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

S. Hackl1, C. Jachmann1, T. Harbig1, M. Witte Paz1 and K. Nieselt1

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/.

© 2023 The Authors.
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.2312/evp.20231067
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 [Hos96] 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

PTM Vision 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 CA 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 alphabetically 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

Infrastructural funding for this project was provided by the following projects of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR261 [project ID 398967434] for TH, the Cluster of Excellence EXC 2124 ‘Controlling Microbes to Fight Infections’ [project ID 390838134] for MWP and KN, and the Cluster of Excellence EXC 2064 ‘Machine Learning: New Perspectives for Science’ [project ID 390727645] for CJ. We thank the CompOmics group at Ghent University, Belgium, for their close cooperation and valuable feedback during the development of PTM Vision.
References


Supplementary Figure 1: Overview of the components of PTMVision: Visualization of the PTM information of the human protein ALDOA. A distance threshold of 8 Å 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.