

HDXBoxeR: An R package for statistical analysis and visualization of multiple Hydrogen-Deuterium Exchange Mass-Spectrometry datasets of different protein states

Maria K. Janowska¹, Katherine Reiter^{1,2}, Pearl Magala¹, Miklos Guttman³, Rachel E. Klevit¹

¹Department of Biochemistry, University of Washington, Seattle, WA United States

²Lyterian Therapeutics, South San Francisco, CA United States

³Department of Medicinal Chemistry, University of Washington, Seattle, WA United States

Abstract.

Summary: Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) is a powerful protein characterization technique that provides insights into protein dynamics and flexibility at the peptide level. However, analyzing HDX-MS data presents a significant challenge due to the wealth of information it generates. Each experiment produces data for hundreds of peptides, often measured in triplicate across multiple time points. Comparisons between different protein states create distinct datasets containing thousands of peptides that require matching, rigorous statistical evaluation, and visualization. Our open-source R package, HDXBoxeR, is a comprehensive tool designed to facilitate statistical analysis and comparison of multiple sets among samples and time points for different protein states, along with data visualization.

Availability and implementation: HDXBoxeR is accessible as the R package (<https://cran.r-project.org/web/packages/HDXBoxeR>) and GitHub: [mkajano/HDXBoxeR](https://github.com/mkajano/HDXBoxeR).

Contact: mjano@uw.edu

Introduction

Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) has emerged as a powerful tool for probing protein structure and dynamics.¹ Traditional bottom-up HDX-MS experiments monitor the incorporation of deuterium into protein backbone amides over time, resulting in a peptide mass change that is measured via mass spectrometry. The rate of deuterium uptake provides insights into protein structure and dynamics, with more unstructured or dynamic regions exchanging deuterium faster than structured or protected ones. Narrow regions of deuterium incorporation can be identified by employing non-specific proteases into the workflow. While this greatly increases sequence coverage and redundancy, it can also generate hundreds of peptides for subsequent analysis. To ensure data reliability, these measurements are usually repeated (triplicate) and taken across multiple time points to capture the kinetics of deuterium exchange within the protein.

Most biological studies involve perturbation to the system to uncover complex and fundamental mechanisms. Thus, HDX-MS experiments often encompass different protein states, such as the impact of ligand binding or mutations, further increasing the complexity of the HDX-MS experiment and resulting data. Extracting meaningful information from such large datasets requires not only accurate matching of peptides across samples but also rigorous statistical analysis to identify peptides that exhibit significant changes in deuterium uptake between different states.

Analysis of the HDX-MS experiments is time-consuming due to the large amount of data. Furthermore, it requires expertise in statistical and data manipulation that limits approachability and accessibility for researchers whose investigations could otherwise benefit from HDX-MS. Considering the growing use of HDX-MS alongside structural tools such as Cryo-Electron Microscopy and X-ray Crystallography, developing and extending pipelines that can overcome this hurdle is critical to make HDX-MS data analysis accessible to a wider range of researchers.

We have developed HDXBoxeR, a comprehensive R package for the post-processing of HDX-MS data. The package provides a facile analysis-to-publication pipeline that incorporates data input, statistical analysis, dataset comparison, visualization, and preparation of formatted outputs necessary for publication.^{2,3}

HDXBoxeR is streamlined to work with HDExaminer (Trajan Scientific and Medical) output, a processing software that is growing in popularity within the HDX community. The HDExaminer software enables data processing, export functionalities, PyMOL script generation, and visualization tools (see Supplementary Table). Despite these important features, there remains a need for further tools to reduce the complexity of large datasets by focusing on peptides and regions that exhibit significant differences across multiple samples, thereby simplifying the analysis. Thus, we developed HDXBoxeR to streamline HDX-MS data processing and visualization, accommodating multiple protein states, time points, and replicates. Other post-processing softwares that accept modified HDExaminer output include Deuterios⁴ or HaDeX.⁵ The functionalities provided by these and other tools (such as MSTools⁶, MEMDX⁷, HDX-Viewer⁸) offer additional and complementary features to HDXBoxeR, but most do not offer customizations for flexible data visualization, curation, and data selection.^{4–20} Aforementioned tools can be accessed either via a web server or as a stand-alone programming package (see supplementary table).

HDXBoxeR is designed to enable users to compare multiple sets of HDX-MS data, offering flexible tools for quick and comprehensive data analysis. HDXBoxeR uses the Welch's T-test and the critical interval statistical framework to identify statistically significant differences in deuterium uptake between datasets recommended by IC-HDX community discussions.³ Many of the visualization tools and data exports are already filtered through the lens of significantly different peptides between the protein sets. Additional functionalities include HDX-summary tables, PyMOL script generation (for significantly different peptides between sets), and the creation of various types of plots. HDXBoxeR ensures that generated data adheres to HDX-MS data publication standards and offers an efficient workflow for HDX-MS data analysis and interpretation.² In addition, the package can accept input from published datasets, potentially allowing for re-analyses of data collected by different groups.

Material and Methods

HDXBoxeR provides a comprehensive tool for data reorganization, statistical analysis, and visualization, substantially expediting the differential HDX-MS analysis process. HDXBoxeR, an R package,^{21–28} facilitates the processing and comparison of multiple protein states, time points, and replicates. The focus of the package is its ability to compare multiple states (samples) simultaneously. By default, all protein states are compared to the first state within the data, but the order of data analysis or reference data can be changed by the user as desired. We have tested the functionalities using a dataset of a total of 8 protein states, 4 time points, and 4 replicates simultaneously, but there is no upper limit on the number of states that can be analyzed. HDXBoxeR uses a hybrid significance test approach³ (the Welch's T-test and the critical interval statistical framework) to assess if the difference in deuterium uptake

between sets is statistically significant.

For comprehensive characterization and analysis, HDXBoxeR uses input from HDExaminer (Trajan Scientific and Medical), eliminating the need for additional tools for data preparation, and further facilitating the analysis. The strength of the package lies in its ability to filter multiple protein states through the lens of significantly different peptides.

Although full use of HDXBoxeR requires familiarity with the R programming language, it is designed to allow use by those with only minimal experience with R. Plot generation and visualization do not require writing new functions. More advanced users can leverage the R programming language to extend their analysis, utilizing the prepared intermediate processing files. We highly recommend using our vignette (tutorial) for users to familiarize themselves with the package. The vignette provides a straightforward workflow, enabling users to follow commands to analyze the data. The software can be downloaded from and installed on GitHub: [mkajano/HDXBoxeR](https://github.com/mkajano/HDXBoxeR) and is available through CRAN.

Functionality

HDXBoxeR facilitates multiple aspects of the final HDX-MS data analysis. In addition to analyzing multiple protein states, the program has the functionality to:

1. Reprocess input data to the format required for data publication.
2. Calculate parameter tables such as back exchange, average peptide lengths, and statistical information for a general HDX summary table for publication.
3. Convert output from HDExaminer (Trajan Scientific and Medical) to a format that is easier to customize for downstream analysis.
4. Return peptides that are significantly different between sets, using Welch's T-tests.
5. Return scripts for PyMOL to show significantly different peptides between sets and display the HDX results onto a protein structure.
6. Enable rapid plot generation, including:
 - a. Generic plot generation such as Woods plots, volcano plots, and average deuteration, uptake plots.
 - b. Welch's and critical interval statistics filtered plots: heat maps, robot plots (modified butterfly plot), (filtered) Woods plots, and significantly different peptides maps.

HDXBoxeR uses a single input to compare whole datasets and provides statistical analysis for all provided states, time points, and technical replicates. Once the statistical analysis is executed, users can save the data as an output, visualize it, or prepare (filtered through the Welch's/critical interval statistics³) PyMOL scripts to display the data. HDXBoxeR provides a suite of plots including deuteration difference, volcano plots, heat maps, 'robot plots' (modified butterfly plots, highlighting peptide ranges and showing peptides that are significantly different between states), or woods plots (with or without statistics filters) (see Figure 1) that can be used for data visualization and publication. The suite of plots allows the user to obtain a more comprehensive understanding of the data and facilitates drawing clear conclusions.

HDXBoxeR also has the functionality to return summary information about the sets: time points, number of replicates, number of peptides, peptide coverage, average peptide length, average redundancy, standard deviation, critical interval, average back exchange, and back exchange range. The information provided complies with Masson et al., Nature Methods volume 16, pages 595–602 (2019).² HDXBoxeR reprocesses HDX-MS data and calculates statistical parameters from it. Data can be exported

automatically in different outputs depending on user needs. Users can prepare a format that is ready for publication or return a simple .csv with all the peptides matched among the sets or provide a verbose output that includes peptides' standard deviation and p-value per peptide.

The uniqueness of the HDXBoxeR package lies in its ability to compare multiple states within a set. The proposed workflow streamlines the analysis of multiple states and offers various options for easy visualization. Another advantage of the HDXBoxeR package is its strong emphasis on visualizing information involving peptides that display significantly different behavior between protein states.

Novel features in the package include:

- Heatmaps focused only on significantly different peptides between protein states.
- Peptide coverage maps color-coded for significantly different peptides between protein states.
- Two options for Woods plots: one showing all peptides and another focusing only on peptides significantly different between (multiple) protein states (also available through Deuterios⁴).
- Robot plots highlight significantly different peptides between protein states. These plots are a modification of butterfly plots. Similar to butterfly plots, robot plots are centered around one axis (here the x-axis), but the y-axis is plotted bi-directionally, allowing for a visual comparison of the data. Specifically, in robot plots, the x-axis represents residue numbers (unlike in butterfly plots, where it is the peptide index), while the positive and negative y-axes represent the percent deuteration. The novelty of the robot plot lies in the explicit drawing of peptide ranges as horizontal bars at y-values corresponding to the peptide's average percent deuteration (similar to woods plots). The standard deviation of the percent deuteration is depicted as a vertical bar. Additionally, robot plots limit the number of peptides plotted by filtering based on p-value and critical interval. Robot plots will display specific peptides only if that peptide was determined to be significantly different from the control state at least at one-time point. The time point at which a peptide was significantly different between the states is denoted by a colored dot in the middle of the peptide. Peptides that are not significantly different (in the time series) will have a grey dot at the middle point of the peptide.
- The workflow ensures consistent color schemes across all plots to denote peptides that are significantly different between states, and facilitates exporting PyMOL scripts using the same color scheme.

Conclusions

HDXBoxeR is a package that provides a complete post-processing workflow and allows for facile and fast data analysis for multiple sets as well as additional functionality to currently available tools.⁴⁻²⁰

Examples of how to use the package can be found in the R vignette that shows examples for the analysis. Our goal is to provide tools for the HDX-MS users to analyze and compare HDX-MS data in a rigorous, statistically-sound manner that will allow them to gain additional insights from their highly information-rich data.

Acknowledgments

Thanks to Natalie L. Stone and Christopher N. Woods for testing the package and providing input.

Conflict of interest

- Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat. Methods* **2019**, *16* (7), 595–602. <https://doi.org/10.1038/s41592-019-0459-y>.
- (3) Hageman, T. S.; Weis, D. D. Reliable Identification of Significant Differences in Differential Hydrogen Exchange-Mass Spectrometry Measurements Using a Hybrid Significance Testing Approach. *Anal. Chem.* **2019**, *91* (13), 8008–8016. <https://doi.org/10.1021/acs.analchem.9b01325>.
 - (4) Lau, A. M.; Claesen, J.; Hansen, K.; Politis, A. Deuteros 2.0: Peptide-Level Significance Testing of Data from Hydrogen Deuterium Exchange Mass Spectrometry. *Bioinformatics* **2021**, *37* (2), 270–272. <https://doi.org/10.1093/bioinformatics/btaa677>.
 - (5) Puchata, W.; Burdukiewicz, M.; Kistowski, M.; Dąbrowska, K. A.; Badaczewska-Dawid, A. E.; Cysewski, D.; Dadlez, M. HaDeX: An R Package and Web-Server for Analysis of Data from Hydrogen–Deuterium Exchange Mass Spectrometry Experiments. *Bioinformatics* **2020**, *36* (16), 4516–4518. <https://doi.org/10.1093/bioinformatics/btaa587>.
 - (6) Kavan, D.; Man, P. MStools—Web Based Application for Visualization and Presentation of HXMS Data. *Int. J. Mass Spectrom.* **2011**, *302* (1), 53–58. <https://doi.org/10.1016/j.ijms.2010.07.030>.
 - (7) Hourdel, V.; Volant, S.; O'Brien, D. P.; Chenal, A.; Chamot-Rooke, J.; Dillies, M.-A.; Brier, S. MEMHDX: An Interactive Tool to Expedite the Statistical Validation and Visualization of Large HDX-MS Datasets. *Bioinforma. Oxf. Engl.* **2016**, *32* (22), 3413–3419. <https://doi.org/10.1093/bioinformatics/btw420>.
 - (8) Bouyssié, D.; Lesne, J.; Locard-Paulet, M.; Albigot, R.; Burlet-Schiltz, O.; Marcoux, J. HDX-Viewer: Interactive 3D Visualization of Hydrogen–Deuterium Exchange Data. *Bioinformatics* **2019**, *35* (24), 5331–5333. <https://doi.org/10.1093/bioinformatics/btz550>.
 - (9) Smit, J. H.; Krishnamurthy, S.; Srinivasu, B. Y.; Parakra, R.; Karamanou, S.; Economou, A. Probing Universal Protein Dynamics Using Hydrogen–Deuterium Exchange Mass Spectrometry-Derived Residue-Level Gibbs Free Energy. *Anal. Chem.* **2021**, *93* (38), 12840–12847. <https://doi.org/10.1021/acs.analchem.1c02155>.
 - (10) Smit, J. H.; Krishnamurthy, S.; Srinivasu, B. Y.; Karamanou, S.; Economou, A. PyHDX: Derivation and Visualization of Protection Factors from Hydrogen-Deuterium Exchange Mass Spectrometry at near Residue Resolution. *bioRxiv* October 2, 2020, p 2020.09.30.320887. <https://doi.org/10.1101/2020.09.30.320887>.
 - (11) Cupp-Sutton, K. A.; Welborn, T.; Fang, M.; Langford, J. B.; Wang, Z.; Smith, K.; Wu, S. The Deuterium Calculator: An Open-Source Tool for Hydrogen–Deuterium Exchange Mass Spectrometry Analysis. *J. Proteome Res.* **2023**, *22* (2), 532–538. <https://doi.org/10.1021/acs.jproteome.2c00558>.
 - (12) Largy, E.; Ranz, M. OligoR: A Native HDX/MS Data Processing Application Dedicated to Oligonucleotides. *Anal. Chem.* **2023**, *95* (25), 9615–9622. <https://doi.org/10.1021/acs.analchem.3c01321>.
 - (13) Liu, S.; Liu, L.; Uzuner, U.; Zhou, X.; Gu, M.; Shi, W.; Zhang, Y.; Dai, S. Y.; Yuan, J. S. HDX-Analyzer: A Novel Package for Statistical Analysis of Protein Structure Dynamics. *BMC Bioinformatics* **2011**, *12 Suppl 1* (Suppl 1), S43. <https://doi.org/10.1186/1471-2105-12-S1-S43>.
 - (14) Zhang, N.; Yu, X.; Zhang, X.; D'Arcy, S. HD-eXplosion: Visualization of Hydrogen-Deuterium Exchange Data as Chiclet and Volcano Plots with Statistical Filtering. *Bioinforma. Oxf. Engl.* **2021**, *37* (13), 1926–1927. <https://doi.org/10.1093/bioinformatics/btaa892>.
 - (15) Lumpkin, R. J.; Komives, E. A. DECA, A Comprehensive, Automatic Post-Processing Program for HDX-MS Data. *Mol. Cell. Proteomics MCP* **2019**, *18* (12), 2516–2523. <https://doi.org/10.1074/mcp.TIR119.001731>.
 - (16) Uhrik, L.; Henek, T.; Planas-Iglesias, J.; Kucera, J.; Damborsky, J.; Marek, M.; Hernychova, L. Study of Protein Conformational Dynamics Using Hydrogen/Deuterium Exchange Mass Spectrometry. *Methods Mol. Biol. Clifton NJ* **2023**, *2652*, 293–318. https://doi.org/10.1007/978-1-0716-3147-8_18.

- (17) Quast, J.-P.; Schuster, D.; Picotti, P. Protti: An R Package for Comprehensive Data Analysis of Peptide- and Protein-Centric Bottom-up Proteomics Data. *Bioinforma. Adv.* **2022**, 2 (1), vbab041. <https://doi.org/10.1093/bioadv/vbab041>.
- (18) Crook, O. M.; Gittens, N.; Chung, C.; Deane, C. M. Inferring Residue Level Hydrogen Deuterium Exchange with ReX. *bioRxiv* April 13, 2024, p 2024.04.12.589190. <https://doi.org/10.1101/2024.04.12.589190>.
- (19) Seetaloo, N.; Kish, M.; Phillips, J. J. HDflex: Software for Flexible High Structural Resolution of Hydrogen/Deuterium-Exchange Mass Spectrometry Data. *Anal. Chem.* **2022**, 94 (11), 4557–4564. <https://doi.org/10.1021/acs.analchem.1c05339>.
- (20) Gessner, C.; Steinchen, W.; Bédard, S.; J. Skinner, J.; Woods, V. L.; Walsh, T. J.; Bange, G.; Pantazatos, D. P. Computational Method Allowing Hydrogen-Deuterium Exchange Mass Spectrometry at Single Amide Resolution. *Sci. Rep.* **2017**, 7 (1), 3789. <https://doi.org/10.1038/s41598-017-03922-3>.
- (21) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
- (22) Schrödinger, LLC. The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8, 2015.
- (23) Wickham, H. *Stringr: Simple, Consistent Wrappers for Common String Operations*; 2023.
- (24) Wickham, H.; François, R.; Henry, L.; Müller, K.; Vaughan, D. *Dplyr: A Grammar of Data Manipulation*; 2023.
- (25) Wickham, H.; Vaughan, D.; Girlich, M. *Tidyr: Tidy Messy Data*; 2023.
- (26) Tunisia, K. M. N. S. & S. D. N. S. C. of. *cancer: A Graphical User Interface for Accessing and Modeling the Cancer Genomics Data of MSKCC*; 2022.
- (27) Spiess, A.-N. *qpcR: Modelling and Analysis of Real-Time PCR Data*; 2018.
- (28) Neuwirth, E. *RColorBrewer: ColorBrewer Palettes*; 2022.