DATE:13/08/22

WEBLEM 4:

Introduction to STCRDAb database and visualisation of structure uing 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 αβ 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 αβ TCRs are annotated. IMGT has a richer (308 experimental structures) and more diverse set of structural data (e.g., γδ 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 αβ and γδ 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.

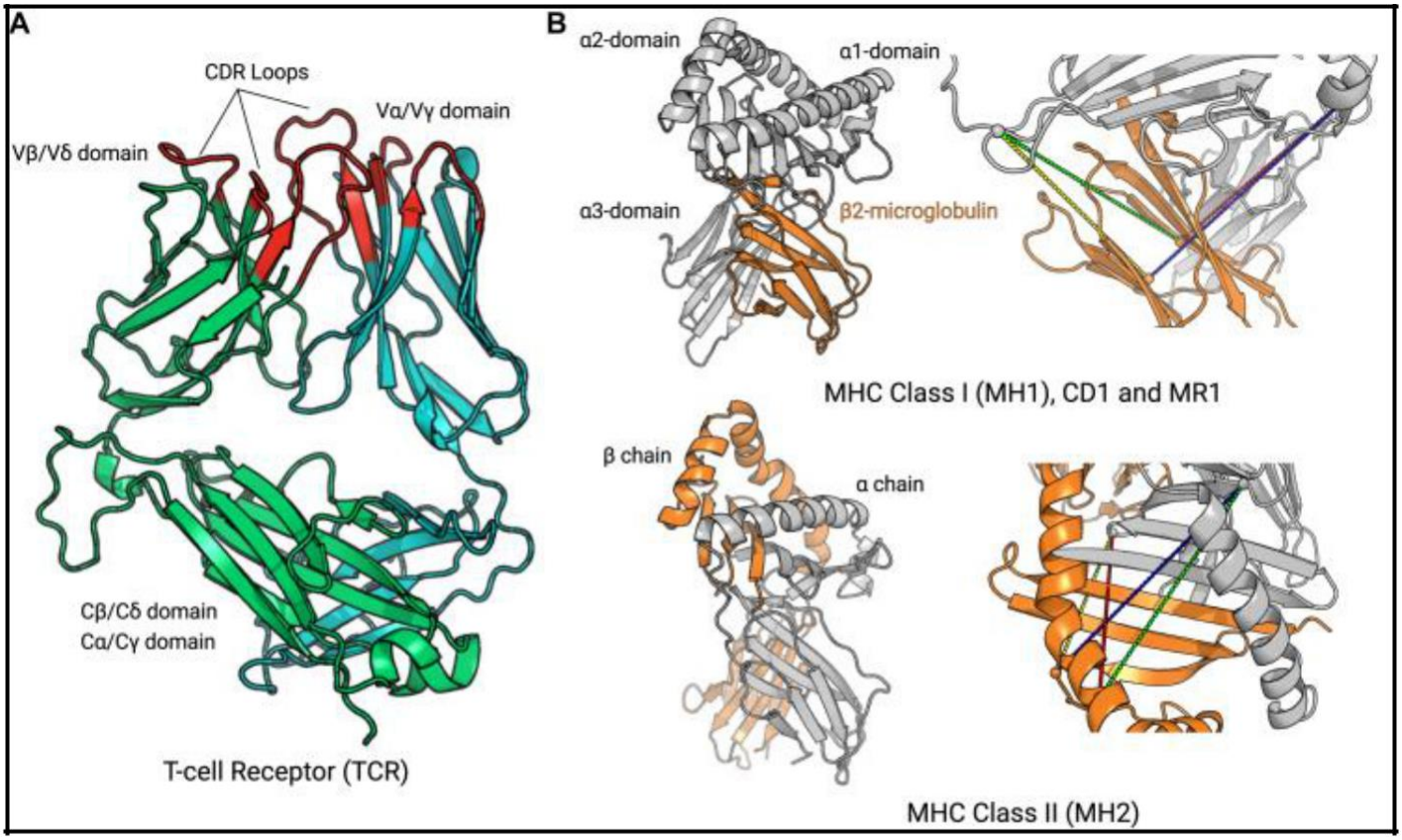

Figure 1.

(Nomenclature and colouring scheme used in STCRab. (A) T-cell receptors (TCRs) are formed from two chains: TCRβ/TCRα (to form αβ TCRs, as shown), or TCRδ/TCRγ (to form γδ 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: α15–β23 (green; 32 Å), α15–β104 (yellow; 32 Å), α51–β23 (red; 32 Å), α51–β104 (blue; 37 Å). To pair MH2, the following distance constraints are used: α29–β64 (green; 34 Å), α29–β39 (yellow; 22 Å), α37–β64 (red; 32 Å), α37–β39 (blue; 28 Å).)

TCR structures

The majority of available TCR structures are αβ TCRs, which are formed of TCRα and TCRβ chains. A small number of TCRs are γδ TCRs, consisting of TCRγ and TCRδ chains. The TCRβ and TCRδ chains are considered to be analogous to antibody heavy chains while the TCRα and TCRγ 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 β–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 β-sheet, flanked by two α helices. MH1, CD1 and MR1 are formed by the pairing of the MHC chain and a β_2 microglobulin, while MH2 is formed by the MHCα and MHCβ 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 αβ TCRs form the majority of the data, followed by single TCR chains, e.g. Vβ only structures, then γδ TCRs. There are also structures that fit none of these categories – for instance, an engineered TCRδ/TCRα receptor.

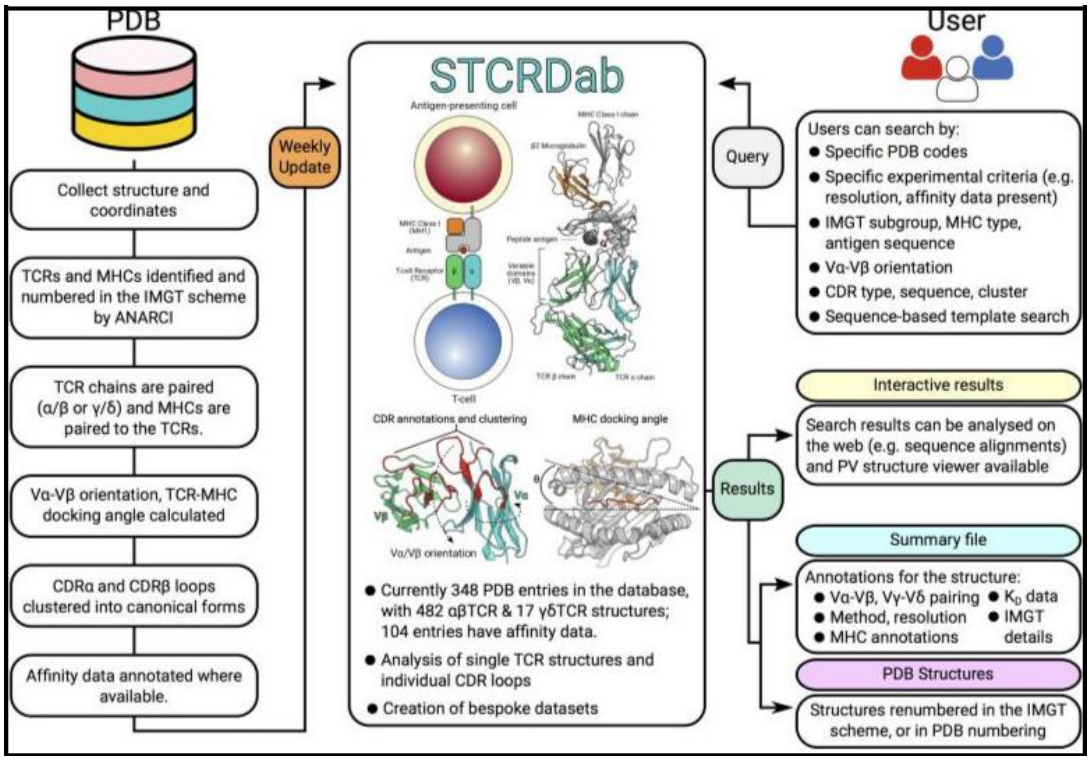

Figure 2.

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.

Vα-Vβ 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 Vα and Vβ domains using six parameters. The effect of Vα-Vβ 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 αβ TCRs. Due to the small amount of data, the TRangle method is currently not used for γδ 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 CDRα1, CDRα2, CDRα3, CDRβ1, CDRβ2 and CDRβ3 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 PDBBind 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

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.

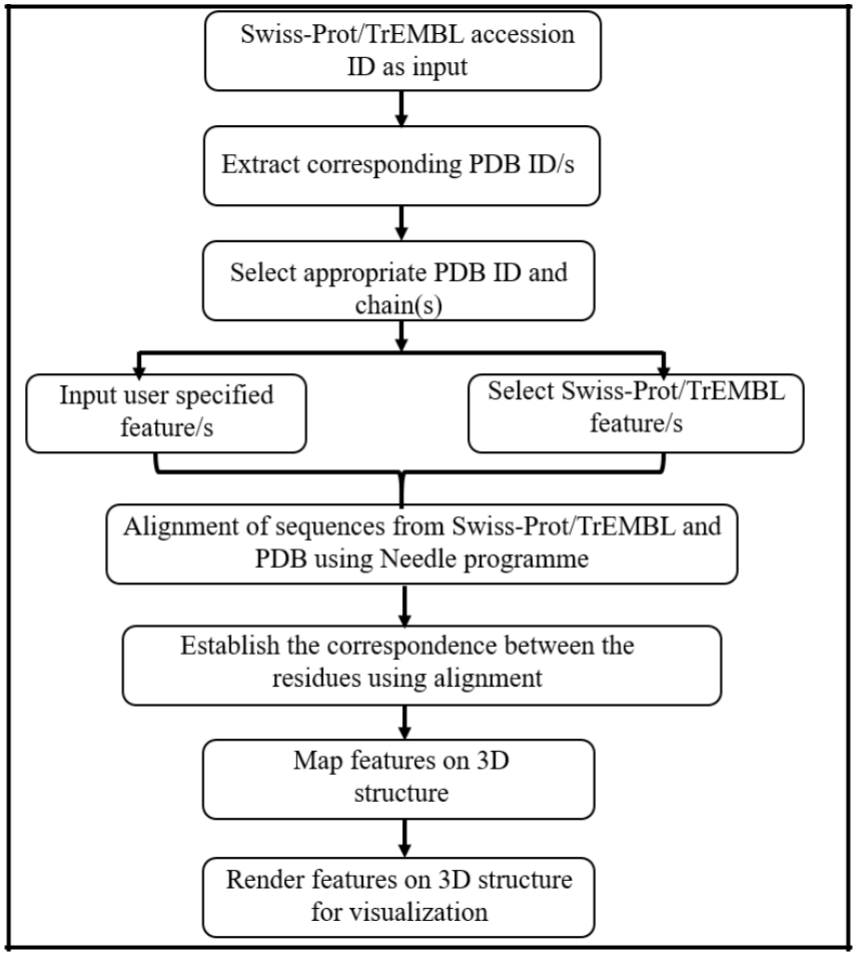

Figure 3. Workflow of PVF3D tool

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

DATE: 13/10/22

WEBLEM 4a

STCRDab Database

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

AIM:

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 β-selection, survival and proliferation of CD4(-)CD8(-) double-negative thymocytes, and subsequent αβ T-cell lineage differentiation. Whereas αβTCR 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 α-chain (pre-Tα) that pairs with any TCR β-chain (TCRβ) following successful TCR β-gene rearrangement. Here we provide the basis of pre-Tα-TCRβ assembly and pre-TCR dimerization. The pre-Tα chain comprised a single immunoglobulin-like domain that is structurally distinct from the constant (C) domain of the TCR α-chain; nevertheless, the mode of association between pre-Tα and TCRβ mirrored that mediated by the Cα-Cβ domains of the αβTCR. 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 α/β or γ/δ 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 αβ and γδ 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:

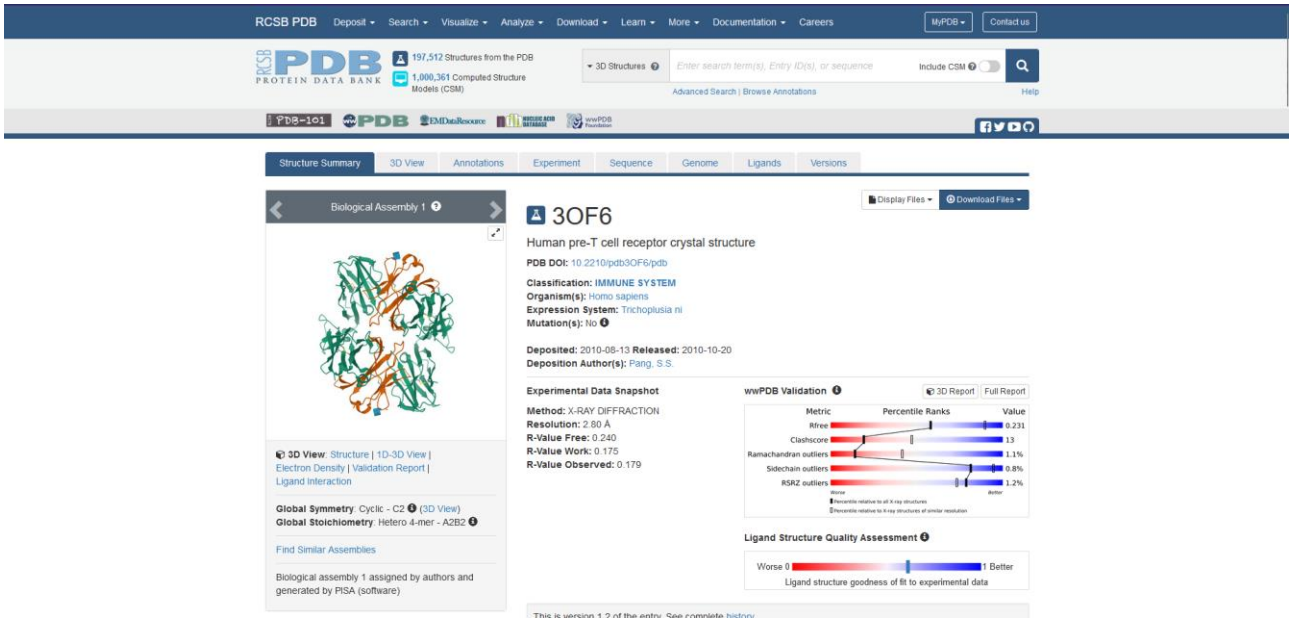

Fig 1. PDB database page for crystal structure of Human pre T-cell receptor

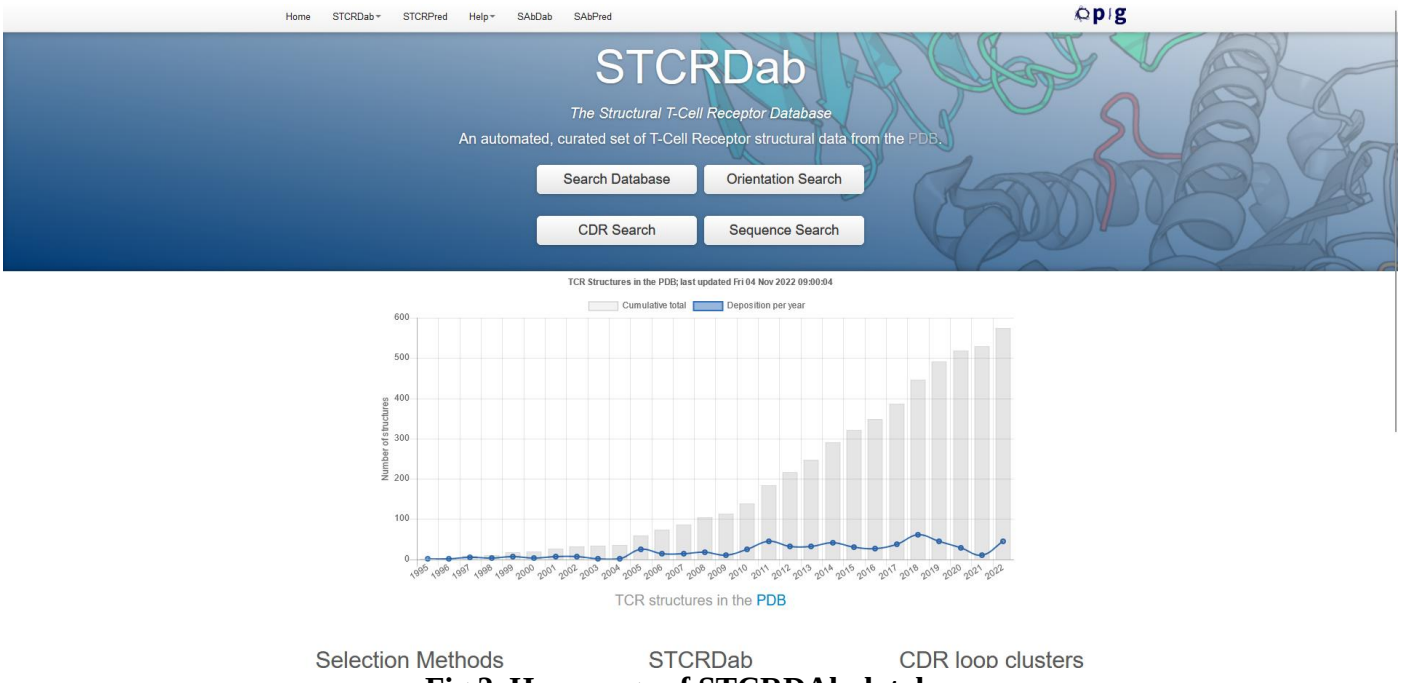

Fig 2. Homepage of STCRDAb database

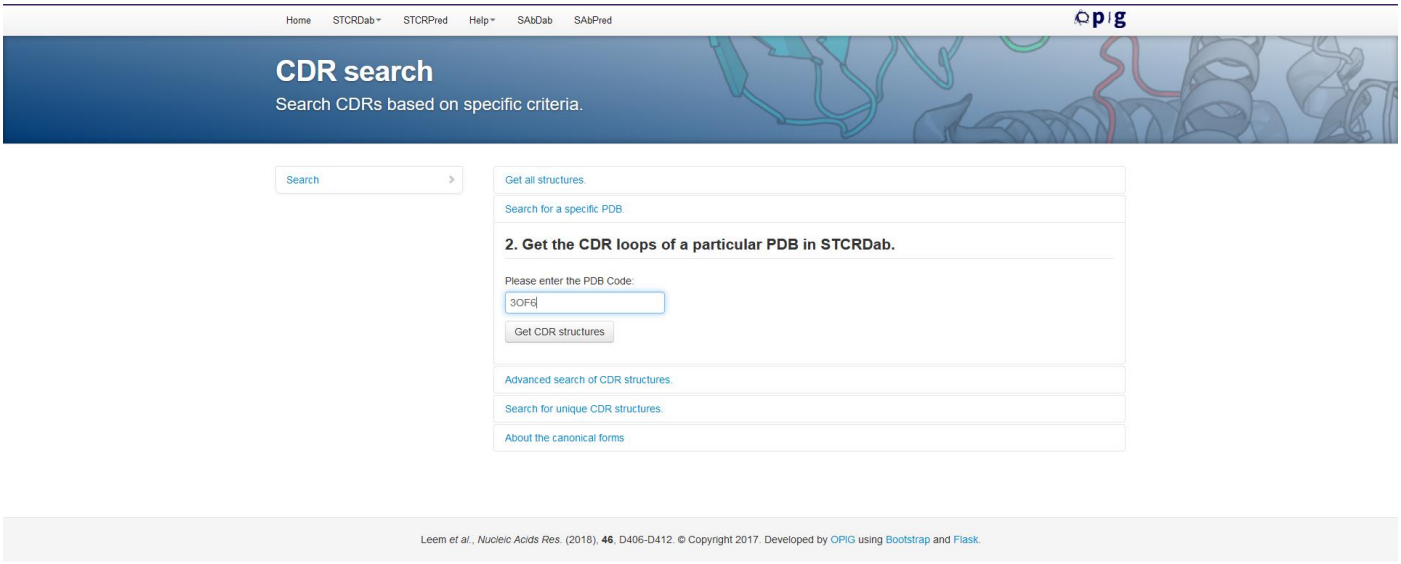

Fig 3. Page for CDR search

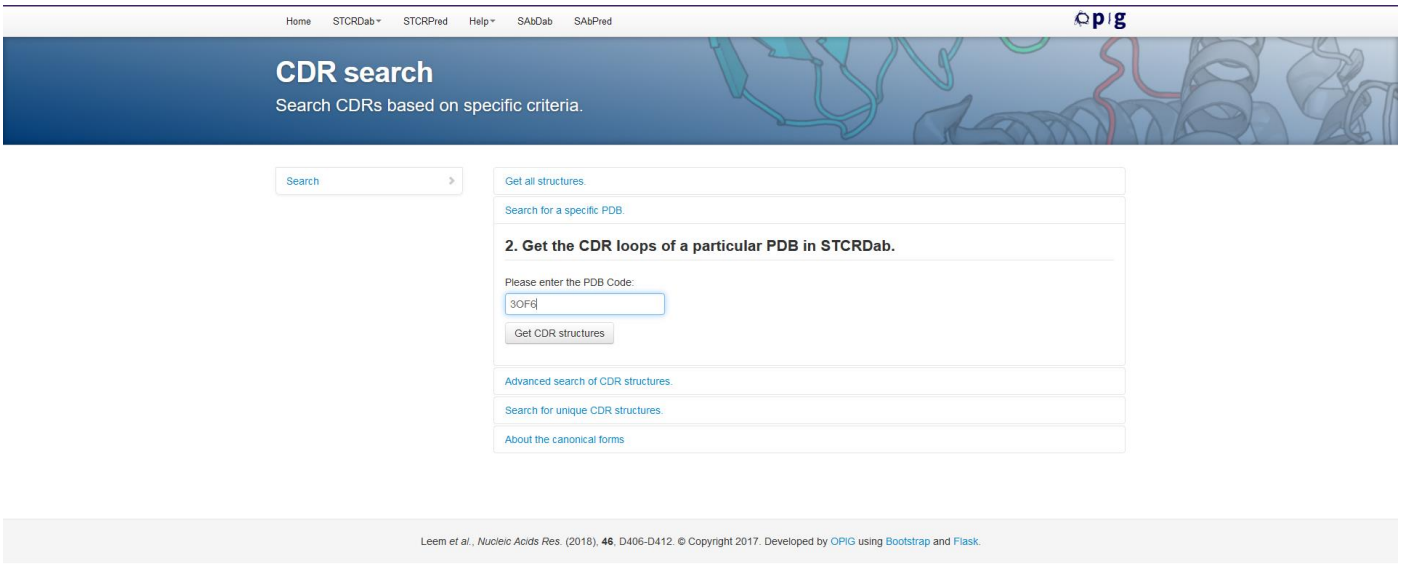

Fig 4. Search for specific PDB (PDB ID: 3OF6)

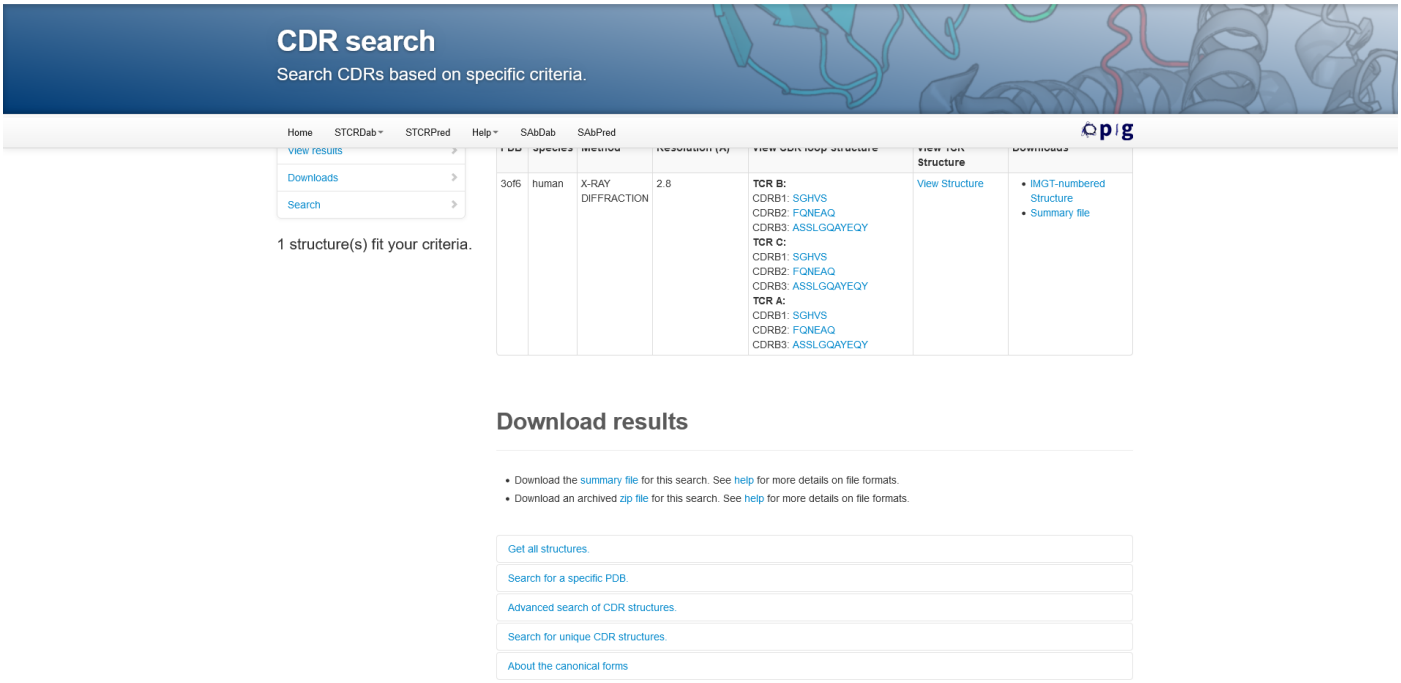

Fig 5. Result for Human pre T-cell receptor

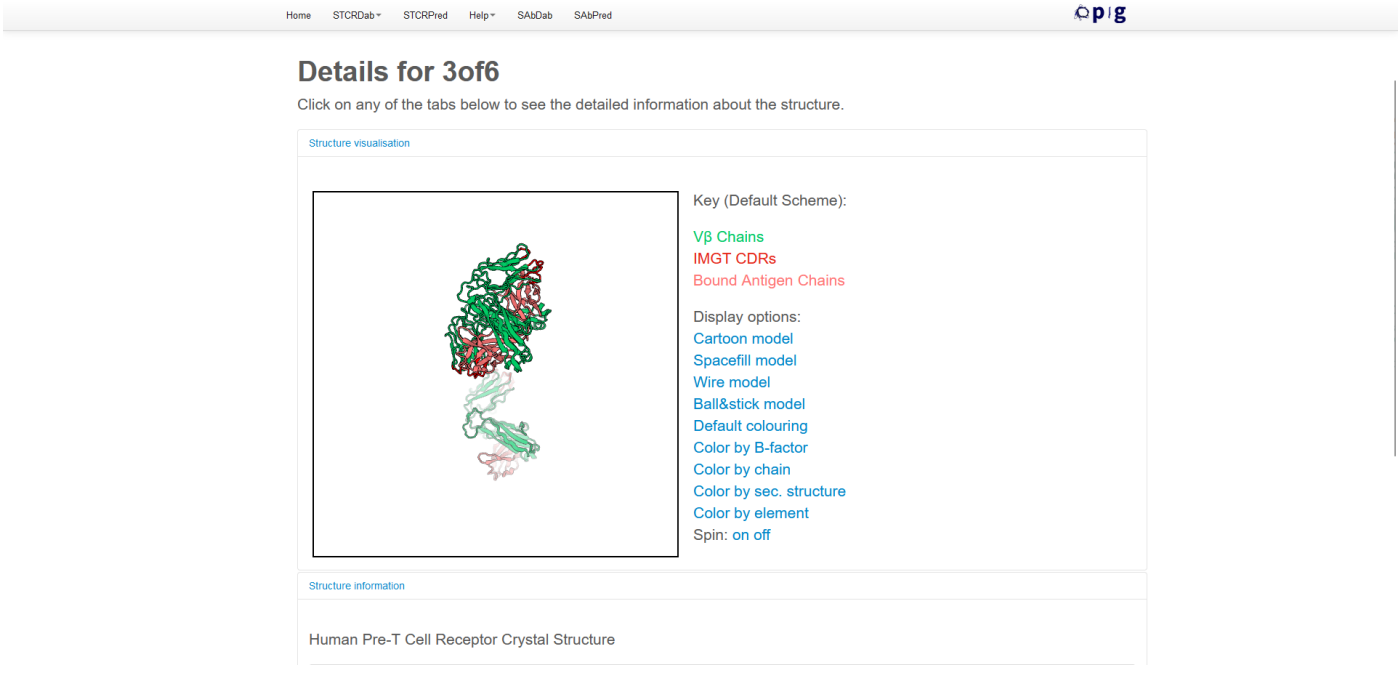

Fig 6. Structure visualization (V in green, V in blue and IMGT CDRs in grey)

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

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	FASTA file	

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	S	G	H	V	S	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
F	Q	N	E	A	Q	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
CDRB3	ASSLGQAYEQY	None	11
CDRB2	FQNEAQ	B2-6-B	6
CDRB1	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	Link
Non-annotated structure from the PDB	Link
Summary file for this TCR	Link
Link in the PDB	Link

Fig 11. Links and files for download

RESULT:

STCRDab Database was use to identify the CDR for the crystal structure of TCR A6 (PDB ID: 4GRm). The following result showed the structure visualisation for Vα chain in green, Vβ 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 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: http://bioinfo.unipune.ac.in/PFV3D/Home)

AIM:

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 chromozome 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 chromozomal 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.

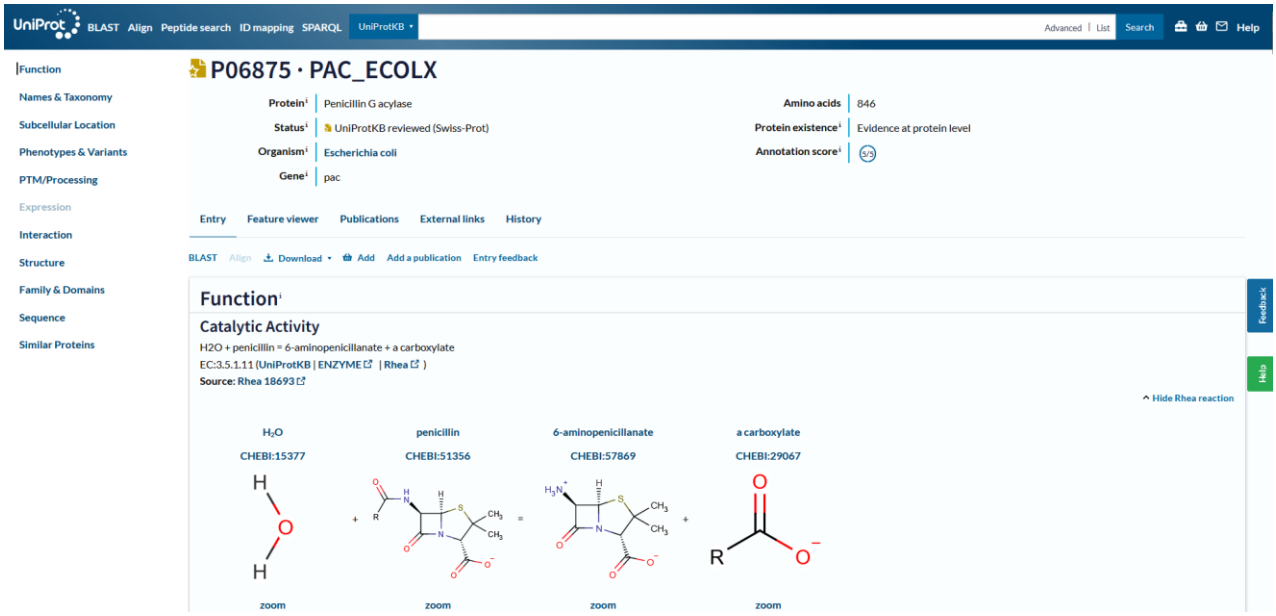

Fig 1. Uniprot page for Penicillin G acylase

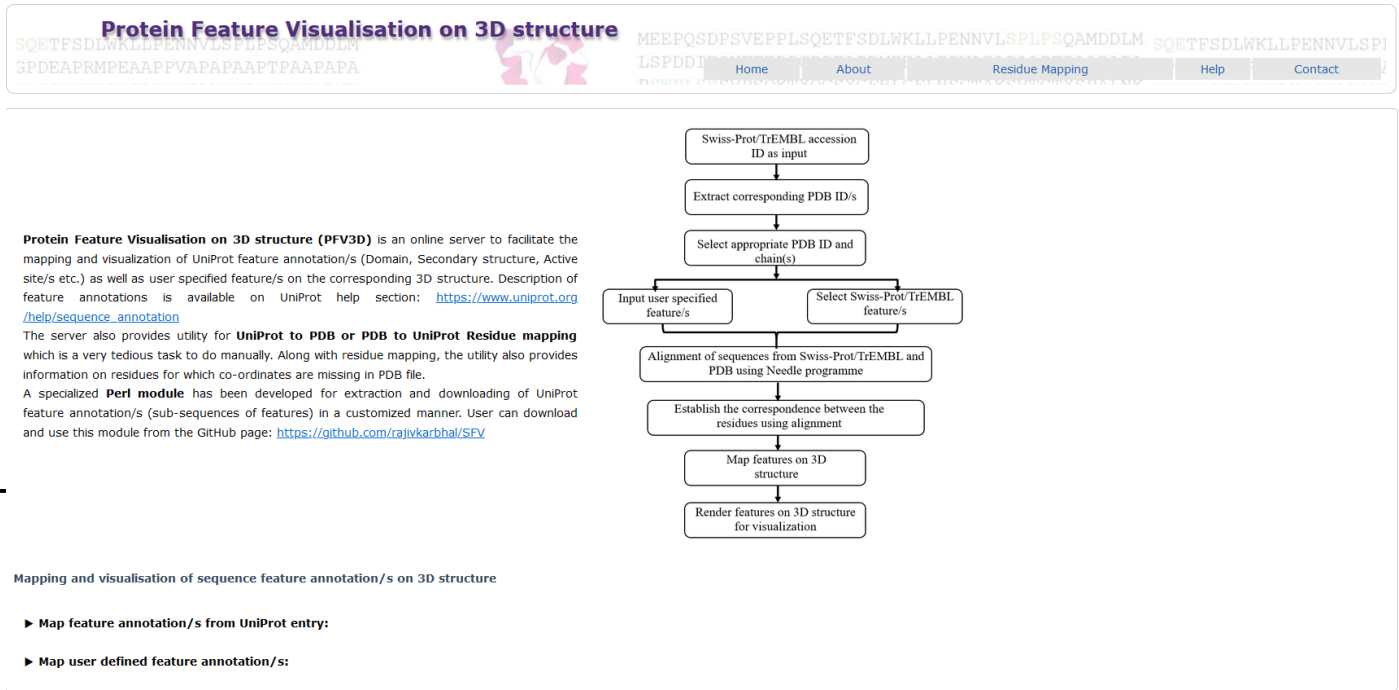

Fig 2. Homepage of PFV3D tool

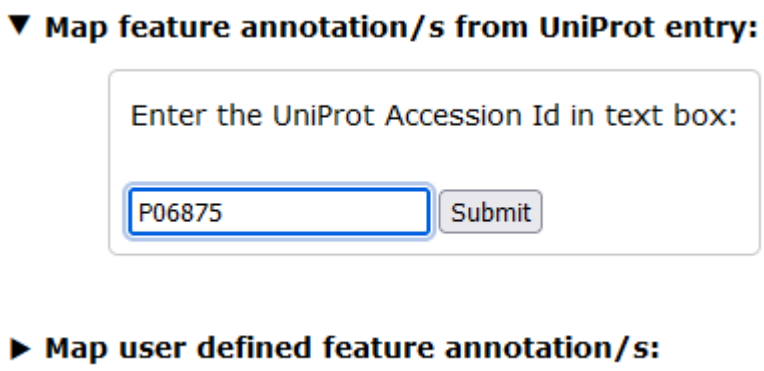

Fig 3. Mapping using feature annotiations for UniProt entry

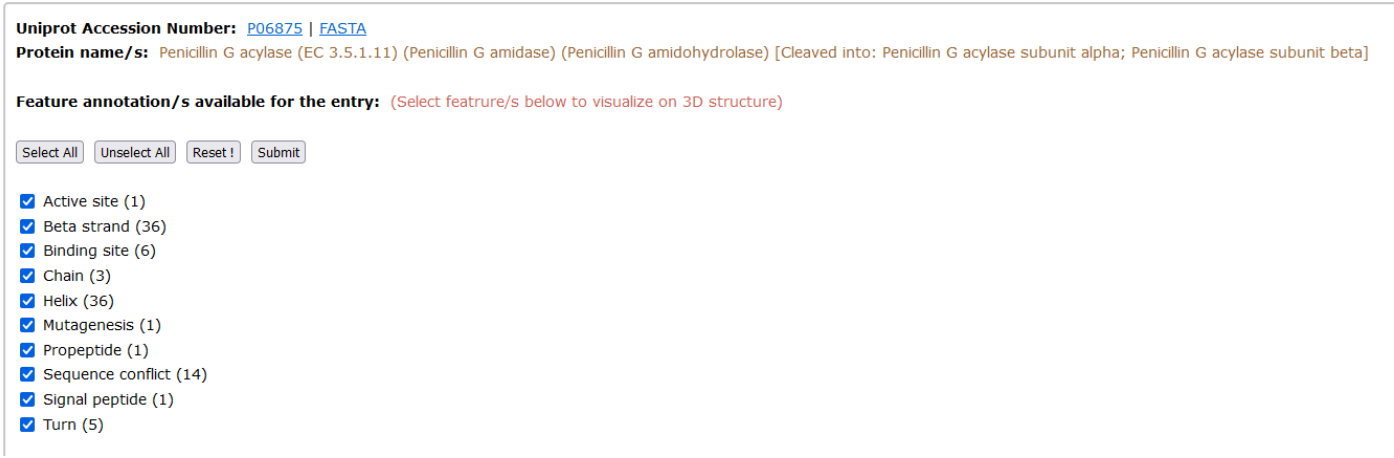

Fig 4. Features selected from available features for the entry (Beta stran, chain and Helix)

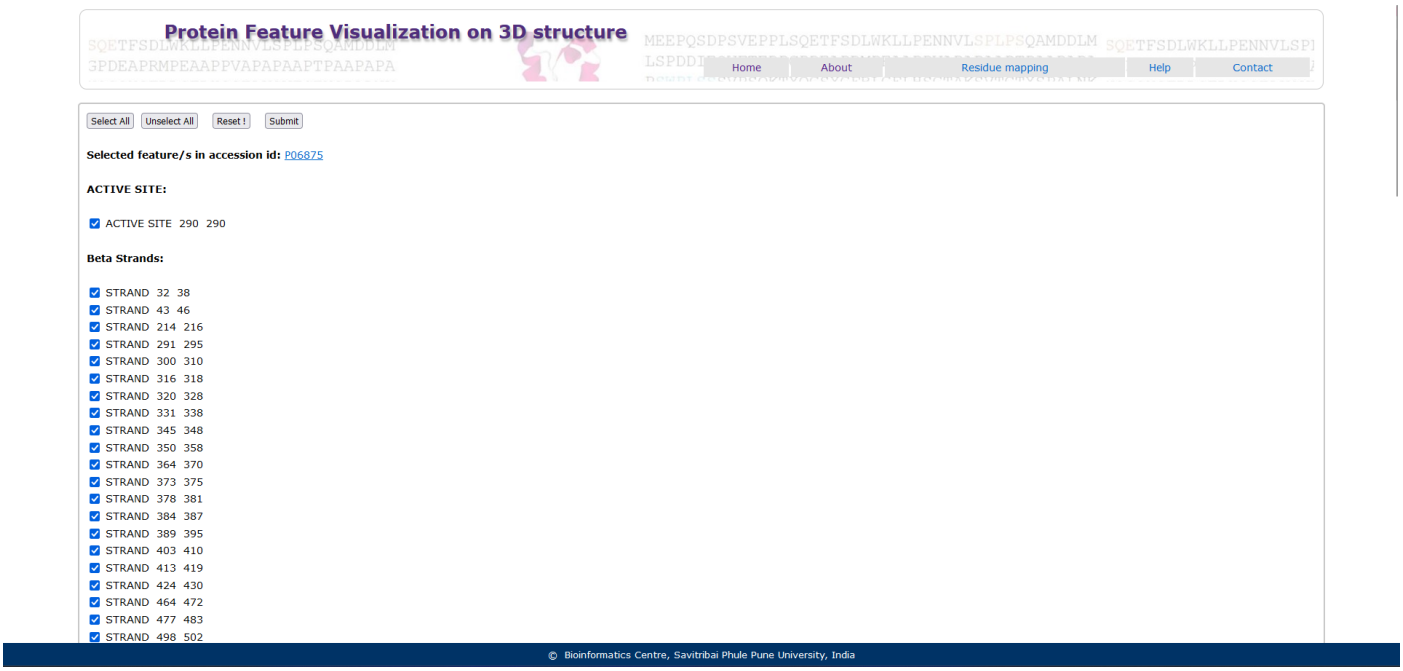

Fig 5. Features selected under beta strand, chain and helix

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

Submit

PDB ID	Resolution ▲▼	Title
◉ 1AI4	2.35	PENICILLIN ACYLASE COMPLEXED WITH 3,4-DIHYDROXYPHENYLACETIC ACID
○ 1AI4	2.35	PENICILLIN ACYLASE COMPLEXED WITH 3,4-DIHYDROXYPHENYLACETIC ACID
○ 1AI5	2.36	PENICILLIN ACYLASE COMPLEXED WITH M-NITROPHENYLACETIC ACID
○ 1AI5	2.36	PENICILLIN ACYLASE COMPLEXED WITH M-NITROPHENYLACETIC ACID
○ 1AI6	2.55	

Fig 6. Selection fo PDB id and Chain id

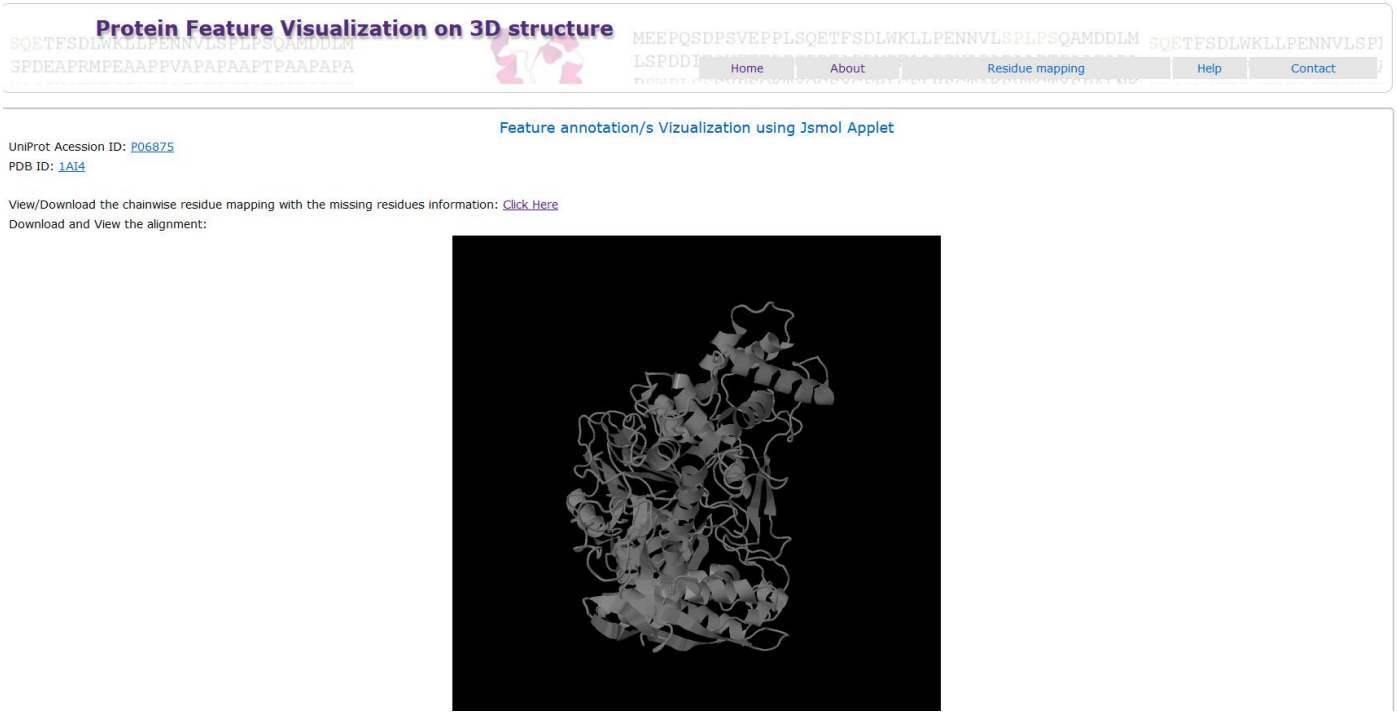

Fig 7. Visualisation of mapped features

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-proteins.htm