

**NAME:** SHALMON N. ANANDAS

**CLASS:** M.Sc PART – II

**COURSE:** BIOINFORMATICS

**ACADEMIC YEAR:** 2022-2023

**ROLL NO:** 92

**PAPER CODE:** GNKPSBI303 (PAPER 3)

**COURSE TITLE:** IMMUNOLOGY AND IMMUNOINFORMATIC

**GURU NANAK KHALSA COLLEGE MATUNGA, MUMBAI-**

**400019**

**DEPARTMENT OF BIOINFORMATICS CERTIFICATE**

This is to certify that Mr. Shalmon N Anandas of M.Sc. Part II Bioinformatics has satisfactorily completed the practical semester III course prescribed by the university of Mumbai during academic year 2022-2023

**TEACHER IN CHARGE**

**HEAD OF DEPARTMENT**

## INDEX

Sr. No.	WEBLEM NAME	PAGE NO	DATE	Signature
1	Introduction to Immunoglobins and its structural features	1-4	08/10/22	
1a	To Browse, search and retrieve “Immunoglobulin G” sequence using UniProt database	5-9	08/10/22	
1b	To Browse, search and retrieve “Immunoglobulin G” sequence using PDB database	10-12	08/10/22	
2	Introduction to Antibody sequence and structure along with Antibody numbering methods such as Kabat, Chothia and other equivalent methods and its importance	13-18	10/10/22	
3	Introduction to SAbDab (Antibody Structure Database) and ABCD (Antibody Sequence Database) Database	19-26	13/10/22	
3a	To study Clostridium Difficile toxin B Crop Domain in complex with Fab Domains of Neutralizing antibody Bezlotoxumab (PDB ID: 4NP4) structure using SAbDab Database	27-38	13/10/22	
3b	To study a Monoclonal antibody “Erenumab” seqnce using ABCD Database	39-43	13/10/22	
4	Introduction to STCRDAb database and visualisation of structure using PFV3D Tool	44-49	13/10/22	
4a	To identify the CDR for the crystal structure of the TK3 TCR in complex with HLA- B*3501/HPVG (PDB ID: 3MV6) using STCRDab Database	50-57	19/10/22	
4b	To study protein feature visualization for Immunoglobulin lambda variable 6-57 on 3D structure using PFV3D tool	58-67	19/10/22	
5	Introduction to Yvis platform for studying variable and constant domains	66-72	29/10/22	
5a	To study variable and constant doamins using topology diagram on Yvis platform	73-81	29/10/22	
6	Introduction to Ag-Ab Interaction Database (AgAbDb)	82-87	23/10/22	
7	Introduction to IEJD	88-98	22/10/22	

	Database		2	
7a	To predict B cell epitope for 1OVZ using Discotope Server 1.1	97-103	22/10/2 2	
7b	To predict peptides for IGLL1_HUMAN binding to MHC class I and class II molecules using Tepitool	104-109	22/10/2 2	
8	To understand various web-based tools for vaccine designing	110-120	16/10/22	

## WEBLEM 1

### Introduction to Immunoglobins and its structural features

#### IMMUNOGLOBULIN (IG)

Immunoglobulins are glycoprotein molecules that are produced by plasma cells in response to an immunogen and which function as antibodies. The immunoglobulins derive their name from the finding that they migrate with globular proteins when antibody-containing serum is placed in an electrical field.

#### GENERAL FUNCTIONS OF IMMUNOGLOBULINS

##### a. Antigen Binding:

Immunoglobulins bind specifically to one or a few closely related antigens. Each immunoglobulin actually binds to a specific antigenic determinant. Antigen binding by antibodies is the primary function of antibodies and can result in protection of the host. The valency of antibody refers to the number of antigenic determinants that an individual antibody molecule can bind. The valency of all antibodies is at least two and in some instances more.

##### b. Effector Functions:

Frequently the binding of an antibody to an antigen has no direct biological effect. Rather, the significant biological effects are a consequence of secondary "effector functions" of antibodies. The immunoglobulins mediate a variety of these effector functions. Usually the ability to carry out a particular effector function requires that the antibody bind to its antigen. Not every immunoglobulin will mediate all effector functions. Such effector functions include:

- i. Fixation of complement - This results in lysis of cells and release of biologically active molecules.
- ii. Binding to various cell types - Phagocytic cells, lymphocytes, platelets, mast cells, and basophils have receptors that bind immunoglobulins. This binding can activate the cells to perform some function. Some immunoglobulins also bind to receptors on placental trophoblasts, which results in transfer of the immunoglobulin across the placenta. As a result, the transferred maternal antibodies provide immunity to the fetus and newborn

#### BASIC STRUCTURE OF IMMUNOGLOBULINS

The basic structure of the immunoglobulins is illustrated in figure 2. Although different immunoglobulins can differ structurally, they all are built from the same basic units.

a. Heavy and Light Chains

All immunoglobulins have a four chain structure as their basic unit. They are composed of two identical light chains (23kD) and two identical heavy chains (50-70kD)

b. Disulfide bonds

- i. Inter-chain disulfide bonds - The heavy and light chains and the two heavy chains are held together by inter-chain disulfide bonds and by non-covalent interactions. The number of inter-chain disulfide bonds varies among different immunoglobulin molecules.
- ii. Intra-chain disulfide binds - Within each of the polypeptide chains there are also intra-chain disulfide bonds.

c. Variable and Constant Region

When the amino acid sequences of many different heavy chains and light chains were compared, it became clear that both the heavy and light chain could be divided into two regions based on variability in the amino acid sequences. These are the:

- i. Light Chain – VL (110 amino acids) and CL (110 amino acids)
- ii. Heavy Chain – VH (110 amino acids) and CH (330-440 amino acids)
- d. Hinge Region

This is the region at which the arms of the antibody molecule forms a Y. It is called the hinge region because there is some flexibility in the molecule at this point.

- e. Domains

Three dimensional images of the immunoglobulin molecule show that it is not straight as depicted in figure 2A. Rather, it is folded into globular regions each of which contains an intra-chain disulfide bond. These regions are called domains.

- i. Light Chains Domains: VL and CL
- ii. Heavy Chain Domains – VH, CH1 – CH3 (or CH4)
- f. Oligosaccharides

Carbohydrates are attached to the CH2 domain in most immunoglobulins. However, in some cases carbohydrates may also be attached at other locations.

## **STRUCTURE OF THE VARIABLE REGION**

- a. Hypervariable (HVR) or complementary determining region(CDR)

Comparisons of the amino acid sequences of the variable regions of immunoglobulins show that most of the variability resides in three regions called the hypervariable regions or the complementarity determining regions as illustrated in figure 3. Antibodies with different specificities (i.e. different combining sites) have different complementarity determining regions while antibodies of the exact same specificity have identical complementarity determining regions (i.e. CDR is the antibody combining site). Complementarity determining regions are found in both the H and the L chains.

- b. Framework regions

The regions between the complementarity determining regions in the variable region are called the framework regions. Based on similarities and differences in the framework regions the immunoglobulin heavy and light chain variable regions can be divided into groups and subgroups. These represent the products of different variable region genes.

## **IMMUNOGLOBULIN FRAGMENTS: STRUCTURE/FUNCTION RELATIONSHIPS**

Immunoglobulin fragments produced by proteolytic digestion have proven very useful in elucidating structure/function relationships in immunoglobulins.

a. Fab

Digestion with papain breaks the immunoglobulin molecule in the hinge region before the H-H inter-chain disulfide bond. This results in the formation of two identical fragments that contain the light chain and the VH and CH1 domains of the heavy chain.

Antigen binding - These fragments were called the Fab fragments because they contained the antigen binding sites of the antibody. Each Fab fragment is monovalent whereas the original molecule was divalent. The combining site of the antibody is created by both VH and VL. An antibody is able to bind a particular

antigenic determinant because it has a particular combination of VH and VL. Different combinations of a VH and VL result in antibodies that can bind a different antigenic determinants.

b. Fc

Digestion with papain also produces a fragment that contains the remainder of the two heavy chains each containing a CH<sub>2</sub> and CH<sub>3</sub> domain. This fragment was called Fc because it was easily crystallized.

Effector functions - The effector functions of immunoglobulins are mediated by this part of the molecule. Different functions are mediated by the different domains in this fragment (figure 5). Normally the ability of an antibody to carry out an effector function requires the prior binding of an antigen; however, there are exceptions to this rule.

c. F(ab')<sub>2</sub>

Treatment of immunoglobulins with pepsin results in cleavage of the heavy chain after the H-H inter-chain disulfide bonds resulting in a fragment that contains both antigen binding sites (figure 6). This fragment was called F(ab')<sub>2</sub> because it is divalent. The Fc region of the molecule is digested into small peptides by pepsin. The F(ab')<sub>2</sub> binds antigen but it does not mediate the effector functions of antibodies.

## **HUMAN IMMUNOGLOBULIN CLASSES, SUBCLASSES, TYPES AND SUBTYPES**

a. Immunoglobulin classes

The immunoglobulins can be divided into five different classes, based on differences in the amino acid sequences in the constant region of the heavy chains. All immunoglobulins within a given class will have very similar heavy chain constant regions. These differences can be detected by sequence studies or more commonly by serological means (i.e. by the use of antibodies directed to these differences).

- i. IgG – Gamma Heavy
- ii. IgM – Mu Heavy
- iii. IgA – Alpha Heavy
- iv. IgD – Delta Heavy
- v. IgE – Epsilon Heavy
- vi. Immunoglobulin Subclasses

The classes of immunoglobulins can be divided into subclasses based on small differences in the amino acid sequences in the constant region of the heavy chains. All immunoglobulins within a subclass will have very similar heavy chain constant region amino acid sequences. Again these differences are most commonly detected by serological means.

a. IgG Subclasses

- a) IgG1 - Gamma 1 heavy chains
- b) IgG2 - Gamma 2 heavy chains
- c) IgG3 - Gamma 3 heavy chains
- d) IgG4 - Gamma 4 heavy chains

b. IgA Subclasses

- a) IgA1 - Alpha 1 heavy chains
- b) IgA2 - Alpha 2 heavy chains

c. Immunoglobulin Types

Immunoglobulins can also be classified by the type of light chain that they have. Light chain types are based on differences in the amino acid sequence in the constant region of the light chain. These differences are detected by serological means.

1.Kappa light chains

2.Lambda light chains

d. Immunoglobulin Subtypes

The light chains can also be divided into subtypes based on differences in the amino acid sequences in the constant region of the light chain.

1. Lambda subtypes

a) Lambda 1

b) Lambda 2

c) Lambda 3

d) Lambda 4

e. Nomenclature

Immunoglobulins are named based on the class, or subclass of the heavy chain and type or subtype of light chain. Unless it is stated precisely, you should assume that all subclass, types and subtypes are present. IgG means that all subclasses and types are present.

g. Heterogeneity

Immunoglobulins considered as a population of molecules are normally very heterogeneous because they are composed of different classes and subclasses each of which has different types and subtypes of light chains. In addition, different immunoglobulin molecules can have different antigen binding properties because of different VH and VL regions.

## REFERENCES:

1. IMMUNOGLOBULINS - STRUCTURE AND FUNCTION. (n.d.).  
Www.microbiologybook.org.  
<https://www.microbiologybook.org/mobile/m.immuno-4.htm>

**WEBLEM 1a****UniProt Database**

(URL:

[https://www.uniprot.org/ \)](https://www.uniprot.org/)**AIM:**

To browse, search & retrieve immunoglobulin sequence using UniProt database.

**INTRODUCTION:**

UniProt is a long-standing collection of databases that enable scientists to navigate the vast amount of sequence and functional information available for proteins. The UniProt Knowledgebase (UniProtKB) is the central resource that combines UniProtKB/Swiss-Prot and UniProtKB/TrEMBL. UniProtKB/Swiss-Prot contains over 550 000 sequences that have been created by our expert biocuration team. For these entries experimental information has been extracted from the literature and organized and summarized, greatly easing scientists access to protein information. UniProtKB/TrEMBL provides a further 60 million sequences that have been largely derived from high throughput sequencing of DNA. These entries are annotated by our rule based automatic annotation systems. We also provide a series of UniRef databases that provide sequence sets trimmed at various levels of sequence identity. Finally we provide the UniProt Archive (UniParc) that provides a complete set of known sequences, including historical obsolete sequences

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added. This includes widely accepted biological ontologies, classifications and cross-references, and clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data.

The UniProt Knowledgebase consists of two sections: a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis, and a section with computationally analyzed records that await full manual annotation. For the sake of continuity and name recognition, the two sections are referred to as "UniProtKB/Swiss-Prot" (reviewed, manually annotated) and "UniProtKB/TrEMBL" (unreviewed, automatically annotated), respectively.

UniProtKB/TrEMBL (unreviewed) contains protein sequences associated with computationally generated annotation and large-scale functional characterization.

UniProtKB/Swiss-Prot (reviewed) is a high quality manually annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusion

## Steps to browse, search & retrieve immunoglobulin sequence using UniProt database:

The screenshot shows the UniProt homepage with a dark blue header. In the top left, the UniProt logo is followed by links to BLAST, Align, Peptide search, ID mapping, and SPARQL. On the right, it shows "Release 2022\_03 | Statistics" and icons for download, email, and help. Below the header, a large white banner with the text "Find your protein" in bold. Underneath, there's a search bar with "UniProtKB" selected and a dropdown menu. To the right of the search bar are "Advanced", "List", and "Search" buttons. Below the search bar, examples of search terms like "Insulin, APP, Human, P05067, organism\_id:9606" are listed. On the far right, there are "Feedback" and "Help" buttons. A note at the bottom left of the page says: "UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)".

### STEP 1: Open homepage of UniProt database

This screenshot is identical to the one above, but the search bar now contains the text "IgE". The rest of the interface, including the header, banner, and footer message, remains the same.

### STEP 2: Entry any immunoglobulin to be searched

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB IgE Advanced List Search Help

Status  
 Reviewed (Swiss-Prot) (2,251)  
 Unreviewed (TrEMBL) (457,696)

Popular organisms  
 Human (541)  
 Mouse (435)  
 Zebrafish (359)  
 Rat (328)  
 Bovine (270)

Taxonomy  
[Filter by taxonomy](#)

## UniProtKB 459,947 results

or search "IgE" as a Gene Ontology, Protein Name, Disease, Author, or Gene Name

BLAST Align Map IDs Download Add View: Cards Table Customize columns Share Feedback Help

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
P0DOX4	IGE_HUMAN	Immunoglobulin epsilon heavy chain		Homo sapiens (Human)	547 AA
P12319	FCERA_HUMAN	High affinity immunoglobulin epsilon receptor subunit alpha	FCER1A, FCE1A	Homo sapiens (Human)	257 AA
P20693	FCER2_MOUSE	Low affinity immunoglobulin epsilon Fc receptor	Fcer2, Fcer2a	Mus musculus (Mouse)	331 AA

**STEP 3: Select any one hit whose sequence is be retrieved**

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced List Search Help

Function  
 Names & Taxonomy  
 Subcellular Location  
 Disease & Variants  
 PTM/Processing  
 Expression  
 Interaction  
 Structure  
 Family & Domains  
 Sequence  
 Similar Proteins

<https://rest.uniprot.org/uniprotkb/P0DOX4.fasta>

## P0DOX4 · IGE\_HUMAN

Immunoglobulin epsilon heavy chain · Homo sapiens (Human) · 547 amino acids · Evidence at protein level · Annotation score: 4/5

Entry Feature viewer Publications External links History

BLAST Align Download Add Add a publication Entry feedback

Function Text  
 FASTA (canonical)  
 FASTA (canonical & isoform)  
 JSON  
 XML  
 RDF/XML  
 GFF

The membrane-bound or secreted glycoproteins produced by B lymphocytes. In the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins play a role in humoral immunity, which results in the elimination of bound antigens.

The antigen-binding domain of one heavy chain, together with that of its associated light chain. Thus, it has remarkable affinity for a particular antigen. The variable domains are segment and can then be subjected to somatic hypermutations which, after exposure to a particular antigen (PubMed:20176268 PubMed:17576170) 3 Publications

**STEP 4: Click on download and FASTA(canonical)**

```
>p|PDDOX4|IGE_HUMAN Immunoglobulin epsilon heavy chain OS=Homo sapiens OX=9606 PE=1 SV=1
QVQLVQSGAEVRKPGASVRSCKASGYTFIDSYVGWIRQAPGHLWNIHWINPNSGGTNY
APRFQQRVTHTRDASFSTAYMDLRLSRLRSDDSAVFYCAKSDPFWSDYNFDYSSSEEGTEVT
YTFSGAWNTLPSVPLTRCKNIPSNATSVTGLCLATGYFPEPVMTWDTGSLNGTTLPAT
TLLSGHYATISLLTVSGAWAKQMFCTRVVAHTPSSTVDNKNTFSCSRDFTPPTVKILOQS
CDGLGHFPPTIQQLCLVSGYTPGTINITWLGGQVMDVDSLSTATESQGELASTESQLTLS
QKHWILSDRTYTCQVTYQGHTFDSTKKCADSNPRGVSAYLSPSPFDLFIRKSPTIITCLV
TDPHLPRALMRSTTKTSGRPAPEVYAFATPEWPGSRDKRTLACLIQNFMPEDISVQNLH
NEVQLPDARHSTTQPRTKGSGFFVFSRLEVTRAEWQEKEDEFICRAVHEAASPSQTQRA
VSVPNGK
```

**FIG 1. FASTA sequence for immunoglobulin ‘IgE’**

## **RESULT:**

Browsing, searching and retrieval of immunoglobulin sequence from UniProt database was demonstrated.

## **CONCLUSION:**

UniProt is a long-standing collection of databases that enable scientists to navigate the vast amount of sequence and functional information available for proteins. It includes amino acid sequence, protein name or description, taxonomic data and citation information, biological ontologies, classifications and cross-references, and clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data.

## **REFERENCE:**

- UniProt. (2019). Uniprot.org. <https://www.uniprot.org/>

- UniProtKB. (2019). Uniprot.org. <https://www.uniprot.org/help/uniprotkb>
- UniProt: the universal protein knowledgebase. (2016). *Nucleic Acids Research*, 45(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>

**WEBLEM 1b****PDB Database**

(URL:

<https://www.rcsb.org/> )**AIM:**

To browse, search & retrieve immunoglobulin structure using PDB database.

**INTRODUCTION:**

RCSB PDB (RCSB.org) is the US data center for the global Protein Data Bank (PDB) archive of 3D structure data for large biological molecules (proteins, DNA, and RNA) essential for research and education in fundamental biology, health, energy, and biotechnology.

The Protein Data Bank (PDB) was established as the 1st open access digital data resource in all of biology and medicine (Historical Timeline). It is today a leading global resource for experimental data central to scientific discovery. Through an internet information portal and downloadable data archive, PDB provides access to 3D structure data for the molecules of life, found in all organisms on the planet.

Knowing the 3D structure of a biological macromolecule is essential for understanding its role in human and animal health and disease, its function in plants and food and energy production, and its importance to other topics related to global prosperity and sustainability.

The enormous wealth of 3D structure data stored in the PDB has underpinned significant advances in our understanding of protein architecture, culminating in recent breakthroughs in protein structure prediction accelerated by artificial intelligence approaches and deep or machine learning methods.

RCSB PDB (Research Collaboratory for Structural Bioinformatics PDB) operates the US data center for the global PDB archive, and makes PDB data available at no charge to all data consumers without limitations on usage.

**Steps to browse, search & retrieve immunoglobulin structure using PDB database:**

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

**PDB** PROTEIN DATA BANK 196,108 Structures from the PDB 1,000,361 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Advanced Search | Browse Annotations Include CSM Help

PDB-101 EMDDataResource MOLBASE WWPDB Foundation

NEW! Computed Structure Models (CSM) Learn more

Welcome Deposit Search Visualize Analyze

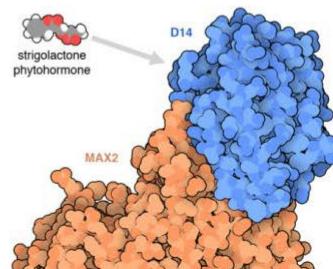
RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive
- Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

COVID-19 Join the

October Molecule of the Month



## STEP 1: Open homepage of PDB database



196,108 Structures from the PDB

1,000,361 Computed Structure Models (CSM)

▼ 3D Structures ?

IgE

Include CSM  Help

## NEW! Computed Structure Models

## Welcome

## Deposit

## Search

## Visualize

## Analyze

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for visualization, and analysis of:

Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive

Computed Structure Models (CSM) from Alpha ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

COVID-19

In Additional Structure Keywords

IMMUNOGLOBULIN FOLD, GLYCOPROTEIN, RECEPTOR, IgE-BINDING PROTEIN, IgE ANTIBODY, IgE-FC, IMMUNE SYSTEM

EFab, IgE, IMMUNE SYSTEM

IgE Fc, Immunoglobulin E, IMMUNE SYSTEM

Antibody, IgE/Fab fragment, IMMUNE SYSTEM

Immunoglobulin E, Antibody, IgE, immune system

immunoglobulin E, IgE, antibody, immune system

IgE-Fc, nanobody, VH, allergy, Immune system

IgE, DARPin, Allergy, Inhibitor, IMMUNE SYSTEM

IgE, allergy, Xolair, Omalizumab, IMMUNE SYSTEM

GALECTIN, GALAPTIN, LECTIN, IgE-BINDING PROTEIN

In Structure Title

IgE-Fc in complex with HAE



## STEP 2: Search for immunoglobulin of interest

Chemical Similarity ?

Return Structures grouped by No Grouping ?

Include Computed Structure Models (CSM)  Count Clear

Search Summary This query matches 601 Structures.

Refinements ?

Structure Determination Methodology  experimental (601)

Scientific Name of Source Organism  Homo sapiens (405)  Mus musculus (49)  synthetic construct (35)  Betula pendula (19)  Human immunodeficiency virus 1 (11)  Dermatophagoides pteronyssinus (10)  Hevea brasiliensis (9)  Actinidia deliciosa (8)  Phleum pratense (8)  Arachis hypogaea (7)  More...

1 to 25 of 601 Structures  Page 1 of 25  Sort by Score

**1JBF** Hairpin Peptide that Inhibits IgE Activity by Binding to the High Affinity IgE Receptor Nakamura, G.R., Starovasnik, M.A., Reynolds, M.E., Lowman, H.B. (2001) Biochemistry 40: 9828-9835

Released 2001-08-22  
Method SOLUTION NMR  
Macromolecule IGE06 (protein)

**7SD2** Murine Fab that recognizes Hev b 8 (profilin for Hevea brasiliensis)

## STEP 3: Out of all hits select immunoglobulin structure of interest

The screenshot shows the RCSB PDB homepage with the search term "1JBF" entered. The results page for entry 1JBF is displayed, featuring a 3D ribbon model of the protein structure. Key information shown includes:

- PDB ID:** 1JBF
- Classification:** PROTEIN BINDING
- Mutation(s):** No
- Deposited:** 2001-06-04 **Released:** 2001-08-22
- Deposition Author(s):** Nakamura, G.R., Starovasnik, M.A., Reynolds, M.E., I
- Experimental Data Snapshot**
- Download Options:** FASTA Sequence, PDBx/mmCIF Format, PDBx/mmCIF Format (gz), PDB Format, PDB Format (gz), PDBML/XML Format (gz)

#### STEP 4: Click on download and then PDB format

#### RESULT:

Browsing, searching and retrieval of immunoglobulin structure from PDB database was demonstrated.

#### CONCLUSION:

Protein Data Bank (PDB) archive of 3D structure data for large biological molecules (proteins, DNA, and RNA) essential for research and education in fundamental biology, health, energy, and biotechnology. Knowing the 3D structure of a biological macromolecule is essential for understanding its role in human and animal health and disease, its function in plants and food and energy production, and its importance to other topics related to global prosperity and sustainability.

#### REFERENCES:

1. *RCSB PDB: Homepage.* (2010). Rcsb.org. <https://www.rcsb.org/>
2. Bank, R. P. D. (n.d.). *RCSB PDB: About RCSB PDB: Enabling Breakthroughs in Scientific and Biomedical Research and Education.* Wwww.rcsb.org. <https://www.rcsb.org/pages/about-us/index>

## WEBLEM 2

### Introduction to Antibody sequence and structure along with Antibody numbering methods such as Kabat, Chothia and other equivalent methods and its importance

Monoclonal antibodies are playing an increasing role in both human and animal health. Different strategies of protein and chemical engineering, including humanization techniques of non-human antibodies were applied successfully to optimize clinical performances of antibodies. Despite the emergence of techniques allowing the development of fully human antibodies such as transgenic Xeno-mice, antibody humanization remains a standard procedure for therapeutic antibodies. An important prerequisite for antibody humanization requires standardized numbering methods to define precisely complementary determining regions (CDR), frameworks and residues from the light and heavy chains that affect the binding affinity and/or specificity of the antibody- antigen interaction. The recently generated deep-sequencing data and the increasing number of solved three- dimensional structures of antibodies from human and non-human origins have led to the emergence of numerous databases. However, these different databases use different numbering conventions and CDR definitions. In addition, the large fluctuation of the variable chain lengths, especially in CDR3 of heavy chains (CDRH3), hardly complicates the comparison and analysis of antibody sequences and the identification of the antigen binding residues.

### Numbering Schemes of Antibody Variable Domains

Antibody engineering methods require precise identification of the residues that have an impact on the interaction and/or affinity of the antibody for its target antigen. For example, as mentioned above, CDR- grafting aims to decrease the immunogenicity of non-human antibodies by engineering the variable regions directed against the target antigen. This method requires an accurate identification of the CDRs and therefore an adequate alignment of antibody sequences from human and non-human species. It has been shown that residues from the framework regions might also exert a strong impact on the antibody affinity. Thus, the precise identification of corresponding positions in human and animal immunoglobulin chains is essential. However, the use of different amino acid numbering schemes currently available in the literature is confusing and might lead to aberrant identification of framework and CDR residues. Therefore, it is of crucial importance to understand the different numbering schemes and, consequently, being able to compare them.

### Kabat numbering scheme

Over the past decades, sequencing and crystallization of antibodies resulted in significant increase of various sequence and structure databases, which made the comparison of the variable regions from human and animal immunoglobulins possible. In 1970, Kabat and Wu aligned 77 Bence-Jones protein and immunoglobulin light chain sequences in order to study the statistical variability in amino acid composition at the sequential positions of the variable antibody regions. They defined the “variability parameter” as the number of different amino acids at a given position divided by the frequency of the most occurring amino acid at that position. This analysis revealed three hypervariable regions in the variable region of the light chains. The presence of highly conserved residues was also demonstrated, such as the two cysteines that form a disulphide bridge at the inner core of the immunoglobulin domain and a tryptophan residue located immediately after CDRL1. Likewise, three corresponding hypervariable regions were also identified in the variable heavy chain domain. Kabat and Wu postulated that these hypervariable regions would cluster at one side of the folded domain to form a surface responsible for specific antigen recognition and referred to these hypervariable regions as “Complementarity Determining Regions” “CDR”-1,-2, and-3. This hypothesis was later confirmed and further investigated to distinguish antigen-contacting or conformational important residues within these CDRs.

In 1979, Kabat et al. were the first to propose a standardized numbering scheme for the variable regions of immunoglobulins. In their compilation of “Sequences of Proteins of Immunological Interest”, the amino acid sequences of the variable region of the light ( $\lambda$ ,  $\kappa$ ) and heavy chain of antibodies, as well as the variable region

of T cell receptors ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) were aligned and numbered. They observed that the analyzed sequences exhibited variable lengths and that gaps and insertions could only be included at precise positions. Interestingly, the points of insertion were located inside the CDRs, except for CDRL2, but also at some positions inside the framework regions. In the numbering schemes, these insertions are identified and annotated with letters (e.g., 27a, 27b...). It is also noticeable that residue L10 is absent in all the  $\lambda$  light chains, while  $\lambda$  and  $\kappa$  chains are being coded by two different genes, located on different chromosomes. Over the last decades, the accumulation of sequences resulted in the creation of the KABATMAN database.

Although the Kabat numbering scheme is often considered as the standard that is widely adopted for numbering antibody residues, it has some important limitations. Firstly, this scheme was built on the alignments of a limited number of sequences from antibodies with the most common sequence lengths. Consequently, sequences with unconventional insertions or deletions in the CDRs or in the framework regions were not included. Therefore, the original Kabat scheme ignores antibody chains of unconventional lengths, with unique insertions or deletions. However, a useful numbering tool named *ABnum* that numbers the amino acid sequences of variable domains according to a much larger and regularly updated database (Abysis), takes into account insertions of variable lengths, particularly in CDR2 by adding an insertion point at position L54. The second main limitation of the Kabat scheme is that it doesn't match very well with the 3D structure of antibodies. Indeed, the hypervariable regions defined by Kabat do not exactly match with the structural antigen-binding loops. The defined insertion points in CDR-L1 (L27) and CDR-H1 (H35) do not fit with their corresponding positions in the structures. In other words, the corresponding residues (topologically aligned) in crystal structures in CDR-L1 and CDR-H1 don't share the same number in the Kabat numbering scheme.

### Chothia numbering scheme

In 1987, Chothia and Lesk introduced a structure-based numbering scheme for antibody variable regions. They aligned crystal structures of antibody variable regions, defined the loop structures that form the CDRs and corrected the position numbers of the insertion points inside CDRL1 and CDRH1 so that they better fit their topological positions. Furthermore, they classified the CDR loops of heavy and light chains in a small number of conserved structures, called “canonical” classes.

Based on the alignment of antibody structures, the Chothia numbering scheme shifts the point of amino acid insertion from position L27 to L30 and from position H35 to H31. It is worth mentioning that the Chothia CDR definition ensures a better correspondence to the structural loops. The loop structure of CDRH3 identified by Chothia matches well the Kabat hypervariable region. In contrast, the other loops are shorter than the hypervariable sequences defined by Kabat, except for CDRH1 which

extends from H26 to H32. In any case, the CDRs defined on the hypervariable amino acids according to Kabat and based on loop topology in Chothia's nomenclature have for some CDR's a shifted location and/or comprise deviating loop lengths.

The Chothia numbering scheme possesses the main advantage that topologically aligned residues from different antibodies are localized at the same position number and that the Chothia CDR definition corresponds in most antibody sequences to the structural antigen-binding loops. However, confusion can also arise given the limited use of this numbering scheme compared to the Kabat or the IMGT numbering schemes (see below). Furthermore, a later study published by Chothia et al. changed the insertion point in CDR L1 from residue L30 to L31. However, while investigating the conformation of the antigen-binding loops, of antibodies present in larger databases, they returned to the initial L30 position in 1997. In a similar way, they initially defined an insertion point at position L93 in  $\lambda$  light chains that was shifted to position L95 in their subsequent study. Finally, an important limitation of this numbering scheme is due to the use of the most common CDR sequence lengths, like the Kabat numbering scheme, and therefore the Chothia scheme ignores sequences with unconventional length. However, similarly to the Kabat numbering scheme, this system could be optimized by defining new insertion points.

For numbering antibody KabatMan database can be used.

## Kabatman Database:

### Kabatman database provides various tools for antibody informatics:

- abYsis - Integrated database and analysis workbench
- abYmod - Antibody modelling software
- abYbank - Antibody sequence and structure data. This includes:
  - AbDb - pre-numbered structures from the PDB
  - SACS - Summary of Antibody Crystal Structures
  - EMBL Ig - antibody sequences from EMBL-ENA
  - Kabat - FASTA formatted sequences from Kabat (2000)
  - AbPDBSeq - FASTA formatted antibody sequences from the PDB
- Humanness (G)- Assess humanness against expressed sequences in Kabat divided into germline families
- PAPS - Predict VH/VL packing angle
- abYdraw - A downloadable software package for drawing antibody cartoons

### The following are included within abYsis, but also available as standalone tools:

- KabatMan: Query the Kabat sequence data
- AbCheck: Test a sequence against the Kabat data for unusual residues
- Chothia canonicals: Identify canonical classes for CDRs from your sequence
- Human subgroups: Assign the human subgroup for your sequence
- Humanness (H): Assess humanness against expressed sequences in Kabat
- Abnum: Apply standard numbering to sequences or structures

[Home](#) > Antibodies

## Antibodies

### Introduction

We provide a large number of tools for antibody informatics:

abYsis	Integrated database and analysis workbench
abYmod	Antibody modelling software
abYbank	Antibody sequence and structure data. This includes: ▪ AbDb - pre-numbered structures from the PDB ▪ SACS - Summary of Antibody Crystal Structures ▪ EMBL Ig - antibody sequences from EMBL-ENA ▪ Kabat - FASTA formatted sequences from Kabat (2000) ▪ AbPDBSeq - FASTA formatted antibody sequences from the PDB
Humanness (G)	Assess humanness against expressed sequences in Kabat divided into germline families

### Information

Our pages of information on antibodies:

- The Kabat Numbering Scheme
- The Chothia Numbering Scheme
- The Martin (Enhanced Chothia) Numbering Scheme
- Table of CDR Definitions
- How to identify the CDRs by looking at a sequence
- Table of mean contact data
- Further information on numbering
- Antibody humanization patents

Information and links to accompany my book chapter Protein Sequence and Structure Analysis of Antibody Variable Domains. In: Antibody Engineering Lab Manual (Ed.: Duebel, S. and Kontermann, R., Springer-Verlag, Heidelberg). [[Information](#)] [[Purchase](#)]

The famous Kabat hook is now available online as a scanned copy via Google Books: [Elvin](#)



**Fig1. Different tools available for Antibodies studies**

## KabatMan:

KabatMan interface lets you create simple queries for more complex cases, you must write queries directly using the in the KabatMan SQL-like query language. The KabatMan interface is shown below:

The screenshot shows the KabatMan homepage. At the top, there's a dark header bar with the URL 'www.bioinf.org.uk' on the left and the UCL logo on the right. Below the header, a navigation bar includes links for 'Prof Andrew C R Martin's Group', 'Home', 'Antibodies' (which is highlighted in red), 'Mutations', 'Other servers', 'Software', and 'Information'. A breadcrumb trail 'Home > Antibodies > KabatMan' is visible above the main content area. The main title 'KabatMan' is centered above a sub-header 'Simple Interface to the Kabat Sequence Database'. Below this, a text block explains the interface: 'This page provides a simple point-and-click interface to the KabatMan database. This interface only lets you create simple queries; for more complex cases, you must write queries directly using the in the KabatMan SQL-like query language.' Three red buttons at the bottom offer links to 'Full query page', 'Query language', and 'Statistics'. A 'News...' link is also present.

**Fig2. Simple Interface to the Kabat Sequence**

## Database KabatMan - Full queries

This page allows you to query the Kabat antibody sequence database using an SQL-like query language. The main deviation between the language used here and SQL is that clauses within the WHERE statement are combined in reverse polish notation. Also, since there is only one table in the database, there is no FROM statement.

The query language uses three statements: SET, SELECT and WHERE. The SELECT and WHERE commands take you into a mode where clauses are specified with no introductory SELECT or WHERE keyword. Conversely, the SET command needs to be specified on each line where variables are to be set and must be given *before* any SELECT or WHERE statement. The QUIT or EXIT command is used to leave the program.

[Home](#) > [Antibodies](#) > [KabatMan](#) > Full queries

## KabatMan - Full queries

Query the Kabat sequence database - Full query language

This page allows you to query the Kabat antibody sequence database using an SQL-like query language. The main deviation between the language used here and SQL is that clauses within the WHERE statement are combined in reverse polish notation. Also, since there is only one table in the database, there is no FROM statement.

Unless you are asking complex queries, you will probably find the [simple point-and-click interface](#) to KabatMan easier!. This only allows you to ask fairly simple queries, but covers the majority of types of queries which people ask. To ask more complex queries, you must use the [KabatMan Query Language](#) directly using the text entry box below.

### News...

#### 05.12.11 Data reading

A small number of sequences with insertions (e.g. 18D-6 light chain, ID 007779) were not being read correctly as the Kabat format was inconsistent. This is now fixed.

**Fig3. KabatMan – Full Query language interface**

# READ THE INSTRUCTIONS FIRST!!!

Or use the simple point-and-click interface.

Please do not just enter a sequence!

Enter your query to KabatMan here:

```
SELECT name, l1  
WHERE len(l1) eq 11 res(L29) eq P AND
```

To submit your query, press here:

To clear the form, press here:

Please cite the following reference in any publications resulting from searches using this

**Fig4. Example of query to find all antibodies with 11 residue CDR-L1s and a proline at the sixth position**

## KabatMan Query Results

KabatMan V2.26  
=====  
Copyright (c) 1994-2019, Dr. Andrew C.R. Martin / University College London / University of Reading.  
This program is copyright. Any copying without the permission of the  
author is prohibited.

### The query was:

```
SELECT name, l1  
WHERE len(l1) eq 11 res(L29) eq P AND
```

### Results were:

```
HIL, SANALPNQYAY  
BEN-27'CL, SGDALPNQYAY  
ITC63B'CL, SGDALPKQYSY  
HCV-65'CL, SGDALPKKYAY  
9604'CL, SGDALKRYAY  
WR1.187'CL, SGDALKQYAH  
WR1.112'CL, SGDALKQYAH  
ITC63B'CL, SGDALKQYSY  
KIR, SGDALKNOYAY  
CAP, SGDALPAEYAY  
1B8'CL, SGDALPQQFAY  
HAN, SGDALKQYAH  
GAR, SGDVLPLKKYAY  
V1 HEP1'CL, SGDALKQYAY
```

**Fig5. Results for the query**

The above results show all antibodies with 11 residue CDR-L1s and a proline at the sixth position.

Thus, KabatMan database is using for querying Kabat sequence data. This information is useful for antibody engineering which requires precise identification of the residues that have an impact on the interaction and/or affinity of the antibody for its target antigen. For example, CDR-grafting aims to decrease the immunogenicity of non-human antibodies by engineering the variable regions directed against the target antigen. This method requires an accurate identification of the CDRs and therefore an adequate alignment of antibody sequences from

human and non-human species. Moreover, it has been shown that residues from the framework regions might also exert a strong impact on the antibody affinity. Thus, the precise identification of corresponding positions in human and animal immunoglobulin chains is essential.

## REFERENCES:

1. *bioinf.org.uk - Prof. Andrew C.R. Martin's group at UCL.* (n.d.).  
Www.bioinf.org.uk. Retrieved October 10, 2022, from  
<http://www.bioinf.org.uk/abs/>
2. Dondelinger, M., Filée, P., Sauvage, E., Quinting, B., Muyltermans, S., Galleni, M., & Vandevenne, M. S. (2018). Understanding the Significance and Implications of Antibody Numbering and Antigen- Binding Surface/Residue Definition. *Frontiers in Immunology*, 9. <https://doi.org/10.3389/fimmu.2018.02278>

## 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:

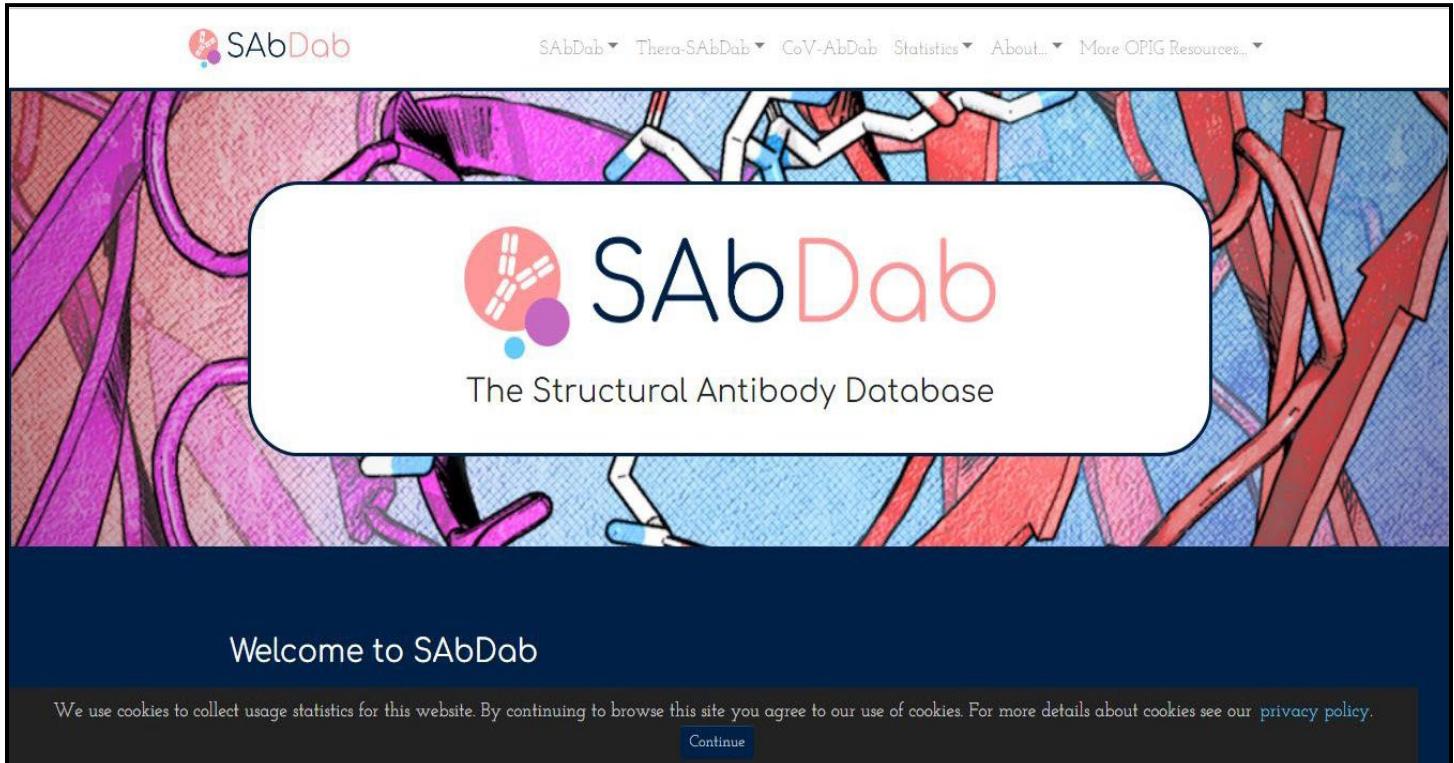
pdb	Hchain	Lchain	model	antigen_chain	antigen_type	...
1ahw	B	A	0	C	protein	...
1ahw	E	D	0	F	protein	...

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**

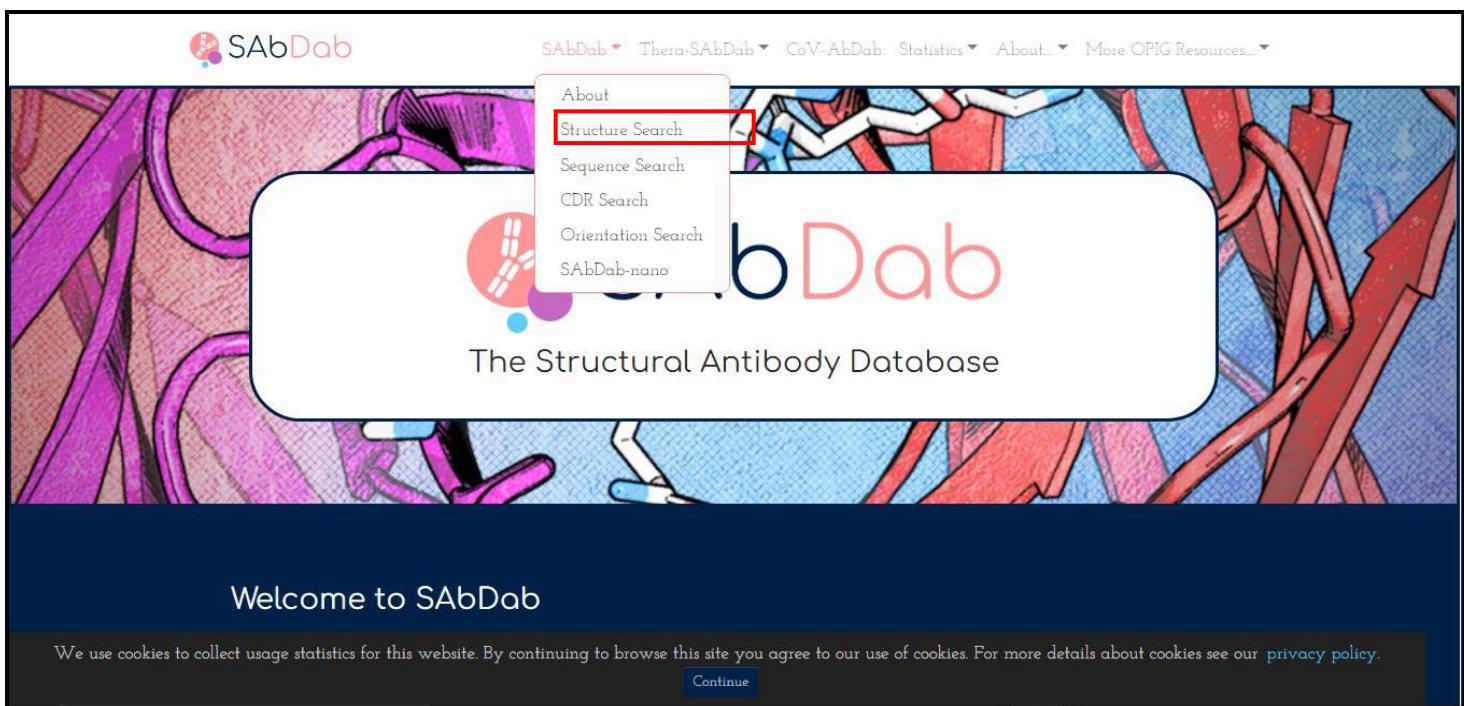


Fig2: Search options under SAbDab database

This screenshot shows the 'Search Structures' page. The main title 'Search Structures' is at the top. Below it, there is a 'Search' button and a 'About' link. The 'About' link is followed by a list of bullet points explaining the search capabilities:

- Find antibody structures that have been deposited in the PDB.
- Use the 'Get all structures' tab to get a list of all antibodies in SAbDab.
- You can search for a specific entry using its PDB code.
- Search for a subset of antibodies by attribute - such as species, experimental method, resolution, residue at a particular position etc.
- Create a non-redundant set of antibody structures by sequence, with specified maximum sequence identity and structure quality cutoffs. Clustering is performed using CD-HIT.
- For more help, see the [About](#) page.

Below this list are five additional links:

- > Get all structures
- > Search for a specific PDB entry
- > Search structures by attribute
- > Search for a non-redundant set of antibodies
- > Create email alert for new additions to SAbDab

**Fig3: Different search options available under Search Structures section**

The screenshot shows the SAbDab website interface. At the top, there is a navigation bar with links: SAbDab, Thera-SAbDab, CoV-AbDab, Statistics, About, and More OPIG Resources. Below the navigation bar is a large, colorful illustration of antibody structures. A white callout box is positioned in the upper right area of the illustration, containing the text "About" and "Search". A dark blue rectangular overlay covers the central part of the illustration. Inside this overlay, the text "About SAbDab" is displayed in white. Below the main content area, there is a sidebar with two buttons: "About SAbDab" (which is highlighted with a light blue background) and "Terminology".

**Fig4:** Search options under Thera-SAbDab database

The screenshot shows the SAbDab website interface. At the top, there is a navigation bar with links: SAbDab, Thera-SAbDab, CoV-AbDab, Statistics, About, and More OPIG Resources. Below the navigation bar is a large, colorful illustration of antibody structures. A white callout box is positioned in the upper right area of the illustration, containing the text "SAbDab" and "SAbDab-nano". A dark blue rectangular overlay covers the central part of the illustration. Inside this overlay, the text "About SAbDab" is displayed in white. Below the main content area, there is a sidebar with two buttons: "About SAbDab" (which is highlighted with a light blue background) and "Terminology".

**Fig5:** Search options under Statistics section

The screenshot shows the SAbDab website's "About SAbDab" page. At the top, there is a navigation bar with links: SAbDab, Thera-SAbDab, CoV-AbDab, Statistics, About..., and More OPIG Resources... A dropdown menu is open over the "About..." link, showing options: SAbDab and Thera-SAbDab. Below the navigation is a large, stylized illustration of antibody structures in various colors (pink, blue, red). A dark blue callout box contains the text "About SAbDab". On the left, there is a sidebar with two buttons: "About SAbDab" (highlighted in light blue) and "Terminology". The main content area contains a brief description of SAbDab and its features.

About SAbDab

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.

Fig6: Search options under About section

The screenshot shows the SAbDab website's "About SAbDab" page. At the top, there is a navigation bar with links: SAbDab, Thera-SAbDab, CoV-AbDab, Statistics, About..., and More OPIG Resources... A dropdown menu is open over the "More OPIG Resources..." link, showing options: SAbPred, OAS Database, and STCRDab. Below the navigation is a large, stylized illustration of antibody structures in various colors (pink, blue, red). A dark blue callout box contains the text "About SAbDab". On the left, there is a sidebar with two buttons: "About SAbDab" (highlighted in light blue) and "Terminology". The main content area contains a brief description of SAbDab and its features.

About SAbDab

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.

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.

**ExPasy**  ABCD Home | Contact

**The ABCD (AntiBodies Chemically Defined) Database**

The ABCD (AntiBodies Chemically Defined) database is a manually curated depository of **sequenced antibodies**, developed by the Geneva Antibody Facility at the University of Geneva, in collaboration with the CALIPHO and Swiss-Prot groups at SIB Swiss Institute of Bioinformatics.

Search by antibody name, species or target ([UniProt](#) or [ChEBI](#) ID)

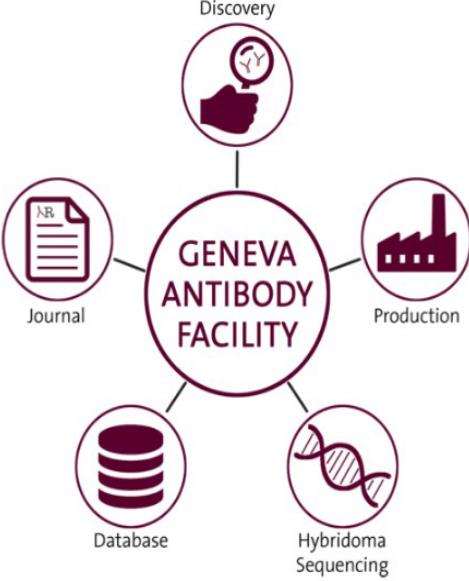
Example searches: 9E10, P07766, 37926, Escherichia coli, Protein tag, Nanobody

The ABCD database is part of a broader project, with the mission of promoting the widespread use of **recombinant antibodies** by academic researchers and, ultimately, the replacement of animal-produced antibodies. This concerted effort also includes the Geneva Antibody Facility (for discovery and production of antibodies) and the scientific journal *Antibody Reports* (publishing technical articles on antibody characterization).

Release information: Release information: Version 12.0 (May 2022)  
23'457 sequenced antibodies, against 4'125 different targets

If you'd like to cite the ABCD database: Lima WC, Gasteiger E, Marcatili P, Duek P, Bairoch A, Cosson P. The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Res.* 2020; 48:D261-D264. doi: 10.1093/nar/gkz714

About us  
Frequently asked questions (FAQ)  
Submit a new Antibody  
Antibodies to Protein tags and Subcellular markers  
 [Coronavirus Resources page](#)



The diagram illustrates the Geneva Antibody Facility's integrated workflow. At the center is a large circle labeled "GENEVA ANTIBODY FACILITY". Surrounding it are four interconnected components: "Discovery" (top), "Production" (right), "Database" (bottom-left), and "Hybridoma Sequencing" (bottom-right). Each component is represented by a circular icon with a specific symbol: a magnifying glass over a hand for Discovery, a factory building for Production, a database server for Database, and a DNA double helix for Hybridoma Sequencing. Lines connect the central facility to each of these four components.

**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.

<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
-

## 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:

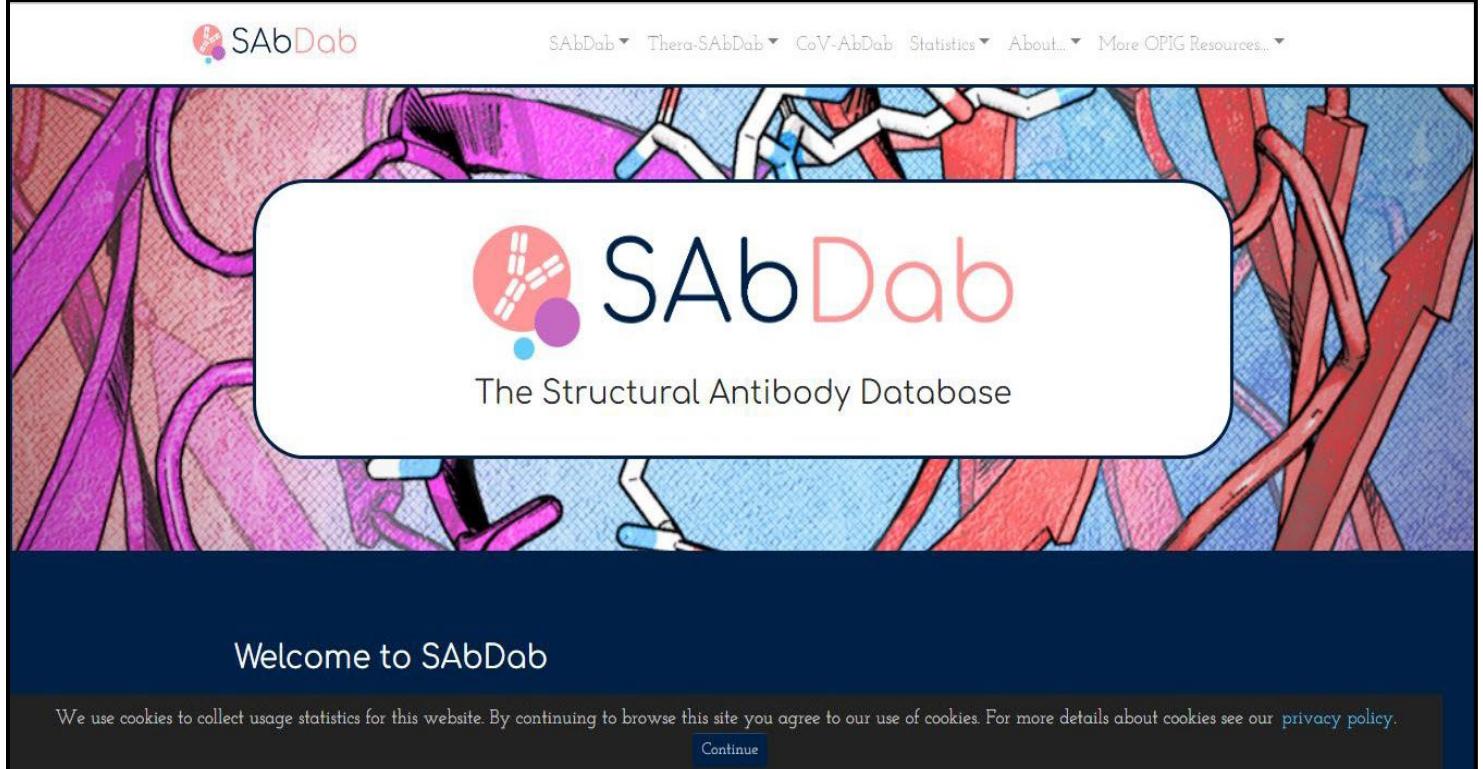


Fig1: Homepage for SAbDab database

# Search Structures

Search

> About

- Find antibody structures that have been deposited in the PDB.
- Use the 'Get all structures' tab to get a list of all antibodies in SAbDab.
- You can search for a specific entry using its PDB code.
- Search for a subset of antibodies by attribute - such as species, experimental method, resolution, residue at a particular position etc.
- Create a non-redundant set of antibody structures by sequence, with specified maximum sequence identity and structure quality cutoffs. Clustering is performed using CD-HIT.
- For more help, see the [About](#) page.

> Get all structures

> Search for a specific PDB entry

> Search structures by attribute

> Search for a non-redundant set of antibodies

> Create email alert for new additions to SAbDab

**Fig2: Different search options available under Search Structures section**

Search

- Create a non-redundant set of antibody structures by sequence, with specified maximum sequence identity and structure quality cutoffs. Clustering is performed using CD-HIT.
- For more help, see the [About](#) page.

> Get all structures

> **Search for a specific PDB entry**

Please enter a PDB code:

4NP4

Get structure

> Search structures by attribute

> Search for a non-redundant set of antibodies

> Create email alert for new additions to SAbDab

**Fig3: Search for 4NP4 PDB query**

# Search Structures

View results

Downloads

Search

## Search results

1 structure(s) fit your criteria. Click on the PDB code to view the structure.

PDB	Species	Method	Resolution	Chain Pairings	Antigens	Downloads
4np4	HOMO SAPIENS	X-RAY DIFFRACTION	2.89 Å	Fv no. 1 VH: H VL: L Fv no. 2 VH: I VL: M	protein	<ul style="list-style-type: none"><li>◦ Structure (as PDB)</li><li>◦ Structure (Chothia)</li><li>◦ Structure (IMGT)</li><li>◦ Summary file</li></ul>

**Fig4: Results for PDB entry 4NP4**

## > Structure details

Details

Visualisation

Fvs

Data in other  
OPIG databases

Downloads

PDB ^

Clostridium Difficile toxin B Crop Domain in complex with Fab Domains of Neutralizing antibody Bezlotoxumab

PDB	4np4
Species	HOMO SAPIENS
Method	X-RAY DIFFRACTION
Resolution	2.89Å
Number of Fvs	2
In complex	True
Light chain type	Kappa
Has constant region	True
Affinity	1.9e-11 M (Method: SPR)

Fig5: Structural information for PDB ID: 4NP4

## > Structure visualisation

Details

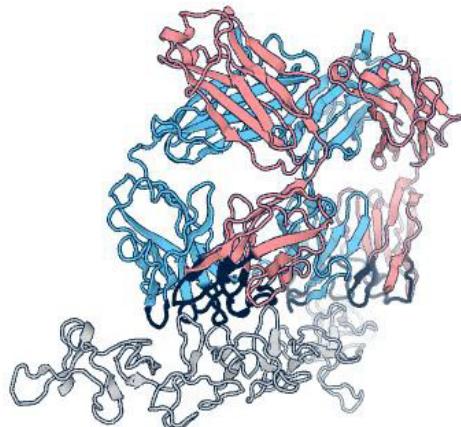
Visualisation

Fvs

Data in other OPIG  
databases

Downloads

PDB ^



Key (Default Scheme):

VH Chains  
VL Chains  
CDRs  
Antigen Chains

Display options:

- Spacefill
- Wire
- Ball&stick
- Cartoon
  
- Default colours
- Colour by B-factor
- Colour by chain
- Colour by sec. structure
- Colour by element
  
- Spin on/off

**FIG6: Structural visualization for PDB ID: 4NP4**

> Fv information

- Details
- Visualisation
- Fvs
- Data in other OPIG databases

This PDB has 2 Fv(s).

- > H/L
- > I/M

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

This PDB has 2 Fv(s).

- Details
- Visualisation
- Fvs
- Data in other OPIG databases
- Downloads
- PDB ^

> H/L

Fv Details	
Heavy chain	H
Light chain	L
Heavy subgroup	IGHV5
Light subgroup	IGKV3
Species	HOMO SAPIENS
In complex?	True
scFv?	False
Has constant domain?	True

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

Numbered Sequences (chothia)

Heavy chain

1	2	3	4	5	6	7	8	9	10	11	12	13
E	V	Q	L	V	Q	S	G	A	E	V	K	K

Light chain

1	2	3	4	5	6	7	8	9	10	11	12	13
E	I	V	L	T	Q	S	P	G	T	L	S	L

Details

Visualisation

Fvs

Data in other OPIG databases

Downloads

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

Antigen Details

Antigen chains A

Antigen type protein

Antigen name toxin b

Antigen species CLOSTRIDIUM DIFFICILE

Antigen sequence

```
MGLIYINDSLYYFKPPVNNLITGFVTVGDDKYYFNPI  
NGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFK  
YFAPANTLDENLEGEAIDFTGKLIDENIYYFDDNYR  
GAVEWKELDGEHYFSPETGKAFKGLNQIGDYKYYFN  
SDGVMQKGTVSINDNKHYFDDSGVMKVGYTEIDGKHF  
YFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEI  
SYSGILNFNNKIYYFDDSFATAVVGWKDLEDGSKYYFD  
EDTAEAYILEHHHHHH
```

Details

Visualisation

Fvs

Data in other OPIG databases

Downloads

PDB ↗

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

	CDR Sequences (chothia definition)
CDRH1	GYSFTSY
CDRH2	YPGDSS
CDRH3	RRNWGNAFDI
CDRL1	RASQSVSSSYLA
CDRL2	GASSRAT
CDRL3	QQYGSSTWT

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

The screenshot shows a sidebar with the following menu items:

- Details
- Visualisation
- Fvs
- Data in other OPIG databases
- Downloads
- PDB ↗

The main content area displays a table titled "Orientation Angles (from ABangle)" with the following data:

	Orientation Angles (from ABangle)
HL	-56.49°
HC1	69.78°
HC2	112.73°
LC1	124.71°
LC2	85.25°
dc	16.41 Å

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

The sidebar menu is identical to Fig8e:

- Details
- Visualisation
- Fvs
- Data in other OPIG databases
- Downloads
- PDB ↗

The main content area shows a section titled "I/M" with a sub-section "Fv Details". The data is presented in a table:

	Fv Details
Heavy chain	I
Light chain	M
Heavy subgroup	IGHV5
Light subgroup	IGKV3
Species	HOMO SAPIENS
In complex?	True
scFv?	False
Has constant domain?	True

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

Details

Visualisation

Fvs

Data in other OPIG databases

Downloads

PDB ↗

Numbered Sequences (chothia)

Heavy chain

1	2	3	4	5	6	7	8	9	10	11	12	13
E	V	Q	L	V	Q	S	G	A	E	V	K	K

Light chain

1	2	3	4	5	6	7	8	9	10	11	12	13
E	I	V	L	T	Q	S	P	G	T	L	S	L

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

Details

Visualisation

Fvs

Data in other OPIG databases

Downloads

PDB ↗

Antigen Details

Antigen chains	A
Antigen type	protein
Antigen name	toxin b
Antigen species	CLOSTRIDIUM DIFFICILE
Antigen sequence	MGLIYINDSLYYFKPPVNNLITGFVTVGDDKYYFNPI NGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFK YFAPANTLDENLEGEAIDFTGKLIDENIYYFDDNYR GAVEWKELDGEHYFSPETGKAFKGLNQIGDYKYYFN SDGVMQKGFVSINDNKHYFDDSGVMKVGYTEIDGKHF YFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEI SYSGILNFNNKIYYFDDSTAVVGWKDLEDGSKYYFD EDTAEAYILEHHHHHH

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

Details	
Visualisation	
Fvs	<b>CDR Sequences (chothia definition)</b>
Data in other OPIG databases	
Downloads	

**CDRH1** GYSFTSY  
**CDRH2** YPGDSS  
**CDRH3** RRNWGNNAFDI  
**CDRL1** RASQSVSSSYLA  
**CDRL2** GASSRAT  
**CDRL3** QQYGSSTWT

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

Details	
Visualisation	
Fvs	<b>Orientation Angles (from ABangle)</b>
Data in other OPIG databases	
Downloads	

**HL** -60.49°  
**HC1** 69.83°  
**HC2** 112.75°  
**LC1** 122.34°  
**LC2** 83.94°  
**dc** 16.43Å

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

> Occurences in auxilliary databases

Details	Occurrences of this structure in other OPIG databases.	
Visualisation	TheraSAbDab	<a href="#">Link</a>
Fvs	CoVAbDab	N/A
Data in other OPIG databases		

**Fig10: Results for Occurrences in auxiliary databases**

The screenshot shows a sidebar on the left with links: 'Visualisation', 'Fvs', 'Data in other OPIG databases', 'Downloads' (which is highlighted in blue), and 'PDB'. The main content area has a header 'Downloads' with a back arrow. Below it, a message says 'Additional links and files for download: see [help](#) for more details.' There are four download links: 'Chothia-numbered structure' (Download), 'IMGT-numbered structure' (Download), 'Non-annotated structure from the PDB' (Download), and 'Summary file for this antibody' (Download).

**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>
-

**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:

**Expsy ABCD**

The ABCD (AntiBodies Chemically Defined) Database

The ABCD (AntiBodies Chemically Defined) database is a manually curated depository of **sequenced antibodies**, developed by the Geneva Antibody Facility at the University of Geneva, in collaboration with the CALIPHO and Swiss-Prot groups at SIB Swiss Institute of Bioinformatics.

Search by antibody name, species or target ([UniProt](#) or [ChEBI](#) ID)

Example searches: 9E10, P07766, 37926, Escherichia coli, Protein tag, Nanobody

The ABCD database is part of a broader project, with the mission of promoting the widespread use of **recombinant antibodies** by academic researchers and, ultimately, the replacement of animal-produced antibodies. This concerted effort also includes the Geneva Antibody Facility (for discovery and production of antibodies) and the scientific journal *Antibody Reports* (publishing technical articles on antibody characterization).

Release information: Release information: Version 12.0 (May 2022)  
23'457 sequenced antibodies, against 4'125 different targets

If you'd like to cite the ABCD database: Lima WC, Gasteiger E, Marcatili P, Duek P, Bairoch A, Cosson P. The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Res.* 2020; 48:D261-D264. doi: 10.1093/nar/gkz714

About us  
Frequently asked questions (FAQ)  
Submit a new Antibody  
Antibodies to Protein tags and Subcellular markers  
[New](#) Coronavirus Resources page

Discovery

Production

Database

Hybridoma Sequencing

Fig1: Homepage for ABCD database

**Expsy ABCD**

The ABCD (AntiBodies Chemically Defined) Database

The ABCD (AntiBodies Chemically Defined) database is a manually curated depository of **sequenced antibodies**, developed by the Geneva Antibody Facility at the University of Geneva, in collaboration with the CALIPHO and Swiss-Prot groups at SIB Swiss Institute of Bioinformatics.

Search by antibody name, species or target ([UniProt](#) or [ChEBI](#) ID)

Example searches: 9E10, P07766, 37926, Escherichia coli, Protein tag, Nanobody

The ABCD database is part of a broader project, with the mission of promoting the widespread use of **recombinant antibodies** by academic researchers and, ultimately, the replacement of animal-produced antibodies. This concerted effort also includes the Geneva Antibody Facility (for discovery and production of antibodies) and the scientific journal *Antibody Reports* (publishing technical articles on antibody characterization).

Release information: Release information: Version 12.0 (May 2022)  
23'457 sequenced antibodies, against 4'125 different targets

If you'd like to cite the ABCD database: Lima WC, Gasteiger E, Marcatili P, Duek P, Bairoch A, Cosson P. The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Res.* 2020; 48:D261-D264. doi: 10.1093/nar/gkz714

About us  
Frequently asked questions (FAQ)  
Submit a new Antibody  
Antibodies to Protein tags and Subcellular markers  
[New](#) Coronavirus Resources page

Discovery

Production

Database

Hybridoma Sequencing

Fig2: Searching monoclonal antibody “Erenumab”

ABCD (AntiBodies Chemically Defined) Database result: 1 hit for erenumab

Identifier	Antibody name	Target	Organism
ABCD_AA791	erenumab	CALCRL, CGRPR, Calcitonin gene-related peptide typ...	Homo sapiens (Human)

**Fig3: Hit Page for “Erenumab”**

ABCD\_AA791 in the ABCD (AntiBodies Chemically Defined) Database

Antigen information	
Target type	Protein
Target link	UniProt: Q16602 Homo sapiens (Human)
Target name	CALCRL, CGRPR, Calcitonin gene-related peptide type 1 receptor, CGRP type 1 receptor, Calcitonin receptor-like receptor
Antibody information	
Antibody name	erenumab
Antibody synonyms	AMG 334, AMG-334
Applications	Surface plasmon resonance, Therapeutic, X-ray crystallography
Cross-references	PDB: 6UMG IMGT/mAb-DB: 618
Publications	PMID: 32049005 PMID: 26559125
Antibody sequence	
If you want to have the protein sequence of this antibody, please check the Publications and Cross-references links (a more comprehensive step-by-step guide on how to find sequences can be found <a href="#">here</a> ). If you have trouble finding it, just send us an email using the <a href="#">contact form</a> .	
Would you like to obtain this antibody?	
It can be produced at the <a href="#">Geneva Antibody facility</a> (for more information, please check <a href="#">here</a> ).	

**Fig4: Antigen and Antibody information for query “Erenumab”**

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search Help

**Q16602 · CALRL\_HUMAN**

Calcitonin gene-related peptide type 1 receptor · *Homo sapiens (Human)* · Gene: CALCRL (CGRPR) · 461 amino acids · Evidence at protein level · Annotation score: 5/5

Entry Feature viewer Publications External links History

BLAST Align Download Add Add a publication Entry feedback

Function Receptor for calcitonin-gene-related peptide (CGRP) together with RAMP1 and receptor for adrenomedullin together with RAMP3 (By similarity). Receptor for adrenomedullin together with RAMP2 (PubMed:22102369, PubMed:30115739). The activity of this receptor is mediated by G proteins which activate adenylyl cyclase (PubMed:22102369, PubMed:30115739). By Similarity 2 Publications

GO Annotations

Fig5: Results for antigen information in UniProt database

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

**PDB** PROTEIN DATA BANK 196,108 Structures from the PDB 1,000,361 Computed Structure Models (CSM) 3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Advanced Search | Browse Annotations Help

PDB-101 wwpdb EMDDataResource NUCLEIC ACID DATABASE wwPDB Foundation

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

Biological Assembly 1 6UMG Display Files Download Files

Crystal structure of erenumab Fab bound to the extracellular domain of CGRP receptor  
PDB DOI: 10.2210/pdb6UMG/pdb  
Classification: MEMBRANE PROTEIN/IMMUNE SYSTEM  
Organism(s): *Homo sapiens*  
Expression System: *Homo sapiens, Escherichia coli BL21(DE3)*  
Mutation(s): No

Deposited: 2019-10-09 Released: 2020-02-12  
Deposition Author(s): Mohr, C.

Experimental Data Snapshot wwPDB Validation Metric Rfree Percentile Ranks Value 0.278

Method: X-RAY DIFFRACTION Resolution: 2.70 Å R Value Free: 0.282

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



**WEBLEM 4:****Introduction to STCRDAb database and visualisation of structure using PFV3D Tool****INTRODUCTION:**

T-cell receptors (TCRs) are proteins of the adaptive immune response. They are expressed on the surfaces of T-cells and typically recognise peptides that are presented by major histocompatibility complex (MHC) molecules. Despite their micromolar binding affinity and potential cross-reactivity, TCRs are selective for foreign peptide-MHC complexes on antigen presenting cells (APCs; 1–3). Upon binding, TCRs can activate the T-cell for direct killing of APCs, or stimulate other components of the adaptive immune system, such as B-cells. The clinical relevance of TCRs has attracted interest in understanding the structural basis of a TCR's activity and exploring the possibility of designing TCRs as novel biotherapeutics. Given the sensitivity of TCR-MHC interactions and the extreme diversity of the TCR repertoire, computational methods are increasingly being used for rational TCR design. TCR structural data is an invaluable resource for designing and developing computational tools, for example, template-based modelling pipelines.

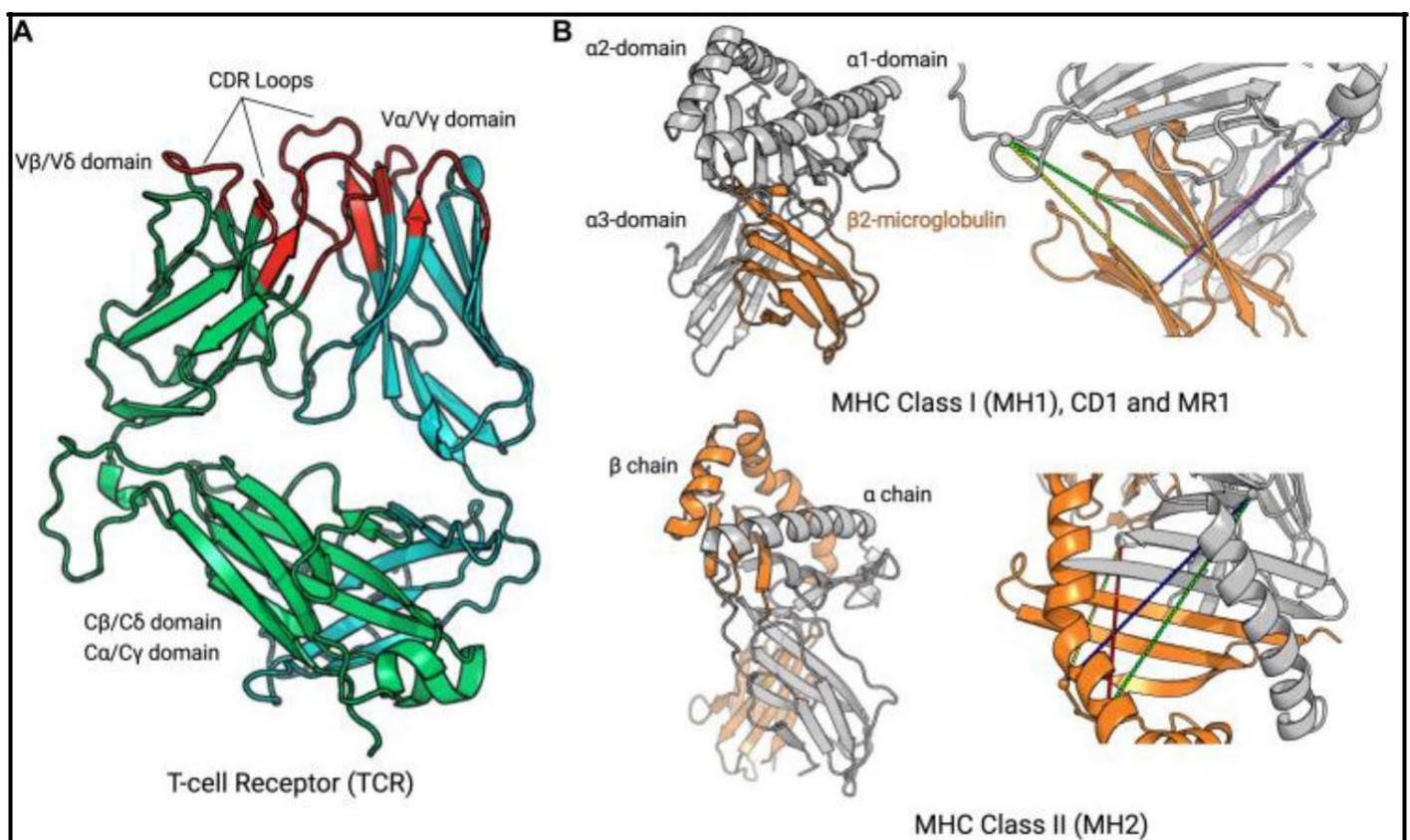
A small number of publicly available databases focus on delivering TCR-specific data. McPAS-TCR is a manually curated database that maps  $\alpha\beta$  TCR sequences to pathogens or epitopes. The database does not contain structural information, making it difficult to determine the importance of specific residues in MHC and antigen binding. There are two databases that contain some TCR structural information: ATLAS and IMGT. ATLAS is a manually curated database, containing a large volume of affinity data; users can view and download one of 87 experimental structures, and retrieve summaries of individual queries. The bulk of the structural data in ATLAS is comprised of homology models of variants of experimental structures. These structures lack annotations that can be useful for further analyses (e.g., numbering). Once again like McPAS-TCR, only  $\alpha\beta$  TCRs are annotated. IMGT has a richer (308 experimental structures) and more diverse set of structural data (e.g.,  $\gamma\delta$  TCRs). However, it is only possible to search based on a limited set of attributes; for example, it is not possible to specify the peptide sequence of the antigen. In addition, IMGT does not allow users to generate bespoke datasets for analysis.

They have developed the Structural TCR Database (STCRDab), building on our Structural Antibody Database. STCRDab is a TCR database that automatically

collects and curates' data on a weekly basis. Users can browse and select both  $\alpha\beta$  and  $\gamma\delta$  TCRs based on a wide range of criteria, such as the sequence of the TCR's complementarity-determining region (CDR) loops, the resolution of the structure, and the type of MHC molecule bound by the TCR. Users can also search by structural annotations, such as the orientation between the TCR's variable domains. STCRDab is linked to SAbDab, so that users can find antibody structures that are similar to TCRs, providing insight into designing TCR-like antibodies and chimaeric antigen receptors. Following a query, users can inspect and download individual or sets of TCR structures. Each search generates a unique zip file, containing a summary of the search and Protein Data Bank (PDB) format files of structures that match the query

### **Structure nomenclature**

STCRDab is primarily focussed on consistently annotating TCR structural data, but also numbers MHC molecules consistently.



**Figure 1.**

(Nomenclature and colouring scheme used in STCRab. (A) T-cell receptors (TCRs) are formed from two chains: TCR $\beta$ /TCR $\alpha$  (to form  $\alpha\beta$  TCRs, as shown), or TCR $\delta$ /TCR $\gamma$  (to form  $\gamma\delta$  TCRs). The residues coloured in red indicate the IMGT-defined CDR loops. This colouring scheme is also used on the website. (B) Major histocompatibility complex (MHC) molecules can be divided into classical and nonclassical MHCs. MH1 and MH2 are considered ‘classical’ MHCs, while CD1 and MR1 are ‘nonclassical’. However, CD1 and MR1 are structurally similar to MH1, whereas MH2 is structurally distinct. To pair MH1, we use the following distance constraints:  $\alpha15-\beta23$  (green; 32 Å),  $\alpha15-\beta104$  (yellow; 32 Å),  $\alpha51-\beta23$  (red; 32 Å),  $\alpha51-\beta104$  (blue; 37 Å). To pair MH2, the following distance constraints are used:  $\alpha29-\beta64$  (green; 34 Å),  $\alpha29-\beta39$  (yellow; 22 Å),  $\alpha37-\beta64$  (red; 32 Å),  $\alpha37-\beta39$  (blue; 28 Å).)

### TCR structures

The majority of available TCR structures are  $\alpha\beta$  TCRs, which are formed of TCR $\alpha$  and TCR $\beta$  chains. A small number of TCRs are  $\gamma\delta$  TCRs, consisting of TCR $\gamma$  and TCR $\delta$ .

chains. The TCR $\beta$  and TCR $\delta$  chains are considered to be analogous to antibody heavy chains while the TCR $\alpha$  and TCR $\gamma$  chains are considered to be analogous to antibody light chains.

Each TCR chain is characterised by two immunoglobulin domains: a variable domain ( $V$ ) and a constant ( $C$ ). Both variable and constant domains have a conserved  $\beta$ -sandwich structure, making it possible to number and compare variable domains from different TCRs. In STCRDab, we use the IMGT numbering as it provides consistent numbering for the CDR loops, and has been used on other occasions for structural analysis of TCRs. On each variable domain, there are three hypervariable loops that have the highest degree of sequence and structural variation, known as the CDRs. Flanking the CDRs, the remaining portions of the TCR structure are collectively known as the TCR's 'framework'.

### MHC structures

APCs use either the 'classical' MHC to present peptide antigens, or the 'nonclassical' MHC-like molecules to present lipid molecules or vitamin B precursors. The classical MHCs can be subdivided into MHC class I

(MH1) and MHC class II (MH2), while the nonclassical MHC-like molecules include cluster of differentiation 1 (CD1) and MHC class I-related protein (MR1). Both classical and nonclassical MHCs have an antigen binding groove formed by a  $\beta$ -sheet, flanked by two  $\alpha$  helices. MH1, CD1 and MR1 are formed by the pairing of the MHC chain and a  $\beta$ 2 microglobulin, while MH2 is formed by the MHC $\alpha$  and MHC $\beta$  chains. As with the TCR structures, the IMGT numbering is used for MHCs.

## DATA SOURCES AND CONTENTS

### TCR structures

As of 7 August 2017, STCRDab contains 348 entries with at least one TCR chain. On average, two TCR structures have been deposited in the PDB per month since 2007.

STCRDab is automatically updated weekly, in line with the PDB updating schedule. Paired  $\alpha\beta$  TCRs form the majority of the data, followed by single TCR chains, e.g. V $\beta$  only structures, then  $\gamma\delta$  TCRs. There are also structures that fit none of these categories – for instance, an engineered TCR $\delta$ /TCR $\alpha$  receptor.

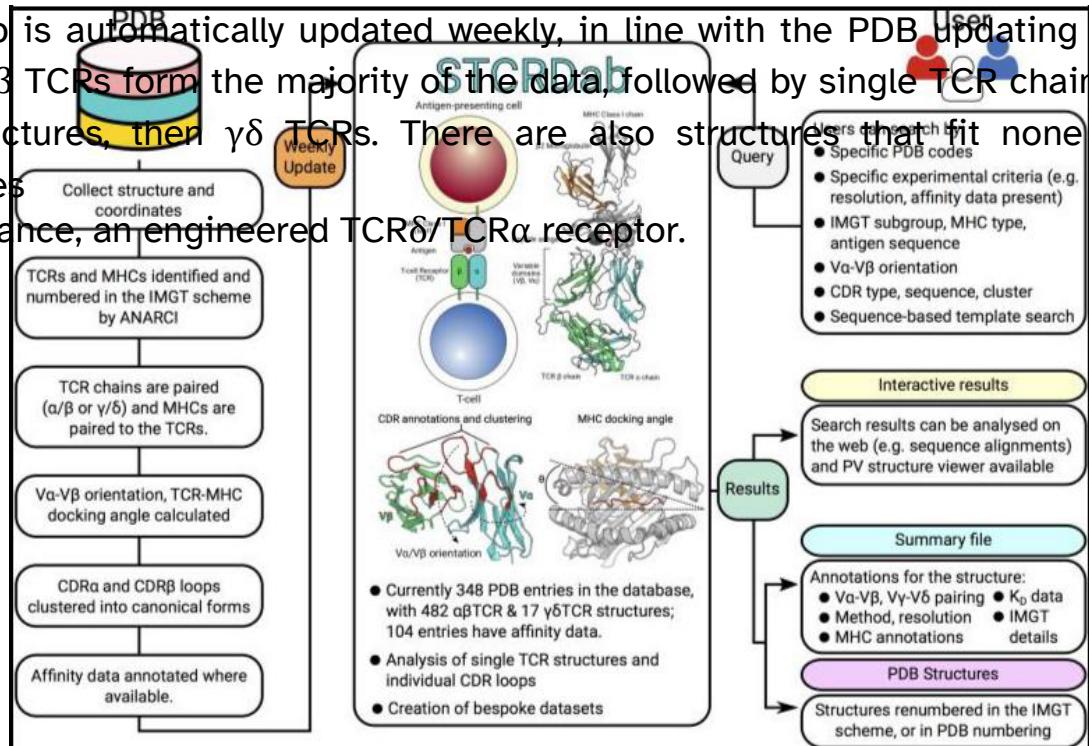


Figure 2.

**Workflow for STCRDab.** Every week, STCRDab automatically detects and numbers newly released TCR structures from the PDB using ANARCI (24). Any MHC or MHC-like molecules are also numbered by ANARCI. Each structure is automatically annotated with several structural properties,

such as its TCR-MHC docking angle (23). Users can submit a variety of queries to STCRDab to retrieve structures. Users are given their results for online analysis, and custom datasets are dynamically generated for download.

### V $\alpha$ -V $\beta$ orientation, docking angle

In order to describe the TCR binding mode with the MHC, we use a TCR-specific version of ABangle (TRangle), and calculate the docking angle between the TCR and the MHC. TRangle describes the relative orientation between the V $\alpha$  and V $\beta$  domains using six parameters. The effect of V $\alpha$ -V $\beta$  orientation on MHC binding is not yet clear, though it can provide the basis for engineering TCR-like antibodies, or antibody-like TCRs. STCRDab automatically calculates the TRangles for  $\alpha\beta$  TCRs. Due to the small amount of data, the TRangle method is currently not used for  $\gamma\delta$  TCRs; however, as data increases, this will become possible.

The docking angle describes how the TCR engages with the MHC. Here, we implement a previously established formula to calculate the docking angle.

### Complementarity-determining region loops and clustering

In STCRDab, the CDR loops are identified using the IMGT definition: CDR1 (IMGT 27–38), CDR2 (IMGT 56–65) and CDR3 (IMGT 105–117). The CDR $\alpha$ 1, CDR $\alpha$ 2, CDR $\alpha$ 3, CDR $\beta$ 1, CDR $\beta$ 2 and CDR $\beta$ 3 loops have been clustered into canonical forms, as has been done for the CDR loops of antibody structures

### TCR binding affinity

The binding affinities of TCR-MHC complexes were manually curated from PDDBind and ATLAS. Where possible, experimental details describing how the affinity was measured (e.g. surface plasmon resonance) were also annotated. For cases where the affinity of a TCR-MHC complex was measured in multiple studies (e.g. PDB: 3qdj), the values from the authors that determined the TCR structure are cited. There are currently 104 entries in STCRDab with a  $K_D$  value. These values should serve as a useful resource for those interested in TCR docking and design.

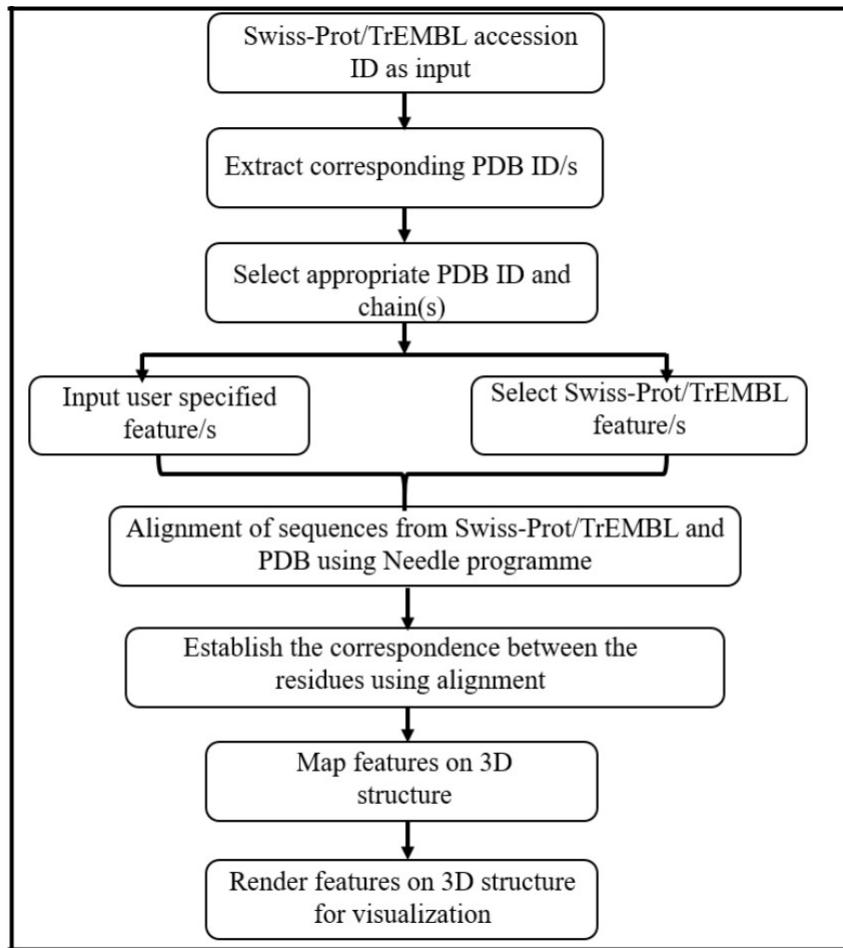
### PFV3D Tool:

Protein Feature Visualisation on 3D structure (PFV3D) is an online server to facilitate the mapping and visualization of UniProt feature annotation/s (Domain, Secondary structure, Active site/s etc.) as well as user specified feature/s on the corresponding 3D structure. Description of feature annotations is available on UniProt help section: [https://www.uniprot.org/help/sequence\\_annotation](https://www.uniprot.org/help/sequence_annotation)

The server also provides utility for UniProt to PDB or PDB to UniProt Residue mapping which is a very tedious task to do manually. Along with residue mapping, the utility also provides information on residues for which co-ordinates are missing in PDB file.

A specialized Perl module has been developed for extraction and downloading of UniProt feature annotation/s (sub-sequences of features) in a customized manner. User can download and use this module from the GitHub page: <https://github.com/rajivkarbhal/SFV>

Mapping and visualisation of sequence feature annotation/s on 3D structure can be done either by submitting UniProt entry or user defined feature annotation.



**Figure 3. Workflow of PVF3D tool**

The server is very easy to use and useful for Understanding of structural motifs (e.g. super-secondary and tertiary structure) through the visualization of secondary structural elements. Visualization of user-specified features on 3D structures (E.g. experimentally validated and predicted epitopes). Detailed understanding of various functional sites of proteins through visualization Visualization of binding/interacting surfaces of subunits in multiprotein complexes

The residue mapping service is used for mapping the UniProt residue on residues of corresponding PDB entry and make an equivalence between them. This utility can be used for, lists the missing residues in the coordinate section of the PDB file. This information is very important in various structural analyses of proteins. User can view

the result of residue mapping on the webpage or download in a text file for downstream processing.

STCRDab automatically collects and curates TCR structural data from the PDB. STCRDab builds upon the foundations of our antibody database, SAbDab, in order to provide consistent annotations, and open a gateway for users to easily access, view, and download custom datasets for analysis. The database aims to act as a resource for the emerging field of computational TCR design, and to help uncover the unique structural properties of TCRs. STCRDab also provides a bridge to the extensive knowledge base of antibody structures in SAbDab, which can potentially be used to inform TCR-like antibody design or antibody-like TCR design. Protein Feature Visualisation on 3D structure (PFV3D) is used to facilitate the mapping and visualization of UniProt feature annotation/s (Domain, Secondary structure, Active site/s etc.) as well as user specified feature/s on the corresponding 3D structure.

## REFERENCES:

1. Leem, J., de Oliveira, S. H., Krawczyk, K., & Deane, C. M. (2017). STCRDab: the structural T-cell receptor database. *Nucleic Acids Research*, 46(D1), D406-D412. <https://doi.org/10.1093/nar/gkx971>
2. PFV3D. (n.d.). Bioinfo.unipune.ac.in. Retrieved October 13, 2022, from <http://bioinfo.unipune.ac.in/PFV3D/Home>
3. PFV3D. (n.d.). Bioinfo.unipune.ac.in. Retrieved October 13, 2022, from <http://bioinfo.unipune.ac.in/PFV3D/About>

**WEBLEM 4a****STCRDab Database**

(URL:

**AIM:** [b/](http://opig.stats.ox.ac.uk/webapps/stcrda)

To identify the CDR for the crystal structure of Human pre-T cell receptor (PDB ID: 3OF6) using STCRDab Database.

**INTRODUCTION:**

The pre-T-cell antigen receptor (pre-TCR), expressed by immature thymocytes, has a pivotal role in early T-cell development, including TCR  $\beta$ -selection, survival and proliferation of CD4(-)CD8(-) double-negative thymocytes, and subsequent  $\alpha\beta$  T-cell lineage differentiation. Whereas  $\alpha\beta$ TCR ligation by the peptide-loaded major histocompatibility complex initiates T-cell signalling, pre-TCR-induced signalling occurs by means of a ligand-independent dimerization event. The pre-TCR comprises an invariant  $\alpha$ -chain (pre-T $\alpha$ ) that pairs with any TCR  $\beta$ -chain (TCR $\beta$ ) following successful TCR  $\beta$ -gene rearrangement. Here we provide the basis of pre-T $\alpha$ -TCR $\beta$  assembly and pre-TCR dimerization. The pre-T $\alpha$  chain comprised a single immunoglobulin-like domain that is structurally distinct from the constant (C) domain of the TCR  $\alpha$ -chain; nevertheless, the mode of association between pre-T $\alpha$  and TCR $\beta$  mirrored that mediated by the C $\alpha$ -C $\beta$  domains of the  $\alpha\beta$ TCR. The pre-TCR had a propensity to dimerize in solution, and the molecular envelope of the pre-TCR dimer correlated well with the observed head-to-tail pre-TCR dimer.

The Structural T-cell Receptor Database (STCRDab) is an online resource that automatically collects and curates TCR structural data from the Protein Data Bank. For each entry, the database provides annotations, such as the  $\alpha\beta$  or  $\gamma\delta$  chain pairings, major histocompatibility complex details, and where available, antigen binding affinities. In addition, the orientation between the variable domains and the canonical forms of the complementarity-determining region loops are also provided. Users can browse and select both  $\alpha\beta$  and  $\gamma\delta$  TCRs based on a wide range of criteria, such as the sequence of the TCR's complementarity-determining region (CDR) loops, the resolution of the structure, and the type of MHC molecule bound by the TCR. Users can also search by structural annotations, such as the orientation between the TCR's variable domains.

**METHODOLOGY:**

- Retrieve PDB ID for structure of 1GM7 from PDB database.
- Open homepage of STCRDab database (URL:  
<http://opig.stats.ox.ac.uk/webapps/stcrdab/>).
- Select CDR search.
- Select search for specific PDB.
- Enter the retrieved PDB ID and click on “Get CDR Structures”.
- Observe and interpret the results.

### **OBSERVATION:**



**Fig 1. PDB database page for crystal structure of Human pre T-cell receptor**



**Fig 2. Homepage of STCRDAb database**

Home STCRDab STCRPred Help SAbDab SAbPred

**CDR search**

Search CDRs based on specific criteria.

Search  >

Get all structures.

Search for a specific PDB.

**2. Get the CDR loops of a particular PDB in STCRDab.**

Please enter the PDB Code:  
3OFE

Get CDR structures

Advanced search of CDR structures.

Search for unique CDR structures.

About the canonical forms

Leem et al., Nucleic Acids Res. (2018), 46, D406-D412. © Copyright 2017. Developed by OPIG using Bootstrap and Flask.

**Fig 3. Page for CDR search**

Home STCRDab STCRPred Help SAbDab SAbPred

**CDR search**

Search CDRs based on specific criteria.

Search  >

Get all structures.

Search for a specific PDB.

**2. Get the CDR loops of a particular PDB in STCRDab.**

Please enter the PDB Code:  
3OFE

Get CDR structures

Advanced search of CDR structures.

Search for unique CDR structures.

About the canonical forms

Leem et al., Nucleic Acids Res. (2018), 46, D406-D412. © Copyright 2017. Developed by OPIG using Bootstrap and Flask.

**Fig 4. Search for specific PDB (PDB ID: 3OFE)**

**CDR search**

Search CDRs based on specific criteria.

Home STCRDab STCRPred Help SAbDab SAbPred

View results	TCR	Species	Method	Resolution	View CDR loop structure	View TCR Structure	Downloads
3of6	human	X-RAY DIFFRACTION	2.8	TCR B: CDRB1: SGHVS CDRB2: FQNEAQ CDRB3: ASSLGGQAYEQY  TCR C: CDRC1: SGHVS CDRC2: FQNEAQ CDRC3: ASSLGGQAYEQY  TCR A: CDRA1: SGHVS CDRA2: FQNEAQ CDRA3: ASSLGGQAYEQY	View Structure	• IMGT-numbered Structure • Summary file	Downloads

1 structure(s) fit your criteria.

**Download results**

- Download the [summary file](#) for this search. See [help](#) for more details on file formats.
- Download an archived [zip file](#) for this search. See [help](#) for more details on file formats.

Get all structures.  
Search for a specific PDB.  
Advanced search of CDR structures.  
Search for unique CDR structures.  
About the canonical forms

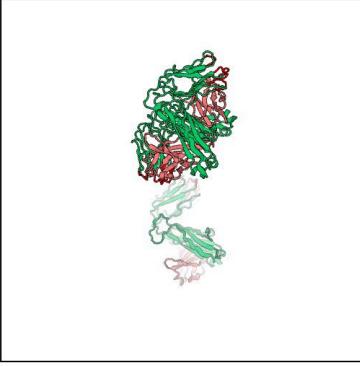
**Fig 5. Result for Human pre T-cell receptor**

Home STCRDab STCRPred Help SAbDab SAbPred

**Details for 3of6**

Click on any of the tabs below to see the detailed information about the structure.

Structure visualisation



Key (Default Scheme):  
 V $\beta$  Chains  
 IMGT CDRs  
 Bound Antigen Chains

Display options:  
 Cartoon model  
 Spacefill model  
 Wire model  
 Ball&stick model  
 Default colouring  
 Color by B-factor  
 Color by chain  
 Color by sec. structure  
 Color by element  
 Spin: on off

Structure information

Human Pre-T Cell Receptor Crystal Structure

**Fig 6. Structure visualization (V in green, V in blue and IMGT CDRs in grey)**

Structure information

Human Pre-T Cell Receptor Crystal Structure

Item	Info
PDB	3of6
Organism	HOMO SAPIENS
Method	X-RAY DIFFRACTION
Resolution	2.8Å
Number of TCRs	3

**Fig 7. Structure information**

Paired chains information

This PDB has 3 TCR(s).

A / NA

B / NA

C / NA

**Fig 8. Paired chain information (A/NA, B/NA, C/NA)**

[A / NA](#)

**TCR Details:**

Item	Info
VB chain	A
VB IMGT details	TRBV7/TRBJ2
Species	human

**Numbered sequence**

Chain type	Chain ID	FASTA/Annotation file	IMGT numbered sequence (Framework/CDR)							
VB	A	<a href="#">FASTA file</a>	B1	B2	B3	B4	B5	B6	B7	B8
			H	M	G	V	S	Q	S	P
			B9	B10	B11	B12	B13	B14	B15	B16
			R	Y	K	V	A	K	R	G
			B17	B18	B19	B20	B21	B22	B23	B24
			Q	D	V	A	L	R	C	D
			B25	B26	B27	B28	B29	B37	B38	B39
			P	I	<b>S</b>	<b>G</b>	<b>H</b>	<b>V</b>	<b>S</b>	L
			B40	B41	B42	B43	B44	B45	B46	B47
			F	W	Y	Q	Q	A	L	G
			B48	B49	B50	B51	B52	B53	B54	B55
			Q	G	P	E	F	L	T	Y
			B56	B57	B58	B63	B64	B65	B66	B67
			<b>F</b>	<b>Q</b>	<b>N</b>	<b>E</b>	<b>A</b>	<b>Q</b>	L	D
			B68	B69	B70	B71	B72	B73	B74	B75
			K	S	G	L	P	S	D	R
			B76	B77	B78	B79	B80	B81	B83	B84
			F	F	A	E	R	P	E	G
			B85	B86	B87	B88	B89	B90	B91	B92
			S	V	S	T	L	K	I	Q
			B93	B94	B95	B96	B97	B98	B99	B100

**Fig 9. Information for A/NA (TCR details and IMGT numbered sequence for framework and CDR(red))**

**CDR Sequences:**

Loop	Sequence	Predicted canonical form	CDR Length
<b>CDRB3</b>	ASSLGQAYEQY	None	11
<b>CDRB2</b>	FQNEAQ	B2-6-B	6
<b>CDRB1</b>	SGHVS	B1-5-A	5

**Fig 10. CDR sequences**

Additional links and files for download: see [help](#) for more details.

Item	Link
IMGT-numbered structure	<a href="#">Link</a>
Non-annotated structure from the PDB	<a href="#">Link</a>
Summary file for this TCR	<a href="#">Link</a>
Link in the PDB	<a href="#">Link</a>

**Fig 11. Links and files for download**

## RESULT:

STCRDab Database was used to identify the CDR for the crystal structure of TCR A6 (PDB ID: 4GRm). The following result showed the structure visualisation for V $\alpha$  chain in green, V $\beta$  chain in blue and IMGT CDRs in grey. Also, two TCRs were obtained for this structure. TCR details and IMGT numbered sequence for framework and CDR which is highlighted in red was given for both TCRs. CDR sequences with its length was also mentioned. Orientation and docking angles i.e angle at which the TCR engages with the MHC as well as TCRs and antibodies with similar orientation with that of query was retrieved.

## CONCLUSION:

STCRDab automatically collects and curates TCR structural data from the PDB. STCRDab builds upon the foundations of our antibody database, SAbDab, in order to provide consistent annotations, and open a gateway for users to easily access, view, and download custom datasets for analysis. In addition, the orientation between the variable domains and the canonical forms of the complementarity-determining region loops are also provided. The database aims to act as a resource for the emerging field of computational TCR design, and to help uncover the unique structural properties of TCRs. STCRDab also provides a bridge to the extensive knowledge base of antibody structures in SAbDab, which can potentially be used to inform TCR-like antibody design or antibody-like TCR design.

## REFERENCES:

1. Leem, J., de Oliveira, S. H., Krawczyk, K., & Deane, C. M. (2017). STCRDab: the structural T-cell receptor database. *Nucleic Acids Research*, 46(D1), D406–D412. <https://doi.org/10.1093/nar/gkx971>
2. Scott, D. R., Borbulevych, O. Y., Piepenbrink, K. H., Corcelli, S. A., & Baker, B. M. (2011). Disparate degrees of hypervariable loop flexibility control T-cell receptor cross-reactivity, specificity, and binding mechanism. *Journal of Molecular Biology*, 414(3), 385–400. <https://doi.org/10.1016/j.jmb.2011.10.006>
3. *STCRDab*. (n.d.). Opig.stats.ox.ac.uk. Retrieved October 13, 2022, from <http://opig.stats.ox.ac.uk/webapps/stcrdab/About>
4. *STCRDab*. (n.d.). Opig.stats.ox.ac.uk. Retrieved October 13, 2022, from <http://opig.stats.ox.ac.uk/webapps/stcrdab/>

**DATE:**  
**13/10/22**

## **WEBLEM 4b**

### **PFV3D tool**

(URL: -

**AIM:** <http://bioinfo.unipune.ac.in/PFV3D/Home> )

To study protein feature visualization for Penicillin F acylase(UniProt Id: P06875) on 3D Structure using PFV3D Tool.

### **INTRODUCTION:**

Penicillin G acylase (PGA) is one of very important industrial enzymes used in the production of polysynthetic beta-lactam antibiotics. This enzyme catalyzes the hydrolysis of the amidic bond of penicillin G with the development of 6-aminopenicillanic acid which serves as the initial substance for the production of semisynthetic penicillins. In the strain Escherichia coli W ATCC 11105 and ATCC 9637, PGA is coded by the pga gene on the chromosome and synthesized as the pre-pro-PGA (pp PGA) precursor, which is transported, with probable participation of the chaperon system, to the periplasmatic space of the cell. Here after a series of proteolytic reactions the active enzyme PGA develops, consisting of two subunits alpha and beta. Expression of the pga gene is subject to several regulatory mechanisms: temperature repression, catabolic repression by glucose, repression by oxygen, and induction by phenylacetic acid (FOK). The formation of active PGA is also influenced at the post-translation level, where an important role is played by intracellular proteolytic reactions and the transport system of pre-pro-PGA across the cytoplasmatic membrane. The chromosomal area of the pga gene of the E. coli W strain was employed for the construction of many recombinant plasmids. These plasmids served to transform suitable host strains, some of which are now used in industry as highly productive microorganisms.

Protein Feature Visualisation on 3D structure (PFV3D) is an online server to facilitate the mapping and visualization of UniProt feature annotation/s (Domain, Secondary structure, Active site/s etc.) as well as user specified feature/s on the corresponding 3D structure. Description of feature annotations is available on UniProt help section. The server also provides utility for UniProt to PDB or PDB to UniProt Residue mapping which is a very tedious task to do manually. Along with residue mapping, the utility also provides information on residues for which co-ordinates are missing in PDB file.

### **METHODOLOGY:**

- Retrieve UniProt accession id for Immunoglobulin lambda-like polypeptide 1.
- Open homepage of PFV3D tool (URL: <http://bioinfo.unipune.ac.in/PFV3D/Home>).
- Enter the UniProt Id and click on search.
- Select the features to be mapped.
- Select the PDB ID and Chain ID to be mapped out of all structures obtained.
- Observe and interpret the results.

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search

**P06875 · PAC\_ECOLX**

Function	Protein <sup>i</sup> : Penicillin G acylase	Amino acids <sup>i</sup> : 846
Names & Taxonomy	Status <sup>i</sup> : UniProtKB reviewed (Swiss-Prot)	Protein existence <sup>i</sup> : Evidence at protein level
Subcellular Location	Organism <sup>i</sup> : Escherichia coli	Annotation score <sup>i</sup> :
Phenotypes & Variants	Gene <sup>i</sup> : pac	
PTM/Processing		

Expression Interaction Structure Family & Domains Sequence Similar Proteins

Entry Feature viewer Publications External links History BLAST Align Add Add a publication Entry feedback

**Function<sup>i</sup>**

**Catalytic Activity**

H<sub>2</sub>O + penicillin = 6-aminopenicillanate + a carboxylate  
EC:3.5.1.11 (UniProtKB | ENZYME | Rhea )  
Source: Rhea 18693<sup>i</sup>

Fig 1. Uniprot page for Penicillin G acylase

**Protein Feature Visualisation on 3D structure**

SQETFSDLWKLLPENNVLSPQAMDDLM MEEPQSDPSVEPPLSQETFSDLWKLPPENNVLSPQAMDDLM LSPDDI Home About Residue Mapping Help Contact

**Protein Feature Visualisation on 3D structure (PFV3D)** is an online server to facilitate the mapping and visualization of UniProt feature annotation/s (Domain, Secondary structure, Active site/s etc.) as well as user specified feature/s on the corresponding 3D structure. Description of feature annotations is available on UniProt help section: [https://www.uniprot.org/help/sequence\\_annotation](https://www.uniprot.org/help/sequence_annotation)

The server also provides utility for **UniProt to PDB or PDB to UniProt Residue mapping** which is a very tedious task to do manually. Along with residue mapping, the utility also provides information on residues for which co-ordinates are missing in PDB file.

A specialized **Perl module** has been developed for extraction and downloading of UniProt feature annotation/s (sub-sequences of features) in a customized manner. User can download and use this module from the GitHub page: <https://github.com/rajivkarbhal/SPV/>

Mapping and visualisation of sequence feature annotation/s on 3D structure

► Map feature annotation/s from UniProt entry:  
► Map user defined feature annotation/s:

Fig 2. Homepage of PFV3D tool

▼ Map feature annotation/s from UniProt entry:

Enter the UniProt Accession Id in text box:

P06875

Submit

► Map user defined feature annotation/s:

Fig 3. Mapping using feature annotations for UniProt entry

Uniprot Accession Number: [P06875](#) | [FASTA](#)

Protein name/s: Penicillin G acylase (EC 3.5.1.11) (Penicillin G amidase) (Penicillin G amidohydrolase) [Cleaved into: Penicillin G acylase subunit alpha; Penicillin G acylase subunit beta]

Feature annotation/s available for the entry: (Select feature/s below to visualize on 3D structure)

Active site (1)  
 Beta strand (36)  
 Binding site (6)  
 Chain (3)  
 Helix (36)  
 Mutagenesis (1)  
 Propeptide (1)  
 Sequence conflict (14)  
 Signal peptide (1)  
 Turn (5)

Fig 4. Features selected from available features for the entry (Beta stran, chain and

**Helix)**

**Protein Feature Visualization on 3D structure**

SQETFSDEMKELIENVLESTLPIQAMDDLM  
GPDEAPRMPEAAPPVAFAPAPAAPTAAAPAPA

MEEPQSDPSVEPLSQETFSDLWKLLENNVLSPLPSQAMDDLM SQETFSDLWKLLENNVLSPLPSQAMDDLM

LSPDDI Home About Residue mapping Help Contact

Select All Unselect All Reset Submit

Selected feature/s in accession id: P06875

ACTIVE SITE:

ACTIVE SITE 290 290

Beta Strands:

- STRAND 32 38
- STRAND 43 46
- STRAND 214 216
- STRAND 291 295
- STRAND 300 310
- STRAND 316 318
- STRAND 320 328
- STRAND 331 338
- STRAND 345 348
- STRAND 350 358
- STRAND 364 370
- STRAND 373 375
- STRAND 378 381
- STRAND 384 387
- STRAND 389 395
- STRAND 403 410
- STRAND 413 419
- STRAND 424 430
- STRAND 464 472
- STRAND 477 483
- STRAND 498 502

© Bioinformatics Centre, Savitribai Phule Pune University, India

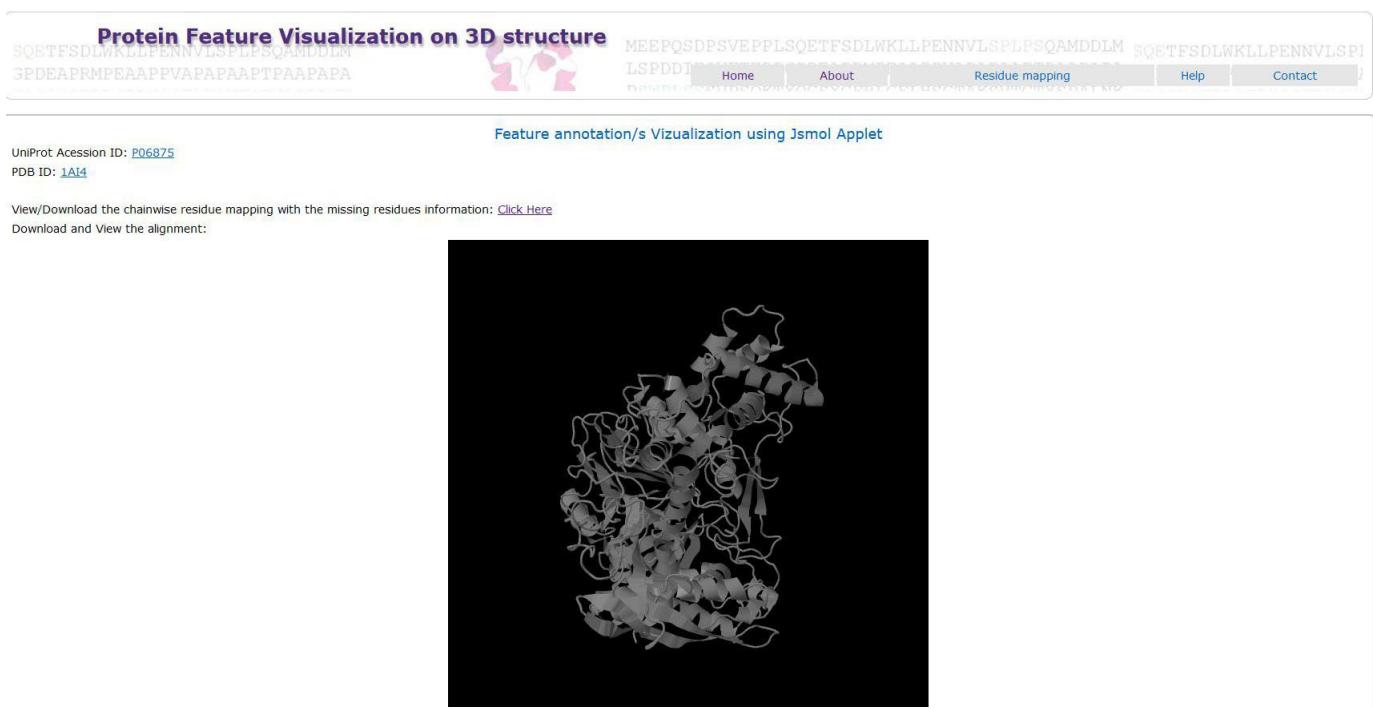
**Fig 5. Features selected under beta strand, chain and helix**

Select PDB id and Chain Id/s from the list below

Submit

PDB ID	Resolution ▲▼	Title
<input checked="" type="radio"/> 1AI4	2.35	PENICILLIN ACYLASE COMPLEXED WITH 3,4-DIHYDROXYPHENYLACETIC ACID
<input type="radio"/> 1AI4	2.35	PENICILLIN ACYLASE COMPLEXED WITH 3,4-DIHYDROXYPHENYLACETIC ACID
<input type="radio"/> 1AI5	2.36	PENICILLIN ACYLASE COMPLEXED WITH M-NITROPHENYLACETIC ACID
<input type="radio"/> 1AI5	2.36	PENICILLIN ACYLASE COMPLEXED WITH M-NITROPHENYLACETIC ACID
<input type="radio"/> 1AI6	2.55	

**Fig 6. Selection fo PDB id and Chain id**



**Fig 7. Visualisation of mapped features**

## RESULT:

PFV3D Tool was used to study protein feature visualization for for Plasma kallikrein of Plasma protein (UniProt Id: P03952) on 3D Structure. Beta strand, chain and helix features of the query were chosen for mapping and structure for catalytic domain of Human Plasma Kallikrein with the Implications for Structure- Based Design of Protease Inhibitors along B chain which have low resolution was selected. The mapped features were visualised in 3D and the alignment between query and structure was also observed.

## CONCLUSION:

Protein Feature Visualisation on 3D structure (PFV3D) is used to facilitate the mapping and visualization of UniProt feature annotation/s (Domain, Secondary structure, Active site/s etc.) as well as user specified feature/s on the corresponding 3D structure. The structures retrieved based on the features mapped can be used ahead for docking studies

## REFERENCES:

1. PFV3D. (n.d.). Bioinfo.unipune.ac.in Retrieved October 13, 2022, from

<http://bioinfo.unipune.ac.in/PFV3D/Home>

2 *PFV3D*. (n.d.). Bioinfo.unipune.ac.in Retrieved October 13, 2022, from

<http://bioinfo.unipune.ac.in/PFV3D/About>

3 *Plasma Proteins / BioNinja*. (n.d.). Retrieved October 13, 2022, from

[https://ib.bioninja.com.au/options/option-d-human-physiology/d3-functions-of-the-liver/plasma-](https://ib.bioninja.com.au/options/option-d-human-physiology/d3-functions-of-the-liver/plasma-proteins.htm)

[proteins.htm](https://ib.bioninja.com.au/options/option-d-human-physiology/d3-functions-of-the-liver/plasma-proteins.htm)





## WEBLEM 5

### **AIM:**

### **Yvis Platform**

Introduction to Yvis platform for studying variable and constant domains.

### **INTRODUCTION:**

The Yvis database is an updated weekly collection of data on antibody PDB structures (in complex with an antigen or not), such as PDB and chain identification, antibody and protein antigen-producing organisms or type gapped sequences of antibody chains, germline information (assigned V and J genes with their identity values), and antigen-antibody putative contacts.

The Yvis script, developed in Python, extracts a list of antibodies PDB structures from SAbDab that is updated weekly. The following data are extracted from this list, processed, and stored in the Yvis database:

(i) PDB and chain identifications, (ii) names of the organisms producing antibody and antigen (when applicable) and (iii) antigen molecule description. When the SAbDab list does not contain the antibody- or antigen-producing organism name, Yvis script extracts this information from the corresponding PDB structure file, acquiring the ORGANISM\_SCIENTIFIC value from SOURCE record after retrieving the molecule ID from COMPND record. After data extraction, Yvis script checks whether the organism names match the UniProt Taxonomy, and correct them if required. Data are manually curated, if the standard name is not found automatically. These standard names facilitate the Yvis database search, reducing the diversity of organism names, for instance by eliminating all synonyms.

The Yvis script submits antibody chain sequences to IMGT/DomainGapAlign to obtain gapped sequences and germline information. Then, it processes the result page and extracts the gapped sequence of the variable domain of each chain, following the IMGT numbering. Moreover, the script extracts and stores the V and J germline genes assigned to the chain sequence, and their identity values.

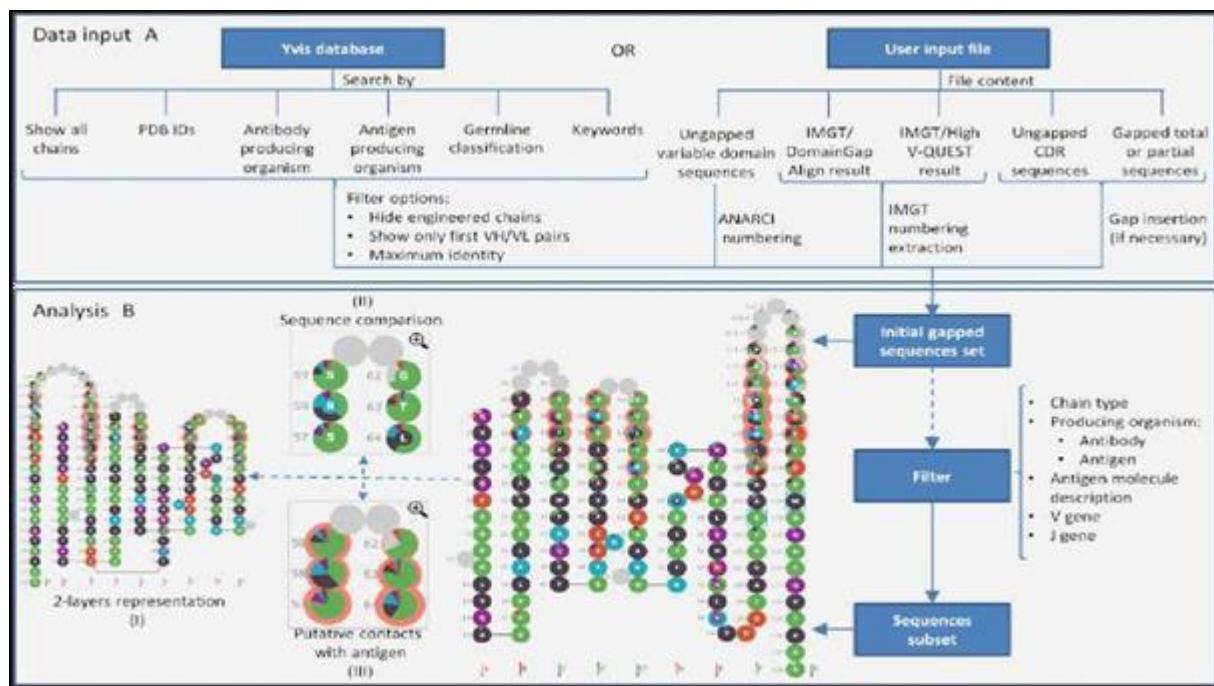
Finally, to obtain information on the putative antibody-antigen contacts, the Yvis script downloads the PDB structure files and extracts the antibody chain amino acids that potentially interact with a peptide or protein antigen using the Biopython PDB module. Then, the distance between each  $\alpha$ -carbon of the antibody and antigen amino acids is calculated. If the distance between two  $\alpha$ -carbons is not higher than 8 Å, the position that contains the amino acid is marked as making a putative contact. This distance is used because it allows including putative direct interactions between antigen and antibody and also water-mediated interactions.

## **analysis**

The Yvis platform integrates resources that allow the analysis of antibody variable domains that have been uploaded as user sequences or selected from the Yvis database. This platform is a web-based application that process sequences in a server or in a user's internet browser, depending on the analyzed data. The server-side application was developed using PHP and Mysql, and the client-side using the JavaScript and D3.js framework.

**The Yvis Platform offers input and search versatility** With the Yvis platform, users can analyze antibody structures stored in the Yvis database or uploaded by them. Different search options (Figure 1) are available to select, from the database, a set of antibody structures to be analyzed. It is possible to show all antibody chains stored in the Yvis database, or to specify a list of PDB identifiers, or a pair of PDB: chain identifiers. Moreover, users can choose to show free or complexed antibodies, and in the latter case, they can indicate the antigen type (hapten, carbohydrate, nucleic acid or protein). For protein antigens, they can indicate the producing organism. Users can also select antibodies with assigned germline V or J genes, or produced by user-selected organisms. In addition, users can search antibodies

by using keywords contained in the literature related to PDB structures. After defining the PDB structure search criteria, the user can apply additional filters to avoid sequence redundancy, such as: (i) to choose only one representative chain of each type (heavy or light) in each PDB structure; (ii) to specify an identity threshold that ensures that none of the filtered sequences has an identity value higher than the user-specified value. This approach was based on Cd-hit. Because of the time requested to analyze and group all sequences, the identity filter is not used by default. All these filters can be combined.



**Fig 1: Yvis platform overview. (A)** The data input box presents two possibilities to input sequences to be analyzed in Yvis (user input files and selection from the Yvis database).

It presents filter options for sequences from the Yvis database (redundancy and engineered chains) and actions taken by the platform to process the user input files. (B) The analysis box presents the options to visualize a multiple sequence alignment of antibody variable domains and the filter possibilities. The user can generate a new subset of sequences to be analyzed, by selecting specific filters. The analysis can be displayed by the Collier de Diamants on one or two layers (I).

Additionally, the user can compare the multiple sequence alignment with a reference sequence (II), and visualize data on putative contacts with the antigen (III).

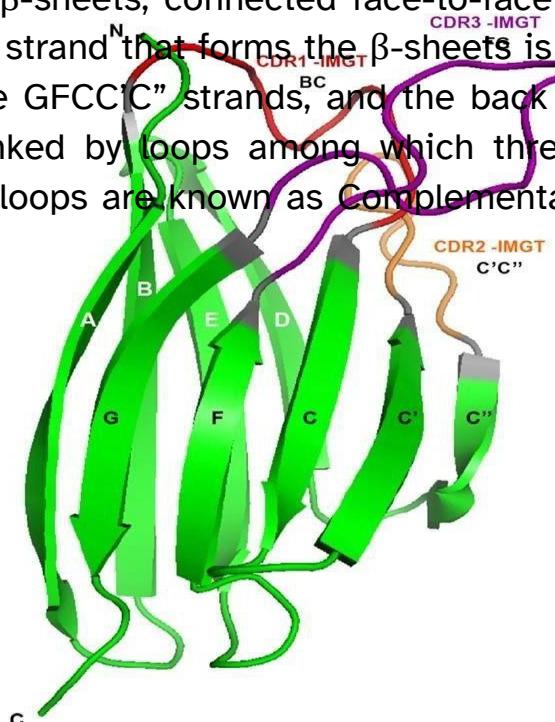
Users can also analyze antibody sequences obtained from an IMGT/DomainGapAlign results file, an IMGT/HighV-QUEST gapped amino acid results file, or a FASTA file containing gapped, or ungapped chain sequences or even CDR sequences (Figure 1A, User Input file). When a user submits an IMGT

results file, the Yvis platform will process it in the user's browser. Moreover, when a user submits ungapped sequences in a FASTA file, the Yvis platform will number them using ANARCI in the Yvis server.

- Yvis Tools:

## 1.1. Antibodies

Antibodies or immunoglobulins are vertebrate immune system proteins produced by B cells and capable of binding to antigens with high specificity and affinity. Most antibodies present a Y-shaped portion, formed by two identical pairs of chains. Each chain pair contains one heavy and one light chain, and each chain has a variable domain and one or more constant domains. The variable domain is the antibody portion that interacts with the antigen. All antibody chains have a variable domain formed by two  $\beta$ -sheets, connected face-to-face by a disulphide bond, as shown in Figure 1. Each strand that forms the  $\beta$ -sheets is identified by a letter. The front sheet contains the GFCC $C''$  strands, and the back sheet contains the ABED strands. Strands are linked by loops among which three are usually involved in antigen binding. These loops are known as Complementarity Determining Regions (CDRs).



**Fig 2: Cartoon representation of the variable domain structure of an antibody chain. Individual letters identify the strands (in green), and the different colours of these letters distinguish the two sheets. The three CDR loops are highlighted in red, orange and purple.**

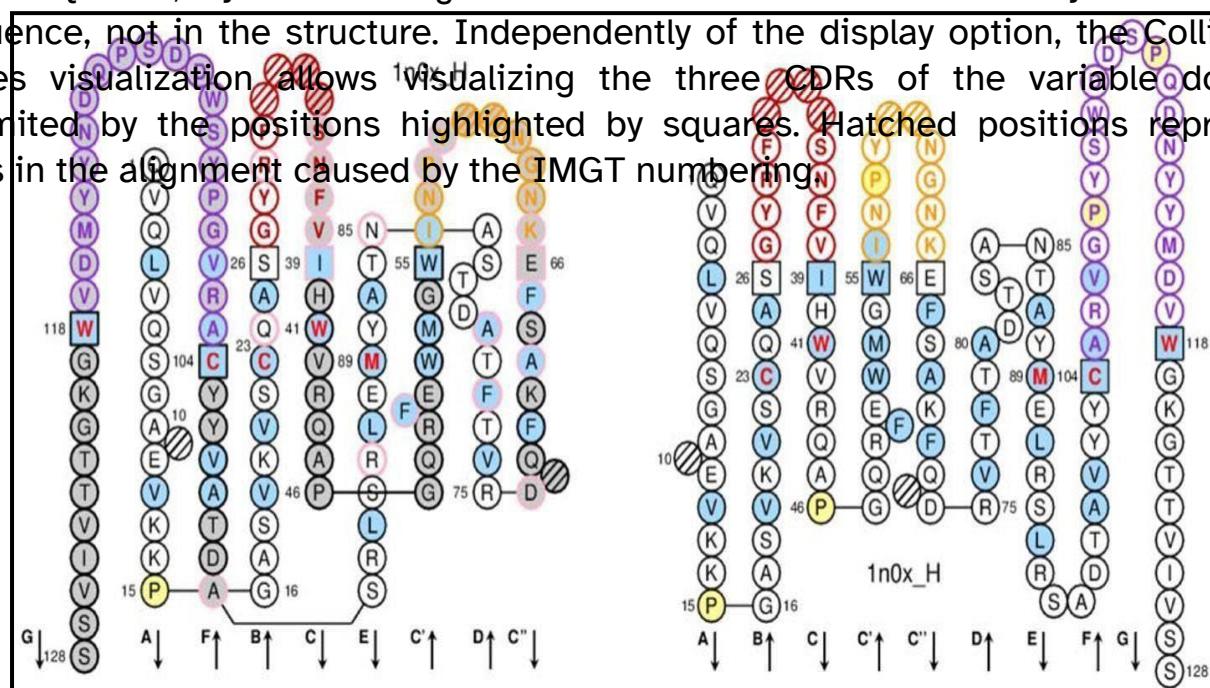
## 1.2. Antibody numbering

To compare the variable domain of antibody chains, some numbering schemas were proposed. They were defined based on the superimposition of antibody

structures showing that there is a high similarity between some parts of the variable domain of antibody chains, known as frameworks (FWs). The numbering schemas allow the identification of FWs and of the hypervariable regions that are usually associated with the antigen binding and are known as CDRs. When a numbering schema is applied to a sequence, some key residues, which are conserved in the numbering definition, are searched, and gaps are inserted to generate a numbered sequence. Antibody numbering is important in antibody analysis because it provides an implicit sequence alignment between any possible variable domain sequence of an antibody chain, thus delimiting the FWs and CDRs.

### 1.3. Antibody data visualization

Data visualization allows data representation in a way that can make easier to understand the data significance. The IMGT/Collier de Perles is a visualization tool that represents the amino acid composition of the variable domain of an antibody chain associated with the conserved 3D structure of antibodies. It displays the variable domain sequence in one or two layers (see Figure 2). As the variable domain of all antibody chains has the same structure formed by two  $\beta$ -sheets connected face-to-face, the Collier de Perles on two layers presents the amino acid sequence closer to its 3D structure, by “superimposing” the strands. The Collier de Perles on one layer shows the same data, but in a way that it is closer to the amino acid sequence, by maintaining the sheets in the same order they have in the sequence, not in the structure. Independently of the display option, the Collier de Perles visualization allows visualizing the three CDRs of the variable domain delimited by the positions highlighted by squares. Hatched positions represent gaps in the alignment caused by the IMGT numbering.



**Fig 3: Collier de Perles visualization in two layers (left image) and one layer (right image).**

Individual letters and arrows identify the strands.

The CDR loops are highlighted in red, orange and purple.

The IMGT/Collier des Perles is a great visualization of the variable domain of an antibody chain, but can only represent one chain per visualization.

**Collier des Diamants**

In Yvis, we offer a new visualization that includes the advantages of the IMGT/Collier de Perles and MSA visualization methods (i.e., representation of antibody sequence closer to the structure to highlight the CDR and FW relations), and the possibility to analyze multiple sequences in the same visualization. This new visualization is based on the IMGT/Collier de Perles (Pearl Necklace) representation. Instead of representing only one amino acid per position, the Collier de Diamants represents multiple amino acids in each position, from a multiple sequence alignment. As each pearl of the necklace is replaced by a new representation with multiple “facets”, this new visualization was called Collier de Diamants (Diamond Necklace).

## Input options:

Users can analyze their own sequences or sequence data stored in the Yvis database.

If the data source is the Yvis database, users can select among the following search criteria and filter options:

- Structures from Protein Data Bank (PDB) that contain antibodies

The screenshot shows the Yvis interface with a search bar at the top containing the text "Protein Data Bank (PDB) that contain antibodies". Below the search bar, there is a section titled "Choose an option to exhibit the antibody variable domain analysis:". This section contains several buttons for filtering results:

- Show all stored antibody PDB sequences (with a question mark icon)
- Filter by PDB IDs (with a question mark icon)
- Filter by antibody-producing organism (with a question mark icon)
- Filter by antigen-producing organism (with a question mark icon)
- Filter by germline classification (with a question mark icon)
- Filter by literature keywords (with a question mark icon)

**Fig 4.1: Search option for Structures from Protein Data Bank (PDB)**

### 1. Show all stored antibody PDB sequences

This screenshot shows a configuration dialog for the "Show all stored antibody PDB sequences" option. It includes the following settings:

- A large "Show all stored antibody PDB sequences" button with a question mark icon.
- A "Analyze" button.
- Checkboxes for:
  - Show engineered antibodies (checked)
  - Show only sequences with the right amino acid in conserved positions (checked)
  - Show only one VH/VL pair for each PDB file (checked)
  - Use identity filter. Cutoff: 0.0 % (unchecked)

**Fig 4.2: Show all stored antibody PDB sequences**

Select this option to show information on all antibody sequences from PDB and stored in Yvis database

## 2. Filter by PDB IDs

The screenshot shows a web-based filtering interface titled "Filter by PDB IDs". At the top, there's a radio button group for "Specify PDB IDs" (selected) and "Specify PDB IDs and chain name". A "Load example" link is also present. Below this, a text input field asks for "Enter PDB IDs separated by commas, semicolons or in new lines". Underneath the input field are several filter checkboxes:

- Show engineered antibodies
- Show only sequences with the right amino acid in conserved positions
- Show only one VH/VL pair for each PDB file
- Use identity filter. Cutoff:  %

A large "Analyze" button is located on the left side of the filter section.

**Fig 4.3: Filter by PDB IDs**

Select this option to show only chains from structures of a user-defined list of PDB identifiers, with or without chain specification.

You can specify a list of PDB IDs by selecting the "Specify PDB IDs" option and inserting in the textbox the PDB IDs separated by commas, semicolons, or by putting each ID in a new line. In this case, Yvis will show the chains stored in the Yvis database that are part of the indicated structures.

If you want to restrict the analysis to specific chains, you should select the "Specify PDB IDs and chain name" option and insert in the textbox a list of chains separated by commas, semicolons, or in new lines. Each chain must be specified by the PDB ID followed by a colon and the chain name.

### Filter options:

Select "Show engineered antibodies" if you want to show, in the results, sequences marked as engineered in the PDB structure file.

Select "Show only sequences with the right amino acid in conserved positions" to restrict the results to sequences that have the correct amino acid residues in the conserved positions of the IMGT numbering: Cysteine 23, Tryptophan 41, Cysteine 104, Leucine 89 and Phenylalanine or Tryptophan 118.

Select "Show only one VH/VL pair for each PDB file" if you want to show, in the results, only the first pair of light and heavy chains of a PDB structure file. Otherwise, all antibody chains of the structures will be shown.

Select "Use identity filter" and set an identity cut-off if you want to analyze only a set of sequences having at most that identity value.

- If using own data, the user can choose among the following input options

**Fig 5.1: Search option for User defined sequences**

## 1. Antibody chain sequences

The screenshot shows a user interface for uploading antibody chain sequences. At the top, there is a section titled "Variable domain sequences" with a question mark icon. Below it, a note says "Upload a FASTA file containing the variable domain sequences. Maximum allowed file size: 2.5MB." There is a "Choose File" button followed by the text "No file chosen". A checked checkbox labeled "Extract germline information with ANARCI" is present. At the bottom are two buttons: "Analyze" and "Example file".

**Fig 5.2: Antibody chain sequences**

Select this option to insert a FASTA file that contains amino acid sequences of variable domains of antibody chains. The Yvis server uses ANARCI to gap sequences.

## 2. CDR sequences

The screenshot shows a user interface for uploading CDR sequences. At the top, there is a section titled "CDR sequences" with a question mark icon. Below it, a note says "Upload a fasta file containing the CDR sequences:". There is a "Choose File" button followed by the text "No file chosen". At the bottom are two buttons: "Analyze" and "Example file".

**Fig 5.3: CDR Sequences**

Select this option to insert a FASTA file containing complementarity-determining region (CDR) amino acid sequences. Choose the type of CDR sequences (CDR1, CDR2, or CDR3; heavy and light chain are treated in the same way). The sequence length must be at most equal to the number of amino acids indicated in each CDR. The Yvis platform will gap sequences according to the chosen CDR.

## **REFERENCES:**

- 1) *Yvis Platform*. (n.d.). NCBI. Retrieved October 29, 2022, from

[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6602444/)

[602444/](#)

- 2) *Yvis Platform.* (n.d.). Retrieved October 29, 2022, from  
<http://bioinfo.icb.ufmg.br/yvis/#analyze>

**WEBLEM 5a:****Yvis Platform**(URL: <http://bioinfo.icb.ufmg.br/yvis/>)**AIM:**

To study variable and constant domain using topology diagram on Yvis Platform.

**INTRODUCTION:**

Y-Vis is a web-based platform that allows the analysis of antibody sequences through a new visualization called Collier des Diamants. This new visualization is based on the IMGT/Collier de Perles representation, and provides information on the amino acids present in the variable domain of an antibody chain together with their position in the conserved beta-strands and loops that define the antibody structure.

Moreover, Collier des Diamants allows visualizing the alignment of multiple antibody chain sequences using the IMGT/Collier de Perles graphical representation, thus providing a new way to analyze antibody variable domain sequences on the basis of the amino acid composition and their positions in the antibody structure.

Yvis allows analyzing user-defined sequences or antibody data from the Yvis database that contains pre- processed information on Protein Data Bank (PDB) antibody structures. User- defined sequences can be uploaded as FASTA files, or as results files.

**METHODOLOGY:**

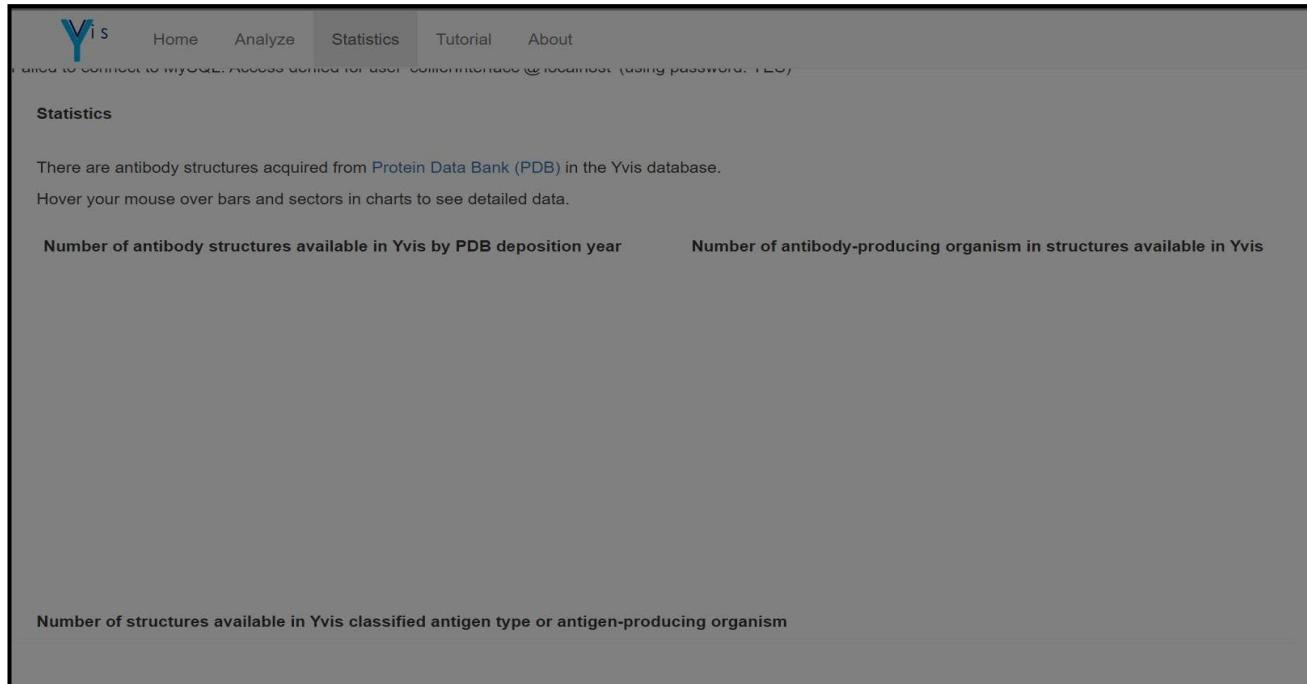
- 1) Open the Yvis platform from google (URL: <http://bioinfo.icb.ufmg.br/yvis/>)
- 2) Click on analyze, go to user defined sequence section
- 3) Click on CDR Sequences, Choose the CDR region (CDR3)
- 4) Download the CDR3 File.
- 5) Upload the CDR3 file and click on analyze.
- 6) Observe and interpret the result.

## OBSERVATIONS:

The screenshot shows the Yvis homepage. At the top is a navigation bar with the Yvis logo, Home, Analyze, Statistics, Tutorial, and About tabs. Below the navigation bar is a section titled "Yvis: AntibodY high-density alignment visualization and analysis platform with integrated database". This section contains a detailed text about the platform's purpose and features, mentioning the IMGT/Collier des Diamants representation and its ability to analyze antibody sequences. It also describes how users can upload FASTA files or IMGT results files, and how they can refine their search using filters and export data.

The screenshot shows the "Analyze" tab of the Yvis platform. The tab is highlighted with a red box. Below the tab, there is a heading "Choose an option to exhibit the antibody variable domain analysis:". Under this heading, there are several filter options: "Structures from Protein Data Bank (PDB) that contain antibodies:", "Show all stored antibody PDB sequences", "Filter by PDB IDs", "Filter by antibody-producing organism", "Filter by antigen-producing organism", "Filter by germline classification", and "Filter by literature keywords". Further down, under "User defined sequences:", there are two options: "Variable domain sequences" and "CDR sequences".

**Fig 2: Analyze page under Yvis Platfrom**



User defined sequences:

### Fig 3: Statistics page under Yvis Platform

Variable domain sequences [?](#)

CDR sequences [?](#)

Choose a CDR:  CDR1 (Max 12 aa)  CDR2 (Max 10 aa)  CDR3 (unlimited)  
 Upload a fasta file containing the CDR sequences:  
 No file chosen  
 Example file

IMGT/DomainGapAlign results file [?](#)

IMGT/HighV-QUEST results file [?](#)

User defined gapped sequences intervals [?](#)

**Fig 4: Search option using CDR Sequences**

**Yis** Home Analyze Statistics Tutorial About

Filter by antigen-producing organism [?](#)

Filter by germline classification [?](#)

Filter by literature keywords [?](#)

User defined sequences:

Variable domain sequences [?](#)

CDR sequences [?](#)

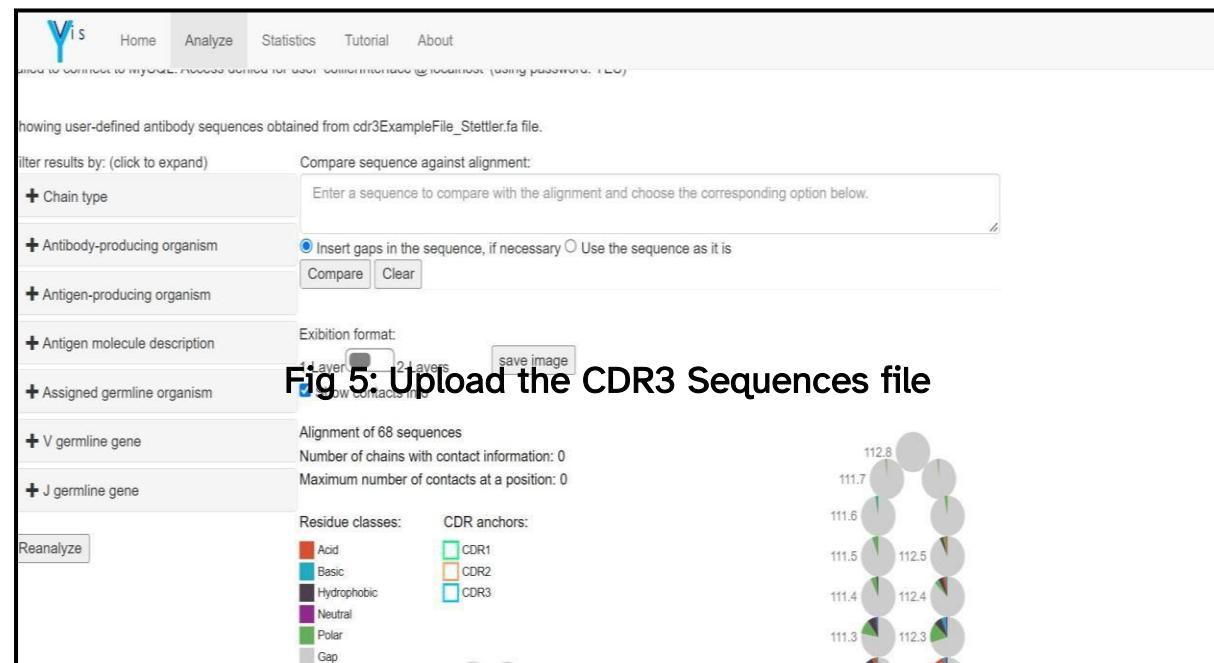
Choose a CDR:  CDR1 (Max 12 aa)  CDR2 (Max 10 aa)  CDR3 (unlimited)  
 Upload a fasta file containing the CDR sequences:

[Choose File](#) cdr3Example...le\_Stettler.fa

[Analyze](#) Example file

IMGT/DomainGapAlign results file [?](#)

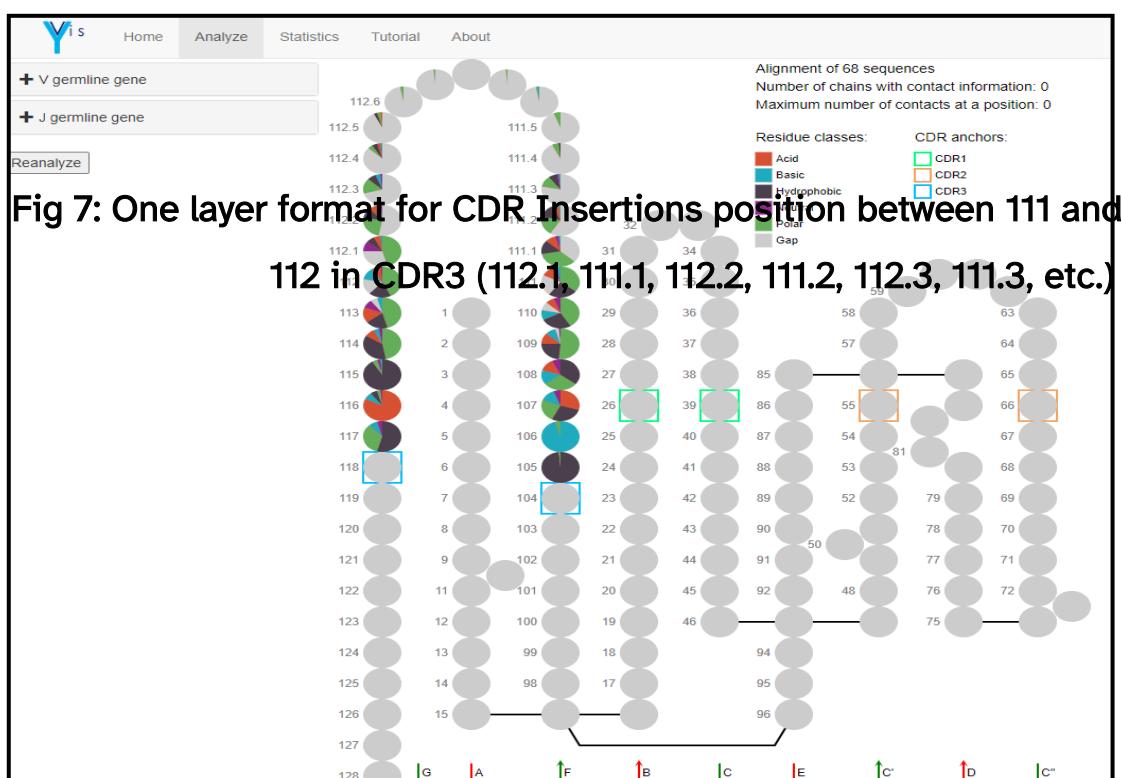
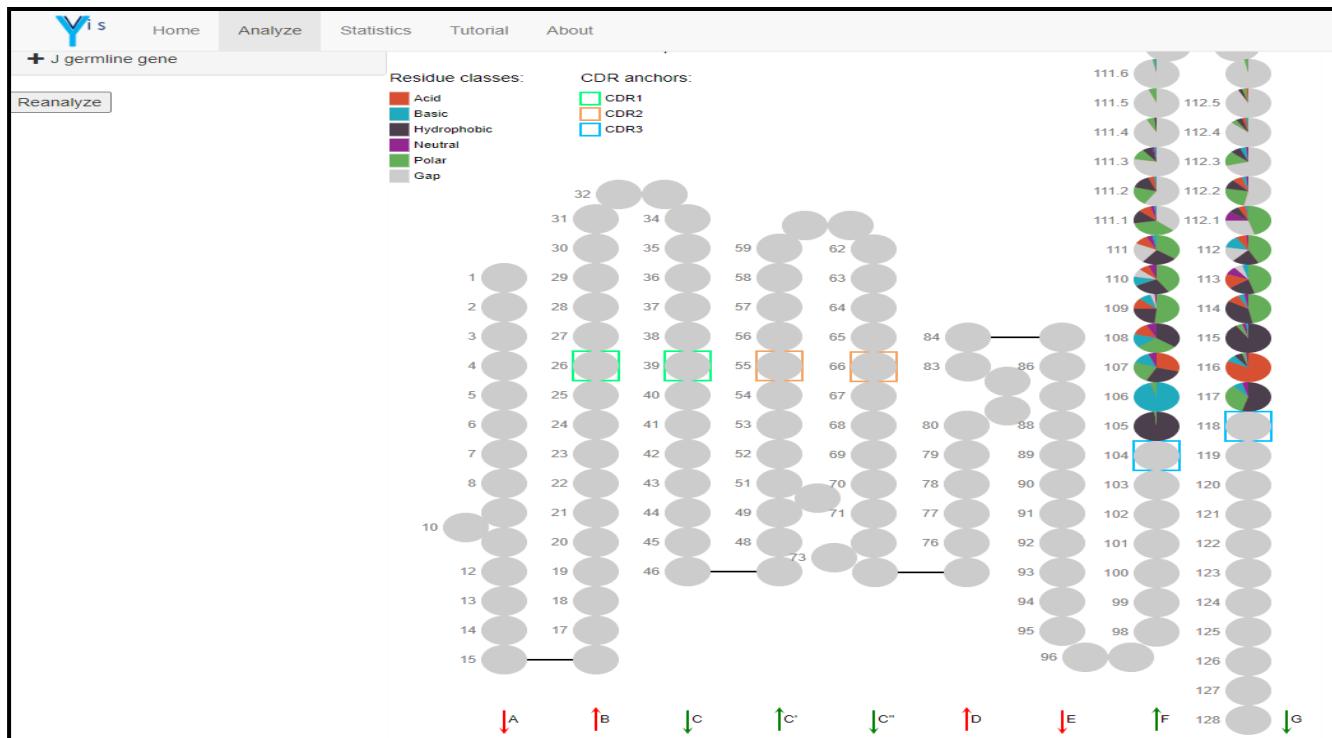
IMGT/HighV-QUEST results file [?](#)



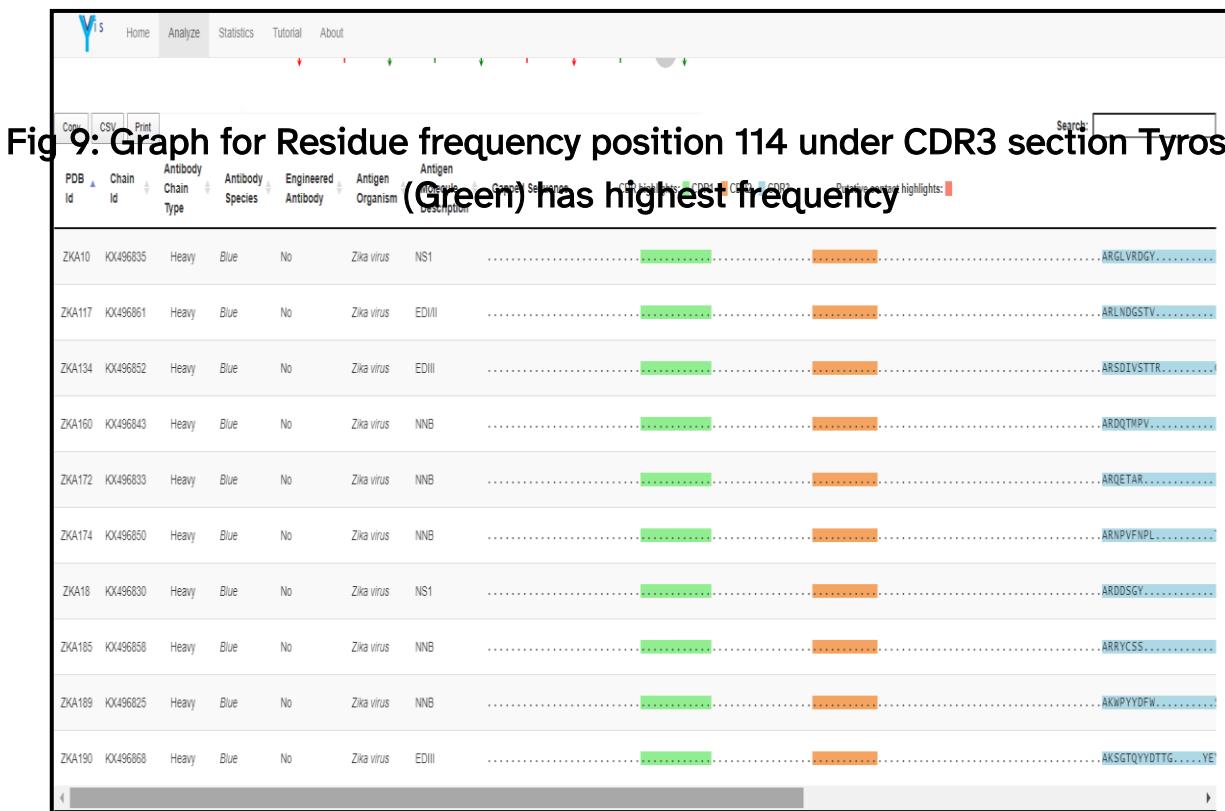
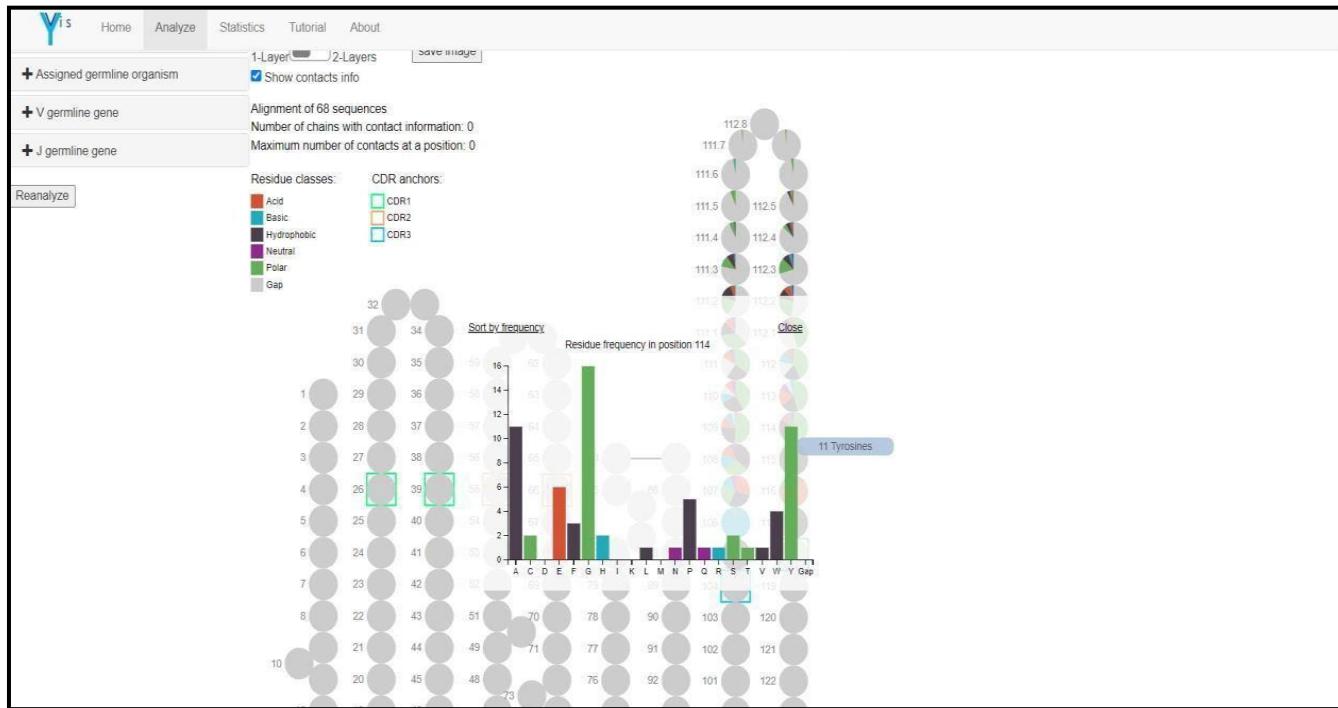
**Fig 6: Result page for CDR3 Sequences fileCDR1**

(Green): 27-38 CDR2(Orange):56-65

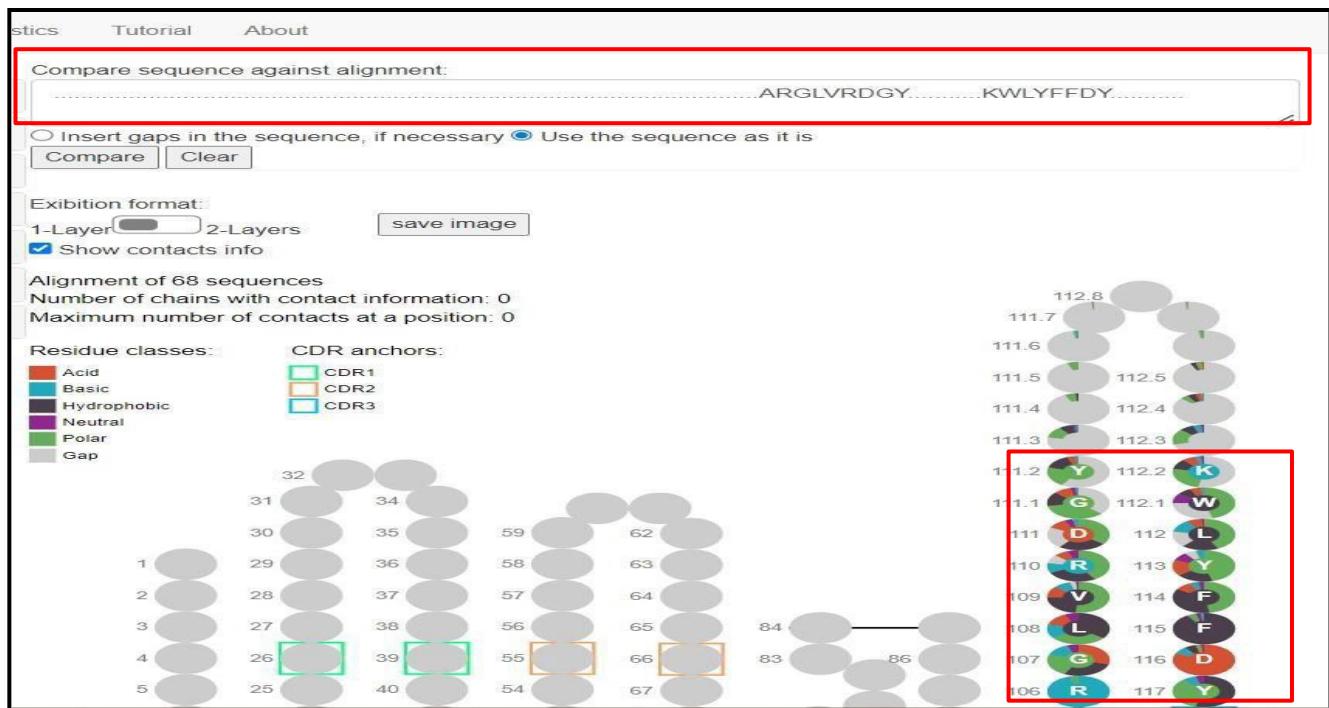
CDR3(Blue):105-118



**Fig 8: Two-layer format for CDR**



**Fig 10: Sequence Alignment for CDR3 region**



**Fig 11: Comparing sequences for PDB ID ZKA10**

## RESULT:

Yvis will present a page containing the search criteria or input data at the top, and the filter options on the left. On the right side of the page, Yvis will present the input for the comparison feature and the Collier de Diamants visualization of sequences. On the bottom of the page, Yvis presents a table with information on each analyzed sequence. These features are explained below.

### 1. Collier de Diamants interpretation

Collier de Diamants uses IMGT/Collier de Perles (Pearl Necklace) representation to present a multiple sequence alignment. Each Collier de Diamants position corresponds to a column in a classic multiple sequence alignment visualization, and it is summarized by a pie chart. In this chart, each “pie slice” (sector) represents the number of sequences with an amino acid of a specific class (defined by a colour) in that position and they are classified according to their chemical properties. Thus, green represents polar amino acids, blue represents basic red represents acidic, and black represents hydrophobic amino acids. In Yvis, gaps are in grey. The Collier de Diamants presents the most frequent classes in a clockwise orientation.

The Collier de Diamants shows the positions as in Collier de Perles, linking

sequences to their 3D structure. Squares indicate the CDR anchors, one position before the CDR start and one after the CDR end (i.e., green for CDR1, orange for CDR2, and blue for CDR3) and allow the quick visualization of the residues that compose each CDR. As the Collier de Diamants uses the IMGT numbering schema to align sequences, CDR1 corresponds to positions 27-38, CDR2 corresponds to positions 56-65, and CDR3 corresponds to position 105-117, regardless of the chain type (heavy or light). CDR3 may contain some insertions when longer than 13 amino acids. In this case, as indicated by the IMGT numbering schema, Yvis insert new position between positions 111 and 112, in the following order: 112.1, 111.1, 112.2, 111.2, 112.3, 111.3, etc.

Like the IMGT/Collier de Perles, Collier de Diamants can be presented in one or two layers. The two- layers version presents the variable domain strands in a position closer to the 3D structure, while the one- layer version has a representation closer to the variable domain sequence. To change the presentation from one to two layers (or vice versa), click on the switch button on the top of Collier de Diamants representation. The strands of the variable domain are identified by letters (A-G) and arrows at the bottom of Collier de Diamants visualization. Moreover, the arrow colour indicates the different sheets of the variable domain.

The Collier de Diamants representation allows visualizing position(s) with a conserved class of residues and their position in the 3D structure. This can be easily done because positions with a conserved residue class will present a dominant sector in the corresponding pie chart. Conversely, variable positions (based on the multiple sequence alignment) will be represented by pie charts with many different sectors.

To identify the amino acids, present in each position, click on each position to open a new chart with the detailed amino acid composition of that position. This is a classical bar chart where each bar represents an amino acid, and its colour corresponds to the amino acid class. The bar height represents the number of sequences that have this amino acid in that position. The user can hover the mouse pointer over the bar to see the exact number of amino acids. Bars can be sorted according to their height or the represented amino acid. By clicking on “Close”, the bar chart is closed, and the Collier de Diamants visualization is back.

Besides the visualization of a multiple sequence alignment, Collier de Diamants can display a quantitative attribute for each position, represented by circles in salmon around each pie chart. In the Yvis platform, this attribute is shown when Yvis database chains from structures of protein antigen-antibody complexes are analyzed. It represents the number of chains with a putative contact at that position. Yvis defines a putative contact when the distance between alfa-carbons of an antibody amino acid is shorter than 8Å. The radius of the contact circle (in salmon) around the pie chart of a given position will be proportionally bigger in function of the number of sequences with a putative contact in that position. By hovering the mouse pointer over a position in the pie chart, we can see the exactly number of antibody chains that have a putative contact in that position. The total number of analyzed sequences with putative contact information is indicated at the top of the Collier de Diamants visualization, as well as the maximum number of contacts at a position.

Contacts are important information in antibody analysis because the positions making putative contacts are usually related to antibody-antigen binding. We can show/hide contact information by selecting/unselecting the “Show contact info” box.

The Collier de Diamants visualization can be saved by clicking the “Save image” button. Yvis will generate a PNG or SVG image that can be downloaded.

## 2. Comparison tool

Uploaded a sequence to be compared with the multiple sequence alignment presented in the Collier de Diamants, Yvis will display, at the center of each pie chart that represents a position, a small circle with the inputted sequence amino acid corresponding to that position. This circle is coloured according to the colour schema used for the pie chart sectors. This allows the easy comparison of the sequence with the multiple sequence alignment, just by comparing the colour of

the small circle and that of the largest sector of the pie chart for that position. Thus, with this representation, divergent sequence positions in the multiple sequence alignment are represented by colours that are different from the one of the predominant slices.

## **CONCLUSION:**

The Yvis platform can be used in different types of antibody analysis. For example, the quick visualization of the most conserved or divergent positions in a set of related antibodies can guide antibody engineering and mutagenesis experiments. In antibody repertoire studies, the Collier de Diamants visualization, coupled with the sequence comparison feature, can be used to compare thousands of antibody sequences with a specific germline sequence. This can give to researchers some insights into the most important mutations that occurred during the antibody affinity maturation process. Therefore, the Yvis platform offers an environment for antibody sequence analysis that helps to formulate hypotheses concerning the key residues in the antibody structure or interactions and improves the understanding of the antibody properties.

## **REFERENCES:**

- 1) *Yvis Platform*. (n.d.). NCBI. Retrieved October 29, 2022, from  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC602444/>
  - 2) *Yvis Platform*. (n.d.). Retrieved October 29, 2022, from  
<http://bioinfo.icb.ufmg.br/yvis/>
-

**WEBLEM 6 INTRODUCTION  
TO AG-AB INTERACTION  
DATABASE (AG-AB DB)**

Antibodies are produced by vertebrates in response to antigens. Antigens are usually foreign molecules of invading pathogens. Antibodies are produced in billions of forms by B cells and are collectively referred to as immunoglobulins (abbreviated as Ig). The clonal selection theory states that all the antibodies produced by an individual B cell have the same antigen-binding site. Furthermore, every B cell produces a single species of antibody having a unique antigen-binding site.

An antibody molecule is a polymer of two light and two heavy chains. The two light chains are identical and are of a length of ~220 amino acids each. Similarly, the two heavy chains are identical with a typical length of ~440 amino acids each. The four chains are held together by various noncovalent and covalent (disulfide) bonds. Every light chain has one variable and one constant region, whereas heavy chains have one variable and two to three constant regions. As a result, two identical antigen-binding sites are formed by the N-terminal variable regions of a pair of light and heavy chains. The tail (Fc) and hinge regions are however formed by the constant regions of two heavy chains. The antigen-binding site of an antibody is referred to as a “paratope”.

There are five classes of antibodies such as IgA, IgD, IgE, IgG, and IgM, which are based on five types of heavy chains such as  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$ , and  $\mu$ . Each of these heavy chains is known to invoke a specific cascade of reactions upon binding to an antigen. However, there are only two types of light chains ( $\kappa$  or  $\lambda$ ) that pair with one of the heavy chains. Therefore, the type of light chain does not seem to affect the properties of the antibody, other than its specificity for the antigen.

**Properties of Antigen–Antibody Interactions:**

The binding of an antigen to an antibody is reversible, and both the molecules can exist independently.

The antigen–antibody interactions are thus mediated by many relatively weak, non-covalent forces

such as hydrogen bonds, hydrophobic interactions, van der Waals forces, and ionic interactions. Of all the forces, van der Waals forces are the weakest and can attract all kinds of molecules. Hydrogen or ion-dipole bonds are formed between oppositely charged atoms, whereas “hydrophobic” interactions are formed between atoms of nonpolar amino acids which do not form electric dipole. These weak forces are effective only when the antigen molecule is close enough to allow some of its atoms to fit into complementary niches on the surface of the antibodies. The attractive forces exerted by ionic and hydrophobic bonds help the molecules to overcome hydration energies. This leads to the expulsion of water molecules and

results in bringing the epitope and paratope closer. This spatial proximity facilitates van der Waals interactions. The overall strength of binding depends on goodness of fit between the epitope and paratope and the total area of contact between them.

Antigen-antibody interactions are highly specific, and understanding the molecular basis of the specificity has been one of the goals of immunology. A large number of high-resolution X-ray structures of several antigens have been solved in the native (uncompleted) form as well as in complex with antibody, and the data are archived in Protein Data Bank (PDB). Analyses of these structures have helped in understanding characteristics of both epitopes (antibody-binding site on antigen) and paratopes (antigen-binding site of antibody), which are complementary to each other and are relational entities.

### **Ag-Ab Interaction Database (AgAbDb):**

Antigen-Antibody Interaction Database (AgAbDb) is an immunoinformatics resource developed at the Bioinformatics Centre, University of Pune, and is available online at <http://bioinfo.net.in/AgAbDb.html>. Antigen-antibody interactions are a special class of protein-protein interactions that are characterized by high affinity and strict specificity of antibodies towards their antigens. Several co-crystal structures of antigen-antibody complexes have been solved and are available in the Protein Data Bank (PDB). AgAbDb is a derived knowledgebase developed with an objective to compile, curate, and analyze determinants of interactions between the respective antigen-antibody molecules. AgAbDb lists not only the residues of binding sites of antigens and antibodies, but also interacting residue pairs. It also helps in the identification of interacting residues and buried residues that constitute antibody-binding sites of protein and peptide antigens. The Antigen-Antibody Interaction Finder (AAIF), a program developed in-house, is used to compile the molecular interactions, viz. van der Waals interactions, salt bridges, and hydrogen bonds. A module for curating water-mediated interactions has also been developed. In addition, various residue level features, viz. accessible surface area, data on epitope segment, and secondary structural state of binding site residues, are also compiled. Apart from the PDB numbering, Wu-Kabat numbering and explicit Definitions of complementarity-determining regions are provided for residues of antibodies.

The molecular interactions can be visualized using the program Jmol. AgAbDb can be used as a benchmark dataset to validate algorithms for prediction of B-cell epitopes. It can as well be used to improve accuracy of existing algorithms and to design new algorithms. AgAbDb can also be used to design mimotopes representing antigens as well as aid in designing processes leading to humanization of antibodies. A user-friendly web-enabled interface for AgAbDb (<http://115.111.37.206:8080/agabdb2>) has been designed and tested for all the web browsers. A “quick search box” is provided on all the web pages of the interface. The “quick search” supports the data base search using the PDB ID or the keywords. This in turn opens a page listing the search results. AgAbDb can be browsed by clicking on the PDB ID. The search results page also provides links to view the antigen-antibody interactions archived in AgAbDb using Jmol, to view the corresponding complex at the RCSB PDB site, and to download the file from the RCSB PDB.

### **AgAbDb: Data Formats and Displays:**

AgAbDb archives data of antigens, antibodies, and molecular interactions under eight categories, viz. Summary, IR: Epitope Paratope, IR: Epitope Segments, Binding Site: IR + BR, Atomic Level Interactions, Water-Mediated Interactions, View Interactions, and Statistics. The tables displaying interaction data under each of these eight categories can be exported as Excel files. The complex, binding site residues of antigen and antibody along with subsets of various interactions can be visualized using Jmol (<http://www.jmol.org/>

). The snapshots of screens based on eight categories are shown in FiAgAbDb records for a complex of the antibody. NC10 Fv and neuraminidase from influenza virus ([ 39], PDB ID: 1A14) are shown. AgAbDb uses PDB ID as a unique identifier to archive interaction data.

This section provides overall information of the complex, the antigen, and the antibody. Data are curated from the PDB and typically lists PDB ID, PubMed ID, resolution, release date, and citation information. The data on antibody includes name, class/ type, scientific and common names of the source, and the PDBchain identifiers for light and heavy chains. The data on antigen includes name, scientific and common names of the source, an antigen type (protein or peptide), and the PDB chain identifier.

### **Epitope–Paratope:**

This section lists all the interacting residues of the binding sites. The residues of antibody (paratope) that are interacting with the residues of antigen (epitope) are provided. For example, the numbers of interacting residues of paratope (NC10 Fv) and epitope (neuraminidase) are 12 and 17, respectively (PDB ID: 1A14). The paratope residues are listed with chain type (heavy or light chain), PDB numbering, and Kabat numbering. It is preferred to have both the numbering systems and their equivalence known as far as antibody numbering is concerned. The table also lists equivalence between the interacting residues of the antigen and antibody. This is one of the unique features of AgAbDb. It is very useful and facilitates interesting analyses as a residue may interact with one or more residues. The residues of both antigen or antibody having minimum and maximum contacts can be identified. For example, Asn400 of the antigen interacts with two residues of CDR2 and one residue of CDR1 of heavy chain. Identification of such important residues or hot spots may have applications in mutation analysis, which is a prerequisite for designing

antigen scaffolds and/or peptide/subunit vaccines. Other immunoinformatics resources, viz. IEDB-3D and IMGT/3Dstructure-DB, do not provide the list of pairs of interacting residues in an explicit fashion.

Generation of such a list using these resources calls for processing of the data through multiple steps. The “IR: Epitope-Paratope” table also lists secondary structural states of interacting residues of antigen, which are obtained from DSSP assignments [ 37]. Analysis pertaining to preference of secondarystructural states of antigens has always been the area of interest and has been used effectively in epitope prediction programs.

### **Binding Site: IR + BR:**

This section lists all the residues of the respective binding sites of the antigens and antibodies. Separate tables for antigen and antibody molecules are generated. In addition to the interacting residues, several residues of epitope are buried under the footprint of an antibody. Such residues are a part of the binding site scaffold and may not directly interact with residues of CDRs and LDRs of an antibody. Similarly, CDR and LDR also have only a few interacting residues while the other residues forming the scaffold, though not interacting explicitly, are used to calculate the area of interface of antibody with antigen.

### **Atomic Level Interactions:**

This section displays various non-covalent interactions between residues of the epitopes and paratopes. For example, NC10 antibody (PDB ID: 1A14) has about 107 non-covalent interactions of the types such as salt bridges (1), hydrogen bonds (7), short van der Waals interactions (2), and van der Waals interactions (97). These interactions are curated using the program AAIF.

### **Statistics:**

This section provides a residue-wise summary of various inter actions. Separate tables are provided forthe antigen (epitope) and antibody (paratope), which list the residues that contribute mamaximallyo the antigen- antibody interactions. This section provides a summary of interactions for every residue and includes data on the total number of interactions, which is a sum of the total number of hydrogen bonds, van der Waals interactions, and salt bridges. The table also lists the total number of residues (from the partner mol ecule) with which a given residue is interacting. This section also helps to quickly enlistwhich of the 20 amino acids are parts of the paratope and epitope. For example, NC10 antibody CDRs have only 7 (S, T, N, F, L, D, Y) amino acids whereas the near aminidase epitope has 11 (S, K, T, N, G,A, D, I Y, P, W) amino acids as characterized in the complex 1A14 .

AgAbDb also helps in analyzing how every CDR partici pates in binding to the epitope. This utility is provided under the “Search” option on the main menu bar. Three CDRs on light chain are termed as LDR 1-

3.There are three LDRs (light chain) and three CDRs (heavy chain). Since the PDB

numbering may or may not be in accordance with the position of a given residue in sequence and/or Kabat scheme of numbering, AgAbDb provides equivalence between PDB and Kabat numbering. “CDR statistics” for NC10 antibody (PDB ID: 1A14) reveals that two of the six CDRs such as LDR2 and CDR1 do not participate in the antigen binding at all. The LDR1, LDR3, CDR2, and CDR3, respectively, have 2, 4, 3, and 3 residues interacting with various residues of the antigen. Of the 107 total interactions, 25, 34, 27, and 21 interactions are contributed by LDR1, LDR3, CDR2, and CDR3, respectively. Thus, AgAbDb can be used to perform various queries and to study the multiple aspects of antigen–antibody interactions.

**Antigen Antibody Interactions Database (AgAbDb)**  
**Bioinformatics Centre**  
**University of Pune, India**

Home   Theory   Search   Predict Epitope   Help   Contact Us

Search

Welcome to Antigen Antibody Interactions Database.

AgAbDb, a derived knowledge base, archives molecular interactions of protein and peptide antigens characterized by co-crystal structures. The interactions are compiled at two levels, viz. residue level and atomic level. The interactions are characterized using AAIF (Antigen-Antibody Interaction Finder) developed in-house. AAIF enlists various non-covalent interactions such as van der Waals, salt bridges, hydrogen bonds and short contacts using distance and geometry-based criteria. Apart from molecular interactions, AgAbDb also archives information pertaining to Ag and Ab description, definition of Complementarity Determining Regions on light chain (LDR) and Complementarity Determining Regions on heavy chain (CDR) residues of the antibody with both, the PDB as well as Wu and Kabat numbering.

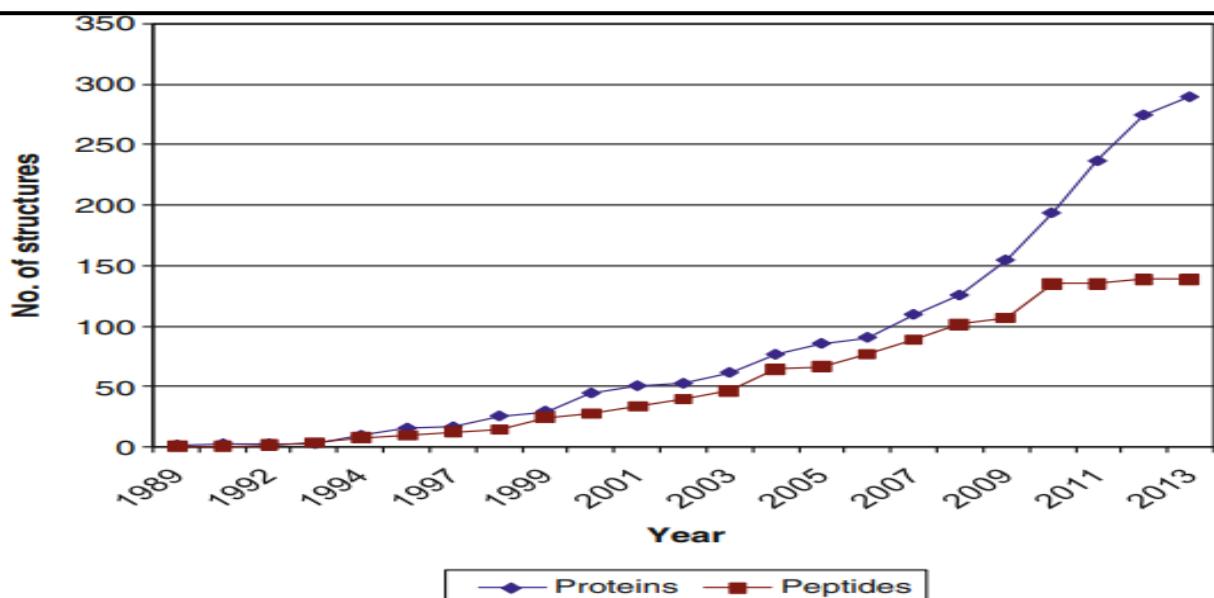
**AgAbDb Statistics**  
The database contains 427\* antigen-antibody complexes.  
• 289 Protein-Ab complexes  
• 138 Peptide-Ab complexes

\*as on Wed Aug 07 16:09:49 IST 2013

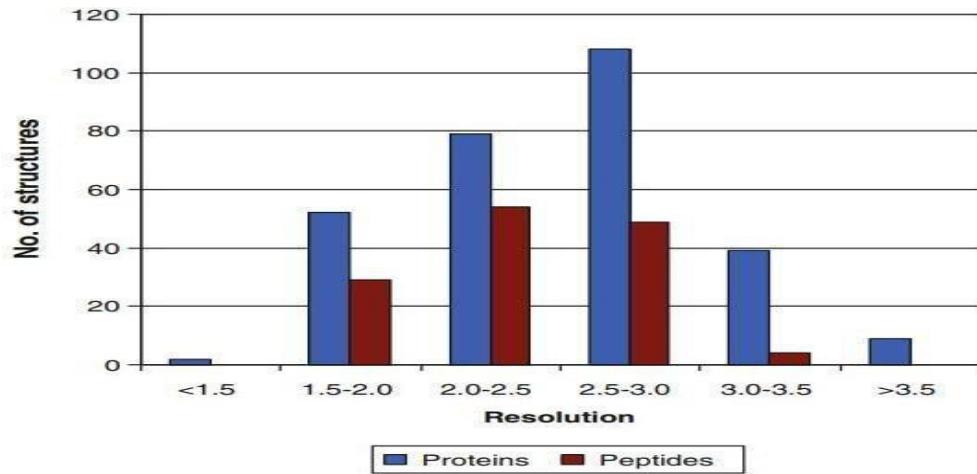
4FQI: Crystal Structure of Fab CR9114 in Complex with a H5N1 influenza virus hemagglutinin  
(click on image to view details)

© Bioinformatics Centre, University of Pune, India

**Fig. 1** A snapshot of the home page of AgAbDb



**Fig. 2** The growth of co-crystal structures of protein- and peptide-antibody complexes in AgAbDb



Resolution graph for 427 complexes

**Fig. 3** Distribution of co-crystal structures of antigen–antibody complexes based on resolution

**Antigen Antibody Interactions Database (AgAbDb)**  
Bioinformatics Centre  
University of Pune, India

**Search** Advanced Search Residue-wise Interaction: Ag & Ab

Residue-wise Interaction: CDR

of protein and peptide structures. The interaction level and atomic level. The (Antigen-Antibody) Interaction enlists various non-covalent bridges, hydrogen bonds etc., geometry-based criteria. Also archives information definition of Complementarity (LDR) and Complementarity (CDR) residues of the antibody Kabat numbering.

**AgAbDb Statistics**

Query to get quick summary about the interacting residues in CDR regions

The Number of Interactions for CDR Region of Antibody

CDR: L-CDR1 From: (Enter PBDID)

Total interacting residues L-CDR1 in 1a14 = 2  
Total interactions in L-CDR1 in 1a14 = 25  
Number of van der Waals interactions in L-CDR1 in 1a14 = 23

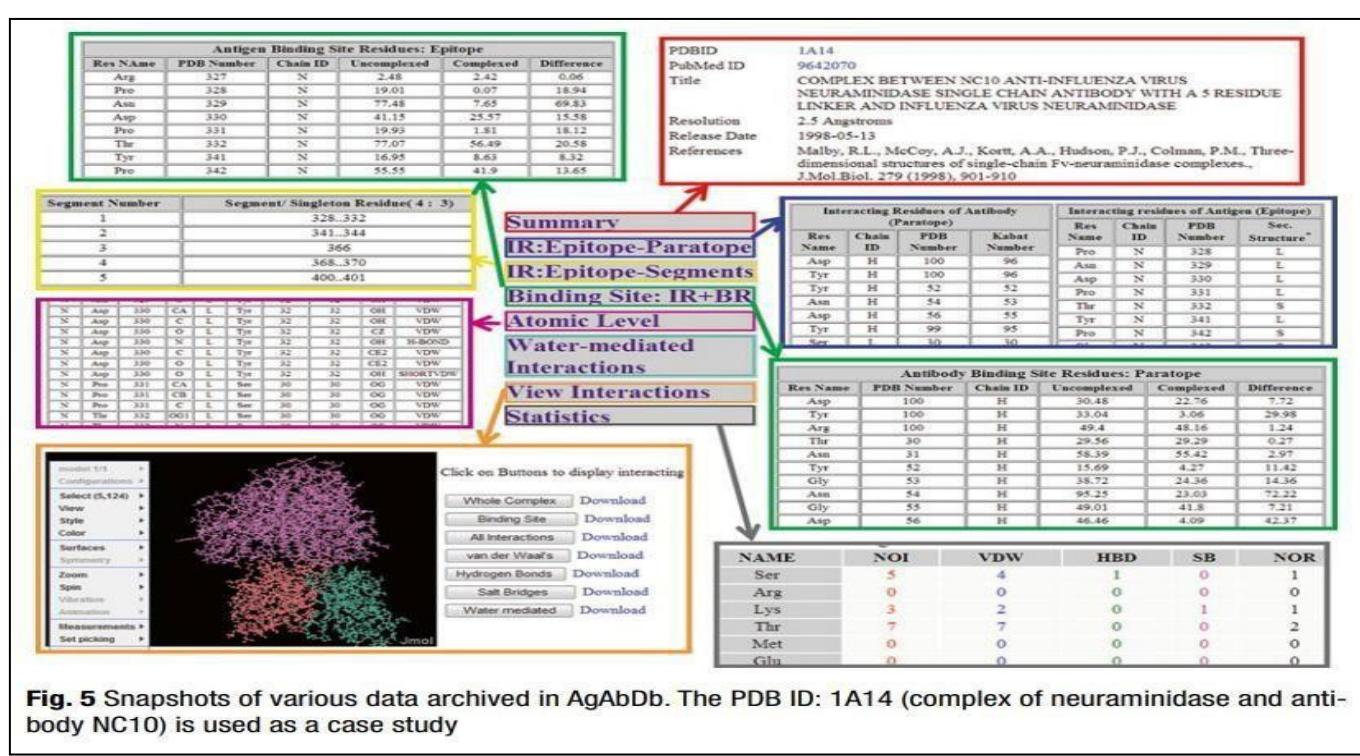
**Antibody Information**

The Number of Interaction for Antibody residue

Residue: Ser From: (Enter Pdb ID)

Total interactions by Ser in 1a14 = 1  
Ser interacting in 1a14 = 1  
Number of interactions by Ser in 1a14 = 5  
Number of van der Waals interactions by Ser in 1a14 = 5  
Number of Hydrogen Bond interactions by Ser in 1a14 = 0  
Number of Bridge interactions by Ser in 1a14 = 0

**Fig. 4** Snapshots of various search strategies in AgAbDb



**Fig. 5** Snapshots of various data archived in AgAbDb. The PDB ID: 1A14 (complex of neuraminidase and antibody NC10) is used as a case study

## RESULT:

Various methods and search analysis were studied while exploring the AG-Ab database. The analysis methods for various Ag-Ab interactions studies in AG-Ab database were

**Advanced search:** This analysis section gives us a complete summary and information about various interactions such as Atomic, water-mediated, Binding sites, IR- Epitope, Paratope and IR- Epitope segments. Residue- wise interaction; Ag &Ab: This analysis section gives us a complete information about the Antigen and Antibody Interaction along with the Number of interactions for antibody/antigen residue, Number of van der Waals interactions by residue in the query PDB ID, Number of Hydrogen Bond interactions by residue in the query PDB ID, Number of salt bridge interactions by residue in the query PDB ID in a tabular format. Residue-Wise interaction; CDR: This analysis section gives us a complex information about the number of interactions for CDR regions of Antibody along with a tabular format which contains information about the Antigen and Antibody with respect to information such as the Res name, Chain ID, PDB Number and Kabat Number respectively.

## CONCLUSION:

AgAbDb can be used as a benchmark dataset to validate algorithms for prediction of B-cell epitopes. It can as well be used to improve accuracy of existing algorithms and to design new algorithms. AgAbDb can also be used to design mimotopes representing antigens as well as aid in designing processes leading to humanization of antibodies. The Database is currently unavailable as it is under maintenance.

## REFERENCE:

Kulkarni-Kale, U., Raskar-Renuse, S., Natekar-Kalantre, G., & Saxena, S. A. (2014). Antigen- Antibody Interaction Database (AgAbDb): a compendium of antigen- antibody interactions. Methods in molecular biology (Clifton, N.J.), 1184, 149–164. [https://doi.org/10.1007/978-1-4939-1115-8\\_8](https://doi.org/10.1007/978-1-4939-1115-8_8)

## WEBLEM 7 INTRODUCTION TO IEDB DATABASE

**AIM:**

Introduction to IEDB Database

**INTRODUCTION:**

The IEDB was established in 2004, and over the past 10 years our team has manually curated almost 16 000 published manuscripts and processed 200 direct submissions. As a result, detailed experimental data regarding more than 120 000 epitopes are now freely and easily accessible to the scientific community via most web browsers as a web-based interface. In addition, if one wishes to view 3D structural data using the Epitope Viewer application, Java 6 or 7 is required. The IEDB's primary curation focus is on data from scientific publications available in PubMed focused on infectious diseases, allergy, autoimmunity and transplantation. Excluded from the primary scope are HIV-derived epitopes captured in the LANLdatabase and cancer epitopes for which there is no resource currently available due to lack of support for such a resource by the National Institutes of Health. As an exception, all publications describing the 3D structure of an epitope in complex with its adaptive immune receptor or major histocompatibility complex(MHC) molecule are included regardless of origin of the epitope in order to provide a complete dataset of this particularly valuable type of information. Details describing the curation process put in place and followed by the curation team, including quality controls for accuracy and consistency, have been discussed previously

The IEDB houses epitope-specific experimental assays. That is, every assay reflects the binding of an epitope-specific T cell receptor (TCR), antibody or MHC molecule to an experimentally tested antigen or epitope. The structure entered as the epitope is limited to the exact entity that was actually tested in the assay or was clearly deduced to be the epitope by the authors. In many cases this is not the minimal epitope and may not be limited to the contact residues of the epitope, but is rather a region containing the epitope. The fields of the IEDB describe the details of these experiments in great detail. First, the epitope structure is designated as either peptidic or non-peptidic. Peptidic epitopes are described by their linear amino acid sequence or as discontinuous amino acids by position within their source protein.

**WORKING:**

IEDB Database icons were designed to highlight the main search components used for epitope related data. Icons were chosen based on a survey of scientists asked to identify the most relevant icon from a set to represent each major search parameter. These icons are similar to ones for hotel, airfare, or car rental on a travel web site, distinguishing the major types of searches possible. These search sections also

serve to restrict search terms to specific database fields and help guide the user as to the types of data that the IEDB contains. For example, in the ‘Host’ section, a variety of hosts including humans, rodents, non- human primates, and an additional nine commonly studied species are presented.

Peptidic epitopes having 3D structural data are described by the residues found to contact the antibody, TCR or MHC molecule. Non-peptidic epitopes are manually curated by staff from the ChEBI team (3) who annotate the complete molecular structures using SMILES annotation. If the epitope was derived from a protein or a larger non-peptidic structure, these are also provided along with the organism in which these structures are found.

**Welcome**

The IEDB is a free resource, funded by a contract from the National Institute of Allergy and Infectious Diseases. It offers easy searching of experimental data characterizing antibody and T cell epitopes studied in humans, non-human primates, and other animal species. Epitopes involved in infectious disease, allergy, autoimmunity, and transplant are included.

The IEDB also hosts tools to assist in the prediction and analysis of B cell and T cell epitopes.

[Learn More](#)

**Summary Metrics**

Peptidic Epitopes	118,403
Non-Peptidic Epitopes	2,035
T Cell Assays	246,033
B Cell Assays	168,187
MHC Ligand Assays	276,441
Epitope Source Organisms	3,148
Restricting MHC Alleles	668
References	15,830

[Provide Feedback](#) | [Help Request](#) | [Solutions Center](#)

**START YOUR SEARCH HERE**

**A**

**Epitope**

- Any Epitopes
- Linear Epitope
- Discontinuous Epitopes
- Non-peptidic Epitopes

Ex: SARS-CoV  
Exact Matches

**Assay**

- Positive Assay Only
- T Cell Assays
- B Cell Assays
- MHC Ligand Assays

**Antigen**

Organism:  
Ex: influenza, peanut

Antigen Name:  
Ex: core, capsid, myosin

**MHC Restriction**

- Any MHC Restriction
- MHC Class I
- MHC Class II
- MHC Nonclassical

**Host**

- Any Host
- Humans
- Rodents
- Non-human Primates
- Other Common Hosts

**Disease**

- Any Disease
- Infectious Disease
- Allergic Disease
- Autoimmune Disease
- Transplant Disease

[Reset](#) [Search](#)

**B**

**Epitope Analysis Resource**

**T Cell Epitope Prediction**

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome, TAP)
- MHC I Immunogenicity

**B Cell Epitope Prediction**

Predict linear B cell epitopes using:

- Antigen Sequence Properties
- Predict discontinuous B cell epitopes using antigen structure via:
- Solvent-accessibility (Disotope)
- Protrusion (EpiPro)

**Epitope Analysis Tools**

Analyze epitope sets of:

- Population Coverage
- Conservation Across Antigens
- Clusters with Similar Sequences
- Location in 3D Structure of Antigen

Data Last Updated: July 27, 2014

**Fig 1: The IEDB 3.0 home page has the most commonly used search parameters centered on the page, shown in box (A), with the highly used analysis tools made more prominent, shown in box (B).**

Once a query such as the one populated in Figure 1 has been executed, the search results are presented on a new page with the current search filters displayed at the top of the results table (Figure 1, box A). Any filter can be removed by a single click on the 'X' next to each parameter. The amount of data present within each of the result set types of Epitopes, Antigens, Assays and References are conveyed by counts and displayed as tabs that allow the user to easily navigate between them (Figure 1, box B). As shown in Figure 1 box C, a search panel added to the left side of the page allows the current result set to be further refined by adding search parameters or to run a new query entirely. These search panels contain the functionality present on the home page plus several additional search features, some of which were previously only present in the IEDB 2.0 'Advanced Search', such as the 'Assay Types.' We plan to continuously monitor the usage of each search parameter to identify additional fields that should be added to or removed from the search panel on the results page.

**IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE**

Home Specialized Searches Analytic Resource Keyword Search Search

**Update Filters**

**Epitope** 

- Any Epitopes
- Linear Epitope
- Discontinuous Epitope
- Non-peptidic Epitope

**Antigen** 

Organism: Escherichia coli

Antigen Name: Ph p 1 ★★ 

**Assay** 

- Positive Assay Only
- T Cell Assays
- B Cell Assays
- MHC Ligand Assays

**MHC Restriction** 

- Any MHC Restriction
- MHC Class I
- MHC Class II
- MHC Nonclassical
- Specific MHC Restriction

**Host** 

- Any Host
- Humans

**Current Filters:**  Positive Assays Only  Host: Homo sapiens (Human)  Disease: Allergic Disease  Antigen: Ph p 1 ★★ **A**

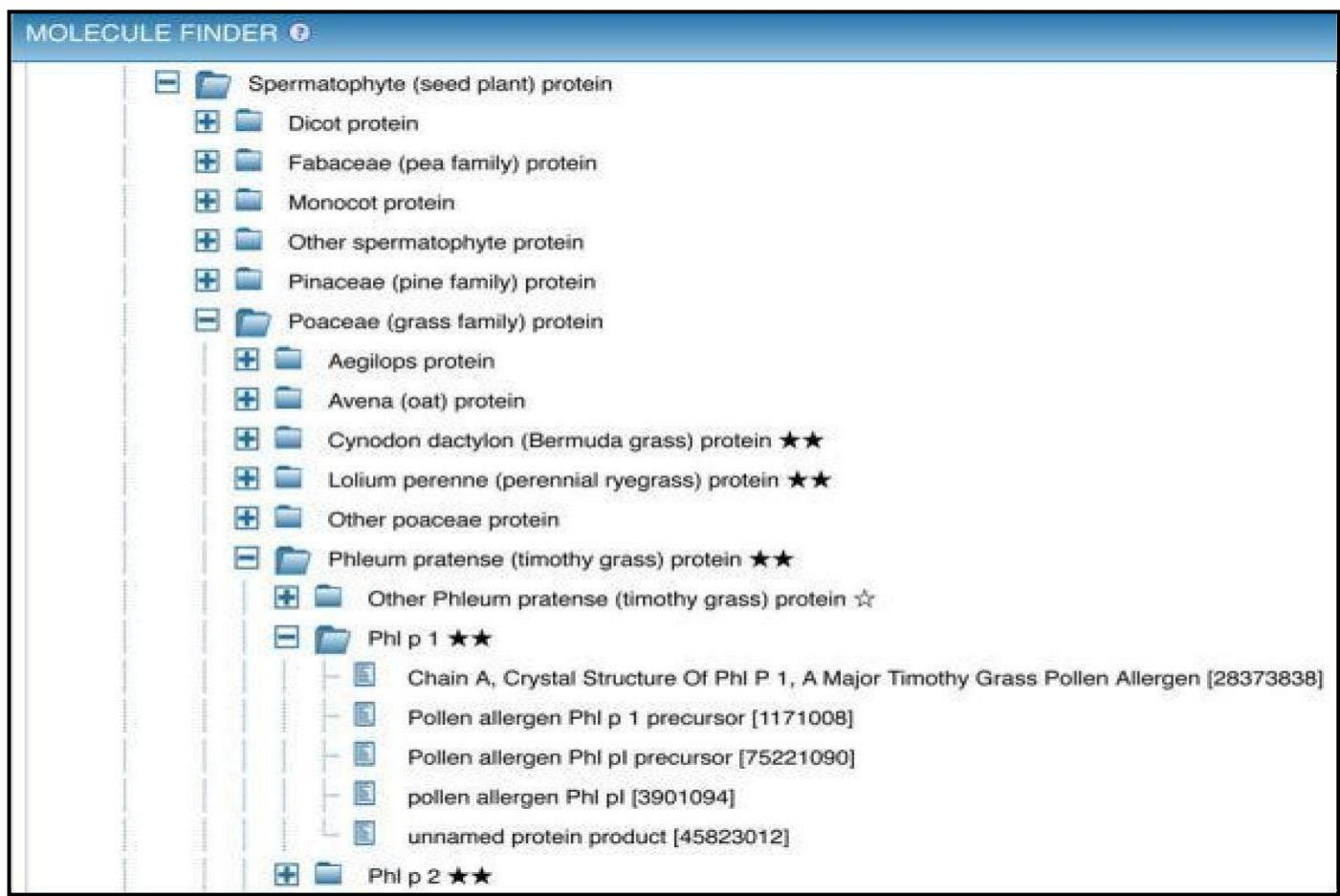
Epitopes (421)	Antigens (5)	Assays (418)	References (10)
Go To Records Starting At: 1 of 1,200 			
Export Epitopes Results 			
121 Records Found			
Details  Epitope 	Antigen 	Organism 	# References  # Assays 
15558 FEIKCTKIPPEACS	Ph p 1	Phleum pratense (timothy grass)	3 15
74106 VHFQDLSGHAFGA	Ph p 1	Phleum pratense (timothy grass)	3 8
127373 WGAAWRIIDTPDKLTG	Ph p 1	Phleum pratense (timothy grass)	3 5
1493 AGEELQFPRIVK	Ph p 1	Phleum pratense (timothy grass)	2 10
39127 APIVYHDLISGHAFGM	Ph p 1	Phleum pratense (timothy grass)	2 2
10750 DVWWVDRKEKGK	Ph p 1	Phleum pratense (timothy grass)	2 9
12313 EDWKWADTSYESK	Ph p 1	Phleum pratense (timothy grass)	2 9
13855 EOKURSAESEEL	Ph p 1	Phleum pratense (timothy grass)	2 9
20115 GHAFGAMMAKKGD	Ph p 1	Phleum pratense (timothy grass)	2 8
24100 HITDNEEDHAPYHIFDLSGHA	Ph p 1	Phleum pratense (timothy grass)	2 2
25389 IAPYHFDLSDKA	Ph p 1	Phleum pratense (timothy grass)	2 16
29549 WRIIDTPDKLTG	Ph p 1	Phleum pratense (timothy grass)	2 3
33539 KSTIWYQHPTDAAG	Ph p 1	Phleum pratense (timothy grass)	2 4
35586 LELOPFRIVKCKY	Ph p 1	Phleum pratense (timothy grass)	2 5
39205 LRBAGELELFPR	Ph p 1	Phleum pratense (timothy grass)	2 9
43602 NEEPIAPYHIFDLSGHA	Ph p 1	Phleum pratense (timothy grass)	2 2
47298 PEGTKYTFPHVEK	Ph p 1	Phleum pratense (timothy grass)	2 5
48043 PKDNKGACGTYKD	Ph p 1	Phleum pratense (timothy grass)	2 11
48725 PNLYLALLVYKVN	Ph p 1	Phleum pratense (timothy grass)	2 10
64695 TKUTTRHVEKGSH	Ph p 1	Phleum pratense (timothy grass)	2 5
66582 TTEGQTTKEAQD	Ph p 1	Phleum pratense (timothy grass)	2 4
69062 VIEPDWKRADTSY	Ph p 1	Phleum pratense (timothy grass)	2 9
71874 VVVAITDNEEIP	Ph p 1	Phleum pratense (timothy grass)	2 3
72440 WGAAWRIIDTPDK	Ph p 1	Phleum pratense (timothy grass)	2 3
73270 WYGSKPTGAGPKD	Ph p 1	Phleum pratense (timothy grass)	2 10

**Fig 2: New results presentation format shows current search filters in box (A), counts returned per data type in box (B) and the new left search panel allowing for continued refinement or editing of one's query, such as by the epitope source, in box (C).**

In addition to the query interface, the presentation of the results has been modified as well. Query results are grouped in four tabs: Epitopes, Antigens, Assays and References that match the current search criteria (Figure 2, box B). These different units of information reflect that some users want to utilize the IEDB as, for example, a way to explore the literature (on the reference tab), while others want to see which specific proteins in an organism have been studied for immune reactivity (on the antigen tab). The amount of data hosted in the IEDB has grown dramatically in the last few years, so that typical queries retrieve a very large number of epitopes. To make sure the most relevant epitopes are immediately visible, results are now sorted by how much information is available, such as the number of references with relevant data, as shown in Figure 2, rather than alphabetically, as was previously done. In addition to the left search panel, users can click on an epitope structure or its source to further narrow the result, using a new ‘filter’ icon present in the results table. Another noteworthy enhancement in the IEDB 3.0 is a new ‘Antigen’ tab which displays all epitopes that belong to the same antigen in one row.

The Molecule Finder has two top-level branches for peptidic and non-peptidic epitopes. Non-peptidic epitopes are assigned to sources in ChEBI and displayed using the ChEBI hierarchy. Peptidic epitopes derived from proteins occurring in

nature have their specific source protein identified by GenPept ([5](#)) entries. The variety of distinct sequences represented in GenPept (e.g. the five versions of Phl p I is necessary and reflective of the heterogeneity of proteins within individual species; however, the large number of entries and lack of standardized nomenclature previously overwhelmed users, and made it difficult to obtain all epitopes belonging to a single antigen.



**Fig 3:** The Molecule Finder provides a hierarchical organization of proteins that allows narrowing the search to epitopes derived from a specific antigen, such as the common allergen Phl p 1. The reference proteome protein ‘Phl p 1’ is the parent of five individual GenPept entries for this protein from Timothy grass.

## DISCOTOPE SERVER 1.1

DiscoTope is the first method to focus explicitly on discontinuous epitopes. We show that the new structure-based method has a better performance for predicting residues of discontinuous epitopes than methods based solely on sequence information, and that it can successfully predict epitope residues that have been identified by different techniques. DiscoTope detects 15.5% of residues located in discontinuous epitopes with a specificity of 95%. At this level of specificity, the conventional Parker hydrophilicity scale for predicting linear B-cell epitopes identifies only 11.0% of residues located in discontinuous epitopes. Predictions by the DiscoTope method can guide experimental epitope mapping in both rational vaccine design and development of diagnostic tools, and may lead to more efficient epitope identification.

Discontinuous epitopes, B-cell epitope, antibody, vaccine design, protein structure, antigen, accessibility, hydrophilicity.

**Table view lists following columns:**

**Chain ID:** The chain id of the protein chain used in prediction (specified by the user)

**Residue ID:** PDB Residue id

**Residue Name:** Name of the residue

**Contact Number:** The residue contact number is the number of C $\alpha$  atoms in the antigen within a distance of 10 Å of the residue's C $\alpha$  atom. A low contact number correlates with localization of the residue close to the surface or in protruding regions of the antigen's structures.

**Propensity Score:** This score tells you about the probability/tendency of being part of an epitope for that particular residue. The propensity is reflected in amino acid epitope log-odds ratios, which were calculated on a set of 75 antigens. The propensity score is calculated by sequentially averaging epitope log-odds ratios within a window of 9 residues. Then the scores are summed up based on the proximity in the 3D structure of the antigen. For any given residue, the sequentially averaged log-odds scores from all residues within 10Å are summed to give the propensity score.

**Discotope Score:** This score is calculated by combining the contact numbers with propensity score. DiscoTope score above the threshold value indicates positive predictions and that below the threshold value indicates negative predictions.

Fig 4: Homepage of IEDB Database



The IEEDB has just launched its updated 3D viewers! Learn more via our help article [here](#).

### Welcome

The Immune Epitope Database (IEEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEEDB also hosts tools to assist in the prediction and analysis of epitopes.

[Learn More](#)

### START YOUR SEARCH HERE

**Epitope** ?

- Any
- Linear peptide
- Discontinuous
- Non-peptidic

Exact M V Ex: SIINFEKL

**Assay** ?

- T Cell
- B Cell
- MHC Ligand

Ex: neutralization   
Outcome:  Positive  Negative

**Epitope Source** ?

Organism

Ex: influenza, peanut

Antigen

Ex: core, capsid, myo

**MHC Restriction** ?

- Any
- Class I
- Class II
- Non-classical

Ex: HLA-A\*02:01

**Host** ?

Any

**Disease** ?

Any

### Epitope Analysis Resource

**T Cell Epitope Prediction** ?

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome,TAP)
- MHC I Immunogenicity

**B Cell Epitope Prediction** ?

Predict linear B cell epitopes using:

Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- EliPro

**Epitope Analysis Tools** ?

Analyze epitope sets of:

Fig 5: B Cell Epitope Prediction section: Discotope server option

## IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

### DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID  
Or upload a PDB file   
 No file chosen

Step 2: Please enter PDB Chain ID

Step 3: Select version

© 2005-2022 | [IEDB Home](#) | [Help](#) | [Contact](#)  
Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services.

Fig 6: Homepage of DiscoTope

## TEPITOOL:

Computational prediction of T-cell epitope candidates is currently being used in several applications including vaccine discovery studies, development of diagnostics and removal of unwanted immune responses against protein therapeutics. There have been continuous improvements on the performance of MHC binding prediction tools but their general adoption by immunologists has been slow due to the lack of user-friendly interfaces and guidelines. Current tools only provide minimal advice on what alleles to include, what lengths to consider, how to deal with homologous peptides and what cutoffs should be considered relevant.

This protocol provides step-by-step instructions with necessary recommendations for prediction of the best T-cell epitope candidates in line with the newly developed online tool called TepiTool. The TepiTool, part of IEDB, provides some of the top MHC binding prediction algorithms for number of species including humans, chimpanzees, bovines, gorillas, macaques, mice and pigs. The TepiTool is freely accessible at <http://tools.iedb.org/tepitool/>.

The binding of a peptide to an MHC molecule is necessary for its ability to activate T cell responses. Peptides bind MHC molecules in the “peptide binding groove”, forming a peptide- MHC complex which in turn is recognized by the T cell receptors. Peptides recognized by T cells are called epitopes. Epitopes bound to class I and class II MHC molecules are recognized by CD8<sup>+</sup> and CD4<sup>+</sup> T cells, respectively.

Generally, MHC binding prediction tools scan amino acid sequences to estimate the binding affinity of each component peptide to a specific MHC. MHC class I molecules have a binding groove that is closed at its ends, limiting the size of its ligands to roughly 8-11 residues in length. Class II molecules, on the other hand, have an open binding groove, allowing them to bind longer peptides, typically 12-20 residues in length. The strength of binding (affinity) of a peptide to an MHC molecule is an important factor that determines potential immunogenicity.

**Fig 7: Homepage of IEDB database**



The IEEDB has just launched its updated 3D viewers! Learn more via our help article [here](#).

### Welcome

The Immune Epitope Database (IEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEDB also hosts tools to assist in the prediction and analysis of epitopes.

[Learn More](#)

### Upcoming Events & News

AAI Exhibitor Booth	May 6-10
FOCUS Exhibitor Booth	June 21-24
<a href="#">Virtual User Workshop</a>	Oct 26-28

\* register [here](#)

[IEDB SARS-CoV-2 Epitope Analysis Videos](#)

### Summary Metrics

Peptidic Epitopes	1,539,160
-------------------	-----------

### START YOUR SEARCH HERE

#### Epitope [?](#)

- Any
- Linear peptide
- Exact M
- Discontinuous
- Non-peptidic



#### Assay [?](#)

- T Cell
- B Cell
- MHC Ligand



Ex: neutralization   
Outcome:  Positive  Negative

#### Epitope Source [?](#)

##### Organism

Ex: influenza, peanut



##### Antigen

Ex: core, capsid, myos

#### Host [?](#)

- Any



#### MHC Restriction [?](#)

- Any
- Class I
- Class II
- Non-classical



#### Disease [?](#)

- Any



### Epitope Analysis Resource

#### T Cell Epitope Prediction [?](#)

Scan an antigen sequence for amino acid patterns indicative of:

MHC I Binding

MHC II Binding

MHC I Processing (Proteasome,TAP)

MHC I Immunogenicity

#### B Cell Epitope Prediction [?](#)

Predict linear B cell epitopes using:

Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

Disotope

ElliPro

#### Epitope Analysis Tools [?](#)

Analyze epitope sets of:

**Fig 8: TCE Prediction under IEEDB Database**

## IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

### T Cell Epitope Prediction Tools

#### T Cell Epitopes - MHC Binding Prediction

These tools predict IC50 values for peptides binding to specific MHC molecules. Note that binding to MHC is necessary but not sufficient for recognition by T cells.

##### [Peptide binding to MHC class I molecules](#)

This tool will take in an amino acid sequence, or set of sequences and determine each subsequence's ability to bind to a specific MHC class I molecule.

##### [Peptide binding to MHC class II molecules](#)

This tool employs different methods to predict MHC Class II epitopes, including a consensus approach which combines NN-align, SMM-align and Combinatorial library methods.

##### [TepiTool](#):

The Tepitool provides prediction of peptides binding to MHC class I and class II molecules. Tool is designed as a wizard with 6 steps as described below. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per your specific needs. You can go back to previous steps to change your selection before submission of the job. Once you submit the job (at the end of step-6), you will not be able to make any more changes and will have to start the prediction all over again with updated input parameters.

#### T Cell Epitopes - Processing Prediction

These tools predict epitope candidates based upon the processing of peptides in the cell.

##### [Proteasomal cleavage/TAP transport/MHC class I combined predictor](#)

This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope.

**Fig 9: Homepage of TCE Prediction**

## **CONCLUSION:**

After catching up on the curation of in-scope journal articles from the past, the focus of IEDB development for the 3.0 release has shifted toward improving query and reporting interfaces. The goal of this release was to provide intuitive ways to extract biologically accurate information from the large amounts of data now stored in the IEDB. We have here described the main new elements of the 3.0 release, all of which were motivated by user feedback gathered over the years. We believe that such development focusing on the usability of the web site is equally important to the introduction of new capabilities which—while often more exciting to implement from a web site developer's perspective—have little value if they are not actually utilized by the user community.

## **REFERENCES:**

1. IEDB Analysis Resource.(n.d.). NCBI. Retrieved October 10, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4384014/>
2. Tepitool. (n.d.). NCBI. Retrieved October 10, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981331/>

**WEBLEM 7A DISCOTOPE  
SERVER 1.1**  
(URL:  
<https://www.iedb.org/>)

**AIM:** To predict B Cell epitope for 2Z91 using Discotope server 1.1.

### **INTRODUCTION:**

DiscoTope is the first method to focus explicitly on discontinuous epitopes. We show that the new structure-based method has a better performance for predicting residues of discontinuous epitopes than methods based solely on sequence information, and that it can successfully predict epitope residues that have been identified by different techniques. DiscoTope detects 15.5% of residues located in discontinuous epitopes with a specificity of 95%. At this level of specificity, the conventional Parker hydrophilicity scale for predicting linear B-cell epitopes identifies only 11.0% of residues located in discontinuous epitopes. Predictions by the DiscoTope method can guide experimental epitope mapping in both rational vaccine design and development of diagnostic tools, and may lead to more efficient epitope identification.

Discontinuous epitopes, B-cell epitope, antibody, vaccine design, protein structure, antigen, accessibility, hydrophilicity.

Table view lists following columns:

**Chain ID:**The chain id of the protein chain used in prediction (specified by the user)

**Residue ID:**PDB Residue id

**Residue Name:**Name of the residue

**Contact Number:**The residue contact number is the number of C $\alpha$  atoms in the antigen within a distance of 10 Å of the residue's C $\alpha$  atom. A low contact number correlates with localization of the residue close to the surface or in protruding regions of the antigen's structures.

**Propensity Score:**This score tells you about the probability/tendency of being part of an epitope for that particular residue. The propensity is reflected in amino acid epitope log-odds ratios, which were calculated on a set of 75 antigens. The propensity score is calculated by sequentially averaging epitope log-odds ratios within a window of 9 residues. Then the scores are summed up based on the proximity in the 3D structure of the antigen. For any given residue, the sequentially averaged log-odds scores from all residues within 10 Å are summed to give the propensity score.

**Discotope Score:**This score is calculated by combining the contact numbers with propensity score. DiscoTope score above the threshold value indicates positive

predictions and that below the threshold value indicates negative predictions.

Immunoglobulins (Ig) or antibodies are glycoproteins that are produced by plasma cells. B cells are instructed by specific immunogens, for, example, bacterial proteins, to differentiate into plasma cells, which are protein-making cells that participate in humoral immune responses against bacteria, viruses, fungi, parasites, cellular antigens, chemicals, and synthetic substances. The immunogen or antigen reacts with a B-cell receptor (BCR) on the cell surface of B lymphocytes, and a signal is produced that directs the activation of transcription factors to stimulate the synthesis of antibodies, which are highly specific for the immunogen that stimulated the B cell. Furthermore, one clone of B cell makes an immunoglobulin (specificity).

Using the crystal structures of anti-ciguatoxin antibody 10C9 Fab in ligand-free form and in complexes with ABCD-ring (CTX3C-ABCD) and ABCDE-ring (CTX3C-ABCDE) fragments of the antigen CTX3Cat resolutions of 2.6, 2.4, and 2.3 angstroms, respectively, we elucidated the mechanism of the interaction between the polycyclic ethers and the antibody. 10C9 Fab has an extraordinarily large and deep binding pocket at the center of the variable region, where CTX3C-ABCD or CTX3C-ABCDE binds longitudinally in the pocket via hydrogen bonds and van der Waals interactions. Upon antigen-antibody complexation, 10C9 Fab adjusts to the antigen fragments by means of rotational motion in the variable region. In addition, the antigen fragment lacking the E-ring induces a large motion in the constant region.

## METHODOLOGY:

1. Open the Homepage of IEDB database from google: <https://www.iedb.org/>
2. On the IEDB Database homepage, under the section B cell Epitope Predictio, Click on Discotope server option.
3. In the Discotope search panel enter the PDB Id 2Z91 along with the chain ID 'A'.
4. Submit the query and interpret the results.

## OBSERVATIONS:

**Fig 1: Homepage of IEDB Database**

The IEDB has just launched its updated 3D viewers! Learn more via our help article [here](#).

### Welcome

The Immune Epitope Database (IEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEDB also hosts tools to assist in the prediction and analysis of epitopes.

[Learn More](#)

### START YOUR SEARCH HERE

**Epitope**

Any  
 Linear peptide  
 Discontinuous  
 Non-peptidic

Exact M▼ Ex: SIINFEKL

**Assay**

T Cell  
 B Cell  
 MHC Ligand

Ex: neutralization   
 Outcome:  Positive  Negative

**Epitope Source**

Organism  
 Ex: influenza, peanut

Antigen  
 Ex: core, capsid, myosin

**MHC Restriction**

Any  
 Class I  
 Class II  
 Non-classical  
 Ex: HLA-A\*02:01

**Host**

Any

**Disease**

Any

### Epitope Analysis Resource

**T Cell Epitope Prediction**

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome, TAP)
- MHC I Immunogenicity

**B Cell Epitope Prediction**

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- ElliPro

**Epitope Analysis Tools**

Analyze epitope sets of:

Fig 2: B Cell Epitope Prediction section: Discotope server option

## IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

### DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID  
Or upload a PDB file

2Z91 (example: 1z40)  
 No file chosen

Step 2: Please enter PDB Chain ID

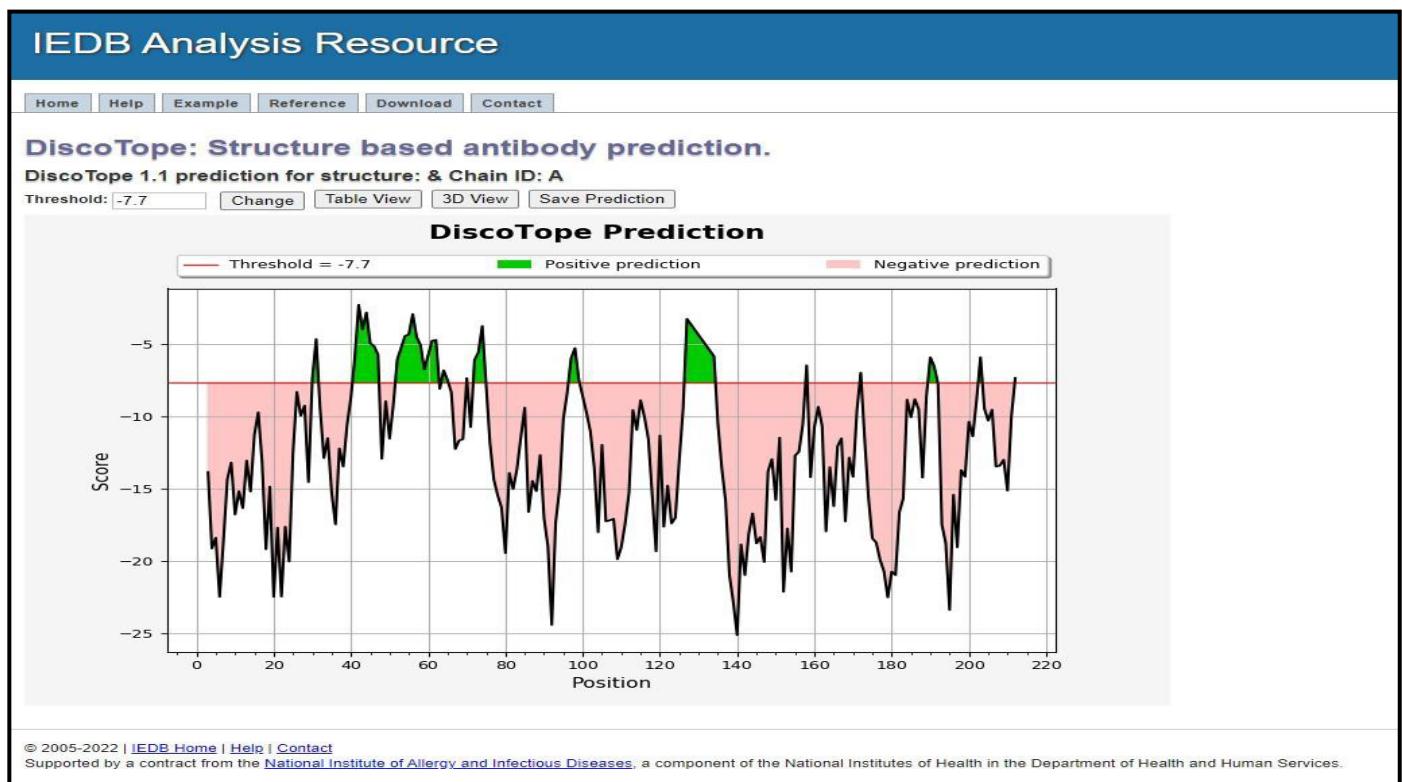
A

Step 3: Select version

1.1 ▼

© 2005-2022 | [IEDB Home](#) | [Help](#) | [Contact](#)  
 Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services.

Fig 3: Search for query (2Z91)in DiscoTope server



**Fig 4: Graphical result for 2Z91 Under DiscoTope server**

# IEDB Analysis Resource

[Home](#) [Help](#) [Example](#) [Reference](#) [Download](#) [Contact](#)

## DiscoTope - Result

**DiscoTope 1.1 prediction for structure: & Chain ID: A**

The positive predictions are displayed in green.

[Chart View](#) [3D View](#) [Save Prediction](#)

Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	Discotope Score
A	3	GLN	14	-6.874	-13.874
A	4	LEU	20	-9.148	-19.148
A	5	LEU	19	-8.914	-18.414
A	6	GLU	24	-10.487	-22.487
A	7	SER	18	-9.596	-18.596
A	8	GLY	16	-6.374	-14.374
A	9	PRO	15	-5.715	-13.215
A	10	ASP	22	-5.799	-16.799
A	11	LEU	22	-4.198	-15.198
A	12	VAL	22	-5.36	-16.36
A	13	LYS	18	-4.079	-13.079
A	14	PRO	20	-5.212	-15.212
A	15	SER	15	-3.813	-11.313
A	16	GLN	13	-3.236	-9.736
A	17	SER	17	-4.691	-13.191
A	18	LEU	24	-7.201	-19.201
A	19	SER	18	-5.878	-14.878

Fig 5: Table view for 2Z91 Under DiscoTope server

A	40	PHE	16	-0.711	-8.711
A	41	PRO	13	0.523	-5.977
A	42	GLY	8	1.685	-2.315
A	43	ASN	11	1.496	-4.004
A	44	LYS	10	2.155	-2.845
A	45	LEU	14	2.034	-4.966
A	46	GLU	14	1.822	-5.178
A	47	TRP	15	1.755	-5.745
A	48	MET	25	-0.449	-12.949
A	49	GLY	23	2.525	-8.975
A	50	TYR	26	1.441	-11.559
A	51	ILE	25	3.34	-9.16
A	52	HIS	18	2.909	-6.091
A	53	TYR	17	3.217	-5.283
A	54	ARG	15	3.031	-4.469
A	55	GLY	16	3.645	-4.355
A	56	THR	12	3.044	-2.956
A	57	THR	17	3.971	-4.529
A	58	ASN	17	3.415	-5.085
A	59	TYR	19	2.727	-6.773
A	60	ASN	14	1.2	-5.8
A	61	THR	11	0.703	-4.797
A	62	SER	10	0.258	-4.742

Fig 5.1: Table view for 2Z91 position prediction (Green Colour) Under DiscoTope server

IEDB Analysis Resource

Home Help Example Reference Download Contact

JSmol-Rendered PDB Structure

Chart View Table View Save Prediction

Model Structure

Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	DiscoTope Score	View
A	30	THR	20	2.574	-7.426	CPK
A	31	SER	10	0.32	-4.68	CPK
A	41	PRO	13	0.523	-5.977	CPK
A	42	GLY	8	1.685	-2.315	CPK
A	43	ASN	11	1.496	-4.004	CPK
A	44	LYS	10	2.155	-2.845	CPK
A	45	LEU	14	2.034	-4.966	CPK
A	46	GLU	14	1.822	-5.178	CPK
A	47	TRP	15	1.755	-5.745	CPK
A	52	HIS	18	2.909	-6.091	CPK
A	53	TYR	17	3.217	-5.283	CPK
A	54	ARG	15	3.031	-4.469	CPK
A	55	GLY	16	3.645	-4.355	CPK

**Fig 6: 3D view for 2Z91 Under DiscoTope server**

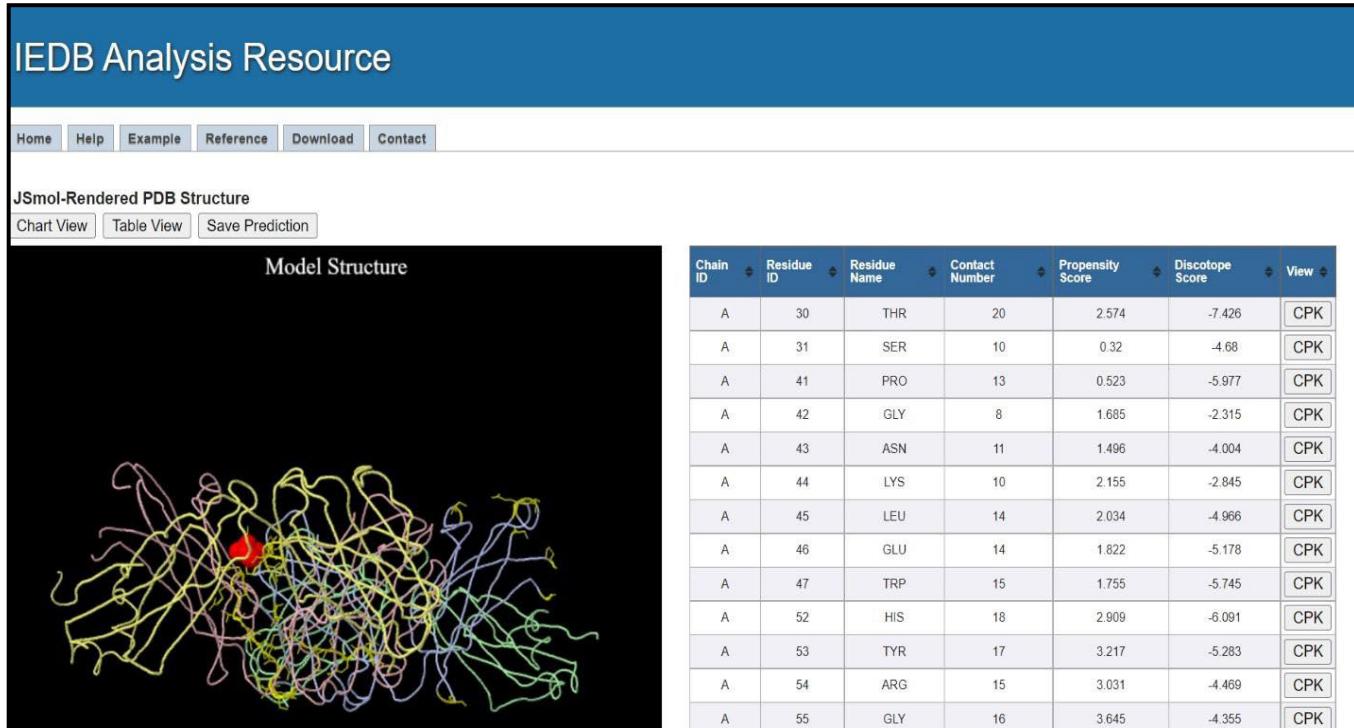


Fig 4: CPK view for Residue ID 30

## RESULTS:

The results are divided into 3 sections: Graphical View/Chart view, Table view & 3D view. Details are discussed below:

### 1. Graphical View/Chart view:

In this, the default threshold value for version 1.1 is -7.7 based on which it states that the predictions above the threshold (red line) are positive predictions (displayed in green) and predictions below the threshold are negative predictions (displayed in orange) indicating which amino acids from 3D structure should be considered as epitope.

### 2. Table View:

The Table view lists the following columns such as:

**Chain ID:** The chain id of the protein chain used in prediction (specified by the user)

**Residue ID:** PDB Residue id

**Residue Name:** Name of the residue

**Contact Number:** The residue contact number is the number of C $\alpha$  atoms in the antigen within a distance of 10 Å of the residue's C $\alpha$  atom. A low contact number correlates with localization of the residue close to the surface or in protruding regions of the antigen's structures.

**Propensity Score:** This score tells you about the probability/tendency of being part of an epitope for that particular residue. The propensity is reflected in amino acid epitope log- odds ratios, which were calculated on a set of 75 antigens. The propensity score is calculated by sequentially averaging epitope log-odds ratios within a window of 9 residues. Then the scores are summed up based on the proximity in the 3D structure of the antigen.

For any given residue, the sequentially averaged log-odds scores from all residues within 10Å are summed to give the propensity score.

- **Discotope Score:**This score is calculated by combining the contact numbers with propensity score. DiscoTope score above the threshold value indicates positive predictions and that below the threshold value indicates negative predictions.
- Whereas, the positive predictions are displayed in green.

### 3. 3D view:

The 3d view uses Jmol to display the structure with positive predictions highlighted in yellow. The side chain of each predicted residue is shown. You can rotate, zoom and manipulate the structure by using different buttons on the mouse. The table lists the predicted epitope residues along with their chain id, residue id, contact number, propensity score and DiscoTope score. Clicking on the CPK button in each residue will highlight this residue in CPK on the 3D viewer.

## CONCLUSION:

The Discotope server helps to predict discontinuous epitopes from 3D structures of proteins in PDB format. These methods achieve highly significant predictive performances suggesting these tools to be a powerful asset in rational epitope discovery.

## REFERENCES:

1. BCE Prediction Discotope. (n.d.). NCBI. Retrieved October 22, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242418/>
2. Immunoglobulin. (n.d.). NCBI. Retrieved October 22, 2022, from <https://www.ncbi.nlm.nih.gov/books/NBK513460/>
3. IEDB Analysis Resource. (n.d.). Retrieved October 22, 2022, from <https://www.iedb.org/>
4. BCE Prediction Discotope. (n.d.). IEDB Analysis Resource. Retrieved October 22, 2022, from <http://tools.iedb.org/discotope/>
5. 2Z91. (n.d.). IEDB Analysis Resource. Retrieved October 22, 2022, from <http://tools.iedb.org/discotope/result/>

**WEBLEM 7B**  
**TEPITOOL**  
(URL:  
<http://tools.iedb.org/tepitool/>

**AIM:**

To predict peptides for LILRA5 binding to MHC class I and class II molecules using Tepitool.

**INTRODUCTION:**

The Tepitool provides prediction of peptides binding to MHC class I and class II molecules. Tool is designed as a wizard with 6 steps as described below. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per your specific needs. You can go back to previous steps to change your selection before submission of the job. Once you submit the job (at the end of step-6), you will not be able to make any more changes and will have to start the prediction all over again with updated input parameters.

Immunoglobulins (Ig) or antibodies are glycoproteins that are produced by plasma cells. B cells are instructed by specific immunogens, for, example, bacterial proteins, to differentiate into plasma cells, which are protein-making cells that participate in humoral immune responses against bacteria, viruses, fungi, parasites, cellular antigens, chemicals, and synthetic substances.

Leukocyte immunoglobulin-like receptor subfamily A member 5 (LILR-A5) also known as CD85 antigen- like family member F (CD85f), immunoglobulin-like transcript 7 (ILT-7), and leukocyte immunoglobulin- like receptor 9 (LIR-9) is a protein that in humans is encoded by the LILRA5 gene. This gene is one of the leukocyte receptor genes that form a gene cluster on the chromosomal region 19q13.4

**METHODOLOGY:**

1. Copy the FASTA sequence for query LILRA5 from the Uniport database. (<https://www.uniprot.org/>)
2. Open the Homepage of IEDB database from google: <https://www.iedb.org/>
3. In the IEDB Database homepage, Click on T cell Epitope Prediction.
4. Under the T cell Epitope Prediction section, select the Tepitool option.
5. Follow the submission Steps:

**STEP 1: SEQUENCE - Provide sequence data:**

Paste the FASTA sequence of query retrieved from Uniprot database

**STEP2: Select SPECIES & ALLELE CLASS:**

Select the host species and MHC allele class.

**STEP 3: Select ALLELES i.e. Specify alleles:**

Select the specific alleles for prediction.(NOTE: For multiple alleles selection, hold CTRL button and select)

**STEP 4: PEPTIDES - Select peptides to be included in prediction:**

Select peptides to be included in prediction as per the option provided by the tool.

**STEP 5: METHOD - Select prediction & peptide selection methods and cutoff values:** Select preferred methods for binding prediction and peptide selection strategy and cutoff values.

**STEP 6: REVIEW: Review selections, enter job details & submit data:**

Review selections, enter job details and submit data.

## OBSERVATIONS:

```
>sp|A6NI73|LIRAS_HUMAN Leukocyte immunoglobulin-like receptor subfamily A member 5 OS=Homo sapiens OX=9606 GN=LILRA5 PE=1 SV=1  
MAPWSHPSAQQLQPVGGDAVSPALMVLCLGLSLGPRTHVQAGNLSKATLWAEPGSVISRG  
NSVTIRCQGTLEAQEYRLVKEGSPEPWDTQNPLEPKNKARFSIPSMTTEHHAGRVRYYYS  
PAGNISEPSDPLIELVVTGFYNKPTLSALPSPVTSQENVTLQCGSRLRFDRFILTEEGDHK  
LSWTLDSQLTPSGQFQALFPVGPVTPSHRWMLRCYGSRRHILQWSEPSDLLEIPVSGAA  
DNLSPSQNKSDSGTASHLQDYAVENLIRNGMAGLILVVLGILIFQDWHSQRSPQAAGR
```

Fig 1: FASTA Sequence for LILRA5 query (UniProt Id:A6NI73 )



The IEEDB has just launched its updated 3D viewers! Learn more via our help article [here](#).

## Welcome

The Immune Epitope Database (IEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEDB also hosts tools to assist in the prediction and analysis of epitopes.

[Learn More](#)

## Upcoming Events &amp; News

AAI Exhibitor Booth	May 6-10
FOCIS Exhibitor Booth	June 21-24
<a href="#">Virtual User Workshop</a>	Oct 26-28

\* register [here](#)

[IEDB SARS-CoV-2 Epitope Analysis Videos](#)

## Summary Metrics

Peptidic Epitopes	1,539,160
-------------------	-----------

## START YOUR SEARCH HERE

Epitope [?](#)

- Any
- Linear peptide
- Discontinuous
- Non-peptidic

Assay [?](#)

- T Cell
  - B Cell
  - MHC Ligand
- Ex: neutralization
- Outcome:  Positive  Negative

Epitope Source [?](#)

- Organism
- Ex: influenza, peanut
- Antigen
- Ex: core, capsid, myos

MHC Restriction [?](#)

- Any
  - Class I
  - Class II
  - Non-classical
- Ex: HLA-A\*02:01

Host [?](#)

- Any

Disease [?](#)

- Any

## Epitope Analysis Resource

T Cell Epitope Prediction [?](#)

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome,TAP)
- MHC I Immunogenicity

B Cell Epitope Prediction [?](#)

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- ElliPro

Epitope Analysis Tools [?](#)

Analyze epitope sets of:

Fig 1.2: Homepage of IEDB database



The IEEDB has just launched its updated 3D viewers! Learn more via our help article [here](#).

## Welcome

The Immune Epitope Database (IEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEDB also hosts tools to assist in the prediction and analysis of epitopes.

[Learn More](#)

## Upcoming Events &amp; News

AAI Exhibitor Booth	May 6-10
FOCIS Exhibitor Booth	June 21-24
<a href="#">Virtual User Workshop</a>	Oct 26-28

\* register [here](#)

[IEDB SARS-CoV-2 Epitope Analysis Videos](#)

## Summary Metrics

Peptidic Epitopes	1,539,160
-------------------	-----------

## START YOUR SEARCH HERE

Epitope [?](#)

- Any
- Linear peptide
- Discontinuous
- Non-peptidic

Assay [?](#)

- T Cell
  - B Cell
  - MHC Ligand
- Ex: neutralization
- Outcome:  Positive  Negative

Epitope Source [?](#)

- Organism
- Ex: influenza, peanut
- Antigen
- Ex: core, capsid, myos

MHC Restriction [?](#)

- Any
  - Class I
  - Class II
  - Non-classical
- Ex: HLA-A\*02:01

Host [?](#)

- Any

Disease [?](#)

- Any

## Epitope Analysis Resource

T Cell Epitope Prediction [?](#)

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome,TAP)
- MHC I Immunogenicity

B Cell Epitope Prediction [?](#)

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- ElliPro

Epitope Analysis Tools [?](#)

Analyze epitope sets of:

Fig 1.3: TCE Prediction under IEDB Database

# IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

## T Cell Epitope Prediction Tools

### T Cell Epitopes - MHC Binding Prediction

These tools predict IC50 values for peptides binding to specific MHC molecules. Note that binding to MHC is necessary but not sufficient for recognition by T cells.

#### Peptide binding to MHC class I molecules

This tool will take in an amino acid sequence, or set of sequences and determine each subsequence's ability to bind to a specific MHC class I molecule.

#### Peptide binding to MHC class II molecules

This tool employs different methods to predict MHC Class II epitopes, including a consensus approach which combines NN-align, SMM-align and Combinatorial library methods.

#### TepiTool:

The TepiTool provides prediction of peptides binding to MHC class I and class II molecules. Tool is designed as a wizard with 6 steps as described below. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per your specific needs. You can go back to previous steps to change your selection before submission of the job. Once you submit the job (at the end of step-6), you will not be able to make any more changes and will have to start the prediction all over again with updated input parameters.

### T Cell Epitopes - Processing Prediction

These tools predict epitope candidates based upon the processing of peptides in the cell.

#### Proteasomal cleavage/TAP transport/MHC class I combined predictor

This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope.

Fig 1.4: Homepage of TCE Prediction

# IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

## T Cell Epitope Prediction Tools

### T Cell Epitopes - MHC Binding Prediction

These tools predict IC50 values for peptides binding to specific MHC molecules. Note that binding to MHC is necessary but not sufficient for recognition by T cells.

#### Peptide binding to MHC class I molecules

This tool will take in an amino acid sequence, or set of sequences and determine each subsequence's ability to bind to a specific MHC class I molecule.

#### Peptide binding to MHC class II molecules

This tool employs different methods to predict MHC Class II epitopes, including a consensus approach which combines NN-align, SMM-align and Combinatorial library methods.

#### TepiTool:

The TepiTool provides prediction of peptides binding to MHC class I and class II molecules. Tool is designed as a wizard with 6 steps as described below. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per your specific needs. You can go back to previous steps to change your selection before submission of the job. Once you submit the job (at the end of step-6), you will not be able to make any more changes and will have to start the prediction all over again with updated input parameters.

### T Cell Epitopes - Processing Prediction

These tools predict epitope candidates based upon the processing of peptides in the cell.

#### Proteasomal cleavage/TAP transport/MHC class I combined predictor

This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope.

Fig 1.5: TepiTool under TCE Prediction Tools

## MHC CLASS I

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 4 5 6

SEQUENCE - Provide sequence data:

Enter sequence(s) in FASTA or PLAIN format.

No format detected.

Or upload file containing sequence(s) Choose File No file chosen

Next

Fig 2: Search Option Under TepiTool

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 4 5 6

SEQUENCE - Provide sequence data:

Enter sequence(s) in FASTA or PLAIN format.

```
>sp|A6NI73|LIRA5_HUMAN Leukocyte immunoglobulin-like receptor subfamily A member 5 OS=Homo sapiens OX=9606 GN=LILRA5 PE=1 SV=1 MAPWSHPSAQQLQPVGGDAVSPALMVLCLGLSLGPRTHVQAGNLSKATLWAEPGSVISRG NSVTIRCQGTLEAQEYRLVKEGSPEPWDTQNPLEPKNKARFSIIPSMTEHHAGRYRCYYYSPAGWSEPSDPLELVVTGFYNKPTLSALPSPVVTSGENVTLQCGSRLRFDRFILTEEGDHK LSWTLDSQLTPSGQFQALFPVGPVTPSHRWMRLCYGSRRHILQVWSEPSDLLEIPVSGAADNLSPSQNKSDSGTASHLQDYAVENLIRMGMAGLILVVLGILIFQDWHSQRSPQAAAGR|
```

FASTA format detected.

Or upload file containing sequence(s) Choose File No file chosen

Next

Fig 2.1: STEP 1- Provide FASTA sequence for query

# IEDB Analysis Resource

Home Help Reference Download Contact

## TepiTool

Steps 1 2 3 4 5 6

### SPECIES & ALLELE CLASS - Select the host species and MHC allele class:

Host species Human

Allele class Class I

Current selections:

No. of sequences 1

Start Over Back Next

Fig 2.2: STEP 2- Select the host species and MHC allele class

# IEDB Analysis Resource

Home Help Reference Download Contact

## TepiTool

Steps 1 2 3 4 5 6

### ALLELES - Specify alleles:

Human - Class I

- Select from list of frequently occurring alleles (Frequency > 1%)
- Select from list of all available alleles
- Select from list of representative alleles from different HLA supertypes
- Use panel of 27 most frequent A & B alleles
- Upload allele file

Alleles

A\*01:01  
A\*02:01  
A\*02:06  
A\*03:01  
A\*11:01  
A\*23:01  
A\*24:02  
A\*25:01  
A\*26:01  
A\*29:02  
A\*30:01

Current selections:

No. of sequences	1
Host species	Human
Allele class	Class I
Selected alleles	
Reset alleles	

Fig 2.3: STEP 3- Mention the specific alleles

## TepiTool

Steps 1 2 3 4 5 6

### PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction	<ul style="list-style-type: none"> <li><input checked="" type="radio"/> Apply default settings for low number of peptides</li> <li><input type="radio"/> Apply default settings for moderate number of peptides</li> <li><input type="radio"/> Apply default settings for high number of peptides</li> <li><input type="radio"/> Custom selection - Select your own settings</li> </ul> <hr/> <p>Handling of duplicate peptides:</p> <ul style="list-style-type: none"> <li>- Duplicate peptides will be removed.</li> </ul> <hr/> <p>Peptide lengths to be considered in prediction:</p> <ul style="list-style-type: none"> <li>- Only peptide length 9 will be included 9mers = 291</li> </ul>
Conservancy analysis (Uses only peptides conserved in specified % of sequences)	N/A (You have only 1 sequence)
<a href="#">Start Over</a> <a href="#">Back</a> <a href="#">Next</a>	

### Current selections:

No. of sequences	1
Host species	Human
Allele class	Class I
Selected alleles	1.A*01:01 2.A*02:01 3.A*02:06

Fig 2.4: STEP 4- Select peptides to be included in prediction

## IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

## TepiTool

Steps 1 2 3 4 5 6

### METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use	<select>IEDB recommended</select>
Selection of predicted peptides	<div style="display: flex; align-items: center;"> <span>Select peptides based on predicted percentile rank</span> <select style="margin-left: 10px;"></select> </div> <div style="margin-top: 10px;">         Select peptides with predicted consensus percentile rank ≤ <input type="text" value="1"/> </div>
<a href="#">Start Over</a> <a href="#">Back</a> <a href="#">Next</a>	

### Current selections:

No. of sequences	1
Host species	Human
Allele class	Class I
Selected alleles	1.A*01:01 2.A*02:01 3.A*02:06
Duplicate peptides	Removed
Peptide lengths selected	9mers
No. of peptides included (Not considering conservancy analysis)	291
Conservancy analysis	Peptides conserved in at least % sequences

Fig 2.5: METHOD - Select prediction & peptide selection methods and cutoff values

# IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

## TepiTool

Steps 1 2 3 4 5 **6**

**REVIEW: Review selections, enter job details & submit data:**

### Summary:

No. of sequences	1
Host species	Human
Allele class	Class I
Alleles	1.A*01:01 2.A*02:01 3.A*02:06
Duplicate peptides	Removed
Peptide lengths selected	9mers
Approx no. of peptides included	291
Peptide overlap	N/A (all possible nmers are included in class I)
Conservancy analysis	Peptides conserved in at least % sequences
Prediction method	IEDB recommended
Peptide selection criterion	Based on predicted consensus percentile rank (Cutoff selected = 1)

Fig 2.6: REVIEW: Review selections, enter job details & submit data

# IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

## TepiTool

Prediction results - concise (Download table ):

Seq # ▾ ▷	Peptide start ▾ ▷	Peptide end ▾ ▷	Peptide ▾ ▷	Percentile rank ▾ ▷	Allele ▾ ▷
1	106	114	MTEHHAGRY	0.01	HLA-A*01:01
1	193	201	GQFQALFPV	0.08	HLA-A*02:06
1	270	278	GMAGLILVV	0.09	HLA-A*02:01
1	253	261	GTASHLQDY	0.1	HLA-A*01:01
1	48	56	TLWAEPGSV	0.12	HLA-A*02:01
1	261	269	YAVENLIRM	0.18	HLA-A*02:06
1	270	278	GMAGLILVV	0.18	HLA-A*02:06
1	18	26	AVSPALMVL	0.19	HLA-A*02:06
1	55	63	SVISRGNSV	0.2	HLA-A*02:06
1	193	201	GQFQALFPV	0.26	HLA-A*02:01
1	143	151	TLSALPSPV	0.36	HLA-A*02:01
1	48	56	TLWAEPGSV	0.37	HLA-A*02:06
1	249	257	KSDSGTASH	0.42	HLA-A*01:01
1	197	205	ALFPVGVPVT	0.48	HLA-A*02:01
1	39	47	VQAGNLSKA	0.48	HLA-A*02:06

Fig 3: Resultpage for LILRA5 under TepiTool

## MHC CLASS II:

### Iedb Analysis Resource

Home Help Reference Download Contact

#### TepiTool

Steps 1 2 3 4 5 6

**SEQUENCE - Provide sequence data:**

Enter sequence(s) in FASTA or PLAIN format.

No format detected.

Or upload file containing sequence(s) Choose File No file chosen

Next

This screenshot shows the 'SEQUENCE - Provide sequence data' step of the TepiTool. It features a large text input area for pasting sequences. A message at the bottom right of this area says 'No format detected.' Below the input area is a section for uploading files, with a 'Choose File' button and a placeholder 'No file chosen'. At the very bottom is a blue 'Next' button.

Fig 4: Search Option Under TepiTool

### Iedb Analysis Resource

Home Help Reference Download Contact

#### TepiTool

Steps 1 2 3 4 5 6

**SEQUENCE - Provide sequence data:**

>sp|P24071|FCAR\_HUMAN Immunoglobulin alpha Fc receptor OS=Homo sapiens  
OX=9606 GN=FCAR PE=1 SV=1  
MDPKQTLLCLVCLGQRQAEQEGDFPMPPFISAKSSPVIPLDGSVKIQCQAIREAYLTQL  
MIINKNSTYREIGRRLKFWNETDPEFVIDHMIDANKAGRYOCQYRIGHTYFRYSDTLELVVT  
GLYGKPFLSADRGVLVLMPPGENISLTCSAAHIPFDRFSLAKEGEELSLPQHQSGEHPANFSL  
GPVDLNNSGIYRCVGNYNRSPYLWSPSPNNALELVTVDSIHQDVTTQNLIRMAVAGLVLVA  
LLAILVENWHSHTALNEASADVAEPWSQMCQPGLTFARTPSVCK

Enter sequence(s) in FASTA or PLAIN format.

FASTA format detected.

Or upload file containing sequence(s) Choose File No file chosen

Next

This screenshot shows the same 'SEQUENCE - Provide sequence data' step as the previous one, but with a detected sequence format. The sequence itself is displayed in a monospaced font within the input area. A message at the bottom right of the input area says 'FASTA format detected.' The rest of the interface, including the file upload section and the 'Next' button, is identical to the previous screenshot.

Fig 4.1: SEQUENCE - Provide sequence data

# IEDB Analysis Resource

Home Help Reference Download Contact

## TepiTool

Steps 1 2 3 4 5 6

**SPECIES & ALLELE CLASS - Select the host species and MHC allele class:**

Host species	Human
Allele class	Class II

Current selections:  
No. of sequences 1

Start Over Back Next

Fig 4.2: SPECIES & ALLELE CLASS - Select the host species and MHC allele class

Steps 1 2 3 4 5 6

**ALLELES - Specify alleles:**

Alleles

Human - Class II

Predict for custom allele set  
 Predict for pre-selected panel of alleles  
 Predict using pre-selected allele sets & methods

Options:  
 Select from list of alleles  
 Upload allele file

Select α and β chains separately when applicable

DR

DRB1*01:01	DRB1*01:02
DRB1*01:03	DRB1*01:04
DRB1*01:05	DRB1*01:06
DRB1*01:07	DRB1*01:08
DRB1*01:09	DRB1*01:10
DRB1*01:11	DRB1*01:12

Current selections:  
No. of sequences 1  
Host species Human  
Allele class Class II  
Selected alleles DRB1\*01:01 DRB1\*01:02 DRB1\*01:03  
Reset alleles

Fig 4.3: ALLELES - Specify alleles

# IEDB Analysis Resource

Home | Help | Reference | Download | Contact

## TepiTool

Steps 1 2 3 4 5 6

### METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use IEDB recommended ▾

Selection of predicted peptides Select peptides based on predicted percentile rank ▾

Select peptides with predicted consensus percentile rank ≤ 10

[Start Over](#) [Back](#) [Next](#)

### Current selections:

No. of sequences	1
Host species	Human
Allele class	Class II
Alleles selected	1.DRB1*01:01 2.DRB1*01:02 3.DRB1*01:03
Duplicate peptides	Removed
Peptide overlap	10 AA residues
Approx no. of peptides included (Not considering conservancy analysis)	58
Conservancy analysis	Peptides conserved in at least % sequences

Fig 4.4: PEPTIDES - Select peptides to be included in prediction

# IEDB Analysis Resource

Home | Help | Reference | Download | Contact

## TepiTool

Steps 1 2 3 4 5 6

### PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction

- Apply default settings for low number of peptides
- Apply default settings for moderate number of peptides
- Apply default settings for high number of peptides
- Custom selection - Select your own settings

#### Handling of duplicate peptides

- Duplicate peptides will be removed.

#### Desired no. of overlapping residues for 15mers

- No. of overlapping residues fixed at 10.

Approximate no. of peptides to be considered for prediction =

58

Conservancy analysis

### Current selections:

No. of sequences	1
Host species	Human
Allele class	Class II
Selected alleles	1.DRB1*01:01 2.DRB1*01:02 3.DRB1*01:03

Fig 4.5: METHOD - Select prediction & peptide selection methods and cutoff values

# IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

## TepiTool

Steps [1](#) [2](#) [3](#) [4](#) [5](#) **6**

**REVIEW: Review selections, enter job details & submit data:**

### Summary:

No. of sequences	1
Host species	Human
Allele class	Class II
Alleles	1.DRB1*01:01 2.DRB1*01:02 3.DRB1*01:03
Duplicate peptides	Removed
Peptide lengths selected	15mers (Only one length for class II)
Approx no. of peptides included	58
Peptide overlap	10 AA residues
Conservancy analysis	Peptides conserved in at least % sequences
Prediction method	IEDB recommended
Peptide selection criterion	Based on predicted consensus percentile rank (Cutoff selected = 10)

Fig 4.6: REVIEW: Review selections, enter job details & submit data

# IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

## TepiTool

Prediction results - concise ([Download table](#) ):

Seq # ▾	Peptide start ▾	Peptide end ▾	Peptide sequence ▾	Consensus percentile rank ▾	Allele ▾
1	269	283	MGMAGLILVVLGILI	2.20	HLA-DRB1*01:01
1	141	155	KPTLSALPSPVVTSG	2.40	HLA-DRB1*01:01
1	264	278	ENLIRMGMAGLILVV	3.30	HLA-DRB1*01:01
1	22	36	ALMVLLCLGLSLGPR	0.60	HLA-DRB1*01:02
1	17	31	DAVPALMVLCLGL	0.60	HLA-DRB1*01:02
1	264	278	ENLIRMGMAGLILVV	3.00	HLA-DRB1*01:02
1	275	289	ILVVLGILIFQDWHS	3.20	HLA-DRB1*01:02
1	269	283	MGMAGLILVVLGILI	3.20	HLA-DRB1*01:02
1	195	209	FQALFPVGPVTPSHR	9.40	HLA-DRB1*01:02
1	189	203	LTPSGQFQALFPVGP	9.40	HLA-DRB1*01:02
1	60	74	GNSVTIRCQGTLEAQ	9.90	HLA-DRB1*01:02
1	65	79	IRCQGTLEAQEYRLV	9.90	HLA-DRB1*01:02
1	264	278	ENLIRMGMAGLILVV	0.14	HLA-DRB1*01:03
1	141	155	KPTLSALPSPVVTSG	6.70	HLA-DRB1*01:03
1	208	222	HRWMLRCYGSRRHIL	7.60	HLA-DRB1*01:03

**Fig 5: Resultpage for LILRA5 under TepiTool**

## **RESULTS:**

### **TepiTool Prediction for structure Immunoglobulin alpha Fc receptor MHC \_\_\_\_\_**

#### **Class I and MHC Class II:**

The prediction results page is divided into 2 sections. The details are given below:

##### **1. Concise results/ Prediction Results:**

The concise results table shows the final list of predicted peptides selected based on the input parameters provided. The table will contain the sequence # of the peptide's source protein in the input sequence set, start and end positions of the peptide within the source protein sequence, the peptide sequence, the selection criterion parameter value (consensus percentile rank), the allele (where applicable). For the Immunoglobulin alpha Fc receptor query, the lower value is 0.01 indicating better binding affinity. Also the concise results table can also be downloaded as .csv file which can be opened using any spreadsheet such as MS Excel for further analysis. This section will be included in the email.

##### **2. Download results details:**

This section provides links for downloading the results details as csv files. It can include the following based on the input parameters:

1. Non-redundant results: This file will contain prediction results with redundant peptides within each sequence removed. Redundant peptides means peptides that overlap with more residues than desired. This result set includes both positive and negative peptides based on the input parameters.
2. Complete results: This file will contain binding predictions of all peptides. This will include the predicted IC50 and percentile rank or other scores depending on the prediction method chosen. In case of IEDB recommended or consensus method, the results will include details from each of the prediction methods employed.
3. Conservancy of peptides (applicable only if conservancy analysis is done): This file will contain conservancy of each peptide in the input sequence set.

## **CONCLUSION:**

Computational prediction of T-cell epitope candidates is currently being used in several applications including vaccine discovery studies, development of diagnostics and removal of unwanted immune responses against protein therapeutics. This protocol provides step-by-step instructions with necessary recommendations for prediction of the best T-cell epitope candidates in line with the newly developed online tool called TepiTool. The TepiTool provides some of the top MHC class I and class II binding prediction algorithms for a number of species including humans, chimpanzees, bovines, gorillas, macaques, mice and pigs. The tool is designed as a user-friendly wizard with well-defined steps which

helps the users to predict the best MHC binding peptides from their sequences of interest.

## REFERENCES:

1. TepITOOL. (n.d.). IEDB Analysis Resource. Retrieved October 22, 2022, from <http://tools.iedb.org/tepitool/>
2. LILRA5. (n.d.). UniProt Database. Retrieved October 22, 2022, from <https://rest.uniprot.org/uniprotkb/A6NI73.fasta>
3. Tepitool. (n.d.). NCBI. Retrieved October 22, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4981331/>



## TO UNDERSTAND VARIOUS WEB-BASED TOOLS FOR VACCINE DESIGNING

### **Introduction:**

Computational methods used in vaccine design have been changing drastically in recent years. In classical immunological research results could be recorded by pen and pencil or in a spreadsheet, but new experimental high throughput methods such as sequencing, DNA arrays, and proteomics have generated a wealth of data that are not efficiently handled and mined by these approaches. This has fueled the rapid growth of the field of Immunological Bioinformatics (or Immuno-informatics) that addresses how to handle these large amounts of data in the field of immunology and vaccine design. Many of the methods have been made available on the Internet and can be used by experimental researchers without expert knowledge of bioinformatics. This review attempts to give an overview over the methods currently available and to point out the strengths and weaknesses of the different methods.

### **Structural antibody database: (SAbDab; <http://opig.stats.ox.ac.uk/webapps/sabdab>)**

It is an online resource containing all the publicly available antibody structures annotated and presented in a consistent fashion. The data are annotated with several properties including experimental information, gene details, correct heavy and light chain pairings, antigen details and, where available, antibody-antigen binding affinity. The user can select structures, according to these attributes as well as structural properties such as complementarity determining region loop conformation and variable domain orientation. Individual structures, datasets and the complete database can be downloaded.

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.

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.

### **Ag-Ab Interaction Database:**

Antigen-Antibody Interaction Database (AgAbDb) is an immunoinformatics resource developed at the Bioinformatics Centre, University of Pune, and is available online at <http://bioinfo.net.in/AgAbDb.htm>. Antigen-antibody interactions are a special class of protein-protein interactions that are characterized by high affinity and strict specificity of antibodies towards their antigens. Several co-crystal structures of antigen-antibody complexes have been solved and are available in the Protein Data Bank (PDB). AgAbDb is

a derived knowledgebase developed with an objective to compile, curate, and analyze determinants of interactions between the respective antigen- antibody molecules. AgAbDb lists not only the residues of binding sites of antigens and antibodies, but also interacting residue pairs. It also helps in the identification of interacting residues and buried residues that constitute antibody-binding sites of protein and peptide antigens. The Antigen-Antibody Interaction Finder (AAIF), a program developed in-house, is used to compile the molecular interactions, viz. van der Waals interactions, salt bridges, and hydrogen bonds. A module for curating water-mediated interactions has also been developed. In addition, various residue-level features, viz. accessible surface area, data on epitope segment, and secondary structural state of binding site residues, are also compiled. Apart from the PDB numbering, Wu-Kabat numbering and explicit definitions of complementarity-determining regions are provided for residues of antibodies. The molecular interactions can be visualized using the program Jmol. AgAbDb can be used as a benchmark dataset to validate algorithms for prediction of B-cell epitopes. It can as well be used to improve accuracy of existing algorithms and to design new algorithms. AgAbDb can also be used to design mimotopes representing antigens as well as aid in designing processes leading to humanization of antibodies.

### **Immune Epitope Database (IEDB):**

The Immune Epitope Database (IEDB) contains >1.6 million experiments representing the adaptive immune response to epitopes, gathered primarily from the literature . These experiments were manually curated following structured curation guidelines, as previously described . This data was obtained from 19 ,500 publications and includes all the literature available from the beginnings of PubMed until now. Historical curation of papers going back to 1952 was completed in 2011 and since, this database has focused on newly published papers. It perform a query of PubMed every two weeks to remain current with new content. The IEDB has approximately 300 unique visitors and 1220 page views per day. The IEDB exists as a free service with the goal of helping further immunological research. Thus, we routinely perform outreach activities to interact with our users to ascertain their needs and gather feedback on existing features. Here we present our efforts toward meeting user needs, as well as extending functionality to keep current with accepted web standards.

Significantly, research is ever-evolving; new experiments are continually created, expanding data quantity and complexity. As the cost of high throughput experiments is decreasing, scientists are publishing greater numbers of experiments per publication, leading to rapid increases in our data. This is reflected in the number of epitopes curated per publication year, which began rapidly

increasing in 2015 Accordingly, the number of experiments captured in the IEDB has also increased by 140% since 2015, now surpassing 1.6 million.

Another factor leading to large amounts of new data is the addition of receptor sequence data to the IEDB schema. Previously, It's only captured full length antibody and T cell receptor (TCR) sequences whenever a 3D structure was available, but we now capture both full length and CDR sequences, as well as gene usage whenever authors provide this. To accommodate this new data, we added new database tables, search panes, results tabs, and details pages, as described in a separate publication.

#### **Structural T-cell Receptor Database (STCRDab):**

The Structural T-cell Receptor Database (STCRDab;

<http://opig.stats.ox.ac.uk/webapps/stcrdab>) is an online resource that automatically collects and curates TCR structural data from the Protein Data Bank. For each entry, the database provides annotations, such as the  $\alpha/\beta$  or  $\gamma/\delta$  chain pairings, major histocompatibility complex details, and where available, antigen binding affinities. In addition, the orientation between the variable domains and the canonical forms of the complementarity-determining region loops are also provided. Users can select, view, and download individual or bulk sets of structures based on these criteria.

Where available, STCRDab also finds antibody structures that are similar to TCRs, helping users explore the relationship between TCRs and antibodies.

## References:

1. Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., & Deane, C. M. (2014, January). SABDAB: The Structural Antibody Database. *Nucleic acids research*. Retrieved October 16, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965125/>
2. SA;, K.-K. U. R.-R. S. N.-K. G. S. (n.d.). Antigen-Antibody Interaction Database (agabdb): A compendium of antigen-antibody interactions. *Methods in molecular biology* (Clifton, N.J.). Retrieved October 16, 2022, from <https://pubmed.ncbi.nlm.nih.gov/25048123/>
3. Leem J;de Oliveira SHP;Krawczyk K;Deane CM; (n.d.). STCRDAB: The structural T-Cell Receptor Database. *Nucleic acids research*. Retrieved October 16, 2022, from <https://pubmed.ncbi.nlm.nih.gov/29087479/>
4. Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2019, January 8). The Immune Epitope Database (IEDB): 2018 update. *Nucleic acids research*. Retrieved October 16, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324067/>