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

Kristina Bridgewater
kbridgw1@umbc.edu

Paul Heiss
pheiss1@umbc.edu

Nathan Vanbourgondien
natha1@umbc.edu

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

Figure 1: Sortable and searchable data table.

Motivation

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.

Background and Related Work

Synthetic biology, meteoritics, and biochemistry are research areas that are interested in the chemical structures of biological material beyond the building blocks of life – **amino acids** – which are organic compounds that form proteins when combined.

These research areas differ dramatically in size and scope, but they all need to compare amino acids individually and in sets. To assist researchers in comparing amino acids structures and their quantifiable properties, the developed tool presents a set of visualization tools that when used interactively expedites the comparison of these data sets properties to one another.

The visualizations selected, as well as the design and flow of the developed application, were determined by the research after interviewing the application's sponsor, Dr. Stephen Freeland.

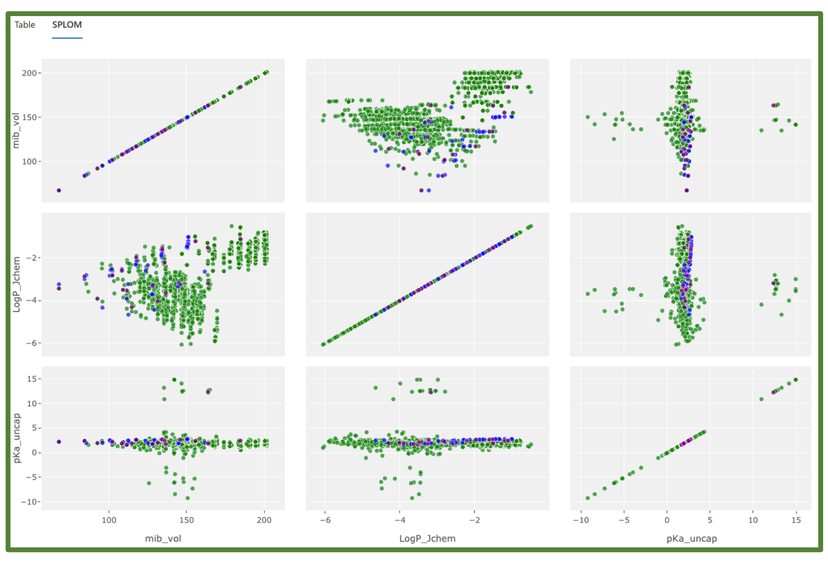

Figure 2: Scatter plot matrix of biotic and abiotic amino acids.

Visualization Tool Design

For data comparisons, three primary visualizations were combined into a stand alone web application tool. The primary visualizations in the tool are:

- Sortable and searchable table (Figure 1)
- Scatterplot Matrix of attributes (Figure 2)
- 3D rotatable model of an amino acid (Figure 4)

supported by the additional visualizations of:

- Tooltip in Scatterplot Matrix (Figure 3)
- Amino acid attribute details (Figure 5)

To integrate the visualizations, the table has a link to the 3D model in Figure 4, and the Scatterplot matrix, commonly referred to as a SPLOM, provides tools for zooming, panning, and selecting data in a box and/or set: all enabling the user in exploring the loaded data.

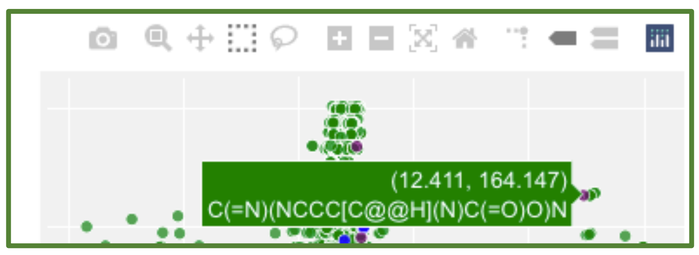

Figure 3: Tooltip from scatter plot matrix.

Processed Data

The intended use of the tool is for users to load their own data sets. The supplied data must be in a CSV format with an ID, a SMILES formula, then N number of attributes. For our static sample data, their were 2,000 rows of data. Attributes loaded were Mib Vol, LogP Jchem, Pka Uncap, and an indicator if the amino acid was abiotic. The sortable table in Figure 1 shows sample data.

SMILES is standard notation representing a chemical structure in a way that can be used by the computer. SMILES formulas utilize the atomic symbol from the periodic table to represent each atom in a complex chemical structure. SMILES formulas are shown in Figures 1, 3, and 5, and 3-dimensionally rendered in Figure 4.

The tool is attribute agnostic empowering users to load other formulas & attributes in their data sets.

Loaded data is filterable and sortable (Figure 1). If the arrow is clicked to the left of the ID, the #D model of the SMILES formula is displayed (Figure 3). Additional exploratory data analysis is accomplished with the scatter plot matrix (SPLOM: Figure 2). This visualization shows the correlation between each numeric attribute of the amino acids in the data.

A "mouse over" on the SPLOM shows the details for the corresponding amino acid (Figure 3). This tooltip is dynamic and based on the user's loaded data. The SPLOM tools for selecting and highlighting various portions of the data enrich the user experience.

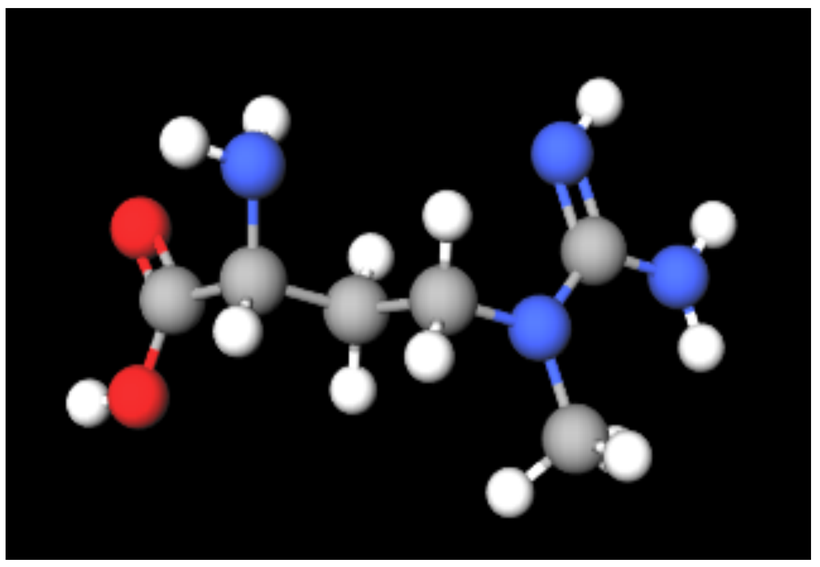

Figure 4: 3D model of a selected amino acid.

If the user clicks on a point int the SPLOM, that data point is highlighted in all the plotted matrices. In addition, a 3D rendering of the amino acid (Figure 4) is loaded with its detailed data (Figure 5) displayed. 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

Figure 5: Record details of selected amino acid.

Longer Term Goals

Hosting the tool on a public web site would be ideal. To accomplish this, work needs to be done to better handle data loading, data visibility across users (e.g. data sharing), and possible other visualizations such as clustering and rank ordering of the data attributes.

Acknowledgements

Our client for this effort is Dr. Stephen Freeland, an Evolutionary Biologist at UMBC. His hope for the project is that the prototype visualization can evolve into a tool utilized by multiple research communities distributed between North America, Europe, and Japan. Interviews with Dr. Freeland influenced the project's design and visualizations.