HDXWizard: A Software for Customizable Hydrogen Deuterium Exchange Data Visualization

Zachary A. Cohen[1], Bindu Y. Srinivasu[1], Daniele Peterle[1], John R. Engen[1], Thomas E. Wales[1]

[1]Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA

Introduction

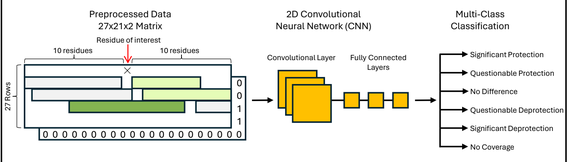
Effective visualization of deuterium incorporation data is key to hydrogen deuterium exchange mass spectrometry (HDX MS). Limitations of some software make implementation of many newer data visualization practices labor-intensive and inflexible. To combat this, we have developed HDXWizard, a python-based application for generating figures with maximal ease and customizability from one or multiple DynamX state or cluster data files. After data import and further processing, data can be visualized in various ways, including traditional deuterium incorporation plots or the more complex peptide-level difference plots. A 2D convolutional neural network is implemented to further localize differences in deuterium uptake and create localized difference plots, which can be exported to PyMOL for visualization.

Houde, D., and Engen, J.R. (2009) Chimica Oggi/Chemistry Today, **27**, 2, 12-15

Methods

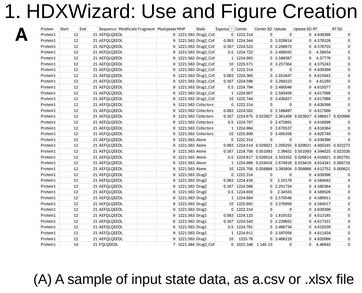
HDXWizard was written in python (3.11.3) with Tkinter (8.6), Numpy (1.24.3), MatPlotLib (3.7.2), PyMuPDF (1.23.7), Biopython (1.82), Tensorflow (2.14.0), Openpyxl (3.1.2), Pandas (2.0.1), and Xlwings (0.30.12).

The 2D convolutional neural network for creating localized difference plots was trained and validated on more than 100,000 residues of HDX difference data, each a 27x21x2 matrix containing a modified peptide difference plot with a multi-class classification output.



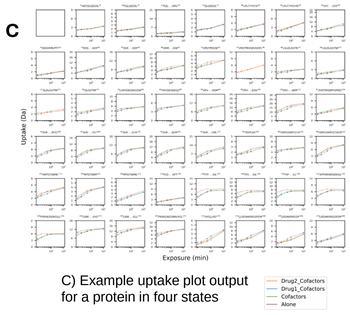
Results and Discussion

1. HDXWizard: Use and Figure Creation

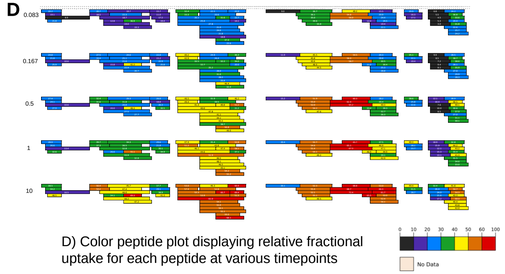


(A) A sample of input state data, as a.csv or .xlsx file

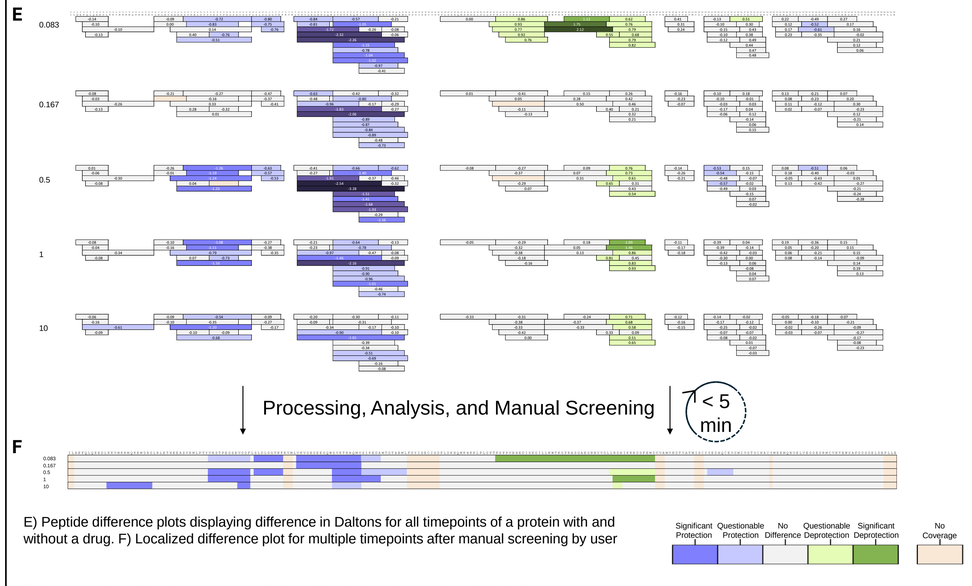
(B) The main window of the HDXWizard software, with options for inputting files and sequences, as well as tools to customize data processing and visualization



C) Example uptake plot output for a protein in four states

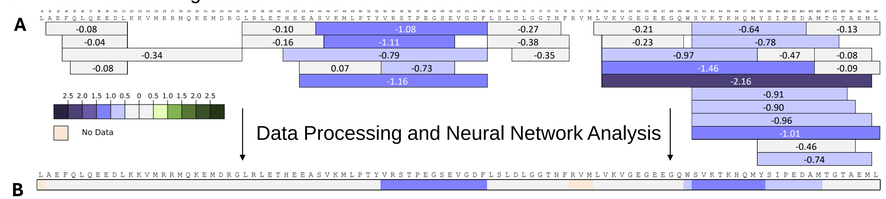


D) Color peptide plot displaying relative fractional uptake for each peptide at various timepoints



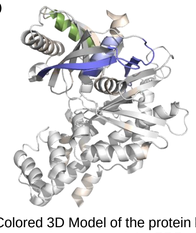
E) Peptide difference plots displaying difference in Daltons for all timepoints of a protein with and without a drug. F) Localized difference plot for multiple timepoints after manual screening by user

2. A Machine Learning Based Interface for Localized Difference Plots



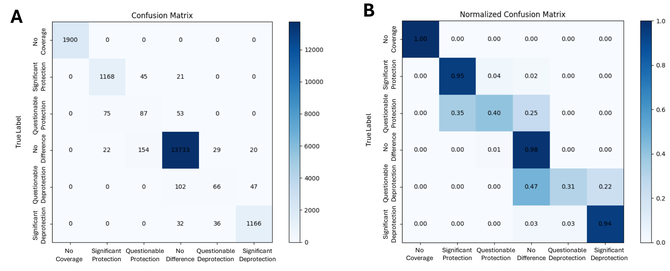
A) Peptide difference plot displaying difference in Daltons between Protein 1 with drug 1 and cofactors minus the protein with only its cofactors
B) Unscreened localized difference plot created using multi-class classification neural network from the above data

(C) User Interface for screening and editing of neural network output



D) Colored 3D Model of the protein based on screened exchange data. PDB: 4RCH

3. Assessing the Accuracy of the Machine Learning Model



A) Confusion matrix for ~18,000 residue test set showing the true label (hand classified) on the y axis and neural network predicted label on the x axis. B) Normalized confusion matrix

Conclusion and Acknowledgements

- HDXWizard provides a platform for the flexible creation of publication ready visualizations of HDX MS data
- The graphical user interface allows users to completely customize all the figures within HDXWizard
- Using a neural network, localized difference plots can be created quickly and easily
- Initial localized difference plots can be screened and exported to PyMOL for visualization on the cartoon structure without exiting the application

This material is based upon work funded by a Northeastern University Undergraduate Research and Fellowships PEAK Experiences Award