DATE: 13/10/2022

WEBLEM: 3

Introduction to SAbDab (Antibody Structure Database) and ABCD (Antibody Sequence Database) Database

Antibodies form the foundations of the vertebrate immune response. These proteins form complexes with potentially pathogenic molecules called antigens and inhibit their function or recruit other components of the immunological machinery to destroy them. In addition to the biological importance of antibodies, their ability to be raised against an almost limitless number of molecules has made them useful laboratory tools and increasingly useful as therapeutic agents in humans. This biopharmaceutical application has motivated the desire to understand how binding, stability and immunogenic properties of the antibody are determined and how they can be modified. Computational analyses and tools are increasingly being employed to aid the antibody engineering process. Many of these tools now use only the antibody data, as opposed to general protein data, because this has been shown to increase performance.

SAbDab (Antibody Structure Database):

Antibodies are the fundamental components of the immune system and represent the largest class of biotherapeutics. Due to the importance of an accurate understanding of the three-dimensional structure of antibodies for the study of their properties and the development of antibody therapeutics, Structural Antibody Database (SAbDab) in 2013, a comprehensive and continuously updated database of experimentally determined antibody structures was released.

Structural Antibody Database (SAbDab), a database devoted to automatically collecting, curating and presenting antibody structural data in a consistent manner for both bulk analysis and individual inspection. SAbDab updates on a weekly basis and provides users with a range of methods to select sets of structures. For example, users can select by species, experimental details (e.g. method, resolution and r-factor), similarity to a given antibody sequence, amino-acid composition at certain positions and antibody–antigen affinity. Entries can also be selected using structural annotations including, for example, the canonical form of the complementarity determining regions (CDR), orientation between the antibody variable domains and the presence of constant domains in the structure. Structures can be inspected individually or downloaded en masse either as the original file from the PDB or as a structure that has been annotated using the Chothia numbering scheme. In all cases, a tab-separated file detailing heavy and light chain pairing, antibody–antigen pairing and all other annotations is generated.

- **Antibody structures:**

Each week, the PDB releases new experimental structures. Using key word searches, it is possible to identify most of those that contain an antibody chain. However, no direct or consistent information is given about chain type, heavy–light chain pairings or antibody–antigen chain pairings. Therefore, SAbDab attempts to apply the Chothia antibody numbering to the sequence of each new chain using ABnum. This automatically detects each chain's type—heavy, light or non-antibody. The process is applied recursively to sequences to identify each variable region of the chain and thus enable the identification of single-chain Fvs (scFvs) that have not been split into separate chains. Those non-antibody chains that belong to a PDB entry containing an unequal number of heavy and light chains are aligned to antibody sequence profiles using MUSCLE. A chain must have a sequence identity of <35% to any antibody sequence profile for it to be considered a potential antigen. Those

that exceed this threshold are flagged for manual inspection. In addition, any structure whose header details contain words similar to 'T-cell' or 'MHC' are flagged for manual inspection before their inclusion in SAbDab.

- **Affinity data:**

SAbDab contains 190 structures with an associated affinity value. In total, 133 are bound to proteins, 38 to peptides and 19 to hapten antigens. This curated data set should serve as a useful benchmarking resource for the antibody–antigen docking prediction community and the antibody engineering community.

- **Complementarity determining regions:**

In SAbDab, the Kabat , Contact and Chothia, CDRs are annotated. The length and sequence of the CDRs, according to these three definitions, is extracted for each structure and recorded in SAbDab. In the database, the Chothia CDRs (16) are further analyzed to assign membership into structural clusters, often referred to as canonical conformations.

- **Accessing the data:**

The data in SAbDab can be accessed and filtered in a number of ways. Details of particular structures can be retrieved and viewed or sets of entries can be selected and downloaded. In addition, the entire structural contents of SAbDab can be downloaded. Downloads For each structure, the following files may be downloaded:

1. The pdb structure file.
2. A Chothia re-numbered structure file.
3. A tab-separated summary file containing information about chain pairings, antigen pairing and other annotations about the structure gathered by SAbDab.

The structure files are available in PDB format. The Chothia re-numbered file contains the coordinates of each atom in the structure. Each antibody residue is renumbered with the Chothia numbering scheme over the variable region of domains. Non-variable region residues are numbered sequentially. Non-antibody chains retain their original residue numbering. The header of each file contains information about the chain types, pairings and antigen pairings. Non-antibody chains retain their original residue numbering. The header of each file contains information about the chain types, pairings and antigen pairings. For instance, the structure 1ahw has two heavy–light chain pairs: B–A and E–D. These associate with protein antigen chains C and F, respectively. Thus, the header contains the lines:

```
REMARK   5 PAIRED_HL HCHAIN=B LCHAIN=A AGCHAIN=C AGTYPE=PROTEIN
REMARK   5 PAIRED_HL HCHAIN=E LCHAIN=D AGCHAIN=F AGTYPE=PROTEIN
```

The summary file is a tab separated, .tsv file containing information about chain pairings and details about the structure, for example, experimental details, antigen affinity and species. The first line is the name of each field. Each following line corresponds to a paired heavy and 21 light antibody chain and details corresponding to that pairing. For instance, the first six fields of the summary file for 1ahw appear as:

```
pdb   Hchain Lchain model antigen_chain antigen_type  ...
1ahw  B      A      0     C             protein       ...
1ahw  E      D      0     F             protein       ...
```

When a user selects any set of structures, they are able to download the files for each structure individually or collectively as a dataset using the 'download all' function. In the latter case, a single zip file is created containing an archive of all the selected structures. A single summary file is also created for all the heavy- and light-chain pairings in the selection. This file may also be downloaded separately without the structural data.

- **CDR search tools:**

SAbDab offers a CDR-specific search functionality. A user may select CDRs using similar criteria as in the advanced search tool ('advanced search' section). In addition, CDR structures can be searched with respect to their CDR type and length in accordance with different CDR definitions and their membership of structural clusters or canonical classes ('complementarity determining regions' section). SAbDab will return a list of the selected CDR structures. These can be inspected individually or downloaded as described in the 'downloads' section. The CDR search tool also allows a nonredundant set of CDR structures to be selected. In this case, only non-identical structures with respect to type, length and sequence are returned. For identical sequences, the structure with the best resolution is returned.

SAbDab continues to be updated weekly and represents the most thoroughly annotated antibody structure database from which researchers can quickly create custom datasets for their studies. Searching SAbDab is now more powerful and faster, with new connections to auxiliary databases that catalogue therapeutic and antigen-specific antibodies. These links will continue to be extended as more such databases become available.

Fig1: Homepage for SAbDab database

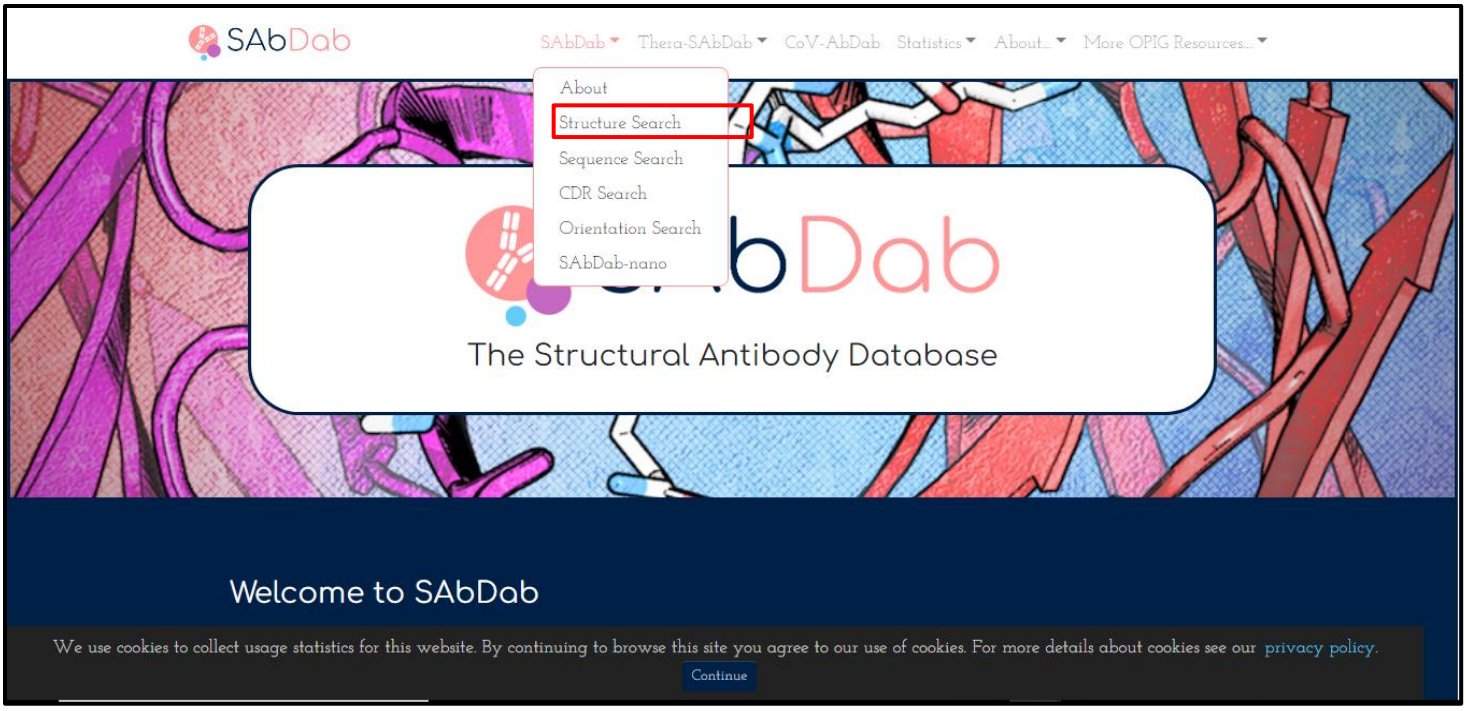

Fig2: Search options under SAbDab database

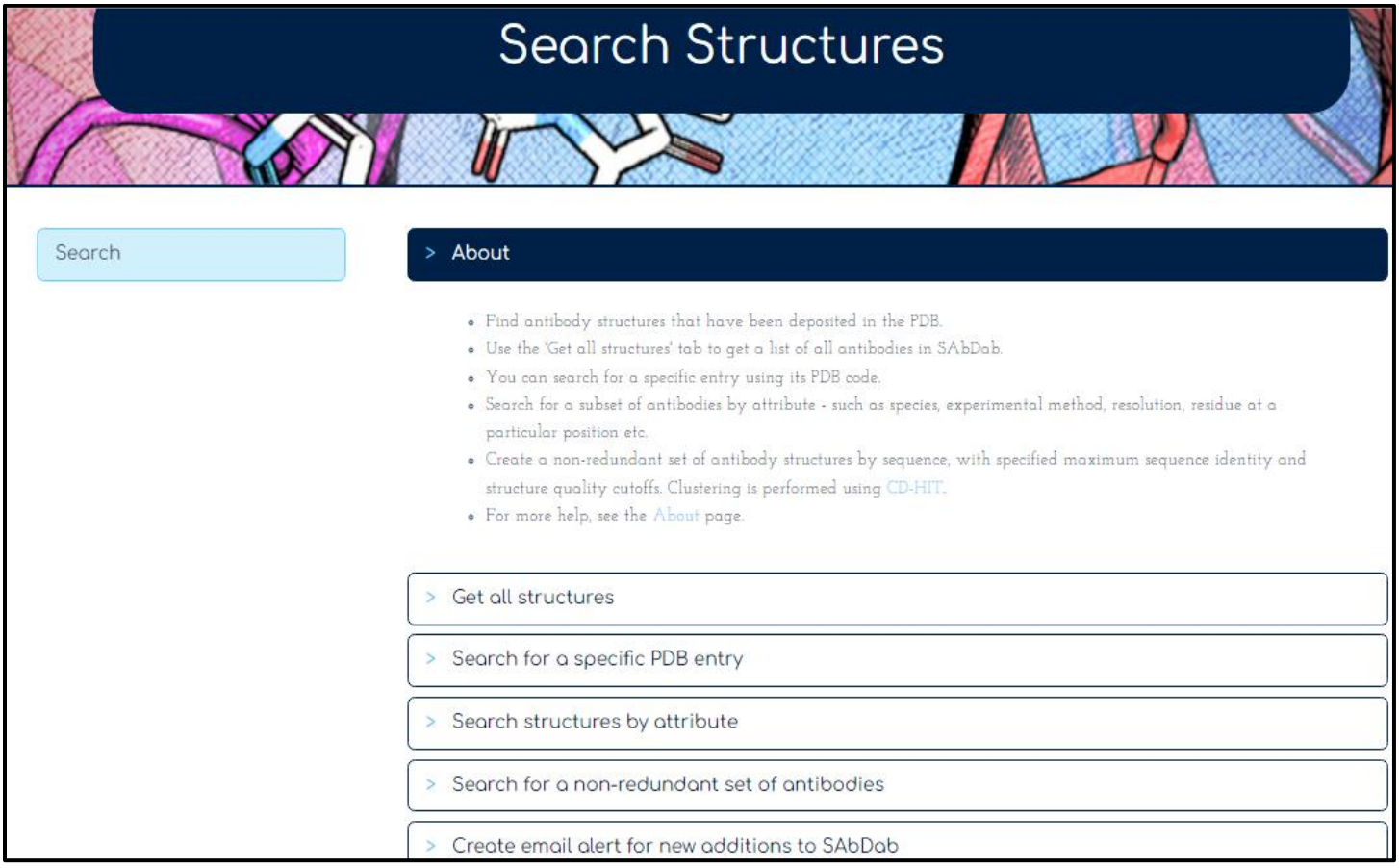

Fig3: Different search options available under Search Structures section

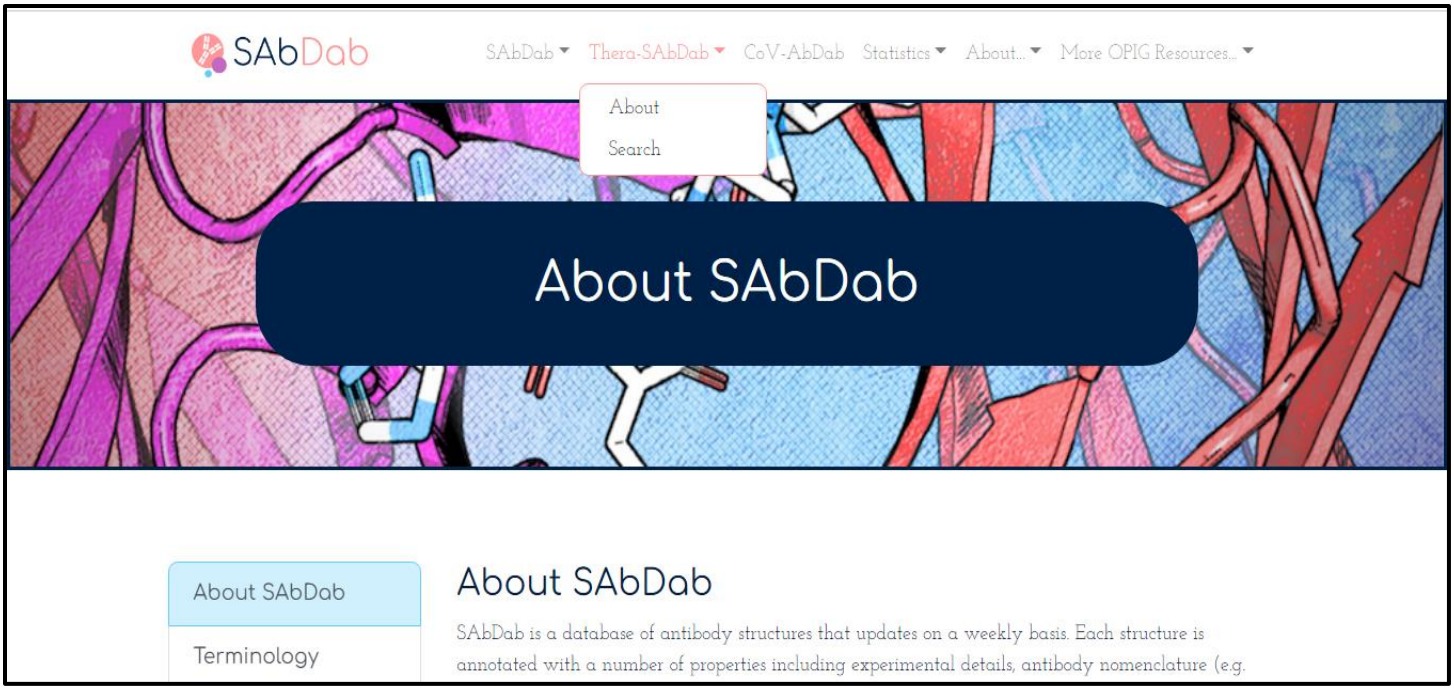

Fig4: Search options under Thera-SAbDab database

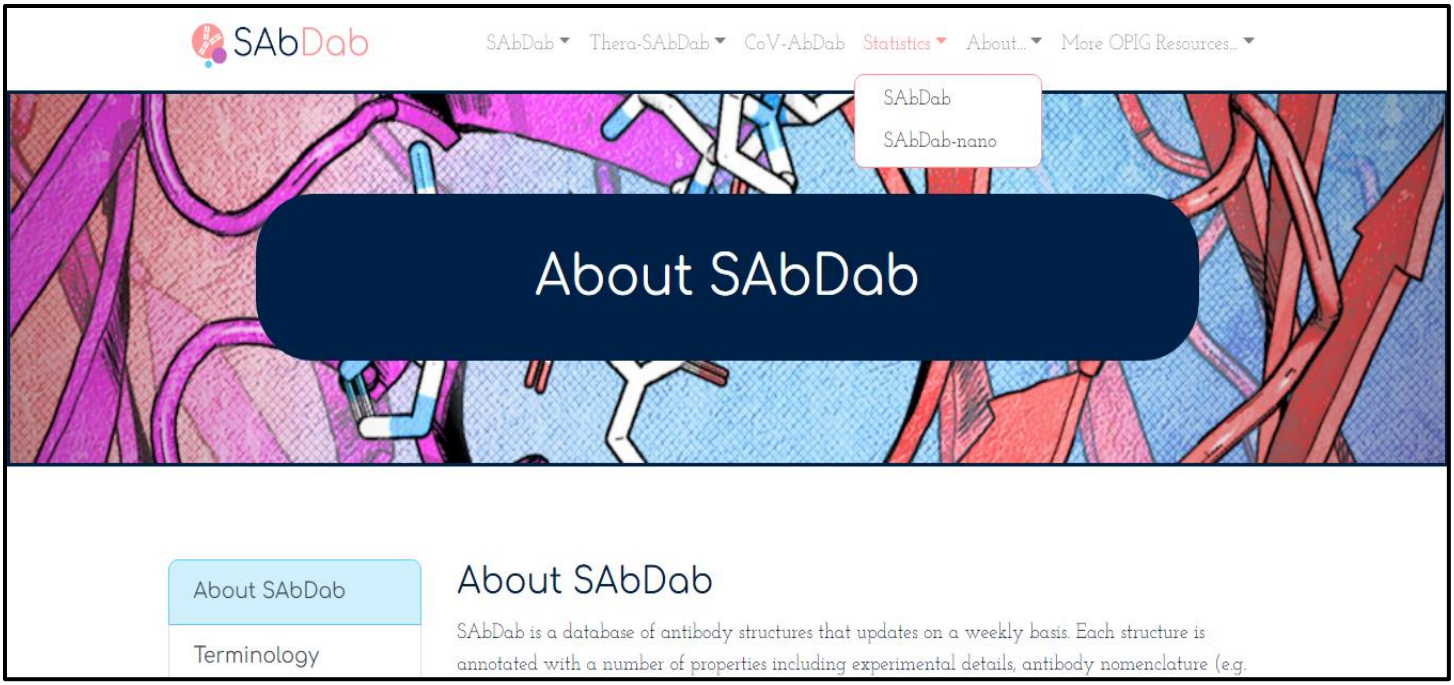

Fig5: Search options under Statistics section

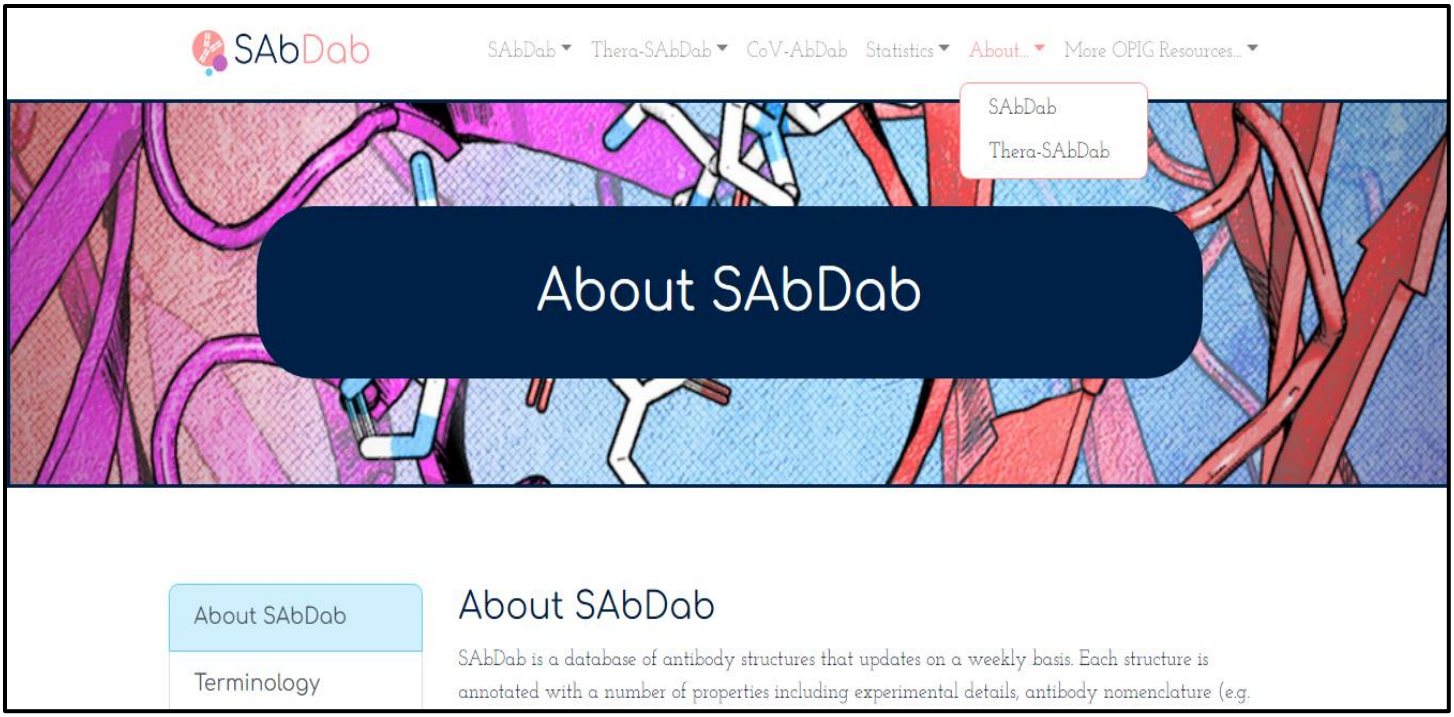

Fig6: Search options under About section

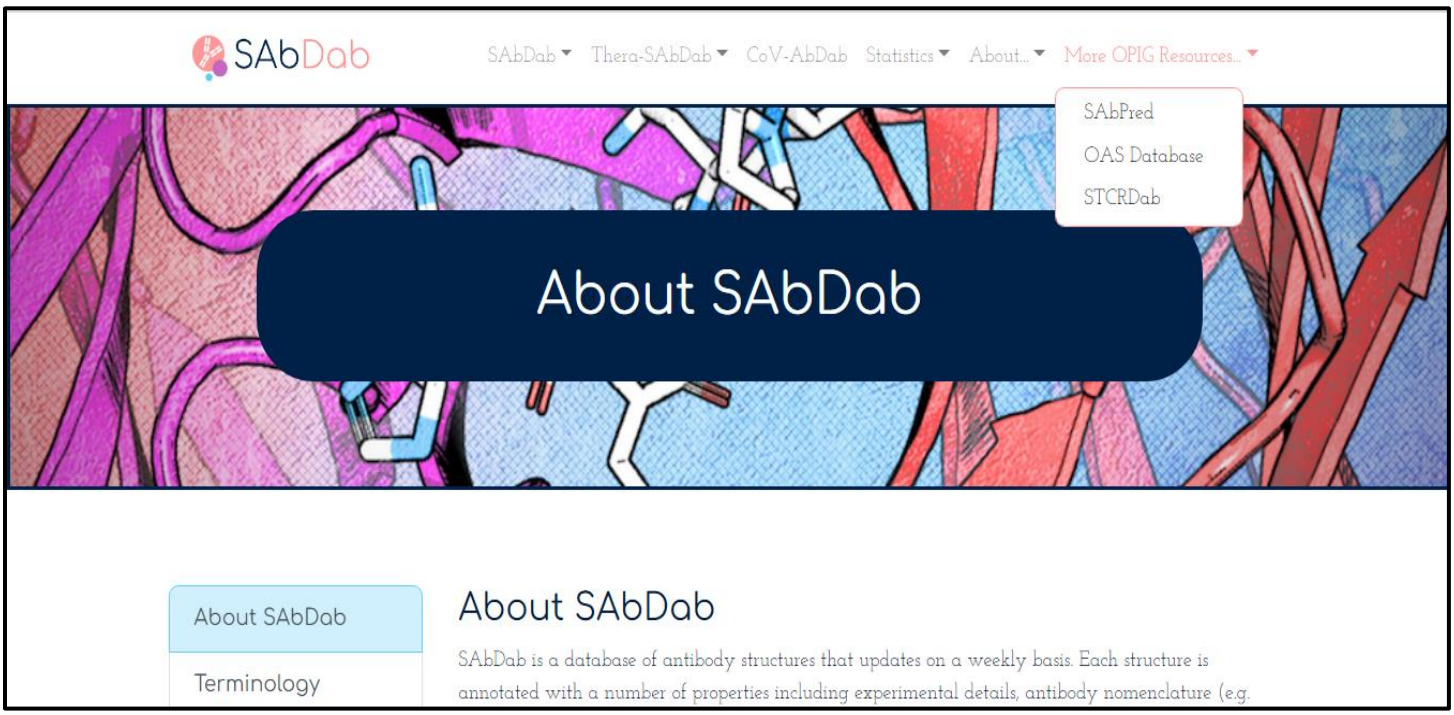

Fig7: Search options under OPIG Resources

ABCD (Antibody Sequence Database) Database:

The ABCD database is, to our knowledge, the first effort to provide freely accessible, curated information on chemically defined antibodies (i.e. antibodies with a known primary amino-acid sequence) connected with their antigenic target, which can be either a protein (linked to an UniProtKB unique identifier (UID) or a chemical entity (linked to a ChEBI UID).

Each ABCD entry corresponds to a unique primary amino-acid sequence, defined by a unique ABCD identifier. For each entry, information about the antigen and about the antibody are provided.

Regarding the antibody, in addition to its ABCD identifier, the following information is given:

i. Recommended name (most frequently, the name provided in the referenced publication) and a list of synonyms.
ii. Technical applications for which the antibody has been used (by no means an exhaustive inventory, as it lists only the applications described on the referenced publications).
iii. At least one bibliographic reference (either a published scientific article—with a PubMed UID or a Digital Object Identifier (DOI)—or a patent, with a link to the WIPO database) in which the antibody sequence is provided. Note that this is not meant to be a comprehensive list of all the publications describing a given antibody.
iv. Cross-references to other databases.
v. Regarding the antigen, the following is given:
vi. Type of target (if a protein or a chemical).
vii. Name of the antigen (and, in the case of a protein, also the species against which the antibody was produced).
viii. Link to UniProtKB (for a protein) or ChEBI (for a chemical) databases.
ix. When available, information about the epitope recognized (for example, a domain or a specific amino-acid subsequence).

The antibody amino-acid sequence can be obtained in the links to the publications and the databases used as source. Alternatively, the information is also available upon request by email. The stored information corresponds to the sequence of the variable region of both the heavy and light chains (or, in the case of camelid antibodies or nanobodies, the sequence of the unique variable chain). When needed, definition of heavy and light chain boundaries, based on alignment with germline sequences, was done using the VBASE2 server.

The ABCD database is populated with data coming from:

i. Sequences published in scientific articles or patents.
ii. 3D structural data.
iii. A few publications and repositories of large-scale phage display or hybridoma sequencing projects.
We only include sequenced antibodies with a known and defined target. However, the source of such information is of variable quality, and we encourage users to verify the reactivity of each antibody that they use.

Database design and implementation:

The ABCD database is developed by the Geneva Antibody Facility team (https://www.unige.ch/medecine/antibodies/), in collaboration with the CALIPHO and Swiss-Prot groups at the Swiss Institute of Bioinformatics. The database is available at the ExPASy web server.The ABCD database website consists of a simple, userfriendly interface. Each antibody page is dynamically linked to external

resources and databases. Entries can be searched by antibody name, antigen name, antigen species, UniProtKB or ChEBI UIDs, epitope information and reference UID (PubMed, DOI or Patent), via a full-text search field.

The ABCD database aims at helping to improve reproducibility in academic research by providing a unique, unambiguous identifier associated to each antibody sequence. It also allows determining rapidly if a sequenced antibody is available for a given antigen.

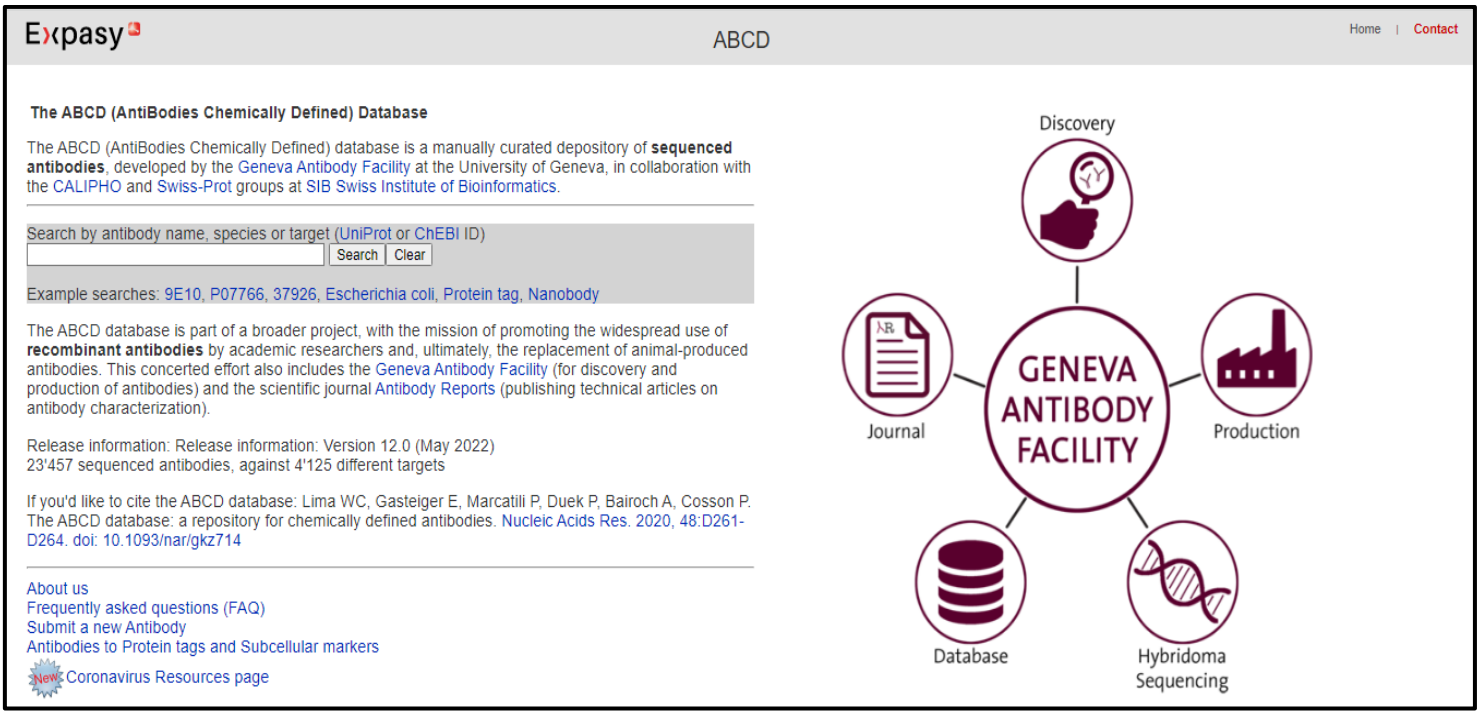

Fig: Homepage for ABCD database

REFERENCES:

1. Dunbar, J., Krawczyk, K., leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., Deane, C.M. (2014). SAbDab: the structural antibody database. *Nucleic Acids Research*, *42*, D1140-D1146. doi:10.1093/nar/gkt1043
2. Schneider, C., Raybould, M.I.J., Deane, C.M. (2022). SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker. *Nucleic Acids Research*, *50*, D1368-D1372. doi: https://doi.org/10.1093/nar/gkab1050
3. Lima, W.C., Gasteiger, E., Marcatili, P., Duek, P., Bairoch, A., Cosson, P. (2019). The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Research*, *48*, D261-D264. doi: 10.1093/nar/gkz714

DATE: 13/10/2022

WEBLEM: 3A

Introduction to Immunoglobulins and its structural features using SAbDab Database

(URL: http://opig.stats.ox.ac.uk/webapps/newsabdab/sabdab/)

AIM:

To study Clostridium Difficile toxin B Crop Domain in complex with Fab Domains of Neutralizing antibody Bezlotoxumab (PDB ID: 4NP4) structure using SAbDab Database.

INTRODUCTION:

Bezlotoxumab is a monoclonal antibody used to reduce the recurrence of Clostridium difficile infections. It is a human monoclonal antibody that binds to Clostridium difficile toxin B and neutralizes its effects. It is used to reduce the recurrence of Clostridium difficile infection in adults receiving antibiotic therapy to treat C. difficile infection and high risk of recurrence. Bezlotoxumab binds to *C.difficile* toxin B, a virulence factor common to practically all *C.difficile*, which prevents the bacteria from infecting host cells. Bezlotoxumab binds two epitopes of toxin B, via two Fab regions, which partially blocks the carbohydrate binding pockets of the toxin resulting in the prevention of toxin B from binding to host cells.

SAbDab is a database of antibody structures that updates on a weekly basis. Each structure is annotated with a number of properties including experimental details, antibody nomenclature (e.g. heavy-light pairings), curated affinity data and sequence annotations. The database is used to inspect individual structures, create and download datasets for analysis, search the database for structures with similar sequences to your query, monitor the known structural repertoire of antibodies. SAbDab has been built by the Oxford Protein Informatics Group (OPIG) under an open-innovation agreement.

METHODOLOGY:

1. Go to SAbDab database (URL: http://opig.stats.ox.ac.uk/webapps/newsabdab/sabdab/).
2. Go to the Search structures page and click on the "Search for a specific PDB entry" tab.
3. Enter the four-digit PDB code of the antibody structure in the search box.
4. Click on “Get Structure”.
5. A results table will be returned.
6. Click on the pdb code in order to open the summary page for the structure.
7. Results are characterized into different sections – Structural details, Visualization, FCs, data in other OPIG databases, downloads and PDB.
8. Interpret the results.

OBSERVATIONS:

Fig1: Homepage for SAbDab database

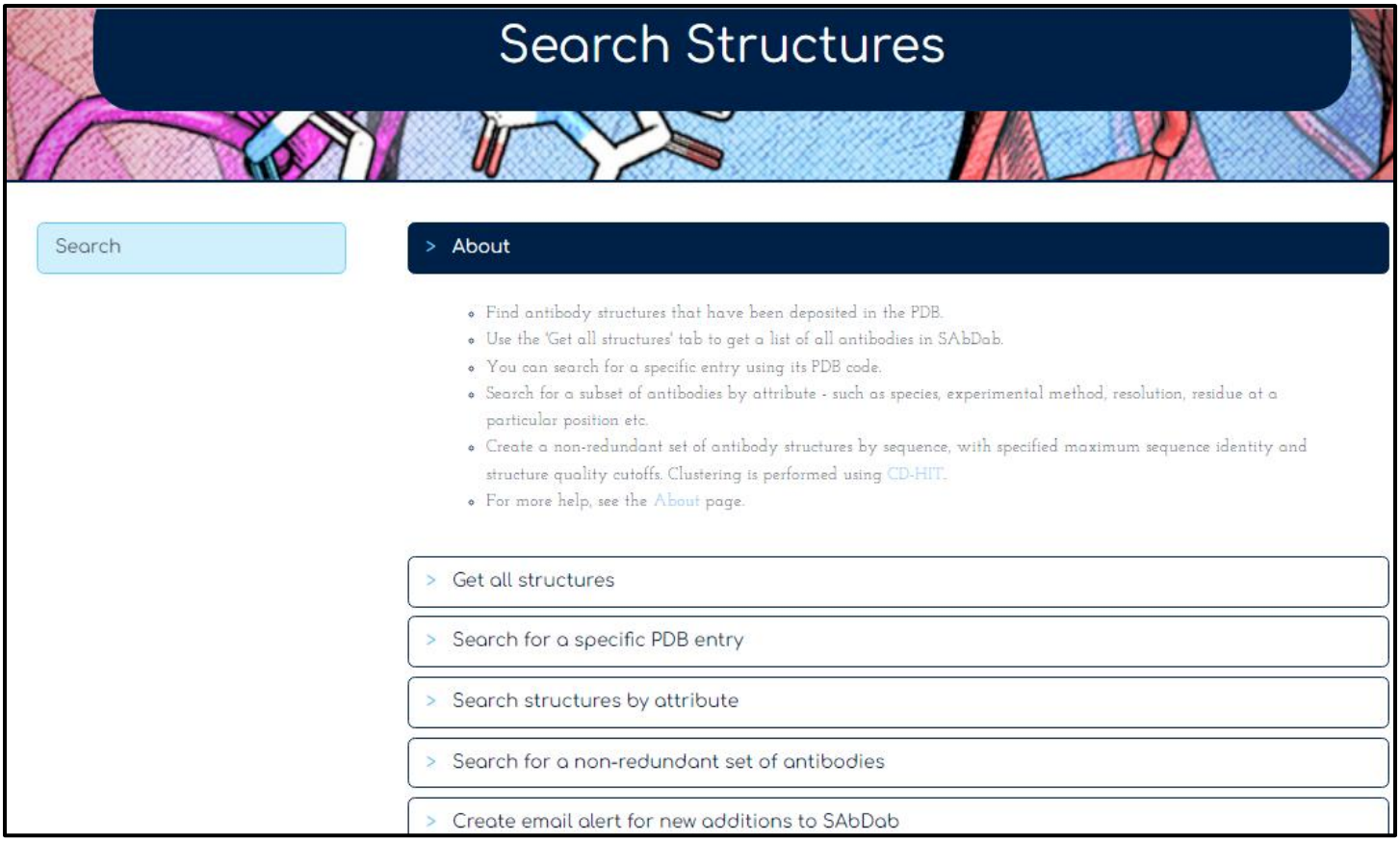

Fig2: Different search options available under Search Structures section

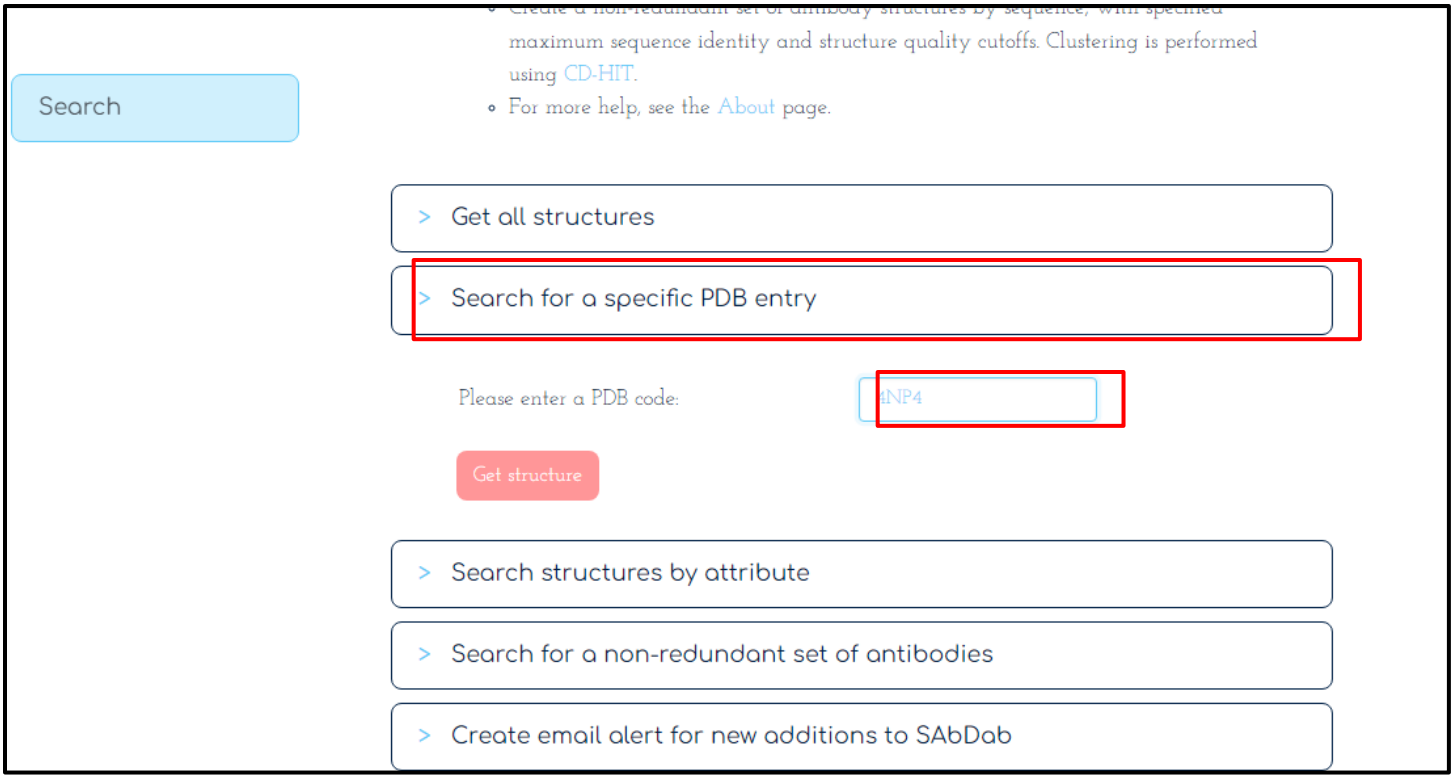

Fig3: Search for 4NP4 PDB query

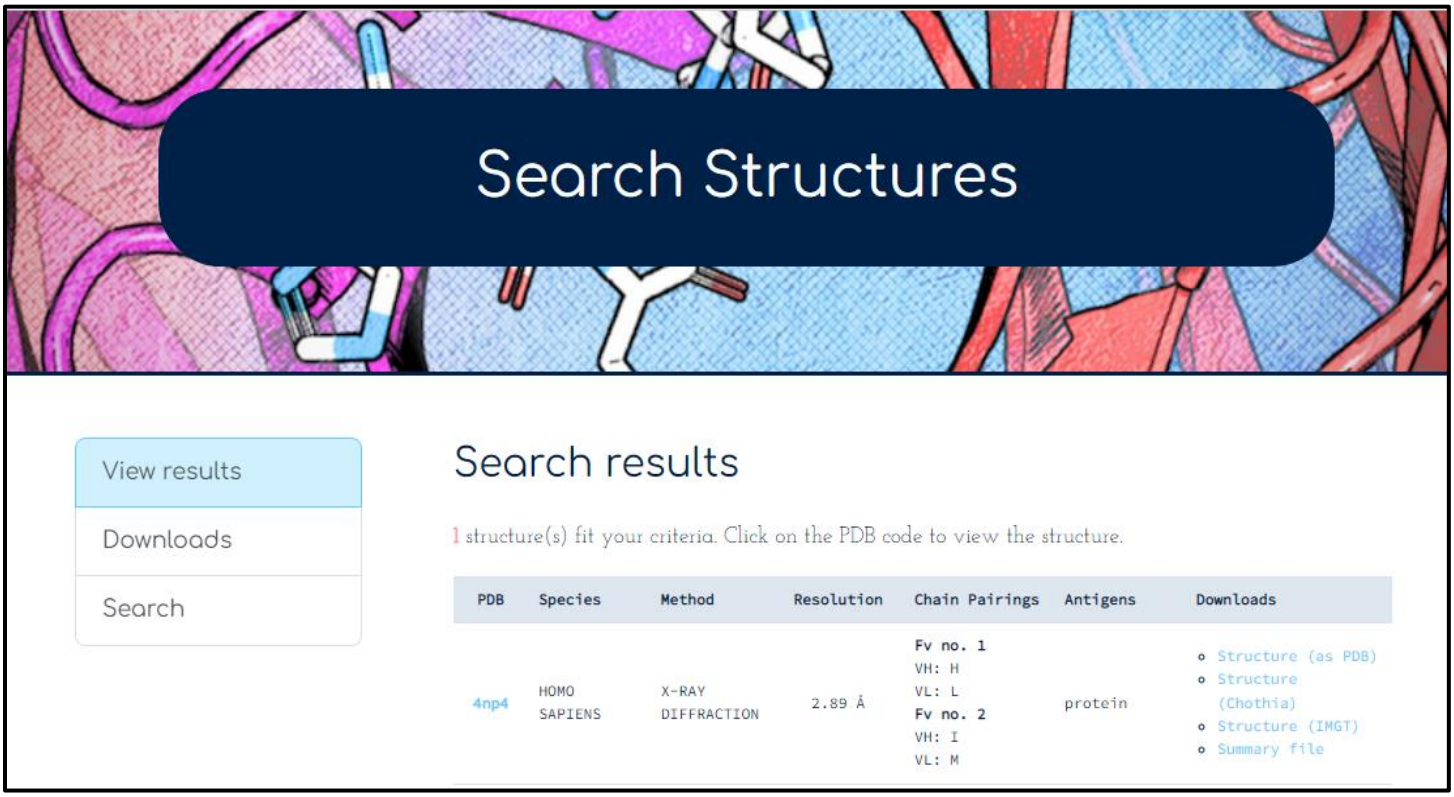

Fig4: Results for PDB entry 4NP4

Fig5: Structural information for PDB ID: 4NP4

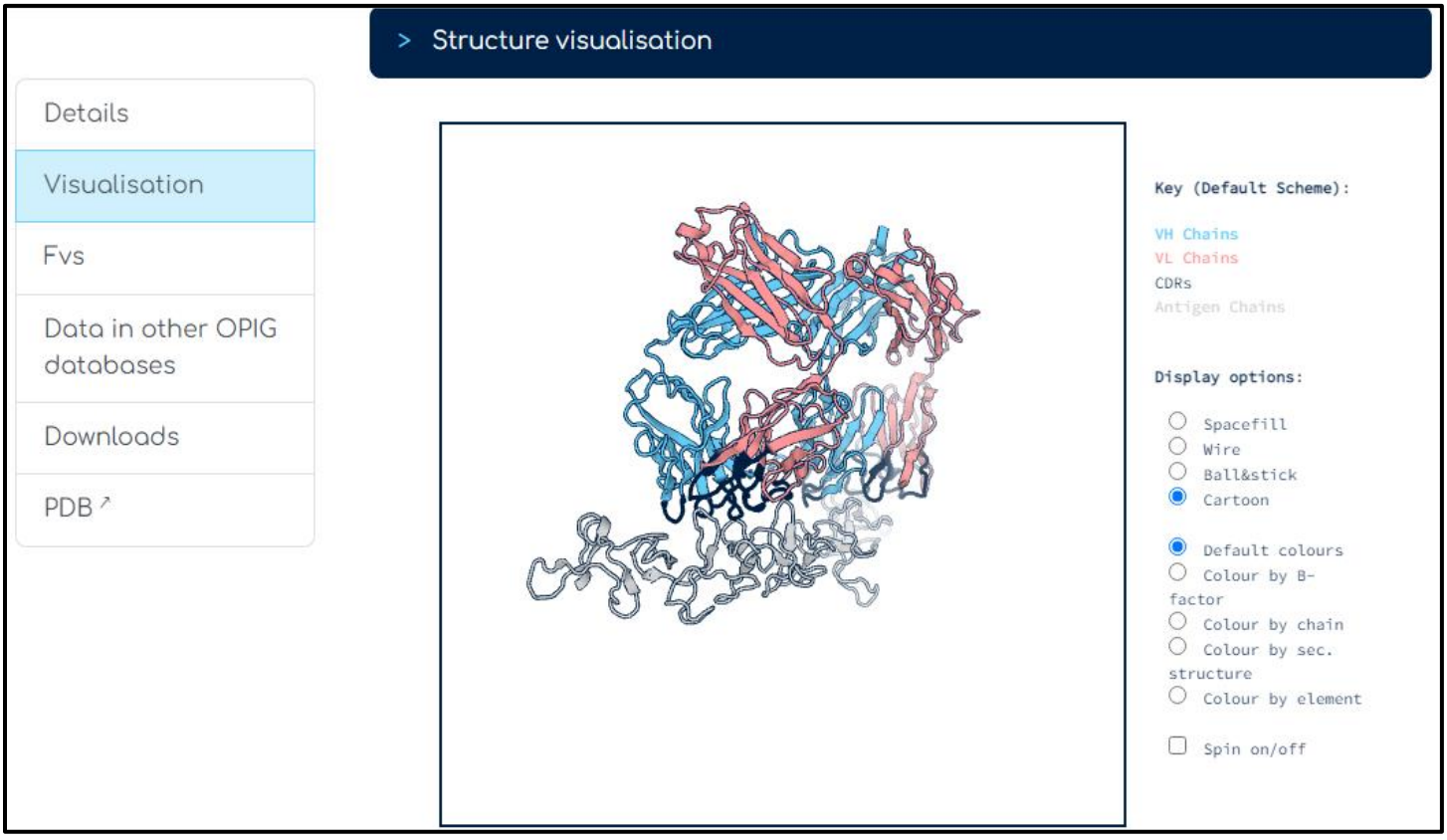

FIG6: Structural visualization for PDB ID: 4NP4

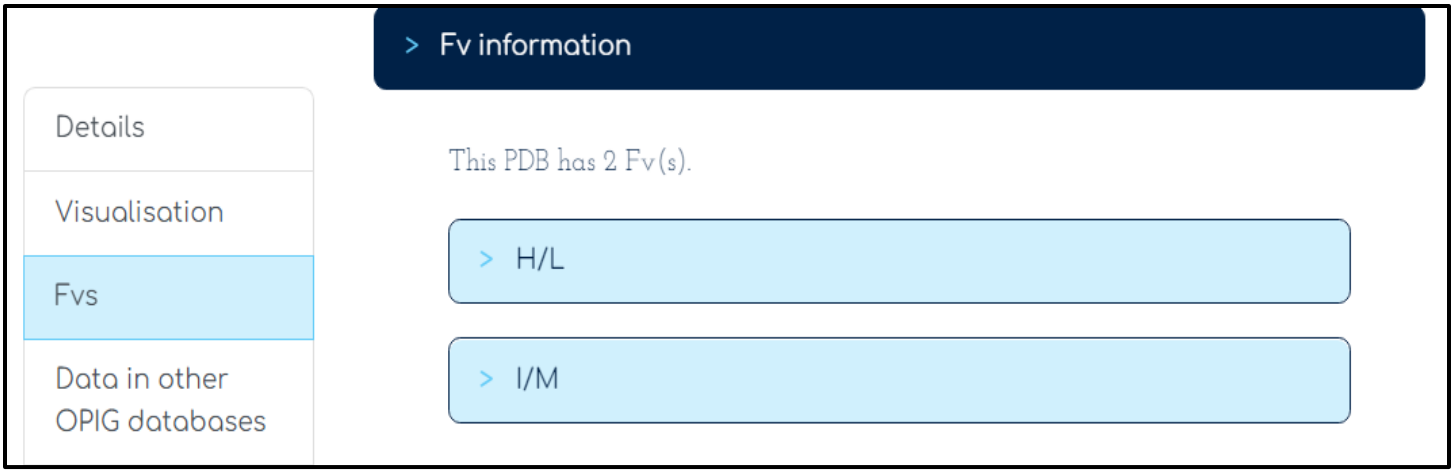

Fig7: Variable fragment (FV) information for PDB ID: 4NP4

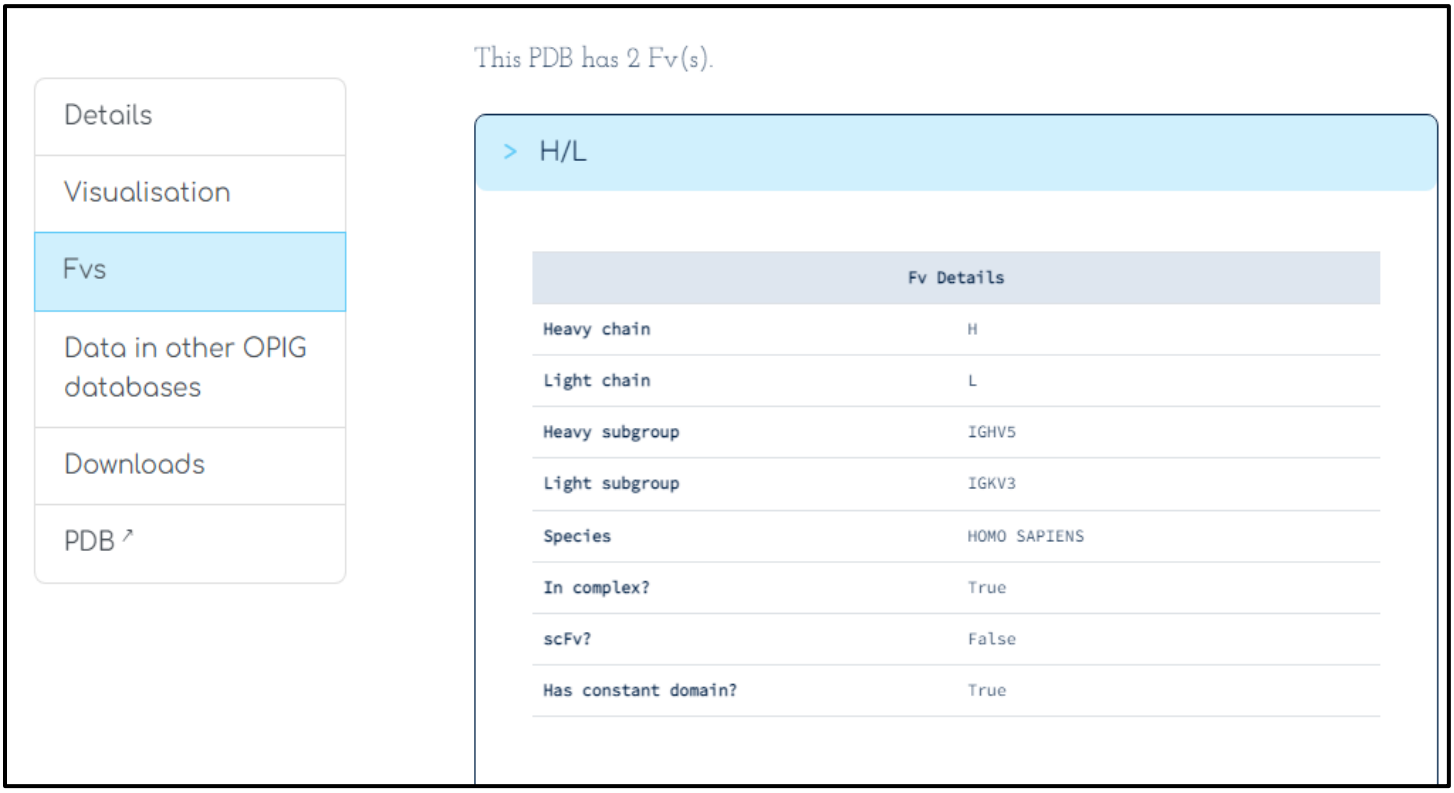

Fig8a: Heavy/Light chain information for PDB ID: 4NP4 under FV(s)

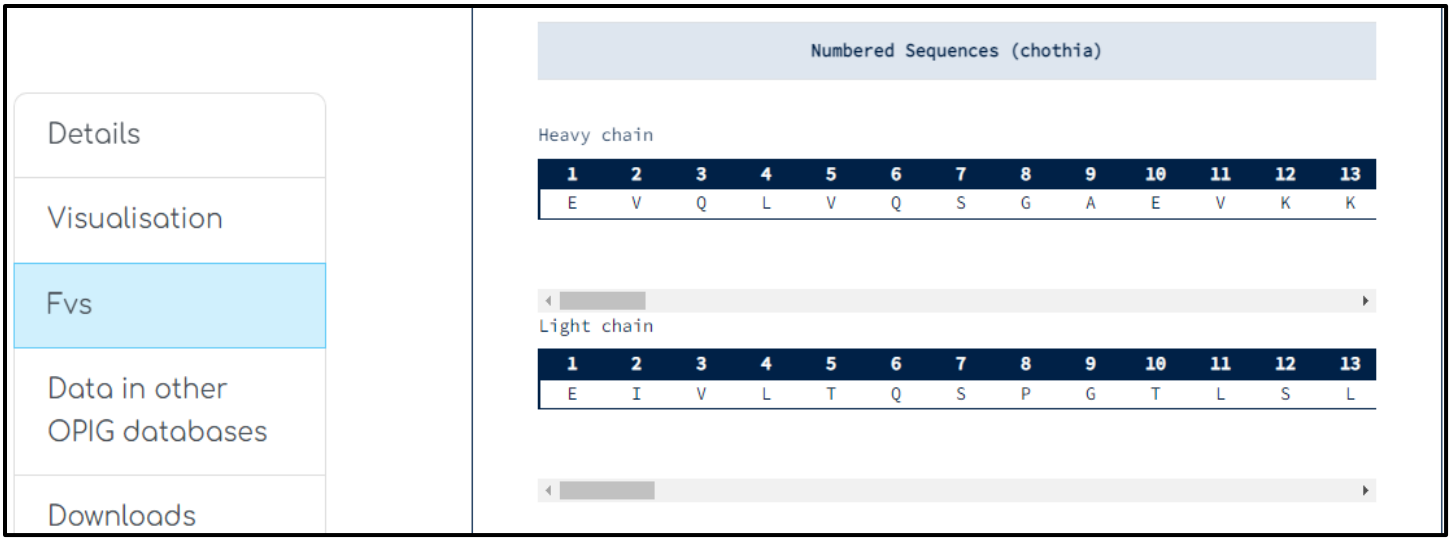

Fig8b: Numbered Sequence (Chothia numbering scheme) information for PDB ID: 4NP4 under FV(s)

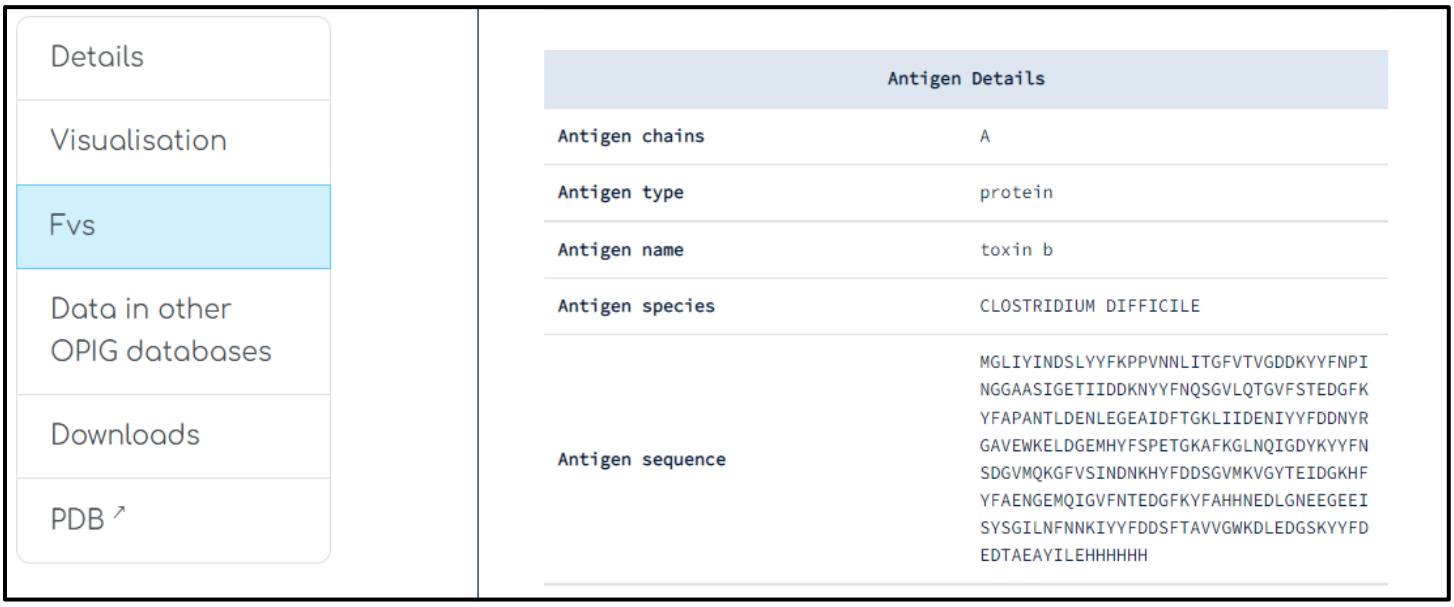

Fig8c: Antigen details for PDB ID: 4NP4 under FV(s)

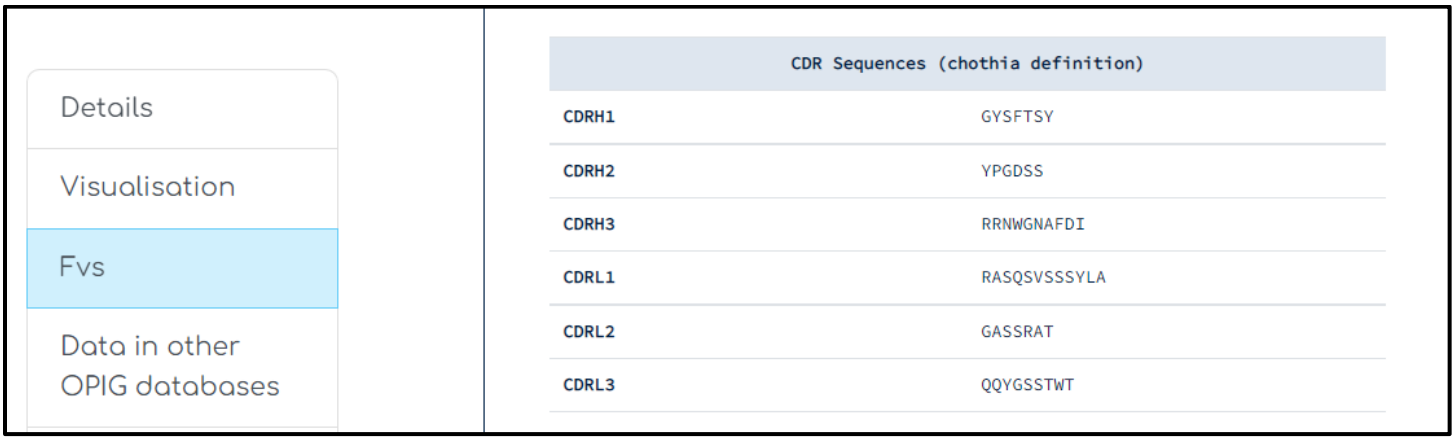

Fig8d: CDR Sequences (Chothia definition) information for PDB ID: 4NP4 under FV(s)

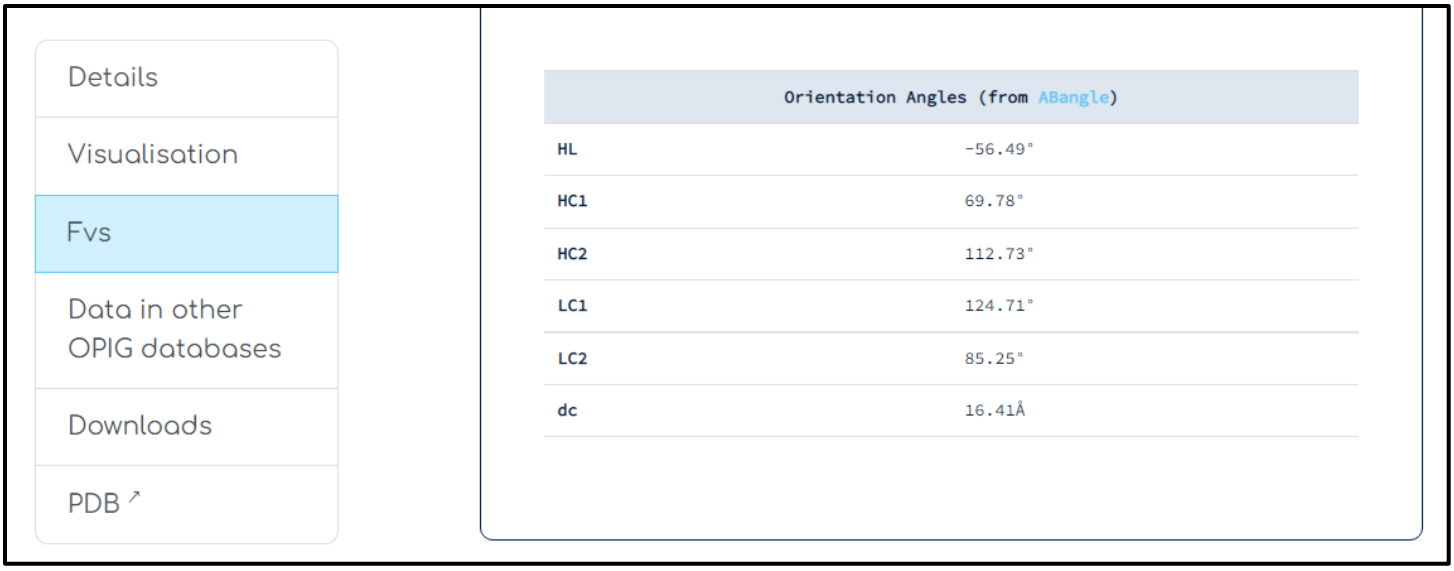

Fig8e: Orientation Angles for PDB ID: 4NP4 under FV(s)

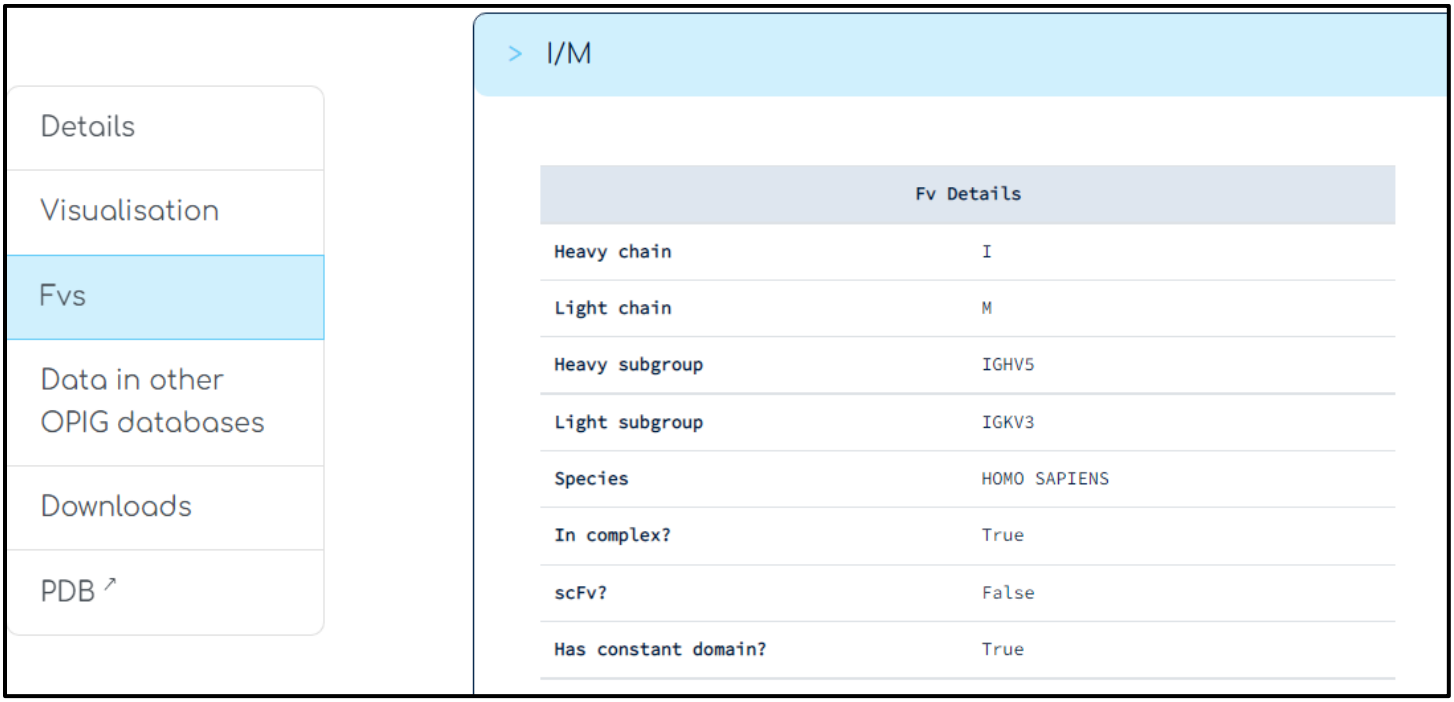

Fig9a: I/M information for PDB ID: 4NP4 under FV(s)

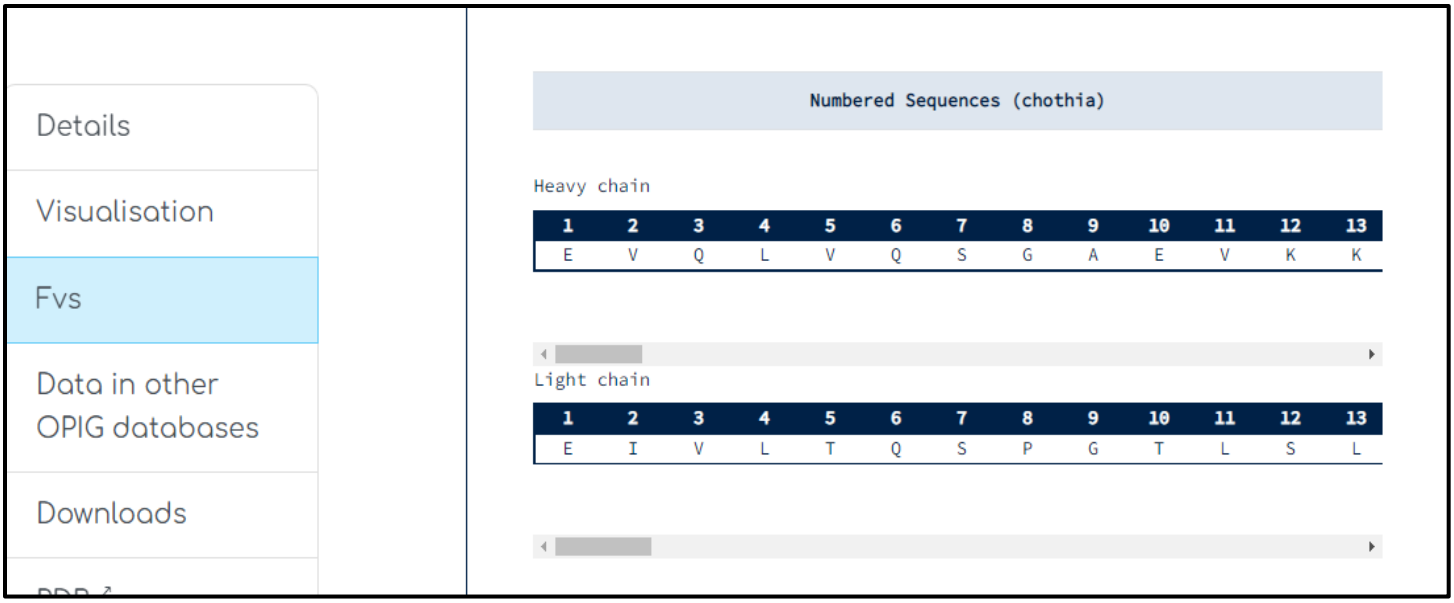

Fig9b: Numbered Sequences (chothia) information for PDB ID: 4NP4

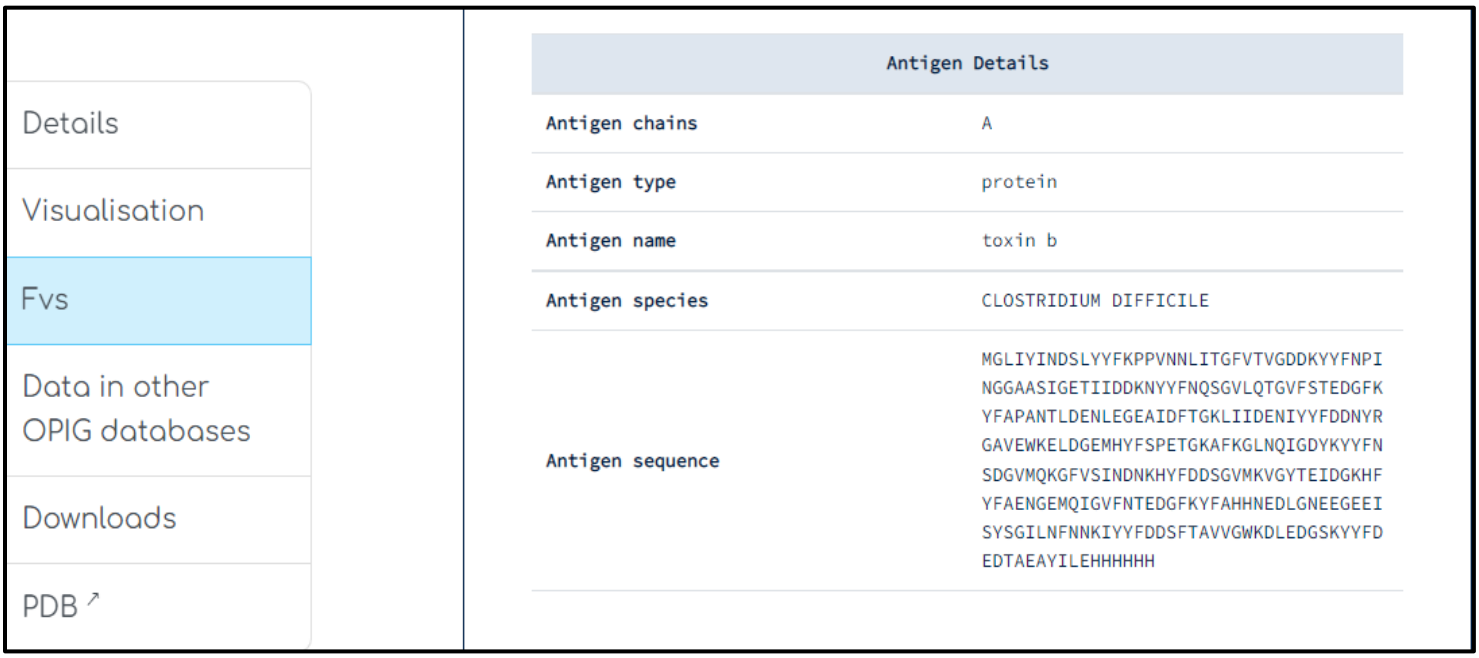

Fig9c: Antigen details for PDB ID: 4NP4 under FV(s)

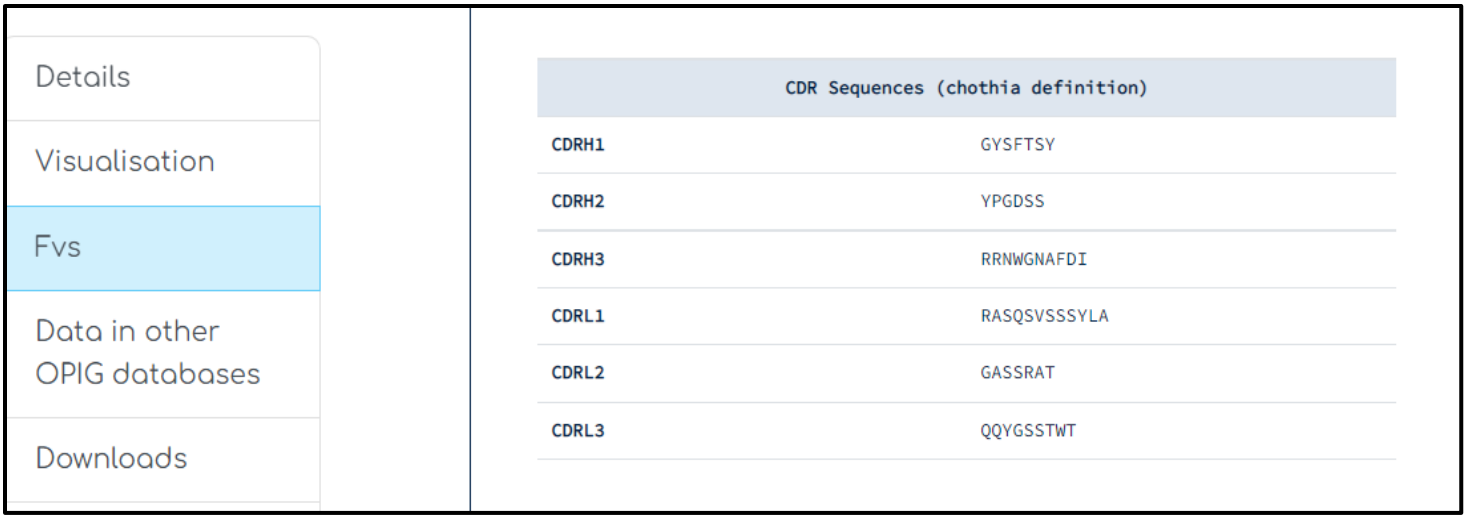

Fig9d: CDR Sequences (chothia definition) information for PDB ID: 4NP4 under FV(s)

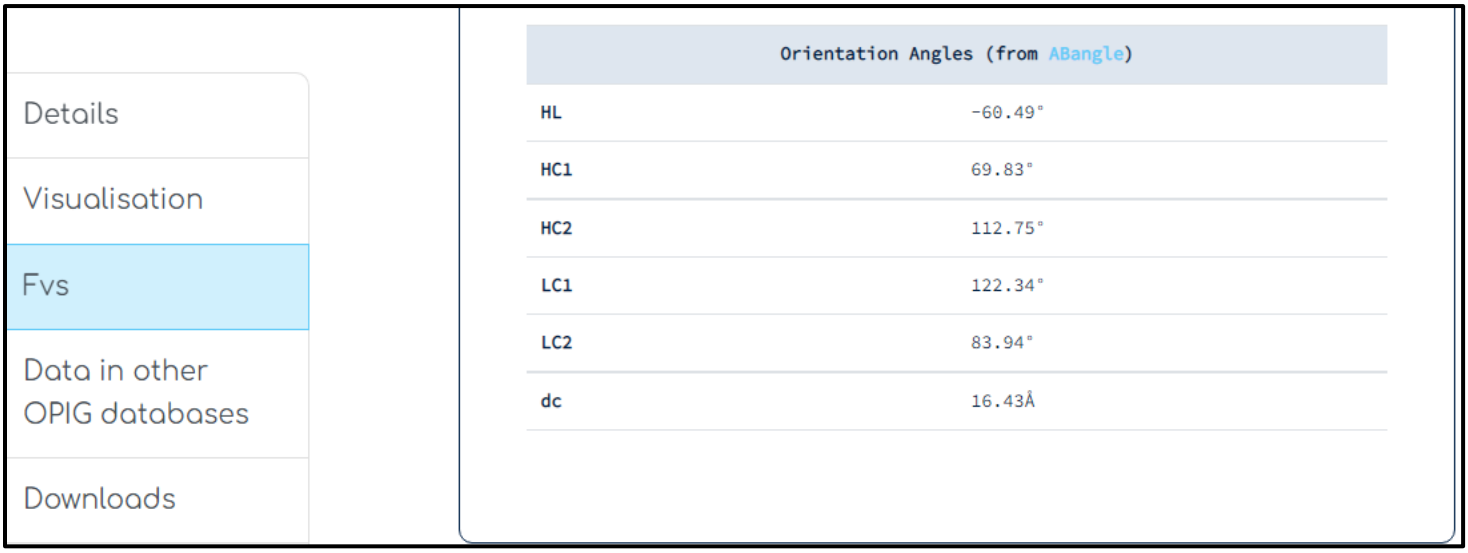

Fig9e: Orientation Angles for PDB ID: 4NP4 under FV(s)

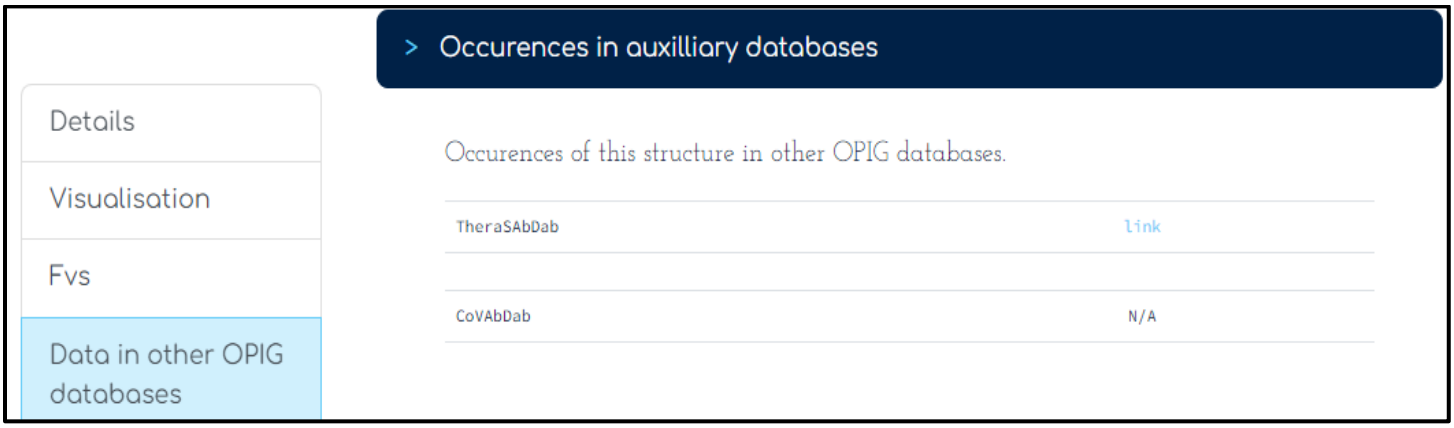

Fig10: Results for Occurrences in auxiliary databases

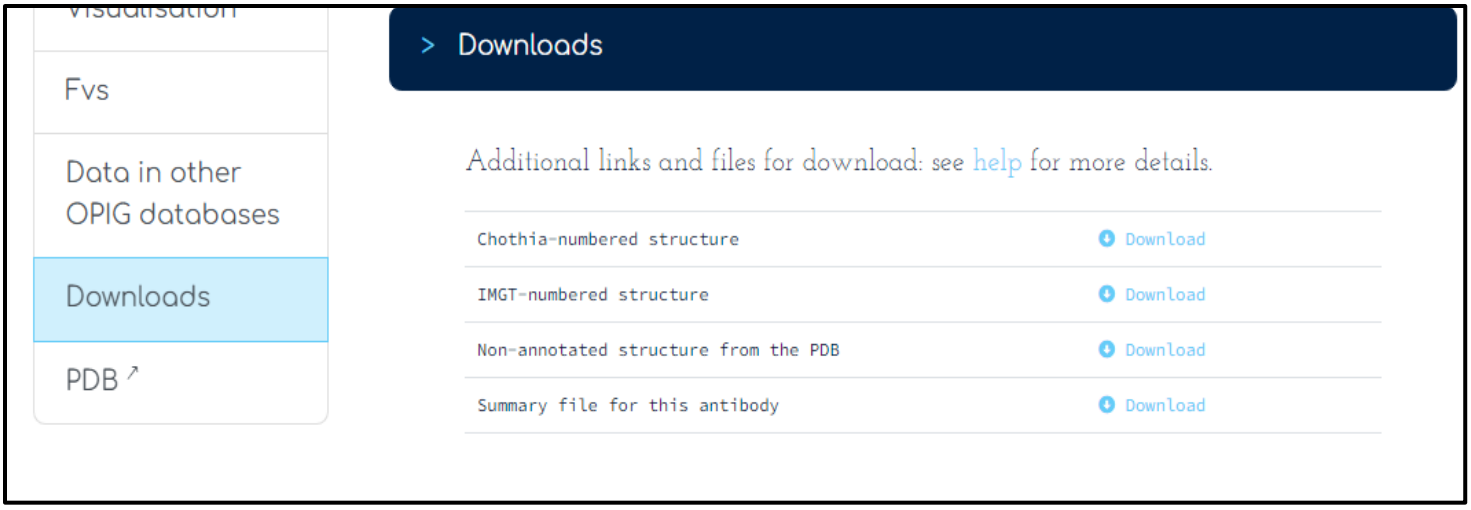

Fig11: Additional links and files for download

RESULTS:

The results are retrieved under different sections. The details are given below:

1. **Header section:**

The header file contains information about the chain types, pairings and antigen pairing. Details regarding the heavy and light chain pairings are generated. The query (PDB id: 4NP4) is of a Homo sapiens species, where the structure has been designed by X-ray diffraction method and the resolution is 2.89Å. The antigen type is a protein. The structure has two heavy-light chain pairs: H/L and I/M. The variable regions of the chain are numbered as per Chothia and IMGT method. The details of the particular structure can be retrieved and downloaded. For each structure, the following files may be downloaded under various section such as,

- The structure in PDB format was deposited.
- The structure in PDB format with the antibody chains numbered using the Chothia numbering.
- The structure in PDB format with the antibody chains numbered using the IMGT numbering.
- A csv summary file containing the information about chain pairings and details about the structure, for example, experimental details, antigen affinity and species.

2. **Details section:**

The data has been fetched from the PDB database. The query structure is of a Clostridium Difficile species. It is a Clostridium Difficile toxin B Crop Domain in complex with Fab Domains of Neutralizing antibody Bezlotoxumab. The experimental method used for designing the structure is X-ray diffraction. The structure information is of a Homo sapiens sample. The number of paired heavy and light chains, that is, Fvs is two. The light chain type is a kappa. The structure has a constant region and the affinity for the structure is 1.9e -11 M.

3. **Visualization:**

The structure can be visualized with heavy chain, light chain, antigen and CDRs annotated in different colors. The color scheme is given wherein heavy chains are indicated in blue, light chains in pink. The CDRs are

indicated in black color and the antigen chains are in the grey color. The query structure is displayed in the Cartoon format.

4. **Variable Fragment (Fvs):**

Information related to variable fragment (Fv) showed that this PDB structure has 2 variable fragments, they are, H/V and I/M. The details about each paired heavy and light chain can be found. These include:

- H and L chain identifiers,
- The Chothia numbered sequence of each chain.
  - The numbering scheme has been provided to annotate equivalent positions in antibodies. The Chothia re-numbered file contains the coordinates of each atom in the structure. Each antibody residue is renumbered with the Chothia numbering scheme over the variable region of domains.
- The details of the antigen and the sequence.
  - The antigen chain type is A. The name of the antigen is toxin b, and species is Clostridium difficile.
- Information about the CDR.
  - The CDR structures are searched according to their type and length of the sequence.
- The orientation angles between the variable heavy and light domains.
  - The orientation angle of the variable domains was described using the ABangle. The distribution of each angle was divided separately.

5. **Data in other OPIG databases:**

The occurrence of this PDB structure has been found in TheraSAbDab database. Link is also provided for the entry of the structure in the TheraSAbDab database.

6. **Downloads:**

In this section, additional links and files for downloading the antibody structure were available. They were as follows:

- Chothia-numbered structure
- IMGT-numbered structure
- Non-annotated structure from the PDB
- Summary file for this antibody.

7. **PDB:**

The antibody structure can be directly accessed to the PDB database. When it is accessed, it goes directly to the summary page of the query.

CONCLUSION:

SAbDab collects, curates and presents antibody structures from the PDB database in a consistent manner. The aim of the database is to provide the antibody research community with a tool to easily create standardized datasets for analysis and to monitor the rapidly increasing amount of available antibody structural data. Detailed information about the structure and a visualization of the antibody and antigen is available. Automated weekly updates keep the data in SAbDab up to date and ensure the longevity of this resource.

REFERENCES:

1. *Bezlotoxumab*. Uses, Interactions, Mechanism of Action. DrugBank Online. (n.d.). Retrieved October 13, 2022, from https://go.drugbank.com/drugs/DB13140
2. Dunbar, J., Krawczyk, K., leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., Deane, C.M. (2014). SAbDab: the structural antibody database. *Nucleic Acids Research*, *42*, D1140-D1146. doi:10.1093/nar/gkt1043
3. Schneider, C., Raybould, M.I.J., Deane, C.M. (2022). SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker. *Nucleic Acids Research*, *50*, D1368-D1372. doi: https://doi.org/10.1093/nar/gkab1050

DATE: 13/10/2022

WEBLEM: 3B

Introduction to Immunoglobulins and its structural features using ABCD Database

(URL: https://web.expasy.org/abcd/)

AIM:

To study a Monoclonal antibody "Erenumab" sequence using ABCD Database.

INTRODUCTION:

Erenumab (Trade name Aimovig) is a human monoclonal antibody designed specifically to bind and antagonize the calcitonin gene-related peptide receptor (CGRPR) as a means to prevent migraines. Studies since 1985 have demonstrated that CGRP levels increase during acute migraine attacks in migraine-suffering patients but normalize after efficacious sumatriptan therapy. Moreover, research has also shown that intravenous administration of CGRP can induce migraine-like attacks in migraine-suffering patients. For all these reasons, the binding and antagonism of CGRP receptors was designed to be mechanism of action for Erenumab to take advantage of in reversing the migraine-inducing activity of natural CGRP.

The ABCD is a database of chemically defined antibodies, i.e. all antibodies with a known primary sequence, and with a known target (to which, most often, a UniProtKB or ChEBI ID can be attributed).The ABCD database provides a comprehensive list of sequenced antibodies with their known targets. Each antibody is assigned a unique ID number that can be used in academic publications to increase reproducibility of experiments.There are increasing concerns about reproducibility of experimental biomedical research, partially attributed to the lack of reliable and standardized biological reagents. Despite being the most widely used class of protein-binding reagents, antibodies are often poorly characterized and ill-defined, and thus contribute largely to the lack of reliability and reproducibility of biomedical research. The ABCD database promotes the use of standardized and well-characterized antibodies in biomedical results. It reduces the need to use poorly-defined antibodies produced in immunized animals.

METHODOLOGY:

1. Go to ABCD Database (https://web.expasy.org/abcd/).
2. Enter the monoclonal antibody name "Erenumab" in the search box.
3. Click on the "Search" option.
4. Results will appear.
5. Click on ID to get a detailed result page of the monoclonal antibody "Erenumab".
6. Interpret the results.

OBSERVATIONS:

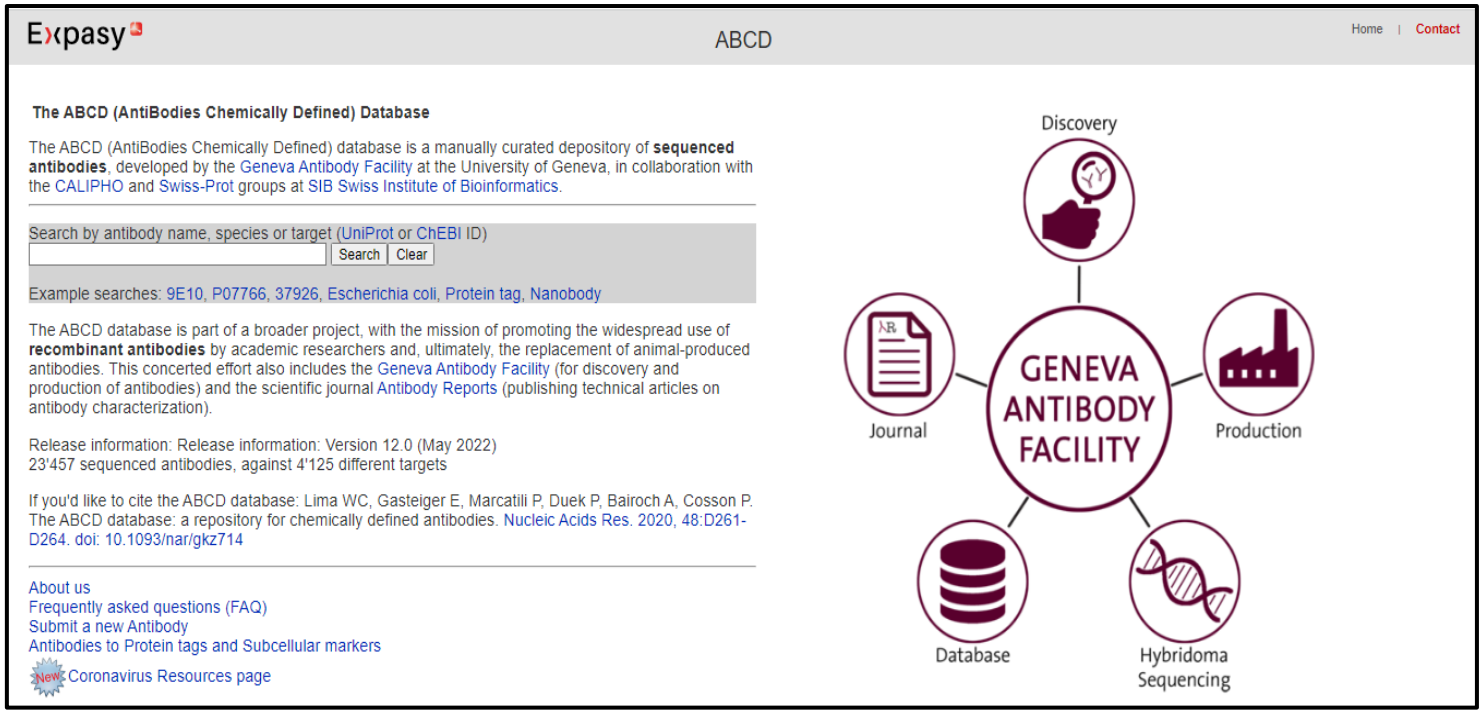

Fig1: Homepage for ABCD database

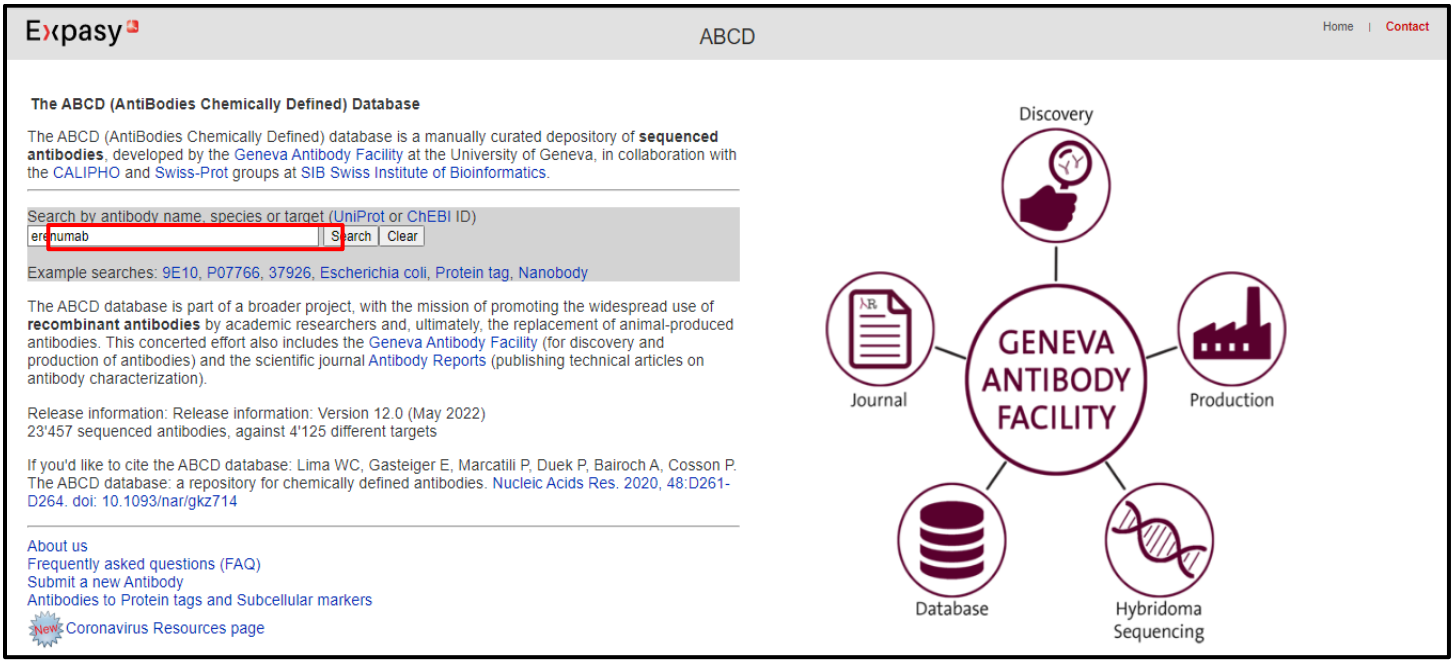

Fig2: Searching monoclonal antibody “Erenumab”

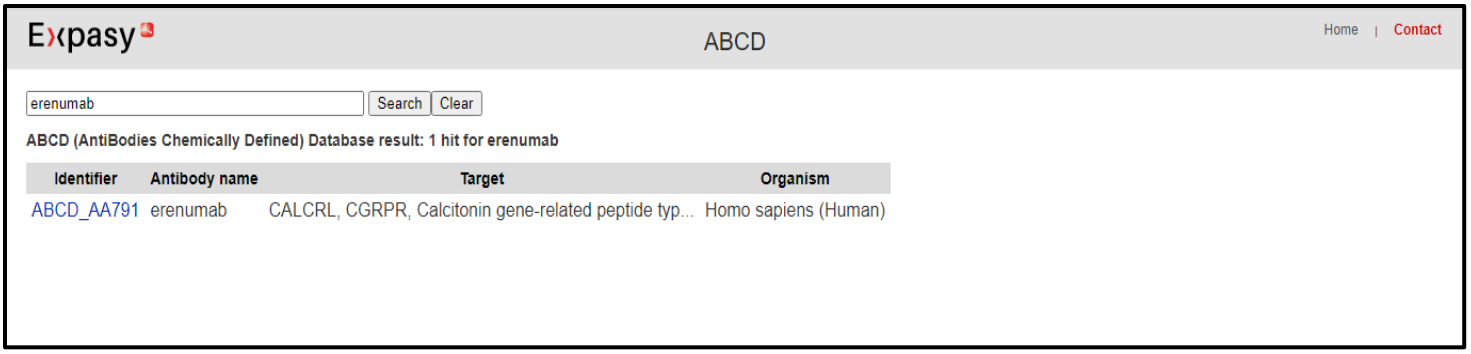

Fig3: Hit Page for "Erenumab"

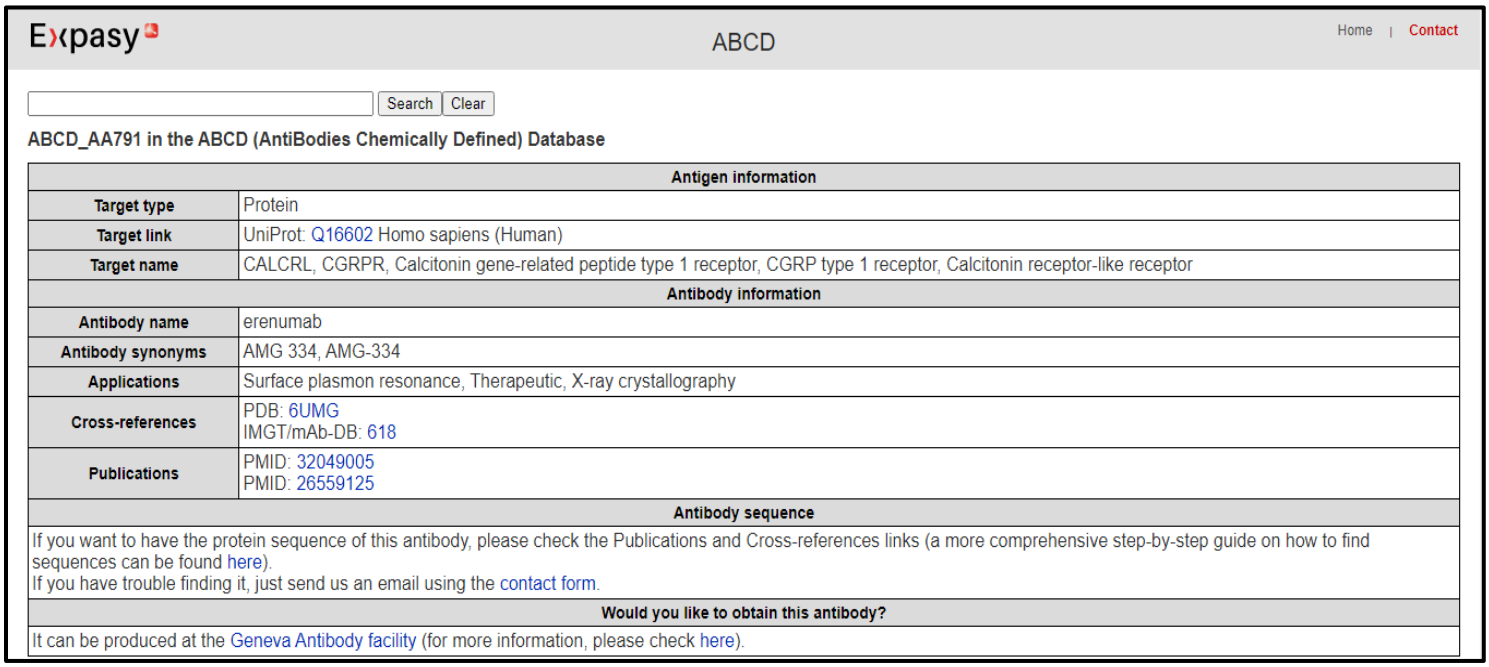

Fig4: Antigen and Antibody information for query "Erenumab"

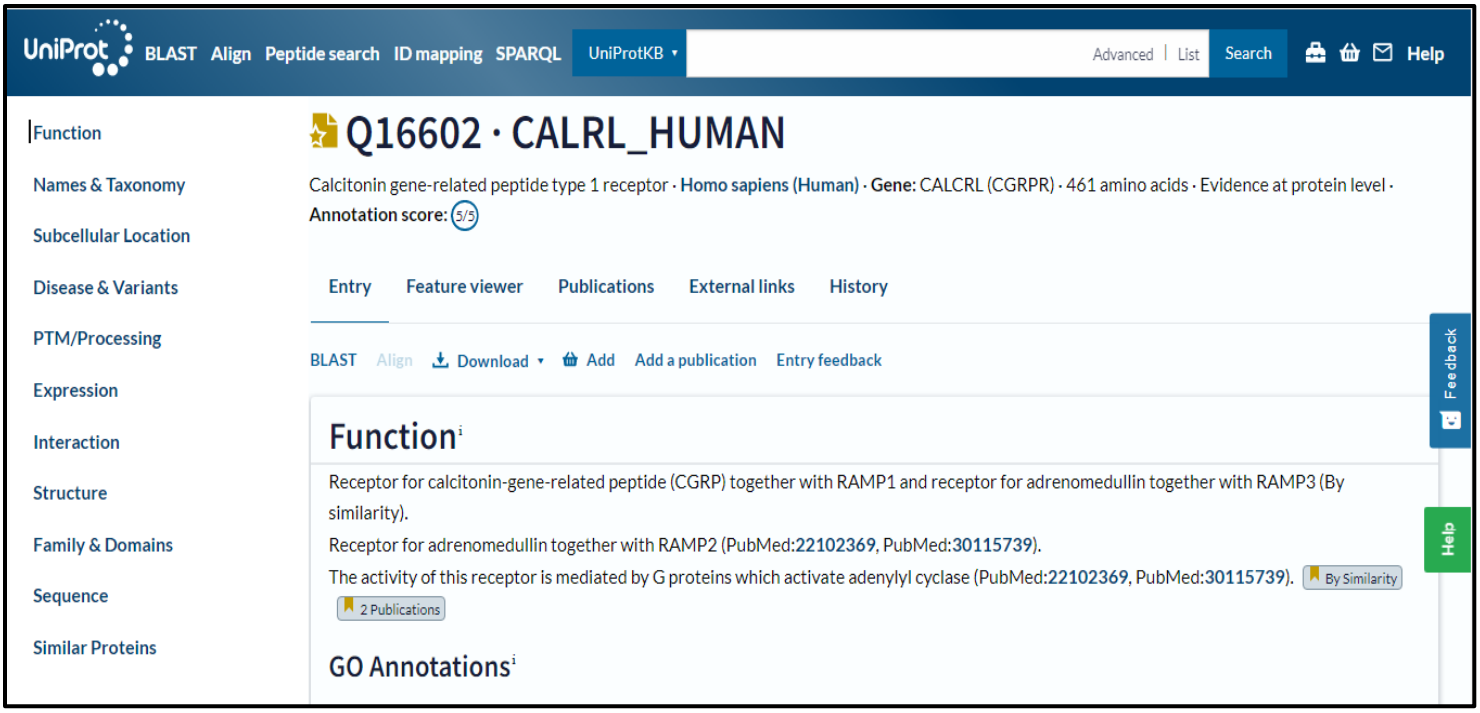

Fig5: Results for antigen information in UniProt database

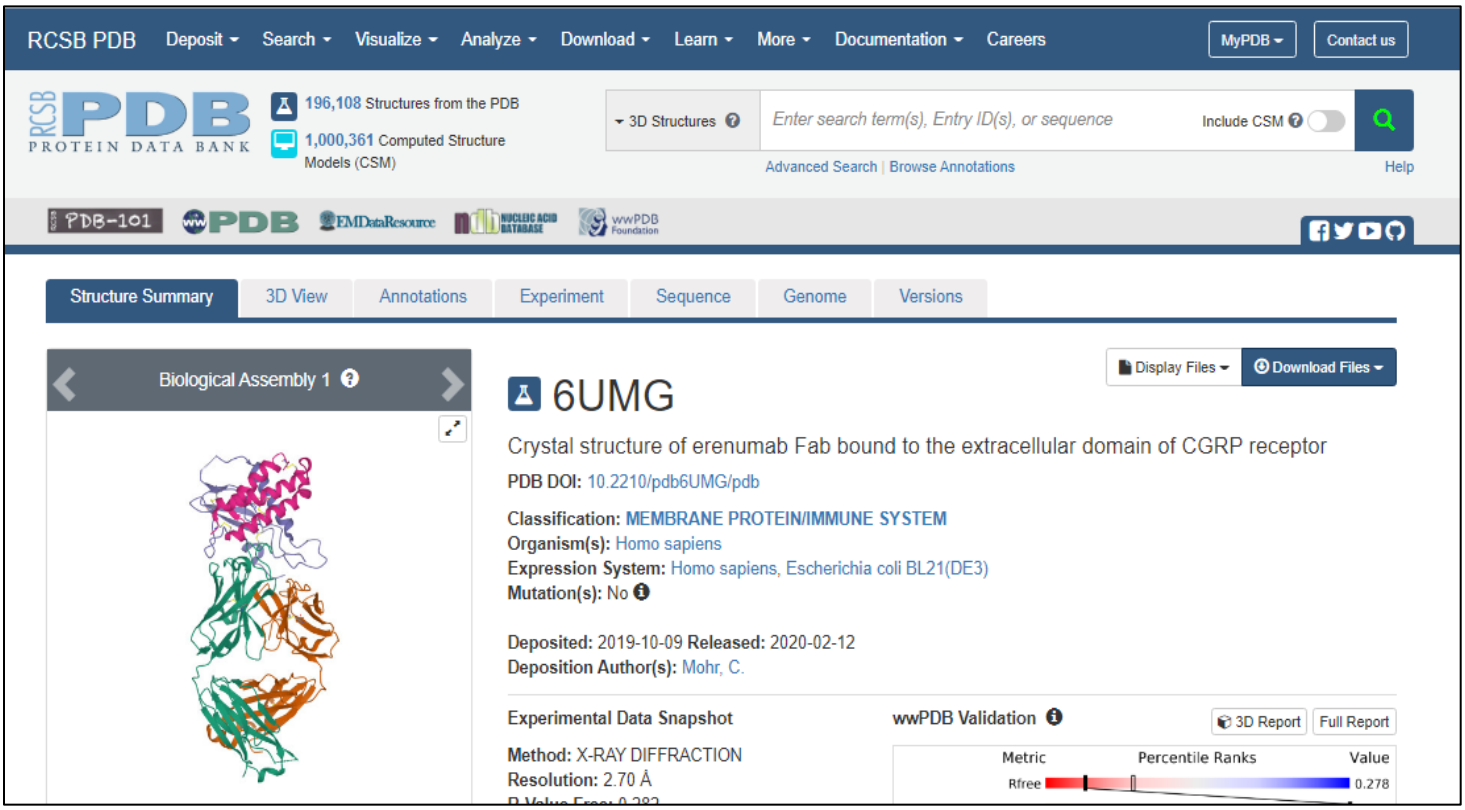

Fig6: Results for antibody information in PDB database

RESULTS:

The ABCD database was used to retrieve the information for the query "Erenumab". The following are the information retrieved for the query.

1. **Antigen information:**

The target type for my particular query is for a Protein "Erenumab" that has been obtained from a Homo sapiens species sample. Target names have been mentioned on which the monoclonal antibody will work.

2. **Antibody information:**

A common name and a list of synonyms have been mentioned. The monoclonal antibodies can be synthesized by various methods such as Surface plasma resonance, therapeutic, X-ray crystallography. The information can be cross-referenced from PDB and IMGT databases. The sequence information searched for my particular monoclonal antibody has been satisfied by showing a link through the Uniprot database. Information about the target (UniProtKB number and description) and about the epitope recognition is also available. Cross-references to original databases and two publications, in which the antibody is described have been provided.

CONCLUSION:

The ABCD database aims at helping to improve reproducibility in academic research by providing a unique, unambiguous identifier associated to each antibody sequence. It also allows determining rapidly if a sequenced antibody is available for a given antigen. The information has been provided for the antigen sequence by providing a link through the Uniprot database. This indicates that the antigen can be used for a particular receptor that can be used further for docking purposes.

REFERENCES:

1. *Erenumab*. Uses, Interactions, Mechanism of Action. DrugBank Online. (n.d.). Retrieved October 13, 2022, from https://go.drugbank.com/drugs/DB14039
2. Lima, W.C., Gasteiger, E., Marcatili, P., Duek, P., Bairoch, A., Cosson, P. (2019). The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Research*, *48*, D261-D264. doi: 10.1093/nar/gkz714