Development of an Exploratory Visualization Tool for Comparison of Biotic and Abiotic Amino Acids

December 3, 2020

Kristina Bridgwater
Paul Heiss
Nathan Vanbourgondien

Motivation

- Our client for this effort is Dr. Stephen Freeland, an Evolutionary Biologist at UMBC.
- The focus of this work is to produce a prototype visualization tool which will allow researchers to more easily compare multiple sets of user-defined amino acids in the Euclidean space.
- These data sets are defined by quantitative measures of amino acid properties like size, charge and hydrophobicity.
- The goal is that this tool can evolve into one used by research communities distributed between North America, Europe, and Japan.

Focus of this project's work is building the framework for a stand alone data viz tool, not data analysis.

Background and Related Works

- The building blocks of life are amino acids:
  - Amino acids are organic compounds that form proteins when combined.
- Research areas of amino acids include:
  - Synthetic biology
  - Meteoritics
  - Biochemistry
- Research areas need to compare amino acids individually and in sets.

The developed tool presents a set of visualizations that interactively expedites the discovery of amino acid properties.

Processed Data

- Data was supplied from the sponsor in a CSV format
  - Data included an ID, a SMILES formula, and 4 attributes.
  - There is not description of the attributes provided, as the sponsor thought that was irrelevant
    - Columns could have been labeled A, B, C and D

- The static sample data set from sponsor consisted of 2,000 rows:
  - Mib Vol
  - LogP Jchem
  - Pka Uncap
  - Biotic / abiotic indicator

- SMILES is standard notation representing a chemical structure.
  - SMILES formulas utilize the atomic symbol from the periodic table to represent each atom in a complex chemical structure.

The developed tool is attribute agnostic empowering users to load and analyze other formulas and attributes.

Technology Stack

- React
  - a web-based framework for developing user interfaces
- Node.js
  - a server based framework for rendering web apps and accessing data
- D3, Plotly and material-table
  - Visualization libraries
- Vercel.com
  - Public facing web service that can pull / recompile code from GIT repo

Tools leveraged JavaScript and enabled integration, albeit with some technical challenges and issues.

Visualization Tool Design

- Three primary visualizations combined into a single web application
  - Visualization selection influenced/guided by sponsor interaction
- Primary visualizations in the developed tool are:
  - Sortable and searchable table
  - Scatterplot Matrix (SPLOM) of attributes
  - 3D rotatable model of an amino acid
- Additional visualizations include
  - Tooltip in SPLOM
  - Amino acid attribute details

Focus of this project's work is building the framework for a stand alone data viz tool, not data analysis.

Viz #1: Sortable and searchable table

Table SPLOM

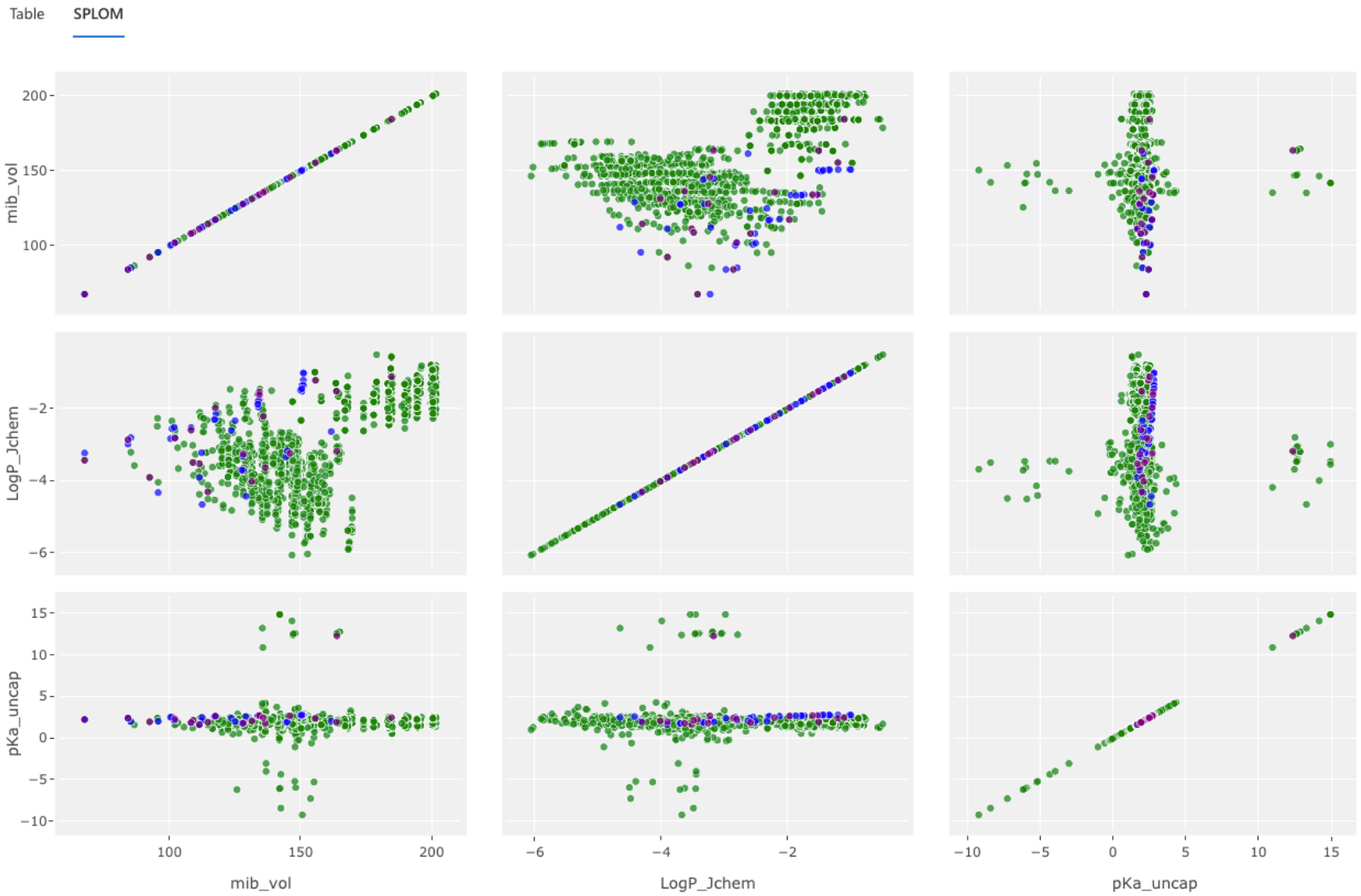
Amino Acid Data Table

ID or SMILES

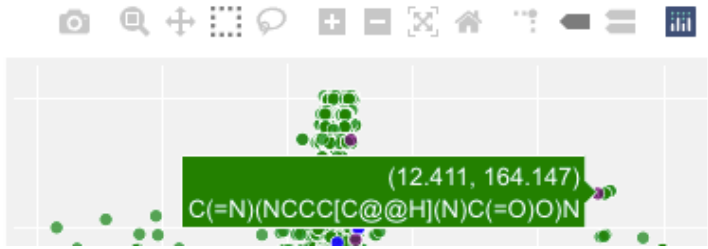
	ID	SMILES	mib_vol	LogP_Jchem	pKa_uncap	Type
>	2	C(=O)(CN)O	67.727	-3.409459573	2.307456882	Computational
>	3	C(C)[C@@H](C)[C@@H](N)C(=O)O	134.504	-1.508406606	2.791308272	Computational
>	5	C(C)(C)C[C@H](N)C(=O)O	134.504	-1.586157172	2.787708882	Computational
>	7	CC(C)[C@@H](N)C(=O)O	117.702	-1.953136464	2.717273382	Computational
>	11	C[C@@H](N)C(=O)O	84.313	-2.84078791	2.474897471	Computational
>	13	C(=N)(NCCC[C@@H](N)C(=O)O)N	164.147	-3.155937377	12.41100429	Computational
>	17	c1(cnc[nH]1)C[C@H](N)C(=O)O	136.792	-3.61602248	1.849972677	Computational
>	19	C(CN)CC[C@@H](N)C(=O)O	146.248	-3.214534721	2.738286245	Computational

8 rows 1-8 of 1977

Viz #2: Scatterplot Matrix (w/ Tooltip)

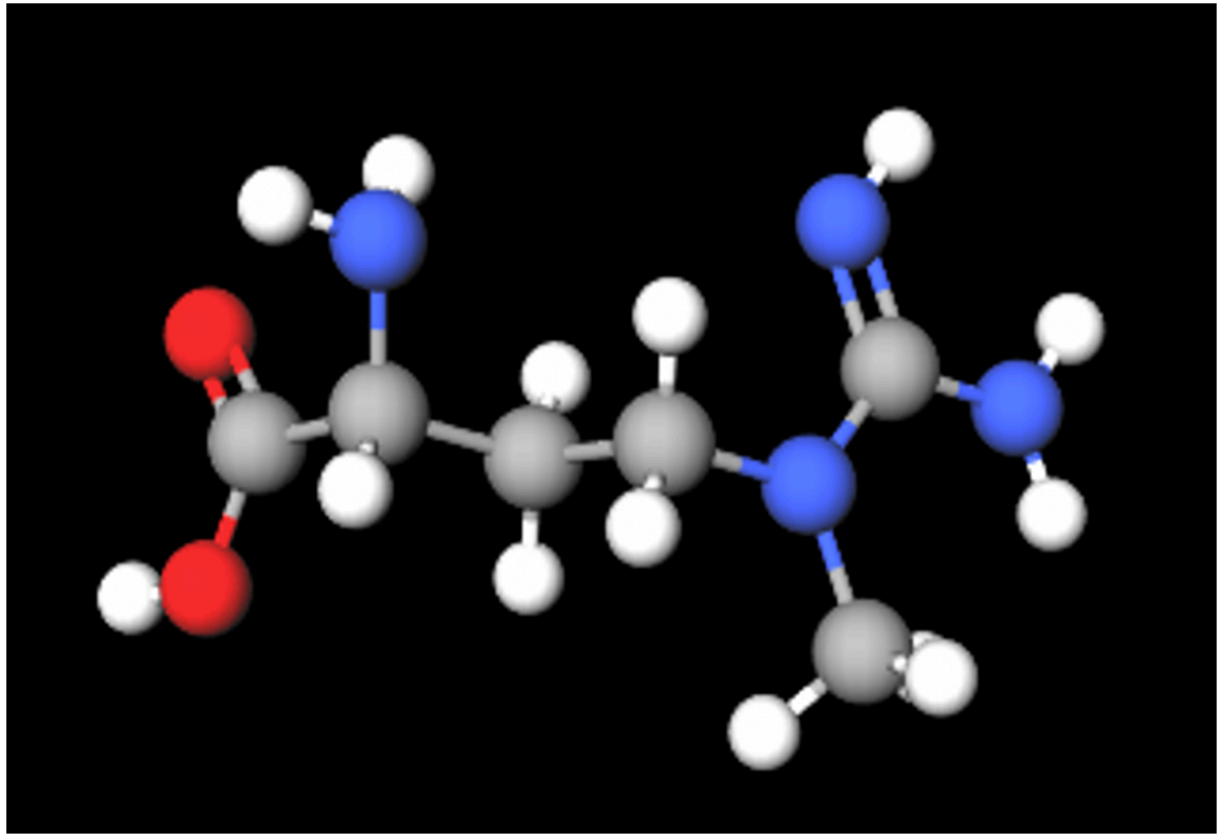


If the user clicks on a point in the SPLOM, that data point is highlighted in all the plotted matrices, and Viz #3 is loaded in the adjacent pane.



When hovering on a data point, a mouse over provides contextual values and displays the tools for manipulating the SPLOM.

Viz #3: 3D Model of Amino Acid (w/ Details)



The 3D rendering supports rotation and zooming with each of the organic compounds uniquely colored.

ID: 4787

SMILES: C1(=C(C=NC1)[C@@H](N)C(=O)O)N

mib_vol: 135.678

LogP_Jchem: -4.88059519

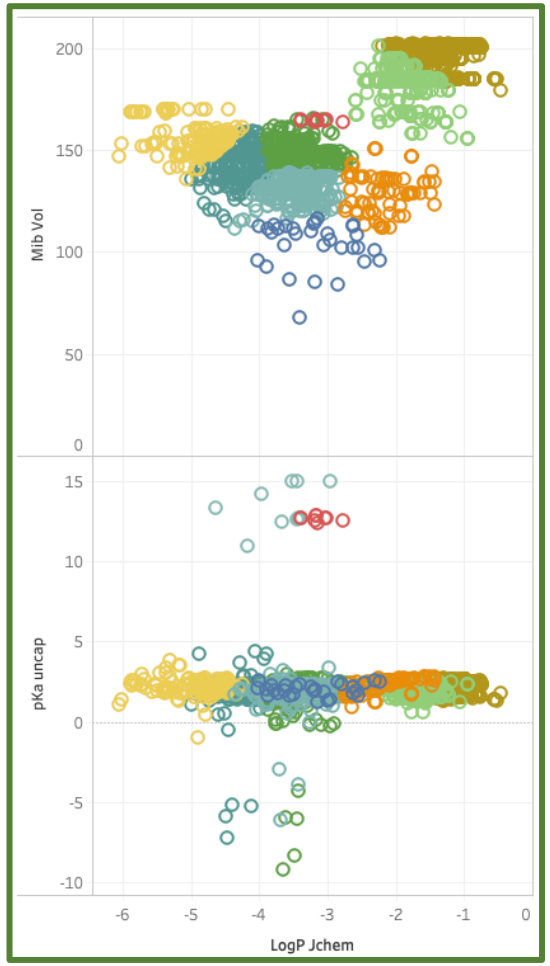
pKa_uncap: 4.254454517

Details of selected amino acid.

Demo

Longer Term Goals

- Improve dynamic data handling data loading
- Provide data visibility across users (e.g. data sharing)
- Other visualizations
  - Clustering
  - Rank ordering of the data attributes
  - Comparisons of selected amino acids
- Drive the SPLOM from the filtered data sets



Deliver the developed tool to research communities in North America, Europe, and Japan.

Q & A