Exploratory Visualization Tool for Comparison of Biotic and Abiotic Amino Acids

User Guide

CMSC 436 - Data Visualization

Professor Don Engel

Group Members: Kristina Bridgwater, Paul Heiss,

Nathan Vanbourgondien

December 10, 2020

1 Introduction

1.1 Purpose

Our client for this effort is Dr. Stephen Freeland, and Evolutionary Biologist at UMBC. Dr. Freeland gave us data sets defined by quantitative measures of amino acid properties such as size, charge, and hydrophobicity. He asked us to create something that will allow researchers to more easily compare multiple sets of user-defined amino acids in the Euclidean space.

1.2 Scope

The focus of our work is to build a stand alone framework for a prototype visualization tool without the inclusion of data analysis. The goal is for the tool to evolve into one used by research communities distributed across North America, Europe, and Japan.

1.3 Visualization Tool Design

The design of the tool is to combine three primary visualizations into a singular web application. The primary visualizations in the developed tool are a sortable and searchable table, a scatterplot matrix (SPLOM) of attributes, and a 3D rotatable model of an amino acid. Additional visualizations include a tooltip in the SPLOM and amino acid attribute details.

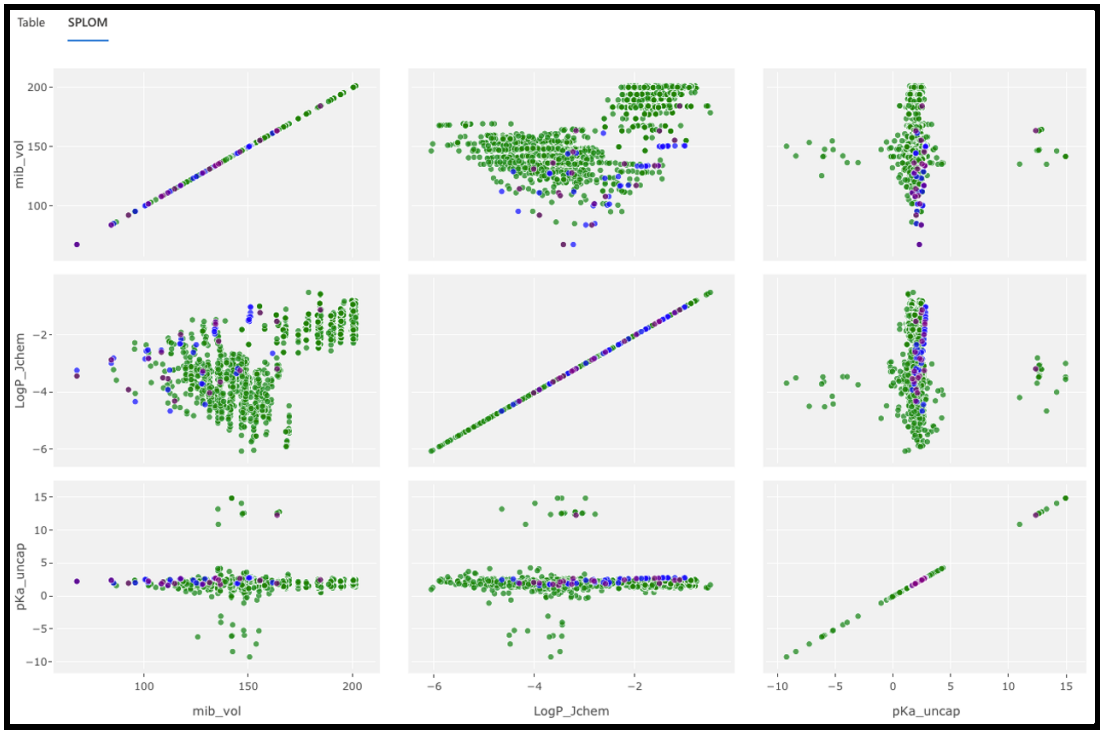
2 Visualization Tool Operations

The tool is hosted using Vercel.com and can be deployed simply by clicking the link associated: https://cmsc-436-group-project.vercel.app/
The tool consists of the following key features.

2.1 Searchable and Sortable Table

The first visualization is the table. When clicking on a column label, data shown in each column is sortable, either alphabetically or numerically, where appropriate. Data can also be filtered for specific values by entering a string in the text box below the column label. A global search for an ID or SMILES formula can be performed using the box in the top right corner.

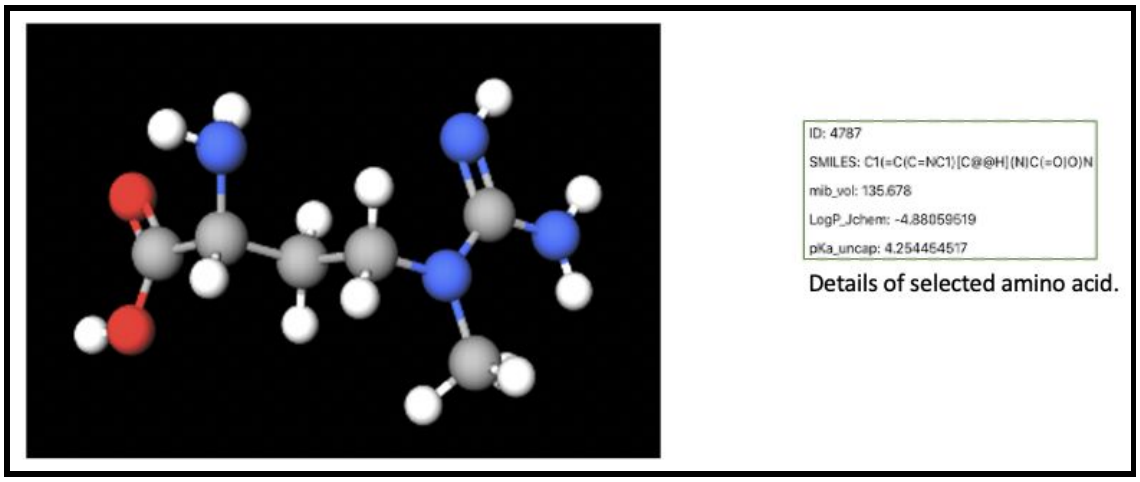
2.2 Scatterplot Matrix (SPLOM) w/ Tooltip



A scatter plot matrix (SPLOM) is a matrix of scatterplots used to visualize the relationship between a pair of variables. For the initial tool release we are plotting three variables, thus our scatter plot matrix is a 3x3 grid containing nine scatterplots. The off diagonal plots show the correlation between two variables, while the diagonal plots are simply the auto-correlations of the variable and show the range of the data. Zoom and panning features have been enabled on the SPLOM to allow the user to visualize all datasets simultaneously, to move to a specific graph (panning), or to enlarge an area of the graph (zoom) at an increased magnification.

An important feature of the scatter plot matrix is that users are provided descriptive information of the plotted data points when they mouse-over a specific data point.

2.3 3D Model of an Amino Acid w/ Details



If a user selects a data point in the SPLOM, a 3D rendering of that amino acid is created, along with displaying the details of the selected amino acid. The 3D rendering supports rotation and zoom with each of the organic compounds uniquely colored.

3 Navigation

Navigation between visualizations appears at the top left of the web page: selecting the Table link navigates to the landing page with the sortable table while selecting the SPLOM link navigates to the scatter plot visualization.