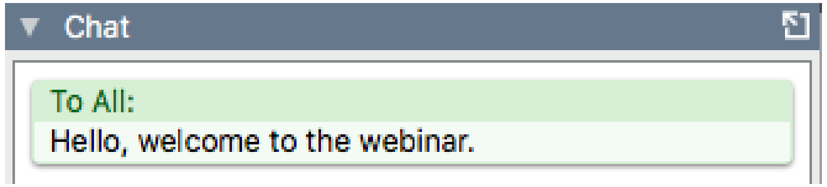
Welcome - webinar instructions

- 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



To All:
Hello, welcome to the webinar.

- Please complete the **feedback survey** which will launch at the end of the webinar

PDBe API webinar series:
Creating complex PDBe API queries

PDBe.org/API

Nurul Nadzirin

PDBe
Protein Data Bank in Europe

✉ pdbhelp@ebi.ac.uk

f proteindatabank

@PDBeurope

▶ proteindatabank

pdbeurope

pdbart

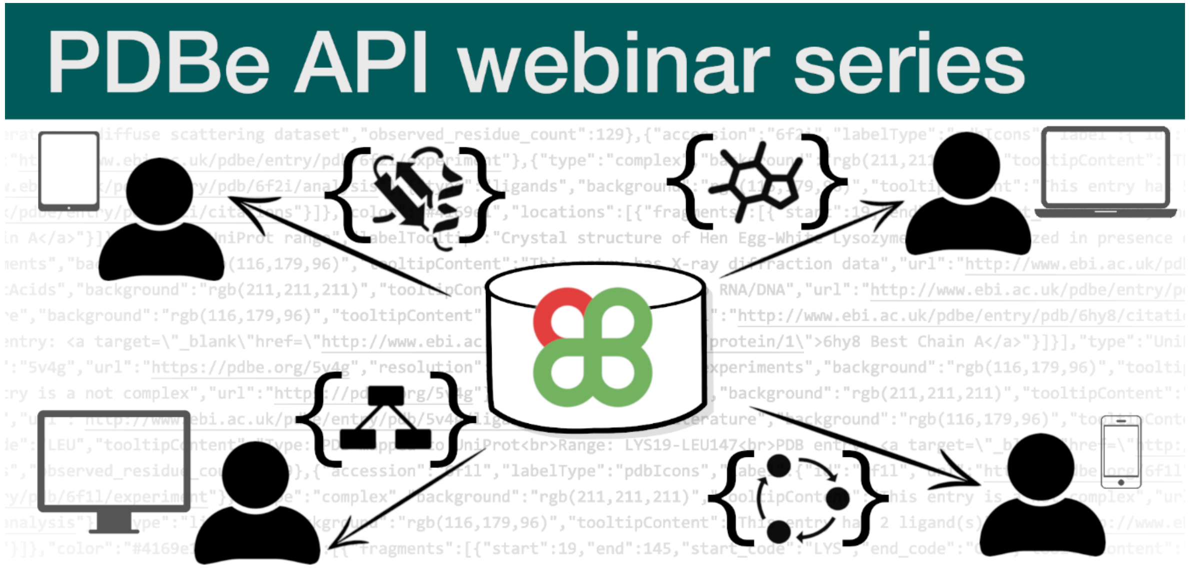
EMBL-EBI

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

To register for future webinars in the series

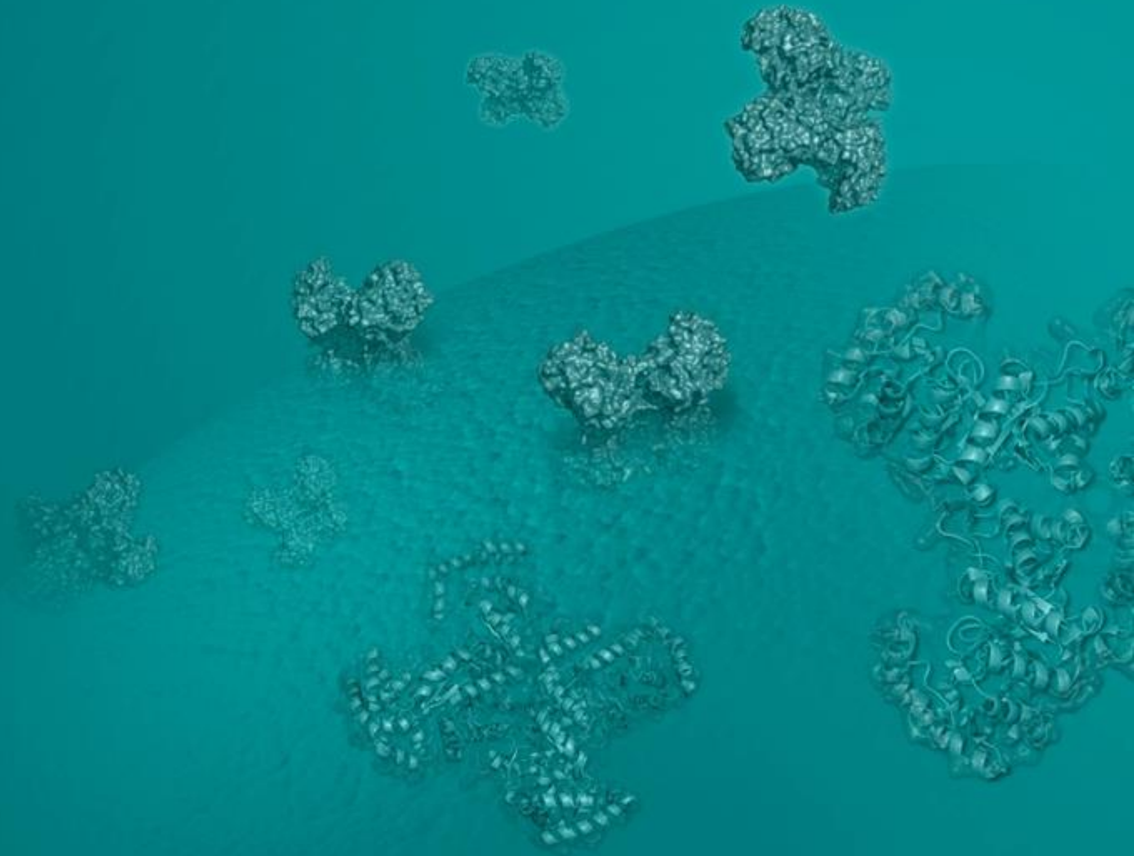
- Webinar videos, slides and interactive jupyter notebooks. To view the materials and sign up for the next in the series, visit https://pdbeurope.github.io/api-webinars/
- Or visit the PDBe events pages at PDBe.org/events
- Remember to register for each webinar individually!!!



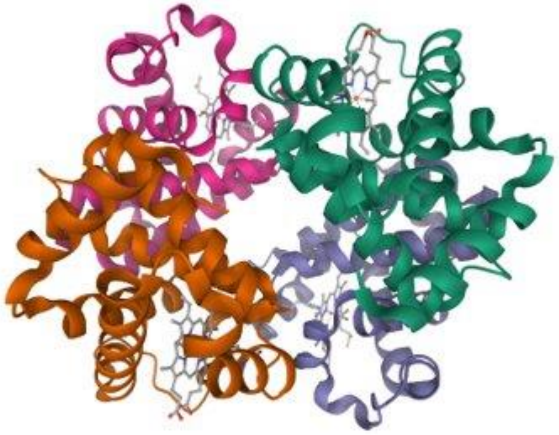
Summary of this webinar

- What type of data is in the PDB?
- The three types of PDBe APIs
- PDBe REST API: Entry-based API
- Categories of endpoints
- How PDBe uses the entry-based API
- Use cases

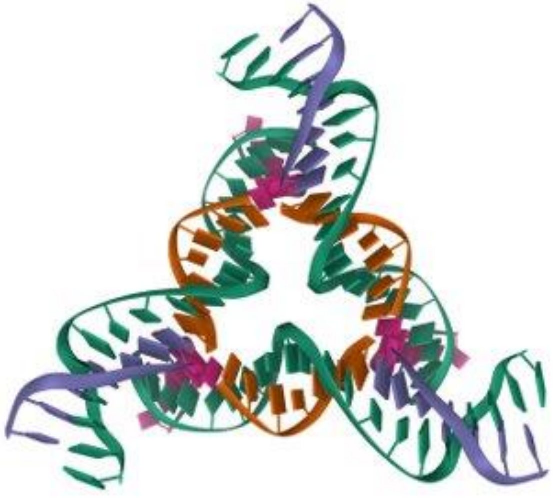
What type of data is in the PDB?



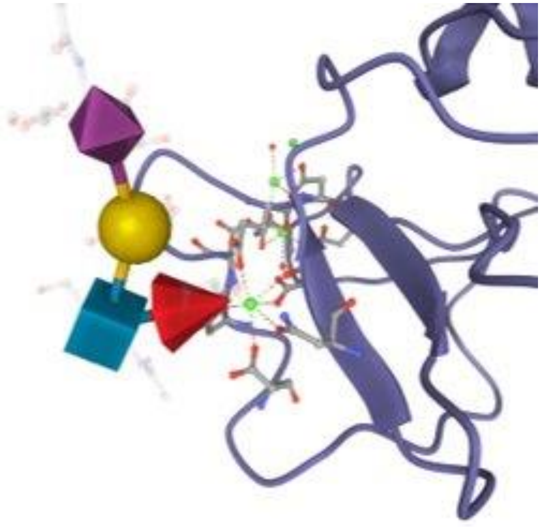
The Protein Data Bank (PDB) is an archive of experimentally determined 3-dimensional structures of biological macromolecules



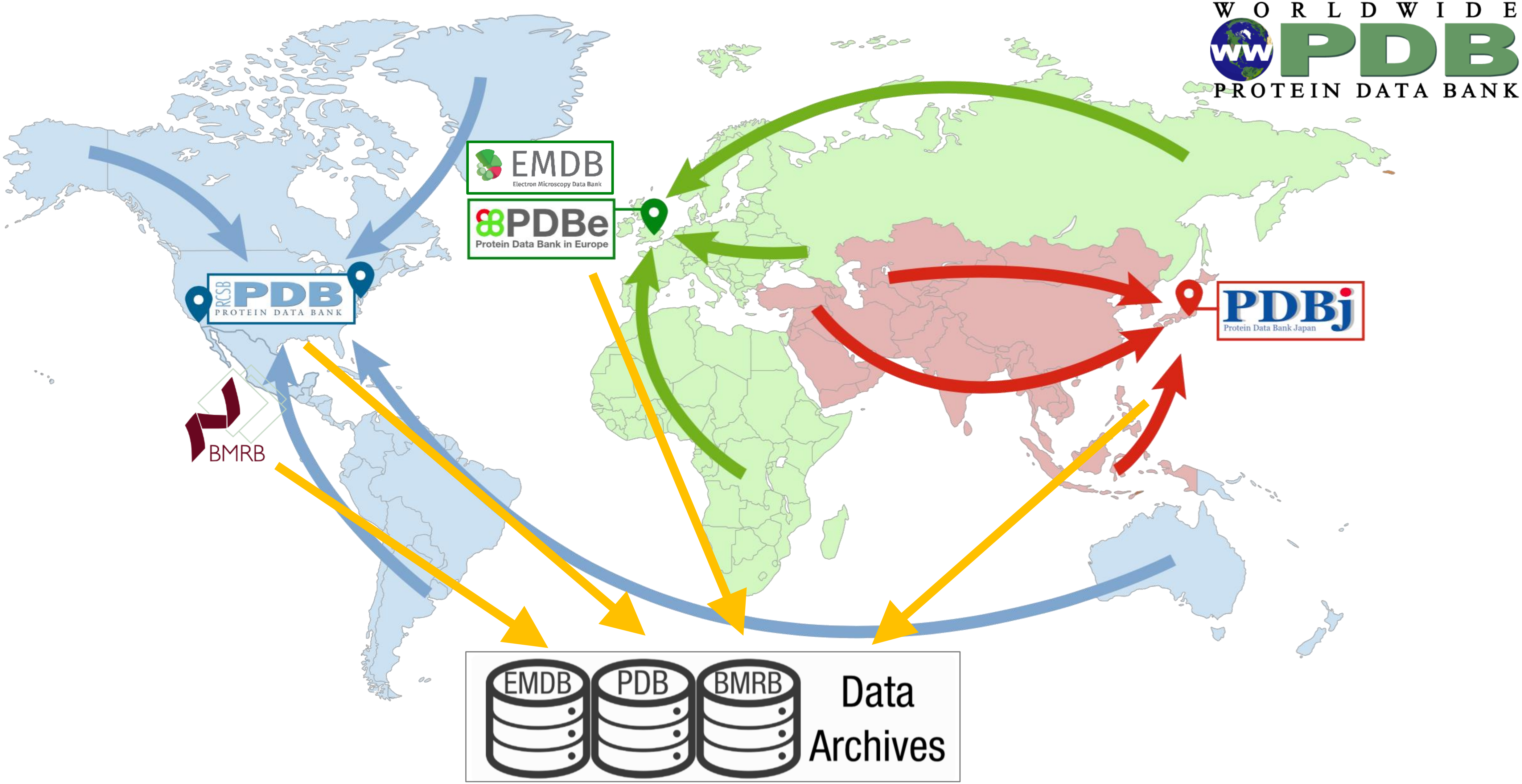
Proteins



DNA/RNA

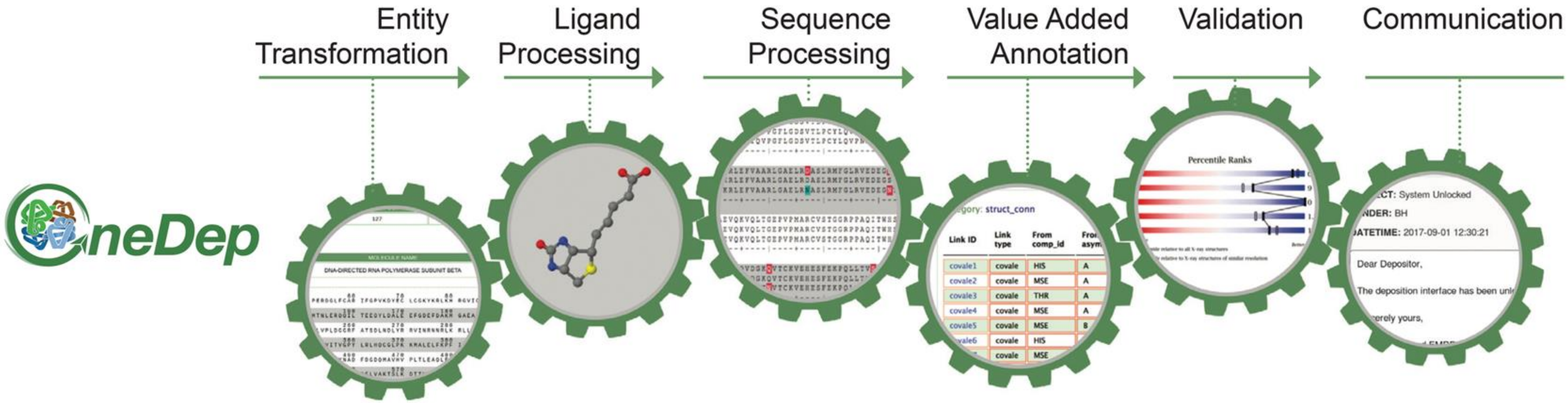


Bound ligands

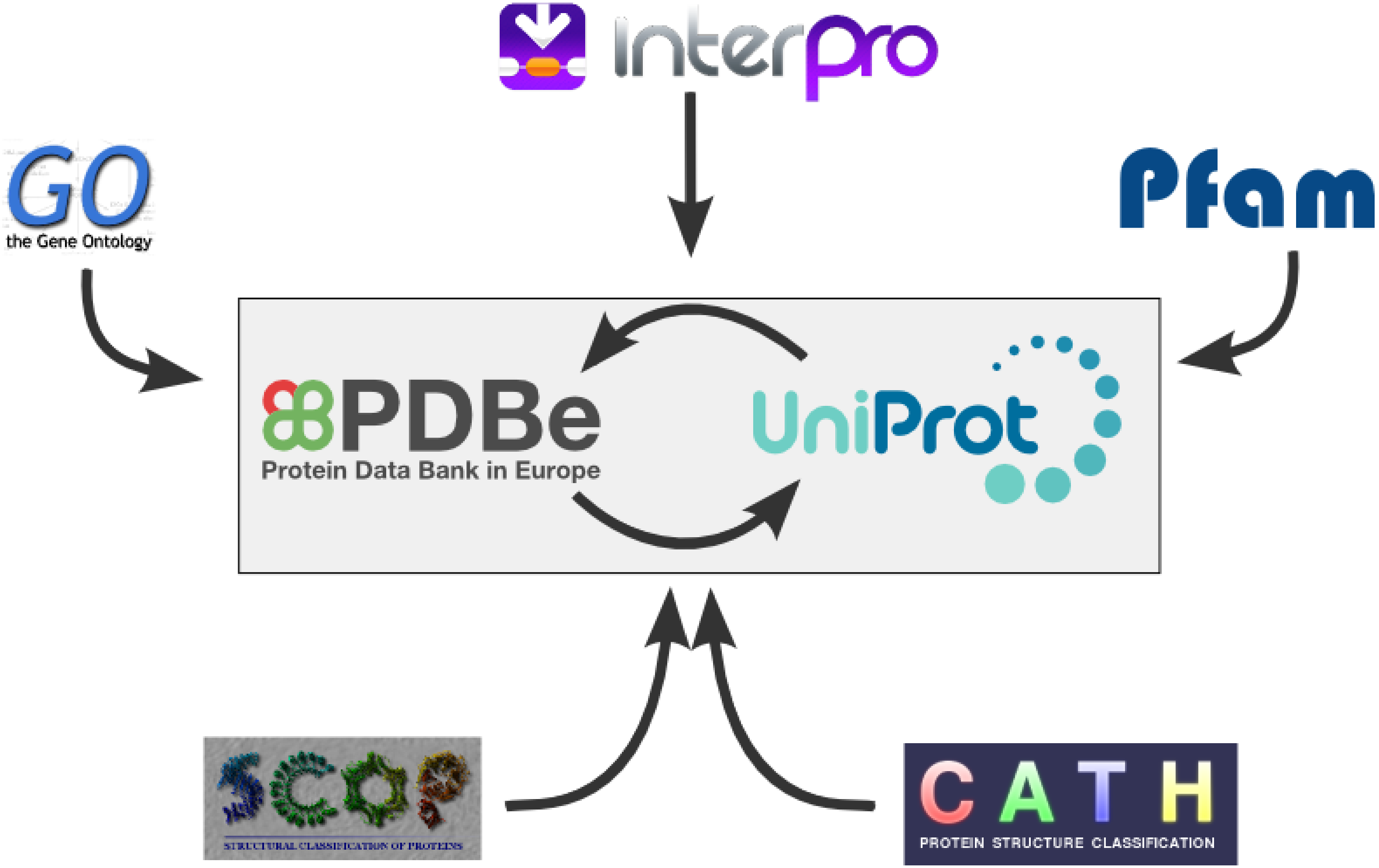

WORLDWIDE PDB PROTEIN DATA BANK
EMDB
Electron Microscopy Data Bank
PDBe
Protein Data Bank in Europe
RCSB PDB PROTEIN DATA BANK
BMRB
PDBj
Protein Data Bank Japan
EMDB
PDB
BMRB
Data Archives

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

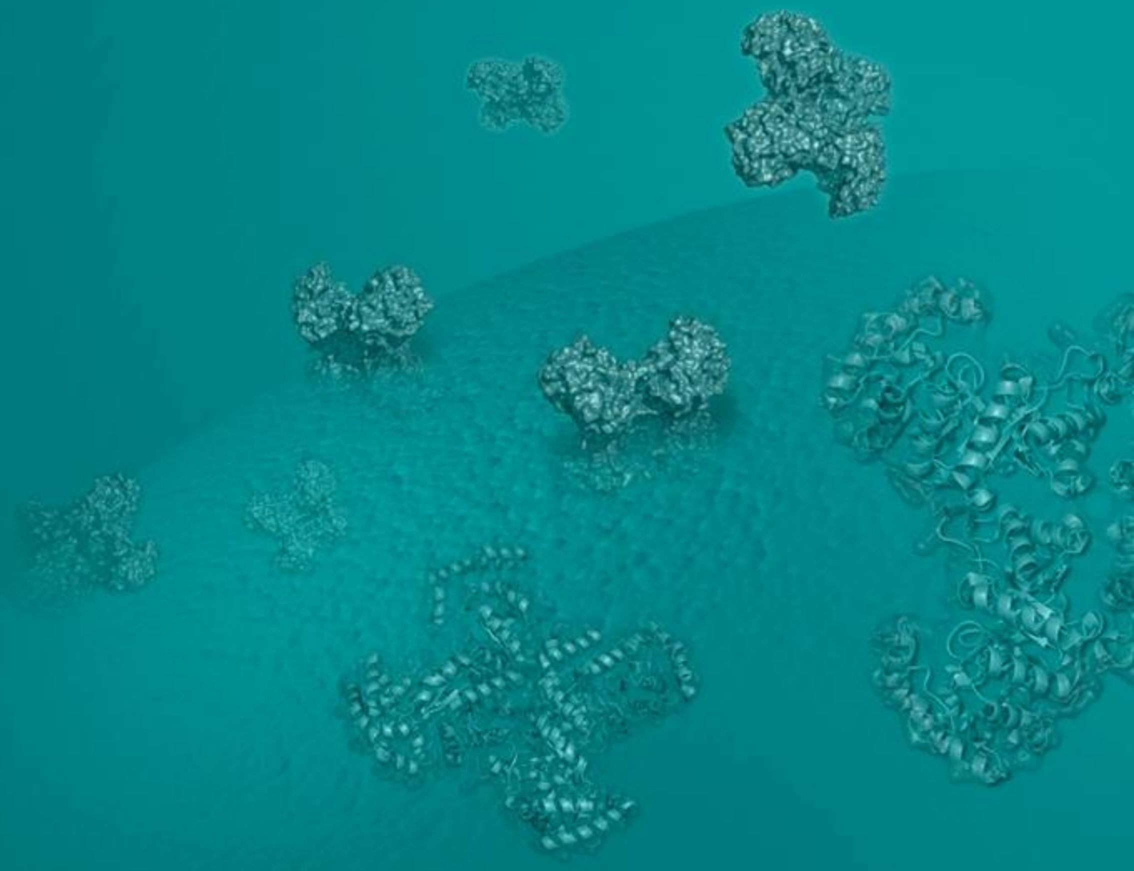


SIFTS - Structure Integration with Function, Taxonomy and Sequence

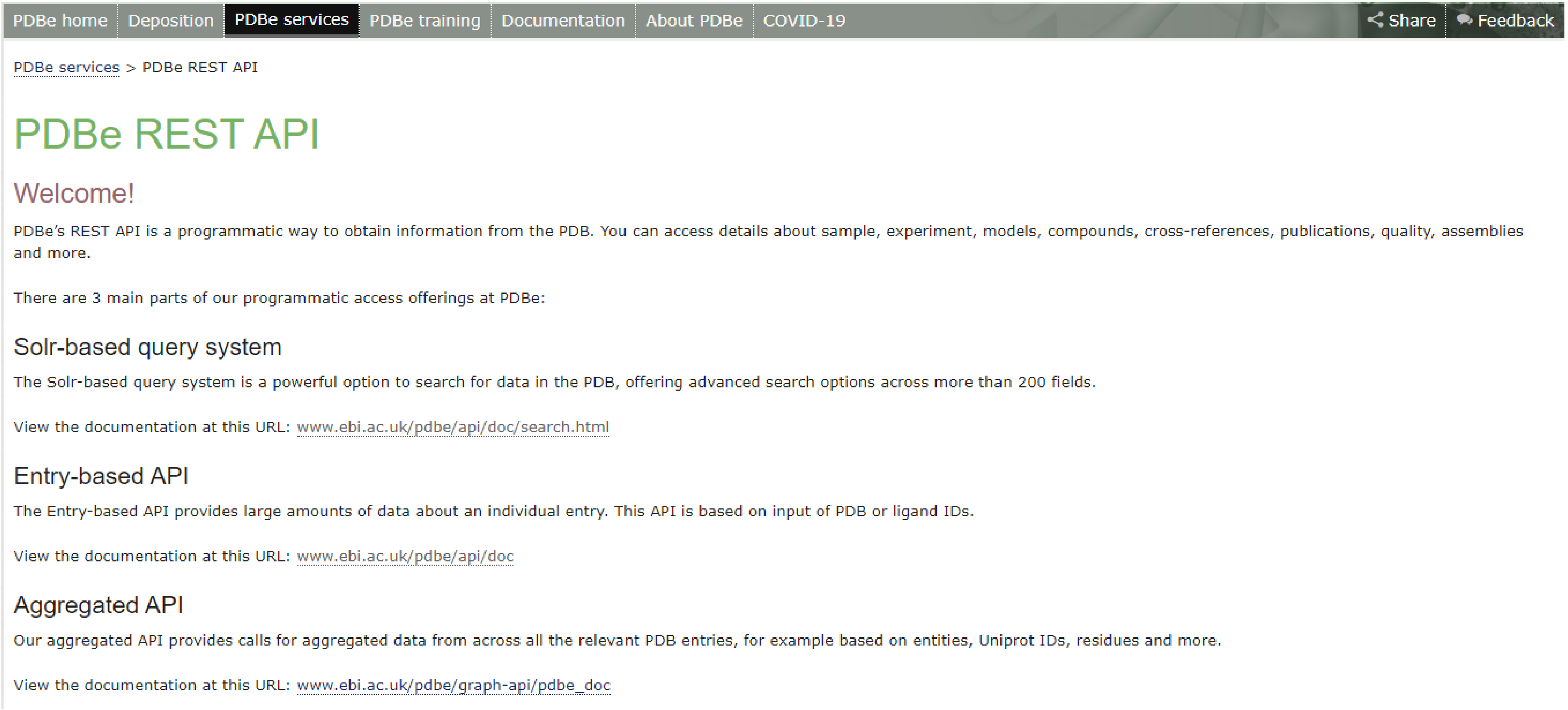

InterPro
GO
the Gene Ontology
Pfam
PDBe
Protein Data Bank in Europe
UniProt
SCOP
STRUCTURAL CLASSIFICATION OF PROTEINS
CATH
PROTEIN STRUCTURE CLASSIFICATION

PDBe.org/SIFTS

The three types of PDBe APIs



Landing page



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.

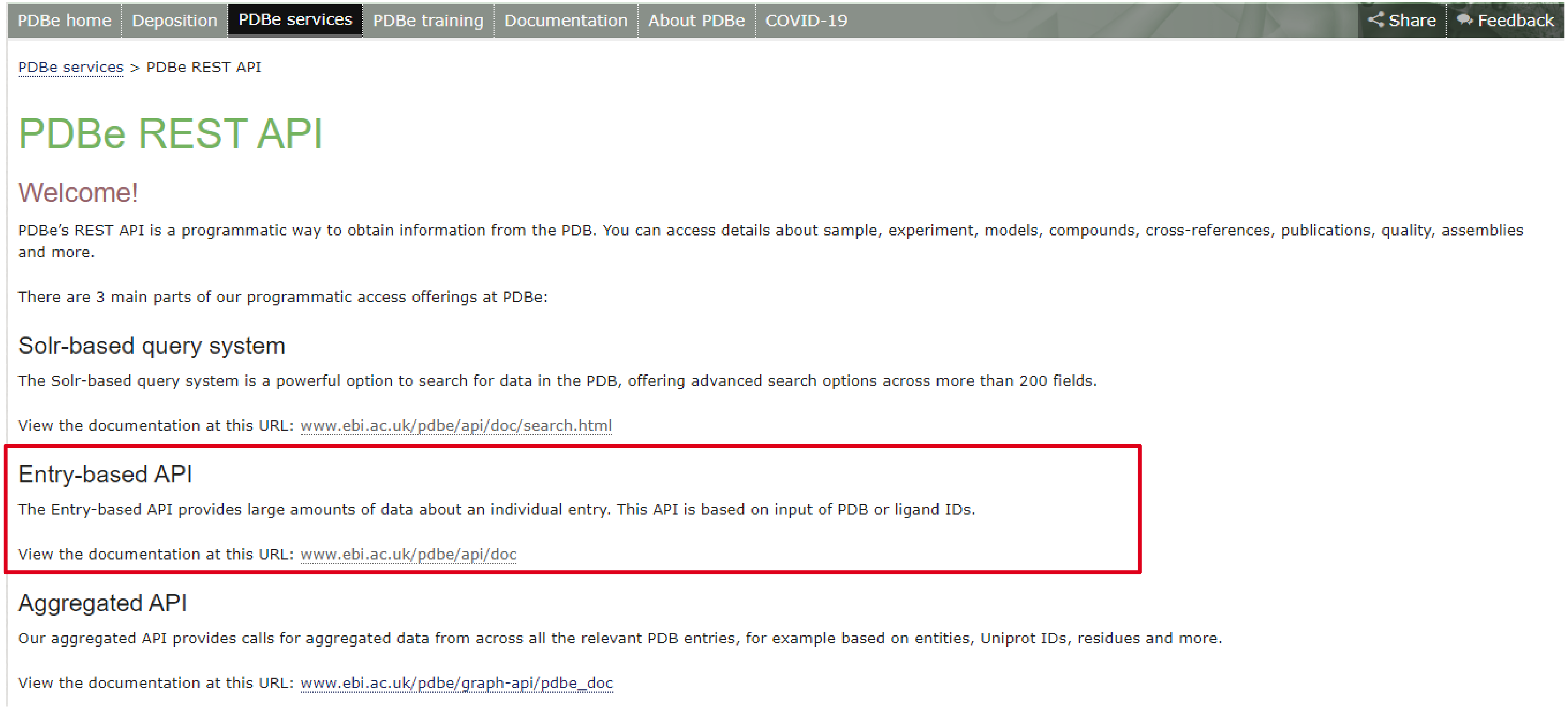
View the documentation at this URL: www.ebi.ac.uk/pdbe/api/doc

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

Landing page



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.org/API

EMBL-EBI

Querying

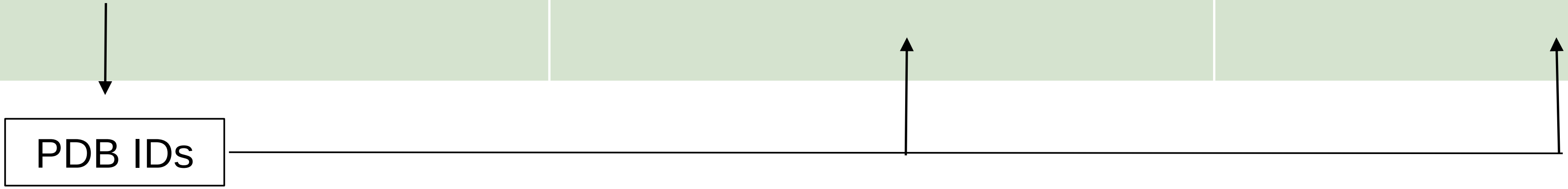
- Solr-based **query system** of data in the relational database
- Underlies the **PDBe search pages**
- **Large number of fields** available for searching
- By default, data **returned per entity** (i.e. unique molecule)
- Lots of data returned, but can also extract IDs and input to relational or graph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



To view the webinar on Solr-based query API, go to https://pdbeurope.github.io/api-webinars/webinars/web2/webinar.html

Querying

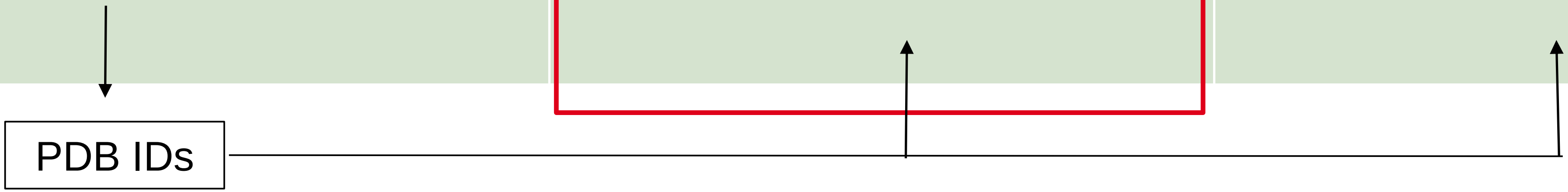
- Solr-based **query system** of data in the relational database
- Underlies the **PDBe search pages**
- **Large number of fields** available for searching
- By default, data **returned per entity** (i.e. unique molecule)
- Lots of data returned, but can also extract IDs and input to relational or graph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

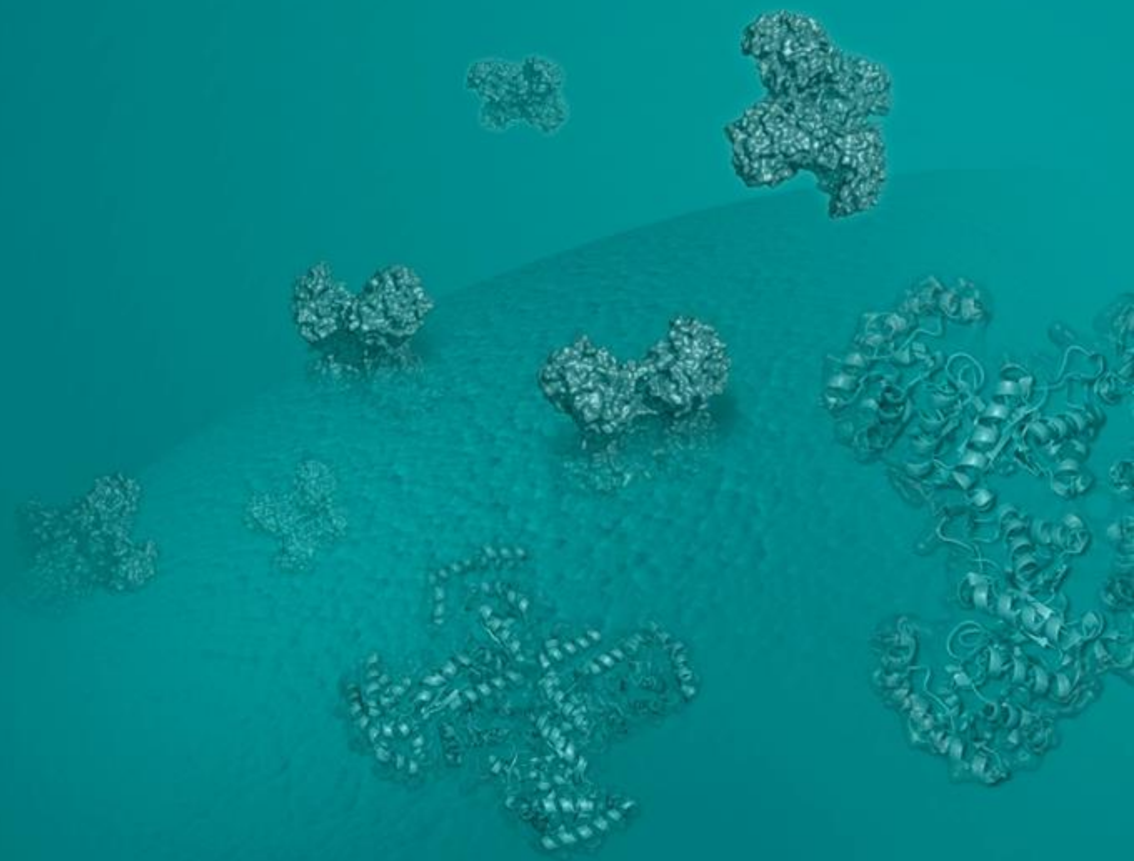
- Created from our newer Neo4J **graph database**
- Used to populate data on the **PDBe-KB aggregated views**
- **Provides aggregated data…**
- **Varied inputs** – get information **per residue** or per atom in ligand
- Incorporates additional data from **PDBe-KB partners**



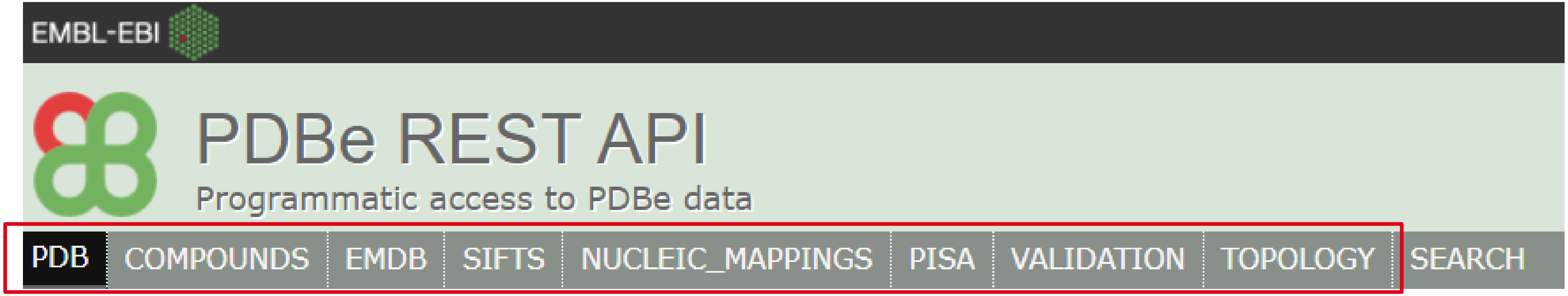
PDB IDs

To view the webinar on Solr-based query API, go to https://pdbeurope.github.io/api-webinars/webinars/web2/webinar.html

PDBe REST API: Entry-based API

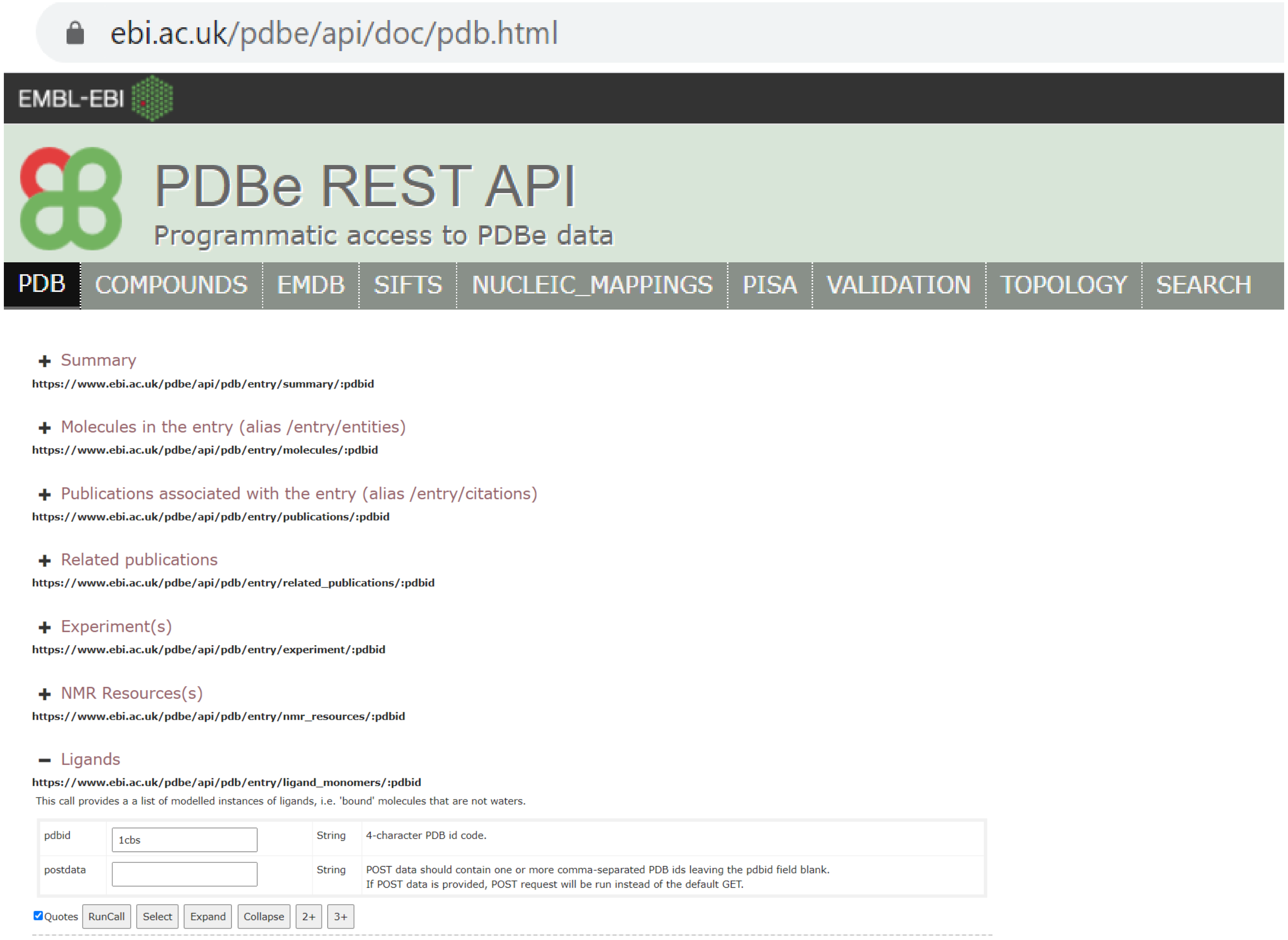


Eight categories/tabs



PDBe REST API

Programmatic access to PDBe data



ebi.ac.uk/pdbe/api/doc/pdb.html

PDBe REST API

Programmatic access to PDBe data

+ Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

+ Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

+ Publications associated with the entry (alias /entry/citations)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/publications/:pdbid

+ Related publications

https://www.ebi.ac.uk/pdbe/api/pdb/entry/related_publications/:pdbid

+ Experiment(s)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/experiment/:pdbid

+ NMR Resources(s)

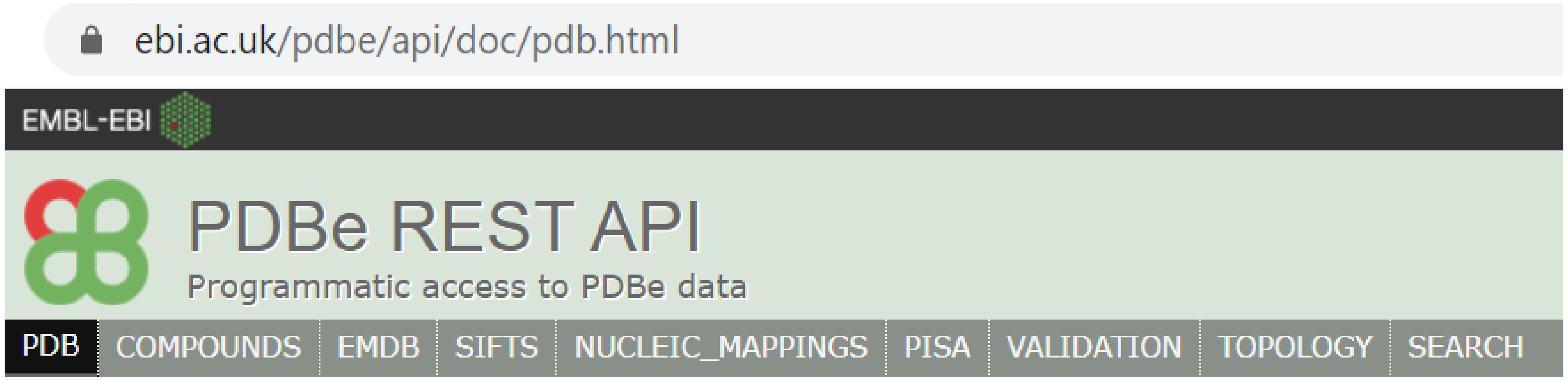
https://www.ebi.ac.uk/pdbe/api/pdb/entry/nmr_resources/:pdbid

− Ligands

https://www.ebi.ac.uk/pdbe/api/pdb/entry/ligand_monomers/:pdbid

This call provides a a list of modelled instances of ligands, i.e. 'bound' molecules that are not waters.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

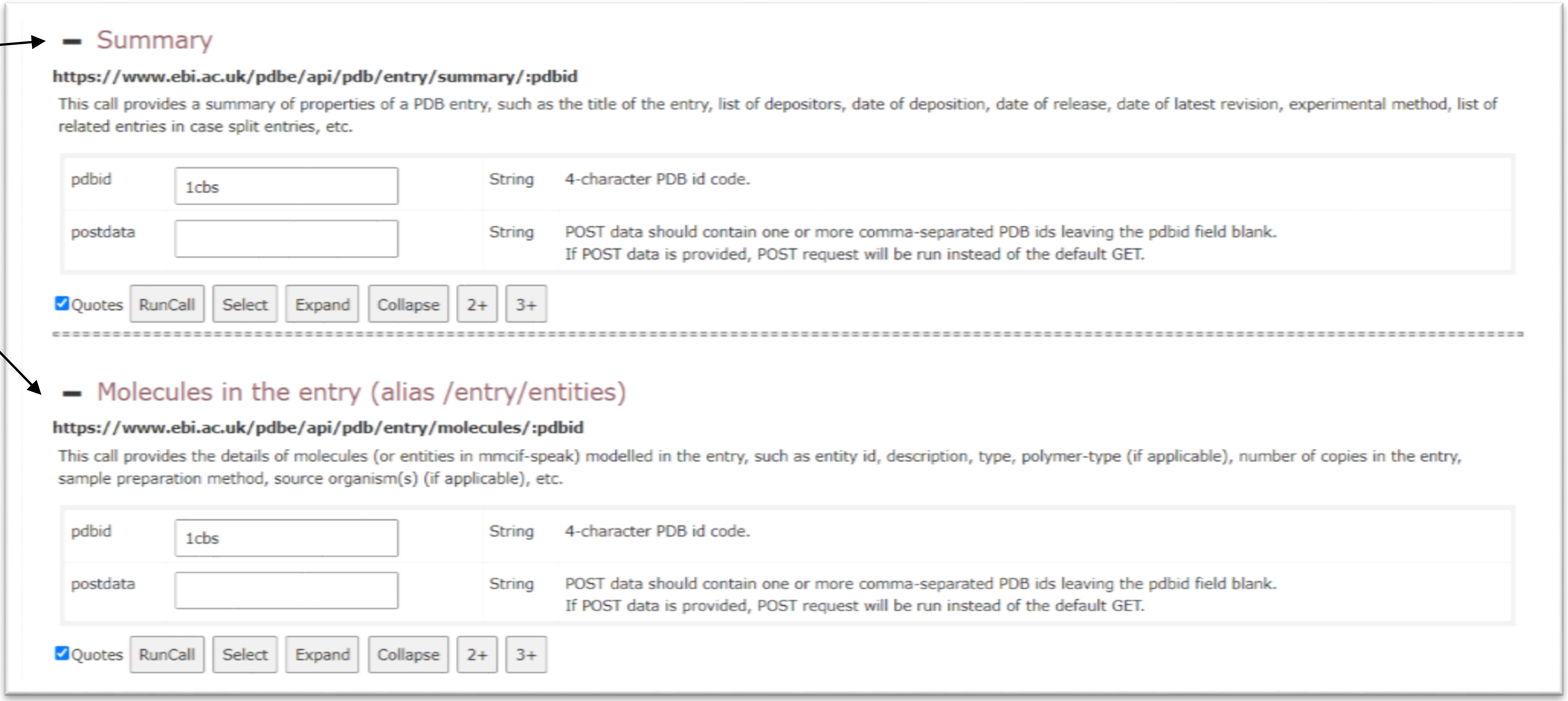


ebi.ac.uk/pdbe/api/doc/pdb.html

PDBe REST API

Programmatic access to PDBe data

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

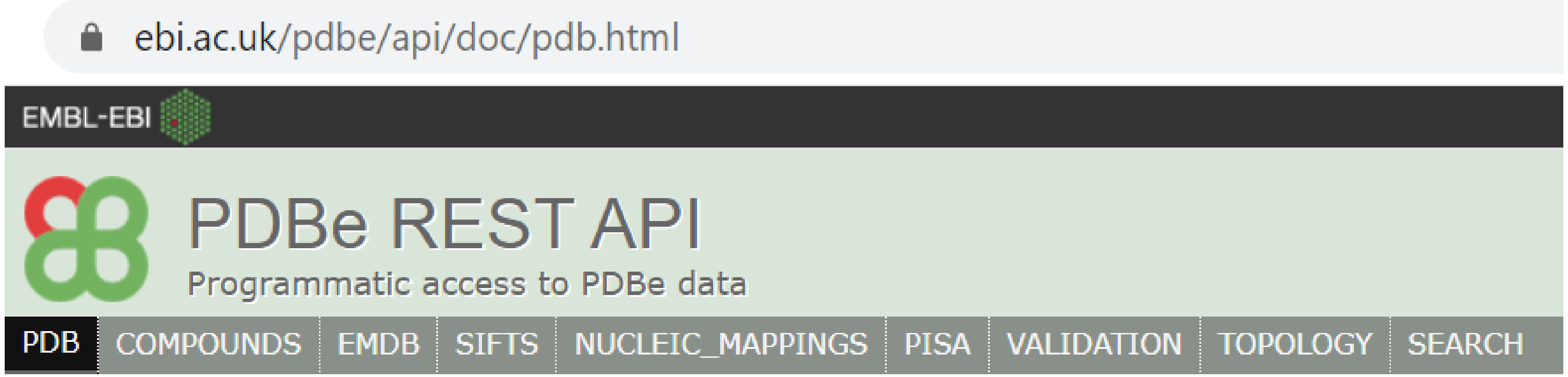
pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

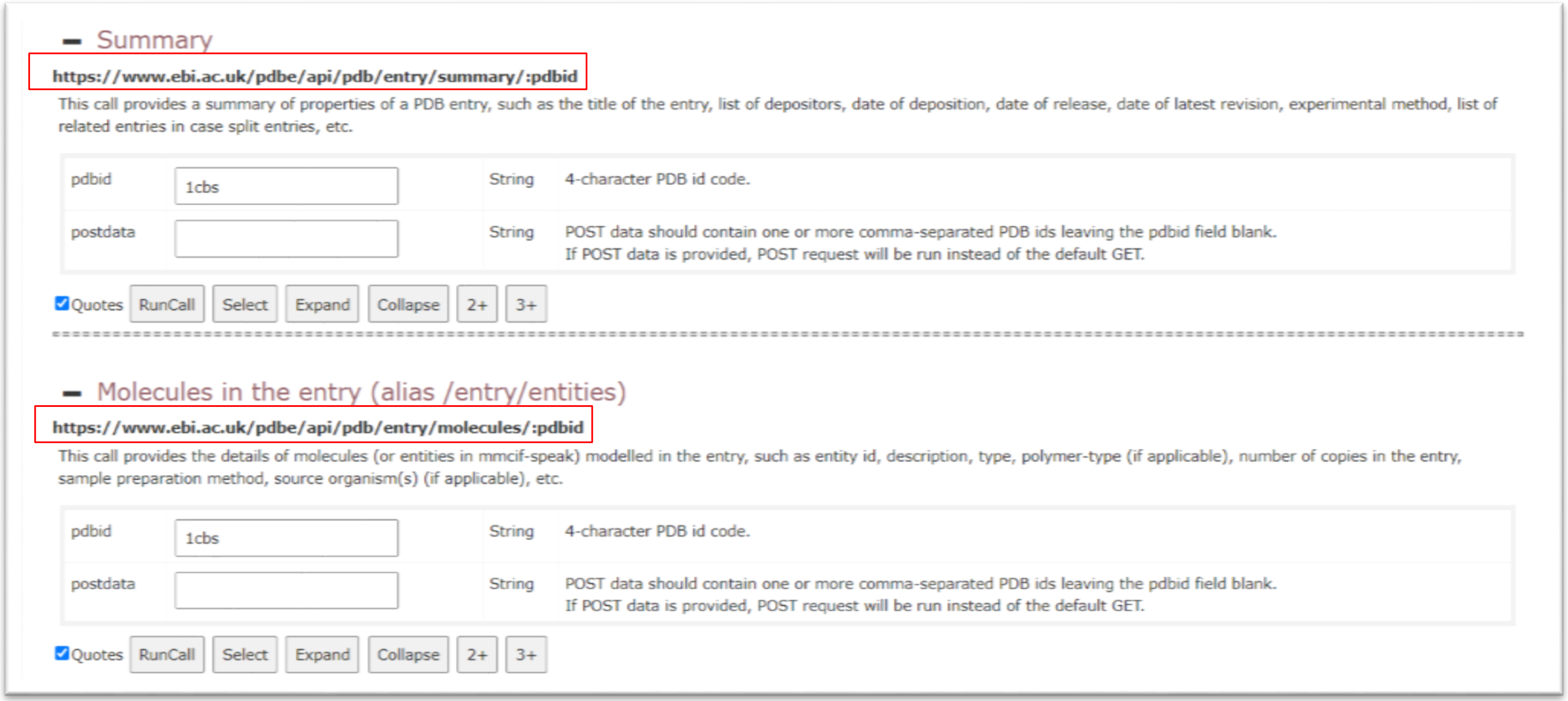
This call provides the details of molecules (or entities in mmcif-speak) modelled in the entry, such as entity id, description, type, polymer-type (if applicable), number of copies in the entry, sample preparation method, source organism(s) (if applicable), etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.



ebi.ac.uk/pdbe/api/doc/pdb.html

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

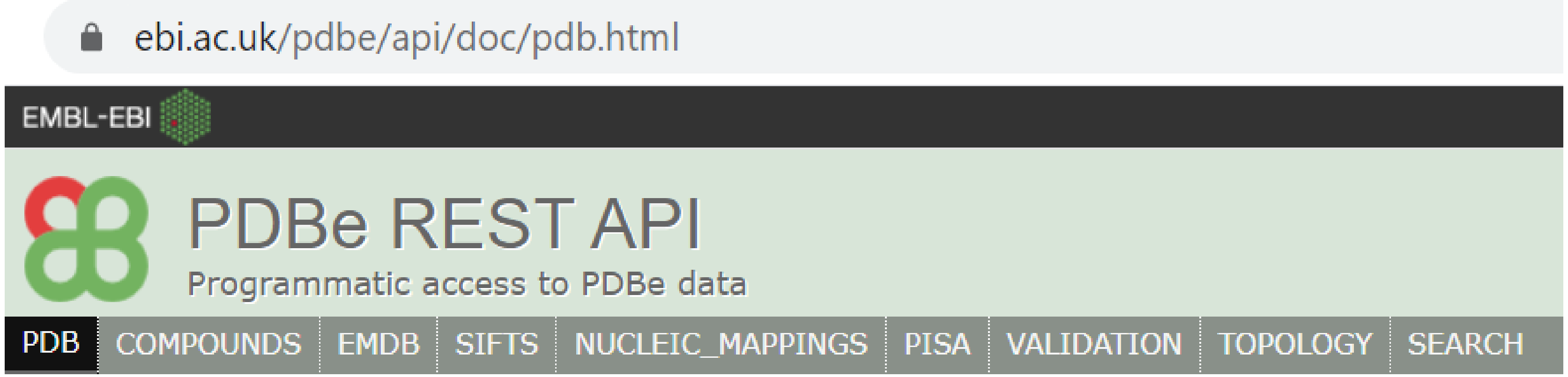
pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

This call provides the details of molecules (or entities in mmcif-speak) modelled in the entry, such as entity id, description, type, polymer-type (if applicable), number of copies in the entry, sample preparation method, source organism(s) (if applicable), etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.



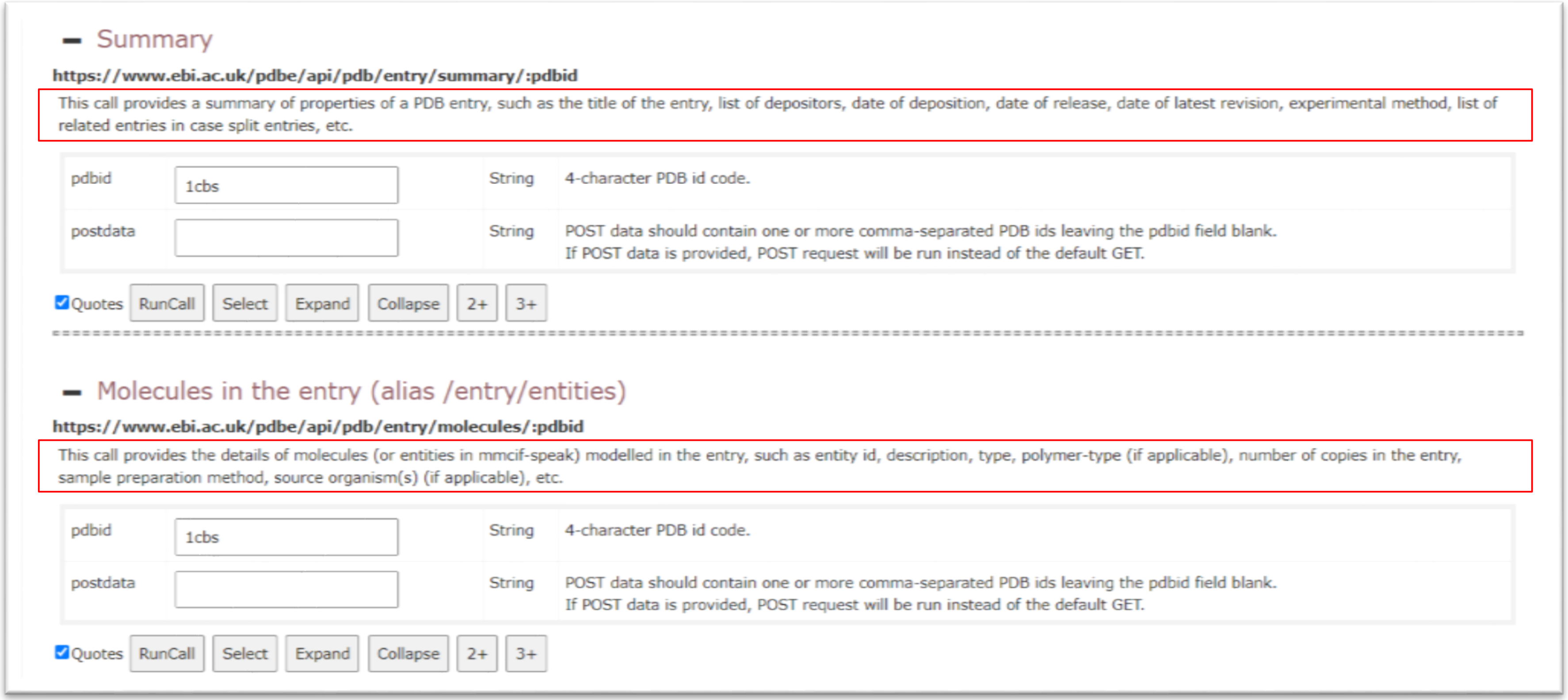
ebi.ac.uk/pdbe/api/doc/pdb.html

EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

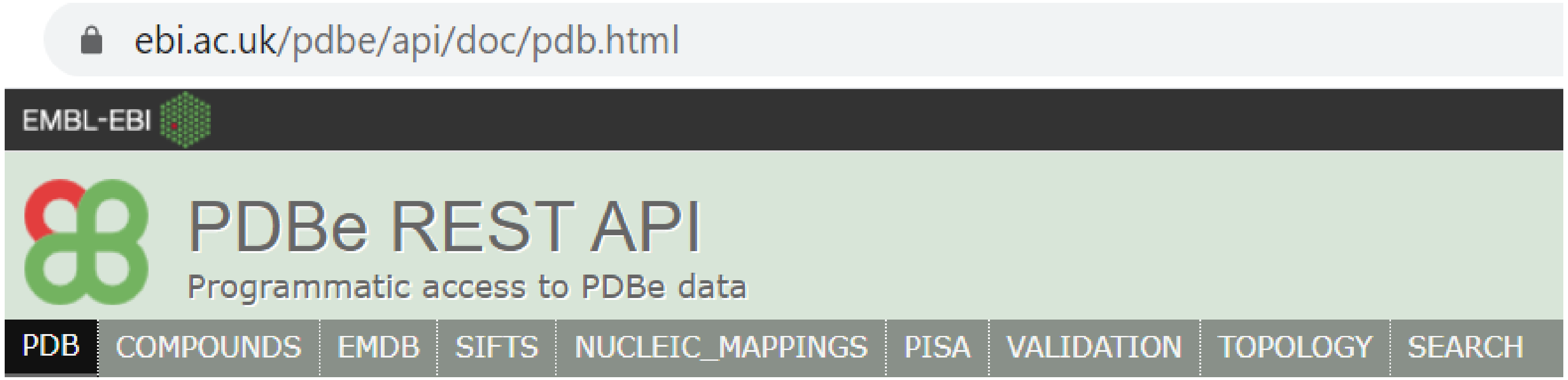
pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

This call provides the details of molecules (or entities in mmcif-speak) modelled in the entry, such as entity id, description, type, polymer-type (if applicable), number of copies in the entry, sample preparation method, source organism(s) (if applicable), etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.



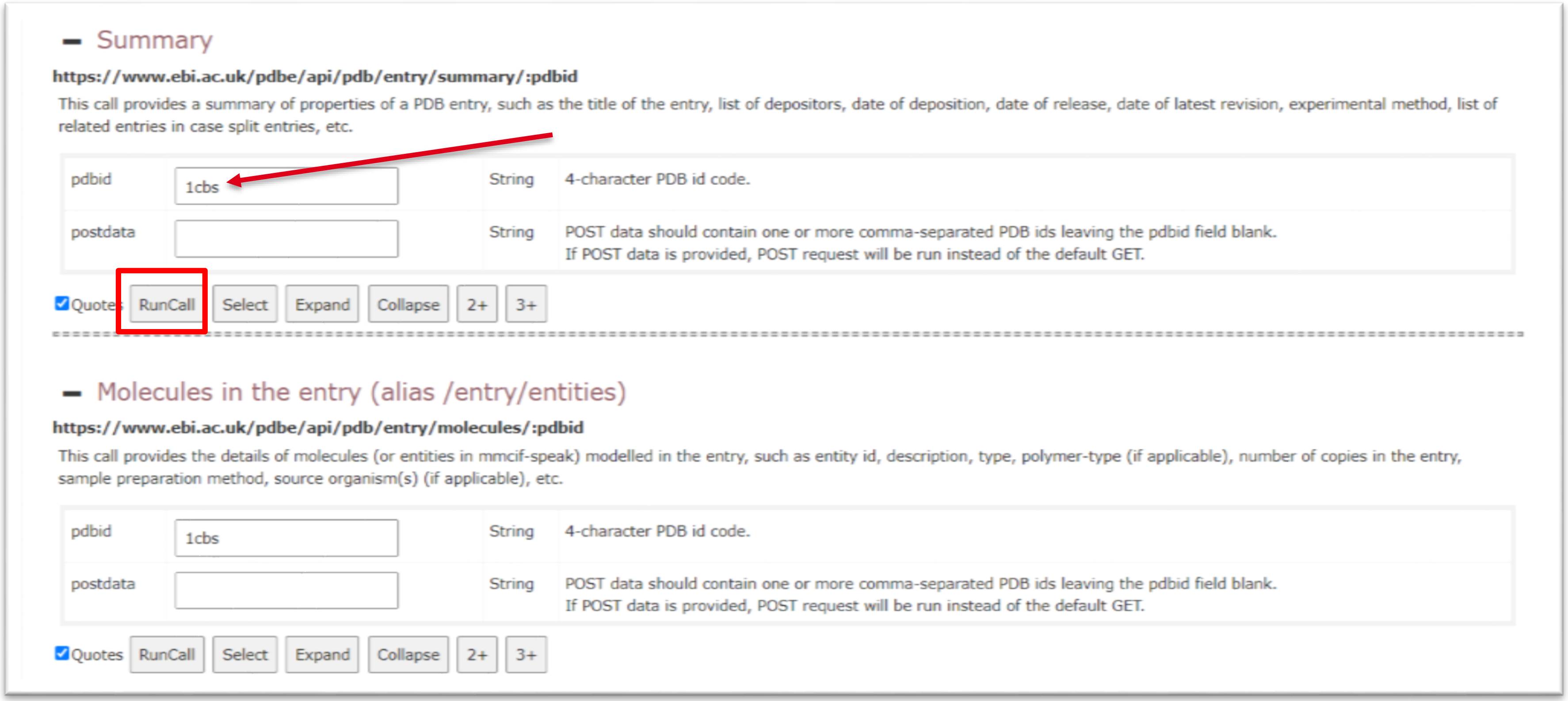
ebi.ac.uk/pdbe/api/doc/pdb.html

EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids 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+

Molecules in the entry (alias /entry/entities)

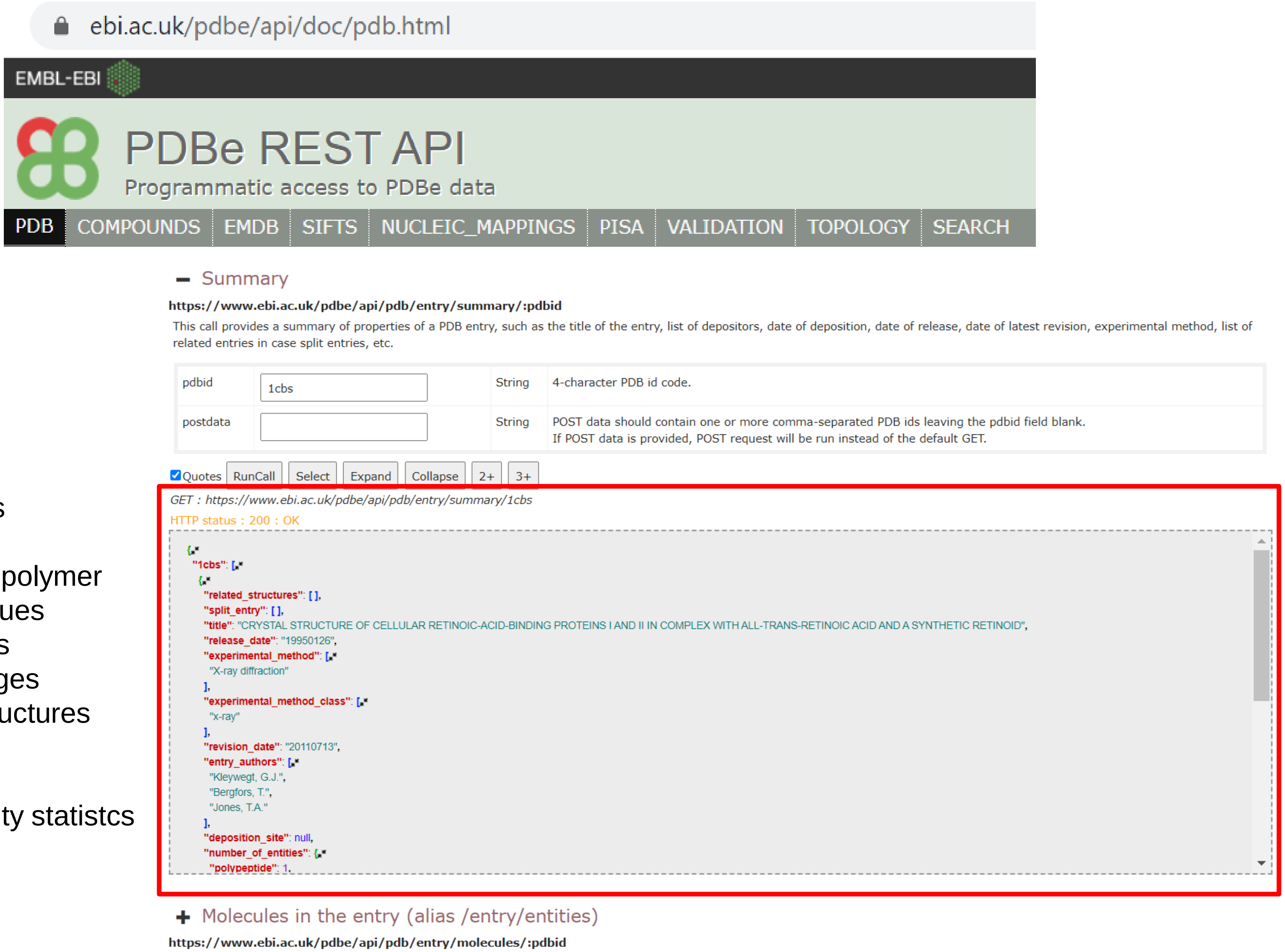
https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

This call provides the details of molecules (or entities in mmcif-speak) modelled in the entry, such as entity id, description, type, polymer-type (if applicable), number of copies in the entry, sample preparation method, source organism(s) (if applicable), etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids 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+

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



PDBe REST API

Programmatic access to PDBe data

− Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

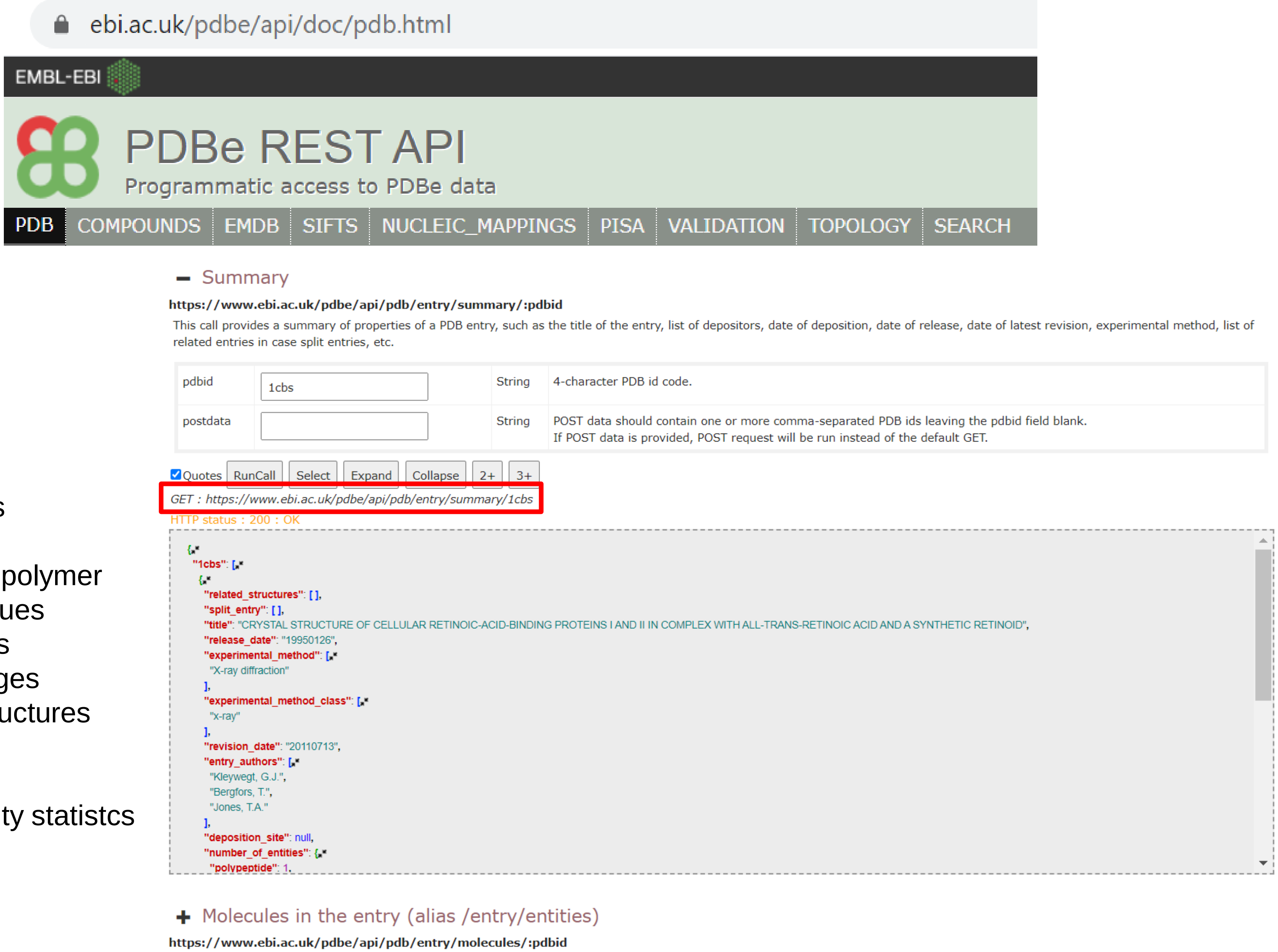
GET : https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/1cbs

HTTP status : 200 : OK

```
{
 "1cbs": [
  {
   "related_structures": [ ],
   "split_entry": [ ],
   "title": "CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID",
   "release_date": "19950126",
   "experimental_method": [
    "X-ray diffraction"
   ],
   "experimental_method_class": [
    "x-ray"
   ],
   "revision_date": "20110713",
   "entry_authors": [
    "Kleywegt, G.J.",
    "Bergfors, T.",
    "Jones, T.A."
   ],
   "deposition_site": null,
   "number_of_entities": {
    "polypeptide": 1,
```

+ Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid



PDBe REST API

Programmatic access to PDBe data

− Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

GET : https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/1cbs

HTTP status : 200 : OK

```
{
 "1cbs": [
  {
   "related_structures": [ ],
   "split_entry": [ ],
   "title": "CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID",
   "release_date": "19950126",
   "experimental_method": [
    "X-ray diffraction"
   ],
   "experimental_method_class": [
    "x-ray"
   ],
   "revision_date": "20110713",
   "entry_authors": [
    "Kleywegt, G.J.",
    "Bergfors, T.",
    "Jones, T.A."
   ],
   "deposition_site": null,
   "number_of_entities": {
    "polypeptide": 1,
```

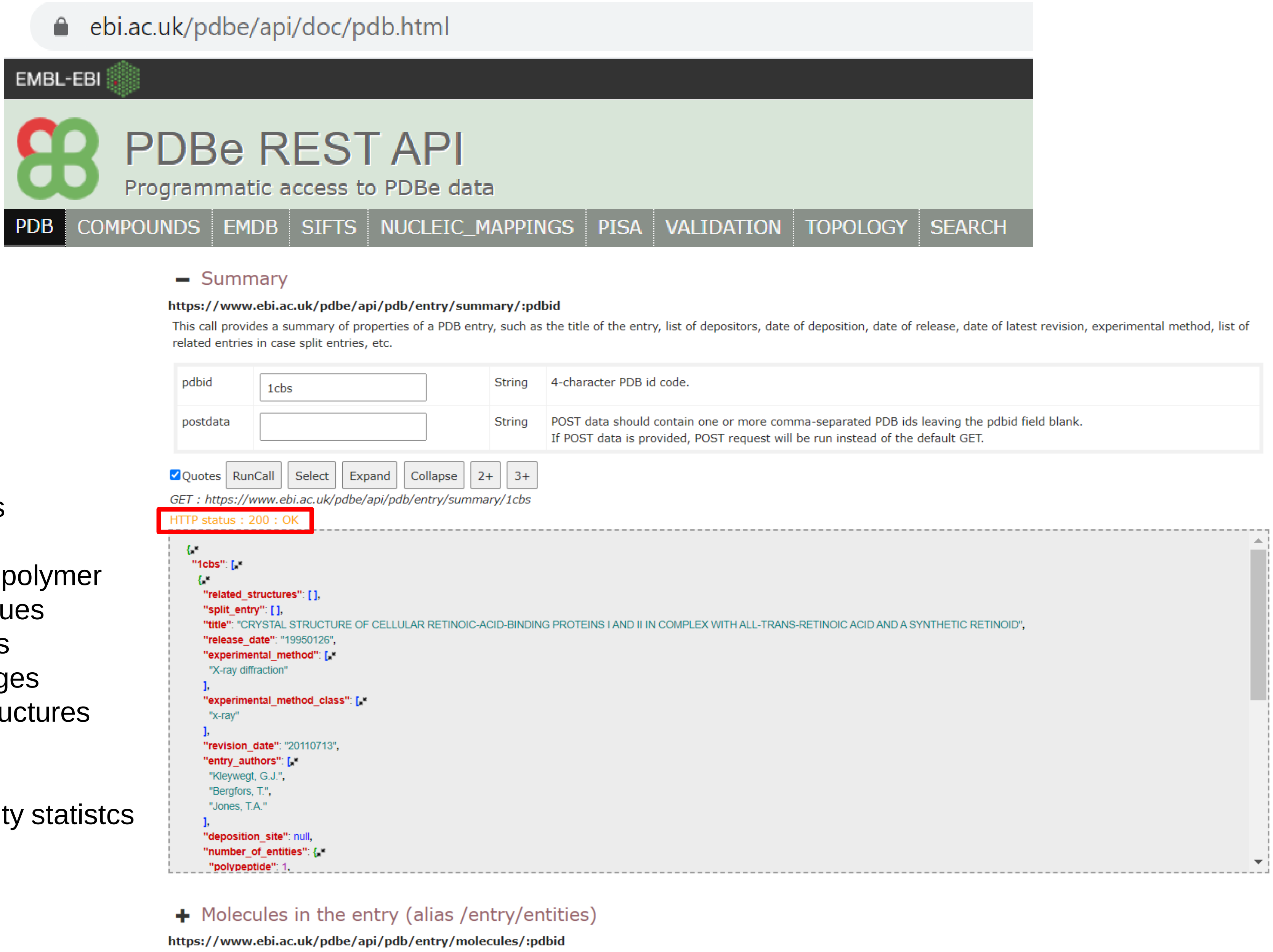
+ Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank

PDBe.org/API

EMBL-EBI



ebi.ac.uk/pdbe/api/doc/pdb.html

PDBe REST API

Programmatic access to PDBe data

− Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

☑ Quotes

GET : https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/1cbs

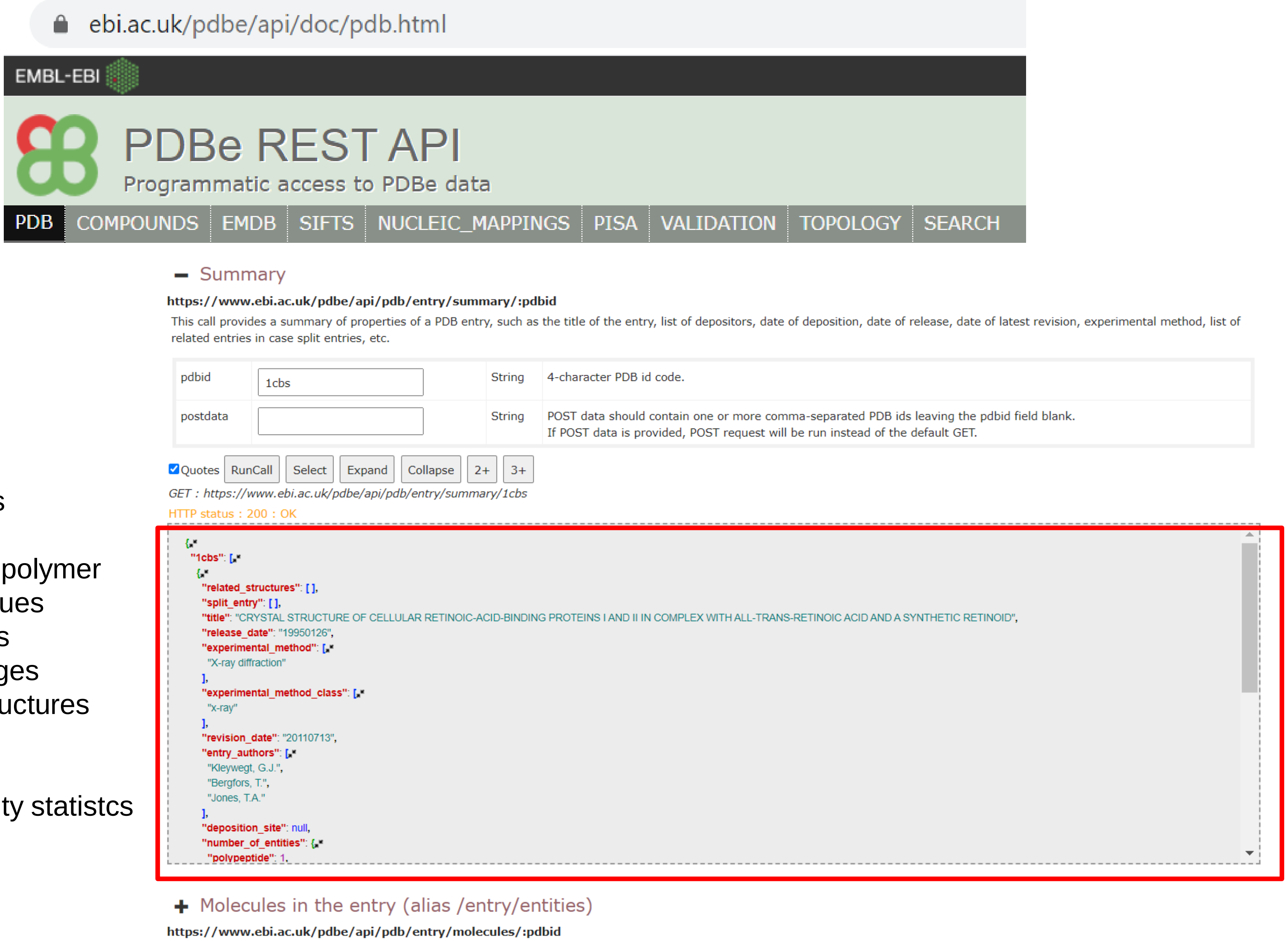
HTTP status : 200 : OK

```
{
 "1cbs": [
  {
   "related_structures": [ ],
   "split_entry": [ ],
   "title": "CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID",
   "release_date": "19950126",
   "experimental_method": [
    "X-ray diffraction"
   ],
   "experimental_method_class": [
    "x-ray"
   ],
   "revision_date": "20110713",
   "entry_authors": [
    "Kleywegt, G.J.",
    "Bergfors, T.",
    "Jones, T.A."
   ],
   "deposition_site": null,
   "number_of_entities": {
    "polypeptide": 1,
```

+ Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



— Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

GET : https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/1cbs

HTTP status : 200 : OK

```
{
 "1cbs": [
  {
   "related_structures": [ ],
   "split_entry": [ ],
   "title": "CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID",
   "release_date": "19950126",
   "experimental_method": [
    "X-ray diffraction"
   ],
   "experimental_method_class": [
    "x-ray"
   ],
   "revision_date": "20110713",
   "entry_authors": [
    "Kleywegt, G.J.",
    "Bergfors, T.",
    "Jones, T.A."
   ],
   "deposition_site": null,
   "number_of_entities": {
    "polypeptide": 1,
```

+ Molecules in the entry (alias /entry/entities)

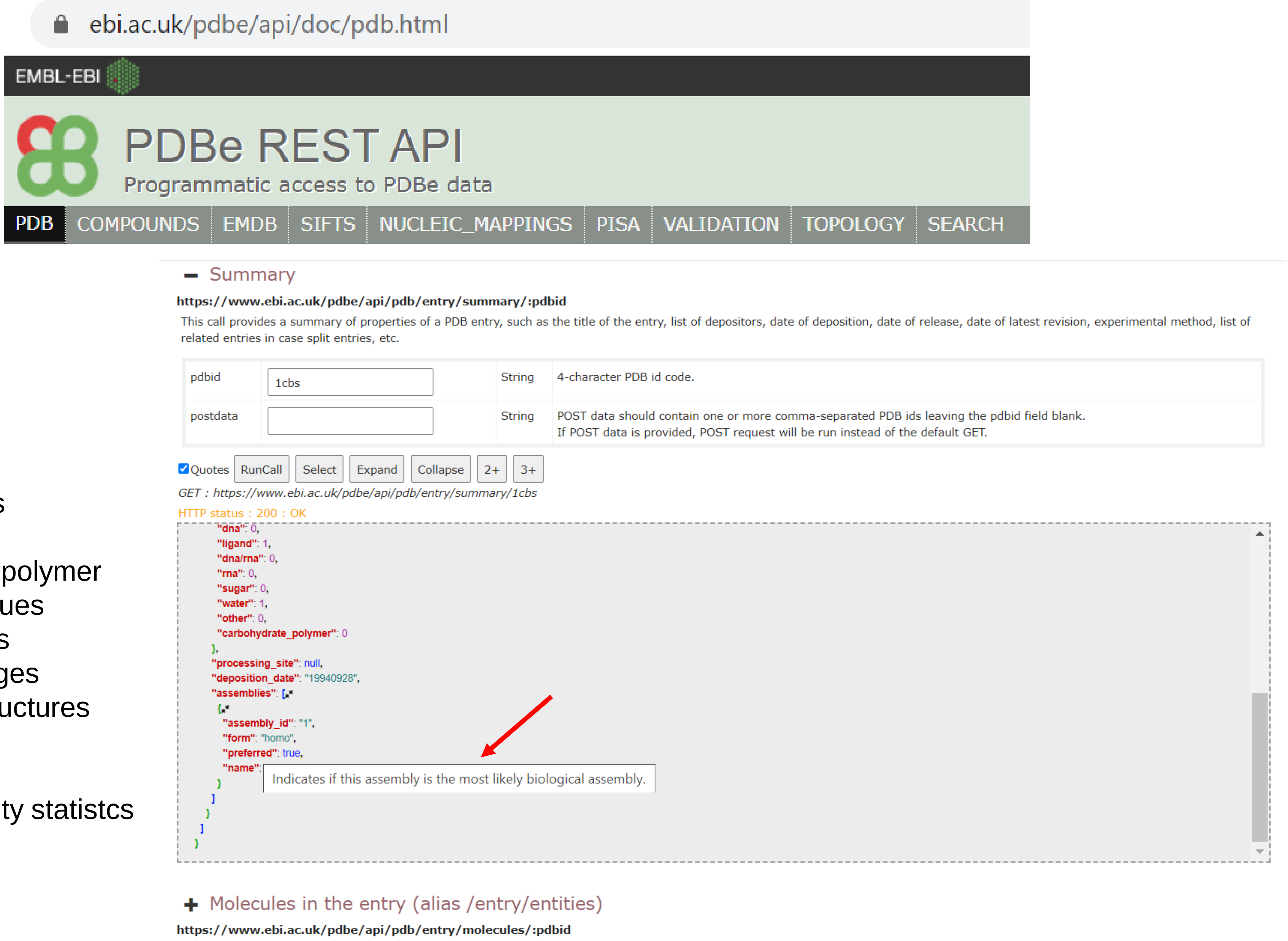
https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank

PDBe.org/API

EMBL-EBI

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



− Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid	1cbs	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids 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 : https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/1cbs

HTTP status : 200 : OK

```
        "dna": 0,
        "ligand": 1,
        "dna/rna": 0,
        "rna": 0,
        "sugar": 0,
        "water": 1,
        "other": 0,
        "carbohydrate_polymer": 0
      },
      "processing_site": null,
      "deposition_date": "19940928",
      "assemblies": [
        {
          "assembly_id": "1",
          "form": "homo",
          "preferred": true,
          "name":
        }
      ]
    }
  ]
}
```

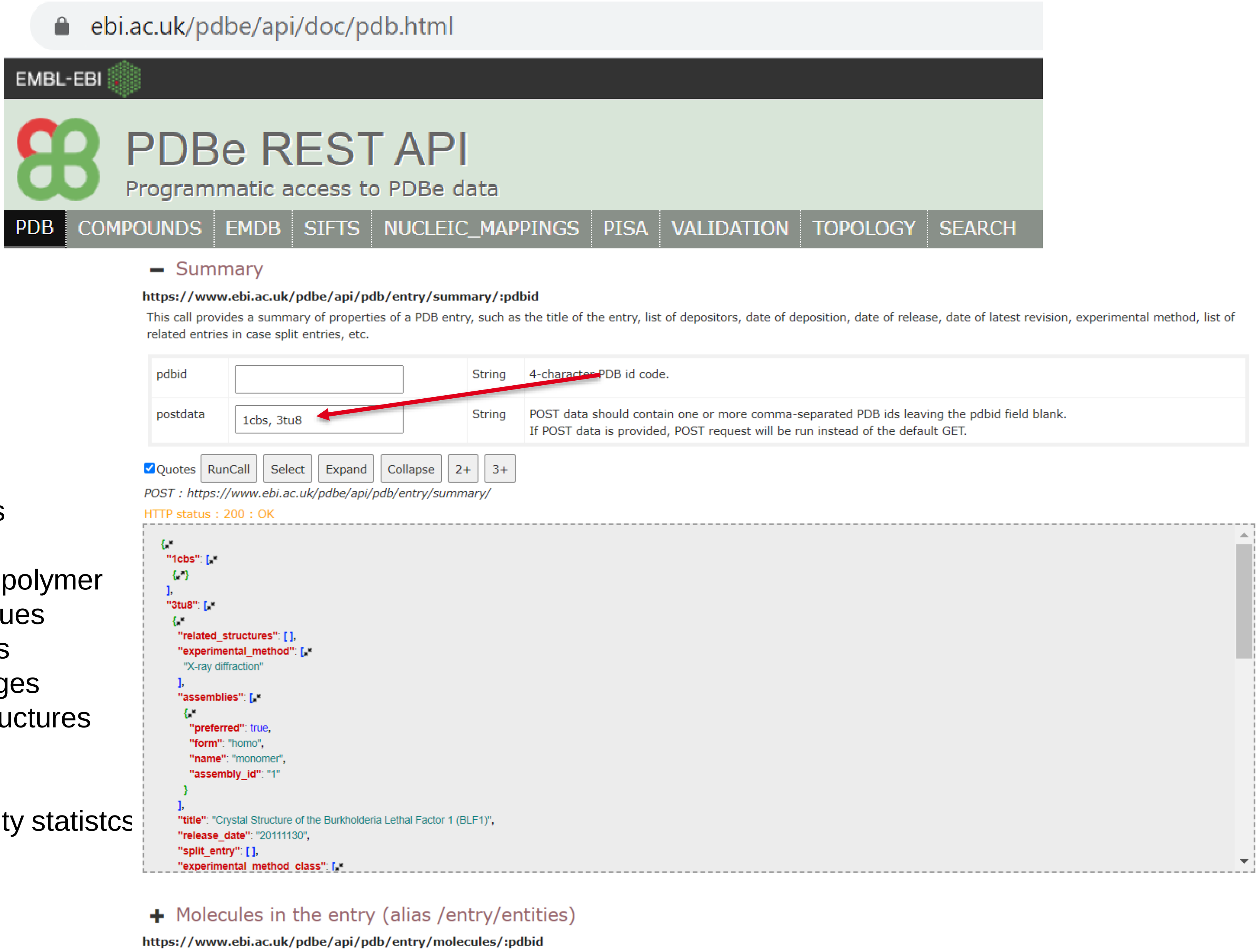
Indicates if this assembly is the most likely biological assembly.

+ Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

PDBe.org/API

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



ebi.ac.uk/pdbe/api/doc/pdb.html

EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

− Summary

https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/:pdbid

This call provides a summary of properties of a PDB entry, such as the title of the entry, list of depositors, date of deposition, date of release, date of latest revision, experimental method, list of related entries in case split entries, etc.

pdbid		String	4-character PDB id code.
postdata	1cbs, 3tu8	String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

POST : https://www.ebi.ac.uk/pdbe/api/pdb/entry/summary/

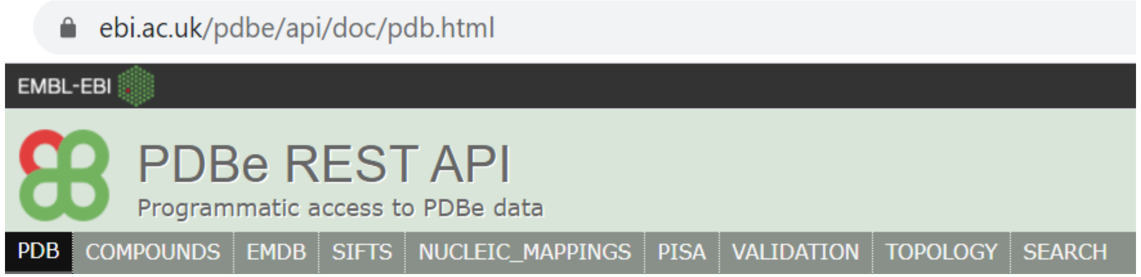
HTTP status : 200 : OK

```
{
 "1cbs": [
  {}
 ],
 "3tu8": [
  {
   "related_structures": [ ],
   "experimental_method": [
    "X-ray diffraction"
   ],
   "assemblies": [
    {
     "preferred": true,
     "form": "homo",
     "name": "monomer",
     "assembly_id": "1"
    }
   ],
   "title": "Crystal Structure of the Burkholderia Lethal Factor 1 (BLF1)",
   "release_date": "20111130",
   "split_entry": [ ],
   "experimental_method_class": [
```

+ Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

The PDBe entry page is powered by the API



ebi.ac.uk/pdbe/api/doc/pdb.html

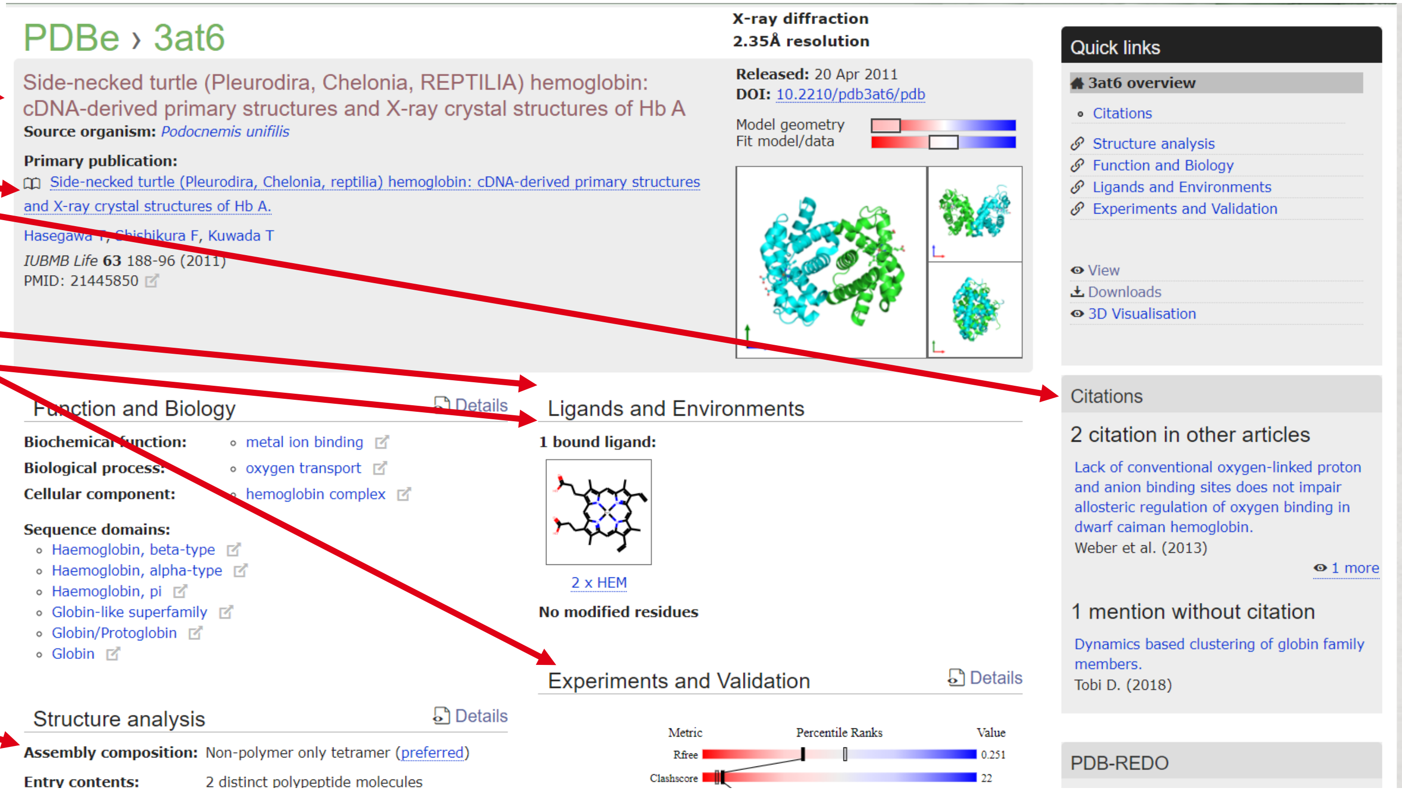
EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

PDB | COMPOUNDS | EMDB | SIFTS | NUCLEIC_MAPPINGS | PISA | VALIDATION | TOPOLOGY | SEARCH

Summary
Molecules
Publications
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



PDBe › 3at6

X-ray diffraction
2.35Å resolution

Side-necked turtle (Pleurodira, Chelonia, REPTILIA) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A

Source organism: *Podocnemis unifilis*

Primary publication:
Side-necked turtle (Pleurodira, Chelonia, reptilia) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A.
Hasegawa T, Shishikura F, Kuwada T
IUBMB Life **63** 188-96 (2011)
PMID: 21445850

Released: 20 Apr 2011
DOI: 10.2210/pdb3at6/pdb

Model geometry
Fit model/data

Quick links

- 3at6 overview
  - Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

- View
- Downloads
- 3D Visualisation

Function and Biology — Details

Biochemical function: metal ion binding
Biological process: oxygen transport
Cellular component: hemoglobin complex

Sequence domains:
- Haemoglobin, beta-type
- Haemoglobin, alpha-type
- Haemoglobin, pi
- Globin-like superfamily
- Globin/Protoglobin
- Globin

Ligands and Environments

1 bound ligand:

2 x HEM

No modified residues

Citations

2 citation in other articles

Lack of conventional oxygen-linked proton and anion binding sites does not impair allosteric regulation of oxygen binding in dwarf caiman hemoglobin.
Weber et al. (2013)

1 more

1 mention without citation

Dynamics based clustering of globin family members.
Tobi D. (2018)

Experiments and Validation — Details

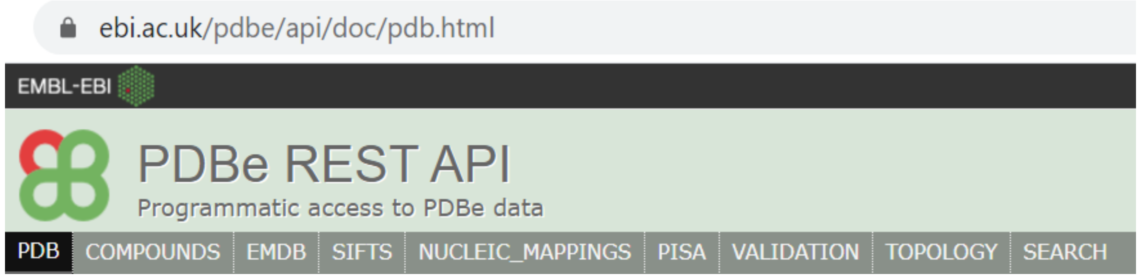
Metric	Percentile Ranks	Value
Rfree		0.251
Clashscore		22

Structure analysis — Details

Assembly composition: Non-polymer only tetramer (preferred)
Entry contents: 2 distinct polypeptide molecules

PDB-REDO

The PDBe entry page is powered by the API



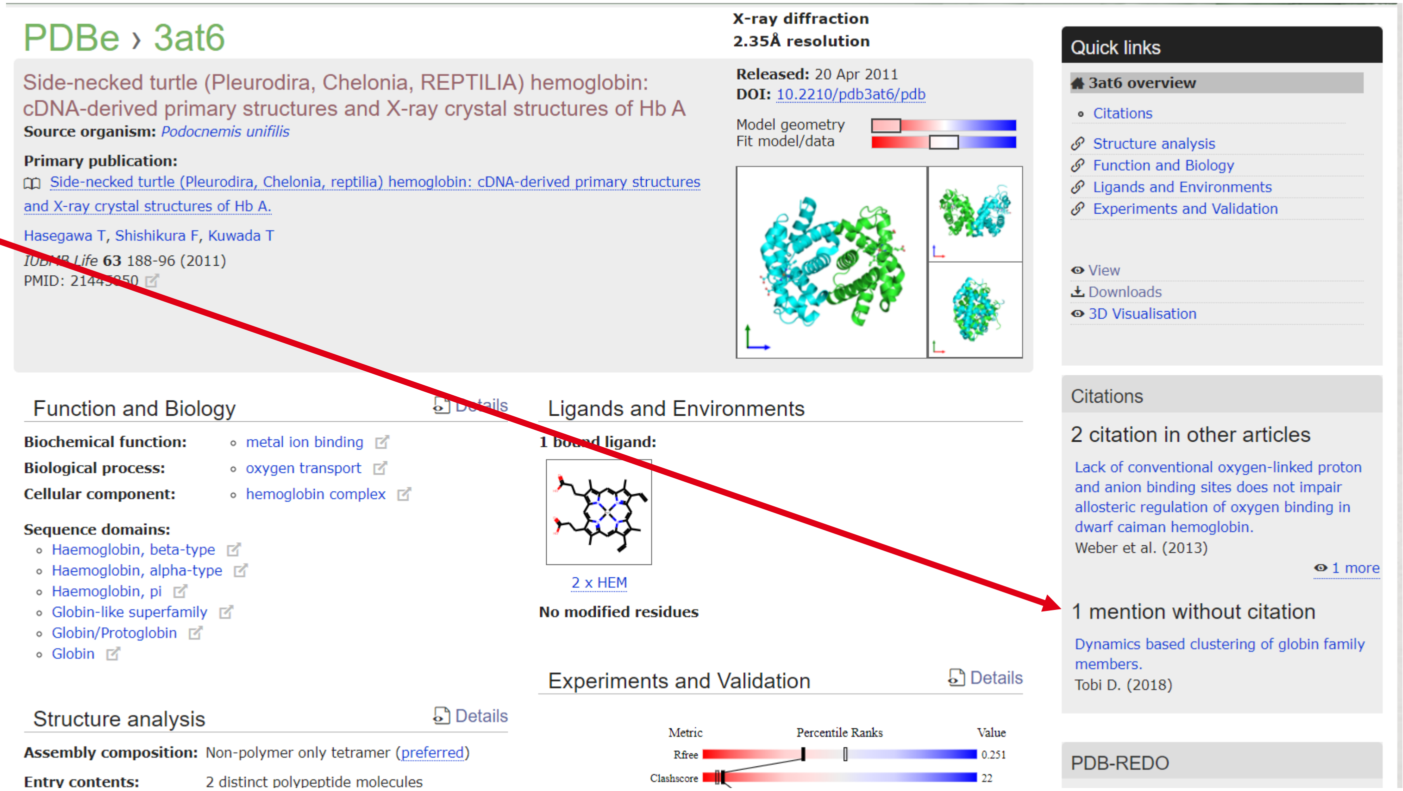
ebi.ac.uk/pdbe/api/doc/pdb.html

EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

- Summary
- Molecules
- Publications
- Related publications
- Experiment
- NMR resoures
- Ligands
- Carbohydrate polymer
- Modified residues
- Release status
- Observed ranges
- Secondary structures
- Binding sites
- Assembly
- Electron density statistcs
- Drugbank



PDBe › 3at6

X-ray diffraction
2.35Å resolution

Side-necked turtle (Pleurodira, Chelonia, REPTILIA) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A

Source organism: *Podocnemis unifilis*

Primary publication:
Side-necked turtle (Pleurodira, Chelonia, reptilia) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A.
Hasegawa T, Shishikura F, Kuwada T
IUBMB Life **63** 188-96 (2011)
PMID: 21445850

Released: 20 Apr 2011
DOI: 10.2210/pdb3at6/pdb

Model geometry
Fit model/data

Quick links

- 3at6 overview
  - Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation
- View
- Downloads
- 3D Visualisation

Function and Biology — Details

Biochemical function: metal ion binding
Biological process: oxygen transport
Cellular component: hemoglobin complex

Sequence domains:
- Haemoglobin, beta-type
- Haemoglobin, alpha-type
- Haemoglobin, pi
- Globin-like superfamily
- Globin/Protoglobin
- Globin

Ligands and Environments

1 bound ligand:
2 x HEM

No modified residues

Citations

2 citation in other articles

Lack of conventional oxygen-linked proton and anion binding sites does not impair allosteric regulation of oxygen binding in dwarf caiman hemoglobin.
Weber et al. (2013)

1 more

1 mention without citation

Dynamics based clustering of globin family members.
Tobi D. (2018)

Experiments and Validation — Details

Metric	Percentile Ranks	Value
Rfree		0.251
Clashscore		22

Structure analysis — Details

Assembly composition: Non-polymer only tetramer (preferred)
Entry contents: 2 distinct polypeptide molecules

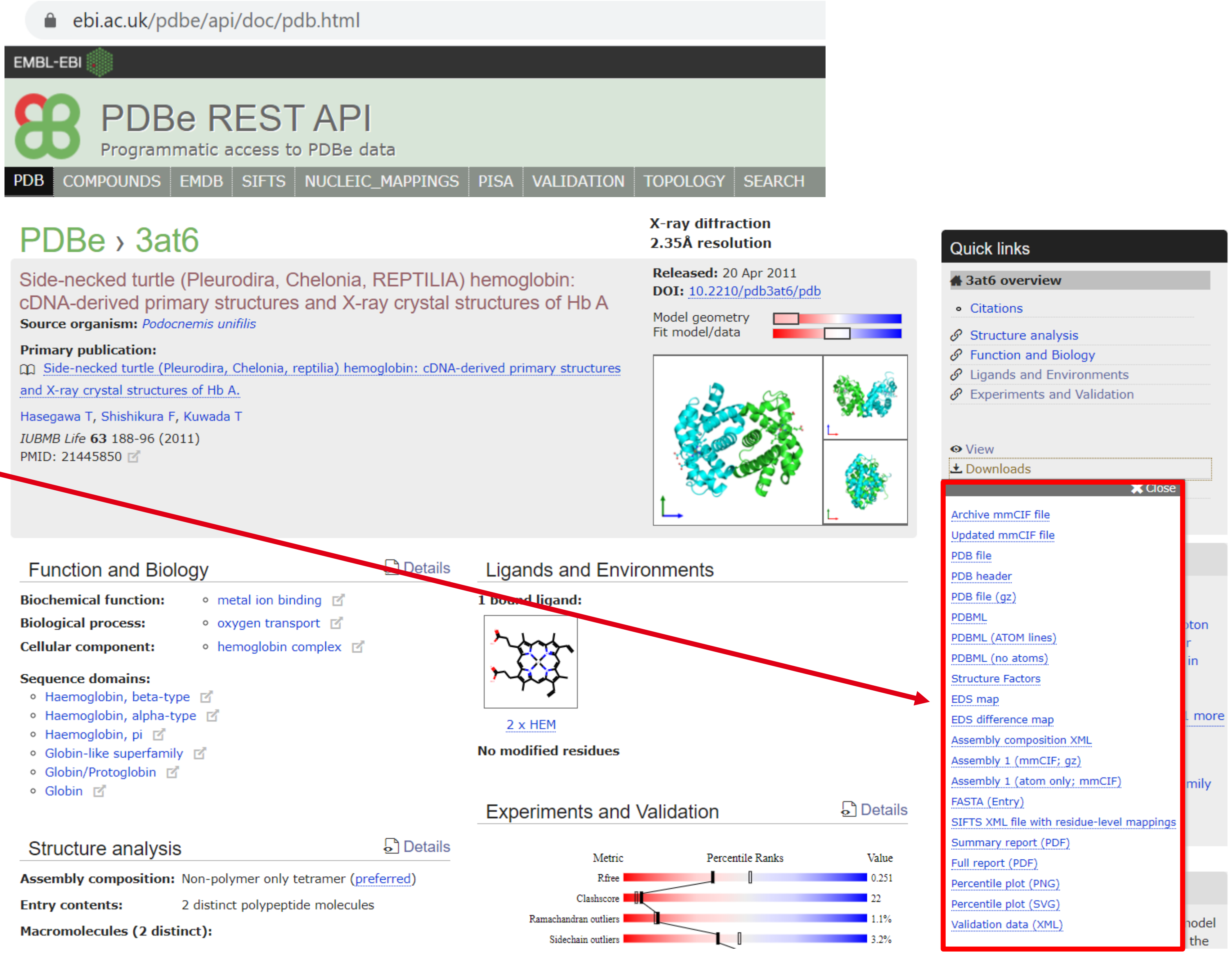
PDB-REDO

PDBe.org/API

EMBL-EBI

The PDBe entry page is powered by the API

Summary
Molecules
Publications
Files
Experiment
NMR resoures
Ligands
Carbohydrate polymer
Modified residues
Release status
Observed ranges
Secondary structures
Binding sites
Assembly
Electron density statistcs
Drugbank



ebi.ac.uk/pdbe/api/doc/pdb.html

EMBL-EBI

PDBe REST API
Programmatic access to PDBe data

PDB | COMPOUNDS | EMDB | SIFTS | NUCLEIC_MAPPINGS | PISA | VALIDATION | TOPOLOGY | SEARCH

PDBe › 3at6

X-ray diffraction
2.35Å resolution

Side-necked turtle (Pleurodira, Chelonia, REPTILIA) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A

Source organism: *Podocnemis unifilis*

Primary publication:
Side-necked turtle (Pleurodira, Chelonia, reptilia) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A.
Hasegawa T, Shishikura F, Kuwada T
IUBMB Life **63** 188-96 (2011)
PMID: 21445850

Released: 20 Apr 2011
DOI: 10.2210/pdb3at6/pdb

Model geometry
Fit model/data

Quick links
3at6 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

View
Downloads
Close

Archive mmCIF file
Updated mmCIF file
PDB file
PDB header
PDB file (gz)
PDBML
PDBML (ATOM lines)
PDBML (no atoms)
Structure Factors
EDS map
EDS difference map
Assembly composition XML
Assembly 1 (mmCIF; gz)
Assembly 1 (atom only; mmCIF)
FASTA (Entry)
SIFTS XML file with residue-level mappings
Summary report (PDF)
Full report (PDF)
Percentile plot (PNG)
Percentile plot (SVG)
Validation data (XML)

pton
r
in
more
mily
nodel
the

Function and Biology

Details

Biochemical function: metal ion binding
Biological process: oxygen transport
Cellular component: hemoglobin complex

Sequence domains:
- Haemoglobin, beta-type
- Haemoglobin, alpha-type
- Haemoglobin, pi
- Globin-like superfamily
- Globin/Protoglobin
- Globin

Ligands and Environments

1 bound ligand:

2 x HEM

No modified residues

Experiments and Validation

Details

Metric | Percentile Ranks | Value
Rfree 0.251
Clashscore 22
Ramachandran outliers 1.1%
Sidechain outliers 3.2%

Structure analysis

Details

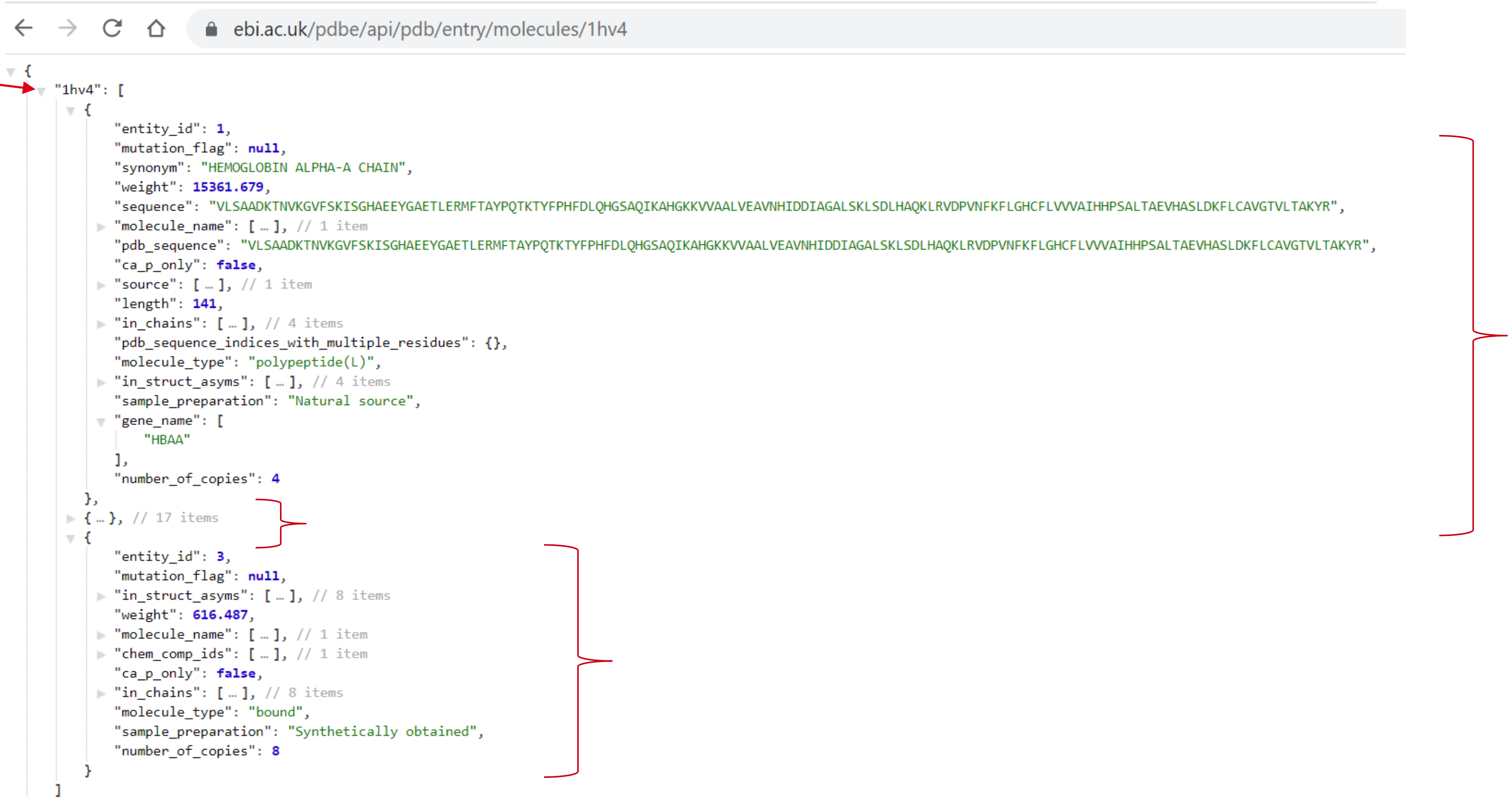
Assembly composition: Non-polymer only tetramer (preferred)
Entry contents: 2 distinct polypeptide molecules
Macromolecules (2 distinct):

API JSON output example: molecules call

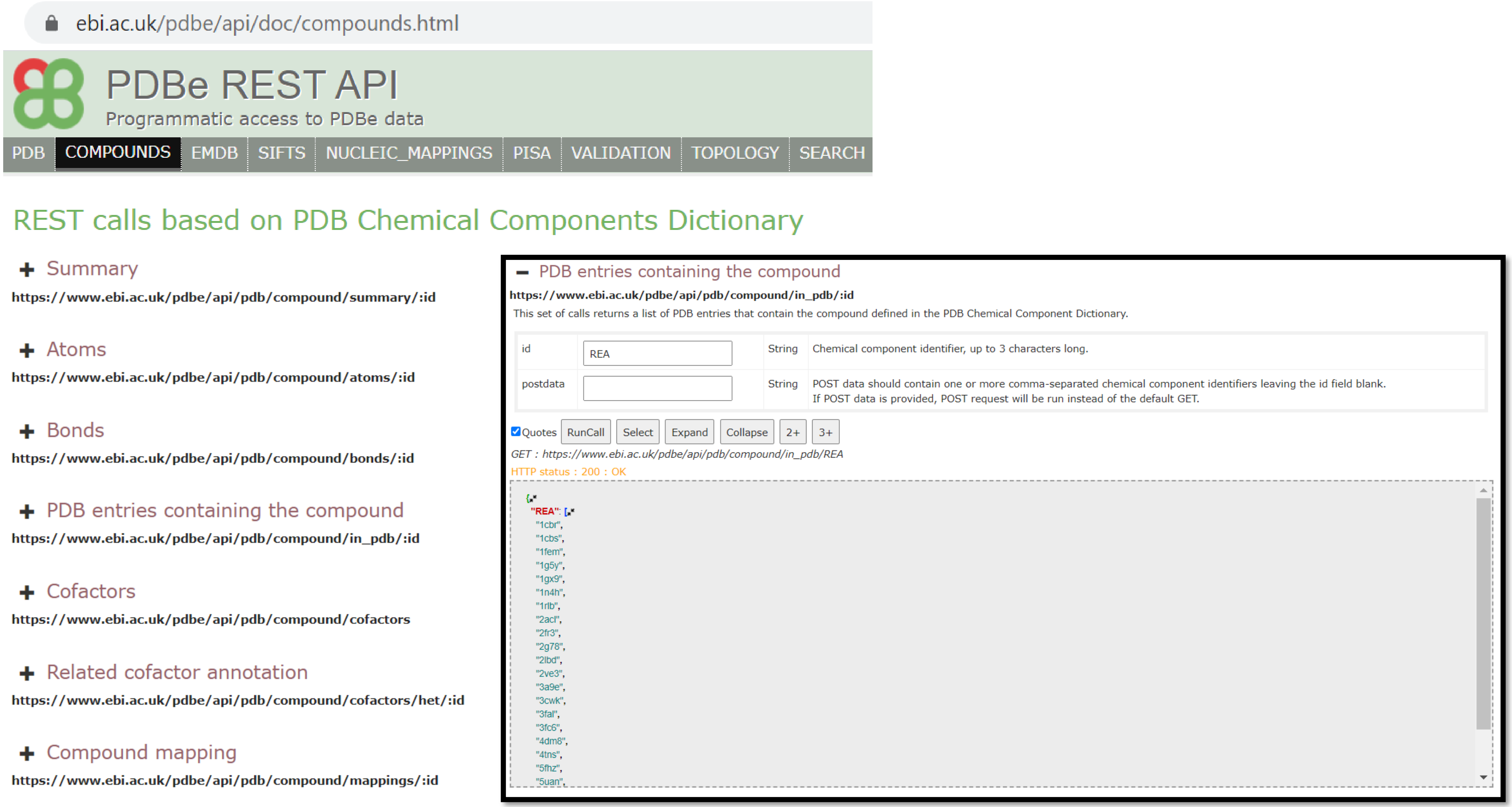
JSON key is always an ID

- PDB ID
- Chem comp ID

Three entities



```
{
  "1hv4": [
    {
        "entity_id": 1,
        "mutation_flag": null,
        "synonym": "HEMOGLOBIN ALPHA-A CHAIN",
        "weight": 15361.679,
        "sequence": "VLSAADKTNVKGVFSKISGHAEEYGAETLERMFTAYPQTKTYFPHFDLQHGSAQIKAHGKKVVAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPSALTAEVHASLDKFLCAVGTVLTAKYR",
      ▶ "molecule_name": [ … ], // 1 item
        "pdb_sequence": "VLSAADKTNVKGVFSKISGHAEEYGAETLERMFTAYPQTKTYFPHFDLQHGSAQIKAHGKKVVAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPSALTAEVHASLDKFLCAVGTVLTAKYR",
        "ca_p_only": false,
      ▶ "source": [ … ], // 1 item
        "length": 141,
      ▶ "in_chains": [ … ], // 4 items
        "pdb_sequence_indices_with_multiple_residues": {},
        "molecule_type": "polypeptide(L)",
      ▶ "in_struct_asyms": [ … ], // 4 items
        "sample_preparation": "Natural source",
        "gene_name": [
            "HBAA"
        ],
        "number_of_copies": 4
    },
  ▶ { … }, // 17 items
    {
        "entity_id": 3,
        "mutation_flag": null,
      ▶ "in_struct_asyms": [ … ], // 8 items
        "weight": 616.487,
      ▶ "molecule_name": [ … ], // 1 item
      ▶ "chem_comp_ids": [ … ], // 1 item
        "ca_p_only": false,
      ▶ "in_chains": [ … ], // 8 items
        "molecule_type": "bound",
        "sample_preparation": "Synthetically obtained",
        "number_of_copies": 8
    }
  ]
```



ebi.ac.uk/pdbe/api/doc/compounds.html

PDBe REST API

Programmatic access to PDBe data

REST calls based on PDB Chemical Components Dictionary

+ Summary

https://www.ebi.ac.uk/pdbe/api/pdb/compound/summary/:id

+ Atoms

https://www.ebi.ac.uk/pdbe/api/pdb/compound/atoms/:id

+ Bonds

https://www.ebi.ac.uk/pdbe/api/pdb/compound/bonds/:id

+ PDB entries containing the compound

https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/:id

+ Cofactors

https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors

+ Related cofactor annotation

https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors/het/:id

+ Compound mapping

https://www.ebi.ac.uk/pdbe/api/pdb/compound/mappings/:id

− PDB entries containing the compound

https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/:id

This set of calls returns a list of PDB entries that contain the compound defined in the PDB Chemical Component Dictionary.

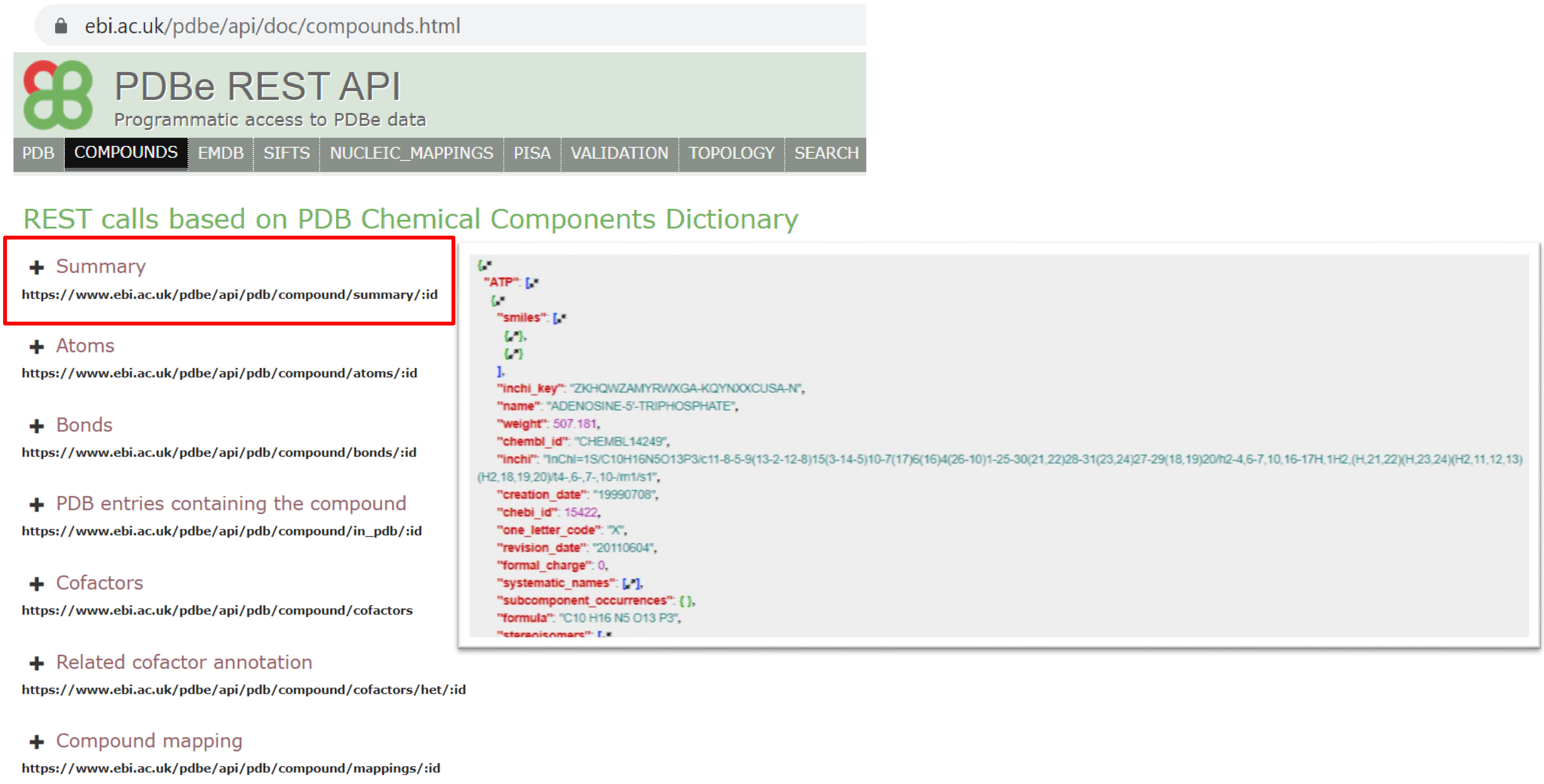
id	REA	String	Chemical component identifier, up to 3 characters long.
postdata		String	POST data should contain one or more comma-separated chemical component identifiers leaving the id field blank. If POST data is provided, POST request will be run instead of the default GET.

☑ Quotes

GET : https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/REA

HTTP status : 200 : OK

```
{
 "REA": [
   "1cbr",
   "1cbs",
   "1fem",
   "1g5y",
   "1gx9",
   "1n4h",
   "1rlb",
   "2acl",
   "2fr3",
   "2g78",
   "2lbd",
   "2ve3",
   "3a9e",
   "3cwk",
   "3fal",
   "3fc6",
   "4dm8",
   "4tns",
   "5fhz",
   "5uan",
```



ebi.ac.uk/pdbe/api/doc/compounds.html

PDBe REST API

Programmatic access to PDBe data

REST calls based on PDB Chemical Components Dictionary

+ Summary

https://www.ebi.ac.uk/pdbe/api/pdb/compound/summary/:id

+ Atoms

https://www.ebi.ac.uk/pdbe/api/pdb/compound/atoms/:id

+ Bonds

https://www.ebi.ac.uk/pdbe/api/pdb/compound/bonds/:id

+ PDB entries containing the compound

https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/:id

+ Cofactors

https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors

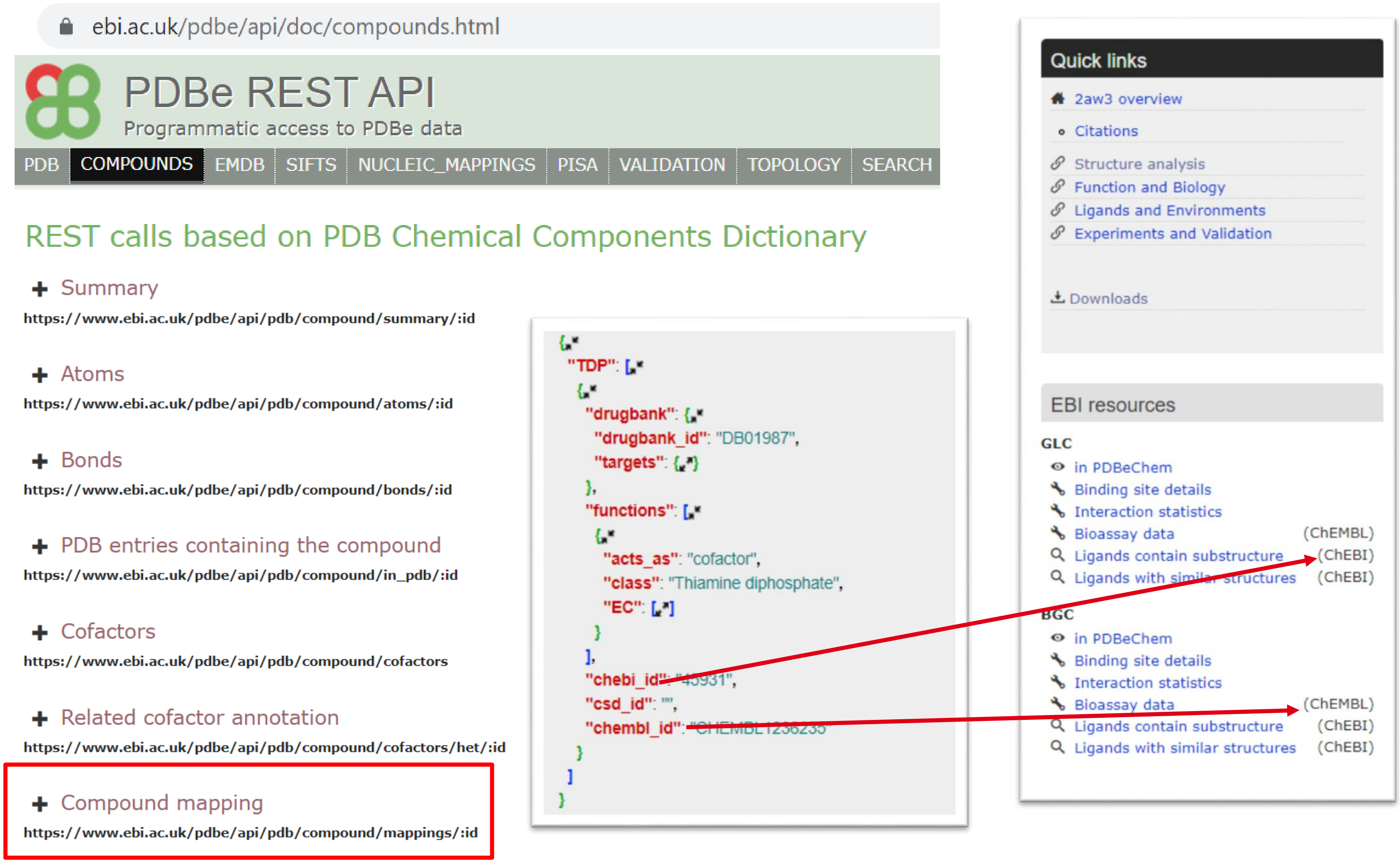
+ Related cofactor annotation

https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors/het/:id

+ Compound mapping

https://www.ebi.ac.uk/pdbe/api/pdb/compound/mappings/:id

```
{
 "ATP": [
  {
   "smiles": [
    {},
    {}
   ],
   "inchi_key": "ZKHQWZAMYRWXGA-KQYNXXCUSA-N",
   "name": "ADENOSINE-5'-TRIPHOSPHATE",
   "weight": 507.181,
   "chembl_id": "CHEMBL14249",
   "inchi": "InChI=1S/C10H16N5O13P3/c11-8-5-9(13-2-12-8)15(3-14-5)10-7(17)6(16)4(26-10)1-25-30(21,22)28-31(23,24)27-29(18,19)20/h2-4,6-7,10,16-17H,1H2,(H,21,22)(H,23,24)(H2,11,12,13)
(H2,18,19,20)/t4-,6-,7-,10-/m1/s1",
   "creation_date": "19990708",
   "chebi_id": 15422,
   "one_letter_code": "X",
   "revision_date": "20110604",
   "formal_charge": 0,
   "systematic_names": [],
   "subcomponent_occurrences": { },
   "formula": "C10 H16 N5 O13 P3",
   "stereoisomers": [
```



ebi.ac.uk/pdbe/api/doc/compounds.html

PDBe REST API

Programmatic access to PDBe data

REST calls based on PDB Chemical Components Dictionary

✚ Summary

https://www.ebi.ac.uk/pdbe/api/pdb/compound/summary/:id

✚ Atoms

https://www.ebi.ac.uk/pdbe/api/pdb/compound/atoms/:id

✚ Bonds

https://www.ebi.ac.uk/pdbe/api/pdb/compound/bonds/:id

✚ PDB entries containing the compound

https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/:id

✚ Cofactors

https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors

✚ Related cofactor annotation

https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors/het/:id

✚ Compound mapping

https://www.ebi.ac.uk/pdbe/api/pdb/compound/mappings/:id

```
{
 "TDP": [
  {
   "drugbank": {
    "drugbank_id": "DB01987",
    "targets": {}
   },
   "functions": [
    {
     "acts_as": "cofactor",
     "class": "Thiamine diphosphate",
     "EC": []
    }
   ],
   "chebi_id": "45931",
   "csd_id": "",
   "chembl_id": "CHEMBL1236235"
  }
 ]
}
```

Quick links

- 2aw3 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation
- Downloads

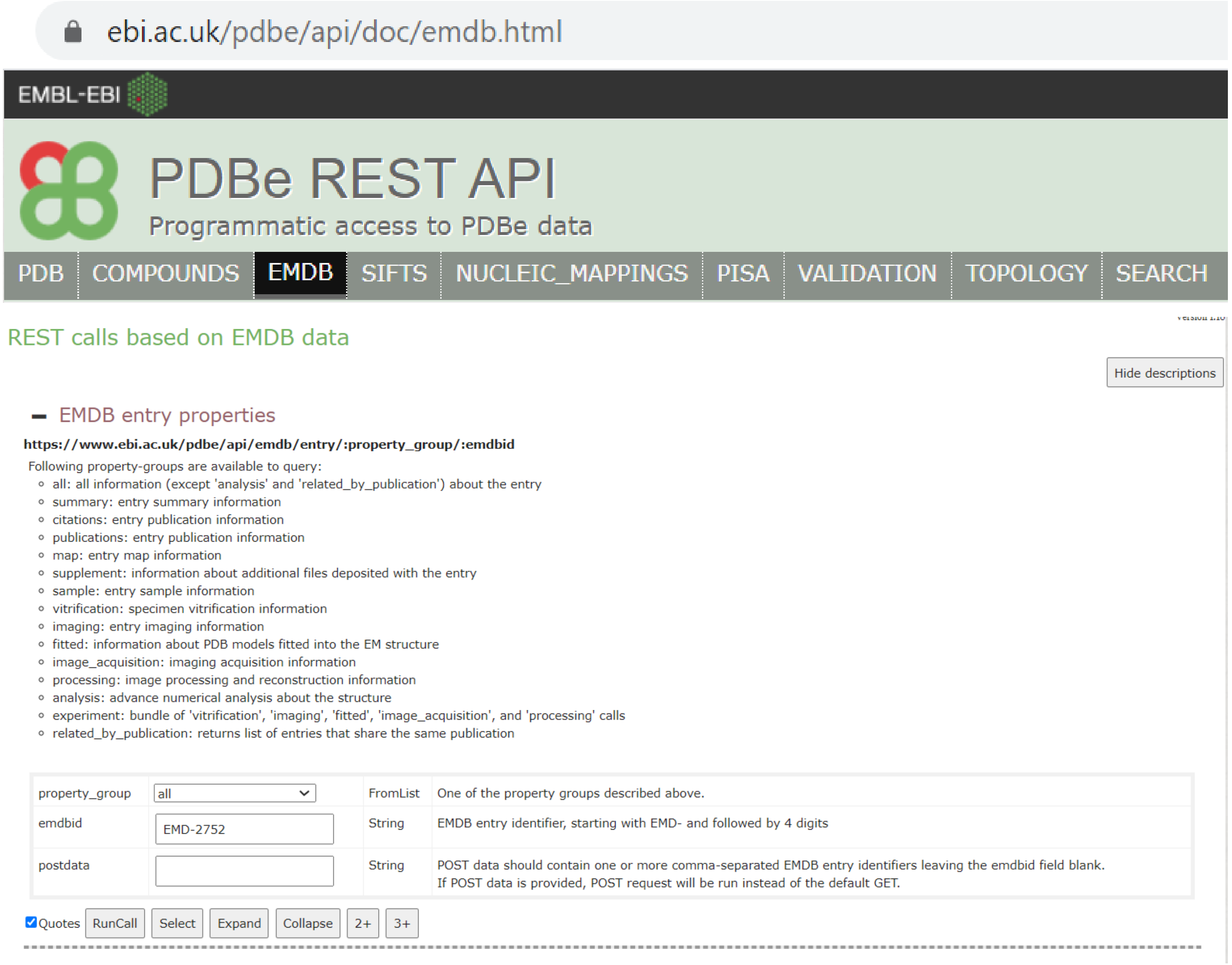
EBI resources

GLC

- in PDBeChem
- Binding site details
- Interaction statistics
- Bioassay data (ChEMBL)
- Ligands contain substructure (ChEBI)
- Ligands with similar structures (ChEBI)

BGC

- in PDBeChem
- Binding site details
- Interaction statistics
- Bioassay data (ChEMBL)
- Ligands contain substructure (ChEBI)
- Ligands with similar structures (ChEBI)



REST calls based on EMDB data

− EMDB entry properties

https://www.ebi.ac.uk/pdbe/api/emdb/entry/:property_group/:emdbid

Following property-groups are available to query:

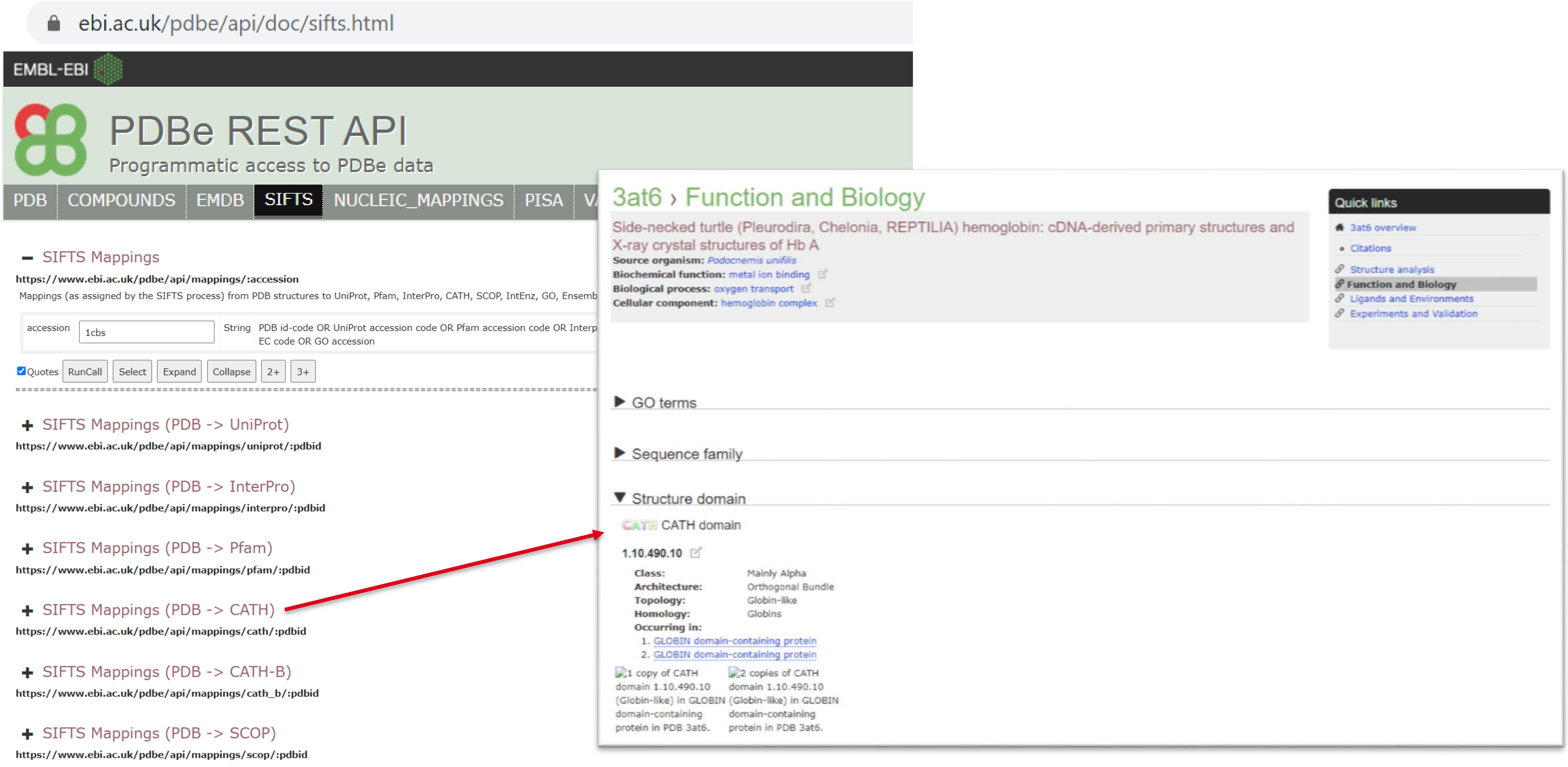
- all: all information (except 'analysis' and 'related_by_publication') about the entry
- summary: entry summary information
- citations: entry publication information
- publications: entry publication information
- map: entry map information
- supplement: information about additional files deposited with the entry
- sample: entry sample information
- vitrification: specimen vitrification information
- imaging: entry imaging information
- fitted: information about PDB models fitted into the EM structure
- image_acquisition: imaging acquisition information
- processing: image processing and reconstruction information
- analysis: advance numerical analysis about the structure
- experiment: bundle of 'vitrification', 'imaging', 'fitted', 'image_acquisition', and 'processing' calls
- related_by_publication: returns list of entries that share the same publication

property_group	all	FromList	One of the property groups described above.
emdbid	EMD-2752	String	EMDB entry identifier, starting with EMD- and followed by 4 digits
postdata		String	POST data should contain one or more comma-separated EMDB entry identifiers leaving the emdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

☑ Quotes

PDBe.org/API

EMBL-EBI



ebi.ac.uk/pdbe/api/doc/sifts.html

EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

PDB | COMPOUNDS | EMDB | SIFTS | NUCLEIC_MAPPINGS | PISA | V

− SIFTS Mappings

https://www.ebi.ac.uk/pdbe/api/mappings/:accession

Mappings (as assigned by the SIFTS process) from PDB structures to UniProt, Pfam, InterPro, CATH, SCOP, IntEnz, GO, Ensemb

accession	1cbs	String	PDB id-code OR UniProt accession code OR Pfam accession code OR Interp EC code OR GO accession

☑Quotes RunCall Select Expand Collapse 2+ 3+

+ SIFTS Mappings (PDB -> UniProt)

https://www.ebi.ac.uk/pdbe/api/mappings/uniprot/:pdbid

+ SIFTS Mappings (PDB -> InterPro)

https://www.ebi.ac.uk/pdbe/api/mappings/interpro/:pdbid

+ SIFTS Mappings (PDB -> Pfam)

https://www.ebi.ac.uk/pdbe/api/mappings/pfam/:pdbid

+ SIFTS Mappings (PDB -> CATH)

https://www.ebi.ac.uk/pdbe/api/mappings/cath/:pdbid

+ SIFTS Mappings (PDB -> CATH-B)

https://www.ebi.ac.uk/pdbe/api/mappings/cath_b/:pdbid

+ SIFTS Mappings (PDB -> SCOP)

https://www.ebi.ac.uk/pdbe/api/mappings/scop/:pdbid

3at6 › Function and Biology

Side-necked turtle (Pleurodira, Chelonia, REPTILIA) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A

Source organism: *Podocnemis unifilis*

Biochemical function: metal ion binding

Biological process: oxygen transport

Cellular component: hemoglobin complex

Quick links

- 3at6 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

▶ GO terms

▶ Sequence family

▼ Structure domain

CATH CATH domain

1.10.490.10

Class:	Mainly Alpha
Architecture:	Orthogonal Bundle
Topology:	Globin-like
Homology:	Globins

Occurring in:

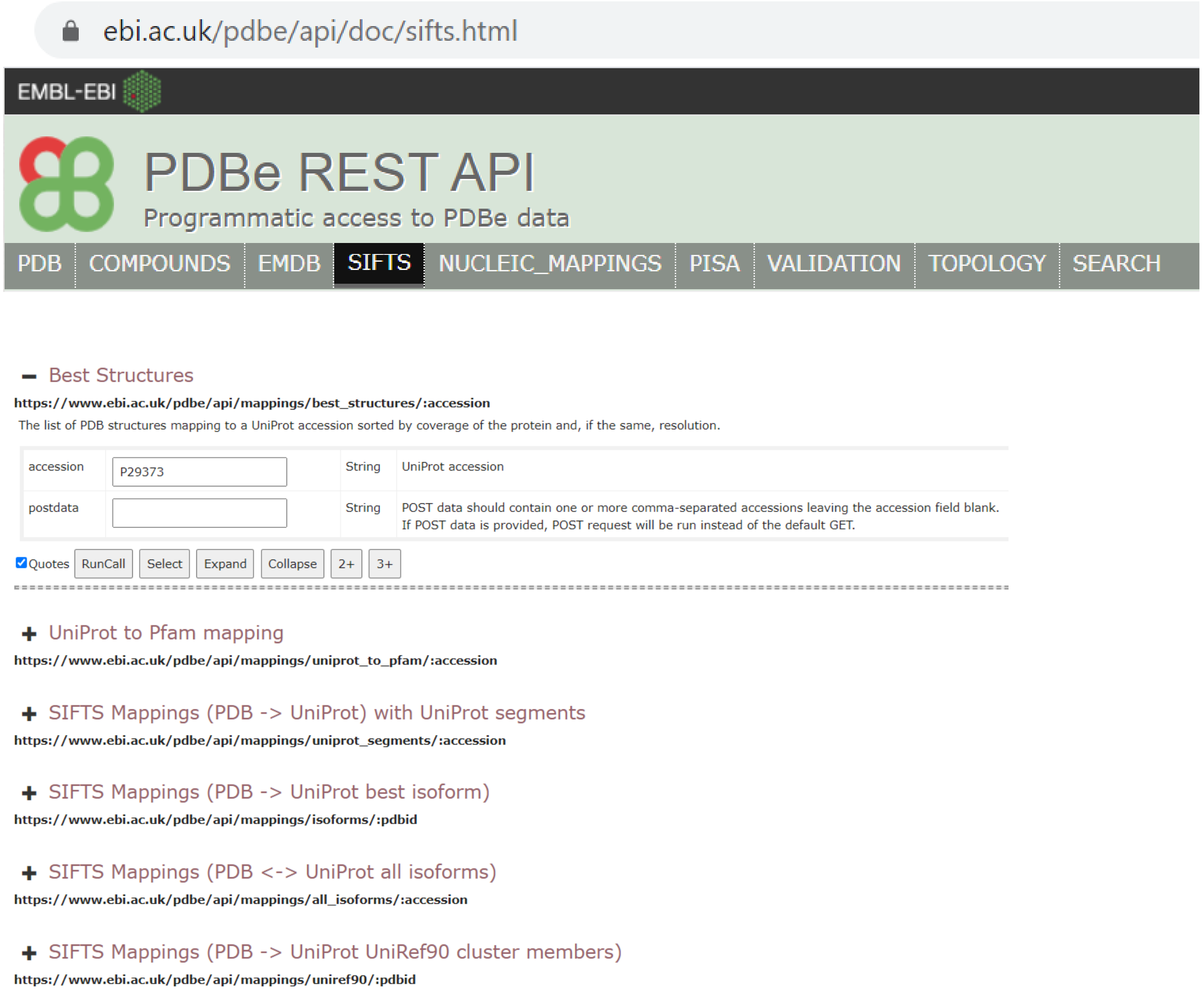
1. GLOBIN domain-containing protein
2. GLOBIN domain-containing protein

1 copy of CATH domain 1.10.490.10 (Globin-like) in GLOBIN domain-containing protein in PDB 3at6.

2 copies of CATH domain 1.10.490.10 (Globin-like) in GLOBIN domain-containing protein in PDB 3at6.

PDBe.org/API

EMBL-EBI



− Best Structures

https://www.ebi.ac.uk/pdbe/api/mappings/best_structures/:accession

The list of PDB structures mapping to a UniProt accession sorted by coverage of the protein and, if the same, resolution.

accession	P29373	String	UniProt accession
postdata		String	POST data should contain one or more comma-separated accessions leaving the accession field blank. If POST data is provided, POST request will be run instead of the default GET.

+ UniProt to Pfam mapping

https://www.ebi.ac.uk/pdbe/api/mappings/uniprot_to_pfam/:accession

+ SIFTS Mappings (PDB -> UniProt) with UniProt segments

https://www.ebi.ac.uk/pdbe/api/mappings/uniprot_segments/:accession

+ SIFTS Mappings (PDB -> UniProt best isoform)

https://www.ebi.ac.uk/pdbe/api/mappings/isoforms/:pdbid

+ SIFTS Mappings (PDB <-> UniProt all isoforms)

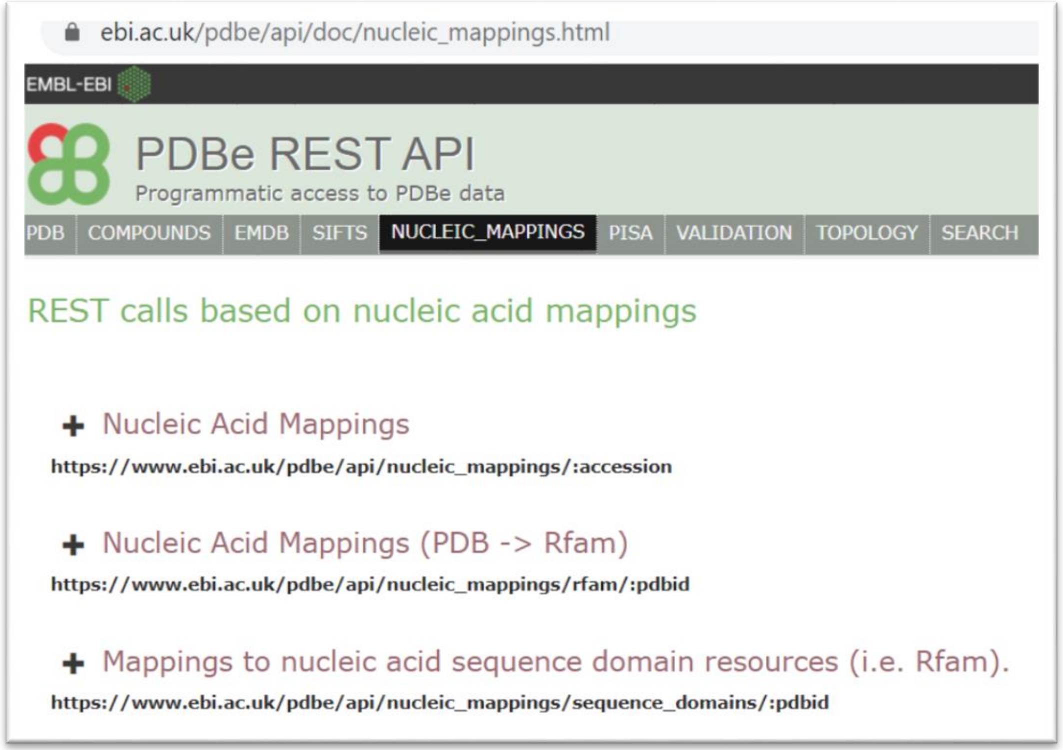
https://www.ebi.ac.uk/pdbe/api/mappings/all_isoforms/:accession

+ SIFTS Mappings (PDB -> UniProt UniRef90 cluster members)

https://www.ebi.ac.uk/pdbe/api/mappings/uniref90/:pdbid

PDBe.org/API

EMBL-EBI



REST calls based on nucleic acid mappings

+ Nucleic Acid Mappings

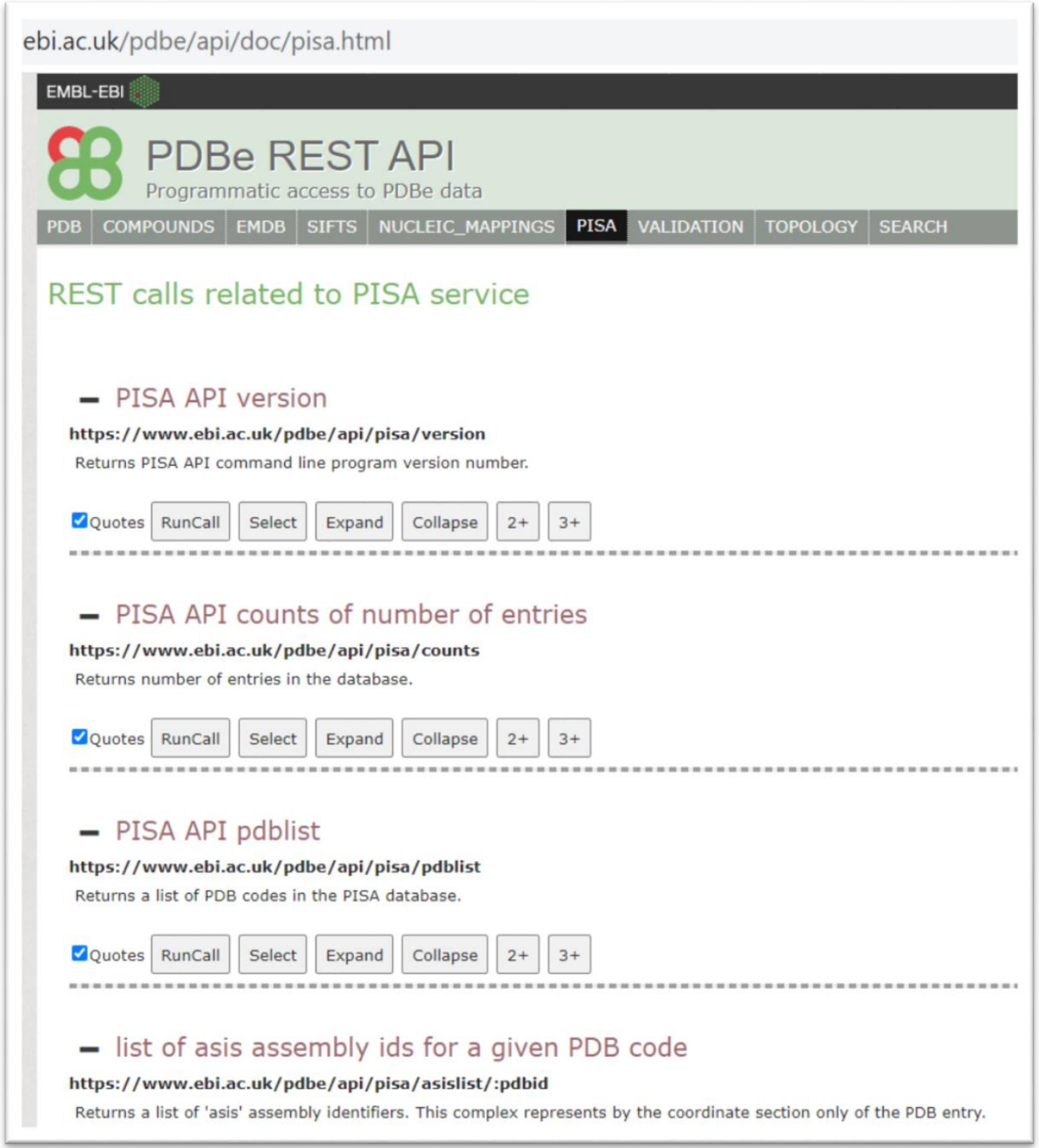
https://www.ebi.ac.uk/pdbe/api/nucleic_mappings/:accession

+ Nucleic Acid Mappings (PDB -> Rfam)

https://www.ebi.ac.uk/pdbe/api/nucleic_mappings/rfam/:pdbid

+ Mappings to nucleic acid sequence domain resources (i.e. Rfam).

https://www.ebi.ac.uk/pdbe/api/nucleic_mappings/sequence_domains/:pdbid



REST calls related to PISA service

— PISA API version

https://www.ebi.ac.uk/pdbe/api/pisa/version

Returns PISA API command line program version number.

☑Quotes RunCall Select Expand Collapse 2+ 3+

— PISA API counts of number of entries

https://www.ebi.ac.uk/pdbe/api/pisa/counts

Returns number of entries in the database.

☑Quotes RunCall Select Expand Collapse 2+ 3+

— PISA API pdblist

https://www.ebi.ac.uk/pdbe/api/pisa/pdblist

Returns a list of PDB codes in the PISA database.

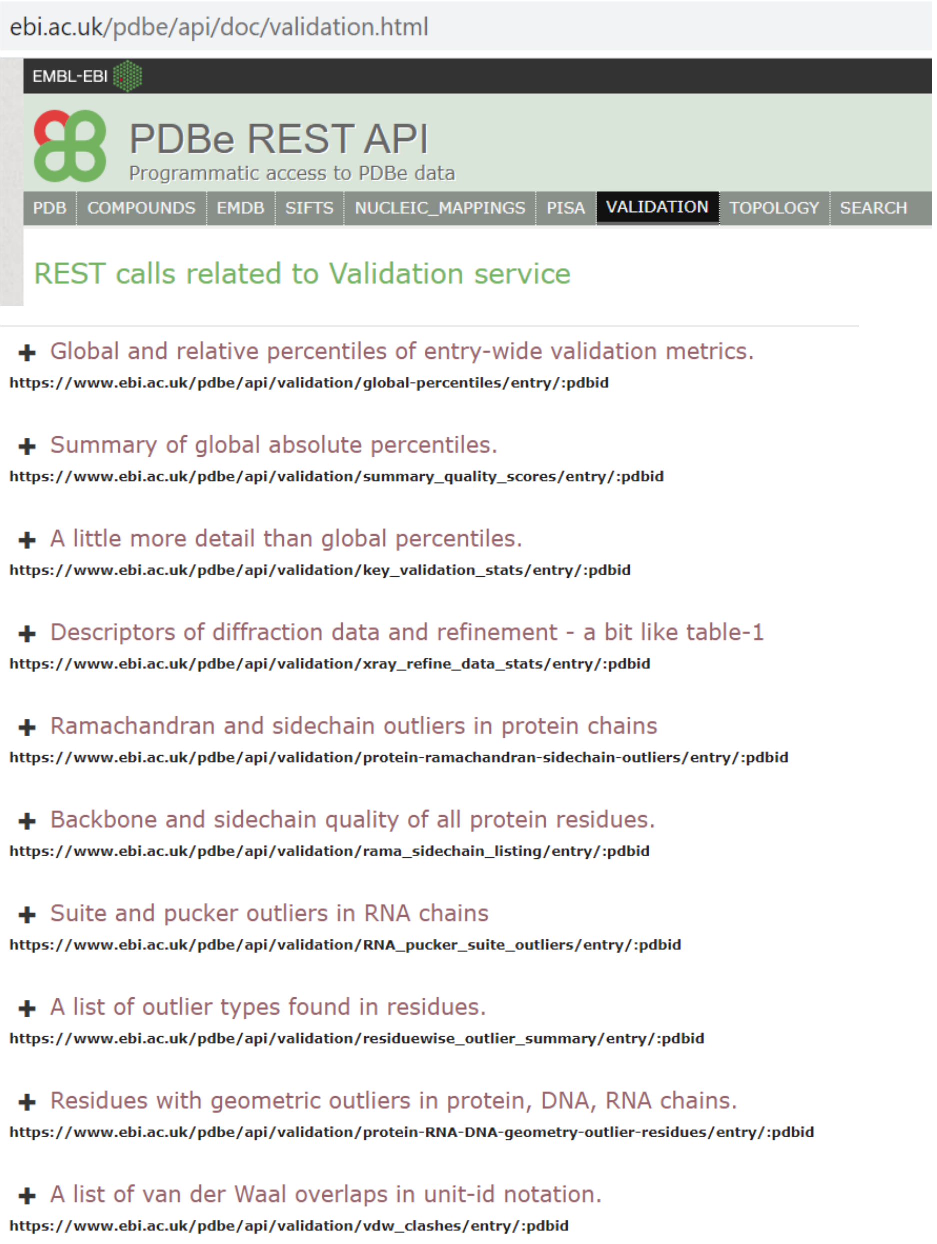
☑Quotes RunCall Select Expand Collapse 2+ 3+

— list of asis assembly ids for a given PDB code

https://www.ebi.ac.uk/pdbe/api/pisa/asislist/:pdbid

Returns a list of 'asis' assembly identifiers. This complex represents by the coordinate section only of the PDB entry.

PDBe.org/API



REST calls related to Validation service

+ Global and relative percentiles of entry-wide validation metrics.
https://www.ebi.ac.uk/pdbe/api/validation/global-percentiles/entry/:pdbid

+ Summary of global absolute percentiles.
https://www.ebi.ac.uk/pdbe/api/validation/summary_quality_scores/entry/:pdbid

+ A little more detail than global percentiles.
https://www.ebi.ac.uk/pdbe/api/validation/key_validation_stats/entry/:pdbid

+ Descriptors of diffraction data and refinement - a bit like table-1
https://www.ebi.ac.uk/pdbe/api/validation/xray_refine_data_stats/entry/:pdbid

+ Ramachandran and sidechain outliers in protein chains
https://www.ebi.ac.uk/pdbe/api/validation/protein-ramachandran-sidechain-outliers/entry/:pdbid

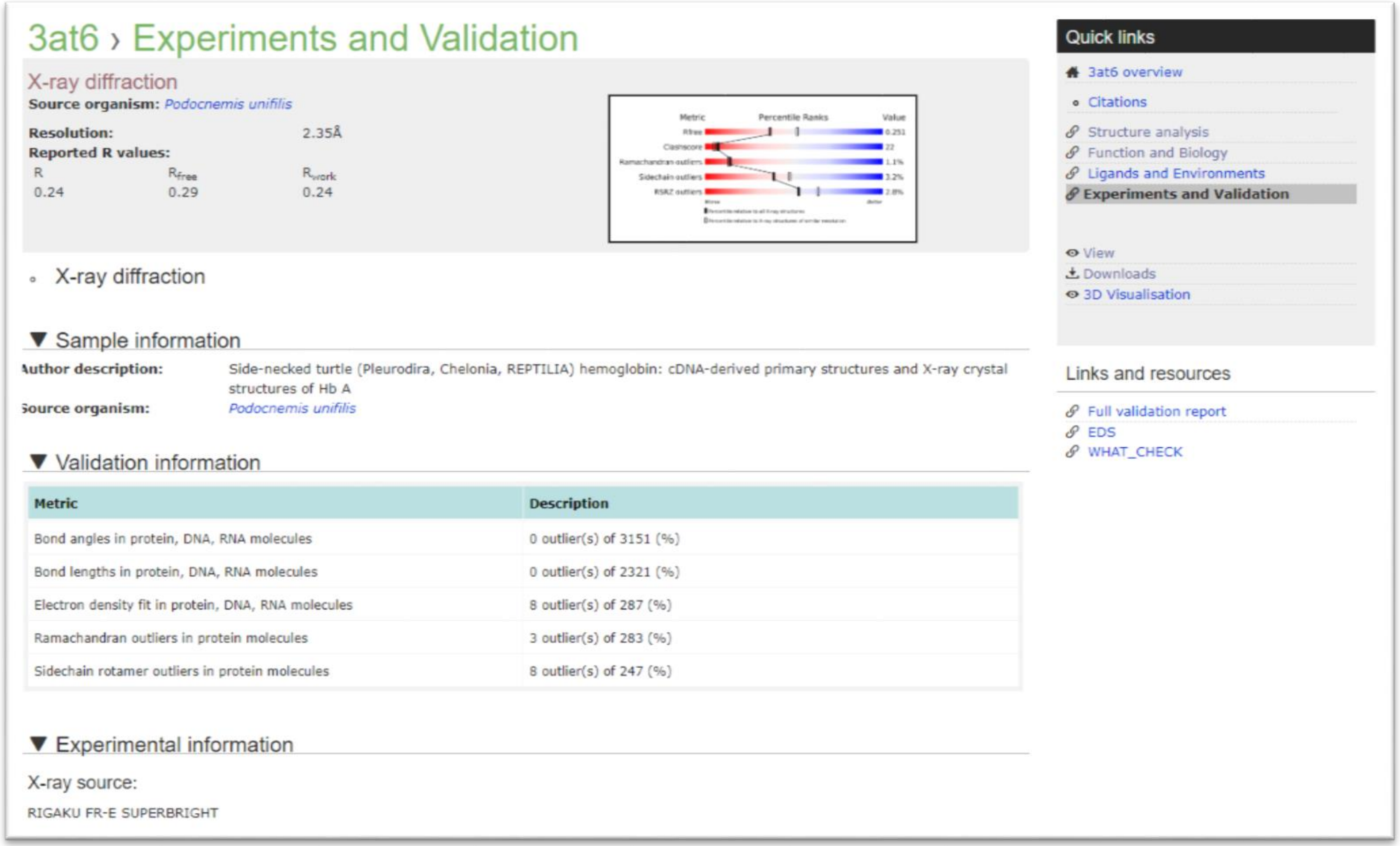
+ Backbone and sidechain quality of all protein residues.
https://www.ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/:pdbid

+ Suite and pucker outliers in RNA chains
https://www.ebi.ac.uk/pdbe/api/validation/RNA_pucker_suite_outliers/entry/:pdbid

+ A list of outlier types found in residues.
https://www.ebi.ac.uk/pdbe/api/validation/residuewise_outlier_summary/entry/:pdbid

+ Residues with geometric outliers in protein, DNA, RNA chains.
https://www.ebi.ac.uk/pdbe/api/validation/protein-RNA-DNA-geometry-outlier-residues/entry/:pdbid

+ A list of van der Waal overlaps in unit-id notation.
https://www.ebi.ac.uk/pdbe/api/validation/vdw_clashes/entry/:pdbid



3at6 › Experiments and Validation

X-ray diffraction

Source organism: *Podocnemis unifilis*

Resolution: 2.35Å

Reported R values:

R	R_{free}	R_{work}
0.24	0.29	0.24

- X-ray diffraction

▼ Sample information

Author description: Side-necked turtle (Pleurodira, Chelonia, REPTILIA) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A

Source organism: *Podocnemis unifilis*

▼ Validation information

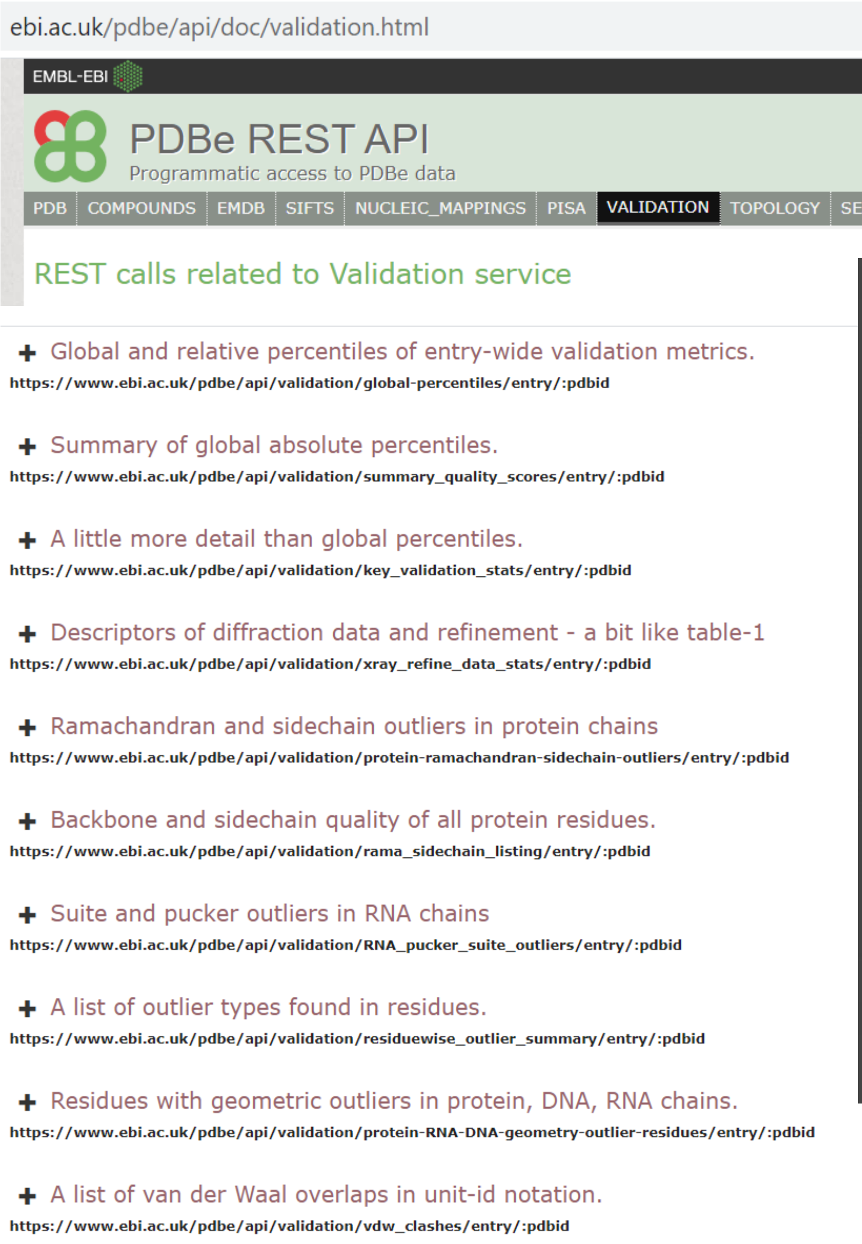
Metric	Description
Bond angles in protein, DNA, RNA molecules	0 outlier(s) of 3151 (%)
Bond lengths in protein, DNA, RNA molecules	0 outlier(s) of 2321 (%)
Electron density fit in protein, DNA, RNA molecules	8 outlier(s) of 287 (%)
Ramachandran outliers in protein molecules	3 outlier(s) of 283 (%)
Sidechain rotamer outliers in protein molecules	8 outlier(s) of 247 (%)

▼ Experimental information

X-ray source:

RIGAKU FR-E SUPERBRIGHT

PDBe.org/API



ebi.ac.uk/pdbe/api/doc/validation.html

EMBL-EBI

PDBe REST API

Programmatic access to PDBe data

PDB | COMPOUNDS | EMDB | SIFTS | NUCLEIC_MAPPINGS | PISA | VALIDATION | TOPOLOGY | SEARCH

REST calls related to Validation service

+ Global and relative percentiles of entry-wide validation metrics.
https://www.ebi.ac.uk/pdbe/api/validation/global-percentiles/entry/:pdbid

+ Summary of global absolute percentiles.
https://www.ebi.ac.uk/pdbe/api/validation/summary_quality_scores/entry/:pdbid

+ A little more detail than global percentiles.
https://www.ebi.ac.uk/pdbe/api/validation/key_validation_stats/entry/:pdbid

+ Descriptors of diffraction data and refinement - a bit like table-1
https://www.ebi.ac.uk/pdbe/api/validation/xray_refine_data_stats/entry/:pdbid

+ Ramachandran and sidechain outliers in protein chains
https://www.ebi.ac.uk/pdbe/api/validation/protein-ramachandran-sidechain-outliers/entry/:pdbid

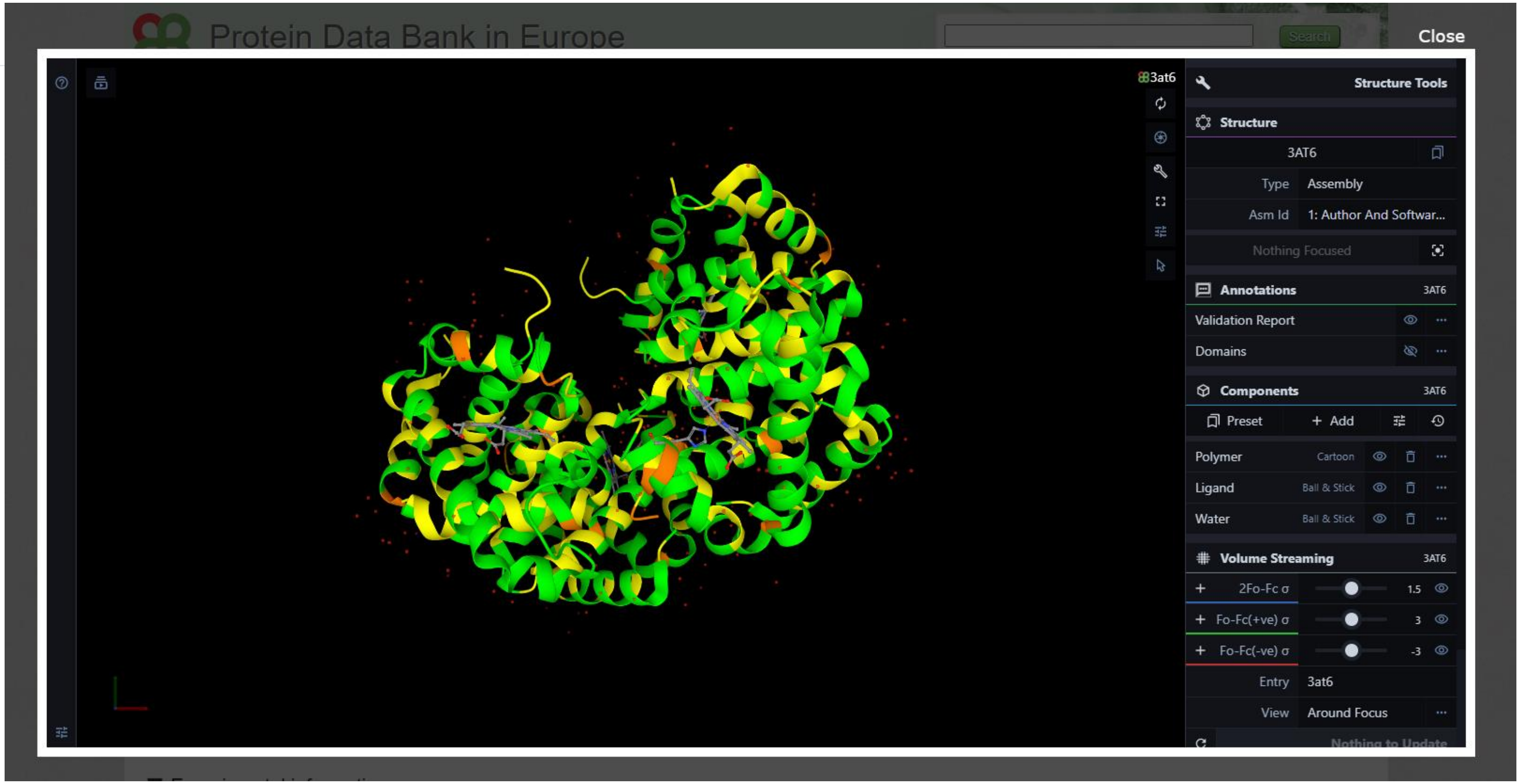
+ Backbone and sidechain quality of all protein residues.
https://www.ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/:pdbid

+ Suite and pucker outliers in RNA chains
https://www.ebi.ac.uk/pdbe/api/validation/RNA_pucker_suite_outliers/entry/:pdbid

+ A list of outlier types found in residues.
https://www.ebi.ac.uk/pdbe/api/validation/residuewise_outlier_summary/entry/:pdbid

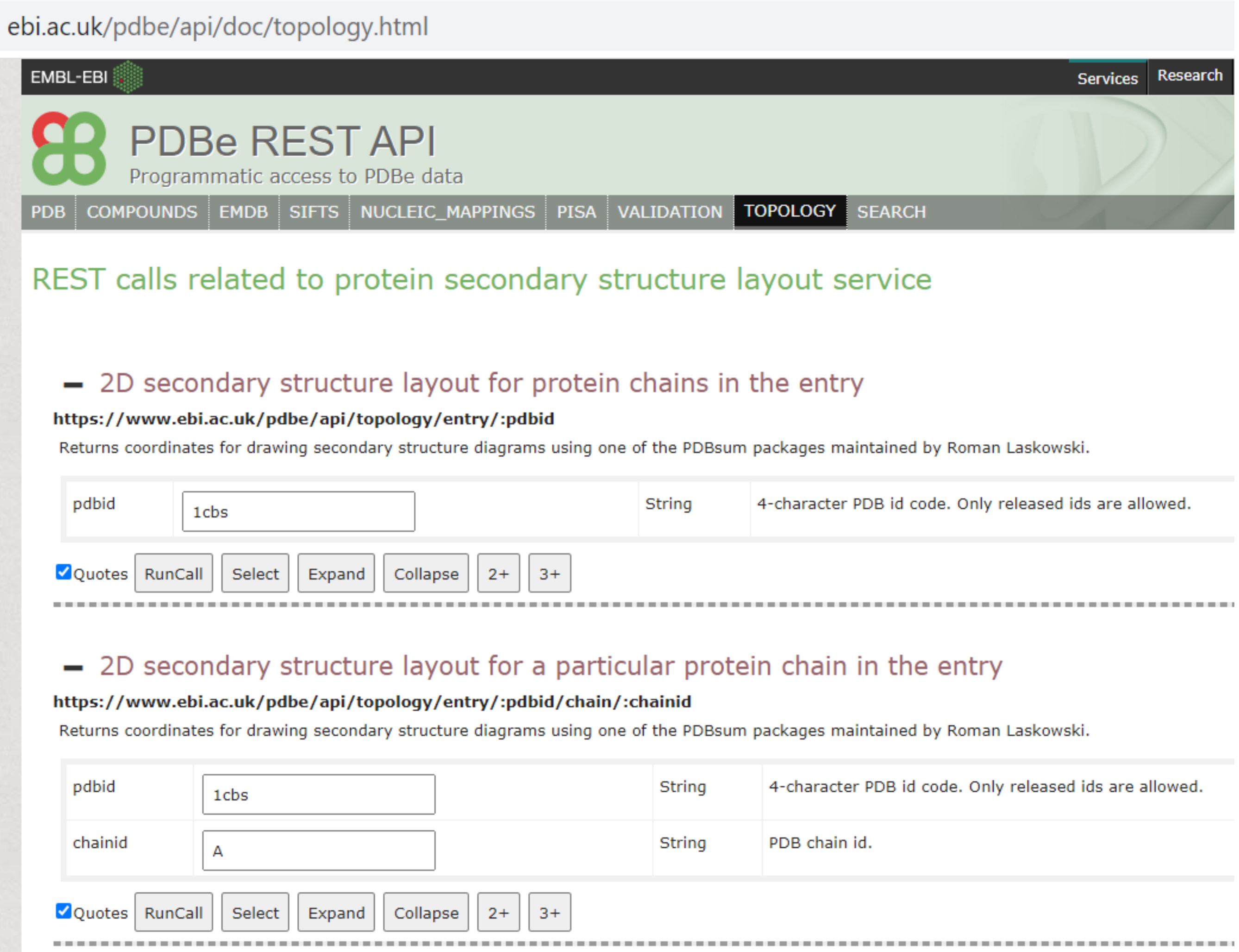
+ Residues with geometric outliers in protein, DNA, RNA chains.
https://www.ebi.ac.uk/pdbe/api/validation/protein-RNA-DNA-geometry-outlier-residues/entry/:pdbid

+ A list of van der Waal overlaps in unit-id notation.
https://www.ebi.ac.uk/pdbe/api/validation/vdw_clashes/entry/:pdbid



PDBe.org/API

EMBL-EBI



ebi.ac.uk/pdbe/api/doc/topology.html

PDBe REST API

Programmatic access to PDBe data

REST calls related to protein secondary structure layout service

– 2D secondary structure layout for protein chains in the entry

https://www.ebi.ac.uk/pdbe/api/topology/entry/:pdbid

Returns coordinates for drawing secondary structure diagrams using one of the PDBsum packages maintained by Roman Laskowski.

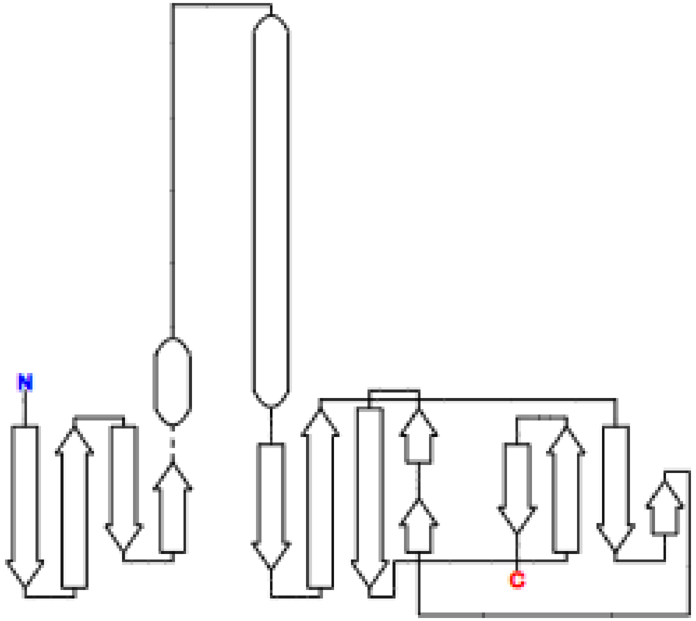
pdbid	1cbs	String	4-character PDB id code. Only released ids are allowed.

– 2D secondary structure layout for a particular protein chain in the entry

https://www.ebi.ac.uk/pdbe/api/topology/entry/:pdbid/chain/:chainid

Returns coordinates for drawing secondary structure diagrams using one of the PDBsum packages maintained by Roman Laskowski.

pdbid	1cbs	String	4-character PDB id code. Only released ids are allowed.
chainid	A	String	PDB chain id.



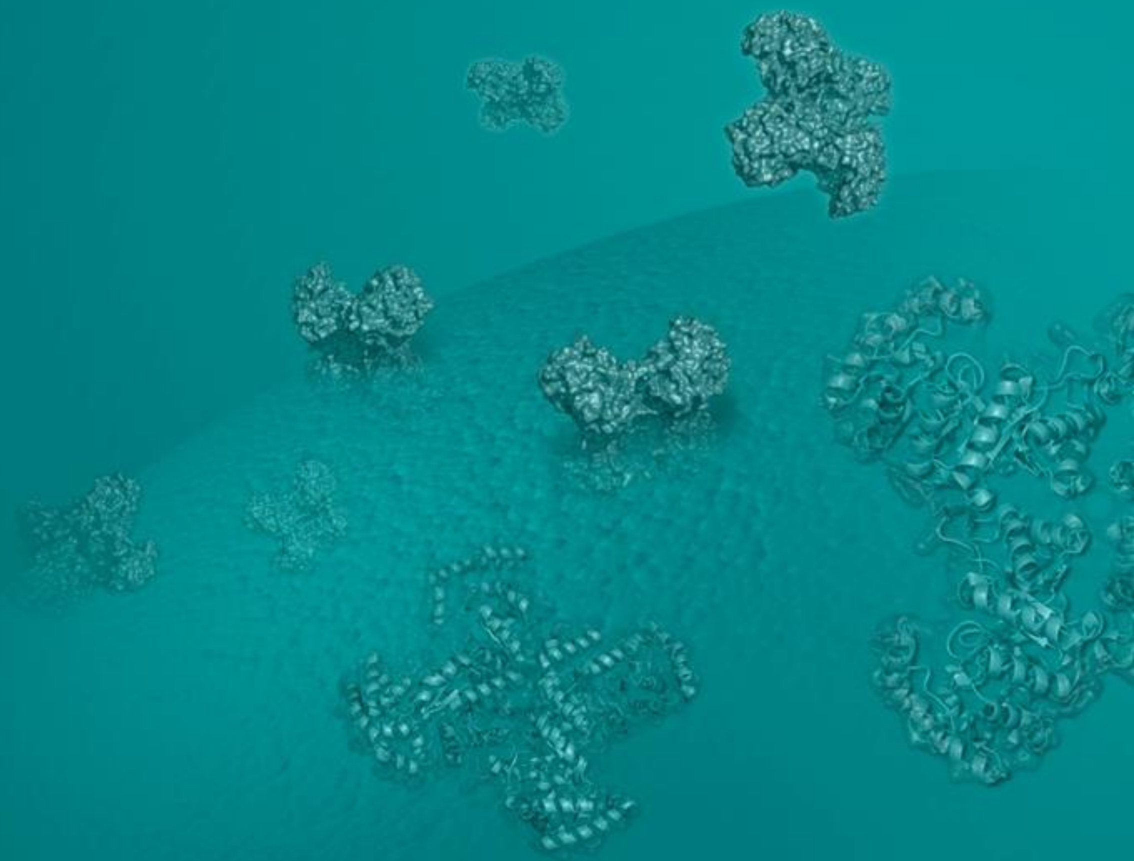
PDBe REST API: Entry-based API

- HTTP GET request for individual queries
- HTTP POST requests for batched queries (not every call allows POST requests)
- JSON output
- HTTP status codes:
  - HTTP 200: OK
  - HTTP 404: What you are looking for is not present (e.g the entry does not have any mappings when calling https://www.ebi.ac.uk/pdbe/api/mappings/4rj1)
  - HTTP 5XX: Something bad happened. Please report.

PDBe REST API: Entry-based API

- No fixed release cycle.
- We keep backwards compatibility (field + value names).
- We use most endpoints in our pages – the majority of calls are subjected to weekly testing.
- Mailing list: pdbe-api-users@ebi.ac.uk (pdbe-api-users-join@ebi.ac.uk)
- We send out emails to the mailing list to test on wwwdev weeks before any release happens.

Case studies



Case studies

- Case study 1: Compile a dataset of ligand-binding domains
- Case study 2: Validation

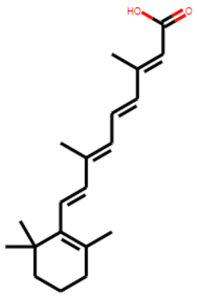
Case study 1: Compiling a dataset of ligand-binding domains using the REST API

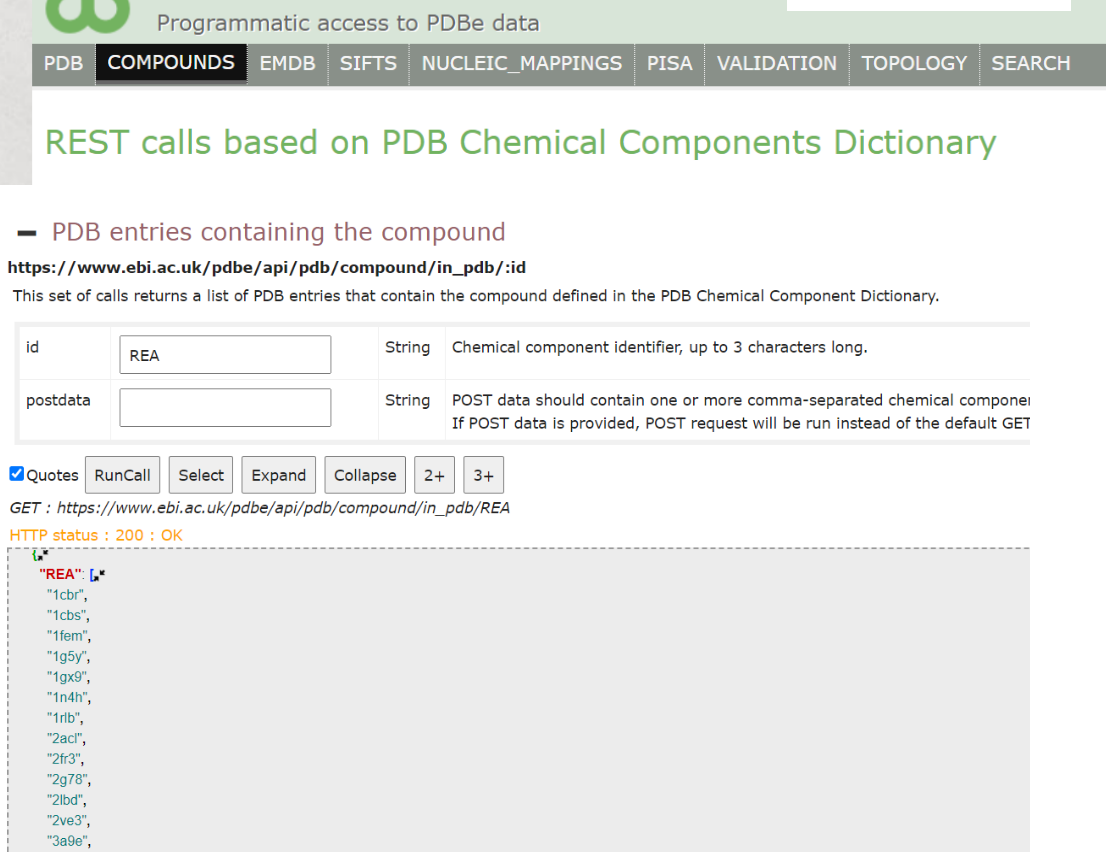
Objective:

- Find PDB entries containing a specific ligand (REA)
- Analyze the domain composition of those entries
- Find CATH domains which interact with the same ligand
- Find a set of chains in the entries that contain CATH domain and bind to REA, too

Let's start with the **compound/in_pdb** endpoint in the **compounds** tab.

Search for retinoic acid (PDBe chemical component 'REA') using this call. It returns a list of 31 PDB IDs.





REST calls based on PDB Chemical Components Dictionary

– PDB entries containing the compound

https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/:id

This set of calls returns a list of PDB entries that contain the compound defined in the PDB Chemical Component Dictionary.

id	REA	String	Chemical component identifier, up to 3 characters long.
postdata		String	POST data should contain one or more comma-separated chemical componei If POST data is provided, POST request will be run instead of the default GET

☑ Quotes RunCall Select Expand Collapse 2+ 3+

GET : https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/REA

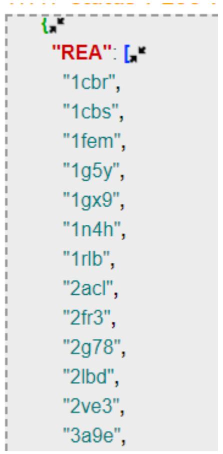
HTTP status : 200 : OK

```
{
 "REA": [
  "1cbr",
  "1cbs",
  "1fem",
  "1g5y",
  "1gx9",
  "1n4h",
  "1rlb",
  "2acl",
  "2fr3",
  "2g78",
  "2lbd",
  "2ve3",
  "3a9e",
```

Case study 1: Compiling a dataset of ligand-binding domains using the REST API

Now let's figure the domain composition of these entries by using the **/mappings** call in the **SIFTS** tab.

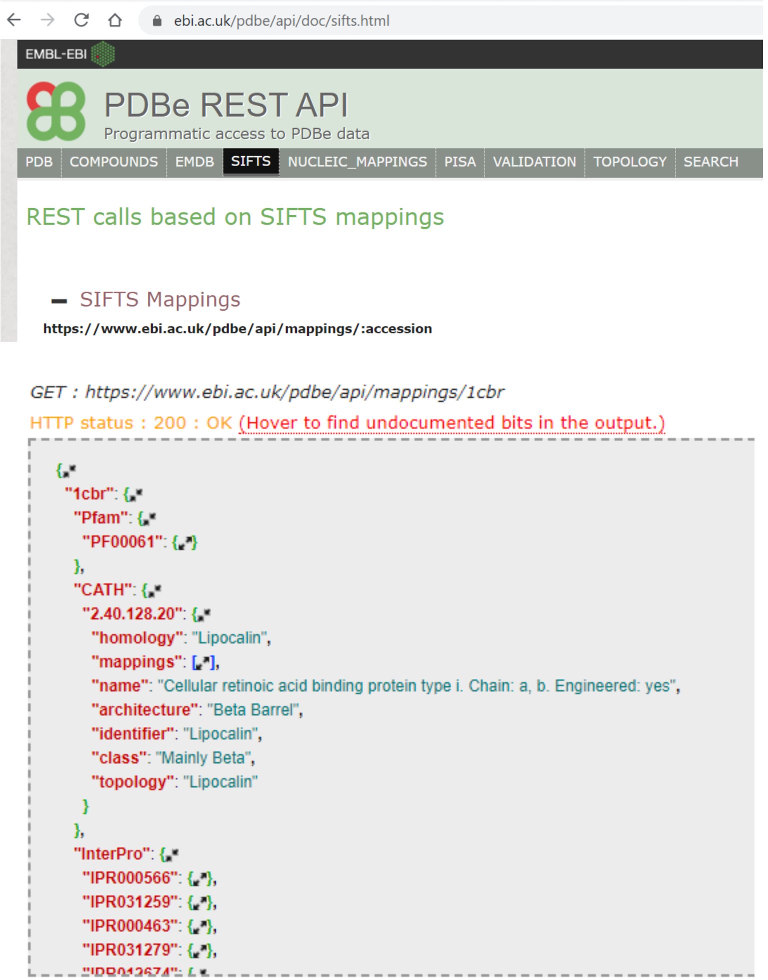
We will make one mapping call for each PDB entry id, programmatically.



```
{
  "REA": [
    "1cbr",
    "1cbs",
    "1fem",
    "1g5y",
    "1gx9",
    "1n4h",
    "1rlb",
    "2acl",
    "2fr3",
    "2g78",
    "2lbd",
    "2ve3",
    "3a9e",
```

```
For id in pdbid_list
   Make get request
https://www.ebi.ac.uk/pdbe/api/mappings/{id}

 or
  Make post request using pdbid_list
```



REST calls based on SIFTS mappings

– SIFTS Mappings

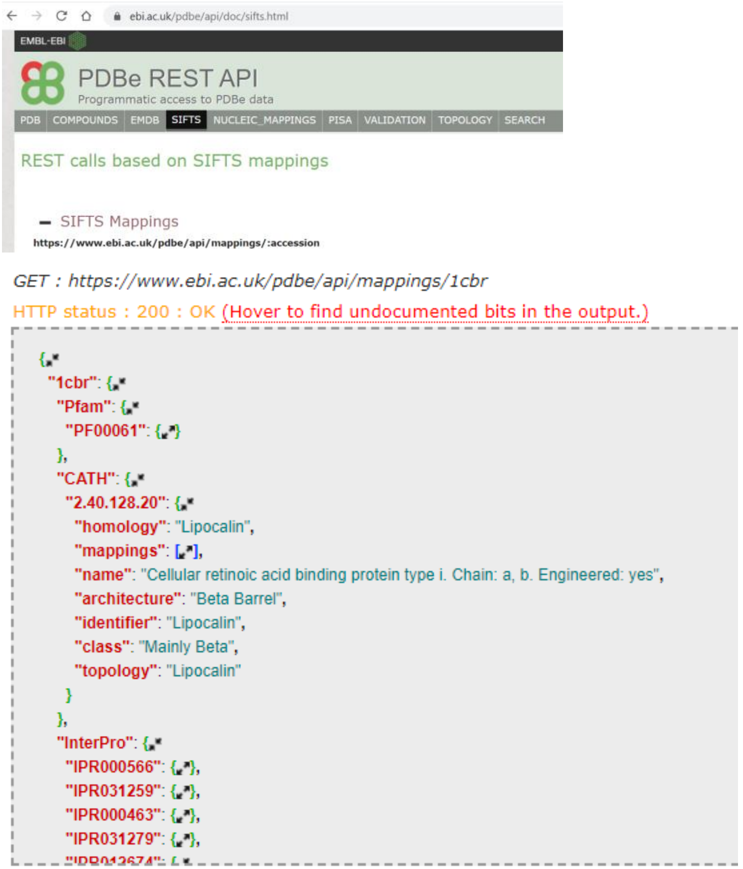
https://www.ebi.ac.uk/pdbe/api/mappings/:accession

GET : https://www.ebi.ac.uk/pdbe/api/mappings/1cbr

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

```
{
  "1cbr": {
    "Pfam": {
      "PF00061": {}
    },
    "CATH": {
      "2.40.128.20": {
        "homology": "Lipocalin",
        "mappings": [],
        "name": "Cellular retinoic acid binding protein type i. Chain: a, b. Engineered: yes",
        "architecture": "Beta Barrel",
        "identifier": "Lipocalin",
        "class": "Mainly Beta",
        "topology": "Lipocalin"
      }
    },
    "InterPro": {
      "IPR000566": {},
      "IPR031259": {},
      "IPR000463": {},
      "IPR031279": {},
      "IPR012674": {
```

Case study 1: Compiling a dataset of ligand-binding domains using the REST API



REST calls based on SIFTS mappings

- SIFTS Mappings

https://www.ebi.ac.uk/pdbe/api/mappings/:accession

GET : https://www.ebi.ac.uk/pdbe/api/mappings/1cbr

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

```
{
  "1cbr": {
    "Pfam": {
      "PF00061": {}
    },
    "CATH": {
      "2.40.128.20": {
        "homology": "Lipocalin",
        "mappings": [],
        "name": "Cellular retinoic acid binding protein type i. Chain: a, b. Engineered: yes",
        "architecture": "Beta Barrel",
        "identifier": "Lipocalin",
        "class": "Mainly Beta",
        "topology": "Lipocalin"
      }
    },
    "InterPro": {
      "IPR000566": {},
      "IPR031259": {},
      "IPR000463": {},
      "IPR031279": {},
      "IPR012674": {
```

JSON output returns each entry's mapping to Pfam, CATH, InterPro, etc

```
For pdbid in json_output:
   Count the type of mappings for {pdbid} based on json_output[pdbid][dom_type]
```

```
31 entries have Pfam mappings.
31 entries have InterPro mappings.
26 entries have CATH mappings.
 4 entries have EC mappings.
31 entries have UniProt mappings.
17 entries have SCOP mappings.
31 entries have GO mappings.
```

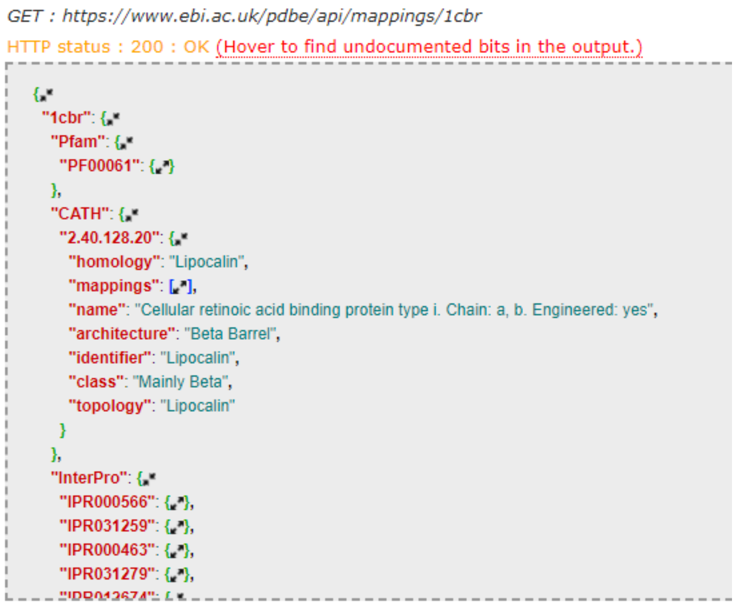
Let us focus on CATH superfamilies.

Let's count how many times each CATH superfamily is mapped onto entries containing the compound of our interest.

Case study 1: Compiling a dataset of ligand-binding domains using the REST API

Let us focus on CATH superfamilies.

Let’s count how many times each superfamily is mapped onto entries containing the compound of our interest.



GET : https://www.ebi.ac.uk/pdbe/api/mappings/1cbr

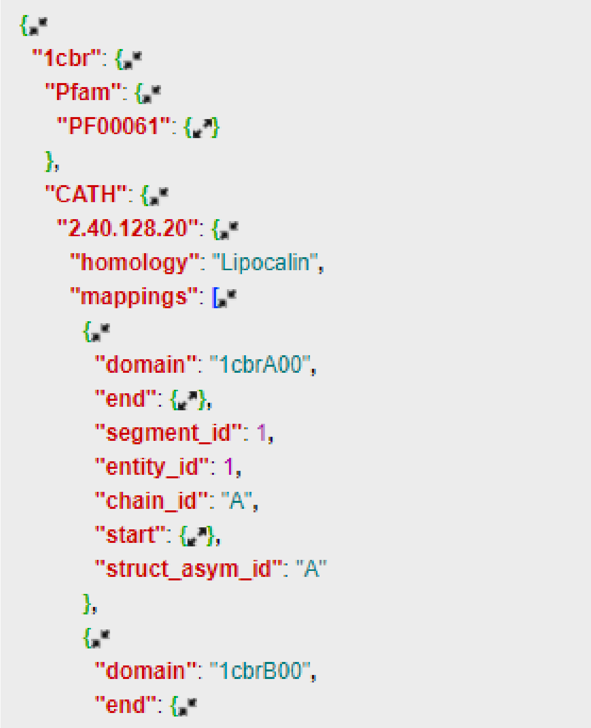
HTTP status : 200 : OK (Hover to find undocumented bits in the output.)

```
{
 "1cbr": {
  "Pfam": {
   "PF00061": {}
  },
  "CATH": {
   "2.40.128.20": {
    "homology": "Lipocalin",
    "mappings": [],
    "name": "Cellular retinoic acid binding protein type i. Chain: a, b. Engineered: yes",
    "architecture": "Beta Barrel",
    "identifier": "Lipocalin",
    "class": "Mainly Beta",
    "topology": "Lipocalin"
   }
  },
  "InterPro": {
   "IPR000566": {},
   "IPR031259": {},
   "IPR000463": {},
   "IPR031279": {},
   "IPR012674": {
```

```
For pdbid in json_output:
  If json_output[pdbid] does not contain “CATH”
    continue
  Else
    Store each CATH superfamily ID based on json_output[pdbid][“CATH”][domid][“homology”]
```

```
 9 entries contain CATH superfamily 2.40.128.20 (Lipocalin).
 2 entries contain CATH superfamily 2.60.40.180 (Immunoglobulin-like).
 3 entries contain CATH superfamily  3.30.50.10 (Erythroid Transcription Factor GATA-1, subunit A).
15 entries contain CATH superfamily 1.10.565.10 (Retinoid X Receptor).
 1 entries contain CATH superfamily 1.10.630.10 (Cytochrome p450).
```

Case study 1: Compiling a dataset of ligand-binding domains using the REST API



```
{
  "1cbr": {
    "Pfam": {
      "PF00061": {}
    },
    "CATH": {
      "2.40.128.20": {
        "homology": "Lipocalin",
        "mappings": [
          {
            "domain": "1cbrA00",
            "end": {},
            "segment_id": 1,
            "entity_id": 1,
            "chain_id": "A",
            "start": {},
            "struct_asym_id": "A"
          },
          {
            "domain": "1cbrB00",
            "end": {
```

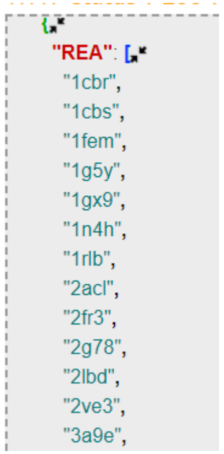
Map pdbid to chain_id and domid

```
For pdbid in json_output:
   for dominfo in json_output[pdbid][“CATH”][domid][“mappings”]
      store dominfo[“chain_id”]
```

```
PDB id 3fal chain D contain domains 1.10.565.10
PDB id 3fc6 chain B contain domains 1.10.565.10
PDB id 3dzy chain A contain domains 3.30.50.10, 1.10.565.10
PDB id 1xls chain B contain domains 1.10.565.10
PDB id 1xiu chain B contain domains 1.10.565.10
```

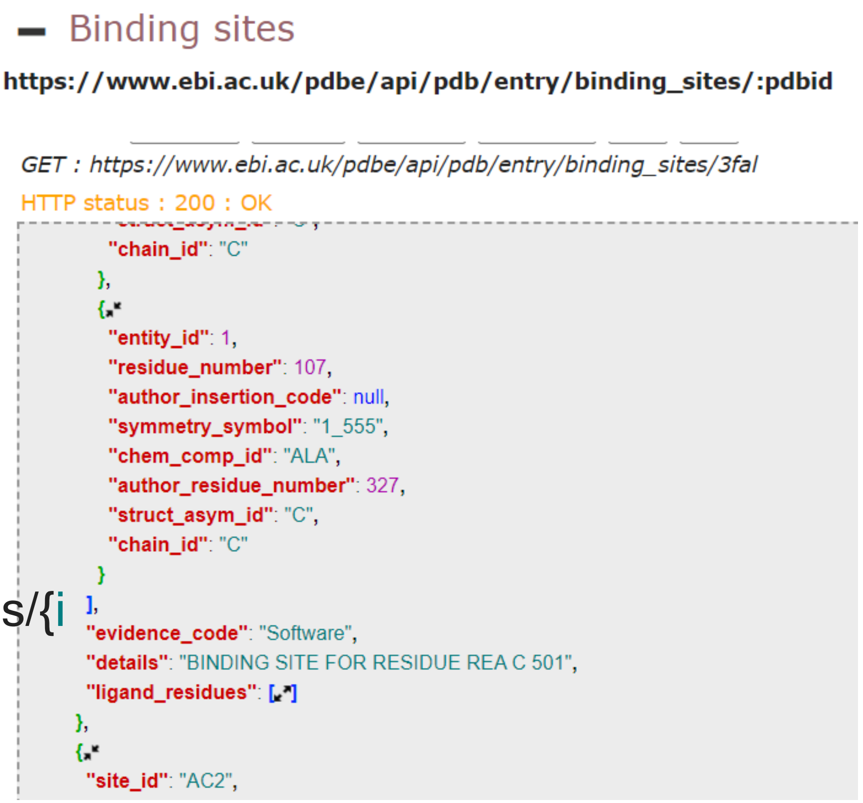
Case study 1: Compiling a dataset of ligand-binding domains using the REST API

Now let us turn our attention to binding sites information. For each of the pdb id we obtained earlier, make get request **to /binding_sites** call in the **pdb tab**



```
{
  "REA": [
    "1cbr",
    "1cbs",
    "1fem",
    "1g5y",
    "1gx9",
    "1n4h",
    "1rlb",
    "2acl",
    "2fr3",
    "2g78",
    "2lbd",
    "2ve3",
    "3a9e",
```

```
For id in pdbid_list
  Make get request
https://www.ebi.ac.uk/pdbe/api/pdb/entry/binding_sites/{id}
```



— Binding sites

https://www.ebi.ac.uk/pdbe/api/pdb/entry/binding_sites/:pdbid

GET : https://www.ebi.ac.uk/pdbe/api/pdb/entry/binding_sites/3fal

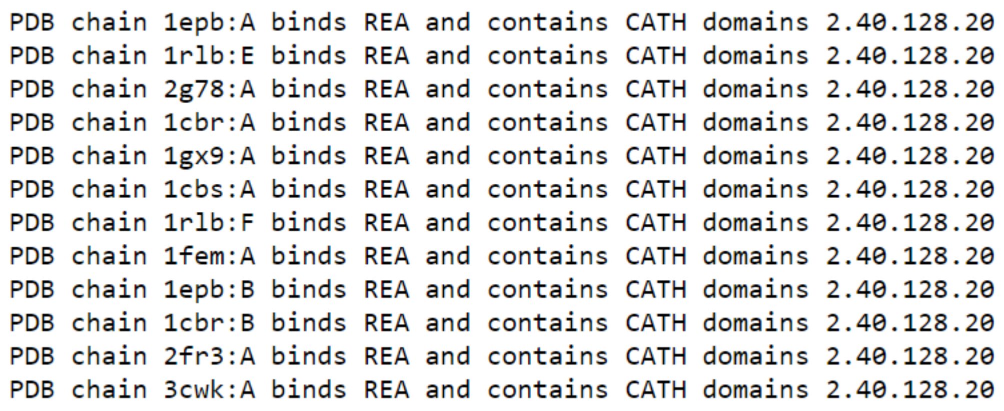
HTTP status : 200 : OK

```
      "chain_id": "C"
    },
    {
      "entity_id": 1,
      "residue_number": 107,
      "author_insertion_code": null,
      "symmetry_symbol": "1_555",
      "chem_comp_id": "ALA",
      "author_residue_number": 327,
      "struct_asym_id": "C",
      "chain_id": "C"
    }
  ],
  "evidence_code": "Software",
  "details": "BINDING SITE FOR RESIDUE REA C 501",
  "ligand_residues": []
},
{
  "site_id": "AC2",
```

Case study 1: Compiling a dataset of ligand-binding domains using the REST API

Now we can find chains common to binding sites and CATH superfamilies to find instances of domains with binding ligands.

residue_info[“chain_id”]
dominfo[“chain_id”]



```
PDB chain 1epb:A binds REA and contains CATH domains 2.40.128.20
PDB chain 1rlb:E binds REA and contains CATH domains 2.40.128.20
PDB chain 2g78:A binds REA and contains CATH domains 2.40.128.20
PDB chain 1cbr:A binds REA and contains CATH domains 2.40.128.20
PDB chain 1gx9:A binds REA and contains CATH domains 2.40.128.20
PDB chain 1cbs:A binds REA and contains CATH domains 2.40.128.20
PDB chain 1rlb:F binds REA and contains CATH domains 2.40.128.20
PDB chain 1fem:A binds REA and contains CATH domains 2.40.128.20
PDB chain 1epb:B binds REA and contains CATH domains 2.40.128.20
PDB chain 1cbr:B binds REA and contains CATH domains 2.40.128.20
PDB chain 2fr3:A binds REA and contains CATH domains 2.40.128.20
PDB chain 3cwk:A binds REA and contains CATH domains 2.40.128.20
```

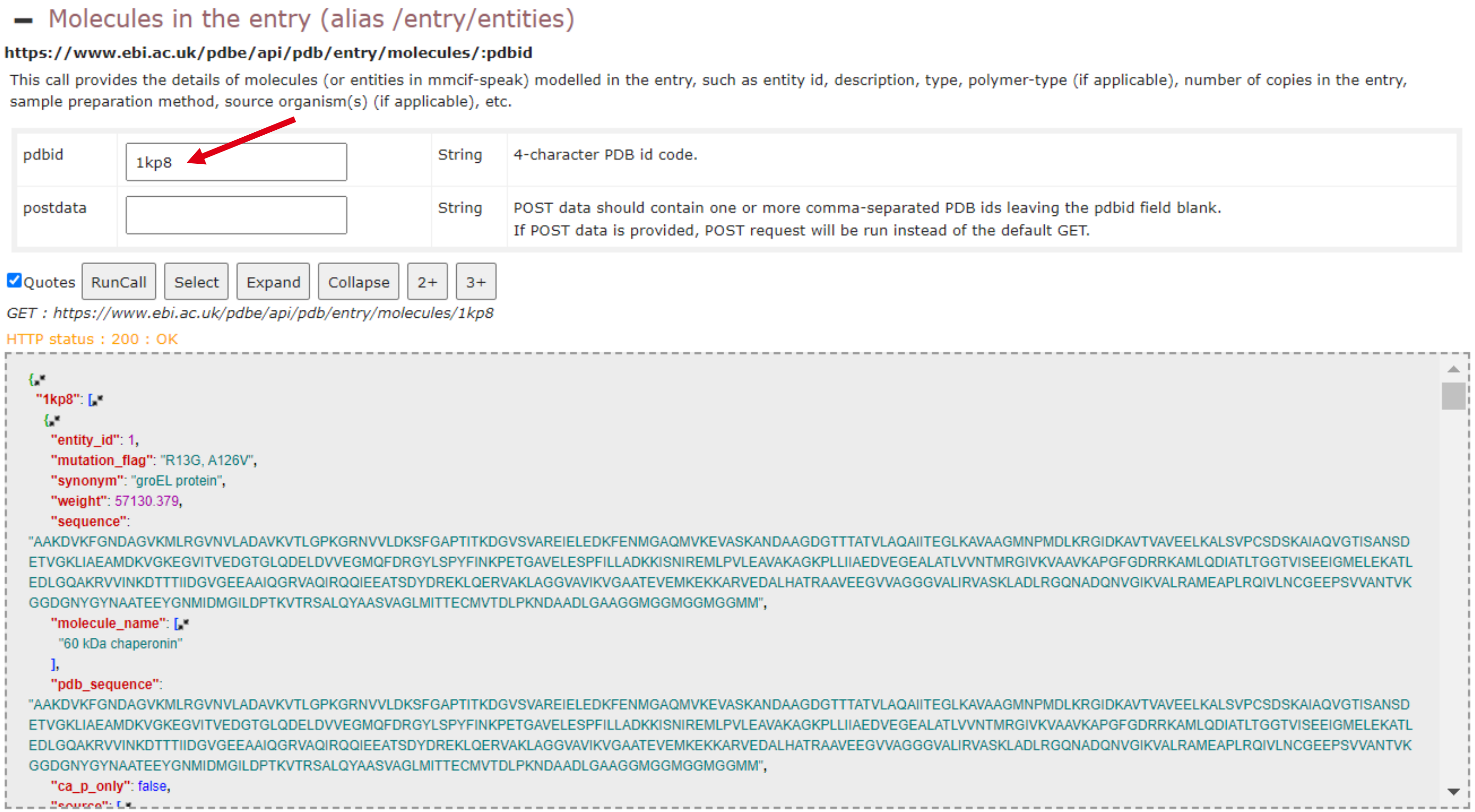
This generates a dataset we can use for futher analyses. These chains bind to REA and contains CATH lipocalin domain.

Case study 2: Backbone and sidechain features of residues across multiple copies of a PDB entry

In validation, an unexpected feature is not always a modelling error. It is useful to compare a strained feature across multiple copies of a protein & generate a quality profile.

In this case study, we will generate a quality profile of multiple copies of the bacterial chaperonin GroEL in PDB entry 1kp8.

/molecules endpoint from **PDB tab**



− Molecules in the entry (alias /entry/entities)

https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/:pdbid

This call provides the details of molecules (or entities in mmcif-speak) modelled in the entry, such as entity id, description, type, polymer-type (if applicable), number of copies in the entry, sample preparation method, source organism(s) (if applicable), etc.

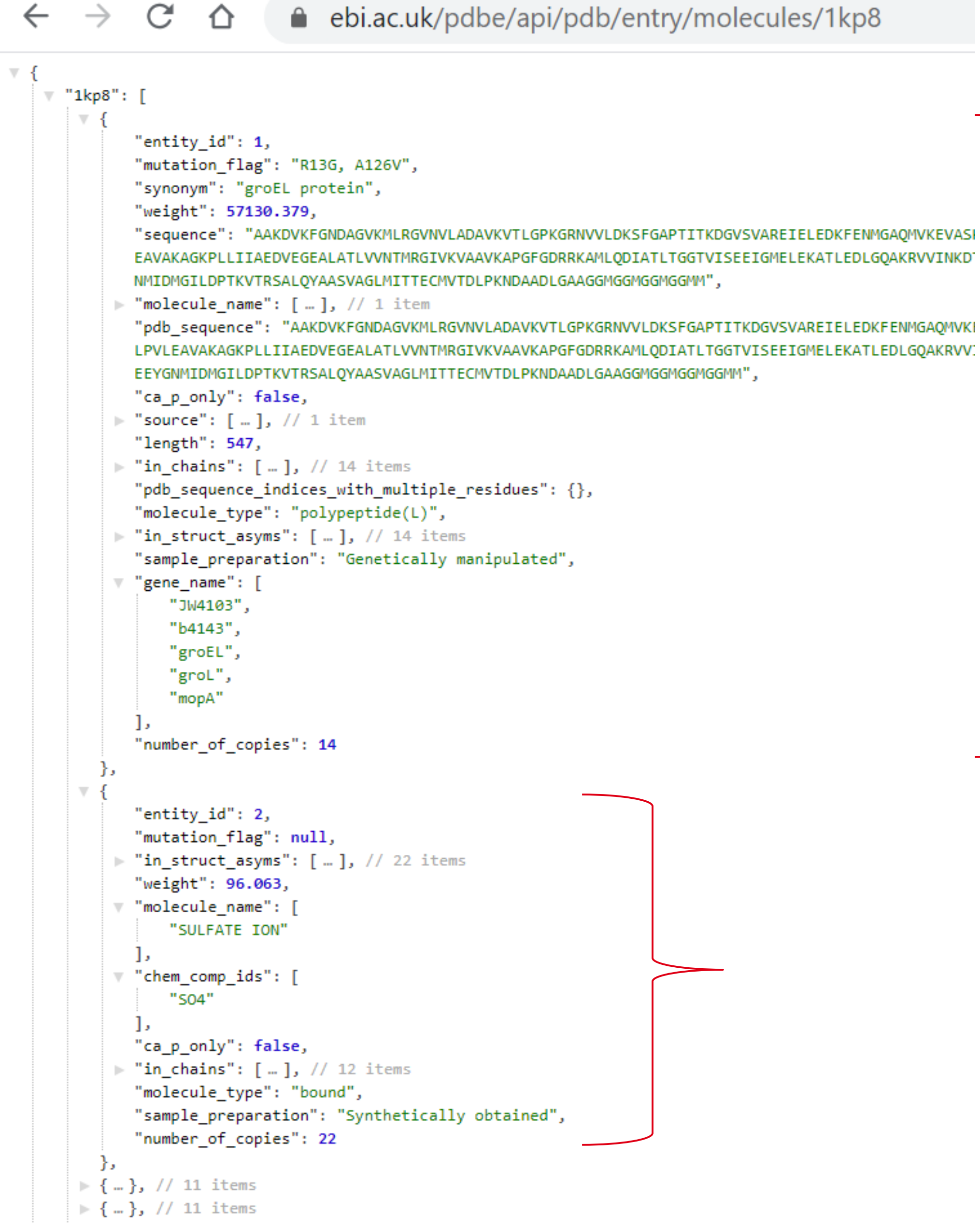
pdbid	1kp8	String	4-character PDB id code.
postdata		String	POST data should contain one or more comma-separated PDB ids leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

☑ Quotes RunCall Select Expand Collapse 2+ 3+

GET : https://www.ebi.ac.uk/pdbe/api/pdb/entry/molecules/1kp8

HTTP status : 200 : OK

```
{
  "1kp8": [
    {
      "entity_id": 1,
      "mutation_flag": "R13G, A126V",
      "synonym": "groEL protein",
      "weight": 57130.379,
      "sequence":
"AAKDVKFGNDAGVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGIDKAVTVAVEELKALSVPCSDSKAIAQVGTISANSD
ETVGKLIAEAMDKVGKEGVITVEDGTGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATL
EDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLADLRGQNADQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVK
GGDGNYGYNAATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGGMGGMGGMM",
      "molecule_name": [
        "60 kDa chaperonin"
      ],
      "pdb_sequence":
"AAKDVKFGNDAGVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGIDKAVTVAVEELKALSVPCSDSKAIAQVGTISANSD
ETVGKLIAEAMDKVGKEGVITVEDGTGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATL
EDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLADLRGQNADQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVK
GGDGNYGYNAATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGGMGGMGGMM",
      "ca_p_only": false,
      "source": [
```



ebi.ac.uk/pdbe/api/pdb/entry/molecules/1kp8

```
{
  "1kp8": [
    {
        "entity_id": 1,
        "mutation_flag": "R13G, A126V",
        "synonym": "groEL protein",
        "weight": 57130.379,
        "sequence": "AAKDVKFGNDAGVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREIELEDKFENMGAQMVKEVAS
        EAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATLEDLGQAKRVVINKD
        NMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGGMGGMGGMM",
      ▶ "molecule_name": [ … ], // 1 item
        "pdb_sequence": "AAKDVKFGNDAGVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREIELEDKFENMGAQMVK
        LPVLEAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATLEDLGQAKRVV
        EEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGGMGGMGGMM",
        "ca_p_only": false,
      ▶ "source": [ … ], // 1 item
        "length": 547,
      ▶ "in_chains": [ … ], // 14 items
        "pdb_sequence_indices_with_multiple_residues": {},
        "molecule_type": "polypeptide(L)",
      ▶ "in_struct_asyms": [ … ], // 14 items
        "sample_preparation": "Genetically manipulated",
      ▼ "gene_name": [
            "JW4103",
            "b4143",
            "groEL",
            "groL",
            "mopA"
        ],
        "number_of_copies": 14
    },
  ▼ {
        "entity_id": 2,
        "mutation_flag": null,
      ▶ "in_struct_asyms": [ … ], // 22 items
        "weight": 96.063,
      ▼ "molecule_name": [
            "SULFATE ION"
        ],
      ▼ "chem_comp_ids": [
            "SO4"
        ],
        "ca_p_only": false,
      ▶ "in_chains": [ … ], // 12 items
        "molecule_type": "bound",
        "sample_preparation": "Synthetically obtained",
        "number_of_copies": 22
    },
  ▶ { … }, // 11 items
  ▶ { … }, // 11 items
```

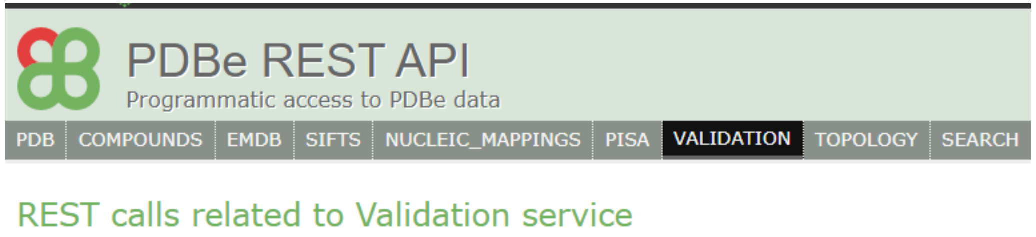
```
For entity_info in json_output[“1kp8”]
  If entity_info[“molecule_type”] = “polypeptide(L)”
    Store entity_info[“entity_id”]
    Store entity_info[“molecule_name”]
    Store entity_info[“in_chains”]
```

```
Molecule no. 1, '60 kDa chaperonin'

modelled in 14 chains A,B,C,D,E,F,G,H,I,J,K,L,M,N.
```

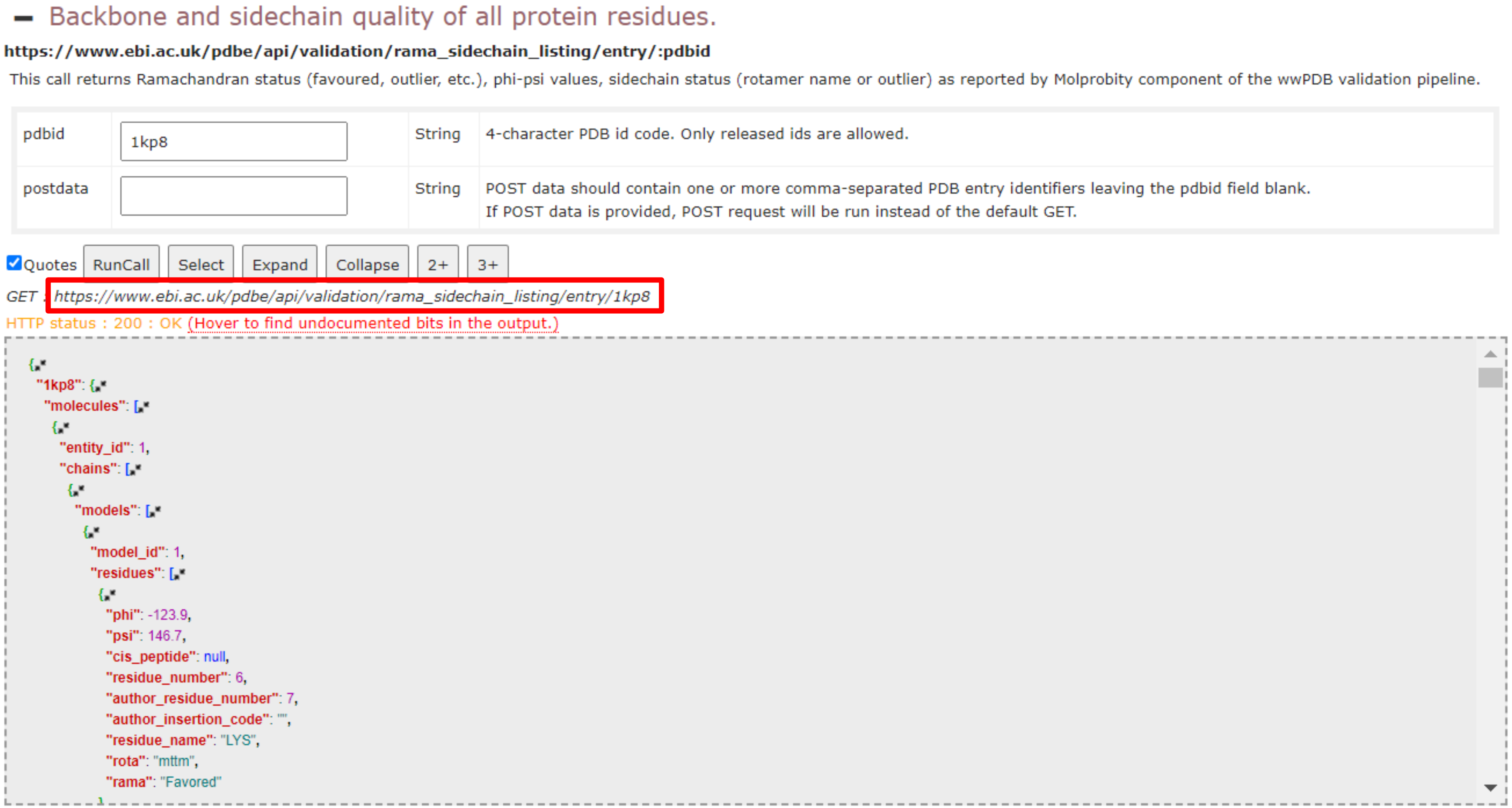
Next, obtain per-residue information on backbone (Ramachandran) and sidechain quality using the call **/validation/rama_sidechain_listing** endpoint

Next, obtain per-residue information on backbone (Ramachandran) and sidechain quality using the call **/validation/rama_sidechain_listing** endpoint



PDBe REST API
Programmatic access to PDBe data

REST calls related to Validation service



– Backbone and sidechain quality of all protein residues.

https://www.ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/:pdbid

This call returns Ramachandran status (favoured, outlier, etc.), phi-psi values, sidechain status (rotamer name or outlier) as reported by Molprobity component of the wwPDB validation pipeline.

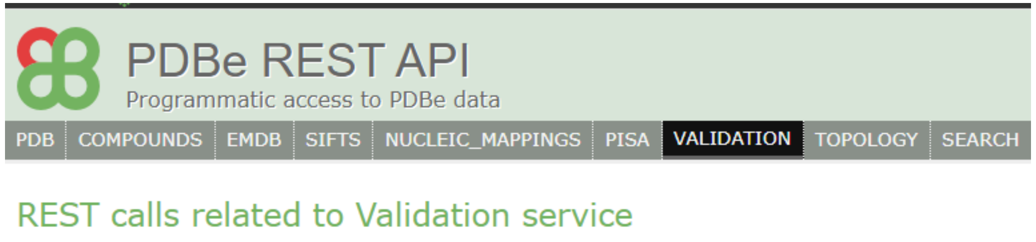
pdbid	1kp8	String	4-character PDB id code. Only released ids are allowed.
postdata		String	POST data should contain one or more comma-separated PDB entry identifiers leaving the pdbid field blank. If POST data is provided, POST request will be run instead of the default GET.

GET : *https://www.ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/1kp8*

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

```
{
 "1kp8": {
  "molecules": [
   {
    "entity_id": 1,
    "chains": [
     {
      "models": [
       {
        "model_id": 1,
        "residues": [
         {
          "phi": -123.9,
          "psi": 146.7,
          "cis_peptide": null,
          "residue_number": 6,
          "author_residue_number": 7,
          "author_insertion_code": "",
          "residue_name": "LYS",
          "rota": "mttm",
          "rama": "Favored"
```

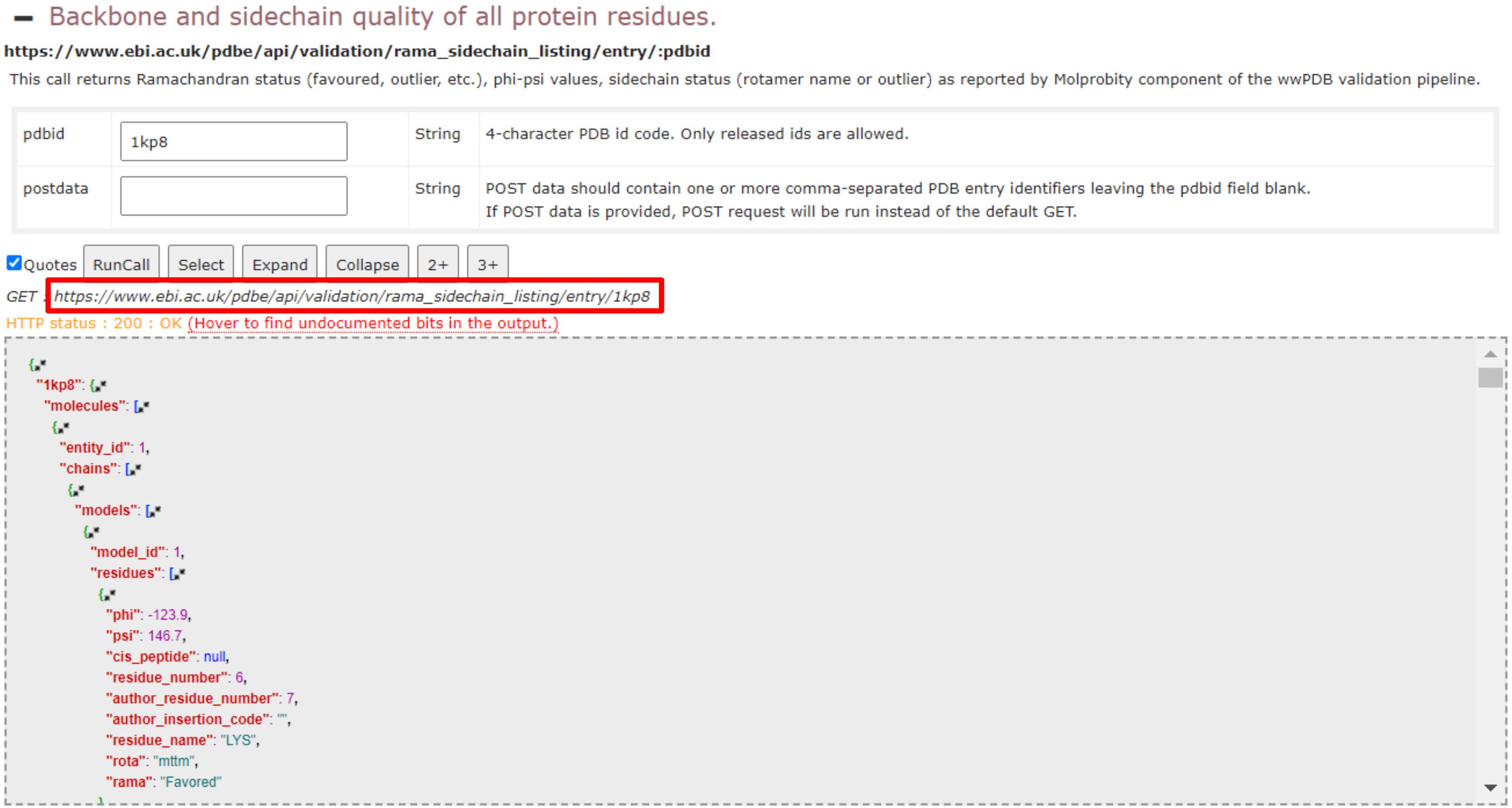
Next, obtain per-residue information on backbone (Ramachandran) and sidechain quality using the call **/validation/rama_sidechain_listing** endpoint



PDBe REST API

Programmatic access to PDBe data

REST calls related to Validation service



– Backbone and sidechain quality of all protein residues.

https://www.ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/:pdbid

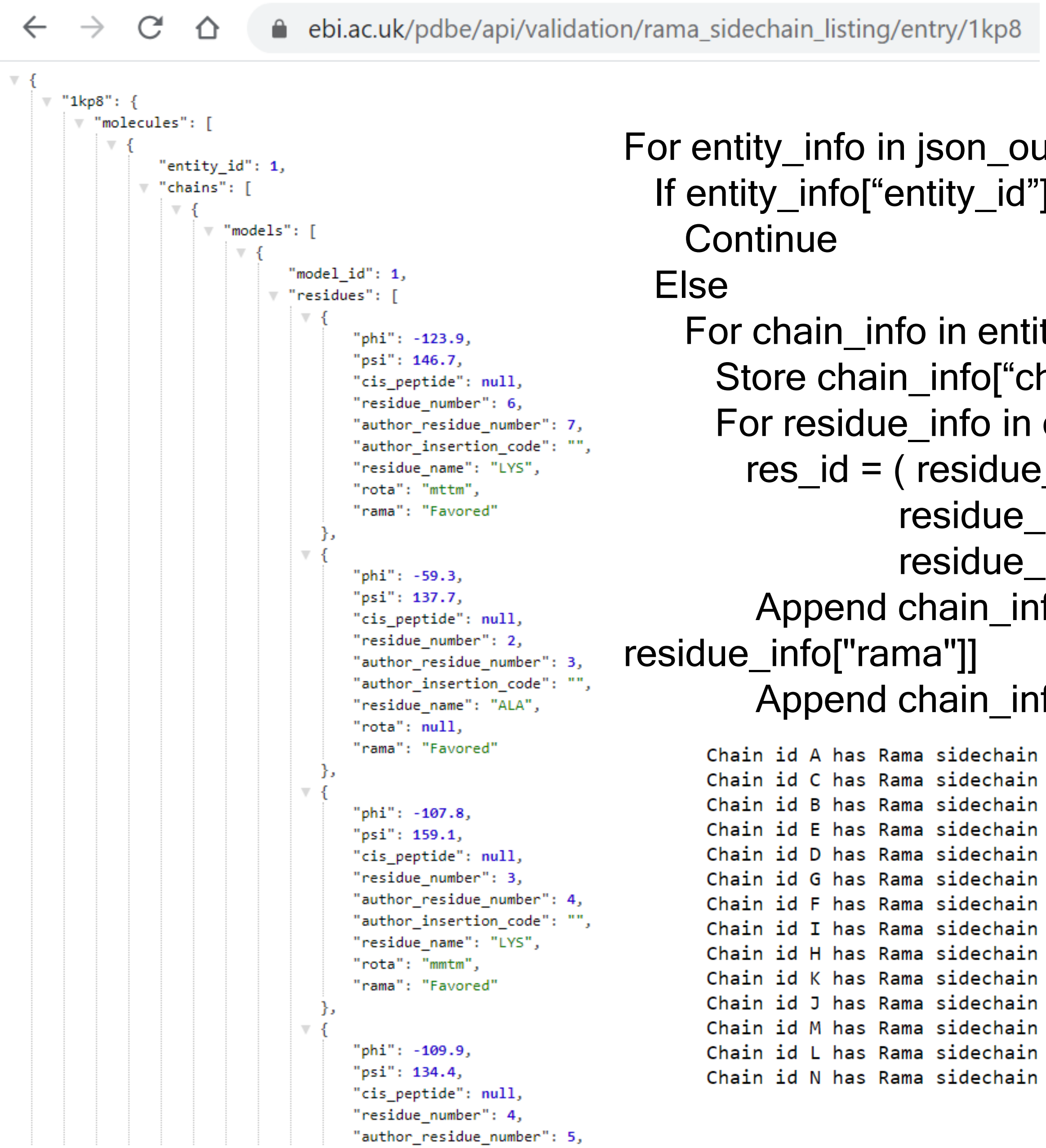
This call returns Ramachandran status (favoured, outlier, etc.), phi-psi values, sidechain status (rotamer name or outlier) as reported by Molprobity component of the wwPDB validation pipeline.

pdbid	1kp8	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.

GET : *https://www.ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/1kp8*

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

```
{
 "1kp8": {
  "molecules": [
   {
    "entity_id": 1,
    "chains": [
     {
      "models": [
       {
        "model_id": 1,
        "residues": [
         {
          "phi": -123.9,
          "psi": 146.7,
          "cis_peptide": null,
          "residue_number": 6,
          "author_residue_number": 7,
          "author_insertion_code": "",
          "residue_name": "LYS",
          "rota": "mttm",
          "rama": "Favored"
```

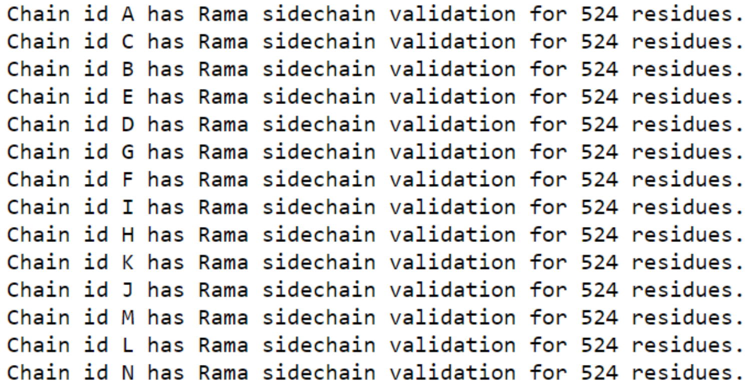


ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/1kp8

```
{
  "1kp8": {
    "molecules": [
      {
        "entity_id": 1,
        "chains": [
          {
            "models": [
              {
                "model_id": 1,
                "residues": [
                  {
                    "phi": -123.9,
                    "psi": 146.7,
                    "cis_peptide": null,
                    "residue_number": 6,
                    "author_residue_number": 7,
                    "author_insertion_code": "",
                    "residue_name": "LYS",
                    "rota": "mttm",
                    "rama": "Favored"
                  },
                  {
                    "phi": -59.3,
                    "psi": 137.7,
                    "cis_peptide": null,
                    "residue_number": 2,
                    "author_residue_number": 3,
                    "author_insertion_code": "",
                    "residue_name": "ALA",
                    "rota": null,
                    "rama": "Favored"
                  },
                  {
                    "phi": -107.8,
                    "psi": 159.1,
                    "cis_peptide": null,
                    "residue_number": 3,
                    "author_residue_number": 4,
                    "author_insertion_code": "",
                    "residue_name": "LYS",
                    "rota": "mmtm",
                    "rama": "Favored"
                  },
                  {
                    "phi": -109.9,
                    "psi": 134.4,
                    "cis_peptide": null,
                    "residue_number": 4,
                    "author_residue_number": 5,
```

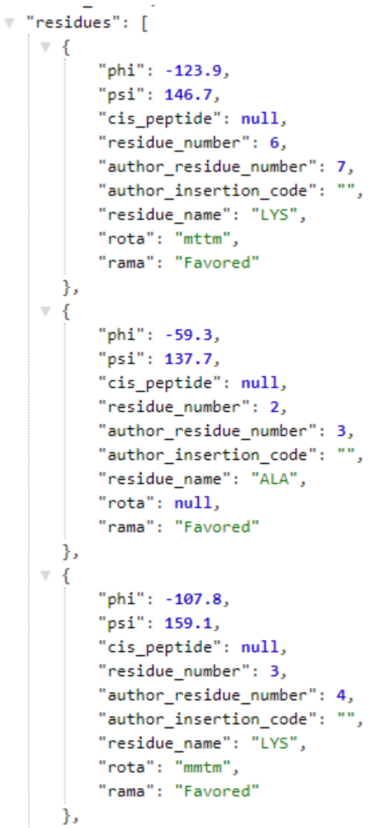
```
For entity_info in json_output[“1kp8”][“molecules”]
  If entity_info[“entity_id”] != 1
    Continue
  Else
    For chain_info in entity_info[“chains”]:
      Store chain_info[“chain_id”]
      For residue_info in chain_info[“models”][0][“residues”]
        res_id = ( residue_info[“residue_name”],
                   residue_info[“author_residue_number”],
                   residue_info[“author_residue_code”])
          Append chain_info[“chain_id”] to Outliers["rama"] [res_id] [ residue_info["rama"]]
          Append chain_info[“chain_id”] to Outliers["rota"] [res_id] [ residue_info["rota"]]
```

Next, store the Rama or sidechain state per residue



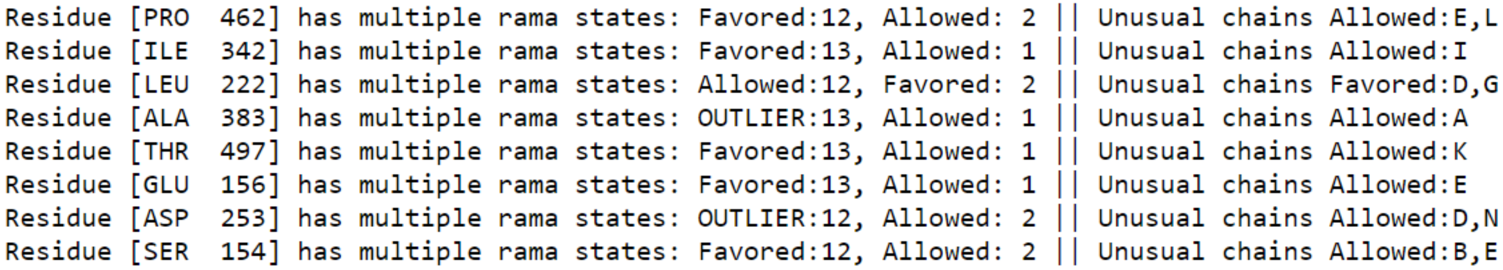
```
Chain id A has Rama sidechain validation for 524 residues.
Chain id C has Rama sidechain validation for 524 residues.
Chain id B has Rama sidechain validation for 524 residues.
Chain id E has Rama sidechain validation for 524 residues.
Chain id D has Rama sidechain validation for 524 residues.
Chain id G has Rama sidechain validation for 524 residues.
Chain id F has Rama sidechain validation for 524 residues.
Chain id I has Rama sidechain validation for 524 residues.
Chain id H has Rama sidechain validation for 524 residues.
Chain id K has Rama sidechain validation for 524 residues.
Chain id J has Rama sidechain validation for 524 residues.
Chain id M has Rama sidechain validation for 524 residues.
Chain id L has Rama sidechain validation for 524 residues.
Chain id N has Rama sidechain validation for 524 residues.
```

Find the residues that have more than one Rama or sidechain state, such that at least one of the states occurs a majority of times



```
▼ "residues": [
  ▼ {
        "phi": -123.9,
        "psi": 146.7,
        "cis_peptide": null,
        "residue_number": 6,
        "author_residue_number": 7,
        "author_insertion_code": "",
        "residue_name": "LYS",
        "rota": "mttm",
        "rama": "Favored"
    },
  ▼ {
        "phi": -59.3,
        "psi": 137.7,
        "cis_peptide": null,
        "residue_number": 2,
        "author_residue_number": 3,
        "author_insertion_code": "",
        "residue_name": "ALA",
        "rota": null,
        "rama": "Favored"
    },
  ▼ {
        "phi": -107.8,
        "psi": 159.1,
        "cis_peptide": null,
        "residue_number": 3,
        "author_residue_number": 4,
        "author_insertion_code": "",
        "residue_name": "LYS",
        "rota": "mmtm",
        "rama": "Favored"
    },
```

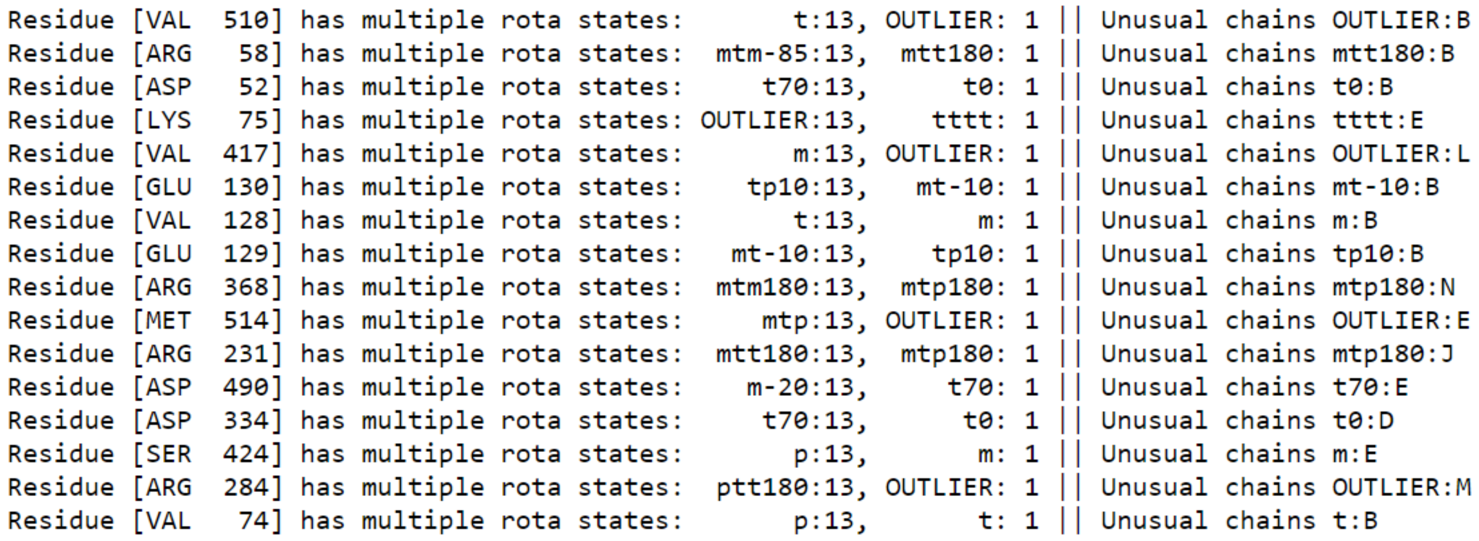
```
For res_id, chain in Outliers[“rama”]
   If length(chain) == 1         // only one state
      Continue
   If length (chain) > 12        // Set 12 chains as min required frequency that a state occurs
      Get the ‘popular’ state and ‘odd’ states
```



```
Residue [PRO  462] has multiple rama states: Favored:12, Allowed: 2 || Unusual chains Allowed:E,L
Residue [ILE  342] has multiple rama states: Favored:13, Allowed: 1 || Unusual chains Allowed:I
Residue [LEU  222] has multiple rama states: Allowed:12, Favored: 2 || Unusual chains Favored:D,G
Residue [ALA  383] has multiple rama states: OUTLIER:13, Allowed: 1 || Unusual chains Allowed:A
Residue [THR  497] has multiple rama states: Favored:13, Allowed: 1 || Unusual chains Allowed:K
Residue [GLU  156] has multiple rama states: Favored:13, Allowed: 1 || Unusual chains Allowed:E
Residue [ASP  253] has multiple rama states: OUTLIER:12, Allowed: 2 || Unusual chains Allowed:D,N
Residue [SER  154] has multiple rama states: Favored:12, Allowed: 2 || Unusual chains Allowed:B,E
```

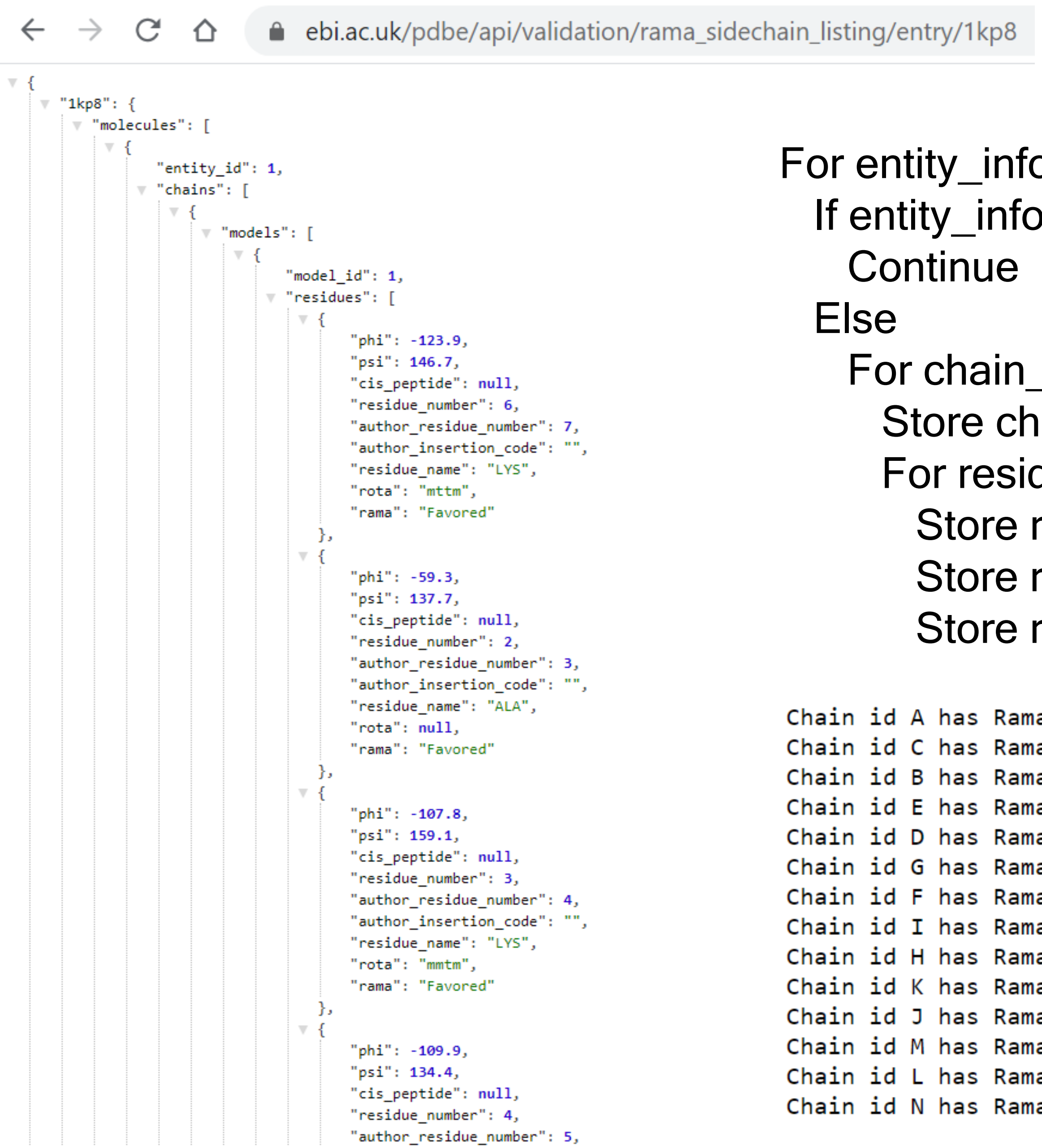
Here we see that Thr K-497, Glu E-156 are in the allowed region whereas residues in other chains are all in the favoured region. These residues need to be reviewed. Residues Ala 383, Asp 253 are nearly always outliers. This could be a genuine outlier with good reason, or it could be a systematic error!

Similarly, let us print residues where only a small minority of chains have a different sidechain rotamer at a residue position.



```
Residue [VAL  510] has multiple rota states:       t:13, OUTLIER: 1 || Unusual chains OUTLIER:B
Residue [ARG   58] has multiple rota states:  mtm-85:13,  mtt180: 1 || Unusual chains mtt180:B
Residue [ASP   52] has multiple rota states:     t70:13,      t0: 1 || Unusual chains t0:B
Residue [LYS   75] has multiple rota states: OUTLIER:13,    tttt: 1 || Unusual chains tttt:E
Residue [VAL  417] has multiple rota states:       m:13, OUTLIER: 1 || Unusual chains OUTLIER:L
Residue [GLU  130] has multiple rota states:    tp10:13,   mt-10: 1 || Unusual chains mt-10:B
Residue [VAL  128] has multiple rota states:       t:13,       m: 1 || Unusual chains m:B
Residue [GLU  129] has multiple rota states:   mt-10:13,    tp10: 1 || Unusual chains tp10:B
Residue [ARG  368] has multiple rota states:  mtm180:13,  mtp180: 1 || Unusual chains mtp180:N
Residue [MET  514] has multiple rota states:     mtp:13, OUTLIER: 1 || Unusual chains OUTLIER:E
Residue [ARG  231] has multiple rota states:  mtt180:13,  mtp180: 1 || Unusual chains mtp180:J
Residue [ASP  490] has multiple rota states:    m-20:13,     t70: 1 || Unusual chains t70:E
Residue [ASP  334] has multiple rota states:     t70:13,      t0: 1 || Unusual chains t0:D
Residue [SER  424] has multiple rota states:       p:13,       m: 1 || Unusual chains m:E
Residue [ARG  284] has multiple rota states:  ptt180:13, OUTLIER: 1 || Unusual chains OUTLIER:M
Residue [VAL   74] has multiple rota states:       p:13,       t: 1 || Unusual chains t:B
```

Note residues Val B-510, Met E-514, Val L-417, Arg M-284 which have rotameric counterparts in all other chains - their sidechain most probably needs remodelling. Also note residues like Lys 75 could be outliers for a good reason, or could just be systematic outliers.



ebi.ac.uk/pdbe/api/validation/rama_sidechain_listing/entry/1kp8

```
{
  "1kp8": {
    "molecules": [
      {
        "entity_id": 1,
        "chains": [
          {
            "models": [
              {
                "model_id": 1,
                "residues": [
                  {
                    "phi": -123.9,
                    "psi": 146.7,
                    "cis_peptide": null,
                    "residue_number": 6,
                    "author_residue_number": 7,
                    "author_insertion_code": "",
                    "residue_name": "LYS",
                    "rota": "mttm",
                    "rama": "Favored"
                  },
                  {
                    "phi": -59.3,
                    "psi": 137.7,
                    "cis_peptide": null,
                    "residue_number": 2,
                    "author_residue_number": 3,
                    "author_insertion_code": "",
                    "residue_name": "ALA",
                    "rota": null,
                    "rama": "Favored"
                  },
                  {
                    "phi": -107.8,
                    "psi": 159.1,
                    "cis_peptide": null,
                    "residue_number": 3,
                    "author_residue_number": 4,
                    "author_insertion_code": "",
                    "residue_name": "LYS",
                    "rota": "mmtm",
                    "rama": "Favored"
                  },
                  {
                    "phi": -109.9,
                    "psi": 134.4,
                    "cis_peptide": null,
                    "residue_number": 4,
                    "author_residue_number": 5,
```

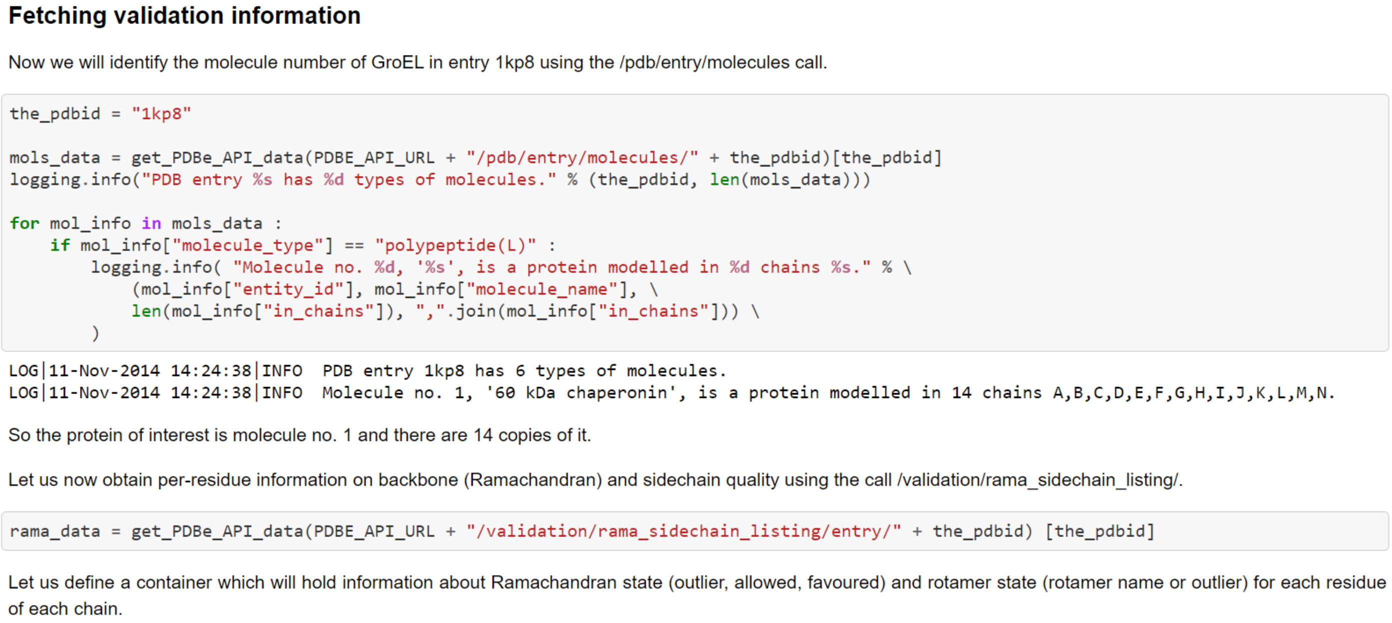
```
For entity_info in json_output["1kp8"]["molecules"]
  If entity_info["entity_id"] != 1
    Continue
  Else
    For chain_info in entity_info["chains"]:
      Store chain_info["chain_id"]
      For residue_info in chain_info["models"][0]["residues"]
        Store residue_info["residue_name"]
        Store residue_info["author_residue_number"]
        Store residue_info["author_residue_code"]
```

```
Chain id A has Rama sidechain validation for 524 residues.
Chain id C has Rama sidechain validation for 524 residues.
Chain id B has Rama sidechain validation for 524 residues.
Chain id E has Rama sidechain validation for 524 residues.
Chain id D has Rama sidechain validation for 524 residues.
Chain id G has Rama sidechain validation for 524 residues.
Chain id F has Rama sidechain validation for 524 residues.
Chain id I has Rama sidechain validation for 524 residues.
Chain id H has Rama sidechain validation for 524 residues.
Chain id K has Rama sidechain validation for 524 residues.
Chain id J has Rama sidechain validation for 524 residues.
Chain id M has Rama sidechain validation for 524 residues.
Chain id L has Rama sidechain validation for 524 residues.
Chain id N has Rama sidechain validation for 524 residues.
```

Case study 2: Backbone and sidechain features of residues across multiple copies of a PDB entry

In validation, an unexpected feature is not always a modelling error. It is useful to compare a strained feature across multiple copies of a protein & generate a quality profile.

In this case study, we will generate a quality profile of multiple copies of the bacterial chaperonin GroEL in PDB entry 1kp8.



Fetching validation information

Now we will identify the molecule number of GroEL in entry 1kp8 using the /pdb/entry/molecules call.

```
the_pdbid = "1kp8"

mols_data = get_PDBe_API_data(PDBE_API_URL + "/pdb/entry/molecules/" + the_pdbid)[the_pdbid]
logging.info("PDB entry %s has %d types of molecules." % (the_pdbid, len(mols_data)))

for mol_info in mols_data :
    if mol_info["molecule_type"] == "polypeptide(L)" :
        logging.info( "Molecule no. %d, '%s', is a protein modelled in %d chains %s." % \
            (mol_info["entity_id"], mol_info["molecule_name"], \
            len(mol_info["in_chains"]), ",".join(mol_info["in_chains"])) \
        )
```

```
LOG|11-Nov-2014 14:24:38|INFO  PDB entry 1kp8 has 6 types of molecules.
LOG|11-Nov-2014 14:24:38|INFO  Molecule no. 1, '60 kDa chaperonin', is a protein modelled in 14 chains A,B,C,D,E,F,G,H,I,J,K,L,M,N.
```

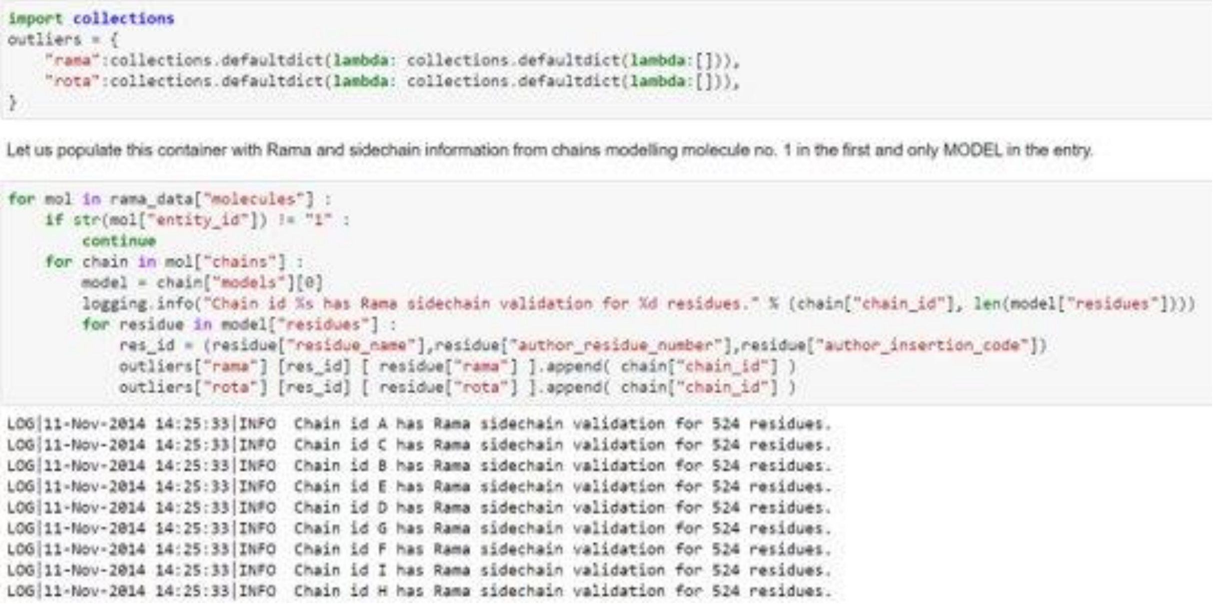
So the protein of interest is molecule no. 1 and there are 14 copies of it.

Let us now obtain per-residue information on backbone (Ramachandran) and sidechain quality using the call /validation/rama_sidechain_listing/.

```
rama_data = get_PDBe_API_data(PDBE_API_URL + "/validation/rama_sidechain_listing/entry/" + the_pdbid) [the_pdbid]
```

Let us define a container which will hold information about Ramachandran state (outlier, allowed, favoured) and rotamer state (rotamer name or outlier) for each residue of each chain.

Case study 2: Backbone and sidechain features of residues across multiple copies of a PDB entry

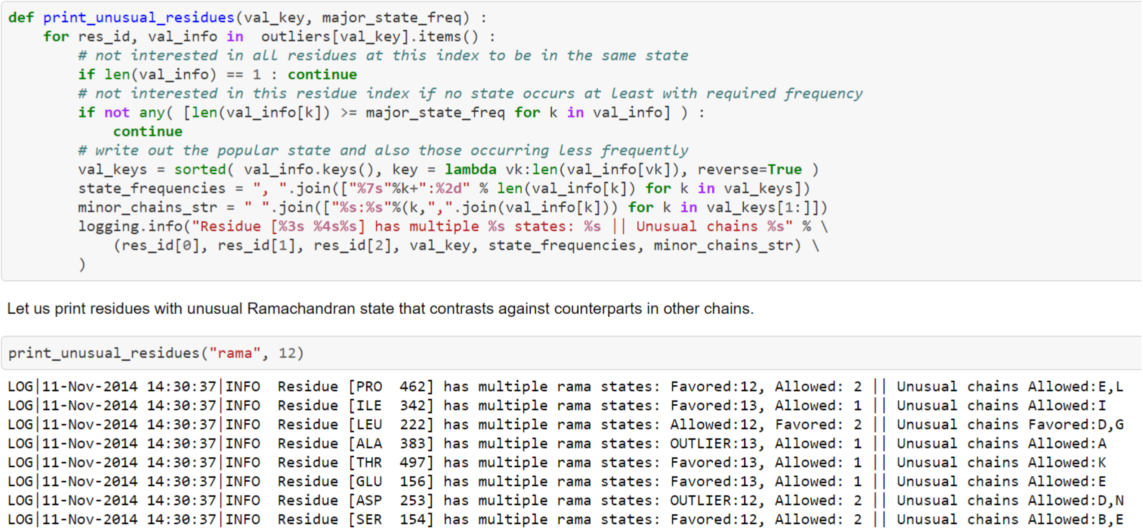


```
import collections
outliers = {
    "rama":collections.defaultdict(lambda: collections.defaultdict(lambda:[])),
    "rota":collections.defaultdict(lambda: collections.defaultdict(lambda:[])),
}
```

Let us populate this container with Rama and sidechain information from chains modelling molecule no. 1 in the first and only MODEL in the entry.

```
for mol in rama_data["molecules"] :
    if str(mol["entity_id"]) != "1" :
        continue
    for chain in mol["chains"] :
        model = chain["models"][0]
        logging.info("Chain id %s has Rama sidechain validation for %d residues." % (chain["chain_id"], len(model["residues"])))
        for residue in model["residues"] :
            res_id = (residue["residue_name"],residue["author_residue_number"],residue["author_insertion_code"])
            outliers["rama"] [res_id] [ residue["rama"] ].append( chain["chain_id"] )
            outliers["rota"] [res_id] [ residue["rota"] ].append( chain["chain_id"] )
```

```
LOG|11-Nov-2014 14:25:33|INFO  Chain id A has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id C has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id B has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id E has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id D has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id G has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id F has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id I has Rama sidechain validation for 524 residues.
LOG|11-Nov-2014 14:25:33|INFO  Chain id H has Rama sidechain validation for 524 residues.
```



```
def print_unusual_residues(val_key, major_state_freq) :
    for res_id, val_info in  outliers[val_key].items() :
        # not interested in all residues at this index to be in the same state
        if len(val_info) == 1 : continue
        # not interested in this residue index if no state occurs at least with required frequency
        if not any( [len(val_info[k]) >= major_state_freq for k in val_info] ) :
            continue
        # write out the popular state and also those occurring less frequently
        val_keys = sorted( val_info.keys(), key = lambda vk:len(val_info[vk]), reverse=True )
        state_frequencies = ", ".join(["%7s"%k+":%2d" % len(val_info[k]) for k in val_keys])
        minor_chains_str = " ".join(["%s:%s"%(k,",".join(val_info[k])) for k in val_keys[1:]])
        logging.info("Residue [%3s %4s%s] has multiple %s states: %s || Unusual chains %s" % \
            (res_id[0], res_id[1], res_id[2], val_key, state_frequencies, minor_chains_str) \
        )
```

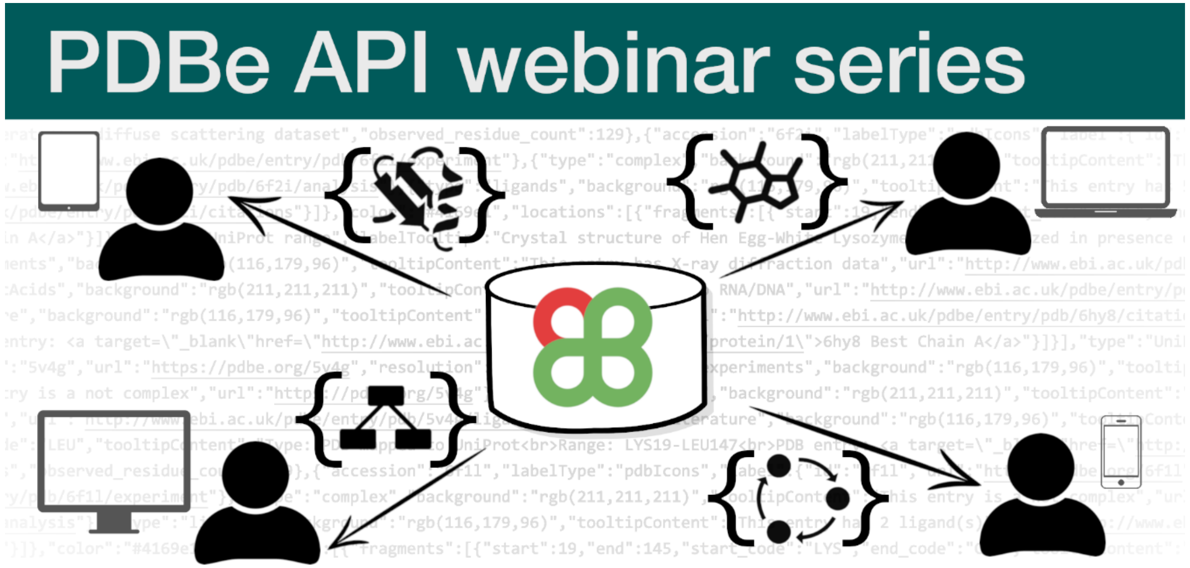
Let us print residues with unusual Ramachandran state that contrasts against counterparts in other chains.

```
print_unusual_residues("rama", 12)
```

```
LOG|11-Nov-2014 14:30:37|INFO  Residue [PRO  462] has multiple rama states: Favored:12, Allowed: 2 || Unusual chains Allowed:E,L
LOG|11-Nov-2014 14:30:37|INFO  Residue [ILE  342] has multiple rama states: Favored:13, Allowed: 1 || Unusual chains Allowed:I
LOG|11-Nov-2014 14:30:37|INFO  Residue [LEU  222] has multiple rama states: Allowed:12, Favored: 2 || Unusual chains Favored:D,G
LOG|11-Nov-2014 14:30:37|INFO  Residue [ALA  383] has multiple rama states: OUTLIER:13, Allowed: 1 || Unusual chains Allowed:A
LOG|11-Nov-2014 14:30:37|INFO  Residue [THR  497] has multiple rama states: Favored:13, Allowed: 1 || Unusual chains Allowed:K
LOG|11-Nov-2014 14:30:37|INFO  Residue [GLU  156] has multiple rama states: Favored:13, Allowed: 1 || Unusual chains Allowed:E
LOG|11-Nov-2014 14:30:37|INFO  Residue [ASP  253] has multiple rama states: OUTLIER:12, Allowed: 2 || Unusual chains Allowed:D,N
LOG|11-Nov-2014 14:30:37|INFO  Residue [SER  154] has multiple rama states: Favored:12, Allowed: 2 || Unusual chains Allowed:B,E
```

To register for future webinars in the series

- Webinar videos, slides and interactive jupyter notebooks. To view the materials and sign up for the next in the series, visit https://pdbeurope.github.io/api-webinars/
- Or visit the PDBe events pages at PDBe.org/events
- Remember to register for each webinar individually!!!



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

Thank you for your attention!
Any questions?

PDBe.org/API

Nurul Nadzirin

✉ pdbhelp@ebi.ac.uk

f proteindatabank

@PDBeurope

▶ proteindatabank

pdbeurope

pdbart

PDBe
Protein Data Bank in Europe

EMBL-EBI