Basic Navigation

Overview

ProteoVision is webserver designed to visualize phylogenetic and structural information about ribosomal proteins in multiple dimensions. ProteoVision in operated via the main Navigation panel for selection/filtering, and contains three main applets i) Alignment Viewer for representation of multiple sequence alignments; ii) PDB topology Viewer for depiction of protein secondary structures; iii) MolStar viewer for visualization of three dimensional structures. Additionally, non-mappable statistical data (*e.g.* amino acids propensities) are visualized in a separate window using Plotly applet. Detailed description of all functions for ProteoVision is listed below. ProteoVision also contains an optional interactive guided tour with a brief description of each functional element.

Selecting phylogenetic group(s) - Caeden

Selecting a protein alignment - Caeden

Selecting a structure for mapping – Petar

Once an alignment is selected the option to select a structure for mapping becomes available. Selecting a structure is a two-step process:

1. First the user selects a 3D structure available from RCSB. Structures from the RCSB are filtered by the polymer of the selected alignment through the https://ribosome.xyz/ website API. Filtered PDB IDs are available in a dropdown menu when using the DESIRE database. When using a custom alignment, no filtering is done on PDBs and the user can write in any 4 letter PDB ID.
2. Once a PDB ID is selected the available chains of that structure are filtered by the polymer identity present in the selected alignment using. No filtering is done when using a custom alignment and all chains of the structure are available. The user can select one chain which will load the topology and MolStar viewers.

Selecting the data to map – Holly

The user has the option to map either provided attributes (described below) or custom data onto the 2D and 3D viewers. Available attributes may be selected from a dropdown menu in the lower right corner of the topology viewer. There is also an option in the sidebar to upload custom mapping data. This data should be uploaded as a .csv; an example file is provided on the ProteoVision site. Once the custom data has been uploaded it will be added as a mapping option in the dropdown menu.

Advanced features

Frequencies – Aparna

Once an alignment has been selected, the user can click the **show amino-acid frequencies** button to display amino acid frequencies for that alignment. Amino acid frequencies are calculated by processing the alignment fasta using Biopython, and the resulting data is displayed as a faceted boxplot through the Plotly graphing library. Every point on an amino acid frequency plot represents the frequency of a single amino acid for a single species within the alignment. For example, the uL10 protein of the bacteria Streptococcus pneumoniae R6 is 9.27% valine. The species and amino acid frequency associated with each point can be viewed by hovering over the point.

If a user chooses from the **Select a substructure** dropdown, the amino acid frequencies will be calculated for specific 3D substructures within the polypeptide. Users to choose to only display helix residues, coil residues, or strand residues.

Masking - Holly

Once a polymer has been selected, the user may select to mask/unmask 2D and 3D residues. This allows for coloration of only selected residue ranges specified by the user. All other residues will be colored in white and have hovering functions disabled. The overall structure of the protein will still be visible.

Range selection - Holly

The user also has the option to cut/uncut 2D and 3D residues. This allows the user to view only the portion of the protein specified by the entered residues. The rest of the protein structure will be removed rather than colored white as in the masking feature.

Synchronization of navigation between the panels

Navigation between all panels is synchronized. Hovering over an alignment position highlights the corresponding residue in the Topology and MolStar viewers. Reversely, hovering over a residue in the Topology or MolStar viewers highlights the alignment and the other structural viewer. When the amino-acid frequencies are shown, hovering over the datapoints on the graph highlights the current species in the alignment viewer.

Guide

ProteoVision offers an interactive guide that demonstrates the steps a user can take to fully utilize ProteoVision capabilities. The guide always starts on the first visit of a user to our website and can be launched at any time with the **Help** button. The user can step through the guide using the keyboard arrow keys or the **Next** and **Previous** buttons on the guide pop-ups. Ending the guide with **Skip tour** button will erase the current session and reset the viewports.

Saving

Saving the alignment (fasta & image) – Petar

The alignment retrieved from the DESIRE database can be downloaded in fasta format with the **Download alignment** button. This will save the entire alignment to the user's machine. The current viewport of the alignment can also be saved as a png image with the **Download alignment image** button. Only the currently viewable region of the alignment will be saved in this way.

Saving secondary structure (svg) - Holly

The secondary structure image may be downloaded as a .svg file by clicking the "S" button in the bottom right corner of the topology viewer.

Saving 3D structure (png) - Holly

The image of the 3D structure may be downloaded as a .png file or viewed in browser by clicking the Screenshot/State button, which appears as a wheel in the upper right corner of the 3D viewer. Upon clicking this button, the user may choose to either keep the default white background or make the background transparent by turning transparency to "On." They may also choose to either include or exclude 3D axes, which show the orientation of the protein.

Saving computed data (csv) – Caeden

Saving frequencies (png) Aparna

The amino acid frequency plot can be saved as a .png file by hovering over the image and clicking on the camera icon in the top center.

Saving a ProteoVision session

At any point, the user can save their progress with the **Save session** button. This will download a json file that holds information about the currently loaded alignment, structure, and custom mapping data. The session file does not save information about the masking and truncation ranges.

ProteoVision Data

Phylogeny (SEREB)

The subset of 152 species from the SEREB (Sparse and Efficient Representation of Extant Biology, Bernier, MBE 2018) database was organized into a phylogenetic browser using a tree topology from the Banfiled lab (https://doi.org/10.1038/nmicrobiol.2016.48)

Alignments – Claudia

Each ribosomal protein has an associated MSA. The MSAs were built using as reference a MSA generated with MATRAS (https://doi.org/10.1093/nar/gkg581) from a multiple structure superimposition. Then, amino acid sequences of species from the SEREB database were added to the reference alignment using MAFFT (https://doi.org/10.1093/bioinformatics/bts578).

2D maps – Anton

Topologies of the protein secondary structures (Laskowski 10.1093/nar/gkn860) were exported into PDB topology viewer as APIs provided by EMBL-EBI PDBe and available at https://www.ebi.ac.uk/pdbe/api/doc/

3D Structures – Anton

3D structures were fetched from the PDBe using the APIs of EMBL-EBI coordinate server https://www.ebi.ac.uk/pdbe/coordinates/. The selection of ranges was implemented using the syntax of the LiteMol's coordinate server https://coords.litemol.org/

Alignment associated data (Fold, Phase)

10.1093/molbev/msx086

Available attributes:

Amino Acid probabilities

Amino acid frequencies in each column of an MSA were adjusted for presence of gaps. Thus, the gap frequencies were prorated, and were treated as a uniform distribution among all possible amino acid characters, such that a single character in a gap counts as 0.05, as described by Bernier et al. (10.1093/molbev/msy101).

Shannon Entropy

The Shannon entropy (as well as all properties listed below) was computed from the gap adjusted probabilities as:

$$H_{SE}(n) = -\sum_{i=1}^{c} p_i(n)\log_2 p_i(n) \cong -\sum_{i=1}^{c} f_i(n)\log_2 f_i(n)$$

Two group comparison (TwinCons)

In case of two groups selected in the phylogeny browser, ProteoVision provides an additional option to compute an in house developed score, TwinCons. TwinCons is computed for a single position of the MSA that compares two pre-defined groups (represented by vectors of the gap adjusted amino acid frequencies) based on their similarity defined by the pre-computed substitution matrix. TwinCons, represents the transformation price between the two vector columns related by the substitution matrix.

Charge, hydropathy, hydrophobicity, polarity, mutability

The physico-chemical properties for each position within an MSA are computed as average properties for a given distribution of the amino acid frequencies. The tabulated values for each property were obtained from the available literature: a) charges (https://doi.org/10.1186/1758-2946-5-39); b) hydropathy (doi: 10.1016/0022-2836(82)90515-0); c) hydrophobicity (doi: 10.1093/**protein**/5.5.**373**.); d) polarity (doi: 10.1093/bioinformatics/8.3.**275**); e) mutability (doi: 10.1016/0022-5193(**68**)90069-6.).

Color Schemes - Petar

Import User supplied Data

Import an external dataset for a pre-selected alignment – Claudia

Once an alignment and a structure have been selected, the option becomes available to upload custom data for mapping onto the selected structure for visualization in the topology and MolStar viewers. The data should be supplied in a csv (comma-separated values) file format. The first row of the csv file contains the headers for each column. An Index column should always be indicated. All other additional columns require a unique header definition. The Index column has the residue number to which the user-supplied data will be mapped. The rest of the columns have the data that will be mapped in the form of numerical values. Once a correct csv file has been uploaded, the selected structure will be colored in the topology and MolStar viewers according to the values in the csv file. Different columns in the csv file will appear as different Annotations in the dropdown menu in the lower right corner of the topology viewer.

Upload an external multiple sequence alignment. – Claudia

The User Upload mode allows the user to use all the RiboVision3 features with an external MSA. An alignment in a fasta file format must be selected and then uploaded with the Upload alignment button. Once an alignment is uploaded it will be displayed in the Alignment viewer. The steps for structure selection, mapping, attribute calculation, and saving are the same as previously described.

Import a saved ProteoVision session file (Petar)