dms-viz: Structure-informed visualizations for deep mutational scanning and other mutation-based datasets

William W. Hannon[1, 2] **and Jesse D. Bloom**[2, 3, 4]

[1]Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, Washington 98109, United States of America, [2]Basic Sciences and Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States of America, [3]Department of Genome Sciences, University of Washington, Seattle, Washington 98109, United States of America, [4]Howard Hughes Medical Institute, Seattle, Washington 98109, United States of America

Summary and Purpose

Many biological questions require an understanding of how mutations impact a protein's functions. Deep-mutational scanning (DMS) offers an approach to characterize the impact of a huge number of mutations in parallel (Fowler and Fields 2014). The wide application of DMS has greatly increased the number of mutation-function datasets (Fowler et al. 2023). For instance, DMS has been used to determine how mutations to viral proteins affect antibody escape (Dadonaite et al. 2023), receptor affinity (Starr et al. 2020), and essential functions such as viral genome transcription and replication (Li et al. 2023). In some cases, the effects of mutations can also be inferred from phylogenies of natural sequences (Bloom and Neher 2023) (Figure 1).

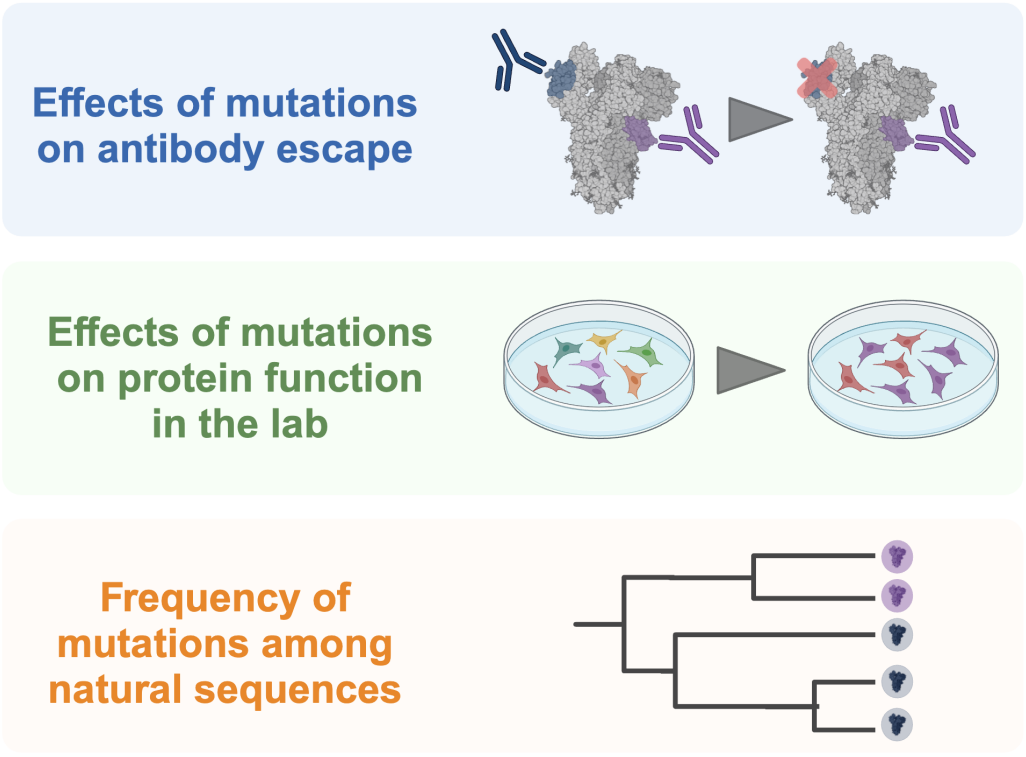

Figure 1: Large mutation-associated datasets can be used to map antibody footprints, assess the impact of mutations on protein function, and identify patterns of selection from natural mutation frequencies.

The mutation-based data generated by these approaches is better understood in the context of a protein's 3D structure. However, current approaches for visualizing mutation data in the context of a protein's structure are often cumbersome and require multiple steps and software. To streamline the visualization

of mutation-associated data in the context of a protein structure, we developed a web-based tool, `dms-viz`. With `dms-viz`, users can straightforwardly visualize mutation-based data such as those from DMS experiments in the context of a 3D protein model in an interactive format. Visit https://dms-viz.github.io/ to use `dms-viz`.

Statement of Need

We wanted `dms-viz` to provide the following functionalities:

1. **Provide structural context**: `dms-viz` simplifies the process of visualizing mutation data with structural context by superimposing mutation measurements on a 3D protein structure. Additionally, it provides extensive control over the visual representation of the 3D structure.
2. **Accommodate diverse data types**: Although analyzing DMS data is a key goal of `dms-viz`, there are many types of mutation data. The tool can handle diverse data types via a command line interface that converts data into a common format.
3. **Display multiple conditions**: With `dms-viz`, multiple experimental conditions can be visualized concurrently; for instance, researchers can easily visualize multiple antibody binding footprints from polyclonal sera (Yu et al. 2022).
4. **Maximize customization**: Every dataset has specific needs for visual representation. Recognizing this, `dms-viz` offers customization with filters (which are important for navigating large and possibly noisy datasets), and tooltips, ensuring that nuances are communicated.
5. **Create compact interactive visualizations**: `dms-viz` creates compact interactive views that can be incorporated into HTML presentation slides (e.g., https://slides.com/).
6. **Share findings with ease**: Users of `dms-viz` can generate shareable URL links to their visualizations. They can also save and share their `JSON` specification files, ensuring that data can be accessed by others.
7. **Preserve data privacy**: `dms-viz` allows users to visualize proprietary structures and analyze sensitive data in their browser without uploading their datasets to a remote server or storing them in a public repository.

Our group previously created a tool called `dms-view` (Hilton et al. 2020) that has some of the functionalities listed above. However, we designed `dms-viz` to be more customizable and comprehensive to handle a wider diversity of experimental designs and questions.

Design and Usage

Using `dms-viz` involves three components. First, using the command line tool `configure-dms-viz`, available as a Python package on PyPI (https://pypi.org/project/configure-dms-viz/), the user formats their data into a `JSON` specification file (see the documentation for details on the `JSON` schema). Then, the user uploads this specification file to `dms-viz.github.io`, a web-based interface written in Javascript, `D3.js`, and `NGL.js` (Rose et al. 2018). Finally, the specification file can either be shared directly or hosted remotely to generate a shareable URL link (Figure 2).

Upon uploading the specification file to `dms-viz`, users will see a visualization composed of four components, as illustrated in Figure 3.

1. ***Context* plot**: Located at the top of the visualization, it allows users to zoom into specific sites on the *Focus* plot while maintaining an overview of the entire dataset.
2. ***Focus* plot**: This plot shows a summarized view of the user's data. Every measured protein site is represented as a point providing a summary statistic of the effects of mutations at that site, and adjacent sites are connected with lines.
3. ***Detail* heatmap**: If the user is interested in the measurements for every mutation at a site, they can click on that site in the *Focus* plot. This will populate a heatmap with each mutation measurement at that site.

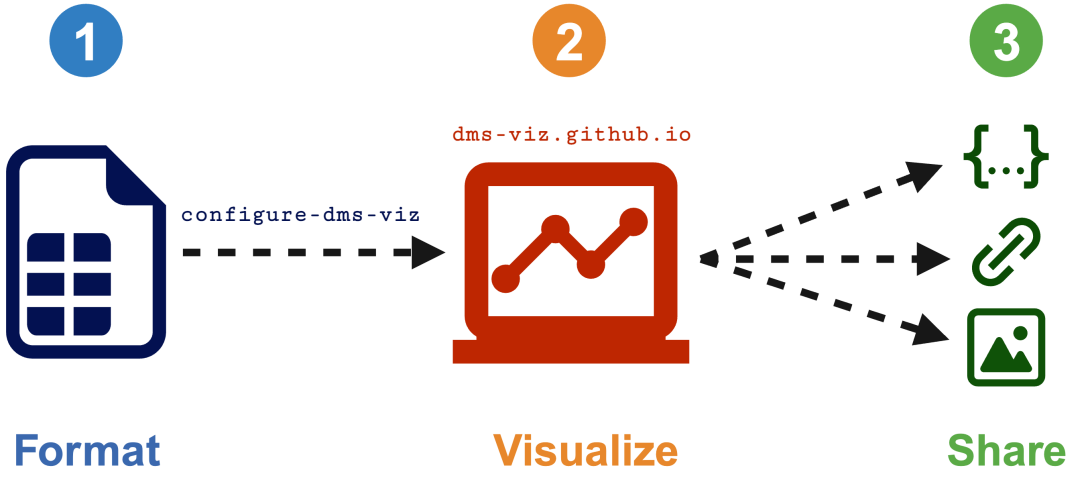

Figure 2: (1) The user formats their data using the command line tool `configure-dms-viz`. (2) The user takes the resulting `JSON` specification file and uploads it to `dms-viz.github.io`. (3) The user shares their results with a `JSON` file, a URL link, or a static image.

4. ***Interactive* structure**: When the user wants structural context for a given set of sites, they can drag a brush over the corresponding points in the *Focus* plot. This action will highlight those sites on an interactive 3D protein model.

To ensure the visualization remains compact, all configuration options are tucked away in a collapsible sidebar. See the documentation at https://dms-viz.github.io/dms-viz-docs/ for more information about how to use `dms-viz` along with detailed tutorials and examples.

Examples

For additional scientific background and a walkthrough of the code that generates these visualizations, visit the documentation.

1. Mapping the neutralization profile of antibodies and sera against HIV envelope

Radford et al. (2023) mapped mutations to HIV envelope (Env) that affect neutralization by polyclonal human serum using a pseudotyping-based deep mutational scanning platform (Radford et al. 2023). See how `dms-viz` can be used to interactively visualize datasets with multiple antibody footprints on a single summary plot here.

2. Using mutation-fitness data to augment structure-guided drug design

Bloom and Neher estimated the fitness effects of mutations to all SARS-CoV-2 proteins by analyzing millions of human SARS-CoV-2 sequences (Bloom and Neher 2023). See how `dms-viz` can be used to enhance structure-guided drug design by merging this data with structural views of a viral target like the SARS-CoV-2 main protease (Mpro) in complex with a bound ligand such as MAT-POS-e194df51-1 from the COVID Moonshot project (Boby et al. 2023) here.

3. Visualizing the pathogenicity of genetic variants of an important tumor suppressor

Matreyek et al. (2018) used variant abundance by massively parallel sequencing (VAMP-seq) to characterize the effect of thousands of mutations on the intracellular abundance of PTEN, a tumor suppressor

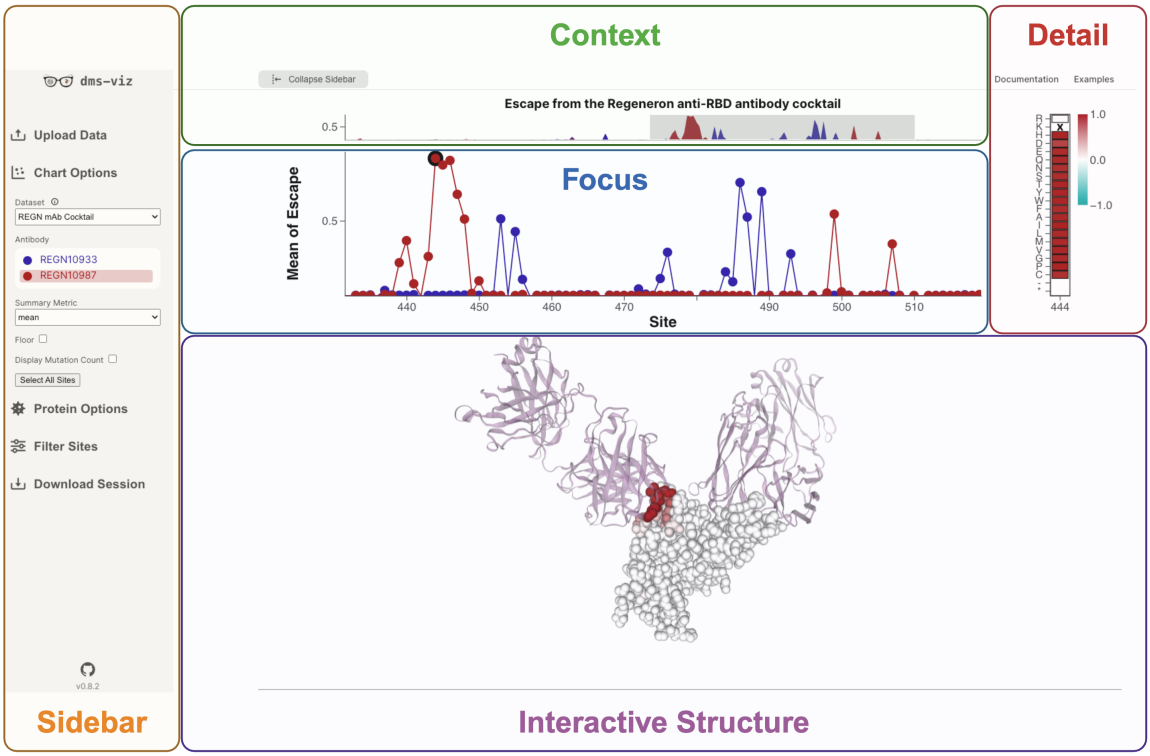

Figure 3: `dms-viz` provides a compact interface for exploring mutation-associated data, in this case, mutation-escape from the constituents of a therapeutic antibody cocktail measured by DMS of the SARS-CoV-2 receptor binding domain (RBD) (Starr et al. 2021). In this example, the structure shown is the SARS-CoV-2 RBD bound to both antibodies in the therapeutic cocktail (PDB: 6XDG).

that is inactivated in many cancers (Matreyek et al. 2018). See how `dms-viz` can be used to identify clinically relevant mutations in human proteins here.

Conclusion

`dms-viz` is a valuable addition to the suite of computational tools available for analyzing, sharing, and visualizing mutation-based datasets, which includes MaveDB, ProtVar, and many others (Esposito et al. 2019). We designed `dms-viz` as a practical and user-friendly approach to visualizing mutation-associated data in the context of protein structures. Because `dms-viz` is capable of handling various data types and has options for both sharing and privacy, it should apply to the visualization of a wide range of datasets.

Code Availability

- The visualization is available at https://dms-viz.github.io/
- The documentation is available at https://dms-viz.github.io/dms-viz-docs/
- The source code for `dms-viz.github.io` is available at https://github.com/dms-viz/dms-viz.github.io
- The source code for `configure-dms-viz` is available at https://github.com/dms-viz/configure_dms_viz

Acknowledgements

This project was envisioned as the successor to `dms-view`. Thank you to Dr. Sarah Hilton and Dr. John Huddleston for laying this groundwork and for providing helpful input. Thank you to the Bloom lab for providing data and guidance. The research reported here was supported in part by NIAID of the National Institutes of Health under award number U19AI171399. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work of JDB was supported in part by the NIH/NIAID under grants R01AI141707 and contract 75N93021C00015. JDB is an Investigator of the Howard Hughes Medical Institute.

Disclosures

JDB is on the scientific advisory boards of Apriori Bio, Aerium Therapeutics, Invivyd, and the Vaccine Company. JDB receives royalty payments as an inventor on Fred Hutch licensed patents related to deep mutational scanning of viral proteins.

References

Bloom, Jesse D., and Richard A. Neher. 2023. "Fitness Effects of Mutations to SARS-CoV-2 Proteins." *Virus Evolution* 9 (2): vead055. https://doi.org/10.1093/ve/vead055.

Boby, Melissa L., Daren Fearon, Matteo Ferla, Mihajlo Filep, Lizbé Koekemoer, Matthew C. Robinson, The COVID Moonshot Consortium, et al. 2023. "Open Science Discovery of Potent Non-Covalent SARS-CoV-2 Main Protease Inhibitors." bioRxiv. https://doi.org/10.1101/2020.10.29.339317.

Dadonaite, Bernadeta, Katharine H. D. Crawford, Caelan E. Radford, Ariana G. Farrell, Timothy C. Yu, William W. Hannon, Panpan Zhou, et al. 2023. "A Pseudovirus System Enables Deep Mutational Scanning of the Full SARS-CoV-2 Spike." *Cell* 186 (6): 1263–1278.e20. https://doi.org/10.1016/j.cell.2023.02.001.

Esposito, Daniel, Jochen Weile, Jay Shendure, Lea M. Starita, Anthony T. Papenfuss, Frederick P. Roth, Douglas M. Fowler, and Alan F. Rubin. 2019. "MaveDB: An Open-Source Platform to Distribute and Interpret Data from Multiplexed Assays of Variant Effect." *Genome Biology* 20 (1): 223. https://doi.org/10.1186/s13059-019-1845-6.

Fowler, Douglas M., David J. Adams, Anna L. Gloyn, William C. Hahn, Debora S. Marks, Lara A. Muffley, James T. Neal, et al. 2023. "An Atlas of Variant Effects to Understand the Genome at Nucleotide Resolution." *Genome Biology* 24 (1): 147. https://doi.org/10.1186/s13059-023-02986-x.

Fowler, Douglas M., and Stanley Fields. 2014. "Deep Mutational Scanning: A New Style of Protein Science." *Nature Methods* 11 (8): 801–7. https://doi.org/10.1038/nmeth.3027.

Hilton, Sarah K., John Huddleston, Allison Black, Khrystyna North, Adam S. Dingens, Trevor Bedford, and Jesse D. Bloom. 2020. "Dms-View: Interactive Visualization Tool for Deep Mutational Scanning Data." *Journal of Open Source Software* 5 (52): 2353. https://doi.org/10.21105/joss.02353.

Li, Yuan, Sarah Arcos, Kimberly R. Sabsay, Aartjan J. W. te Velthuis, and Adam S. Lauring. 2023. "Deep Mutational Scanning Reveals the Functional Constraints and Evolutionary Potential of the Influenza A Virus PB1 Protein." bioRxiv. https://doi.org/10.1101/2023.08.27.554986.

Matreyek, Kenneth A., Lea M. Starita, Jason J. Stephany, Beth Martin, Melissa A. Chiasson, Vanessa E. Gray, Martin Kircher, et al. 2018. "Multiplex Assessment of Protein Variant Abundance by Massively Parallel Sequencing." *Nature Genetics* 50 (6): 874–82. https://doi.org/10.1038/s41588-018-0122-z.

Radford, Caelan E., Philipp Schommers, Lutz Gieselmann, Katharine H. D. Crawford, Bernadeta Dadonaite, Timothy C. Yu, Adam S. Dingens, Julie Overbaugh, Florian Klein, and Jesse D. Bloom. 2023. "Mapping the Neutralizing Specificity of Human Anti-HIV Serum by Deep Mutational Scanning." *Cell Host & Microbe* 31 (7): 1200–1215.e9. https://doi.org/10.1016/j.chom.2023.05.025.

Rose, Alexander S., Anthony R. Bradley, Yana Valasatava, Jose M. Duarte, Andreas Prlic, and Peter W. Rose. 2018. "NGL Viewer: Web-based Molecular Graphics for Large Complexes." *Bioinformatics (Oxford, England)* 34 (21): 3755–58. https://doi.org/10.1093/bioinformatics/bty419.

Starr, Tyler N., Allison J. Greaney, Amin Addetia, William W. Hannon, Manish C. Choudhary, Adam S. Dingens, Jonathan Z. Li, and Jesse D. Bloom. 2021. "Prospective Mapping of Viral Mutations That Escape Antibodies Used to Treat COVID-19." *Science (New York, N.Y.)* 371 (6531): 850–54. https://doi.org/10.1126/science.abf9302.

Starr, Tyler N., Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H. D. Crawford, Adam S. Dingens, Mary Jane Navarro, et al. 2020. "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding." *Cell* 182 (5): 1295–1310.e20. https://doi.org/10.1016/j.cell.2020.08.012.

Yu, Timothy C., Zorian T. Thornton, William W. Hannon, William S. DeWitt, Caelan E. Radford, Frederick A. Matsen, and Jesse D. Bloom. 2022. "A Biophysical Model of Viral Escape from Polyclonal Antibodies." *Virus Evolution* 8 (2): veac110. https://doi.org/10.1093/ve/veac110.