBe API**
- Sep 29th Creating complex PDBe API queries
- Oct 6th Using the PDBe graph API
- Oct 13th PDBe tools in github
- Oct 20th Data visualisation at PDBe

Summary of the webinar

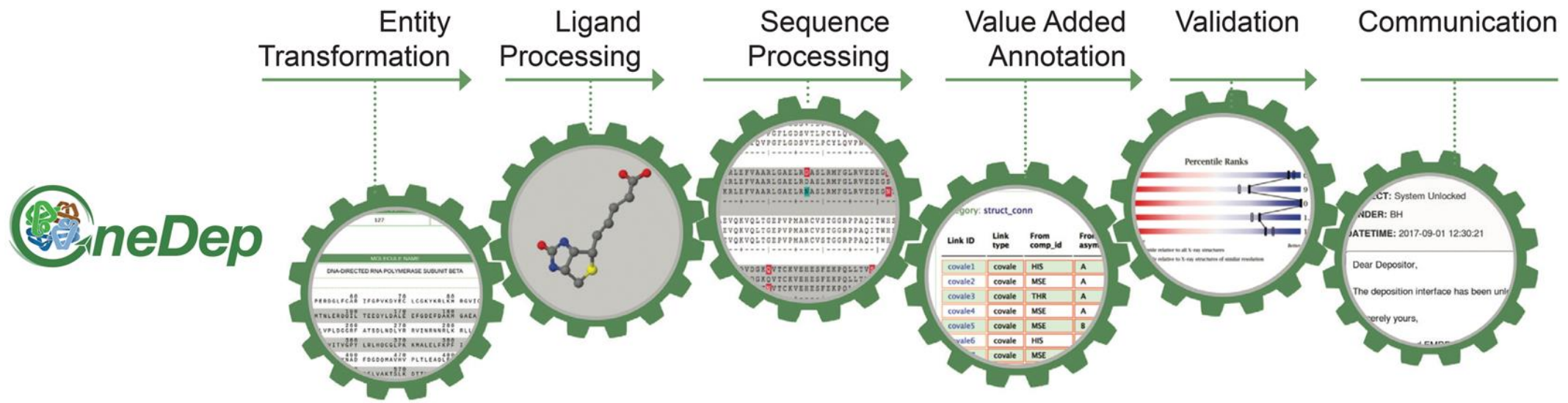
- PDBe in the wwPDB
- PDBe search
- Advanced search
- Introduction to documentation page
- Usage examples



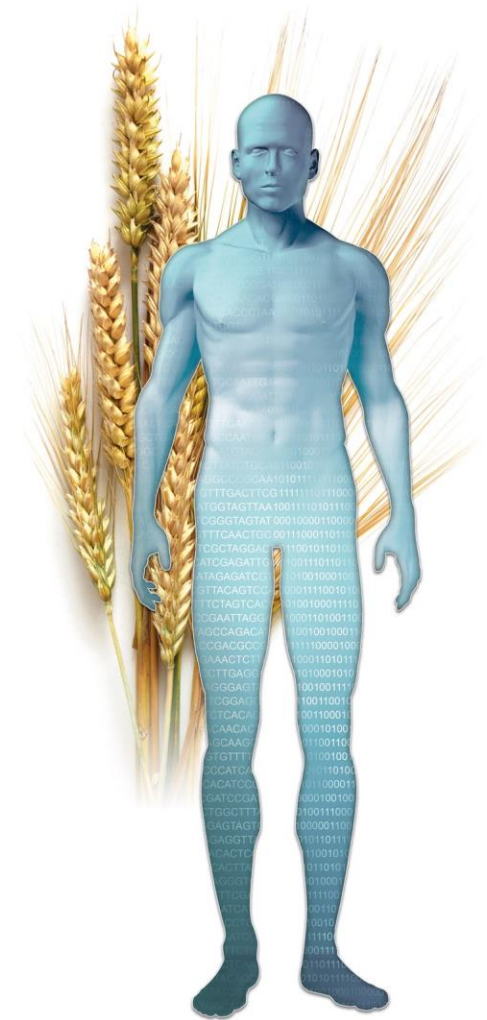
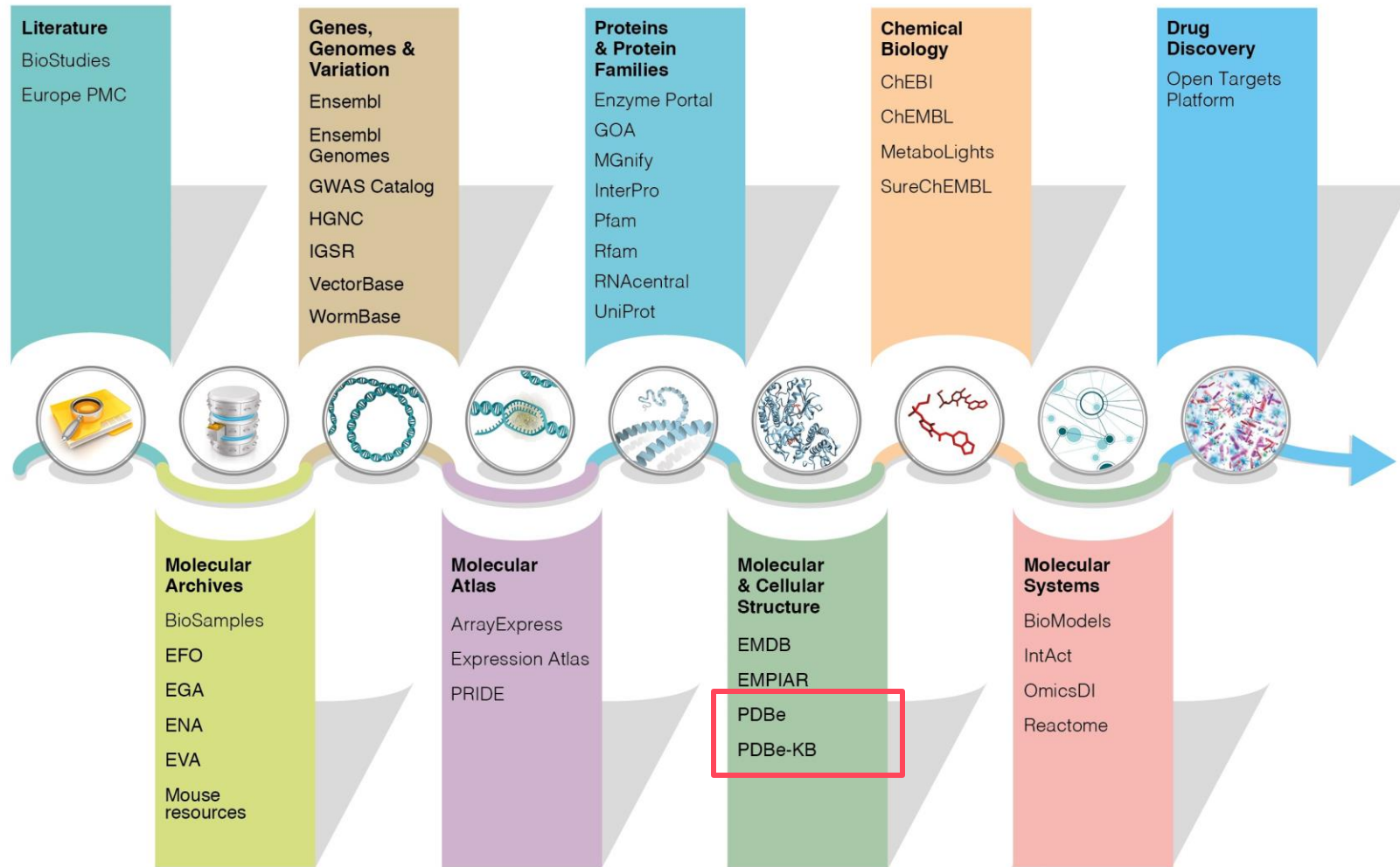
 Data Archives

Data added through deposition

- Provided by depositor or added during curation
- Data added to define experimental methods, sequence information, connectivity, validation and more...
- Added to the archive mmCIF file or validation XML

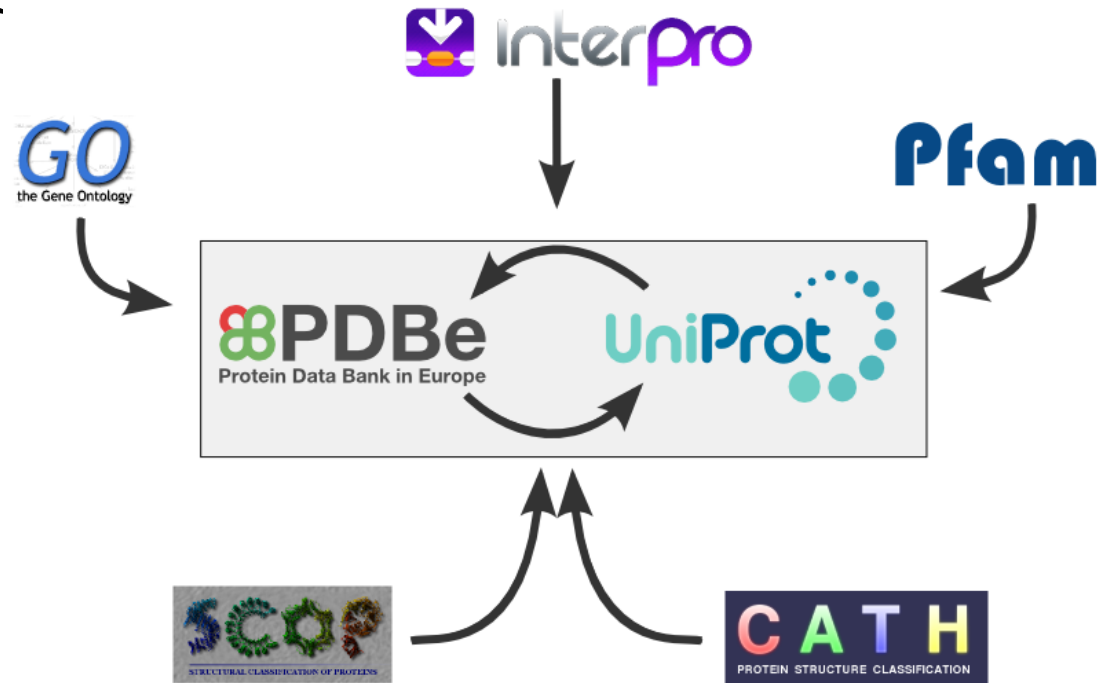


At the heart of EMBL-EBI resources



PDBe – add additional data

- At PDBe we integrate data from other data resources, including from within EBI
- An example
 - SIFTS - Structure Integration with Function, Taxonomy, and Sequence



Why search the PDB?

- Provides a convenient way of finding entries in the PDB



Entries 1 to 3 of 3 | Sort by | 10 /page

Select all entries on this page

6nbc human methemoglobin state 1 determined using single-particle cryo-EM at 200 keV

Released: 13 Mar 2019

Model geometry: Data not analysed

Herzik Jr MA, Wu M, Lander GC
Nat Commun (2019) [PMID: 30833564]

Source organism: *Homo sapiens*

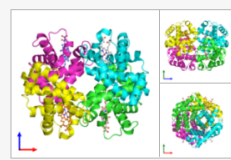
Assembly composition: protein/protein complex

Bound ligands: HEM

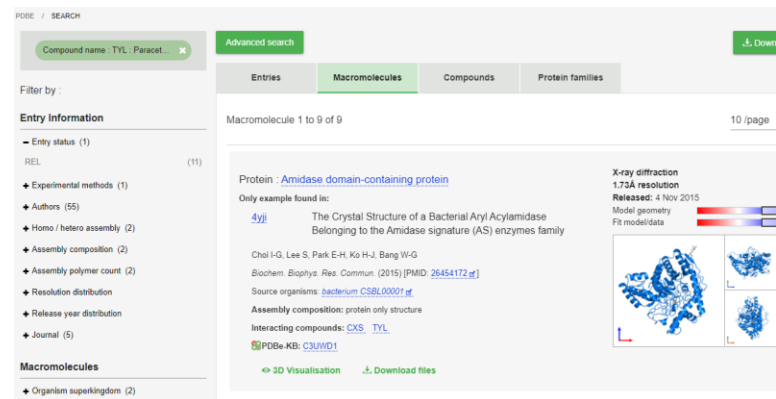
PDBx-KB: P69905 P68871

3D Visualisation | Download files

Electron Microscopy
2.8Å resolution



Hemoglobin solved by Electron Microscopy



PDBE / SEARCH | Compound name: TYL Paracetamol | Advanced search | Download

Filter by: Entries | **Macromolecules** | Compounds | Protein families

Macromolecule 1 to 9 of 9 | 10 /page

Entry Information

- Entry status (1)
- REL (11)
- Experimental methods (1)
- Authors (55)
- Homo / hetero assembly (2)
- Assembly composition (2)
- Assembly polymer count (2)
- Resolution distribution
- Release year distribution
- Journal (5)

Macromolecules

- Organism superkingdom (2)

Protein: **Amidase domain-containing protein**

Only example found in: **4yji** The Crystal Structure of a Bacterial Aryl Acylamidase Belonging to the Amidase signature (AS) enzymes family

Choi I-G, Lee S, Park E-H, Ko H-J, Bang W-G
Biochem Biophys Res Commun (2015) [PMID: 26454172]

Source organisms: *Bacterium CSBL00001*

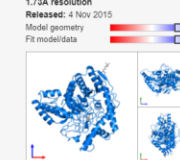
Assembly composition: protein only structure

Interacting compounds: CXS TYL

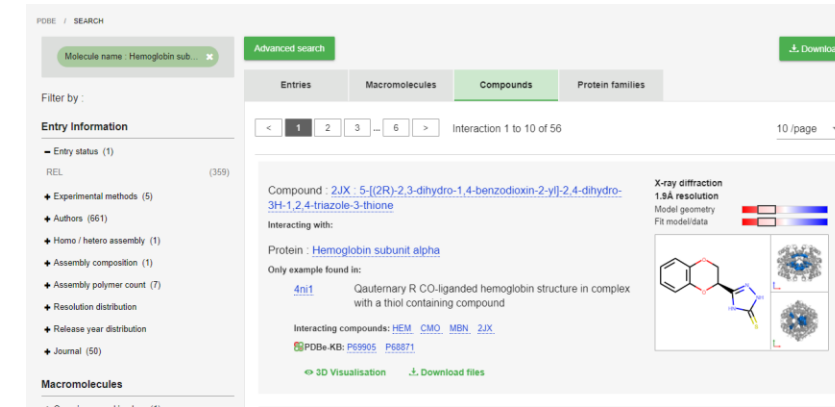
PDBx-KB: C3UWD1

3D Visualisation | Download files

X-ray diffraction
1.73Å resolution
Released: 4 Nov 2015



Which proteins bind to paracetamol



PDBE / SEARCH | Molecule name: Hemoglobin sub... | Advanced search | Download

Filter by: Entries | Macromolecules | **Compounds** | Protein families

Interaction 1 to 10 of 56 | 10 /page

Entry Information

- Entry status (1)
- REL (359)
- Experimental methods (5)
- Authors (561)
- Homo / hetero assembly (1)
- Assembly composition (1)
- Assembly polymer count (7)
- Resolution distribution
- Release year distribution
- Journal (50)

Macromolecules

- Organism superkingdom (1)

Compound: **2JX : 5-[(2R)-2,3-dihydro-1,4-benzodioxin-2-yl]-2,4-dihydro-3H-1,2,4-triazole-3-thione**

Interacting with: **Hemoglobin subunit alpha**

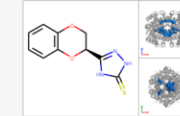
Only example found in: **4ni1** Quaternary R CO-liganded hemoglobin structure in complex with a thiol containing compound

Interacting compounds: HEM CMO MBN 2JX

PDBx-KB: P69905 P68871

3D Visualisation | Download files

X-ray diffraction
1.8Å resolution



Which ligands bind to hemoglobin

PDBe search - introduction

- Search is available directly from PDBe's home page – pdbe.org

The screenshot shows the PDBe.org website home page. At the top left is the EMBL-EBI logo. The main header features the PDBe logo and the text "Protein Data Bank in Europe" with the tagline "Bringing Structure to Biology". A search bar is prominently displayed on the right side of the header, highlighted with a red box. Below the search bar are example search terms: "Examples: hemoglobin, BRCA1_HUMAN" and a link to "Advanced search". The navigation menu includes "PDBe home", "Deposition", "PDBe services", "PDBe training", "Documentation", "About PDBe", and "COVID-19". On the right side of the navigation menu are "Share" and "Feedback" options. Below the navigation menu, there is a section for "New PDBe-KB COVID-19 Data Portal" with a small image and text describing the portal. To the right of this section is a "Popular" section with links to "PDBe-KB", "EMsearch", "News", and "Events".

PDBe's search - autocomplete

- PDBe search features an auto complete which provides results in categories

Protein Data Bank in Europe

Author	Enzyme	GO mapping	Gene
Hemmings AM (72)	With a heme protein as acceptor (168)	GO:0020037 : heme binding (4976)	hemB (23)
Hemsworth GR (64)	1.6.2.4 : NADPH-hemoprotein reductase (151)	GO:0005833 : hemoglobin complex (461)	hemL (20)
Hemmi H (38)	NADPH-hemoprotein reductase (151)	GO:0031838 : haptoglobin-hemoglobin ... (338)	hemH (18)
Hemmig R (19)	NADPH:hemoprotein oxidoreductase (151)	GO:0031721 : hemoglobin alpha binding (332)	hemF (17)
Hemrika W (19)	1.14.14.18 : Heme oxygenase (biliverdi... (85)	GO:0030492 : hemoglobin binding (293)	HEM2 (14)
Christensen HEM (12)	Heme oxygenase (biliverdin-producing) (85)	GO:0044179 : hemolysis in other organ... (179)	hemC (11)
Hemann MT (9)	1,8-cineole,[reduced NADPH-hemopro... (64)	GO:0007599 : hemostasis (162)	HEMK2 (7)
Hemkens MD (9)	Albendazole,[reduced NADPH-hemopr... (64)	GO:0006783 : heme biosynthetic process (159)	hemE (7)
Hemmings BA (9)	Quinine,[reduced NADPH-hemoprotei... (64)	GO:1902033 : regulation of hematopoi... (134)	hemA (5)
Hemmis CW (7)	Sterol,[reduced NADPH-hemoprotein r... (50)	GO:0004392 : heme oxygenase (decyc... (106)	hemO (5)
More...	More...	More...	More...

Journal	Ligand	Molecule name	Organism
Am. J. Hematol. (2)	HEM : HEME (4877)	Hemagglutinin (640)	Streptococcus Hemolyticus (146)
	HEM : 3-[(5Z,10Z,14Z,19Z)-18-(2-carb... (4877)	Hemagglutinin HA1 chain (460)	Hemiptera (138)
	HEM : [3,3'-(7,12-diethenyl-3,8,13,17-t... (4877)	Hemagglutinin HA2 chain (456)	Hemophilus Pertussis (31)

Summary of the webinar

- PDBe in the wwPDB
- PDBe search
- **Advanced search**
- Introduction to documentation page
- Usage examples

PDBe's search – advanced search

- PDBe's search provides an advanced search feature to build complex queries

The image shows a screenshot of the Protein Data Bank in Europe (PDBe) website. The main header features the EMBL-EBI logo and navigation links for Services, Research, Training, and About us. The main content area displays the PDBe logo and the text "Protein Data Bank in Europe" with the tagline "Bringing Structure to Biology". A search bar is visible with a "Search" button and a link to "Advanced search". Below the search bar, there are links for "Share" and "Feedback". A green-bordered inset box provides a magnified view of the search bar area, highlighting the "Advanced search" link with a red border.

EMBL-EBI

Services Research Training About us

Protein Data Bank in Europe
Bringing Structure to Biology

Search

Examples: [hemoglobin](#), [BRCA1_HUMAN](#)

[Advanced search](#)

Share Feedback

Services Research Training About us

Search

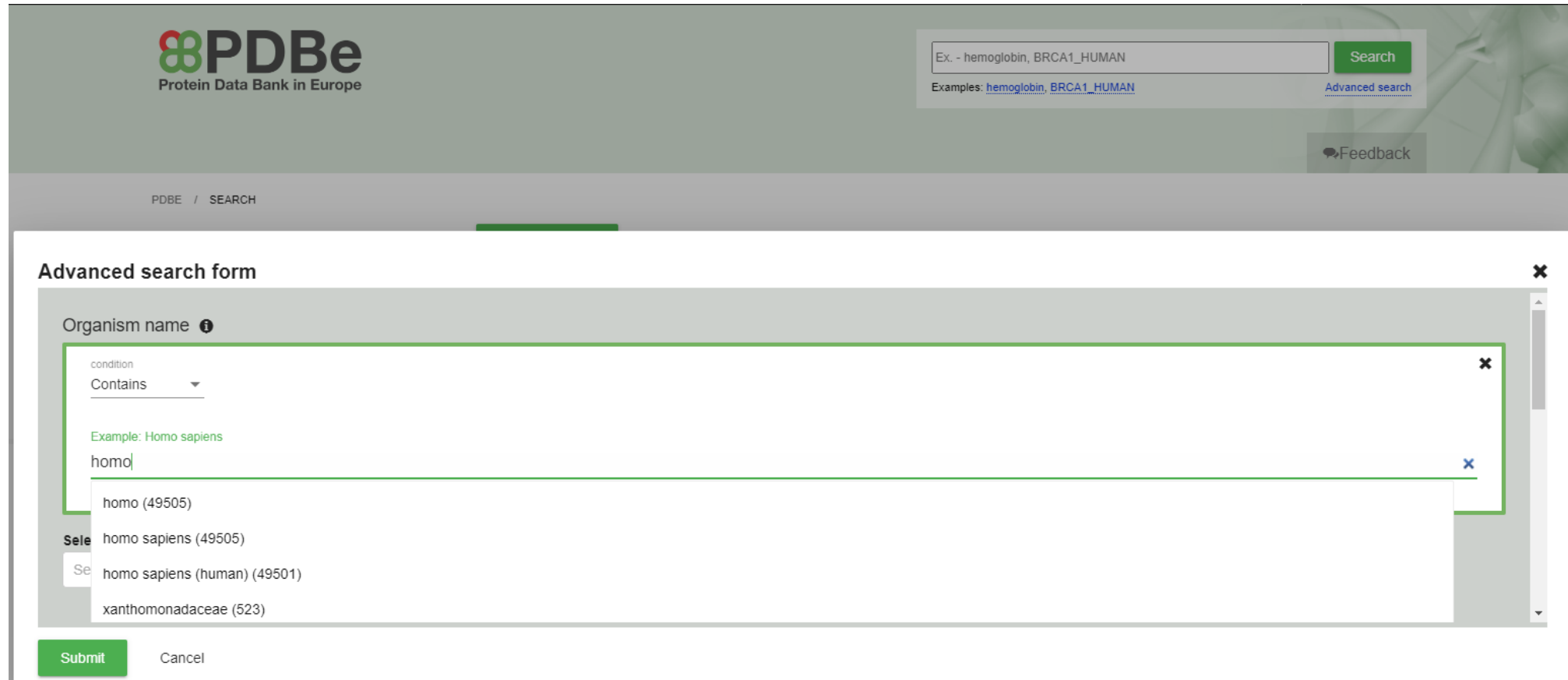
Examples: [hemoglobin](#), [BRCA1_HUMAN](#)

[Advanced search](#)

Share Feedback

PDBe search – advanced search

- Advanced search contains the same auto complete functionality



The screenshot displays the PDBe website's advanced search interface. At the top left is the PDBe logo (Protein Data Bank in Europe). To the right is a search bar with a 'Search' button and a 'Feedback' link. Below the search bar is a breadcrumb trail: 'PDBE / SEARCH'. The main content area is titled 'Advanced search form' and features a search criteria section for 'Organism name'. A dropdown menu is open, showing a list of organisms with their corresponding counts. The text 'homo' is entered in the search field, and the dropdown lists 'homo (49505)', 'homo sapiens (49505)', 'homo sapiens (human) (49501)', and 'xanthomonadaceae (523)'. The 'Submit' button is highlighted in green.

Organism name ⓘ

condition
Contains

Example: Homo sapiens
homo

homo (49505)

homo sapiens (49505)

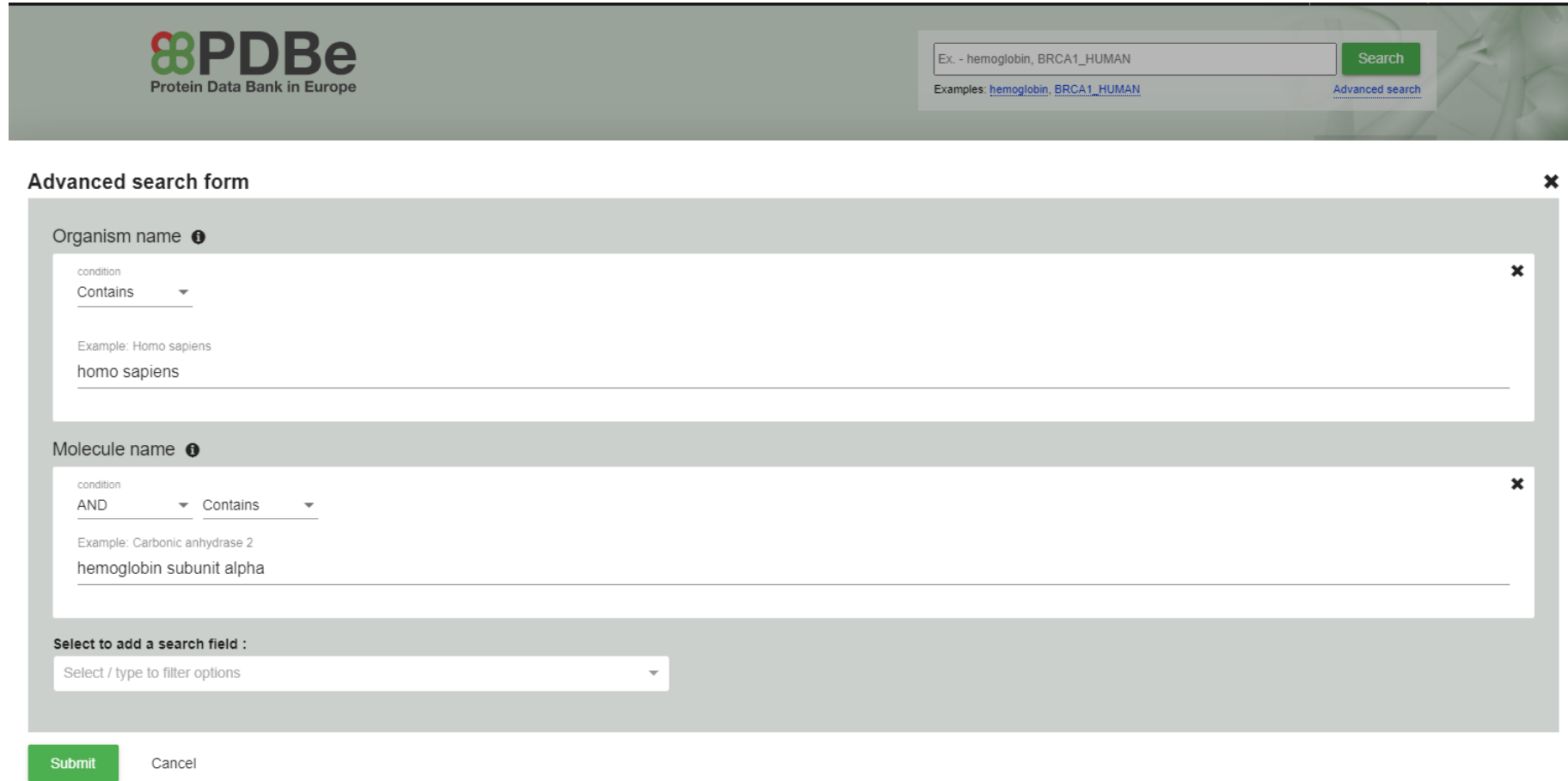
homo sapiens (human) (49501)

xanthomonadaceae (523)

Submit Cancel

PDBe search – advanced search

- Allowing complex queries to be made easily



The screenshot displays the PDBe (Protein Data Bank in Europe) advanced search interface. At the top left is the PDBe logo. A search bar at the top right contains the example query "Ex. - hemoglobin, BRCA1_HUMAN" and a green "Search" button. Below the search bar are links for "Examples: hemoglobin, BRCA1_HUMAN" and "Advanced search".

The main section is titled "Advanced search form" and contains two search criteria:

- Organism name**: A dropdown menu is set to "Contains". The text input field contains "homo sapiens". An example "Example: Homo sapiens" is shown above the input.
- Molecule name**: A dropdown menu is set to "AND" and another to "Contains". The text input field contains "hemoglobin subunit alpha". An example "Example: Carbonic anhydrase 2" is shown above the input.

At the bottom of the form, there is a section "Select to add a search field:" with a dropdown menu showing "Select / type to filter options". At the very bottom are "Submit" and "Cancel" buttons.

PDBe search – advanced search

- Operator can be changed from AND and the regular expression can be

Advanced search form

Molecule name ⓘ

condition
Contains

Example: Carbonic anhydrase 2
Hemoglobin subunit alpha

Organism name ⓘ

condition
AND

OR

NOT

IGNORE

Select / type to filter options

Submit Cancel

Advanced search form

Molecule name ⓘ

condition
Contains

Example: Carbonic anhydrase 2
Hemoglobin subunit alpha

Organism name ⓘ

condition
AND

Contains

Equal to

Starts with

Ends with

Select to add a search field

Select / type to filter options

Submit Cancel

PDBe search – advanced search

Non human Hemoglobin subunit alpha

Advanced search form

Molecule name ⓘ

condition
Contains ▼

Example: Carbonic anhydrase 2
Hemoglobin subunit alpha

Organism name ⓘ

condition
NOT ▼ Equal to ▼

Example: Homo sapiens
homo sapiens

Hemoglobin subunit alpha from either human or sperm whale

Advanced search form

Molecule name ⓘ

condition
Contains ▼

Example: Carbonic anhydrase 2
Hemoglobin subunit alpha

Organism name ⓘ

condition
AND ▼ Contains ▼

Example: Homo sapiens
homo sapiens

Organism name ⓘ

condition
OR ▼ Contains ▼

Example: Homo sapiens
sperm whale

PDBe's search – wild card search

- Wild card searching works in the advanced search
- Single character (matches a single character) - ?
 - The search string te?t would match both test and text.
- Multiple characters (matches zero or more sequential characters) - *
 - The wildcard search:
 - tes* - would match test, testing, and tester.
 - You can also use wildcard characters in the middle of a term. For example: te*t - would match test and text.
 - *est - would match pest and test.

PDBe search - results

- The results page
 - Allows the search to be changed
 - Presents PDB search results with one panel per PDB entry

The screenshot shows the PDBe search results page. At the top, the PDBe logo and search bar are visible. The search bar contains the text 'Ex. - hemoglobin, BRCA1_HUMAN' and a 'Search' button. Below the search bar, there are tabs for 'Entries', 'Macromolecules', 'Compounds', and 'Protein families'. The 'Entries' tab is selected. The search filters are 'Organism name : homo sapiens' and 'Molecule name : hemoglobin ...'. The results are displayed in a list format. The first result is for entry 6I5y, titled 'Carbonmonoxy human hemoglobin A in the R2 quaternary structure at 140 K: Light (2 min)'. It includes details about the authors, source organism, assembly composition, and bound ligands. The second result is for entry 5e83, titled 'CRYSTAL STRUCTURE OF CARBONMONOXY HEMOGLOBIN S (LIGANDED SICKLE CELL HEMOGLOBIN) COMPLEXED WITH GBT440, CO-CRYSTALLIZATION EXPERIMENT'. Both results include a 3D visualization and download options. The page also features a 'Filter by' section on the left and a 'Download' button at the top right.

PDBe search – results – changing the query

- Additional filters can be easily added using the facets

The image displays the PDBe search interface. The main search results page shows a search for "Molecule name : Hemoglobin subunit alpha" and "Organism name : homo sapiens". The results are filtered by "Experimental method: Electron Microscopy", showing 3 entries. An inset shows the search results for entry 6nbc, which is a human methemoglobin state 1 determined using single-particle cryo-EM at 200 keV.

Search Criteria:
Molecule name : Hemoglobin subunit alpha
AND Organism name : homo sapiens

Filter by:
Entry Information
- Entry status (1)
REL (284)
- Experimental methods (4)
X-ray diffraction (277)
Electron Microscopy (3)
Neutron Diffraction (2)
Solution NMR (2)
+ Authors (503)
+ Homo / hetero assembly (1)
+ Assembly composition (1)

Search Results (Entries 1 to 10 of 284):
4ni1 Qaternary R CO-liganded hemoglobin structure in complex with a thiol containing compound
Safo MK, Meadows J, Ko T-P, Nakagawa A, Zapol W
ACS Chem. Biol. (2014) [PMID: 25061917]
Source organism: [Homo sapiens](#)
Assembly composition: protein/protein complex
Bound ligands: [HEM](#) [CMO](#) [2JX](#) [MBN](#)
PDBe-KB: [P69905](#) [P68871](#)
[3D Visualisation](#) [Download files](#)

Search Results (Entries 1 to 3 of 3):
6nbc human methemoglobin state 1 determined using single-particle cryo-EM at 200 keV
Herzik Jr MA, Wu M, Lander GC
Nat Commun (2019) [PMID: 30833564]
Source organism: [Homo sapiens](#)
Assembly composition: protein/protein complex
Bound ligands: [HEM](#)
PDBe-KB: [P69905](#) [P68871](#)
[3D Visualisation](#) [Download files](#)
Electron Microscopy
2.8Å resolution
Released: 13 Mar 2019
Model geometry
Fit model/data
Data not analysed

Search Results (Entries 1 to 1 of 1):
5ni1 CryoEM structure of haemoglobin at 3.2 Å determined with the Volta phase plate
Electron Microscopy
3.2Å resolution

What data is available in the search?

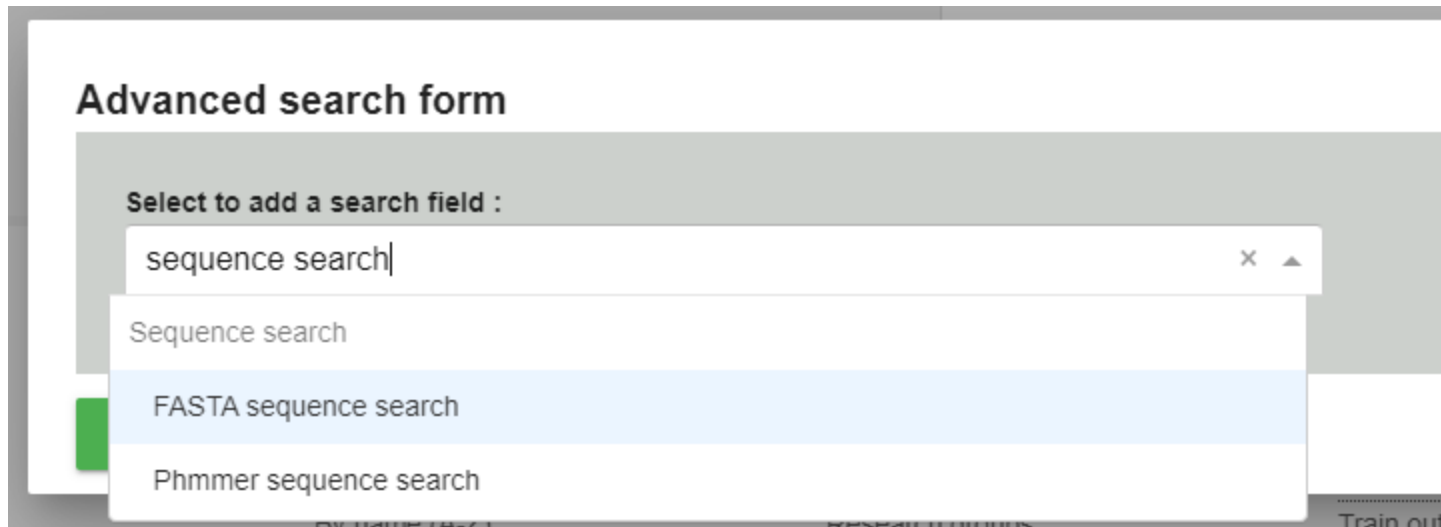
- Information about the entry
 - PDB ID
 - Date of release
 - Experimental method
- Related database identifiers
 - EMDB, BMRB
- Source and taxonomy information
- Experimental details for the entry
- Crystallographic cell parameters
- Simplified quality information
- Primary citation
- Likely biological assembly
- Chemical components details
- Cross references
 - UniProt, Pfam, SCOP, CATH, InterPro, GO, Rfam
- Any information you see on the search results page

The screenshot shows the PDB search results page for the query 'hemoglobin, BRCA1_HUMAN'. The page is divided into several sections:

- Search Bar:** Contains the search query and a search button. Examples of search terms are provided: 'hemoglobin, BRCA1_HUMAN'.
- Filters:** A sidebar on the left shows filters for 'Organism name: homo sapiens' and 'Molecule name: hemoglobin...'. There is a 'remove all filters' button.
- Advanced Search:** A green button labeled 'Advanced search' is visible.
- Download:** A green button labeled 'Download' is visible.
- Navigation:** A navigation bar shows 'Entries', 'Macromolecules', 'Compounds', and 'Protein families'. Below it, there are pagination controls showing 'Entries 1 to 10 of 284' and a 'Sort by' dropdown menu.
- Entry Information:** A section on the left lists various filters for the search results, such as 'Entry status (1)', 'REL (284)', 'Experimental methods (4)', 'Authors (503)', 'Homo / hetero assembly (1)', 'Assembly composition (1)', 'Assembly polymer count (7)', 'Resolution distribution', 'Release year distribution', and 'Journal (41)'. Below this, there are sections for 'Macromolecules' and 'Materials'.
- Search Results:** The main area displays a list of search results. The first result is for entry 6I5y, titled 'Carbonmonoxy human hemoglobin A in the R2 quaternary structure at 140 K: Light (2 min)'. It includes the authors 'Shibayama N, Park SY, Ohki M, Sato-Tomita A', the journal 'Proc. Natl. Acad. Sci. U.S.A. (2020) [PMID: 32071219]', the source organism 'Homo sapiens', and the assembly composition 'protein/protein complex'. It also lists bound ligands 'HEM, CMO' and PDB IDs 'P69905, P68871'. There are links for '3D Visualisation' and 'Download files'. To the right of the entry, there is a summary of X-ray diffraction data: 'X-ray diffraction 1.65Å resolution', 'Released: 19 Feb 2020', and a color scale for 'Model geometry' and 'Fit model/data'.
- Second Result:** The second result is for entry 5e83, titled 'CRYSTAL STRUCTURE OF CARBONMONOXO HEMOGLOBIN S (LIGANDED SICKLE CELL HEMOGLOBIN) COMPLEXED WITH GBT440, CO-CRYSTALLIZATION EXPERIMENT'. It includes the authors 'Patskovska L, Patskovsky Y, Bonanno JB, Almo SC' and X-ray diffraction data: 'X-ray diffraction 1.8Å resolution', 'Released: 20 Jul 2016', and a color scale for 'Model geometry' and 'Fit model/data'.

PDBe search – sequence search

- PDBe's search offers two protein sequence searches to search against the PDB in the advanced search
- Both use EBI's sequence search programs
 - FASTA – <https://www.ebi.ac.uk/Tools/sss/fasta/> - Heuristic approach
 - Phmmer - <https://www.ebi.ac.uk/Tools/hmmer/search/phmmer> - hidden Markov model approach



PDBe search – sequence search

- Search results from sequence search provide the
 - Alignment
 - E-value
 - Percentage identity
 - Target chain ID in the entry

PDBE / SEARCH

FASTA sequence search : MMYK... x

Advanced search

Download

Filter by :

Entry Information

- Entry status (1)
- REL (30)
- + Experimental methods (1)
- + Homo / hetero assembly (1)
- + Assembly composition (1)
- + Assembly polymer count (2)
- + Journal (6)

Macromolecules

- + Organism superkingdom (2)
- + Molecule name (7)
- + Molecule type (1)
- + Gene names (5)
- + Interacting ligands (22)

Function and Biology

- + Biological function (8)
- + Biological process (3)

Entries 1 to 30 of 30

E-value (desc) 100 /page

Select all entries on this page

5hi0 The Substrate Binding Mode and Chemical Basis of a Reaction Specificity Switch in Oxalate Decarboxylase

X-ray diffraction
2.602Å resolution
Released: 6 Apr 2016
Model geometry
Fit model/data

Zhu W, Easthon LM, Reinhardt LA, Tu C, Cohen SE, Silverman DN, Allen KN, Richards NGJ
Biochemistry (2016) [PMID: [27014926](#)]

Source organism: *Bacillus subtilis subsp. subtilis str. 168*

Assembly composition: protein only structure

Bound ligands: [CO](#) [OXL](#) [NA](#)

PDBe-KB: [O34714](#)

3D Visualisation Download files

E-value: 0.0021
Target: 5HI0_A
Identity percentage: 27.2%

Sequence Alignment:

```
Query : MMYKEPFGVKVDFETGIIIEGAKKSVRRLLSDMEGYFVDERAHKELVEKEDPVVVEVYAVEQEEKEG-----DI
Return : LVFDDG5FSNSTFQLTDWLAHTPKVEVIAANFGVTKE-EISNLPGKEKYIFE-NQLPGSLKDDIVEGPNGEVPPYFTYRLLLEQEPVIESEGGKVIADST
```

Summary of the webinar

- PDBe in the wwPDB
- PDBe search
- Advanced search
- Introduction to documentation page
- Usage examples

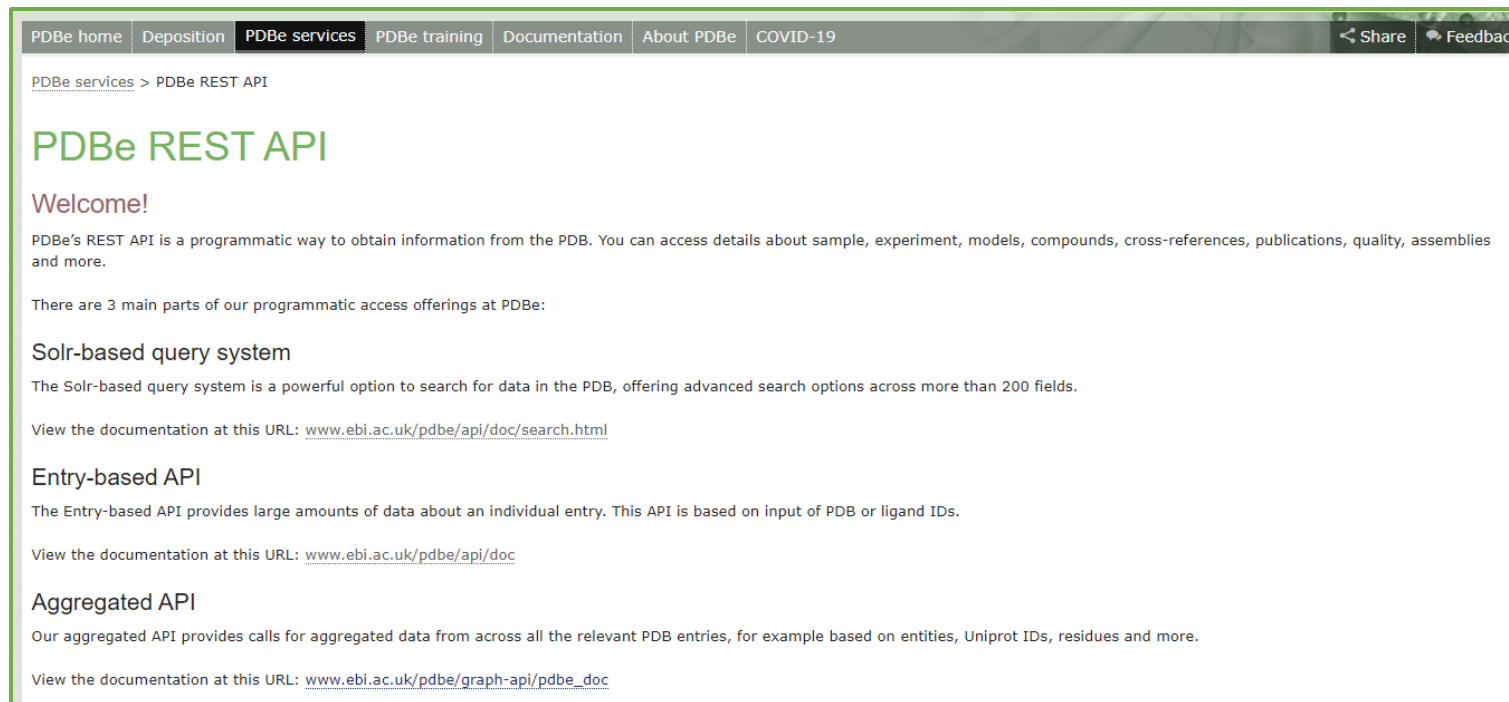
PDBe search – Solr powered



- PDBe's search is powered by Apache Solr
- 214 data items indexed
- 126 are available through advanced search
- 76 are available through the auto complete functionality
- 37 are available through facets in search results

PDBe's APIs

- All data for PDBe's webpages are served through APIs
- Search is no exception to this
- PDBe's API are available at <https://pdbe.org/api>



The screenshot shows the PDBe REST API page. The navigation bar includes links for PDBe home, Deposition, PDBe services (selected), PDBe training, Documentation, About PDBe, and COVID-19. There are also Share and Feedback buttons. The main content area is titled 'PDBe REST API' and includes a 'Welcome!' message. It describes the REST API as a programmatic way to obtain information from the PDB and lists three main parts of programmatic access offerings: Solr-based query system, Entry-based API, and Aggregated API. Each part includes a brief description and a link to its documentation.

PDBe home | Deposition | **PDBe services** | PDBe training | Documentation | About PDBe | COVID-19 | Share | Feedback

PDBe services > PDBe REST API

PDBe REST API

Welcome!

PDBe's REST API is a programmatic way to obtain information from the PDB. You can access details about sample, experiment, models, compounds, cross-references, publications, quality, assemblies and more.

There are 3 main parts of our programmatic access offerings at PDBe:

Solr-based query system

The Solr-based query system is a powerful option to search for data in the PDB, offering advanced search options across more than 200 fields.

View the documentation at this URL: www.ebi.ac.uk/pdbe/api/doc/search.html

Entry-based API

The Entry-based API provides large amounts of data about an individual entry. This API is based on input of PDB or ligand IDs.

View the documentation at this URL: www.ebi.ac.uk/pdbe/api/doc

Aggregated API

Our aggregated API provides calls for aggregated data from across all the relevant PDB entries, for example based on entities, Uniprot IDs, residues and more.

View the documentation at this URL: www.ebi.ac.uk/pdbe/graph-api/pdbe_doc

PDBe's search - API

PDBe home | Deposition | **PDBe services** | PDBe training | Documentation | About PDBe | COVID-19 Share Feedback

[PDBe services](#) > [PDBe REST API](#)

PDBe REST API

Welcome!

PDBe's REST API is a programmatic way to obtain information from the PDB. You can access details about sample, experiment, models, compounds, cross-references, publications, quality, assemblies and more.

There are 3 main parts of our programmatic access offerings at PDBe:

Solr-based query system

The Solr-based query system is a powerful option to search for data in the PDB, offering advanced search options across more than 200 fields.

View the documentation at this URL: www.ebi.ac.uk/pdbe/api/doc/search.html

Entry-based API

The Entry-based API provides large amounts of data about an individual entry. This API is based on input of PDB or ligand IDs.

View the documentation at this URL: www.ebi.ac.uk/pdbe/api/doc

Aggregated API

Our aggregated API provides calls for aggregated data from across all the relevant PDB entries, for example based on entities, Uniprot IDs, residues and more.

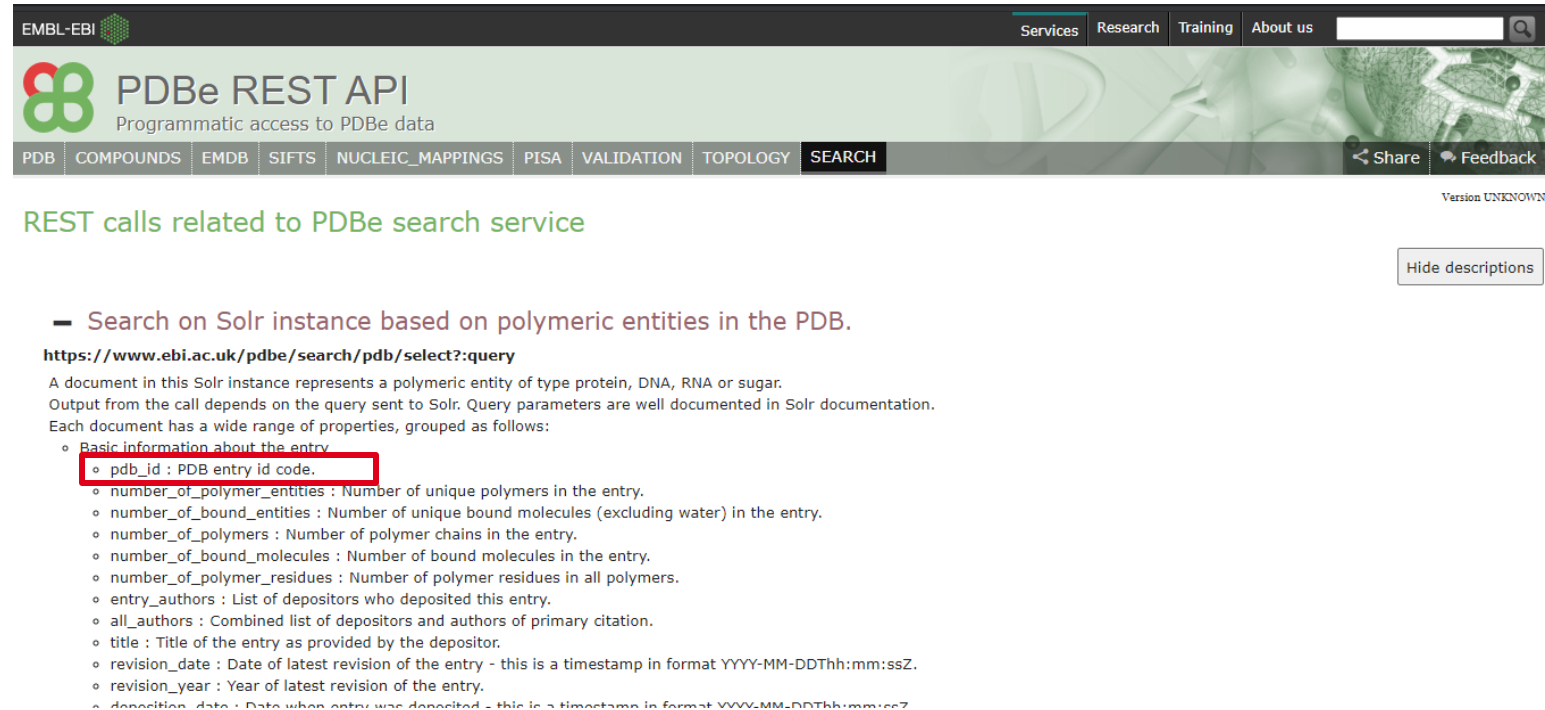
View the documentation at this URL: www.ebi.ac.uk/pdbe/graph-api/pdbe_doc

For more information, visit the following links:

- [Frequently asked questions \(FAQ\)](#)
- [Webinar: Introduction to programmatic access at PDBe](#)
- [Get started - worked out examples](#)
- [PDBe API user mailing list](#)

PDBe's search – API documentation

- The documentation provides a list of fields which are available for querying
- For example
 - `pdb_id` : PDB entry id code
- Field name is before the colon
- A description on what data is in the field is provided after the colon



EMBL-EBI

Services Research Training About us

PDBe REST API
Programmatic access to PDBe data

PDB COMPOUNDS EMD SIFTS NUCLEIC_MAPPINGS PISA VALIDATION TOPOLOGY SEARCH

REST calls related to PDBe search service

Version UNKNOWN

Hide descriptions

— Search on Solr instance based on polymeric entities in the PDB.

<https://www.ebi.ac.uk/pdbe/search/pdb/select?:query>

A document in this Solr instance represents a polymeric entity of type protein, DNA, RNA or sugar. Output from the call depends on the query sent to Solr. Query parameters are well documented in Solr documentation. Each document has a wide range of properties, grouped as follows:

- Basic information about the entry
 - `pdb_id` : PDB entry id code.
 - `number_of_polymer_entities` : Number of unique polymers in the entry.
 - `number_of_bound_entities` : Number of unique bound molecules (excluding water) in the entry.
 - `number_of_polymers` : Number of polymer chains in the entry.
 - `number_of_bound_molecules` : Number of bound molecules in the entry.
 - `number_of_polymer_residues` : Number of polymer residues in all polymers.
 - `entry_authors` : List of depositors who deposited this entry.
 - `all_authors` : Combined list of depositors and authors of primary citation.
 - `title` : Title of the entry as provided by the depositor.
 - `revision_date` : Date of latest revision of the entry - this is a timestamp in format YYYY-MM-DDThh:mm:ssZ.
 - `revision_year` : Year of latest revision of the entry.
 - `deposition_date` : Date when entry was deposited - this is a timestamp in format YYYY-MM-DDThh:mm:ssZ.


PDBe's search – API documentation

- And a input box for an interactive query tool

- scop_fold : A list of SCOP fold names for the SCOP domains mapped to the chains of this entity.
- scop_class : A list of SCOP class names for the SCOP domains mapped to the chains of this entity.
- CATH
 - cath_code : A list of CATH codes for the CATH domains mapped to the chains of this entity.
 - cath_class : A list of CATH classes for the CATH domains mapped to the chains of this entity.
 - cath_architecture : A list of CATH architectures for the CATH domains mapped to the chains of this entity.
 - cath_topology : A list of CATH topologies for the CATH domains mapped to the chains of this entity.
 - cath_homologous_superfamily : A list of CATH superfamilies for the CATH domains mapped to the chains of this entity.
- InterPro
 - interpro_accession : A list of InterPro accessions mapped to the chains of this entity.
 - interpro_name : A list of names of InterPro accessions mapped to the chains of this entity.
- GO
 - go_id : A list of GO ids mapped to this entity.
 - biological_cell_component : A list of GO cellular component names mapped to this entity.
 - biological_function : A list of GO biological functions mapped to this entity.
 - biological_process : A list of GO biological processes mapped to this entity.
- HomoloGene
 - homologous_pdb_entity_id : A list of concatenated pdb and entity ids that are homologous to the entity according to the HomoloGene resource, e.g. ['1brq_1', '2wqa_2'].

query	<input type="text" value="q=pfam_name:Lipocalin&w"/>	String	Options string allowed by Solr syntax. Details about constructing Solr queries can be found from webpages such as this .
postdata	<input type="text"/>	String	Send one or more query options in post data instead of appending to URL.

Quotes

 PDBe is a member of  PDB  EMDataBank

This project is also funded by  West-Life  BBSRC

EMBL-EBI 

News

Services

By topic

Research

Overview

Training

Overview

Industry

Overview

About us

Overview

PDBe's search – API documentation

- Running the call gives
 - the URL which was used
 - The results in the same page

- biological_cell_component : A list of GO cellular component names mapped to this entity.
- biological_function : A list of GO biological functions mapped to this entity.
- biological_process : A list of GO biological processes mapped to this entity.
- HomoloGene
 - homologus_pdb_entity_id : A list of concatenated pdb and entity ids that are homologous to the entity according to the HomoloGene resource, e.g. ['1brq_1', '2wqa_2'].

query	<input type="text" value="q=pfam_name:Lipocalin&w"/>	String	Options string allowed by Solr syntax. Details about constructing Solr queries can be found from webpages such as this .
postdata	<input type="text"/>	String	Send one or more query options in post data instead of appending to URL.

Quotes

GET : https://www.ebi.ac.uk/pdbe/search/pdb/select?q=pfam_name:Lipocalin&wt=json

HTTP status : 200 : OK (Hover to find undocumented bits in the output.)

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "q": "pfam_name:Lipocalin",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 639,
    "start": 0,
    "docs": [
      {
        "abstracttext_unassigned": [
          "In the present study we examine the thermodynamics of binding of two related pyrazine-derived ligands to the major urinary protein, MUP-I, using a combination of isothermal titration calorimetry (ITC), X-ray crystallography, and NMR backbone (15)N and methyl side-chain (2)H relaxation measurements. Global thermodynamics data derived from ITC indicate that binding is driven by favorable enthalpic contributions, rather than the classical entrop-
```

PDBe's search – API documentation

- The search can be changed directly in the query box
- Any field in red in the “docs” section can be used in the search – these fields are the same as those in the documentation section above
- Here the search has been changed to “all assembly type:tetramer”

query	<input type="text" value="q=all_assembly_type:tetra"/>	String	Options string allowed by Solr syntax. Details about constructing Solr queries can be found from webpages such as this .
postdata	<input type="text"/>	String	Send one or more query options in post data instead of appending to URL.

Quotes

GET : https://www.ebi.ac.uk/pdbe/search/pdb/select?q=all_assembly_type:tetramer&wt=json

HTTP status : 200 : OK (Hover to find undocumented bits in the output.)

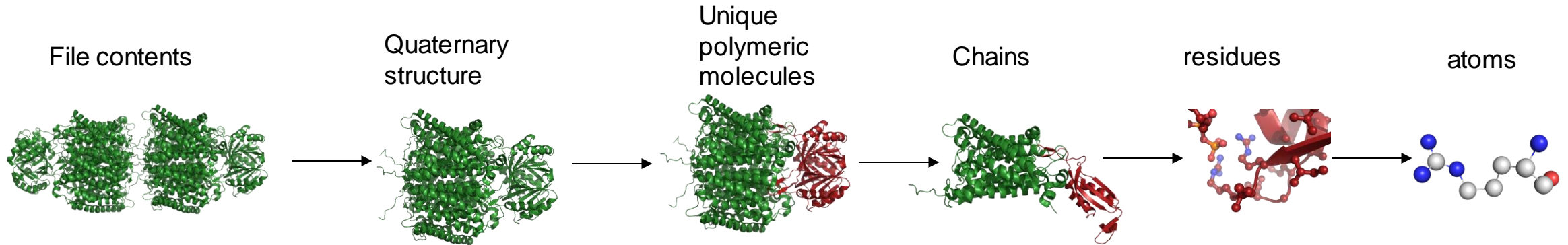
```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "q": "all_assembly_type:tetramer",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 30550,
    "start": 0,
```

PDBe search – Solr powered



- A “document” is the unit of search and index
- Within a document there are fields – properties of the document
- Comparing Solr to a database
 - Document = A table row
 - Fields = columns
- So for PDB entries what is the optimal “document”?

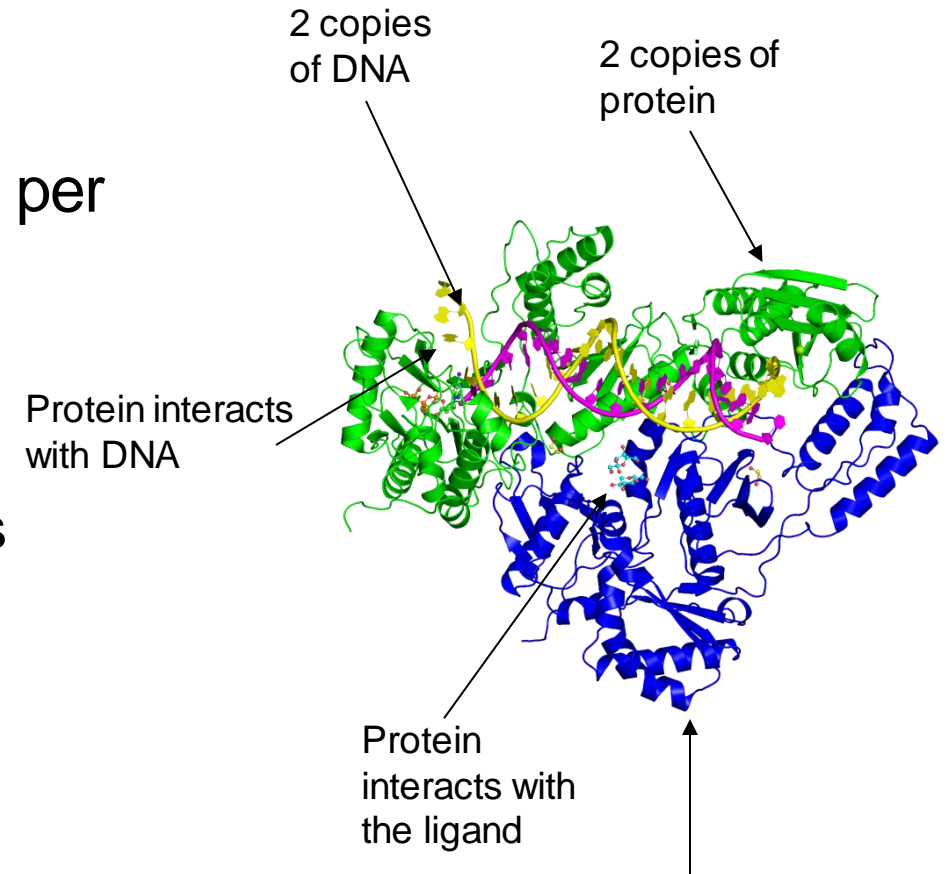
Hierarchy of a PDB entry



- Document must allow users to
 - Find all molecules in an entry
 - Access properties of a molecule – i.e. sequence
 - Access properties of an entry – i.e. quaternary structure
 - Not create too large an index so that the search is slow and returns unnecessary documents
- A PDBe we use “Unique polymeric molecules”

PDBe's search – API structure

- Advantages of using unique polymeric molecule per document
 - Type of molecule
 - The sequence
 - Number of copies in the entry and their chain IDs
 - If the sequence has modified residues
- And:
 - Which other polymers interact with this molecule
 - Which ligands interact with this molecule



```
>pdb|5tx1|A C
MVPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRT
QDFWEVQLGIPHPAGLKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNNVLPQGWKGSIPAIFQSSMTKILEP
FKKQNPDIVIYQYMDLVVGSLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPFLWYGELHPDKWTVQPIVLPKSDSNTVND
TKLVGKINHAQOTVDSKYDQLEKLRGTYALTEVTRITTEAELELAEMDETIVKRVHGVVYDREKVLTAETQKAGGQUTVAT
```

PDBe's search – API structure

- So if we change the search to `pdb_id:6l5y` – haemoglobin with 2 unique polymer molecules
- Then we get 2 results

query	<input type="text" value="q=pdb_id:6l5y&wt=json"/>	String	Options string allowed by Solr syntax. Details about constructing Solr queries can be found from webpages such as this .
postdata	<input type="text"/>	String	Send one or more query options in post data instead of appending to URL.

Quotes

RunCall

Select

Expand

Collapse

2+

3+

GET : https://www.ebi.ac.uk/pdbe/search/pdb/select?q=pdb_id:6l5y&wt=json

HTTP status : 200 : OK (Hover to find undocumented bits in the output.)

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "q": "pdb_id:6l5y",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 2,
    "start": 0,
    "docs": [

```

Summary of the webinar

- PDBe in the wwPDB
- PDBe search
- Advanced search
- Introduction to documentation page
- Usage examples

PDBe's search – URL construction

- PDBe's search is a Solr search
- https://www.ebi.ac.uk/pdbe/search/pdb/select?q=all_assembly_type:tetramer&wt=json
- Parts of this URL:
 - Search URL: <https://www.ebi.ac.uk/pdbe/search/pdb/select?>
 - Query: [q=all_assembly_type:tetramer](#)
 - Return type: [wt=json](#) – options are json, xml – default is json
- By default this returns 100 documents
 - Rows: [rows=100](#)
- More details about Solr syntax can be found in the Solr help pages
 - <https://lucene.apache.org/solr/guide/common-query-parameters>

PDBe's search – accessing data programmatically

- GET and POST calls supported
 - GET -
https://www.ebi.ac.uk/pdbe/search/pdb/select?q=all_assembly_type:tetramer&fl=pdb_id,experimental_method&rows=100&wt=json
 - With “&” separating parameters
 - POST
 - URL = <https://www.ebi.ac.uk/pdbe/search/pdb/select?>
 - data is posted as a JSON with parameters a key value pairs

```
{'fl': 'pdb_id,experimental_method',  
  'q': 'all_assembly_type:tetramer',  
  'rows': 100,  
  'wt': 'json'}
```

PDBe's search – accessing data programmatically

- By default each result / “document” contains every field
- This is useful for developing queries but usually you only need a few fields
- The Solr parameter Field List - “**fl**” allows only selected fields to be returned – speeding up return of results
- E.g.
 - [q=all assembly type:tetramer](#)
 - [fl=pdb_id,experimental_method](#)

```
▼ "response": {
  "numFound": 30550,
  "start": 0,
  ▼ "docs": [
    ▼ {
      ▼ "experimental_method": [
        "X-ray diffraction"
      ],
      "pdb_id": "3rqh"
    },
    ▼ {
      ▼ "experimental_method": [
        "X-ray diffraction"
      ],
      "pdb_id": "3vjr"
    },
    ▼ {
      ▼ "experimental_method": [
        "X-ray diffraction"
      ],
      "pdb_id": "3dpi"
    },
    ▼ {
      ▼ "experimental_method": [
        "X-ray diffraction"
      ],
      "pdb_id": "1slw"
    },
  ]
}
```

PDBe's search – exclusive searches - prosite patterns

- More advanced queries are possible in the API than the advanced search
- It's possible to look for prosite (<https://prosite.expasy.org/>) patterns in sequences
- For example
 - [ST]-x(2)-[DE] à Casein kinase II phosphorylation site
 - [q=molecule sequence:/.*?\[ST\].\[DE\].*/](#)
 - [RK]-x(2,3)-[DE]-x(2,3)-Y à Tyrosine kinase phosphorylation site
 - [q=molecule sequence:/.*?\[RK\].{2,3}\[DE\].{2,3}Y.*/](#)
 - G-{EDRKHPFYW}-x(2)-[STAGCN]-{P} à N-myristoylation site
 - [q=molecule sequence:/.*?G{^EDRKHPFYW}.{2}\[STAGCN\]\[^P\].*/](#)
- Prosite rules changed into regular expression rules
 - Prosite x(2,3) = any 2 or 3 residues - which becomes `.{2,3}`
 - Prosite {} = not any of these residues - which becomes `[^`
 - Prosite [] = any one of the residues - which remains `[]`

PDBe's search – exclusive searches - antibodies

- Antibody data is currently only available in the search API
- Text search for “antibody” gives 3414 PDB entries

PDBE / SEARCH

Text : antibody ✕ Advanced search Download

Filter by :

Latest PDB release

- Entries released this week (2)
- new (5)
- revised (1)

Entries | Macromolecules | Compounds | Protein families

< 1 2 3 ... 342 > Entries 1 to 10 of 3413 Sort by 10 /page

Select all entries on this page

<input type="checkbox"/> 4xcf	Crystal structure of human 4E10 Fab in complex with its peptide epitope on HIV-1 gp41; crystals cryoprotected with	X-ray diffraction 1.43Å resolution
---	--	---------------------------------------

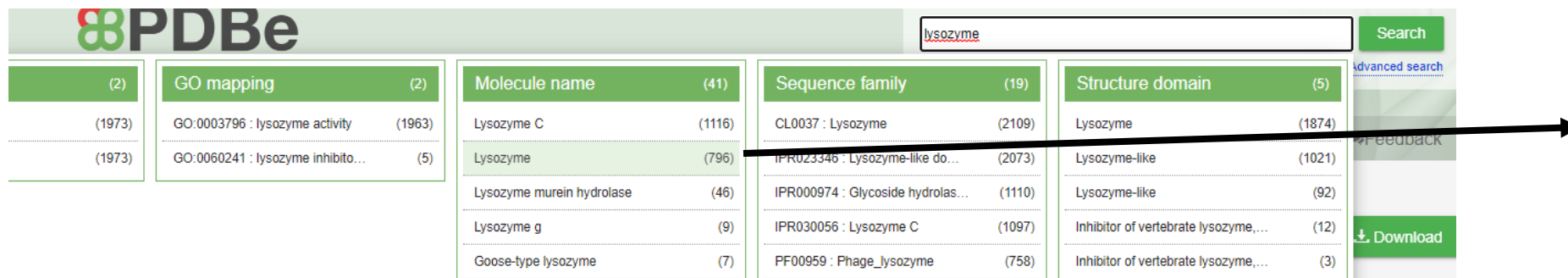
PDBe's search – worked example

- There is a flag for antibodies in the search API
 - “antibody_flag” – if this is set to Y then an entity is an antibody
 - So this can be searched for using
 - [q=antibody_flag:Y](#)
 - However, each PDB entry can contain multibody antibodies so it is necessary to group the results by “pdb_id” to get the number of PDB entries
 - This can be done using the following
 - [q=antibody_flag:Y](#)
 - [group=true](#)
 - [group.field=pdb_id](#)
 - [group.mincount=1](#)
 - [group.ngroups=true](#)
 - [fl=pdb_id](#)
 - Then “ngroups” is the number of PDB IDs
 - This 4931 is significantly more than the 3414 from a free text search

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 4,
    "params": {
      "q": "antibody_flag:Y",
      "group.mincount": "1",
      "fl": "pdb_id",
      "group.ngroups": "true",
      "wt": "json",
      "group.field": "pdb_id",
      "group": "true"
    }
  },
  "grouped": {
    "pdb_id": {
      "matches": 9313,
      "ngroups": 4931,
      "groups": [
        /
      ]
    }
  }
}
```

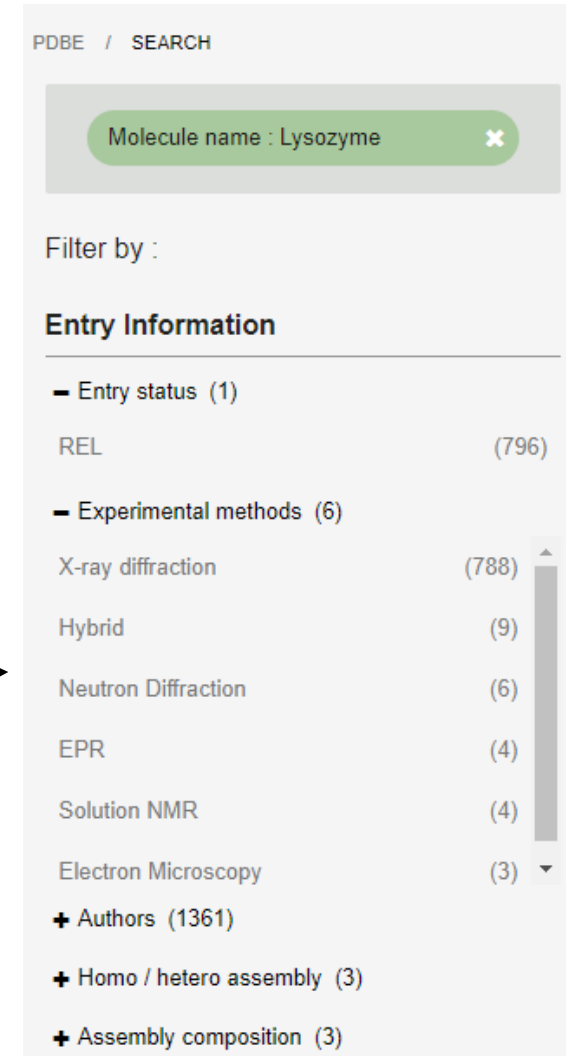
PDBe's search – worked example of faceting

- At PDBe we use faceting to display how many entries fit into each class
- For example
 - After searching for “Lysozyme” we can find the number of entries solved by each experimental method



The screenshot shows the PDBe search interface with the search term 'lysozyme' entered. The results are faceted into several categories:

Count	Facet	Count
(2)	GO mapping	(2)
(1973)	GO:0003796 : lysozyme activity	(1963)
(1973)	GO:0060241 : lysozyme inhibito...	(5)
(41)	Molecule name	(41)
(1116)	Lysozyme C	(1116)
(796)	Lysozyme	(796)
(46)	Lysozyme murein hydrolase	(46)
(9)	Lysozyme g	(9)
(7)	Goose-type lysozyme	(7)
(19)	Sequence family	(19)
(2109)	CL0037 : Lysozyme	(2109)
(2073)	IPR023346 : Lysozyme-like do...	(2073)
(1110)	IPR000974 : Glycoside hydrolas...	(1110)
(1097)	IPR030056 : Lysozyme C	(1097)
(758)	PF00959 : Phage_lysozyme	(758)
(5)	Structure domain	(5)
(1874)	Lysozyme	(1874)
(1021)	Lysozyme-like	(1021)
(92)	Lysozyme-like	(92)
(12)	Inhibitor of vertebrate lysozyme,...	(12)
(3)	Inhibitor of vertebrate lysozyme,...	(3)



The screenshot shows the 'Filter by' section of the PDBe search interface. The search term 'Lysozyme' is entered in the search box. The 'Filter by' section is expanded to show 'Experimental methods' with a count of 6. The methods listed are:

Method	Count
REL	(796)
X-ray diffraction	(788)
Hybrid	(9)
Neutron Diffraction	(6)
EPR	(4)
Solution NMR	(4)
Electron Microscopy	(3)
Authors	(1361)
Homo / hetero assembly	(3)
Assembly composition	(3)

PDBe's search – worked example of faceting

- We can repeat the same query using the search API

- [q=all_molecule_names:Lysozyme](#)
- [group=true](#)
- [group.field=pdb_id](#)
- [group.ngroups=true](#)
- [json.nl=map](#)
- [facet=true](#)
- [facet.mincount=1](#)
- [facet.field=experimental_method](#)
- [group.facets=true](#)

```
..
  "facet_counts": {
    "facet_queries": {},
    "facet_fields": {
      "experimental_method": {
        "X-ray diffraction": 789,
        "Hybrid": 9,
        "Neutron Diffraction": 6,
        "EPR": 4,
        "Solution NMR": 4,
        "Electron Microscopy": 3
      }
    }
  },
  ..
```

PDBe's search – worked example of pivoting

- We can pivot the results so we get
 - We use the query `q=all_molecule_names:Lysozyme`
 - `group=true`
 - `group.field=pdb_id`
 - `group.ngroups=true`
 - `json.nl=map`
 - `facet=true`
 - `facet.mincount=1`
 - `group.facets=true`
 - `facet.pivot=experimental_method`

```
''
  "grouped": {
    "pdb_id": {
      "matches": 797,
      "ngroups": 796,
      "groups": []
    }
  },
  "facet_counts": {
    "facet_queries": {},
    "facet_fields": {},
    "facet_ranges": {},
    "facet_intervals": {},
    "facet_heatmaps": {},
    "facet_pivot": {
      "experimental_method": [
        {
          "field": "experimental_method",
          "value": "X-ray diffraction",
          "count": 789
        },
        {
          "field": "experimental_method",
          "value": "Hybrid",
          "count": 9
        },
        {
          "field": "experimental_method",
          "value": "Neutron Diffraction",
          "count": 6
        },
        {
          "field": "experimental_method",
          "value": "EPR",
          "count": 4
        },
        {
          "field": "experimental_method",
          "value": "Solution NMR",
          "count": 4
        }
      ]
    }
  }
}
```



PDBe's search – worked example of faceting

- If we want more complex examples we can find the number of components in the assembly per experimental method in entries solved between 2014 to 2020
 - [q=release_year:\[2014 TO 2019\]](#)
 - [group=true](#)
 - [group.field=pdb_id](#)
 - [group.ngroups=true](#)
 - [json.nl=map](#)
 - [facet=true](#)
 - [facet.mincount=1](#)
 - [group.facets=true](#)
 - [facet.pivot=release_year,experimental_method,assembly_num_component](#)
 - [facet.sort=release_year asc](#)
- Pivots are given as a comma separated list and in the results you get a hierarchy of pivots

```
''
  "grouped": {
    "pdb_id": {
      "matches": 147867,
      "ngroups": 63413,
      "groups": []
    }
  },
  "facet_counts": {
    "facet_queries": {},
    "facet_fields": {},
    "facet_ranges": {},
    "facet_intervals": {},
    "facet_heatmaps": {},
    "facet_pivot": {
      "release_year,experimental_method,assembly_num_component": [
        {
          "field": "release_year",
          "value": 2014,
          "count": 27242,
          "pivot": [
            {
              "field": "experimental_method",
              "value": "Electron Microscopy",
              "count": 5185,
              "pivot": [
                {
                  "field": "assembly_num_component",
                  "value": 1,
                  "count": 4
                },
                {
                  "field": "assembly_num_component",
                  "value": 2,
                  "count": 15
                }
              ]
            }
          ]
        }
      ]
    }
  }
}
```

PDBe's search – teaching resources

- API tutorials are available in Python on PDBe's GitHub page
- <https://github.com/PDBEurope/pdbe-api-training>
- A specific search API tutorial is available
- https://github.com/PDBEurope/pdbe-api-training/blob/master/api_tutorials/6_PDB_search.ipynb

2791 lines (2791 sloc) | 366 KB

PDBe API Training

This interactive Python notebook will guide you through various ways of programmatically accessing Protein Data Bank in Europe (PDBe) data using REST API

The REST API is a programmatic way to obtain information from the PDB and EMDB. You can access details about:

- sample
- experiment
- models
- compounds
- cross-references
- publications
- quality
- assemblies and more... For more information, visit <https://www.ebi.ac.uk/pdbe/pdbe-rest-api>

Notebook #6

This notebook is the sixth in the training material series, and focuses on getting information from the PDBe search API

1) Making imports and setting variables

First, we import some packages that we will use, and set some variables.

Note: Full list of valid URLs is available from <https://www.ebi.ac.uk/pdbe/api/doc/>

```
In [1]: import requests # used for getting data from a URL
from pprint import pprint # pretty print
import pandas as pd # used for turning results into mini databases
from solrq import Q # used to turn result queries into the right format

search_url = "https://www.ebi.ac.uk/pdbe/search/pdb/select?" # the rest of the URL used for PDBe's search API.
```

2) a function to get data from the search API

Let's start with defining a function that can be used to GET data from the PDBe search API.

```
In [2]: def make_request(search_dict, number_of_rows=10):
    """
    makes a get request to the PDBe API
    :param dict search_dict: the terms used to search
    :param number_of_rows: number of rows to return - limited to 10
    """
```

PDBe search – sequence search

- Search results from sequence search is also available through the API

FASTA sequence search : MMYK... ✕

Advanced search

Download

Filter by :

Entry Information

- Entry status (1)
- REL (30)
- + Experimental methods (1)
- + Homo / hetero assembly (1)
- + Assembly composition (1)
- + Assembly polymer count (2)
- + Journal (6)

Macromolecules

- + Organism superkingdom (2)
- + Molecule name (7)
- + Molecule type (1)
- + Gene names (5)
- + Interacting ligands (22)

Function and Biology

- + Biological function (8)
- + Biological process (3)
- + Biological cell component (2)

Entries 1 to 30 of 30

E-value (asc) 100 /page

Select all entries on this page

[2gc0](#) The crystal structure of phosphoglucose isomerase from *Pyrococcus furiosus* in complex with 5-phospho-D-arabinonohydroxamate and zinc

Berrisford JM, Rice DW, Baker PJ
J. Mol. Biol. (2006) [PMID: [16580686](#)]

Source organism: [Pyrococcus furiosus](#)

Assembly composition: protein only structure

Bound ligands: [ZN](#) [PAN](#)

PDBe-KB: [P83194](#)

3D Visualisation Download files

X-ray diffraction
2Å resolution
Released: 11 Apr 2006
Model geometry
Fit model/data

E-value: 6.4e-81
Target: 2GC0_B
Identity percentage: 100%

Sequence Alignment:

Query : MMYKEPFGVKVDFETGII EGAKKSVRRRLSDMEGYFVDERAWKELVEKEDPVVYEVYAVEQEKEGDLNFATTVLYPGKVGKEFFFTKGHFHAKLDRAE

Return : MMYKEPFGVKVDFETGII EGAKKSVRRRLSDMEGYFVDERAWKELVEKEDPVVYEVYAVEQEKEGDLNFATTVLYPGKVGKEFFFTKGHFHAKLDRAE

PDBe's search – sequence search

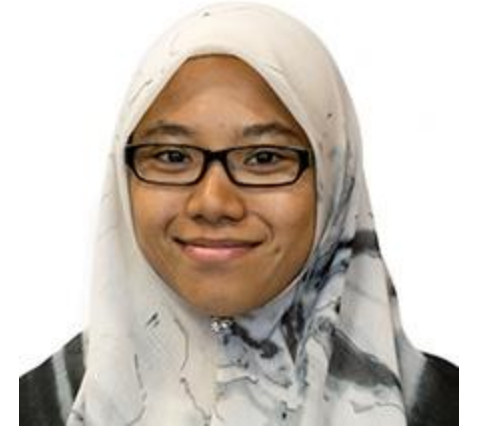
- PDBe's FASTA and pHMMER searches are available through the search API
- The sequence search results are joined to the usual results and a bit of post processing is required
- This is covered in
 - https://github.com/PDBEurope/pdbe-api-training/blob/master/api_tutorials/7_PDB_sequence_search.ipynb

PDBe's search

- PDBe provides an advanced search powered by Solr
- PDBe provides 3 distinct APIs
 - Search
 - Entry based
 - Aggregated API
- PDBe's search system is accessible directly via the search API
- Complex queries are possible, including FASTA and HMMER searches
- Tutorials are available at <https://github.com/PDBEurope/pdbe-api-training>

PDBe API webinar series

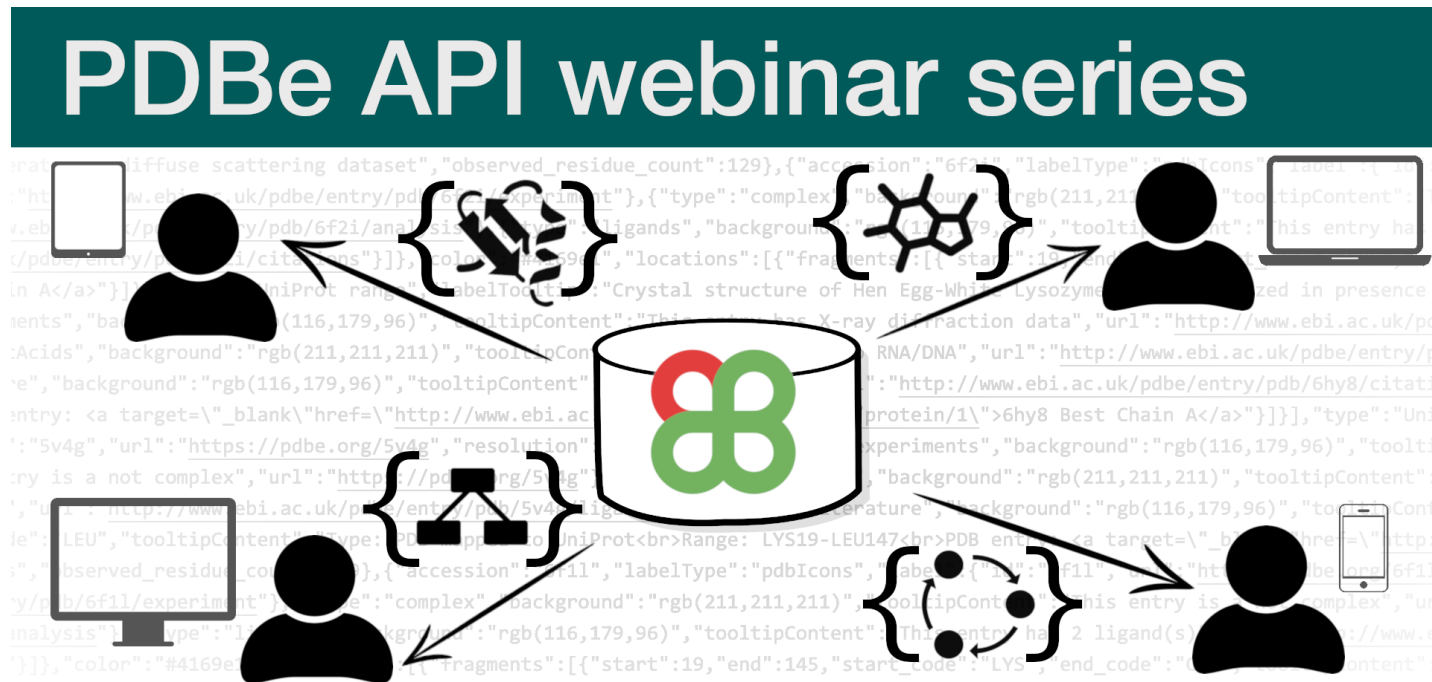
- Sep 15th Introduction to PDBe programmatic access
- Sep 22nd Searching with the PDBe API
- Sep 29th Creating complex PDBe API queries**
- Oct 6th Using the PDBe graph API
- Oct 13th PDBe tools in github
- Oct 20th Data visualisation at PDBe



Nurul Nadzirin

To register for future webinars in the series

- See the full list of upcoming webinars at bit.ly/PDBeAPIwebinars
- Or visit the PDBe events pages at PDBe.org/events
- Remember to register for each webinar individually!!!



Thank you for your attention!

Any questions?

[PDBe.org/API](https://pdbe.org/API)



John Berrisford



pdhelp@ebi.ac.uk



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



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



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



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