DATE: 29/10/2022

WEBLEM 5

Yvis Platform

AIM:

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 theV 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 α-carbon of the antibody and antigen amino acids is calculated. If the distance between two α-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.

Yvis resources: integrated tools for high-density antibody data visualization and 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 assignedgermline 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.

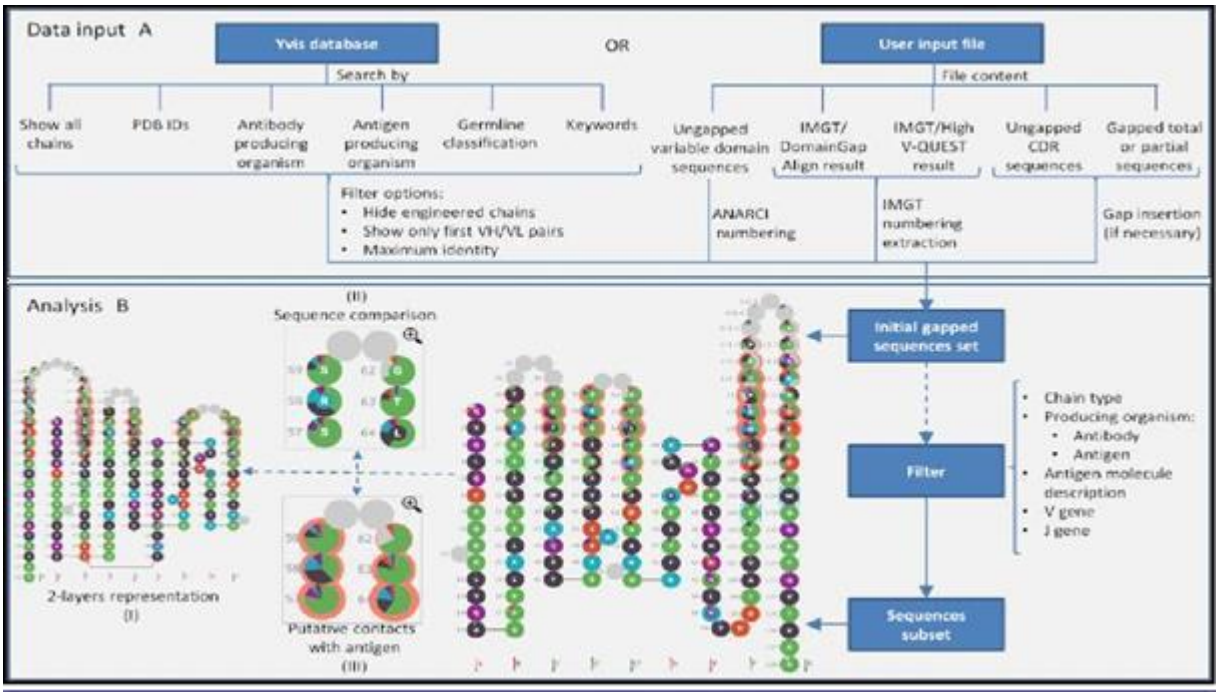

Fig 1: Yvis platform overview. (A) The data input box presents two possibilities to input sequences to be analyzed in Yvis (user input files and selection from the Yvis database).
It presents filter options for sequences from the Yvis database (redundancy and engineered chains) and actions taken by the 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 β-sheets, connected face-to-face by a disulphide bond, as shown in Figure 1. Each strand that forms the β-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).

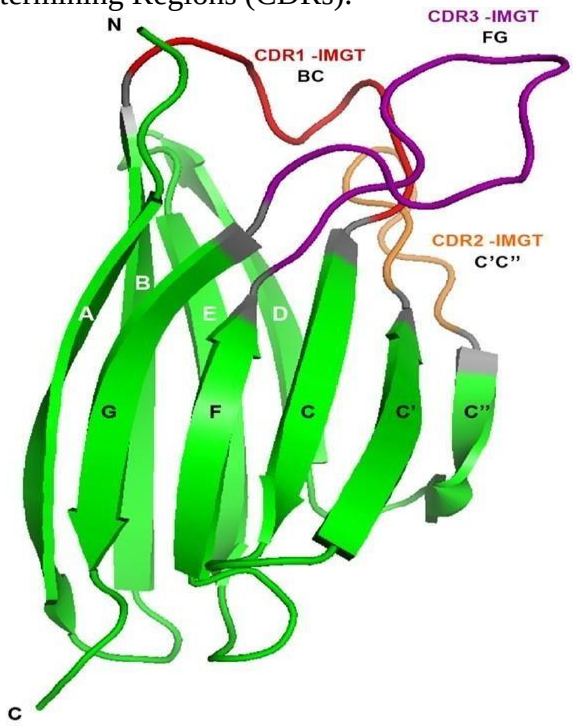

Fig 2: Cartoon representation of the variable domain structure of an antibody chain. Individual letters identify the strands (in green), and the different colours of these letters distinguish the two sheets. The three CDR loops are highlighted in red, orange and purple.

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 β-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.

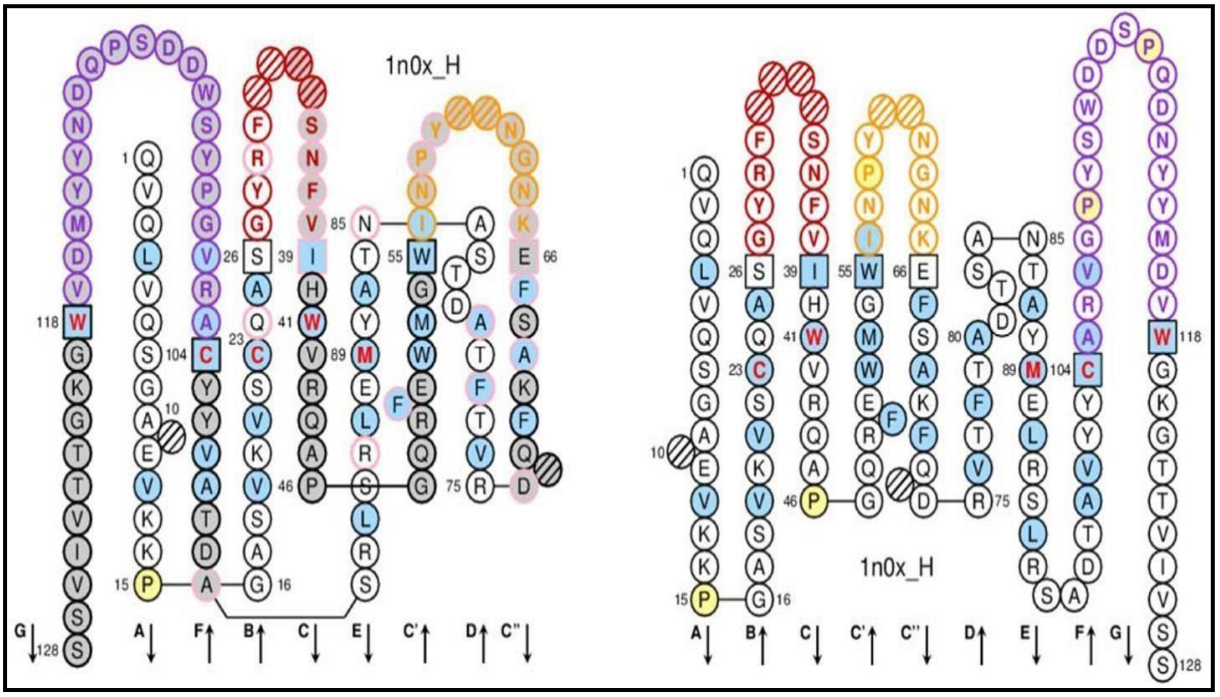

Fig 3: *Collier de Perles* visualization in two layers (left image) and one layer (right image). Individualletters and arrows identify the strands. The CDR loops are highlighted in red, orange and purple.

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

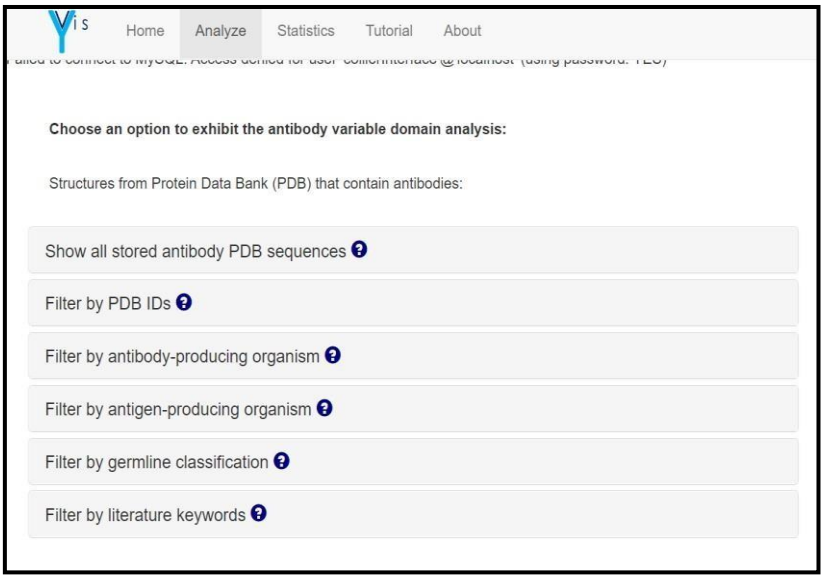

Fig 4.1: Search option for Structures from Protein Data Bank (PDB)

1. **Show all stored antibody PDB sequences**

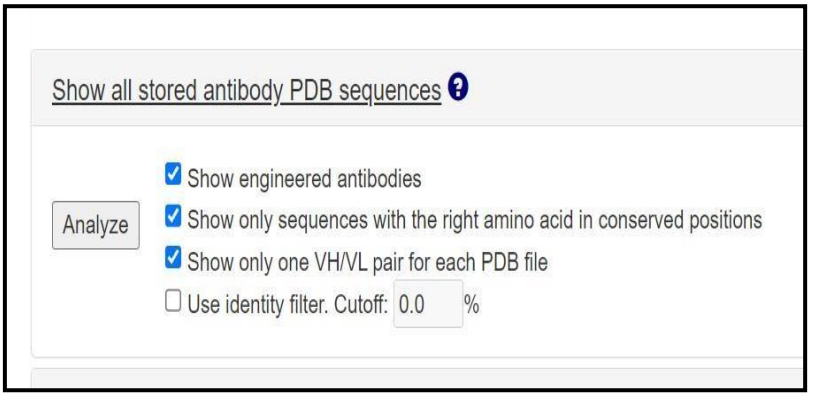

Fig 4.2: Show all stored antibody PDB sequences

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

2. Filter by PDB IDs

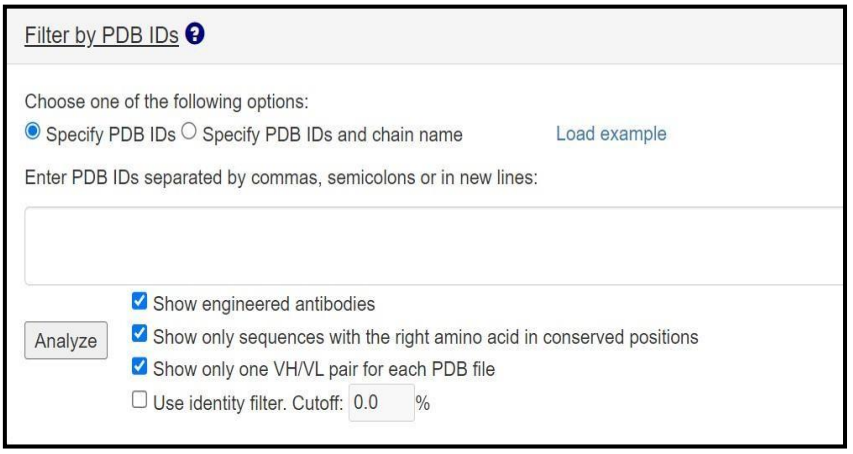

Fig 4.3: Filter by PDB IDs

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

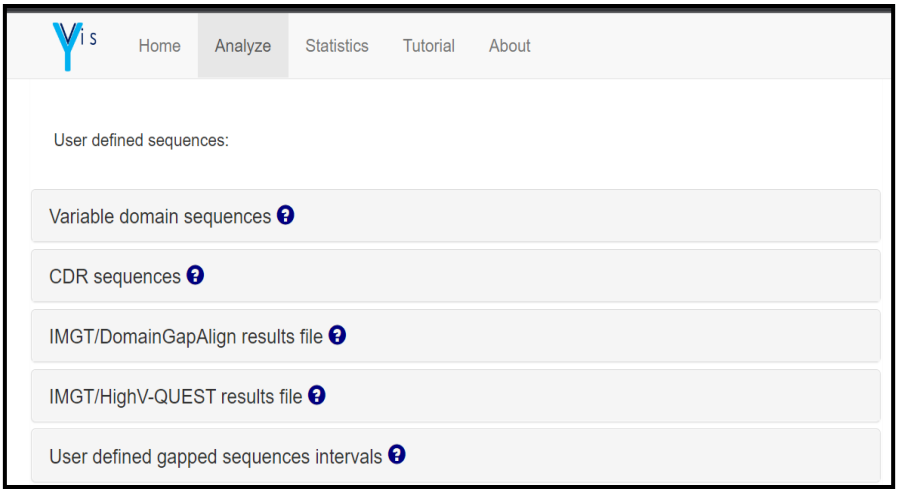

Fig 5.1: Search option for User defined sequences

1. **Antibody chain sequences**

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**

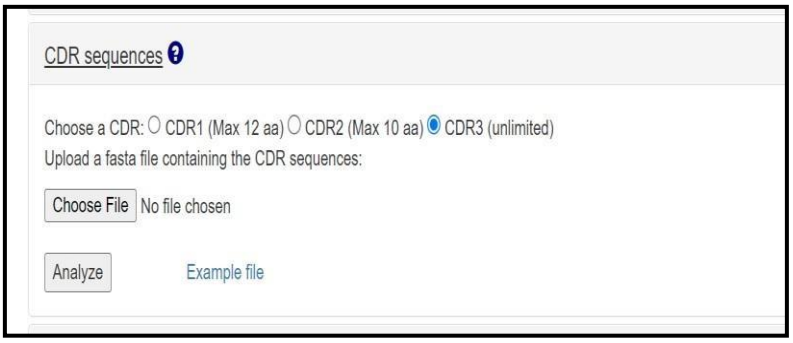

Fig 5.3: CDR Sequences

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/PMC6602444/

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

DATE: 29/10/2022

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 platfrom 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:

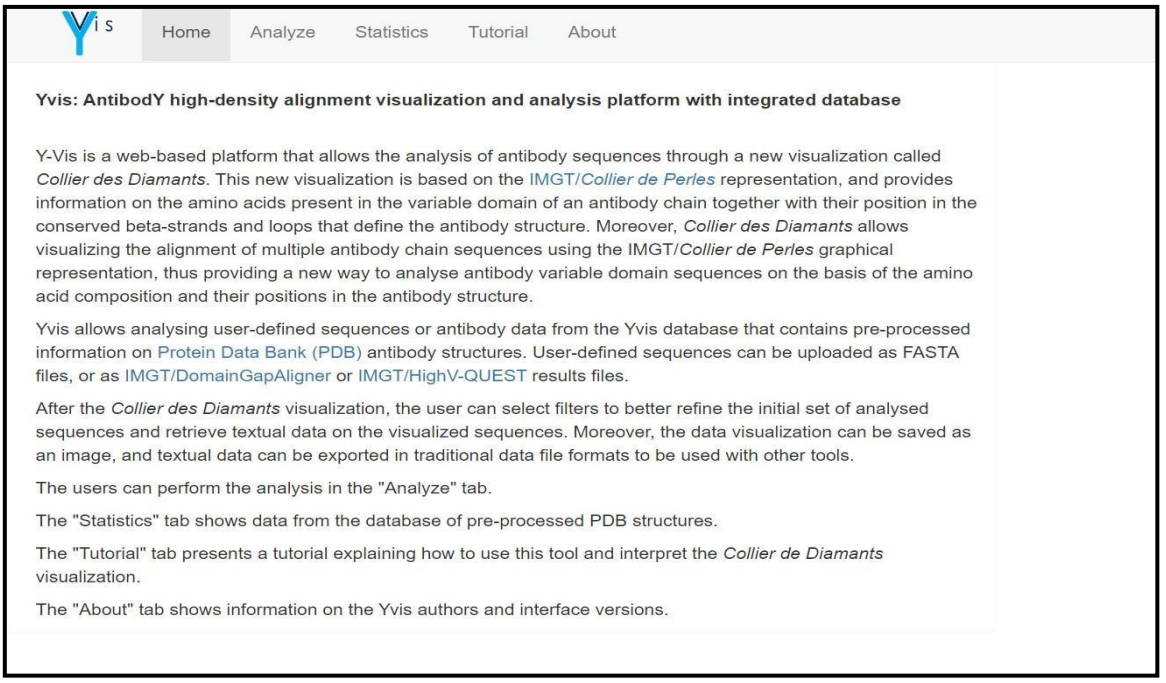

Fig 1: Homepage of Yvis Platfrom

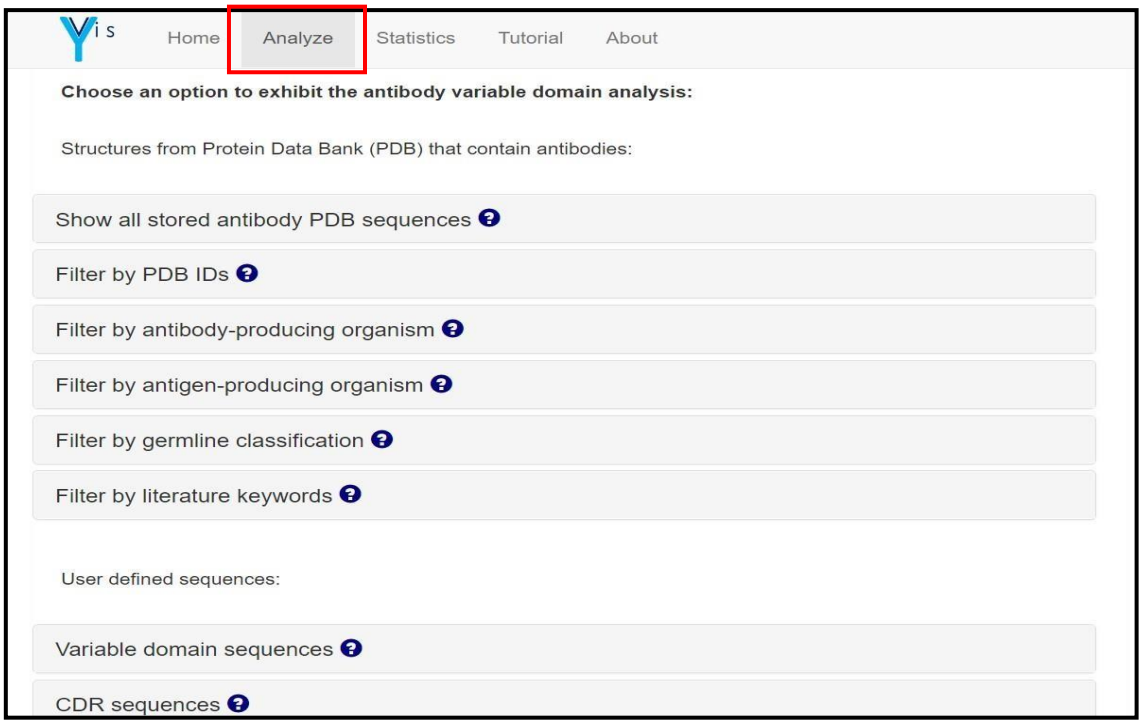

Fig 2: Analyze page under Yvis Platfrom

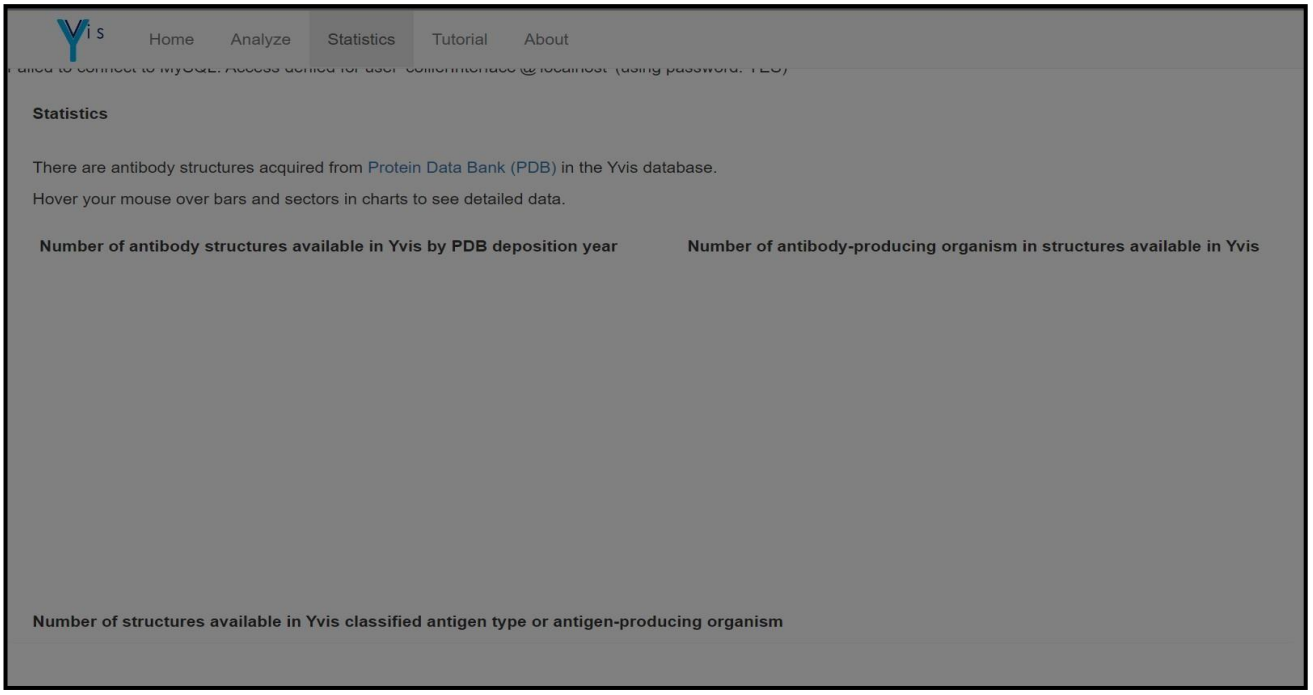

Fig 3: Statistics page under Yvis Platfrom

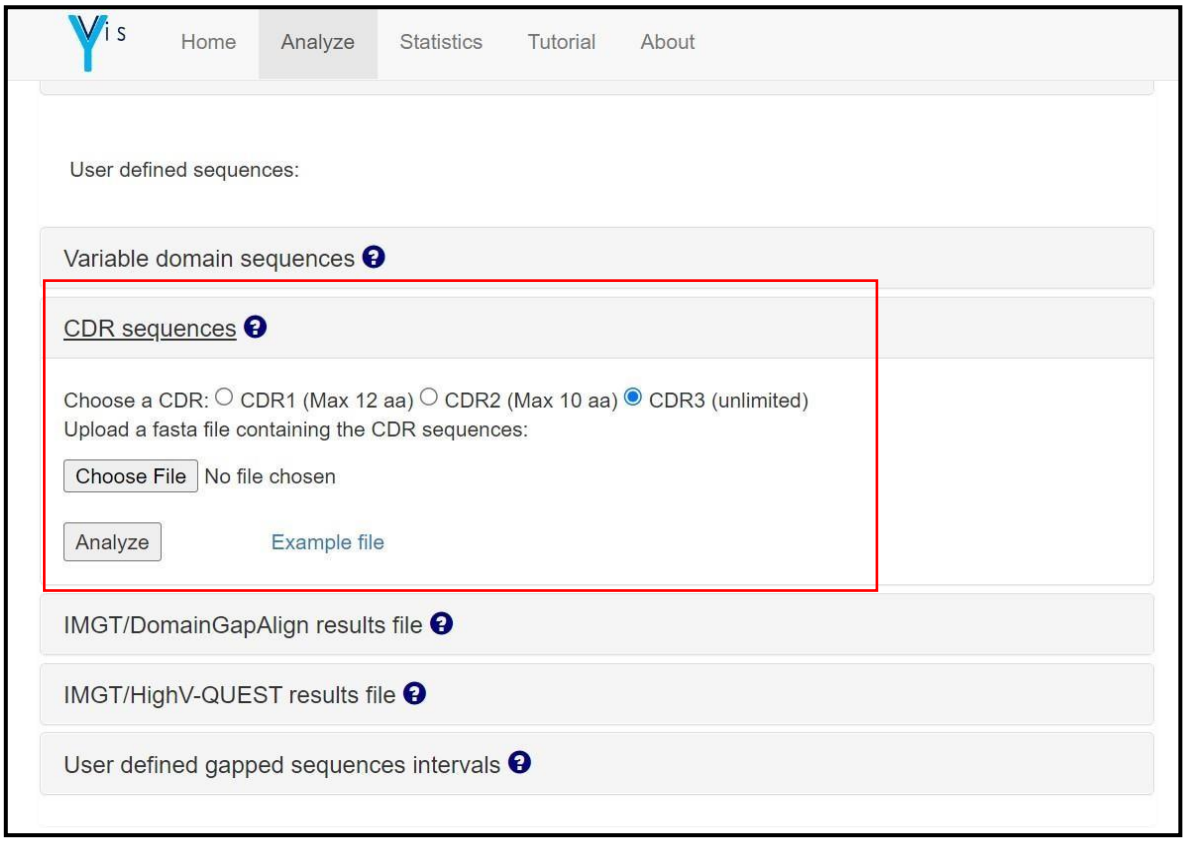

Fig 4: Search option using CDR Sequences

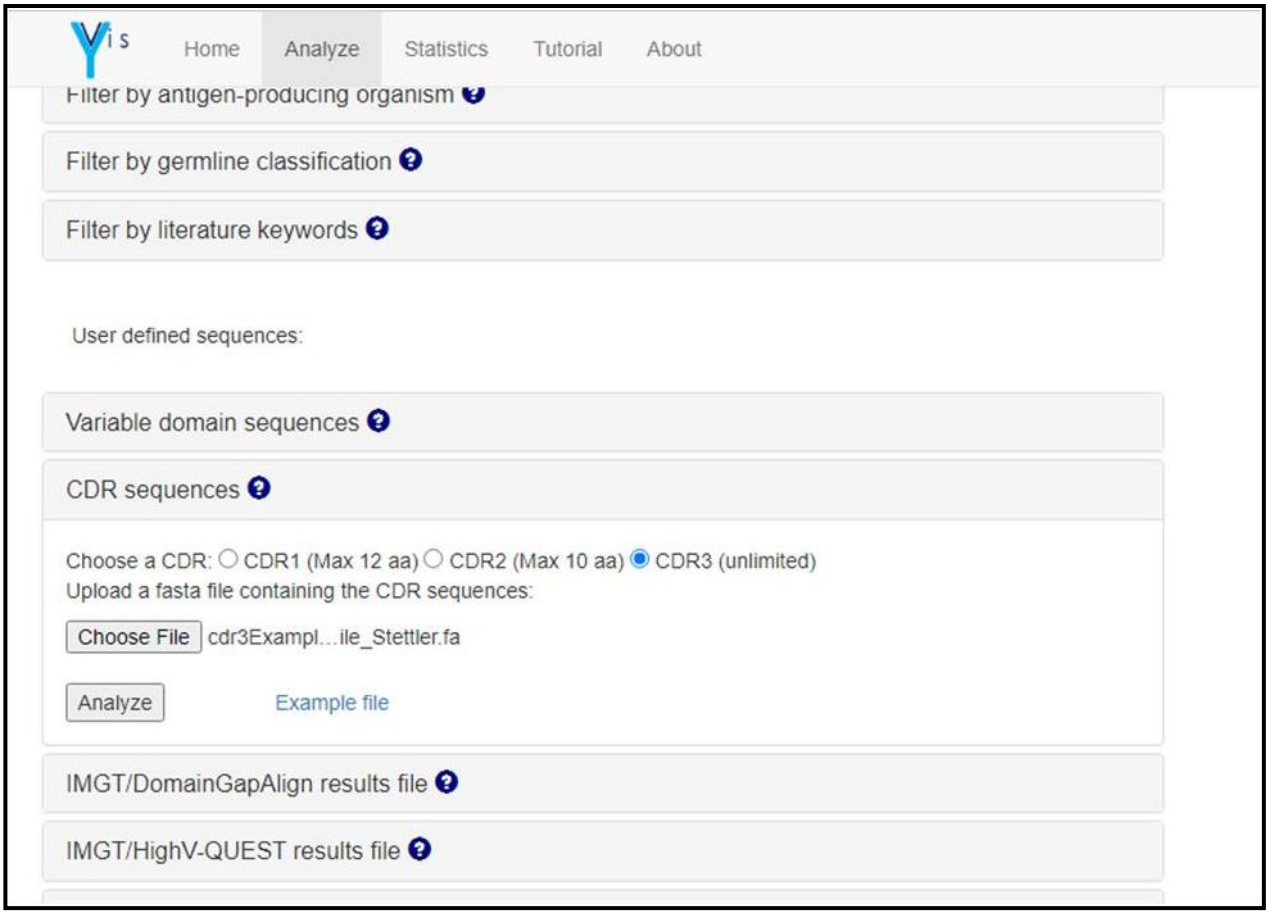

Fig 5: Upload the CDR3 Sequences file

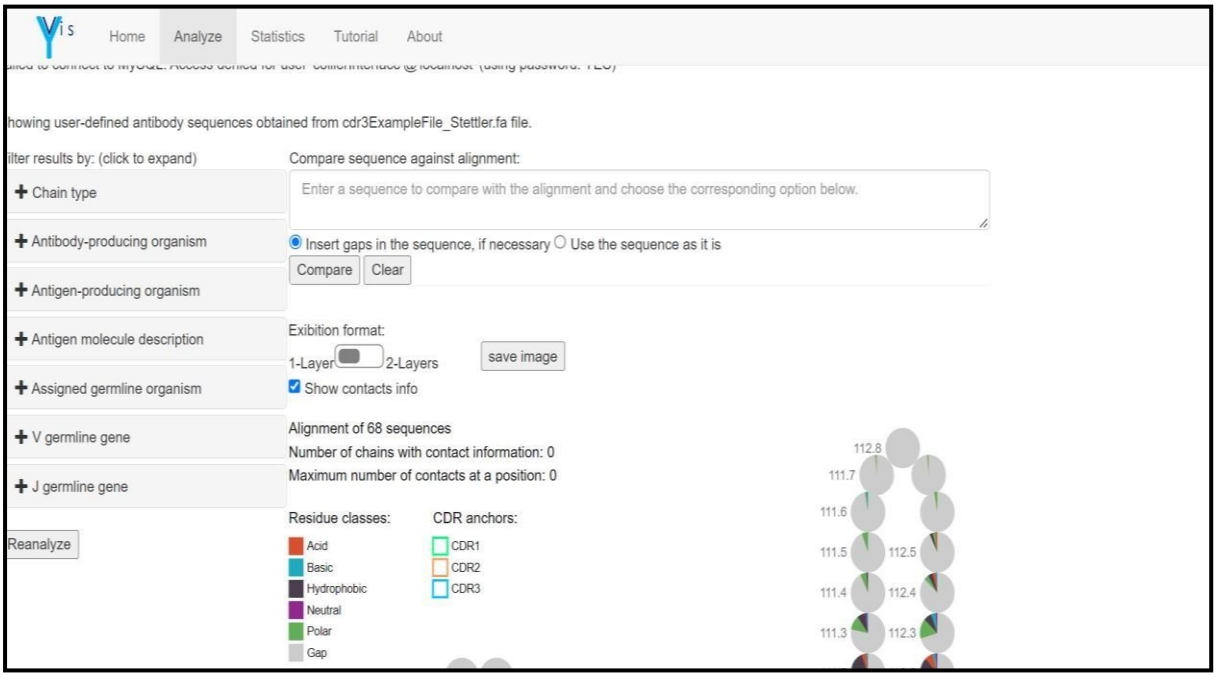

Fig 6: Result page for CDR3 Sequences fileCDR1 (Green): 27-38

CDR2(Orange):56-65 CDR3(Blue):105-118

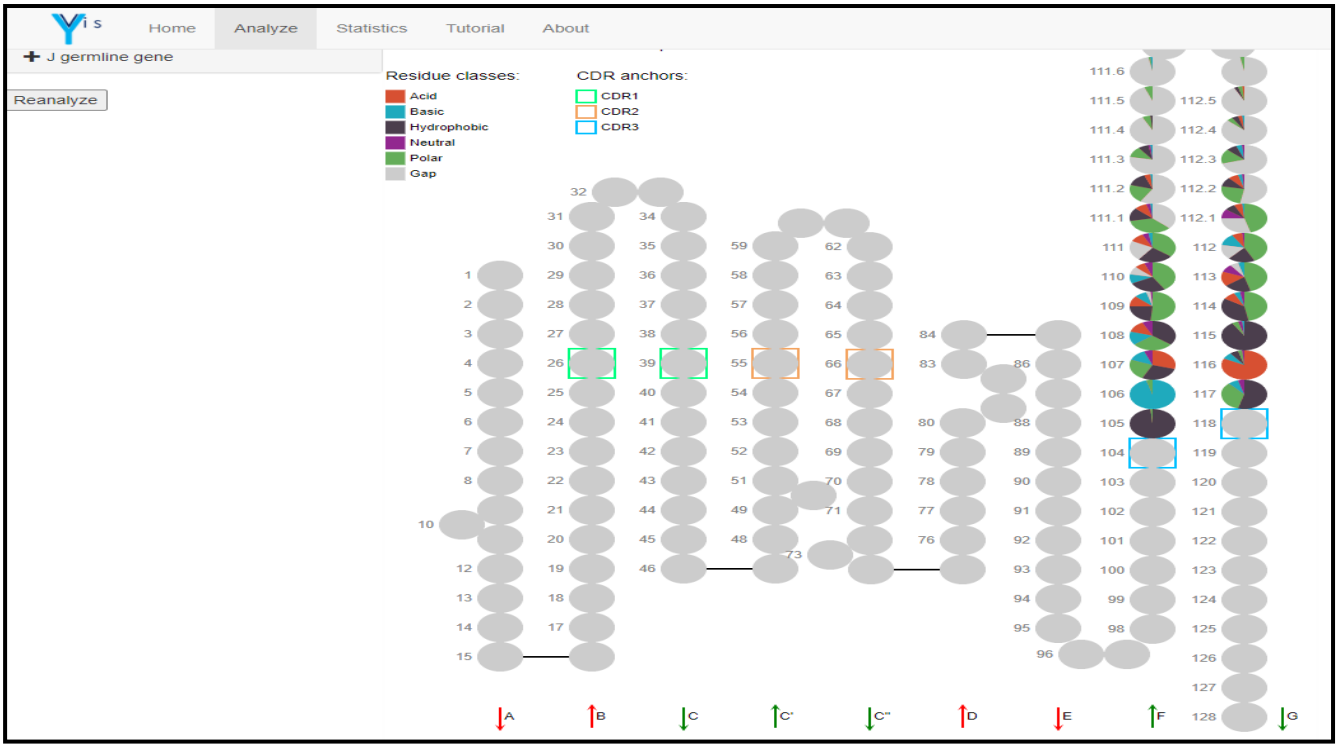

Fig 7: One layer format for CDR Insertions position between 111 and 112 in CDR3 (112.1, 111.1, 112.2, 111.2, 112.3, 111.3, etc.)

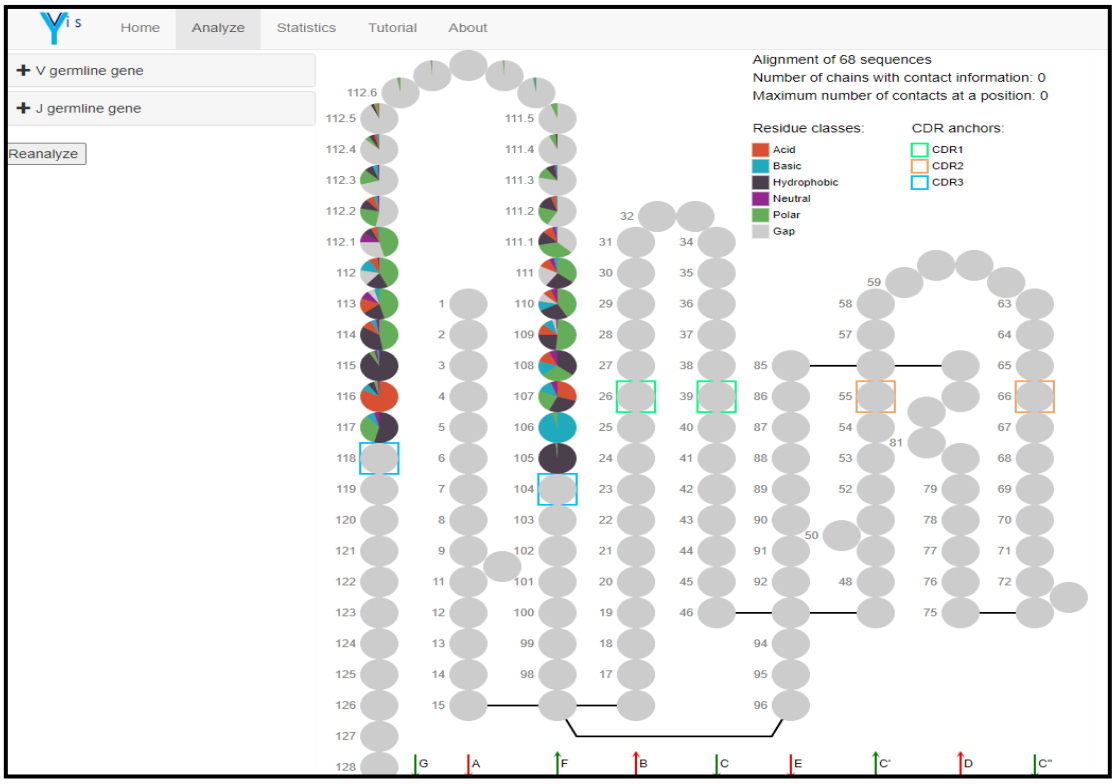

Fig 8: Two-layer format for CDR

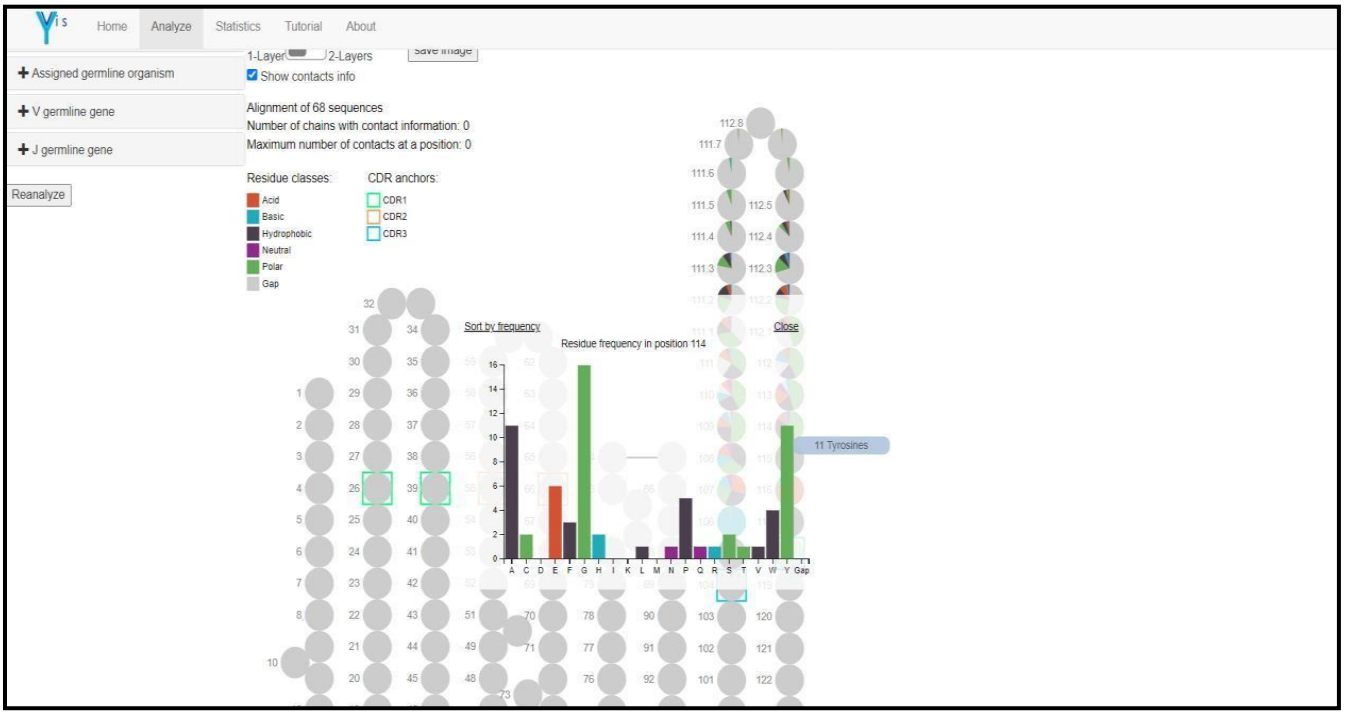

Fig 9: Graph for Residue frequency position 114 under CDR3 section Tyrosines (Green) has highest frequency

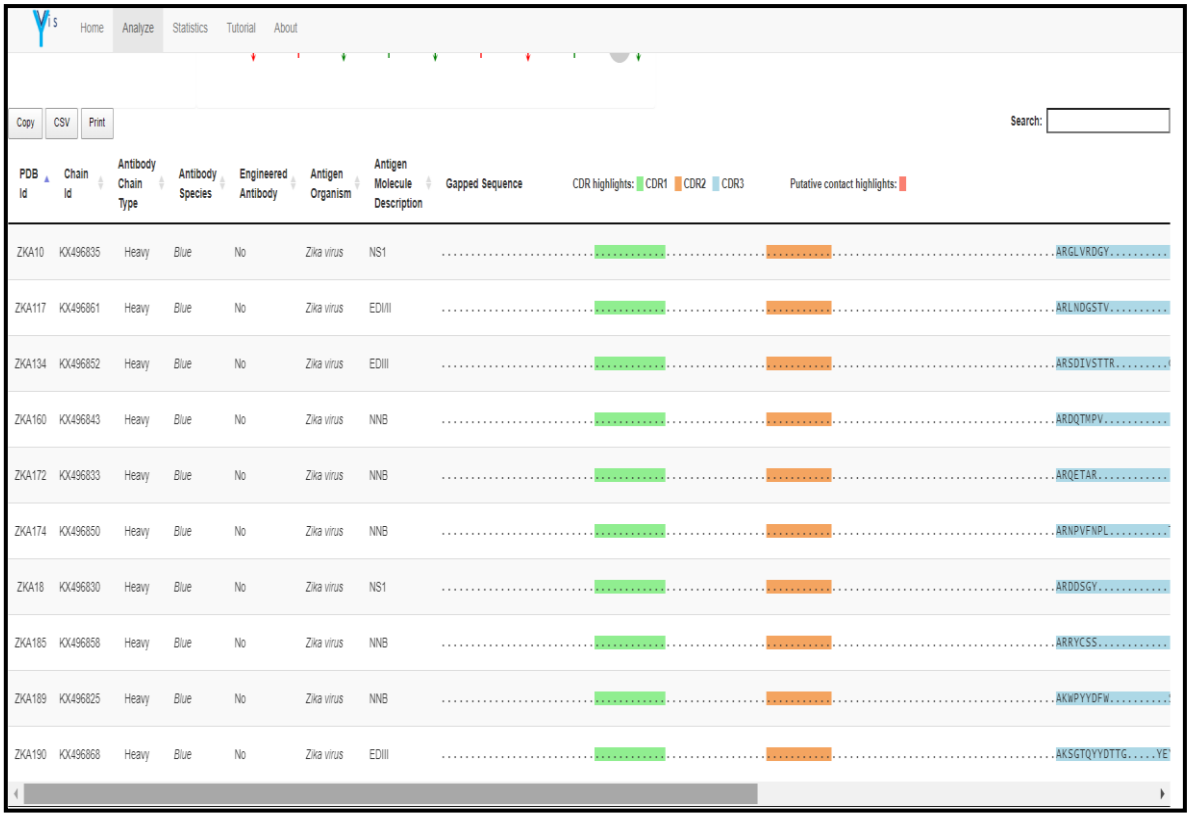

Fig 10: Sequence Alignment for CDR3 region

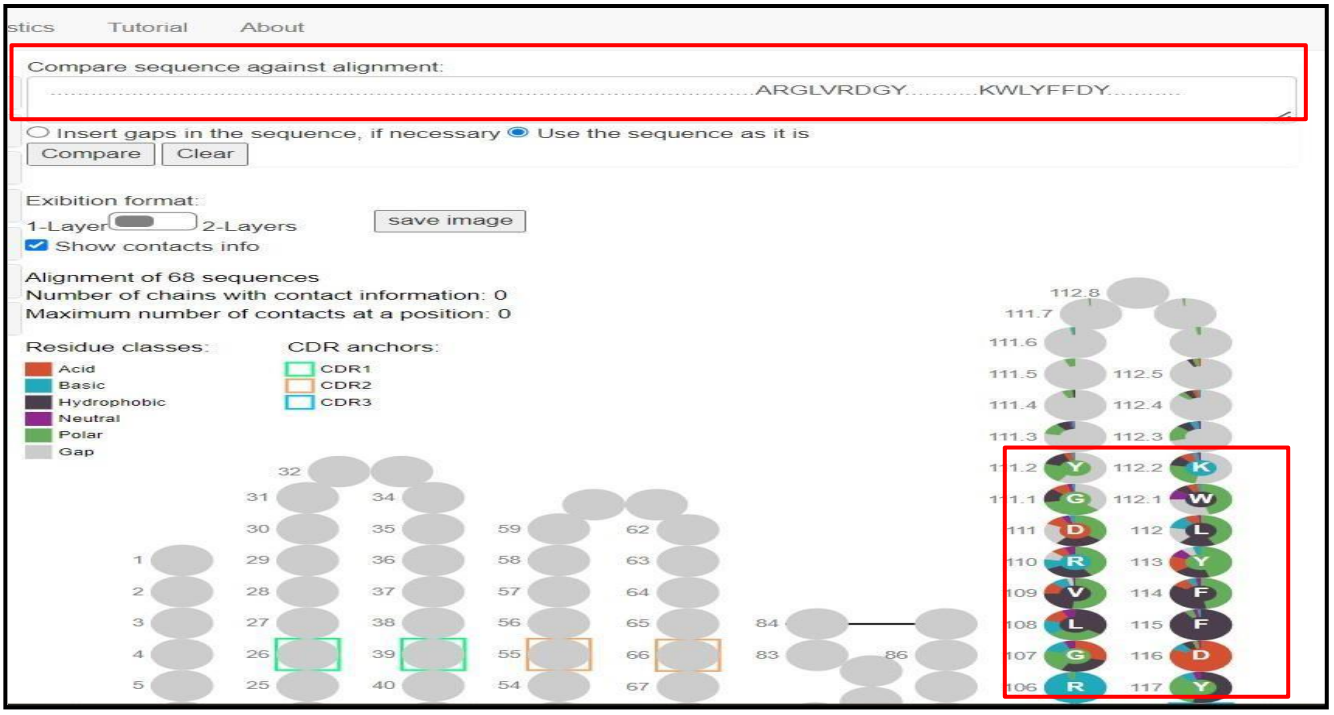

Fig 11: Comparing sequences for PDB ID ZKA10

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 informationis indicated at the top of the Collier de Diamants visualization, as well as the maximum number of contactsat 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/PMC6602444/

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