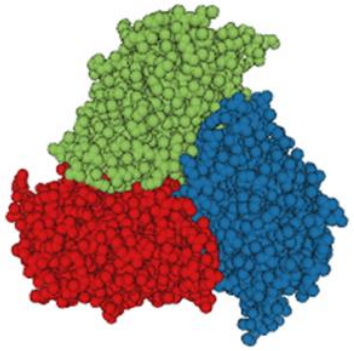
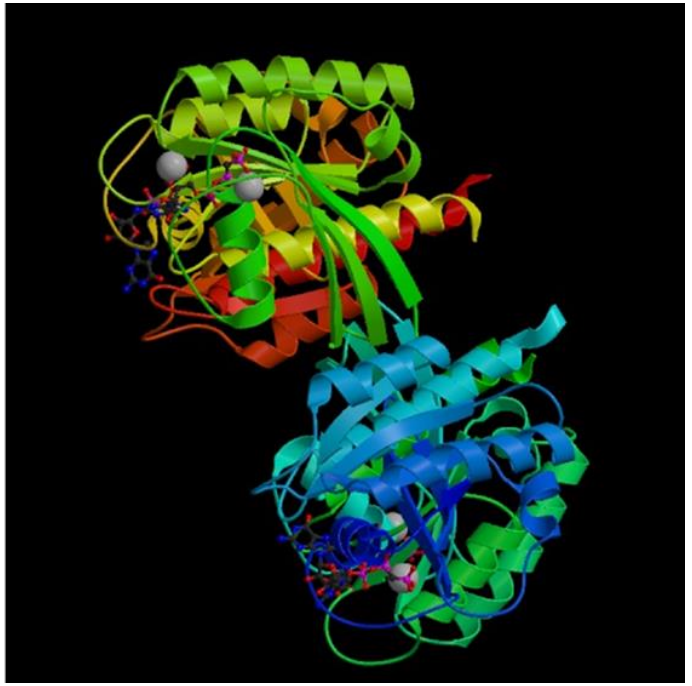
XLPM Map Viewer

A PROTEIN-PROTEIN INTERACTION MAP VIEWER

Proteins





Protein Data Bank

- Protein data bank is a database for all known existing proteins

- PDB is a source of information for protein sequencing, modeling and mapping

- You can download data files for any protein: PDB files, FASTA Files, Mapping files

- http://www.rcsb.org/pdb/home/home.do

- These provided materials are used to develop software tools to help biologists and other users understanding proteins

XLPM Algorithm

- Cross Link protein-protein mapping is an algorithm developed by students at the Bioinformatics lab

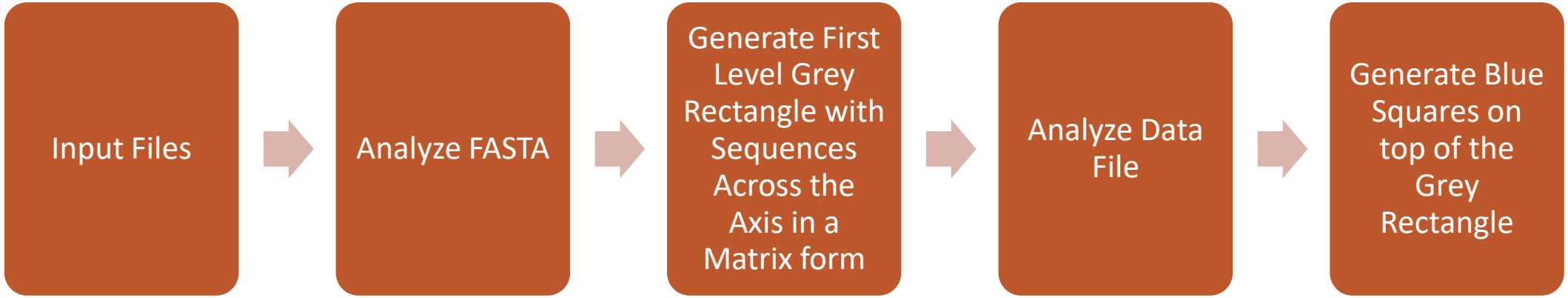
- It analysis the data coming out of what is called cross link analysis of two proteins

- Proteins → Peptides → Amino Acids → Residues → Molecules → Atoms

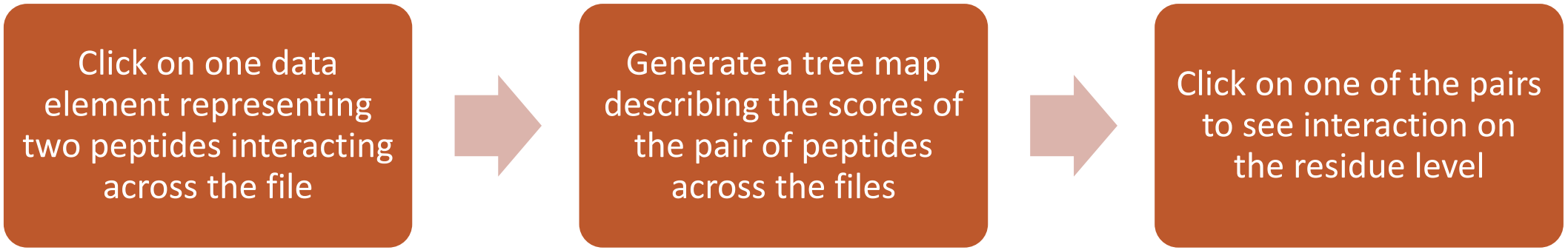
- It assigns a score two every two interacting peptides and based on this score the user can infer some conclusions

- The algorithm produces a text file with all the analysis data having a specific format

Data Flow


Input Files
Analyze FASTA
Generate First Level Grey Rectangle with Sequences Across the Axis in a Matrix form
Analyze Data File
Generate Blue Squares on top of the Grey Rectangle

Data Flow


Click on one data element representing two peptides interacting across the file
Generate a tree map describing the scores of the pair of peptides across the files
Click on one of the pairs to see interaction on the residue level

Visualization Details

- Inputs to this system are FASTA files and XLPM Data files

- The visualization is a representation of the data file on three levels including the molecular level

- The first level is a heat map for showing the overall interactions between the parts of the two protein sequences

- In the first level only the highest scored interaction is shown on the heat map for each pair of peptides

Visualization Details

- Clicking on any of the point on the heat map opens a level 2 visualization for that particular peptide pair

- Level 2 is a tree map showing all the analyzed spectra that found the interaction between the peptide pair chosen in the first level

- Selecting any of the spectra in the tree map opens a third level of visualization

- The third level of visualization is a heat map showing the cross-linking between pairs of residues

3D Models

- Understanding PDB files and FASTA files is so important to be able to parse it
- Using three.js WebGL based library to generate the molecules
- User enters either the Pdb Id or uploads a pdb file
- The system queries the PDB webservice for the PDB files
- Start parsing the PDB files and extracting coordinates, colors and sizes of Atoms

Data Types

- Ordered Data

- Flat Files

- After parsing the data I structure them in tables to be able to query easily

Sequence Table	Block Table	Precursor Info Table	Data Points Table
• Id • uniqueNumber • Sequence • Type	• Id • Fragment_1 • Fragment_2 • Score	• Block Id • Fragment_1 • Fragment_2 • Score	• Id • BlockId • Sequence_1 • Sequence_2 • Value

Applied Tufte

- Scaling is applied on the first and the third levels on color opacity
- Size Scaling is applied on the treemap third level
- Clear labeling is applied through mouse overs to avoid any ambiguity
- Data variation is clear on the three levels
- Also data ink ratio is maximized
- Avoided junk as much as possible

Demo

www.i3akef.net/xlpm

https://iaebeid@bitbucket.org/iaebeid/xlpmviewer.git

References

Mihir Jaiswal et. al. (2014) BMC Bioinfo 15(suppl 11):S16.

Michael Bostock et. al. (2011) IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)

"GLmol - Molecular Viewer on WebGL/Javascript." GLmol - Molecular Viewer on WebGL/Javascript. Web. 31 Mar. 2015. <http://webglmol.sourceforge.jp/index-en.html>.

https://github.com/typicaljoe/taffydb

https://github.com/jakiestfu/Snap.js/