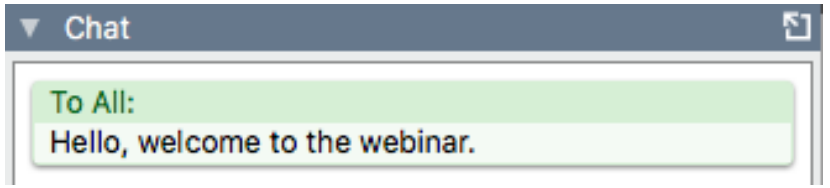
Welcome - webinar instructions

- GoToTraining works best in **Chrome** or on Linux, **Firefox**
- All **microphones will be muted** while 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

PDBe API webinar series: Introduction to the PDBe API

PDBe.org/API

David Armstrong

PDBe
Protein Data Bank in Europe

pdbhelp@ebi.ac.uk
proteindatabank
@PDBeurope
proteindatabank
pdbeurope
pdbart

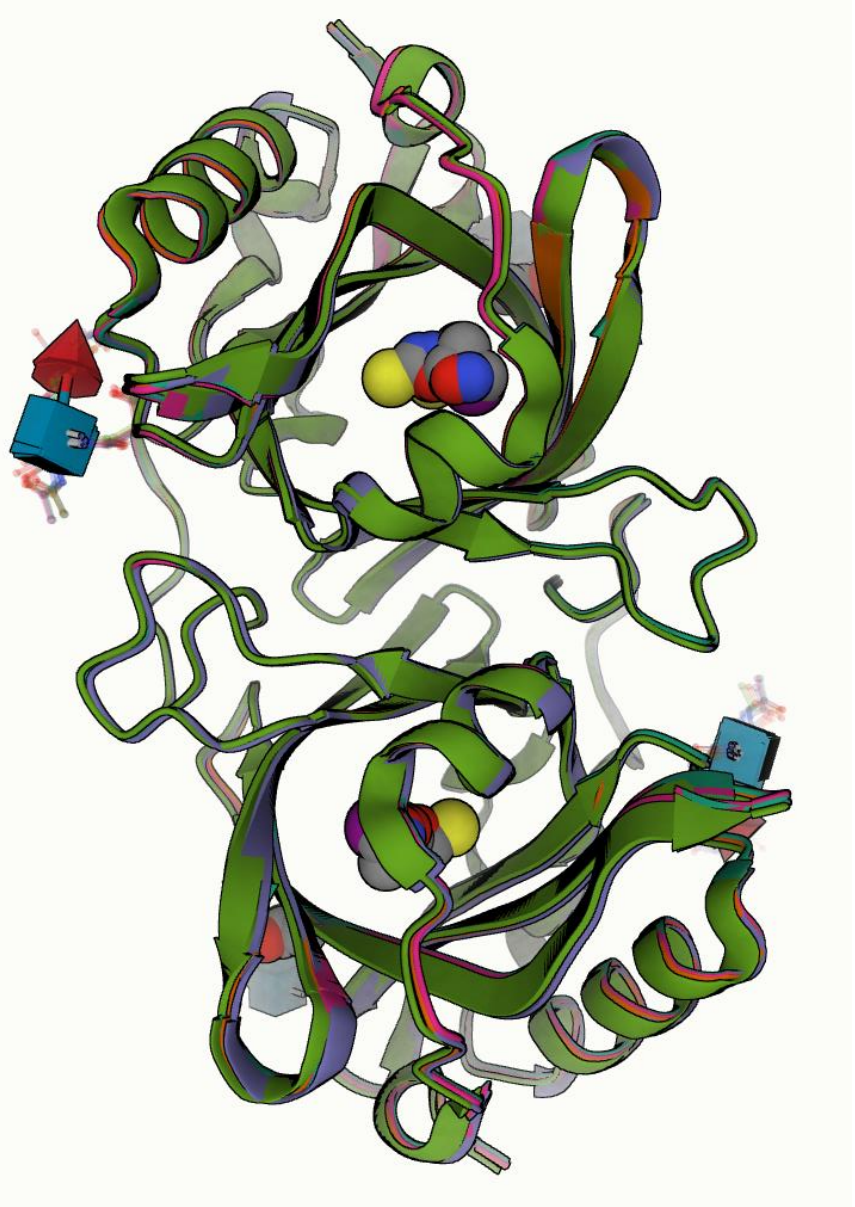
EMBL-EBI

PDBe API webinar series

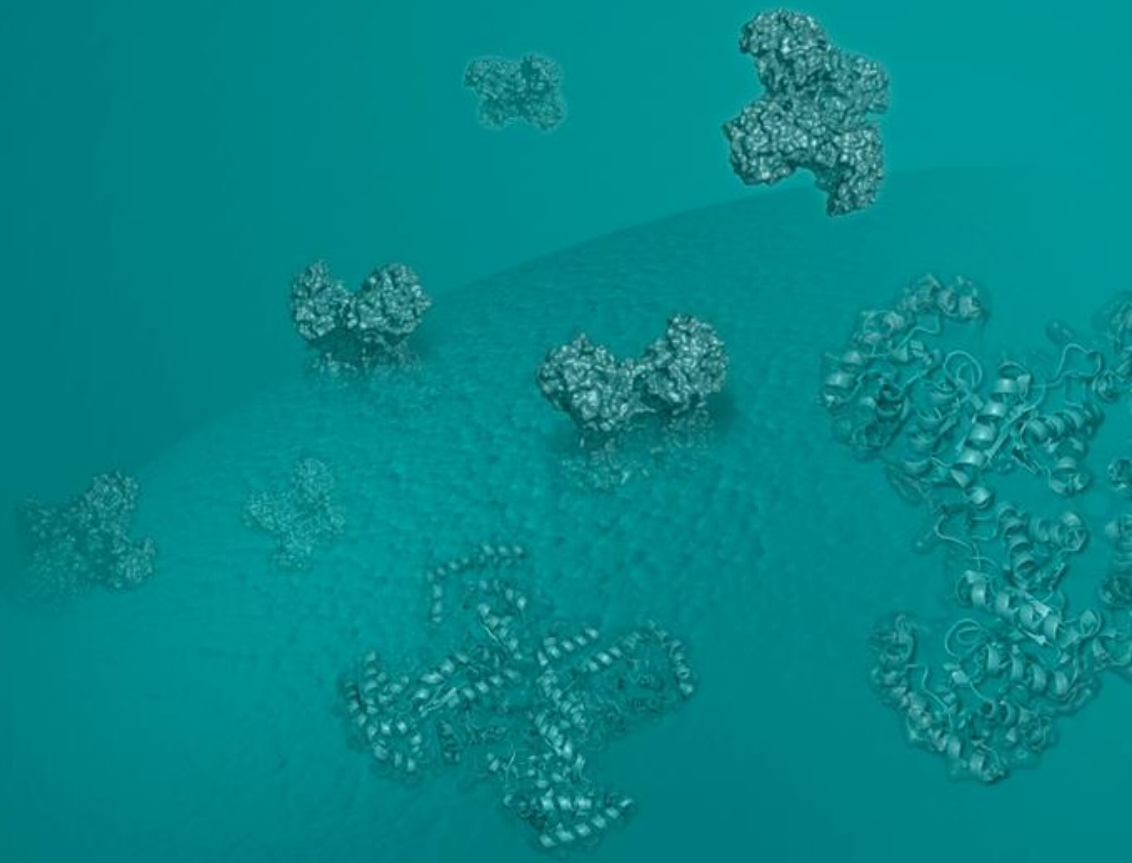
Sep 15th	Introduction to PDBe programmatic access
Sep 22nd	Searching with the PDBe API
Sep 29th	Creating complex PDBe API queries
Oct 6th	Using the PDBe graph API
Oct 13th	PDBe tools in github
Oct 20th	Data visualisation at PDBe

Summary of this webinar

- What type of data is in the PDB?
- Additional functional data added by PDBe
- Why use the PDBe APIs?
- What types of API are available from PDBe
- How we use APIs on PDBe pages
- Outline of the rest of the webinar series

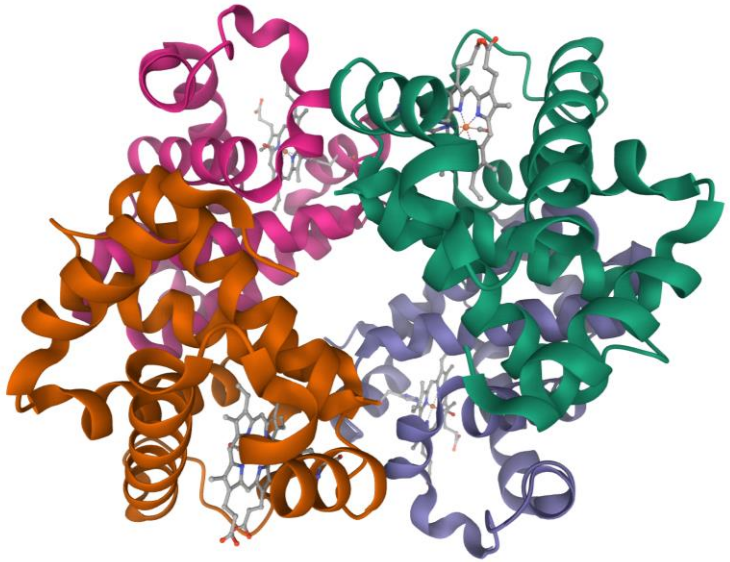


What type of data is in the PDB?

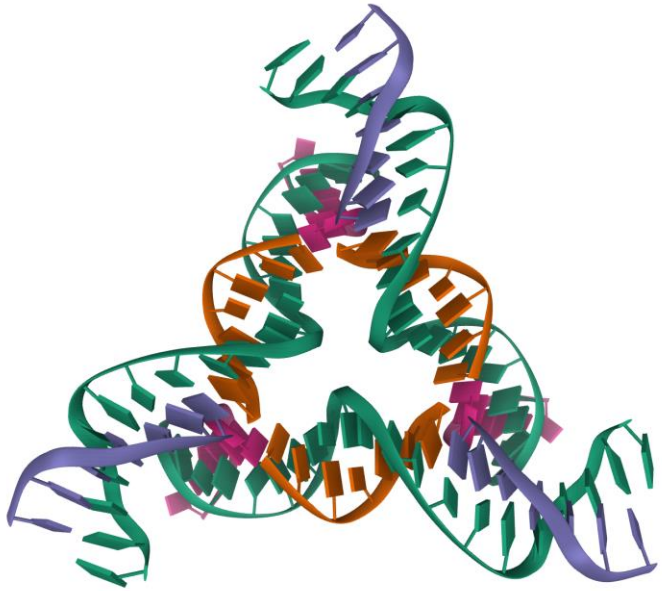


EMBL-EBI

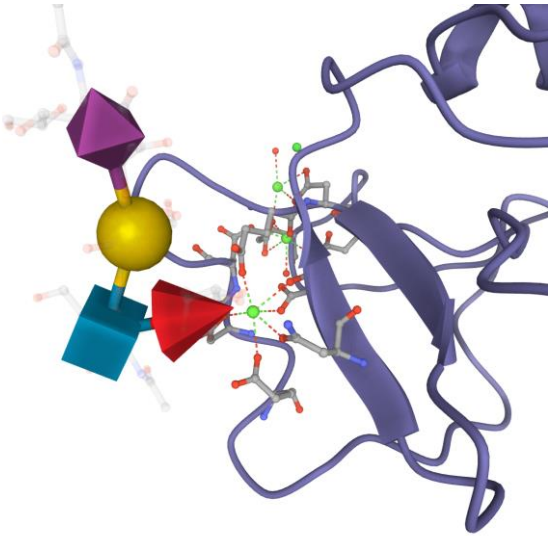
The Protein Data Bank (PDB) is an archive of **experimentally determined** 3-dimensional structures of biological macromolecules



Proteins



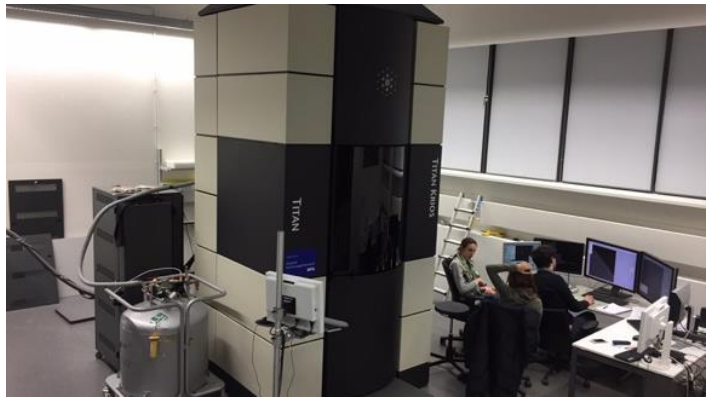
DNA/RNA



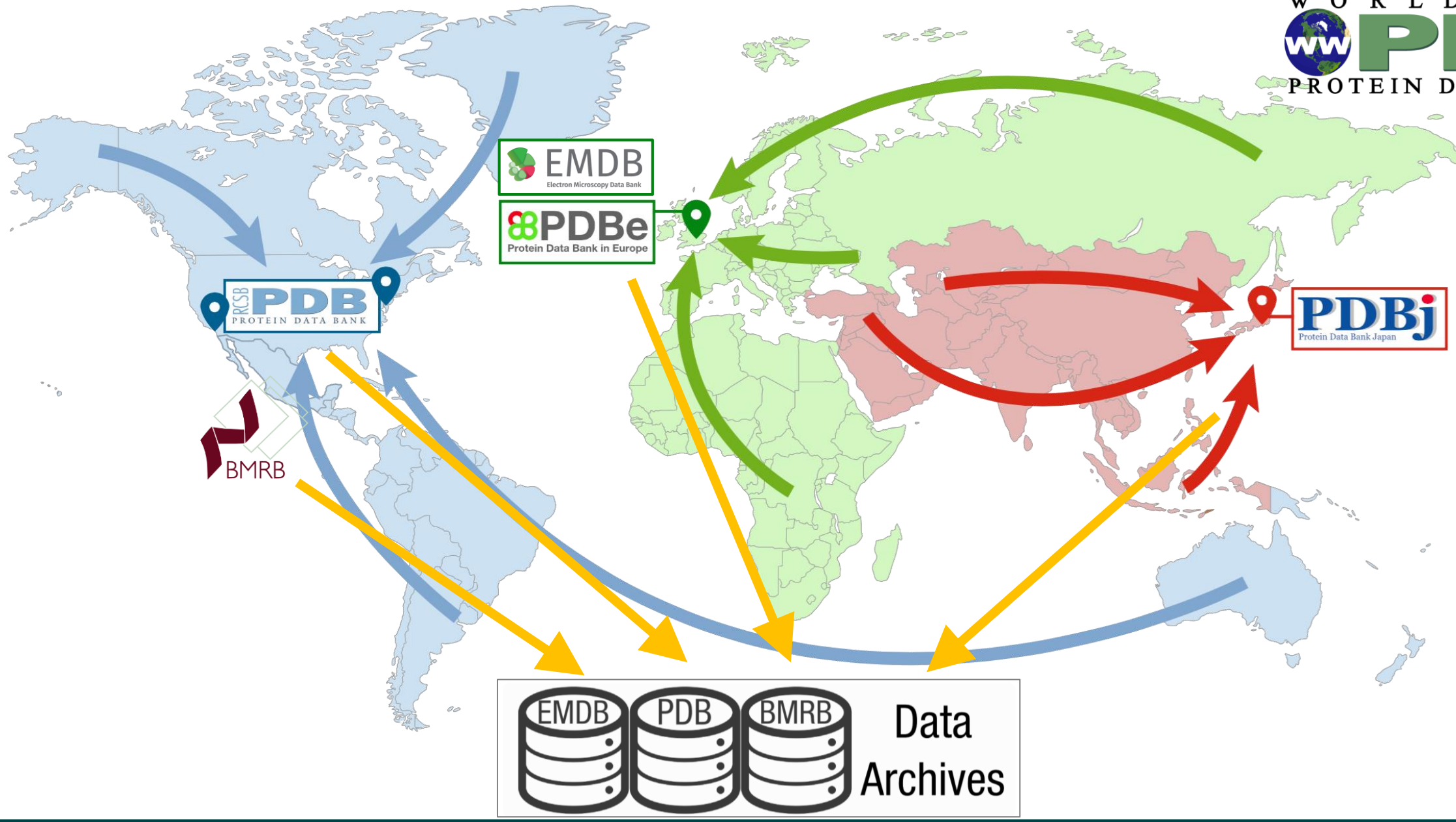
Bound ligands

Methods of solving the structures

Method	Total	2020
X-ray crystallography	89%	82%
NMR Spectroscopy	7.5%	3.5%
Electron microscopy	3.5%	15%



https://youtu.be/CpNb17k4oXI

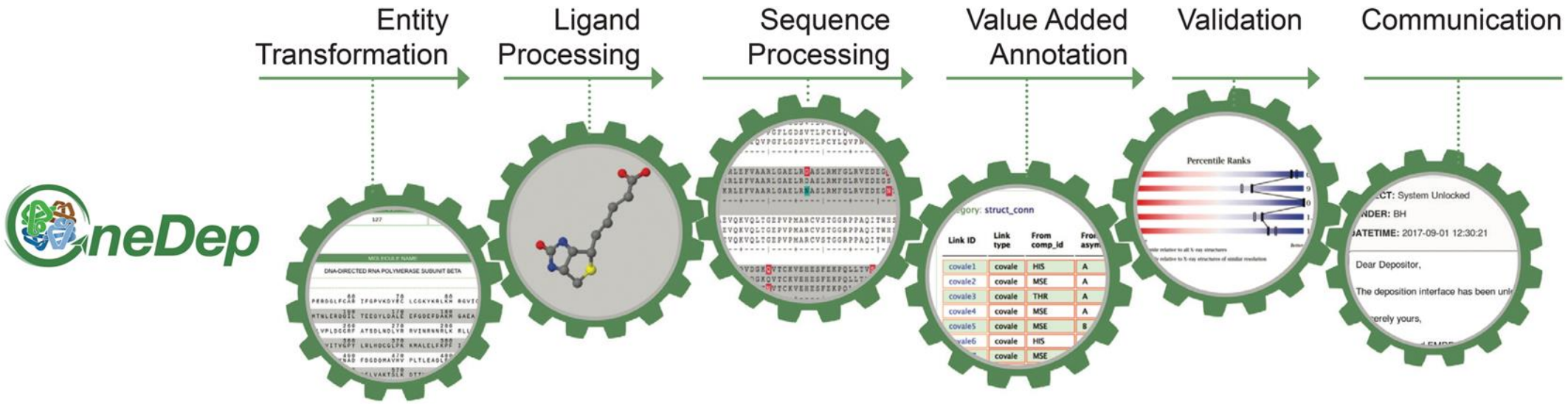


PDBe.org/API

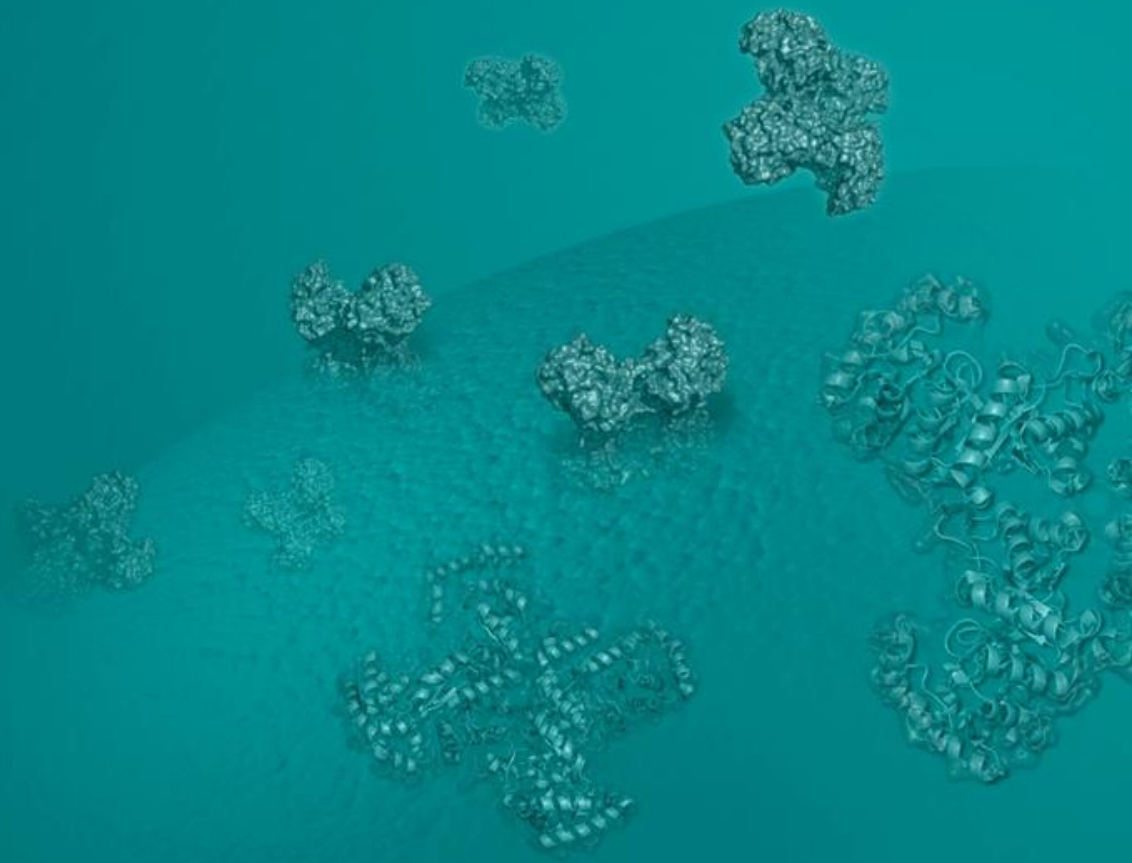
EMBL-EBI

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



PDBe: adding more functional data



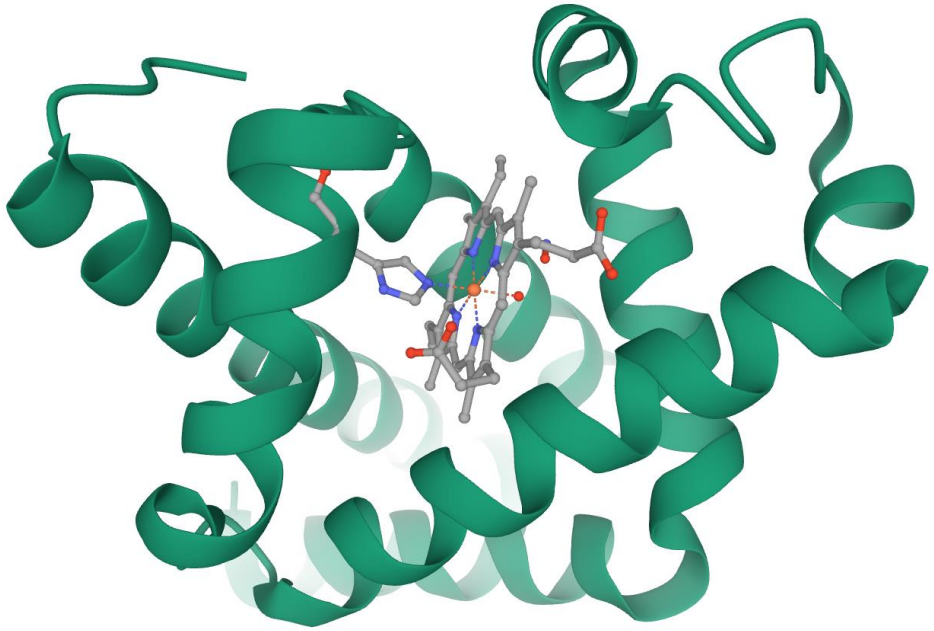
EMBL-EBI

Metadata gives scientific meaning to coordinates

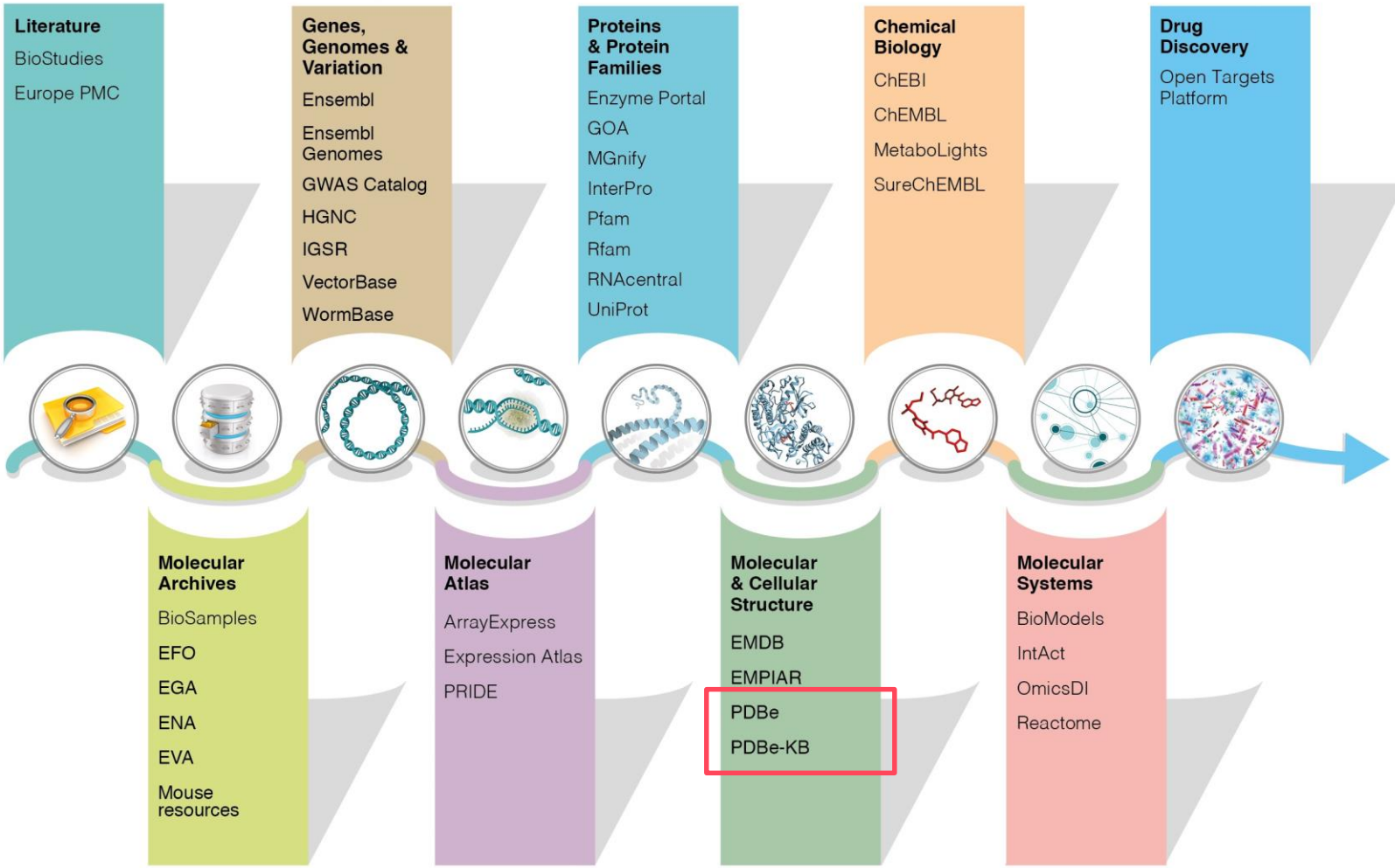
- “Coordinates by themselves just specify shape and are not necessarily of intrinsic biological value, unless they can be related to other information.”

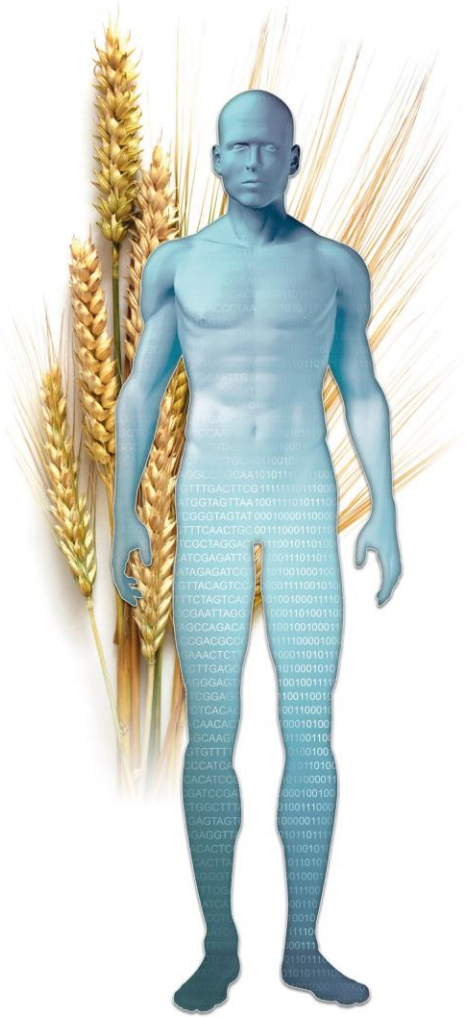
  Gerstein (2000) *Nature Structural Biology*, 7 960-963

- Some metadata is added in deposition process
  - Helps to link this to even more data after release
  - Gives more biological and chemical meaning to the structures in the PDB



At the heart of EMBL-EBI resources

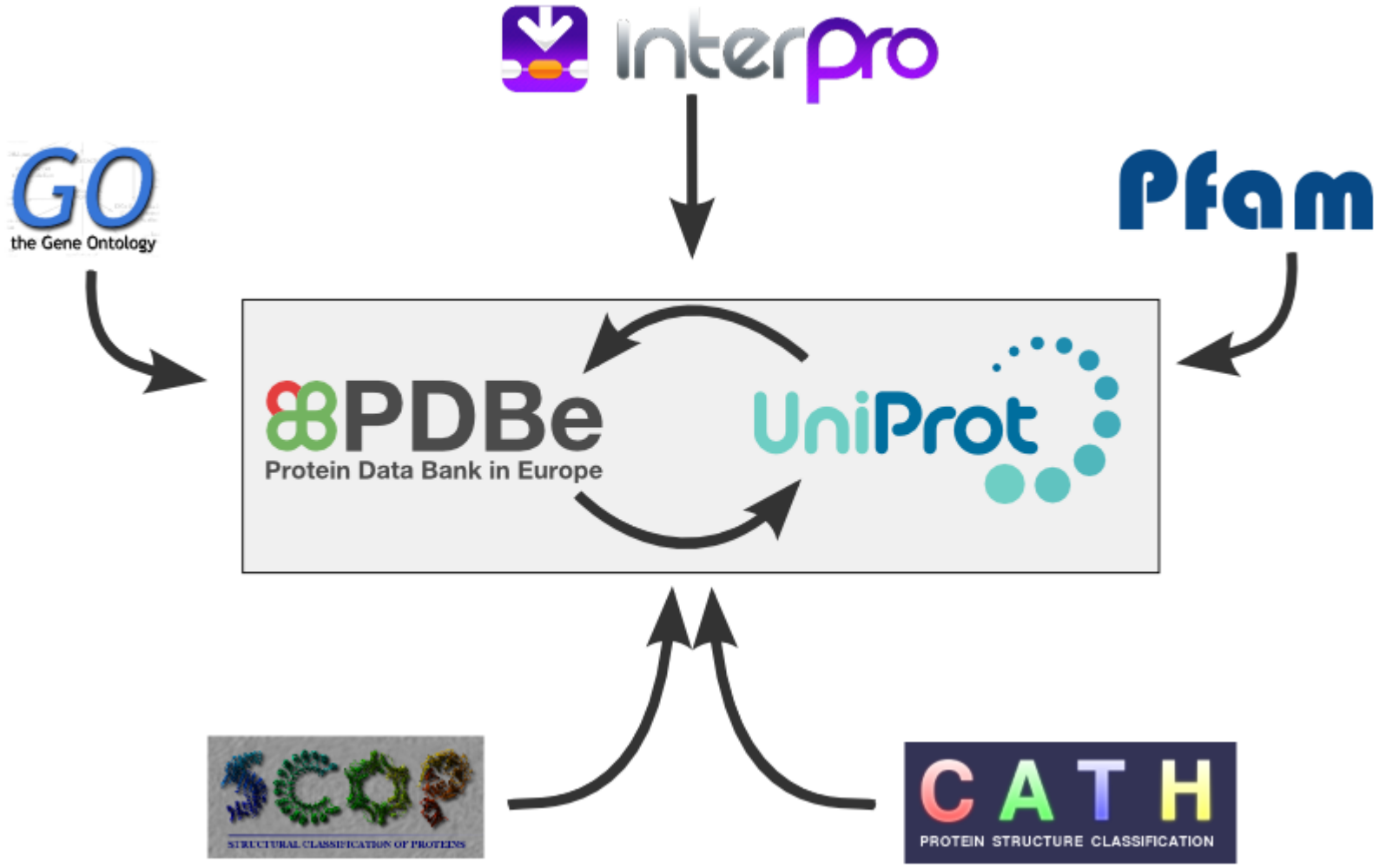




PDBe.org/API

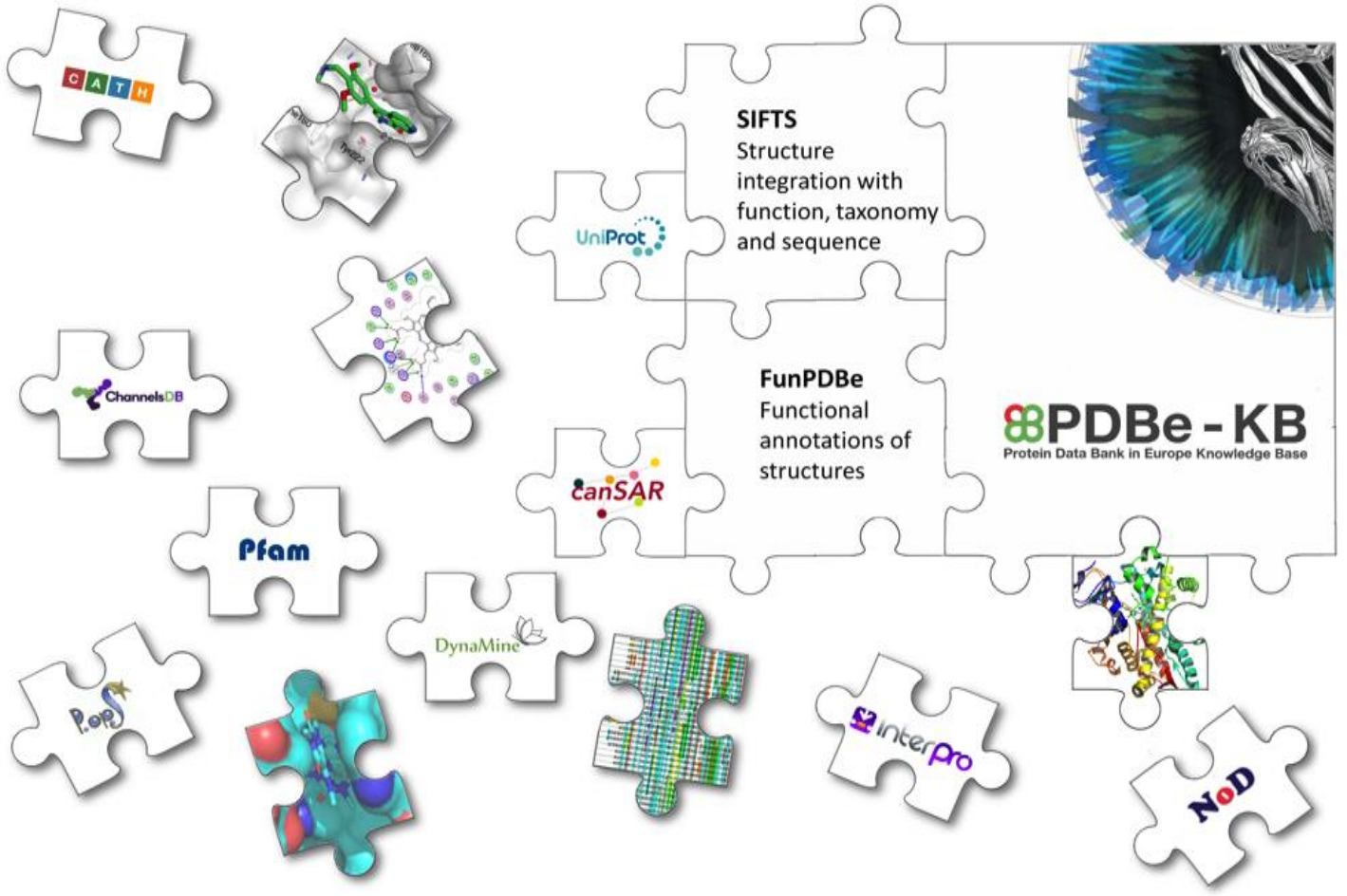
EMBL-EBI

SIFTS - Structure Integration with Function, Taxonomy and Sequence



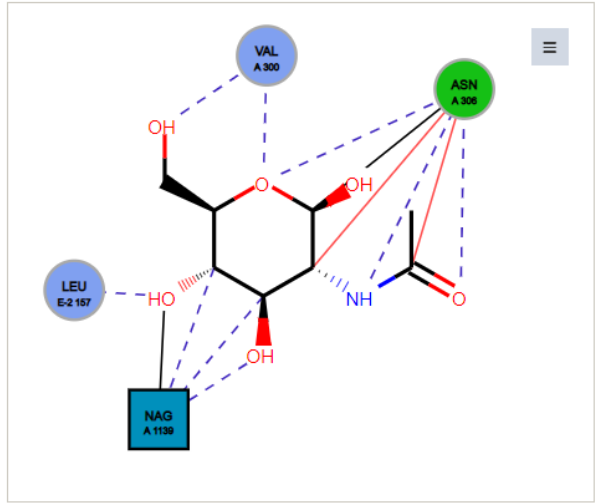
PDBe-KB: adding more functional data

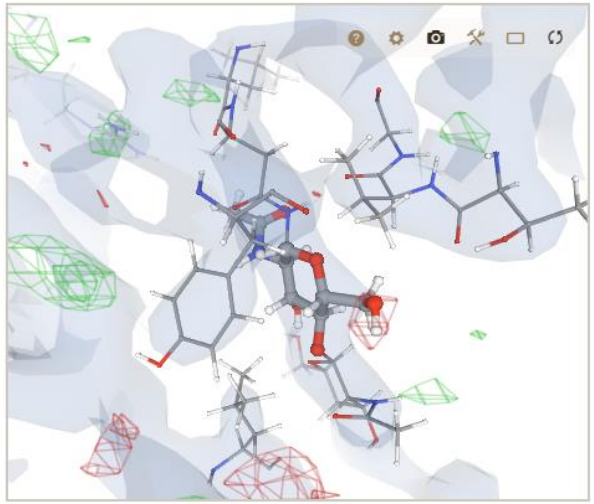
- PDBe-KB consortium
  - Structural and functional annotations for macromolecular structures
  - Collaboration between PDBe-KB and world-leading providers of structural bioinformatics data
- Includes data on domains, variation, drug-binding and much more
- Combined with SIFTS, vastly increases value of PDB data



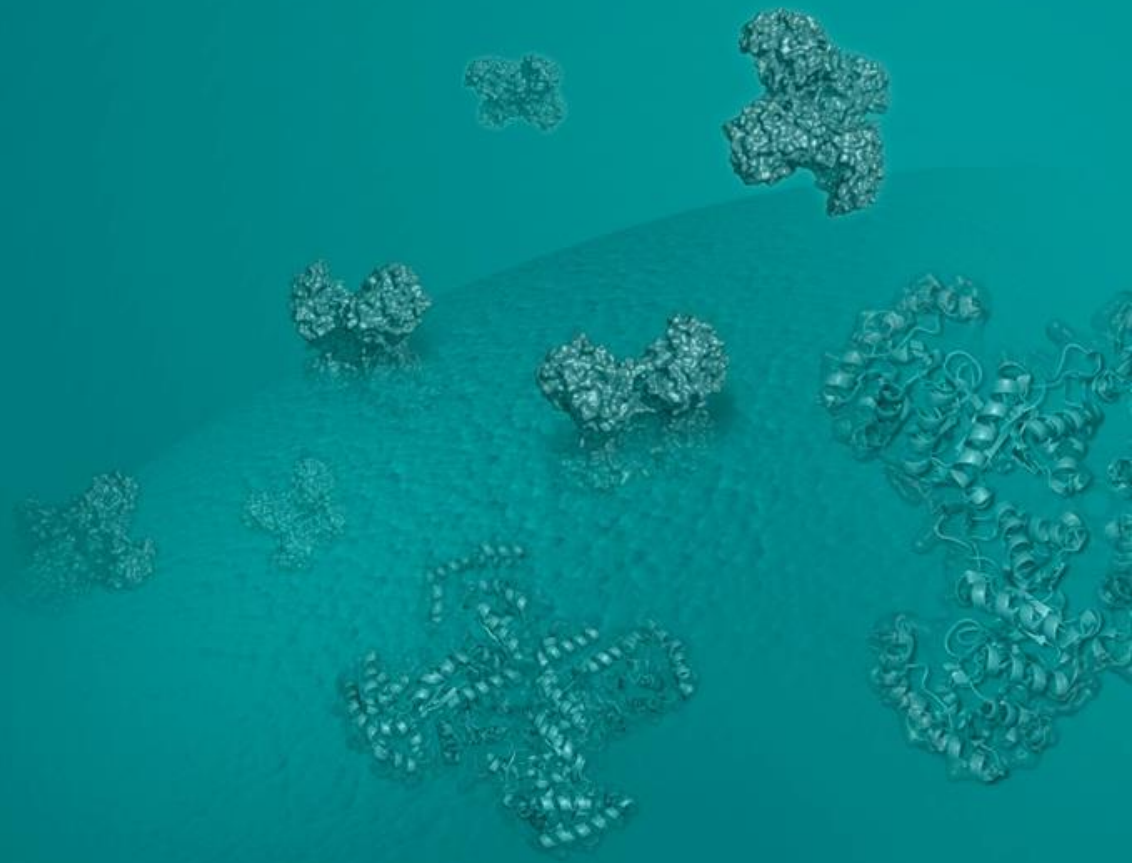
Additional chemistry information

- Add functional annotations for ligands in PDB
  - E.g. Drugbank links, Cofactor annotation
  - Embedded into search and pages
- PDBe interactions software pipeline
  - Calculate and define interactions
  - Generate interactive visualisations
- Software comprising the pipeline available in github
  - Will be discussed in more detail in a future webinar



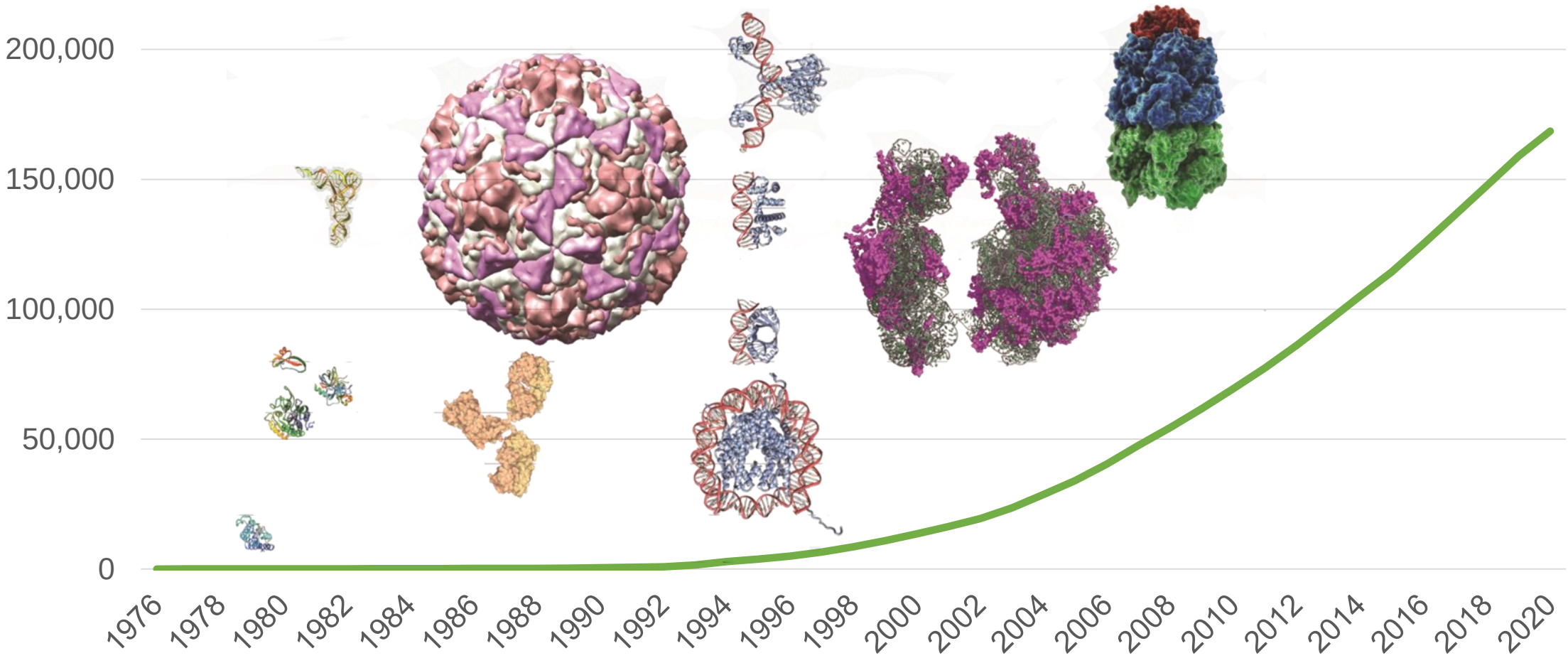


Why use PDBe APIs?



EMBL-EBI

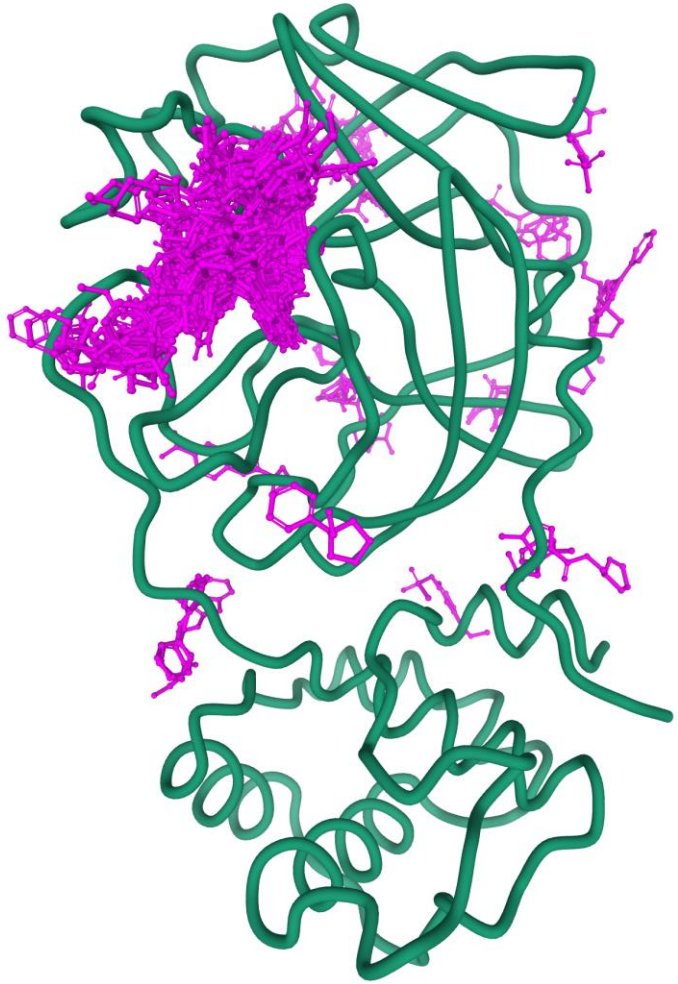
Growth of data in the PDB



In May 2014, we passed the 100,000 milestone. As of today, ~170,000 structures

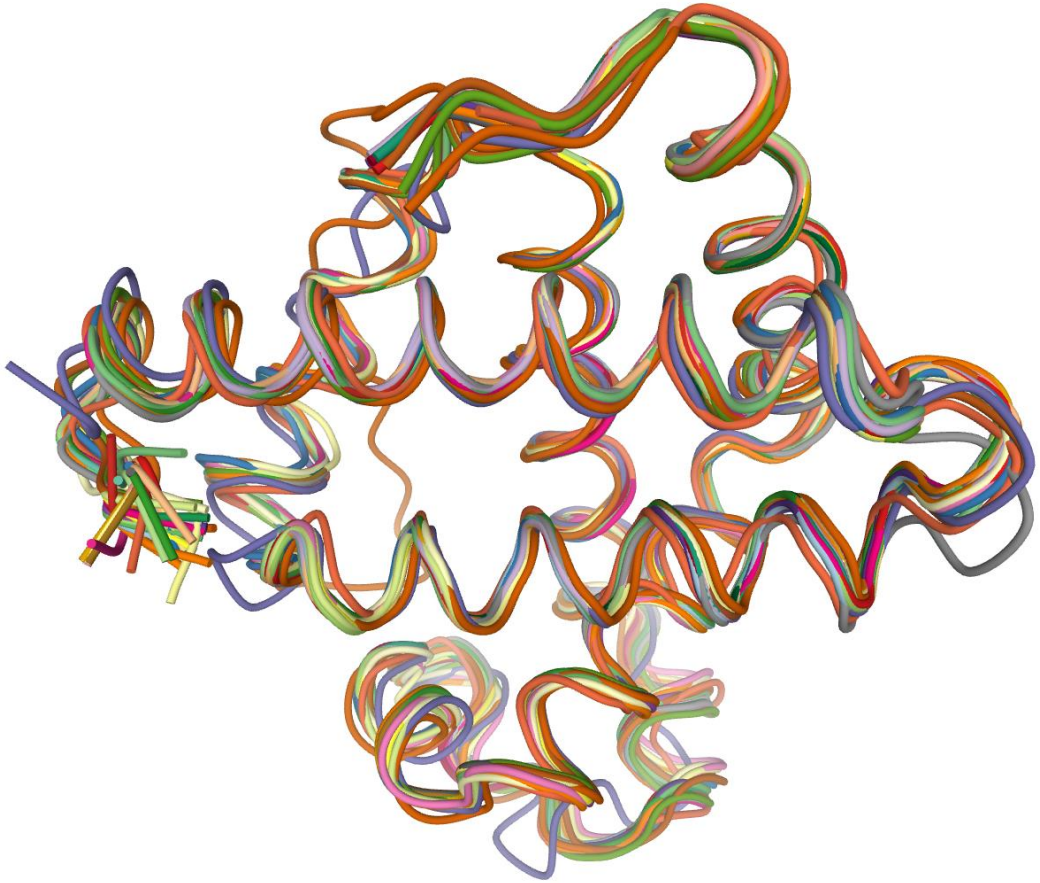
Redundancy of protein structures

- Though 170k structures, smaller number of proteins
  - Around 50k Uniprot accessions (unique organisms)
  - Even more overlap across multiple organisms
  - Also a high level of structural similarity
- Can have multiple structures with different conditions
  - Multiple ligands bound
  - Different experimental techniques and resolutions
  - Complexes with other macromolecules
- Need to consider a broad range of structures to gain maximum functional insight
  - The API can be vital in this



Large scale data analysis

- Large numbers of structures
  - Unfeasible to access individual entries
  - Programmatic access allows collation of huge amounts of data
- Repetition
  - Repeating data analysis at a later date
  - Doing the same analysis with any input
- Getting data for software pipelines
- Visualisation of data using web components
  - Serve data on demand to users



Getting coordinates and experimental data

- API does not provide coordinates
  - High volume of data – not in database
  - However, API helps direct access of coordinates
- FTP (PDBe.org/FTP)
  - Can get coordinate and structure factor files
- Download service
  - Available at www.ebi.ac.uk/pdbe/download/docs
- Model Server and Volume Server
  - Allows streaming of specific data using Mol*

Residues Inside a Sphere `/ambientResidues`

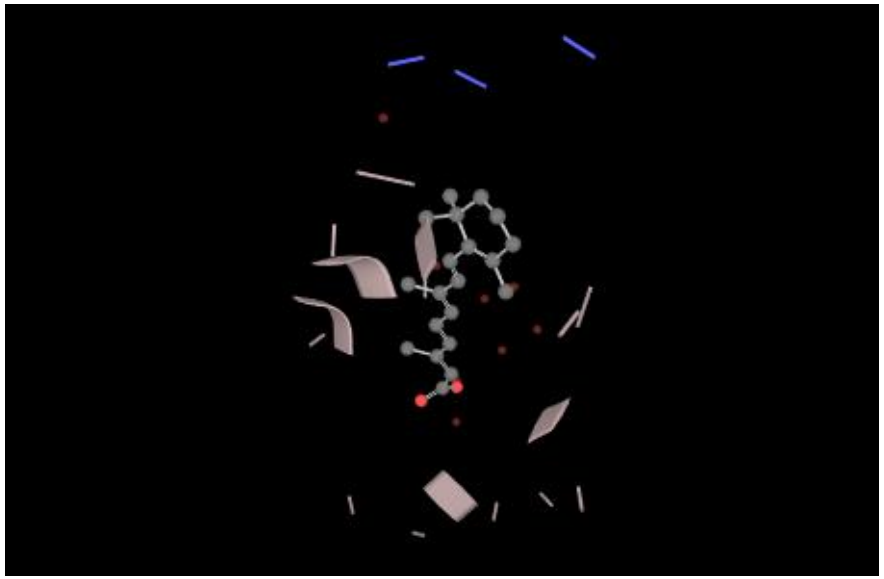
Identifies all residues within the given radius from the source residue.

Example

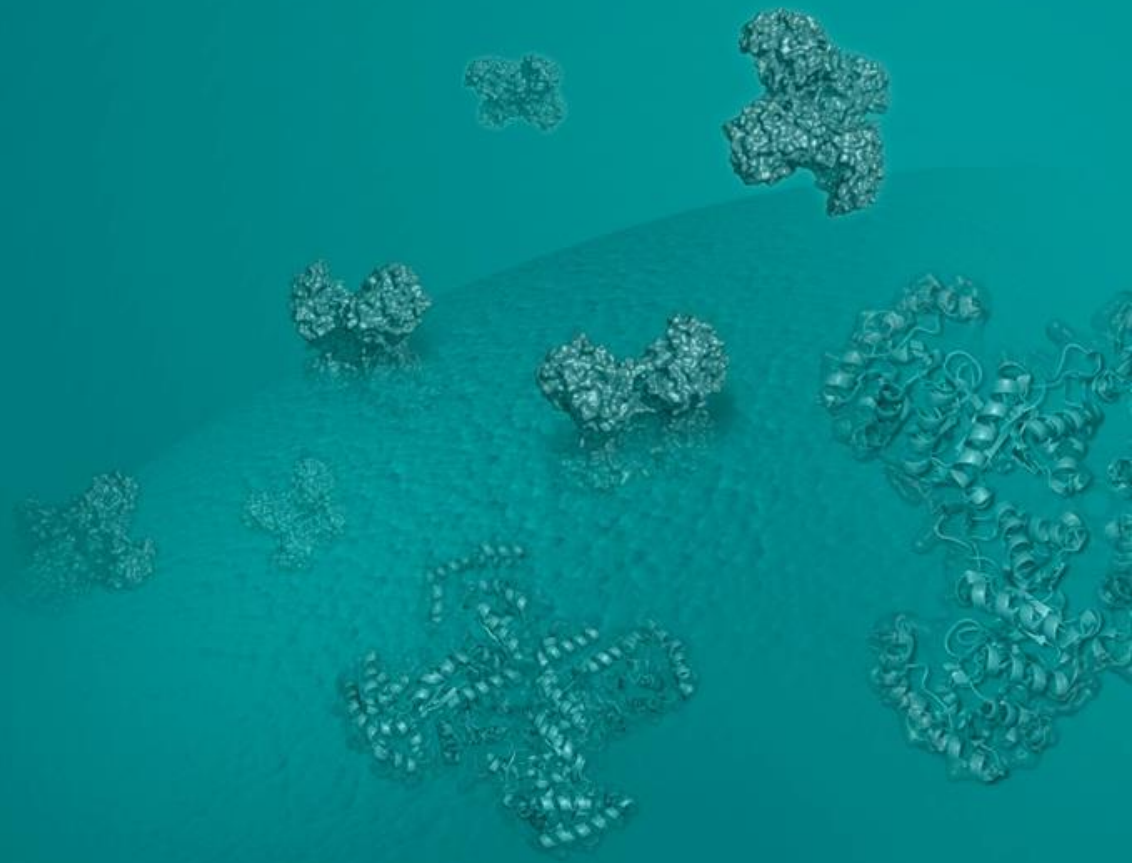
`/1cbs/ambientResidues?authAsymId=A&authName=REA&authSeqNumber=200&radius=5`

Parameters

Name	Type	Default	Description
entityId	String		Corresponds to the '_entity.id' or '*.label_entity_id' field, depending on the context.
asymId	String		Corresponds to the '_atom_site.label_asym_id' field.
authAsymId	String		Corresponds to the '_atom_site.auth_asym_id' field.
name	String		Residue name. Corresponds to the '_atom_site.label_comp_id' field.
authName	String		Author residue name. Corresponds to the '_atom_site.auth_comp_id' field.
insCode	String		Corresponds to the '_atom_site.pdbx_PDB_ins_code' field.
seqNumber	Integer		Residue seq. number. Corresponds to the '_atom_site.label_seq_id' field.
authSeqNumber	Integer		Author residue seq. number. Corresponds to the '_atom_site.auth_seq_id' field.
radius	Float	5	Value in Angstroms.



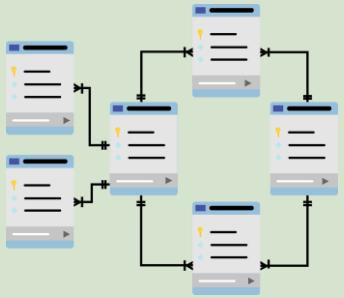
What type of APIs do the PDBe offer?



EMBL-EBI

Querying

- Solr-based **query system** of data in the relational database
- Underlies the **PDBe search pages**
- **Large number of fields** available for searching
- By default, data **returned per entity** (i.e. unique molecule)
- Lots of data returned, but can also extract IDs and input to relational or graph APIs



Entry-based API

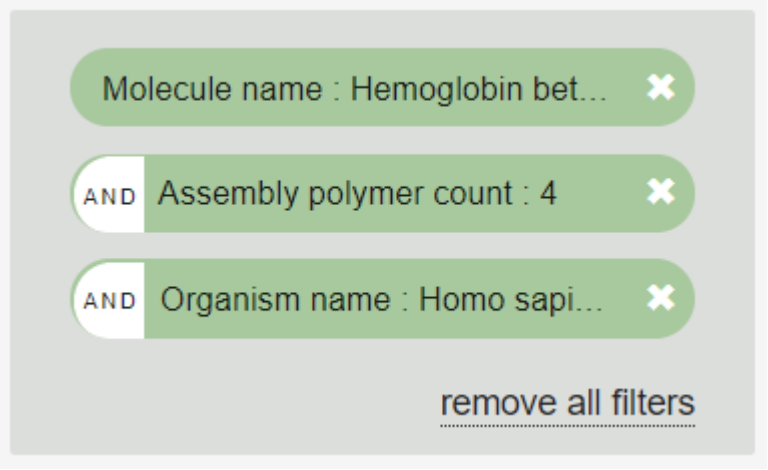
- Created from the Oracle **relational database**
- Used to populate data on **PDBe entry pages**
- Contains **entry-specific** and ligand-specific data
- Includes **mapping information** through SIFTS
- Specific calls related to **validation** from the wwPDB validation report

Aggregated API

- Created from our newer Neo4J **graph database**
- Used to populate data on the **PDBe-KB aggregated views**
- **Provides aggregated data…**
- **Varied inputs** – get information **per residue** or per atom in ligand
- Incorporates additional data from **PDBe-KB partners**

Querying using programmatic access

- Based on a Solr query system
  - Allows creation of complex queries with Solr syntax
- Underlies our PDBe search pages
  - URLs at these pages contain the API queries
  - Can easily check a query on the UI and then implement programmatically
  - But, even more fields available through the API
- Returns data per entity (i.e. per unique molecule)
  - Can filter fields to a subset of data required
  - Can then input into entry-based or aggregated APIs



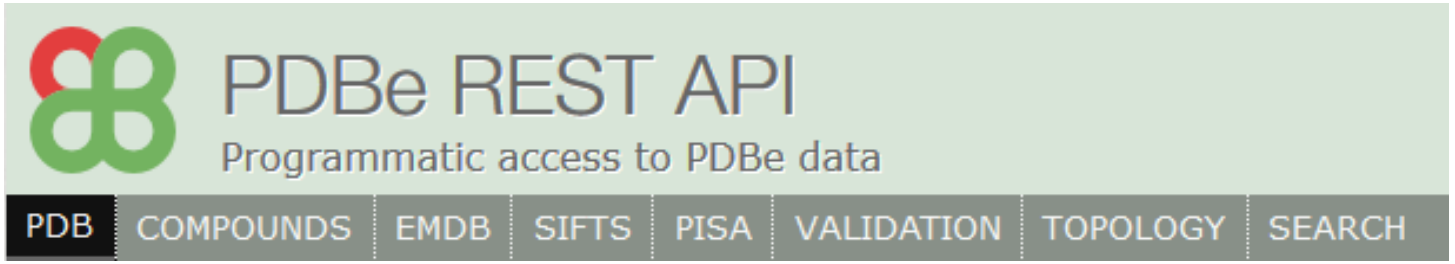
q_all_molecule_names = Hemoglobin beta chain

q_assembly_type = 4

q_organism_name = Homo sapiens

What can the Entry-based API give you?

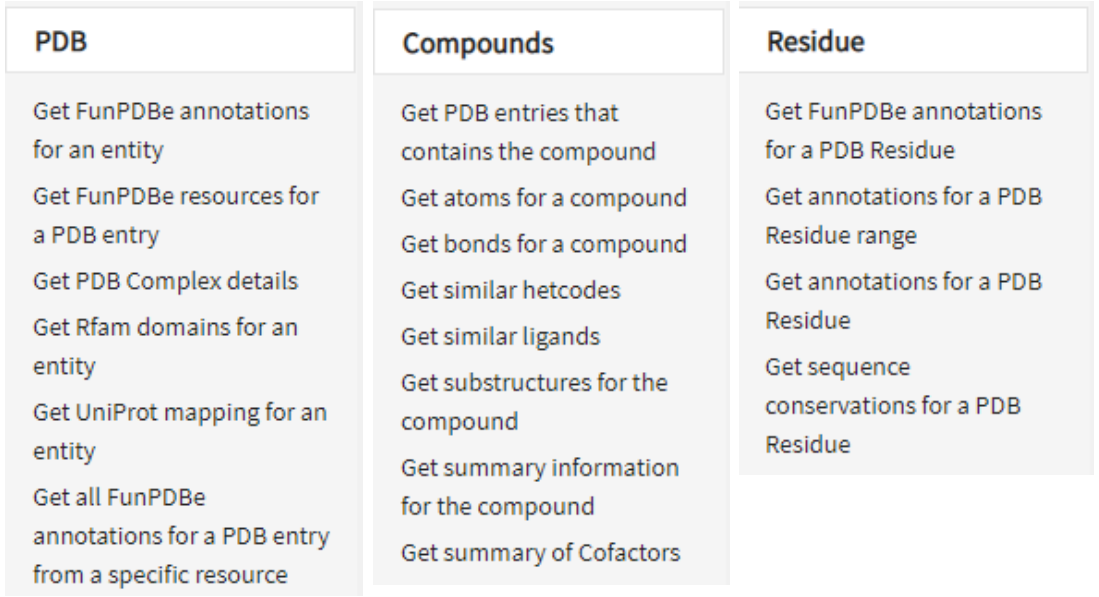
- These REST calls are based on different subsets of data
  - These are grouped for ease of use



- Available on the documentation page
  - Thorough set of example calls
  - Give URL and output JSON on the page
- Querying using the search API also available from this page
  - Can run example search queries within the page

What can the Aggregated API give you?

- More varied inputs
  - Includes per-entity and per-residue calls
- More data from FunPDBe and PDBe-KB partners
- Other data not in the entry-based API
  - E.g. atom-level ligand interaction data



PDB

- Get FunPDBe annotations for an entity
- Get FunPDBe resources for a PDB entry
- Get PDB Complex details
- Get Rfam domains for an entity
- Get UniProt mapping for an entity
- Get all FunPDBe annotations for a PDB entry from a specific resource

Compounds

- Get PDB entries that contains the compound
- Get atoms for a compound
- Get bonds for a compound
- Get similar hetcodes
- Get similar ligands
- Get substructures for the compound
- Get summary information for the compound
- Get summary of Cofactors

Residue

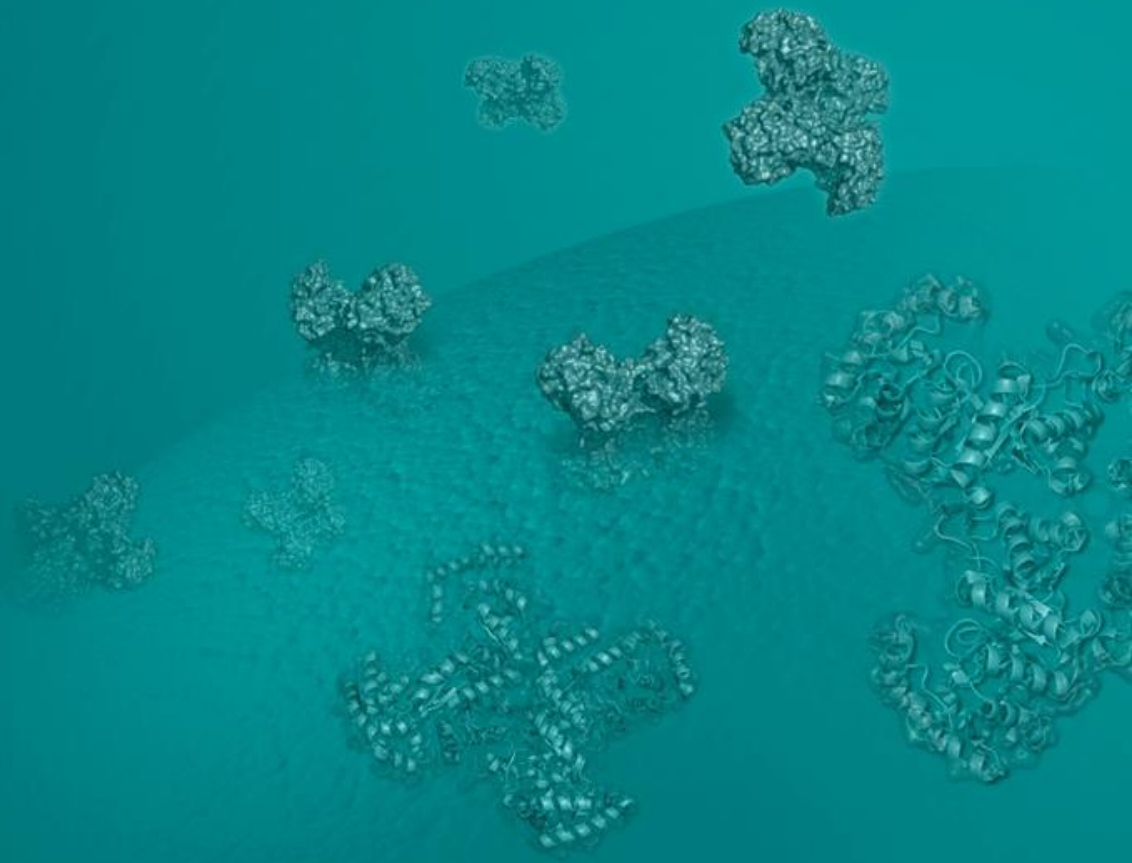
- Get FunPDBe annotations for a PDB Residue
- Get annotations for a PDB Residue range
- Get annotations for a PDB Residue
- Get sequence conservations for a PDB Residue

Residue - Get annotations for a PDB Residue

Get mappings (as assigned by the SIFTS process) for a PDB Residue to UniProt, Pfam, InterPro, CATH, SCOP, IntEnz, GO, Ensembl and HMMER accessions (and vice versa).

```
https://www.ebi.ac.uk/pdbe/graph-api/residue_mapping/:pdbId/:entityId/:residueNumber
```

How do we use the APIs at PDBe?



EMBL-EBI

Finding validation information for PDB entry

A list of outlier types found in residues.

http://www.ebi.ac.uk/pdbe/api/validation/residuewise_outlier_summary/entry/:pdbid

A residue can have many types of geometric or experimental-data-based outliers. This call lists all kinds of outliers found in a residue. For residues with no recorded outlier, there is no information returned.

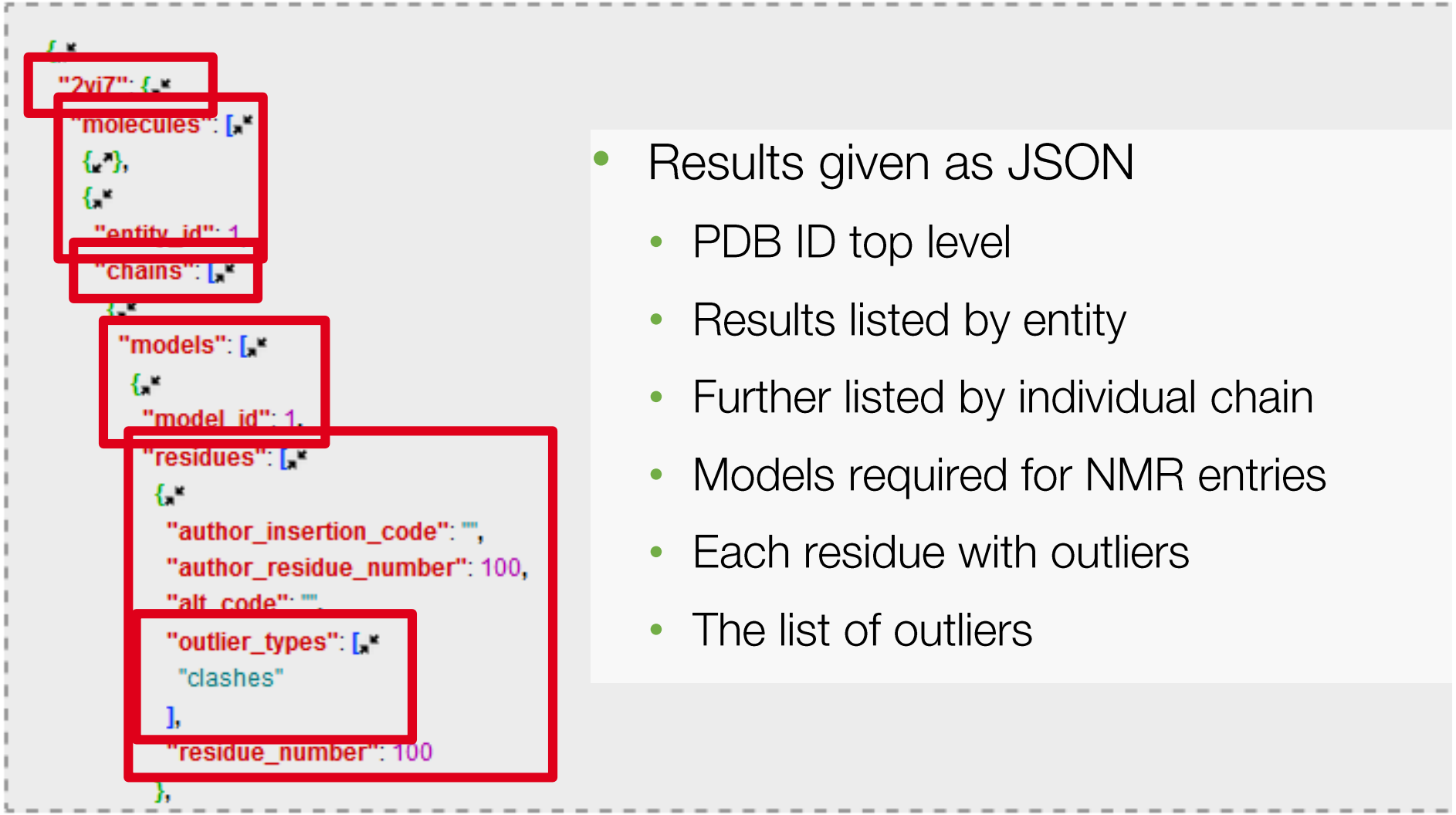
pdbid	2yi7	String	4-character PDB id code. Only released ids are allowed.
postdata		String	POST data should contain one or more comma-separated PDB entry identifiers leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

☑Quotes RunCall Select Expand Collapse 2+ 3+

GET : http://www.ebi.ac.uk/pdbe/api/validation/residuewise_outlier_summary/entry/2yi7


```
{
 "2yi7": {
  "molecules": [
   {
    "entity_id": "1",
    "chains": [
     {
      "models": [
       {
        "model_id": 1,
        "residues": [
         {
          "author_insertion_code": " ",
          "author_residue_number": "177",
          "alt_code": " ",
          "outlier_types": [
           "RSRZ"
          ],
          "residue_number": 177
         },
         {
```

Finding validation information for PDB entry



- Results given as JSON
  - PDB ID top level
  - Results listed by entity
  - Further listed by individual chain
  - Models required for NMR entries
  - Each residue with outliers
  - The list of outliers

Mol* – powered by PDBe REST API

- Lightweight structure viewer at PDBe
- Uses API to get structure and annotation data
  - Including residue-level validation information
- Gets coordinates/maps from model/volume servers


```
"author_insertion_code": "",
"author_residue_number": 120,
"alt_code": "",
"outlier_types": [
 "clashes"
```

```
"author_insertion_code": "",
"author_residue_number": 156,
"alt_code": "",
"outlier_types": [
 "bond_lengths",
 "planes"
```

```
"author_insertion_code": "",
"author_residue_number": 191,
"alt_code": "",
"outlier_types": [
 "clashes",
 "bond_lengths",
 "bond_angles",
 "sidechain_outliers"
```

Example: Finding structure mappings to Uniprot ID

UniProt - Get PDB structure mappings for a UniProt accession

This call provides details on mapped PDB structures for a UniProt accession.

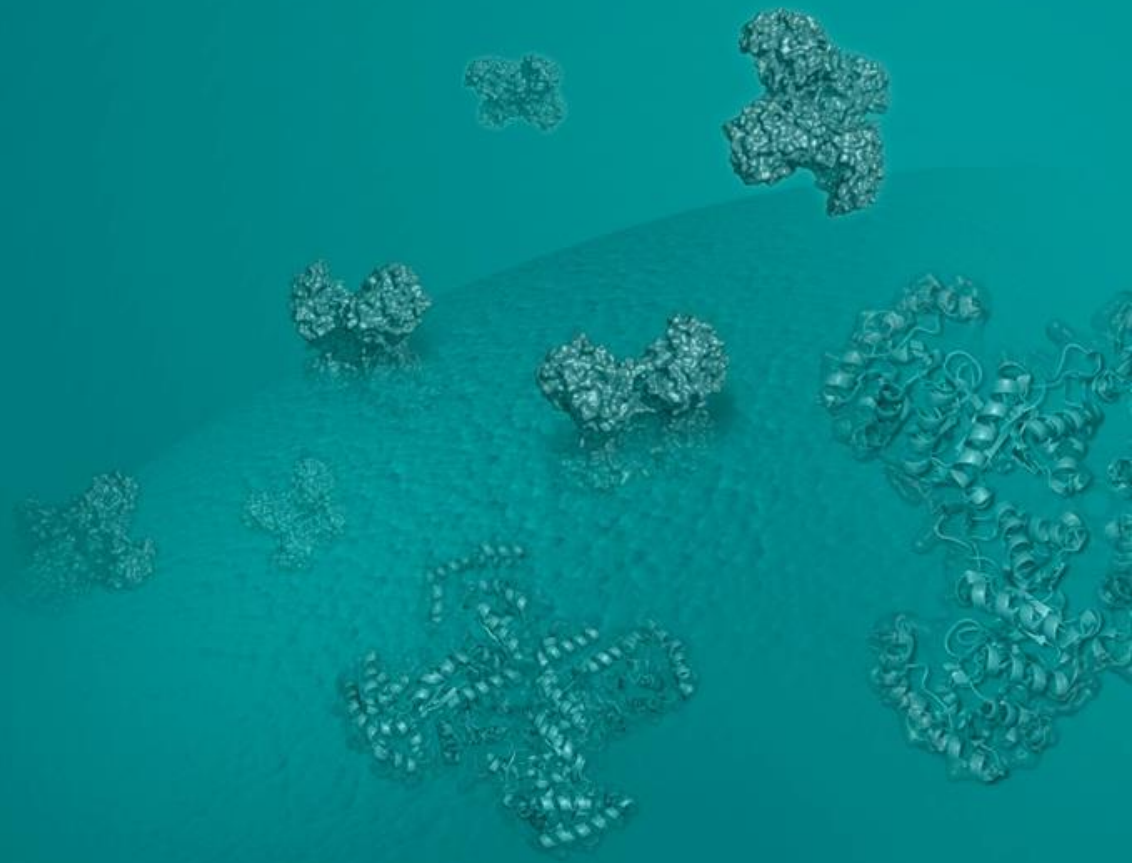
```
https://www.ebi.ac.uk/pdbe/graph-api/uniprot/unipdb/:accession
```

Example success response JSON:


```
{
    "P07550": {
        "sequence":
"MGQPGNGSAFLLAPNGSHAPDHDVTQERDEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVMGLAVVPFGAAHILMKMWTFGNFWCEFWTSI
        "length": 413,
        "dataType": "UNIPDB",
        "data": [
            {
                "name": "2rh1",
                "accession": "2rh1",
                "dataType": "bestChain",
                "entityId": 1,
                "bestChainId": "A",
                "residues": [
                    {
                        "startIndex": 1,
                        "endIndex": 28,
```

Example: Finding structure mappings to Uniprot ID

Outline of future webinars in this series



EMBL-EBI

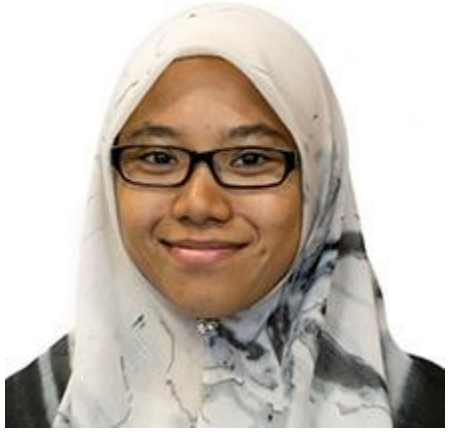
PDBe API webinar series

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

John Berrisford

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

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

Sreenath Nair

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

Lukas Pravda

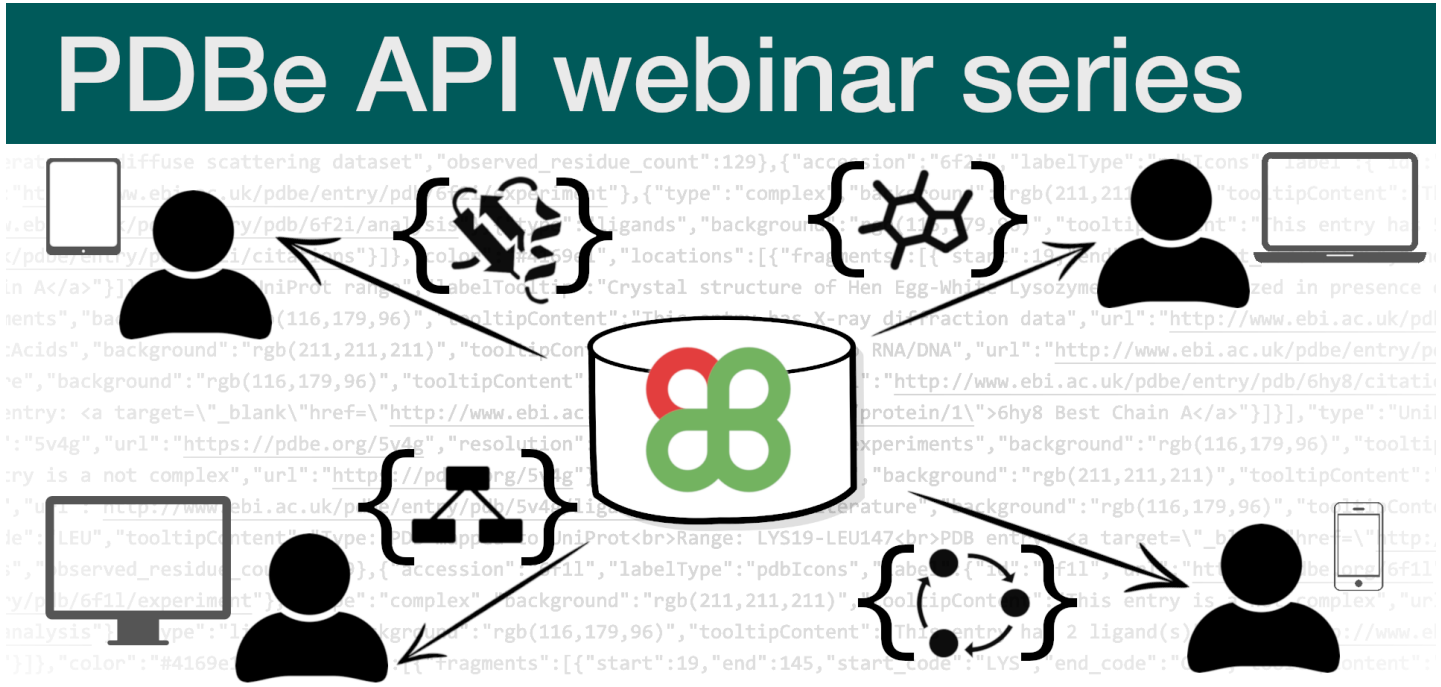
PDBe API webinar series

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

Mandar Deshpande

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

David Armstrong

PDBe
Protein Data Bank in Europe

pdbhelp@ebi.ac.uk
proteindatabank
@PDBeurope
proteindatabank
pdbeurope
pdbart

EMBL-EBI