ANNOTED: Taming Protein Beasts Through Interactive Visualization

Aaron Watters and Vikram Mulligan
Flatiron Institute
awatters@flatironinstitute.org

Introduction
ANNOTED is an interactive visualization designed to address the requirements of the Bio+MedViz Challenge 2022 [1]. Briefly, it is intended to permit visualization of proteins involved in disease, and to allow users to explore sites at which amino acid residues are modified in potentially disease-relevant ways. It is published as a public web application at https://aaronwatters.github.io/bioviz2022/challenge.html with open source code available at https://github.com/AaronWatters/bioviz2022 as a Github repository.

This document describes the design and implementation of ANNOTED and some aspects of the visualization which could be improved. Some figures with screenshots from the interface follow the main text below.

Design choices
We build the interface using the jQueryUI HTML5 support library [2] augmented with a canvas based 3d vector graphics interactive plugin [3]. This framework represents a large amount of information in graphical and textual form, combined with event handling for interactivity. The result is also suitable for inclusion in any web page. Source data is compiled into Javascript compatible JSON by a preprocessor script implemented in Python.

The general shape of the protein is outlined by a **polyline scaffolding** and detail for each residue is attached to the scaffolding using standard amino acid colorization [4]. Residue representations represent the protein backbone as well as the direction of the side chain. Annotated residues are marked with circles that respond to hover and click events.

Interactions
The user selects one of the six proteins in the data set using a **dropdown list**. The protein appears in a 3d presentation with amino acid residues **colorized by type**. Residues with associated **annotations are marked with circles**. The user selects the types of annotation of interest by checking **annotation type checkboxes**. If the protein has residues marked pathogenic the user may opt to view only pathogenic residue annotations.

By dragging over the protein display region the user can **rotate** the 3d protein representation. A checkbox also enables or disables **automatic rotation**. A “fit” button centers and **fits** the protein in the view area.

When the user **hovers** over the annotation mark for a residue, information about the annotations for the residue (the known chemical modifications, as well as any known disease relevance) appear in the right panel. The user may opt to focus on a residue by **clicking** its annotation mark. The annotation information for the clicked residue also appears in the right panel.

By adjusting the **radius slider** the user suppresses detail for residues further than the selected radius from the clicked residue. This allows users to focus on sites of chemical modification near a selected residue.

Task 1 Workflow: using the interface to visualize modifications
By selecting only the annotation type of interest and then hovering or clicking on the highlighted residues we can view their characteristics and study their geometric relationships by rotating the figure. For example, Figure 1 shows the interface configured to show O-linked glycosylation annotations for protein P94312. We see that all such annotations are Serine residues on one side of the exterior of the protein.

Task 2 Workflow: Using the interface to relate a pathogenic mutation to proximal modifications
When “pathogenic only” is checked the interface only marks residues identified as pathogenic. To view nearby residues to a pathogenic residue click the pathogenic residue and then uncheck “pathogenic only” to show all annotations. Then adjust the radius slider to restrict the display detail to only nearby residues. Figure 2 shows the interface focussed on a pathogenic residue with nearby annotations highlighted.

Possible Enhancements
It would be good to be able to switch between the vector graphics representation of this interface and a more realistic volumetric representation for the protein. The current interface would also benefit from better depth perception.

We designed this interface in the absence of any realistic user story use cases. The interface would benefit from feedback from researchers or analysts attempting to solve a concrete problem by studying these protein annotations using this tool.

References
[1] http://biovis.net/2022/biovisChallenges_vis/
[2] https://jqueryui.com/
[3] https://github.com/AaronWatters/jp_doodle
[4] http://acces.ens-lyon.fr/biotic/rastop/help/colour.htm

Figure 1: Here the interface displays O-linked glycosylation annotations for protein P94312. By rotating the figure and hovering over the selected annotations we can see that all such annotations are Serine residues on one side of the exterior of the protein.

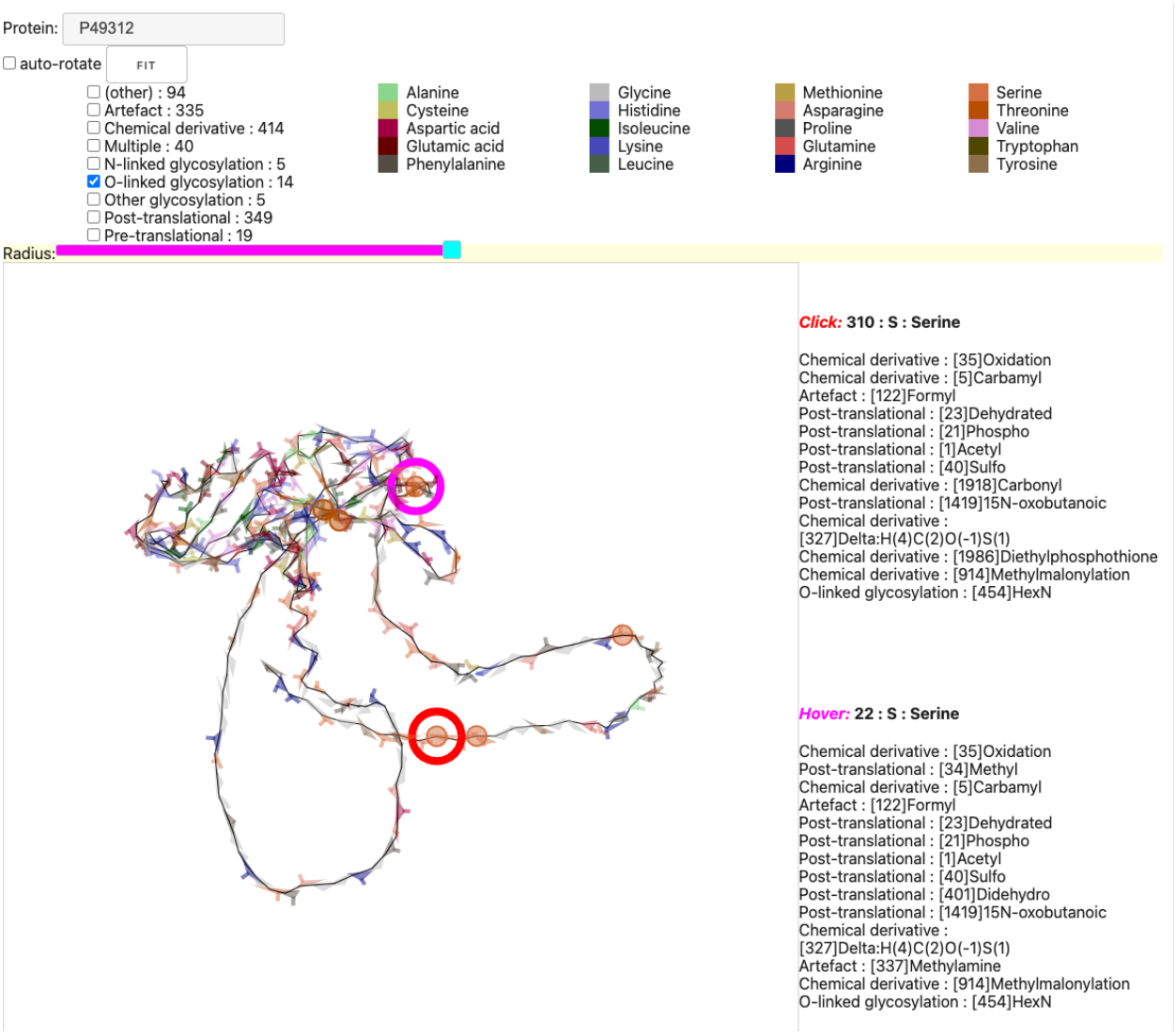

Figure 2: Below the interface displays annotations near a residue where mutations are known to be associated with disease.