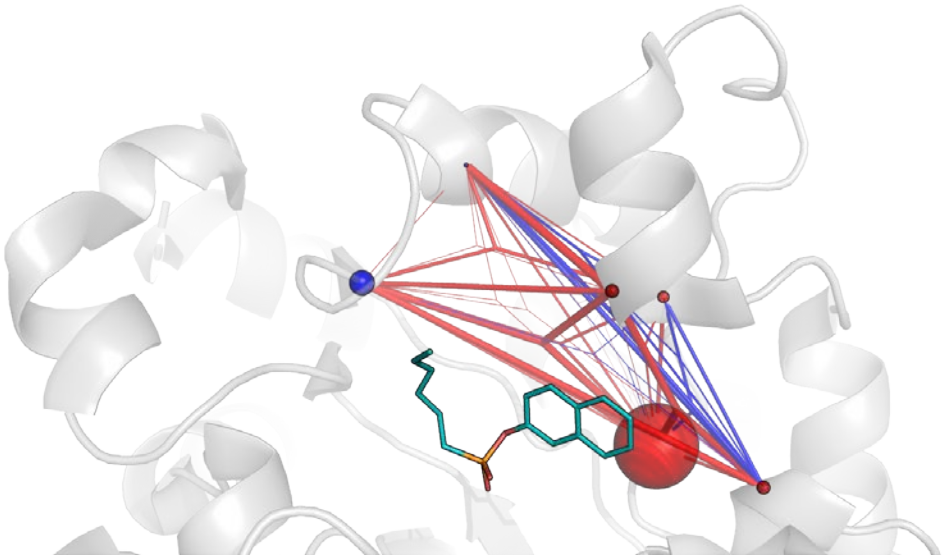
John Chen 2020/05/15

Overview

The 'visualize_epistasis.py' script adds a command to pymol that can read in a set of epistasis data and visualize it on a target crystal structure. First order epistatic effects (mutation effects at a single position) are displayed using the scale of the alpha carbon. Second order epistatic effects (interaction between two positions) are displayed by the thickness of the line connecting the two positions. Similarly, higher order effects (interaction between three or more positions) are displayed by the thickness of the lines connecting all positions at a central point. Epistatic effects that are positive are colored in red while negative effects are colored in blue.

The following shows the 1st, 2nd and 3rd order epistatic effects mapped on to the phosphotriesterase structure (PDB:4pcp).



Usage

Data preparation

The pymol software is designed to be used with the output of the 'simplex_regression.py' script, which analyzes the epistatic effects using a linear model approach. Interaction data should be prepared in two column, comma separated format, with the 1st column indicating the interacting positions in the protein and the 2nd column representing the magnitude of the epistatic effect. Additional columns may be present, but only the first two columns are considered by the program. If inputting custom data from a different source, simply start entering the data starting on the third row and leave the first two rows empty (the program expects the data to start on the third row). This can be done by arranging the data in a spreadsheet program (Excel, LibreOffice) and saving as a '.csv' file with comma as the delimiter. The input data is shown below.

indices	effect	R2

INTERCEPT	0.751995	0
233	0.111109	0.016065
254	0.683495	0.607938
271	0.07588	0.007493
272	0.109818	0.015694
306	-0.13771	0.02468
313	0.014318	0.000267
233|254	0.000297	1.15E-07
233|271	-0.02723	0.000965
233|272	-0.0108	0.000152
233|306	0.111943	0.016307
233|313	-0.10053	0.013151
254|271	-0.13799	0.024781
254|272	-0.05889	0.004513
254|306	0.004786	2.98E-05
254|313	-0.02658	0.000919

Make sure that the position numbers match the positions in the PDB file, as PDB files are not guaranteed to use the same numbering scheme even if it is the same protein.

Loading the script

To load the command into pymol type the following:

run visualize_epistasis.py

This needs to be done every time a new pymol session is started.

Preparing the structure

Before using the new 'visualize_epistasis' command, clean up the structure in pymol to make it easier to handle later. Some recommended things to do (type the command in quotes into the python command line):

1. Hide unwanted atoms using 'hide everything'
2. Display the target molecule 'show cartoon'
3. Turn all parts of the structure into loop representation, as this gives the most space for visualization. Type the command 'cartoon loop'

Using the command

The run the 'visualize_epistasis' command using default settings, simply supply the path to the data file and the pymol object you wish to target by typing the following command into the pymol command line:

visualize_epistasis 'file_path' 'target_structure'

For example:

visualize_epistasis PXE_Terms.csv, 4pcp

Note that in this case, the data file 'PXE_Terms.csv' is in the same folder as the pymol session. In the case that it is not, either change the pymol directory to the location of the file using 'cd path/of/file/' or load the file's full path 'C:/Users/path/to/file/PXE_Terms.csv'. The target structure 4pcp is the name of the pymol object that houses the 4pcp structure. The command will apply the visualization on all chains in the structure. Also note that the scales are normalized to the minimum and maximum absolute values of epistatic effects, which the smallest value assigned to the minimum scale and the highest value assigned to the maximum scale.

The command creates a selection for each epistatic interaction in the data. These can be shown or hidden by clicking on their labels in the pymol selections panel.

The 'visualize_epistasis' command has the following inputs:
visualize_epistasis file, obj, max_atom_scale=2, min_atom_scale=0.5, max_line_scale=2, min_line_scale=0.5, pos_color=[1,0,0], neg_color=[0,0,1], backbone_color=[0.8,0.8,0.8]

- 'file' is the path to the data file. Required.
- 'obj' is the name of the pymol object to target. Required.

The following inputs are optional:

- 'min_atom_scale' and 'max_atom_scale' set the limit for atom scale (first order effects) in visualization.
- 'min_line_scale' and 'max_line_scale' set the limit for line scale (second and higher order effects) in visualization.
- 'pos_color' and 'neg_color' set the color used for positive and negative epistatic values, respectively. The inputs are given in RGB format, where each value is a range from 0 to 1.
- 'backbone_color' sets the color of the cartoon representation. Input is in RGB format.

Before explaining specifically how to input each option, I will provide a brief note on PyMol command formatting. When running commands from the python command line, the command has the following structure.

command argument1, argument2, argument3

To make use of optional arguments either supply the arguments in the same order as listed above such as the following:

visualize_epistasis PXE_Terms.csv, 4pcp, 4, 0, 3, 1

The above will set max and min atom scale to 4 and 0, respectively, and set max and min line scale to 3 and 1, respectively. The settings that occur after are left at their default settings.

To provide an option that is later in the list, but without providing earlier options, specify the option by name like the following:

```
visualize_epistasis PXE_Terms, 4pcp, max_line_scale = 3, neg_color = [1, 0.5, 0]
```

The above will set the max line scale to 3 and the color for the negative effects to a slight orange color, without affecting any of the other settings. When providing arguments by name, the order does not matter so long as all ‘named’ arguments are provided after the unnamed or ‘positional’ arguments. For example, the following will give the same results as the example above.

```
visualize_epistasis PXE_Terms, 4pcp, neg_color = [1, 0.5, 0], 4pcp, max_line_scale = 3
```

A combination of both methods can be used, so long as the arguments are in the correct order up to the point before providing named arguments, after which the order does not matter.

```
visual_epistasis PXE_Terms, 4pcp, 4, 0, neg_color=[1, 0.5, 0], max_line_scale=3
```