PROTEIN BEASTS

VISUALIZING PROTEIN RESIDUE CHEMICAL MODIFICATIONS

Laura Garrison, Hauke Bartsch, & Stefan Bruckner
University of Bergen, Bergen, Norway
Mohn Medical Imaging and Visualization Centre, Bergen, Norway

MOTIVATION

Proteins are essential biological structures in living organisms that are comprised of residues, i.e., linked amino acids. During protein synthesis, chemical modifications, such as post-translational modifications, on these residues can impact the protein's structure and function. Such modifications can have positive, negative, or neutral effects on the protein that more broadly can impact the overall fitness of the organism. Understanding the chemical basis of these modifications, as well as **where** and the **frequency** in which they occur, has the potential to help researchers identify and treat rare diseases. However, the data capturing this information is complex. Visualizing these data clearly, with minimal clutter, is critical to helping experts identify sites with high numbers of multple types of modifications.

REDESIGN CHALLENGE

Information about the occurrence and abundance of different chemical modifications in a protein sequence is difficult to extract with the **existing visualization (Fig. 1)**, which is cluttered and uses non-optimal visual encodings for the available screen space and intended user tasks.

We propose a **redesign** with the following design requirements:

1. Declutter the existing visualization.
2. Enable identification and comparison of multiple modifications at a single site.
3. Improve accessibility for colorblind users.
4. Enable comparison between mouse and human models of a protein.

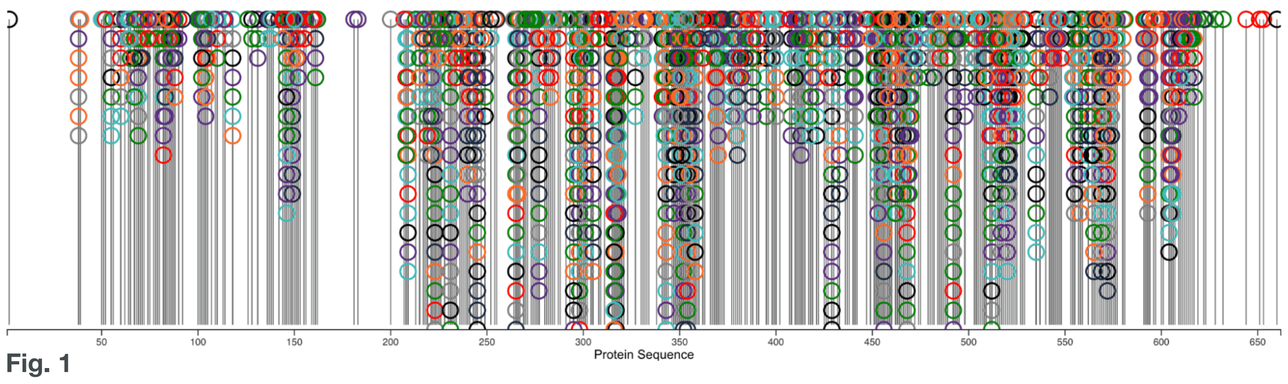
Our redesigned visualization (Fig. 2), consists of three linked views to facilitate identification and comparison of mutated protein site characteristics at different levels of granularity. An interactive prototype is available through the QR code below, which we created in Tableau[1].

Although our current design excludes the three-dimensional structure of proteins, this could be incorporated with a linked structural view to explore connections between, e.g., sites with higher diversity, that appear unconnected in our two-dimensional view.

(1) https://www.tableau.com/

Visualize protein site modification at three linked levels of granularity.

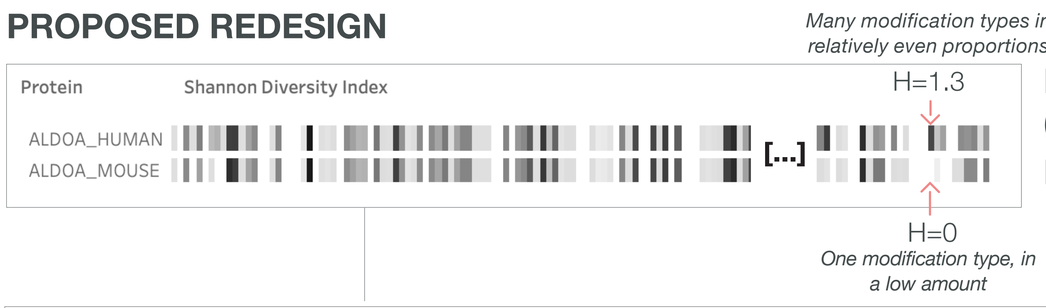
ORIGINAL VISUALIZATION



The original visualization is cluttered with some non-optimal visual encoding choices.

Fig. 1

PROPOSED REDESIGN



PER-SITE MODIFICATION DIVERSITY (richness and evenness)

$H = -\sum p_i \ln(p_i)$

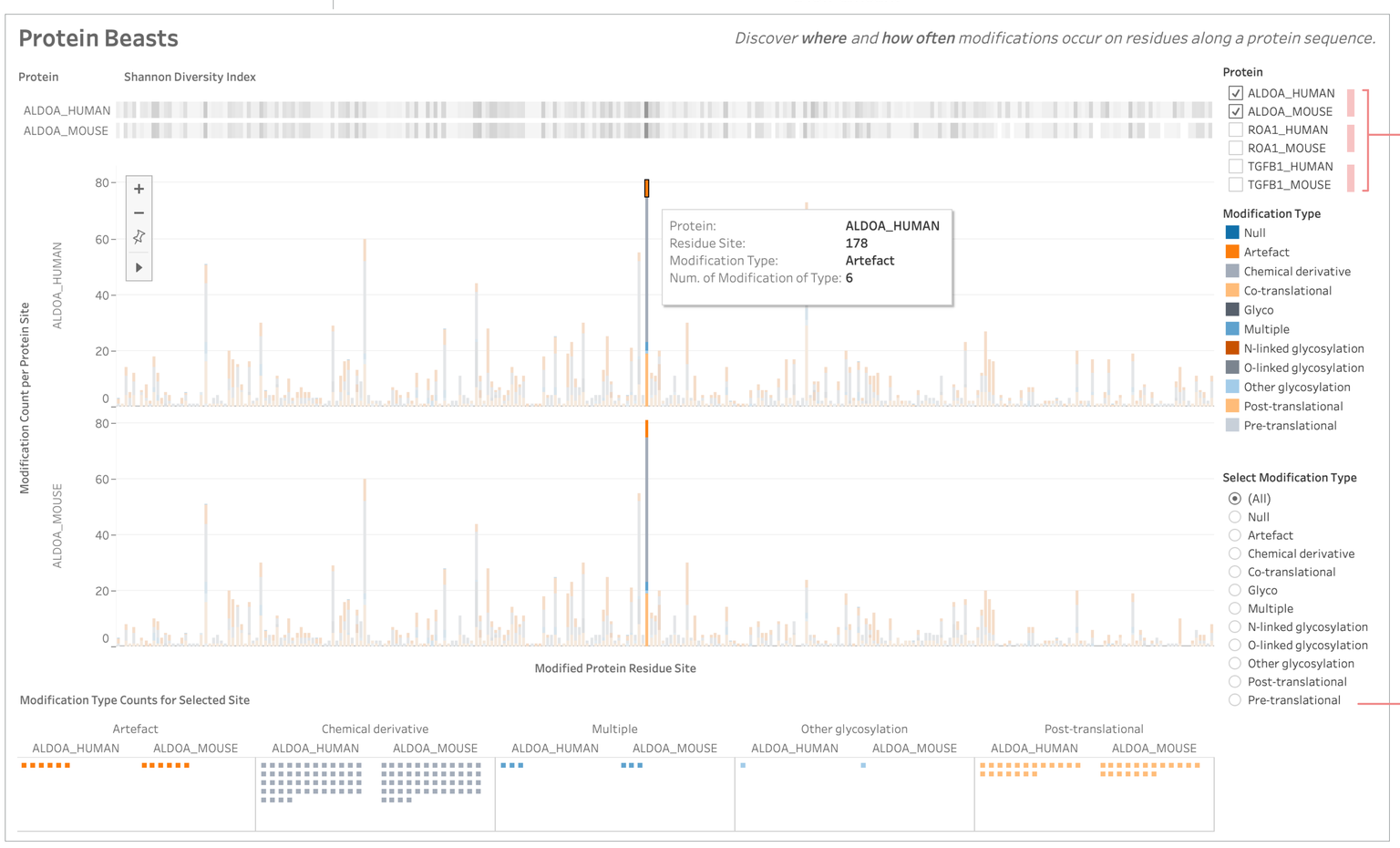


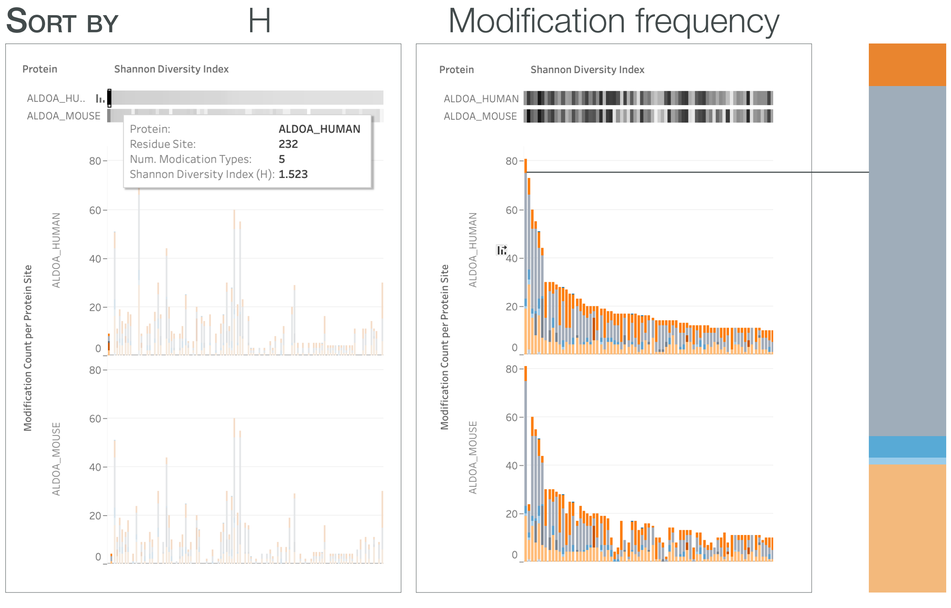
Fig. 2

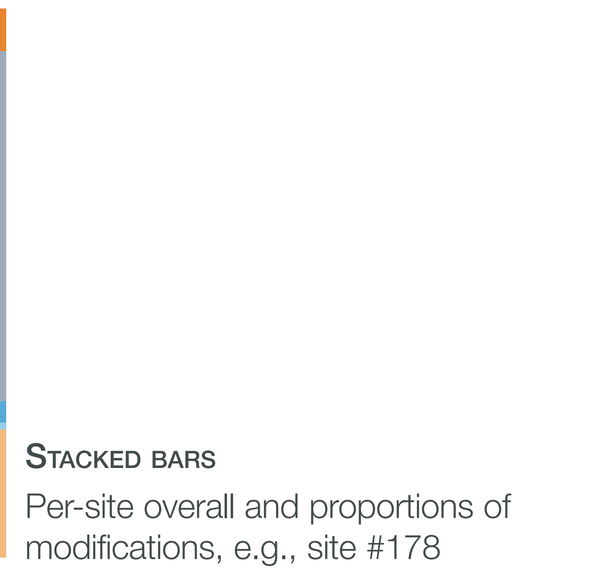
Site-focused **unit visualizations** to compare modification types within and between species

SUBSET

- Modification type (all or single type)
- Protein and species (pairs or individual)

SORT BY H Modification frequency





STACKED BARS

Per-site overall and proportions of modifications, e.g., site #178

ACKNOWLEDGEMENTS

This research is supported by the University of Bergen and the Trond Mohn Foundation in Bergen (#813558, Visualizing Data Science for Large Scale Hypothesis Management in Imaging Biomarker Discovery (VIDI). Parts of this work have been carried out in the context of the Mohn Medical Imaging and Visualization Centre (MMIV) and the Center for Data Science (CEDAS) in Bergen, Norway.

REFERENCES

[1] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. Comput Graph Forum, pp.39–63, 2013.
[2] L. A. Garrison, I. Kolesar, I. Viola, H. Hauser, and S. Bruckner. Trends & opportunities in visualization for physiology: A multiscale overview. Comput Graph Forum, 41(3):609–643, 2022. doi: 10.1111/cgf.14575
[3] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In Proc ACM Human Factors in Computing Systems, pp. 1191–1200, 2015. doi: 10.1145/2702123.2702275
[4] L. Howorko, J. M. Boedianto, B. Daniel, et al. The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. Visual Informatics, 2(3):155–165, 2018. doi: 10.1016/j.visinf.2018.09.002
[5] T. Munzner. Visualization analysis and design. CRC press, 2014.
[6] S. Nusrat, T. Harbig, and N. Gehlenborg. Tasks, techniques, and tools for genomic data visualization. In Comput Graph Forum, vol. 38, pp. 781–805, 2019. doi: 10.1111/cgf.13727
[7] D. Park, S. Drucker, R. Fernandez, and N. Elmqvist. Atom: A grammar for unit visualizations. IEEE Trans Vis Comput Graph, 24(12):3032 – 3043, 2017. doi: 10.1109/TVCG.2017.2785807
[8] B. Saket, A. Endert, and C . Demiralp. Task-based effectiveness of basic visualizations. IEEE Trans Vis Comput Graph, 25(7):2505–2512, 2018. doi: 10.1109/TVCG.2018.2829750
[9] K. Schatz, J. J. Franco-Moreno, M. Sch afer, A. S. Rose, V. Ferrario, J. Pleiss, P.-P. V azquez, T. Ertl, and M. Krone. Visual analysis of large-scale protein-ligand interaction data. In Comput Graph Forum, vol. 40, pp. 394–408, 2021. doi: 10.1111/cgf.14386