PTMVision: An Interactive Visualization Platform for Post-Translational Modifications of Proteins

S. Hackl[1] ID, C. Jachmann[1] ID, T. Harbig[1] ID, M. Witte Paz[1] ID and K. Nieselt[1] ID

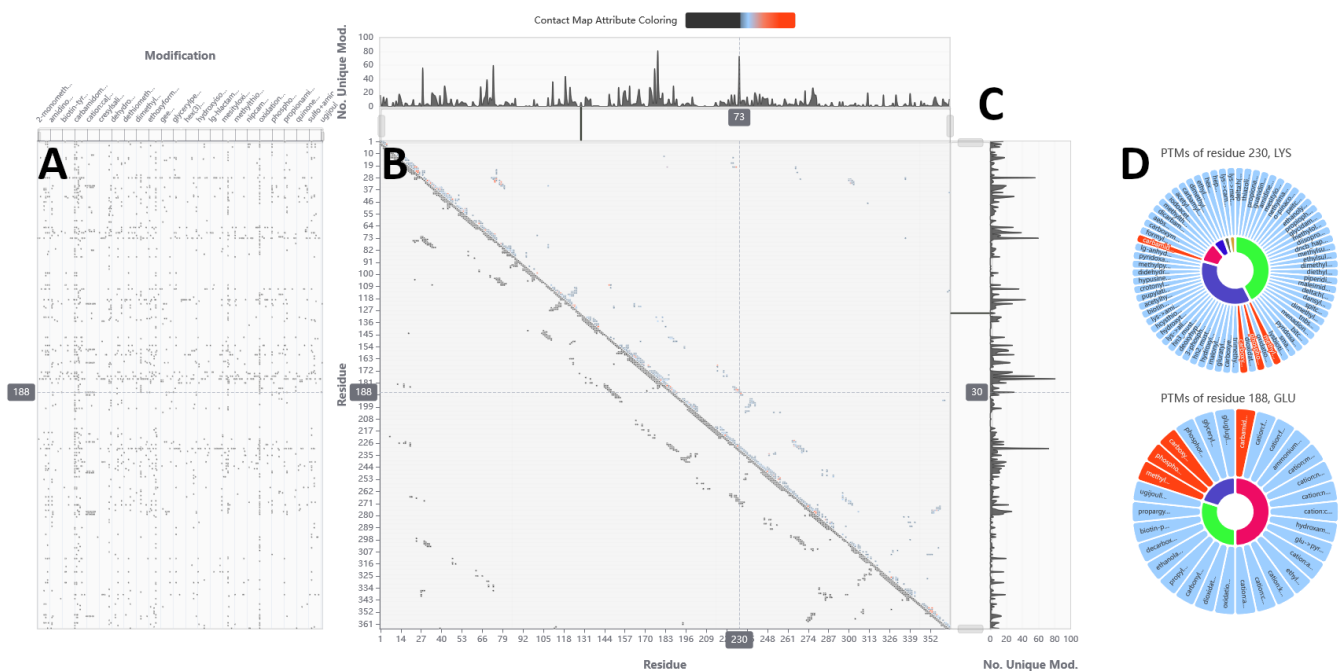[1] University of Tübingen, Institute for Bioinformatics and Medical Informatics, Tübingen, Germany

Figure 1: *PTM data of the human protein ALDOA visualized using* `PTMVision`. *(A) A presence-absence heatmap shows the PTM composition of the protein along the primary sequence (rows of the matrix, in the columns the PTMs are shown). (B) A contact map shows the residues of close proximity (here, at most 7 Å) in the lower left triangle and the residues that are in contact and additionally share PTMs in the upper right triangle. The similarity (using the Jaccard index) of the residues concerning their PTMs is indicated using a colorscale from blue to red (above the contact map). (C) On both axes of the contact map, additional user-defined annotations can been shown along the primary sequence, and a line plot depicts the total number of PTMs per site. (D) Sunburst charts show the composition of PTMs of the two residues currently selected in the contact map for comparison: The inner ring displays Unimod classes and the outer ring indicates shared (red) and unique (blue) modifications.*

Abstract

In recent years, proteins have been shown to carry many more post-translational modifications (PTMs) than originally thought. The visualization of proteins along with their PTMs facilitates exploration and understanding of the effects of these PTMs on the protein structure, function, and interactions with other proteins. Therefore, we developed `PTMVision`, *an interactive web-based visualization. We combine information about PTMs in the primary sequence with a two-dimensional representation of the protein's tertiary structure using a presence-absence map and a modified contact map that relates PTMs with the spatial arrangement of proteins without the need of a 3D structure. The prototype of* `PTMVision` *is part of the TueVis Visualization Server and is available at* *https://ptmvision-tuevis.cs.uni-tuebingen.de/.*

1. Introduction

Post-translational modifications (PTMs) are chemical modifications that occur after the translation of a protein. PTMs are known to play a key role in a variety of biological processes such as signal transduction [GK99], protein degradation [Hoc96] and transcriptional activation [DMRA16]. Recent advances in machine learning have enabled the untargeted identification of PTMs across large protein datasets and have shown that proteins are modified more frequently than previously expected [BGH*21,DGV*21,GMD19]. Current visualization techniques for PTMs include mapping them on 3D structures [BWZ*22] or the use of lollipop plots [CNNW19], which highlight the occurrence of PTMs on the primary sequence. However, multiple modifications can occur on the same site, resulting in cluttered visualizations making detailed interpretation impossible. This issue was addressed in the Bio+MedVis redesign Challenge @IEEE 2022. Our submission, *PTM Vision*, employs summary visualizations and interactive zooming to avoid cluttering. We also incorporate a contact map, which is a visualization widely used by domain experts, to enable users to examine PTM interactions with neighboring residues and modifications.

2. Tasks Identification

Based on the tasks defined by Nusrat *et al.* [NHG19] and the topology presented by Brehmer *et al.* [BM13], we identified the following tasks as relevant in the context of PTMs:

Identify: The visualization should enable single locus tasks, such as identifying the position of PTMs in the primary structure (Task 1) of a protein or identifying a modification at a position (Task 2). In addition, we aim to enable users to condense the information relevant to their research from a large number of PTMs, e.g. by highlighting specific structural domains or disease-associated positions along the primary protein sequence (Task 3), as well as by filtering PTMs according to specific properties such as co-occurrence (Task 4).

Summarize: Users should be able to summarize and visualize the PTM composition of a specific site in a protein (Task 5).

Compare: The tool should allow pattern comparison, such as the occurrence of two modifications in the entire protein sequence (Task 6) or the PTM composition of two contact sites (Task 7).

3. Visualization

PTMVision combines information about PTMs on the primary sequence with a two-dimensional representation of the protein's tertiary structure. This is achieved with a contact map aligned to a presence-absence heatmap via the positions of the primary sequence (see Figure 1). The two aligned matrices encode two complementary aspects of the modification landscape: The presence-absence heatmap shows an overview of the PTMs along the primary structure (Task 1 and 2) while the contact map includes the 3D information of the protein while avoiding the problems of 3D visualizations. The axes of the contact map correspond to the positions in the primary sequence and filled cells indicate that two residues are in contact, i.e., their Cα atoms are found at proximity, with a user-defined threshold (maximal distance in Angstrom, Å). Since the contact map is symmetric, it can be divided into two

views (see Figure 1 B). While the lower left triangle shows all contacts within the protein, the upper right triangle shows only contacts between residues carrying PTMs and satisfying a user-defined property (Task 4). For the current prototype, we implemented the share of common PTMs of two residues as the Jaccard index, which represents the intersection of both sets of modifications divided by their union. The contact map allows a visual assessment of the PTM-protein structure relationship.

For the visualization of the PTMs in the primary sequence of the protein, we chose a heatmap where each column corresponds to a PTM and each row corresponds to a position in the primary sequence (see Figure 1 A). The PTMs are sorted alphanumerically to allow searching for PTMs of interest. The visualization summarizes the composition of PTMs at a given position (Task 5) and the distribution of a PTM across the primary sequence (Task 1). Furthermore, these patterns can be compared to see e.g. whether two positions of the protein are susceptible to the same modifications (comparing two rows of the heatmap) or whether two PTMs have the same distribution across amino acids (comparing two columns of the heatmap) (Task 6). The Unimod [CC04] entry of the modification is available via clicking the cells, including information on additional physico-chemical properties and cross-references.

Linked zooming is implemented for all axes displaying the primary sequence. The x-axis of the presence-absence heatmap can be zoomed independently. Next to the contact map, a heatmap-like track displays user-chosen annotations of the protein primary sequence, such as secondary structure, structural domains, or sites of interest, e.g., based on associations with disease (Task 3). A track containing a line plot (Figure 1 C) displays the number of unique PTMs per primary sequence position (Task 5), which can be used to easily navigate to points of interest. Lastly, two sunburst diagrams can be displayed (Figure 1 D) by clicking on cells in the contact map. These show the PTM composition of the chosen residue pair in contact (Task 2 and 7), distinguishing shared and unique modifications as well as visually aggregating the composition by a user-supplied classification.

4. Conclusion and Future Work

We have developed a new interactive visual analytics tool for the exploration of the PTMs of a protein. For the future, we aim at extending the tool to allow for comparative PTM visualization across e.g. species or conditions. Furthermore, we plan to include further filtering and sorting options, for example, via chemical features of the modifications (Task 4).

Funding and Acknowledgments

References

[BGH*21] BOUWMEESTER R., GABRIELS R., HULSTAERT N., MARTENS L., DEGROEVE S.: Deeplc can predict retention times for peptides that carry as-yet unseen modifications. *Nature Methods 18*, 11 (2021), 1363–1369. 2

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics 19*, 12 (2013), 2376–2385. 2

[BWZ*22] BLUDAU I., WILLEMS S., ZENG W.-F., STRAUSS M. T., HANSEN F. M., TANZER M. C., KARAYEL O., SCHULMAN B. A., MANN M.: The structural context of posttranslational modifications at a proteome-wide scale. *PLOS Biology 20*, 5 (05 2022), 1–23. 2

[CC04] CREASY D. M., COTTRELL J. S.: Unimod: Protein modifications for mass spectrometry. *PROTEOMICS 4*, 6 (2004), 1534–1536. 2

[CNNW19] CRUZ E. R., NGUYEN H., NGUYEN T., WALLACE I. S.: Functional analysis tools for post-translational modification: a post-translational modification database for analysis of proteins and metabolic pathways. *The Plant Journal 99*, 5 (2019), 1003–1013. 2

[DGV*21] DEGROEVE S., GABRIELS R., VELGHE K., BOUWMEESTER R., TICHSHENKO N., MARTENS L.: ionbot: a novel, innovative and sensitive machine learning approach to lc-ms/ms peptide identification. *bioRxiv* (2021). 2

[DMRA16] DRAZIC A., MYKLEBUST L. M., REE R., ARNESEN T.: The world of protein acetylation. *Biochimica et Biophysica Acta (BBA)- Proteins and Proteomics 1864*, 10 (2016), 1372–1401. 2

[GK99] GRAVES J. D., KREBS E. G.: Protein phosphorylation and signal transduction. *Pharmacology & therapeutics 82*, 2-3 (1999), 111–121. 2

[GMD19] GABRIELS R., MARTENS L., DEGROEVE S.: Updated ms^{2}pip web server delivers fast and accurate ms^2 peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. *Nucleic acids research 47*, W1 (2019), W295–W299. 2

[Hoc96] HOCHSTRASSER M.: Ubiquitin-dependent protein degradation. *Annual review of genetics 30*, 1 (1996), 405–439. 2

[NHG19] NUSRAT S., HARBIG T., GEHLENBORG N.: Tasks, techniques, and tools for genomic data visualization. In *Computer Graphics Forum* (2019), vol. 38, Wiley Online Library, pp. 781–805. 2
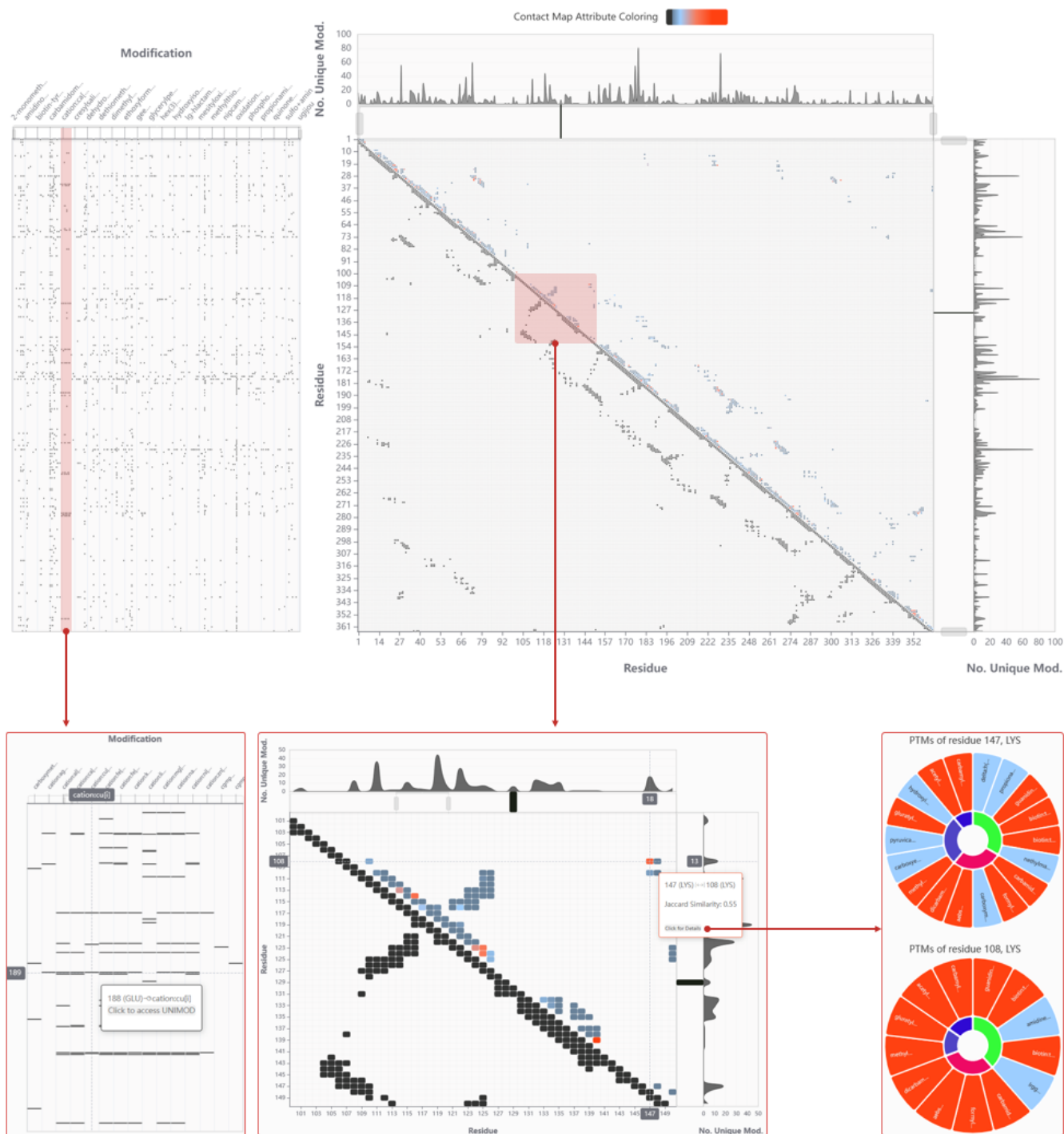
Supplementary Figure 1: *Overview of the components of* `PTMVision`*: Visualization of the PTM information of the human protein ALDOA. A distance threshold of 8Å was chosen for the contact map. The upper visualization shows the initial view of* `PTMVision` *for the given protein. User interactions are indicated by highlighted areas in red. The connecting arrows point to enlarged sections of the same visualization. The cells of the presence/absence map (left) and of the contact map (middle) show detailed information via tooltips when hovered. The colors of the cells found in the upper triangle of the contact map encode the number of PTMs shared among the neighboring amino acids. Moreover, users can click on a cell of the contact map to display sunburst charts (right) and get a detailed view of the shared PTMs.*