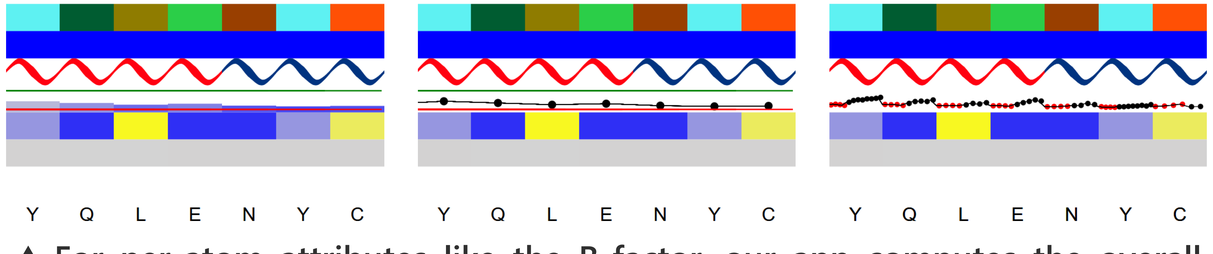
Marco Schäfer, Arne Cremer, Merve Aktürk, Michael Krone

Big Data Visual Analytics in Life Sciences, University of Tübingen

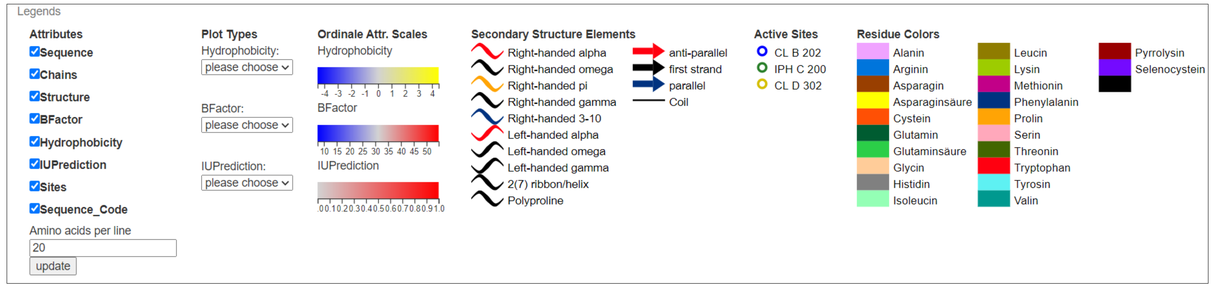
Towards an Enhanced Interactive Protein Sequence Diagram

Motivation: Sequence diagrams (SD) are a common way to visualize the amino acid sequence of proteins [1]. Additional rows in the SD can show attributes per amino acid like B-factor, hydrophobicity, binding sites, or secondary structure (typically by color-coding). SDs are often rendered only as static images.

Contribution: We present an interactive web-based SD that was implemented using the JavaScript library D3 [2]. Our enhanced SD visualizes not only the attributes stored in the RCSB Protein Data Bank [3], but also information provided by external analysis tools, such as the predicted intrinsic disorder by IUPred [4]. Furthermore, we enhanced the basic idea of SDs by adding per-atom information instead of showing only per-amino-acid attributes.



▲ **For per-atom attributes like the B-factor, our app computes the overall minimum and maximum values, and the average value per amino acid. Different visualizations (bar charts or line plots) are offered that show either the summary statistics or the individual values per atom. In the image to the right, backbone and sidechain atoms are easily distinguished.**



Outlook: Our prototypical visualization application is still work in progress, therefore, there are plenty of options for future work:

- Integrate more attributes to support a wider range of analysis tasks
- Explore additional visualizations to show more detailed information
- Display which residues are bonded or in contact to residues of another chain (e.g., hydrogen bonds or disulfide bridges)
- Link our interactive SD to a 3D molecular structure viewer such as *Mol** [5]
- Comparative visualization of proteins using a multi-sequence alignment [6]
- Improve the user interface of our current prototype
- Evaluate our proposed per-atom visualization concepts with domain experts

Acknowledgments: This work was funded by German Research Foundation (DFG) Project ID 391088465 and by Carl-Zeiss-Stiftung.

References:

[1] Kozlíková, Krone, Falk, Lindow et al., "Visualization of Biomolecular Structures: State of the Art Revisited," *Comput. Graph. Forum*, 36(8), 2017.
[2] Bostock, Ogievetsky, J. Heer, "D3 Data-Driven Documents," *IEEE TVCG*, 17(12), 2011.
[3] Berman, Westbrook, Feng, Gilliland, Bhat, Weissig, Shindyalov, Bourne, "The Protein Data Bank," *Nucl. Acids Res.*, 28(1), 2000.
[4] Dosztányi, Csizmók, Tompa, Simon, "The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins," *J. Mol. Biol.*, 347(4), 2005.
[5] Sehnal, Rose, Koca, Burley, Velankar, "Mol*: Towards a Common Library and Tools for Web Molecular Graphics," in *EG MolVA*, 2018.
[6] Schulz, Schatz, Krone, Braun, Ertl, Weiskopf, "Uncertainty Visualization for Secondary Structures of Proteins," in *IEEE PacificVis*, 2018.

VCBM 2020
VISUAL COMPUTING FOR BIOLOGY AND MEDICINE