OXFORD

Structural bioinformatics

Deuteros: software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry

Andy M. C. Lau (iD) **, Zainab Ahdash[†], Chloe Martens[†] and Argyris Politis***

Department of Chemistry, King's College London, London SE1 1DB, UK

*To whom correspondence should be addressed.
[†]The authors wish it to be known that these authors contributed equally.
Associate Editor: Alfonso Valencia

Abstract

Summary: Hydrogen deuterium exchange-mass spectrometry (HDX-MS) has emerged as a powerful technique for interrogating the conformational dynamics of proteins and their complexes. Currently, analysis of HDX-MS data remains a laborious procedure, mainly due to the lack of streamlined software to process the large datasets. We present Deuteros which is a standalone software designed to be coupled with Waters DynamX HDX data analysis software, allowing the rapid analysis and visualization of data from differential HDX-MS.

Availability and implementation: Deuteros is open-source and can be downloaded from https://github.com/andymlau/Deuteros, under the Apache 2.0 license. Written in MATLAB and supported on both Windows and MacOS. Requires the MATLAB runtime library. According to the Wellcome Trust and UK research councils' Common Principles on Data Policy on data, software and materials management and sharing, all data supporting this study will be openly available from the software repository.

Contact: argyris.politis@kcl.ac.uk

1 Introduction

Hydrogen deuterium exchange mass spectrometry (HDX-MS) is a structural technique which has garnered attention for its ability to assess protein-protein and protein-ligand interactions, protein folding, and the associated dynamics of these processes (Jensen and Rand, 2016; Konermann *et al.*, 2011; Masson *et al.*, 2017; Mistarz *et al.*, 2016). The basis of HDX-MS relies on the exchange of labile amide hydrogens of the protein backbones, for bulk deuterium within in solution. The protein of interest is allowed to undergo exchange in a deuterium-rich buffer for a set number of timepoints and then quenched. The protein is then enzymatically cleaved to the peptide level, and the mixture is subjected to liquid chromatography coupled to mass spectrometry (LC-MS). Using LC-MS, the mass of the peptide acquired through deuteration can be determined via a database search. Peptides which participate in hydrogen bonding of amide hydrogens result in lesser exchange (Marcsisin and Engen, 2010). Additionally, those which comprise the accessible surfaces of

proteins, may experience relatively greater deuteration, than those found in the protein interior (Marcsisin and Engen, 2010). In differential HDX-MS (ΔHDX-MS), peptides from a reference state are compared with those from an altered state (which could be for instance a mutation or a ligand) to report on regions of the protein which are affected by structural or conformational perturbations.

After data acquisition, the analysis of HDX-MS data using Waters Instrumentation consists of several steps: (i) peptide identification using the Protein Lynx Global Server (Waters Corp.), (ii) peptide mass assignment using DynamX (Waters Corp.) and (iii) visualization and interpretation of data. In particular, interpretation of the results can be challenging due to the size of the datasets involved. A typical HDX-MS experiment will result in the order of 10^2 peptides depending on the system size and complexity. Several visualization methods have been introduced to provide clarity on the ensemble of peptides generated from HDX-MS. Waters offers in DynamX, 'butterfly' and 'difference' plots which while useful during

data analysis, have several shortcomings. They do not show the identity of each peptide, their lengths or regional peptide redundancy. As an alternative, the 'Woods' plot (Woods and Hamuro, 2001) (developed separately from Waters), provides a per-timepoint breakdown of the peptide ensemble, with each subplot displaying peptide length, start and end residues, global coverage and a y-axis metric which may be absolute uptake (in Daltons) or relative fractional uptake (RFU). HDX data can additionally be visualized on uptake maps, and molecular representations where uptake and other data are projected onto 3-dimensional structures of the system (Kavan and Man, 2011).

To simplify the HDX-MS analysis workflow, we have developed Deuteros for the rapid visualization of differential HDX-MS data. Development of the software was primarily motivated by the lack of data representation methods for differential HDX-MS. Deuteros has been designed to be used post-Dynamx, for the downstream statistical filtering and visualization of data from differential HDX-MS. It provides alternative visualization capabilities in the form of statistically filtered 'Woods' plots and through PyMOL-compatible scripts, which allow differential HDX-MS data to be projected onto three-dimensional structures of proteins. A combination of these visualization methods provides users with the ability to effortlessly identify biologically interestingly regions of proteins. Finally, inputs to Deuteros have been standardized to csv files, allowing the software to be potentially compatible with any instrumentation.

We have applied our software to a comparison of the wild-type xylose transporter (XylE) and a E153Q mutant. XylE is a secondary membrane transporter protein tasked with the role of shuttling xylose sugar across bacterial cell membranes (Quistgaard *et al.*, 2013). A member of the Major Facilitator Superfamily (MFS), XylE operates through an alternating-access mechanism, transitioning between inward-facing and outward-facing conformational states in a highly dynamic fashion (Wisedchaisri *et al.*, 2014).

2 How does it work?

Deuteros is a standalone MATLAB application available to both MacOS and Windows and requires the MATLAB runtime library. Deuteros is designed to assist interpretation of HDX-MS data that has been analyzed using DynamX (Waters Corp.) and provides complementary data visualization capabilities. Deuteros requires two inputs: the 'state' and 'difference' files exported from DynamX. Deuteros consists of four steps including data input and three visualization stages: flattened data maps, Woods plot (with statistical peptide filtering) and output to PyMOL.

Input data: The DynamX 'state' file contains a per-protein, per-peptide, per-time point aggregation of peptide deuterium uptake data from the ΔHDX-MS conditions. State files contain information including m/z, maximum possible deuterium uptake, observed deuterium uptake, standard deviation, retention time, and any residue modifications reported. Users should only enable proteins and states of interest and disable all others within the DynamX session file. The 'difference' file contains a per-peptide, per-timepoint comparison of peptide deuterium uptake from two user defined states. The difference file can only be generated from DynamX when two or more states are loaded into the dataset. Users should ensure that the correct comparison is made by selecting the correct states within DynamX. A video tutorial and example datasets have been provided alongside the software.

Visualization: Deuteros produces flattened data maps including coverage, residue-level redundancy, deuterium uptake heat maps and Woods plots. Coverage, redundancy and various deuterium uptake styles can also be exported from Deuteros, to be projected onto atomic models of the protein of interest in the PyMOL molecular graphics viewer (Schrödinger, 2015). Users can simply 'drag and drop' these files into PyMOL (for MacOS, or PyMOL version 2.0 and above for Windows), or alternatively copy and paste the contents of the file into the PyMOL command line.

Statistics: Confidence limits are calculated as (Houde *et al.*, 2011):

$$0 \pm \left(\frac{\sigma_t}{\sqrt{N}} \right) \cdot \alpha \tag{1}$$

where σ_t is the standard deviation of the mean uptake for timepoint t, N is the number of sample replicates and α is the critical value desired. By default, Deuteros provides critical values for 98 and 99% confidence limits (6.965 and 9.925) for a two-tailed t-test with $df = 2$ degrees of freedom. For 'sum' data, where peptide deuterium uptake differences from each timepoint are aggregated together to better identify potential peptides that are conformationally active, the confidence limits are calculated using:

$$0 \pm \left(\sqrt{\frac{\sum_{t=1}^{n} \sigma_t^2}{\sqrt{N_n}}} \right) \cdot \alpha \tag{2}$$

where σ_t^2 is the variance of all peptides for timepoint t, N is the number of timepoint observations for the variance, n is the number of timepoints and α is the critical value as in (1). Errors are propagated as a simple sum of variables as according to (Houde *et al.*, 2011).

Application: To showcase the capabilities of Deuteros, we imported state and difference files from the wild-type and E153Q mutant XylE membrane transporter Martens *et al.* (2018). The coverage map of XylE indicate a 92.5% coverage, with the largest non-covered region around residues 230–240 (Fig. 1a). The redundancy map expands on the coverage map, displaying the same coverage, but with a white-magenta color gradient to represent peptide redundancy. Reviewing the map shows that the highest redundancy of the XylE dataset was at 12 peptide copies around residues 465 (Fig. 1b). The N-terminal, residues 90, 180, 260 and 360 have only a single peptide representing these regions.

The Woods plot section displays a per-timepoint breakdown of the differential dataset in a grid layout. Deuteros can display a maximum of approximately 8 timepoints simultaneously before individual Woods plots become too crowded, depending on the screen size and resolution. Woods plots first apply confidence filtering to all peptides in each timepoint (Fig. 1c). Peptides with differential deuteration outside of the user selected confidence limits are non-significant and are shown in grey. The significant peptides are shown as red for deprotected and blue for those that are protected. While only one set of confidence limits are applied to the data, two boundaries are shown on each Woods plot as a visual aid for users to view which peptides might be significant, should they wish to tighten or relax the filter used. The legend section displays the confidence limits as ± Da (to two decimal places) values around 0 (or no difference). To facilitate interpretation, significant peptides can also be exported as a *csv* file containing a per-peptide per-timepoint breakdown of the ΔHDX-MS data. Users may also take advantage of the in-built MATLAB data cursor which displays the residue number and differential uptake of a peptide by clicking on the peptide within the graphical user interface.

The PyMOL export section consists of options for formatting the data from the linear coverage map and Woods plot sections, for visualization in PyMOL through *pml* files. Coverage and
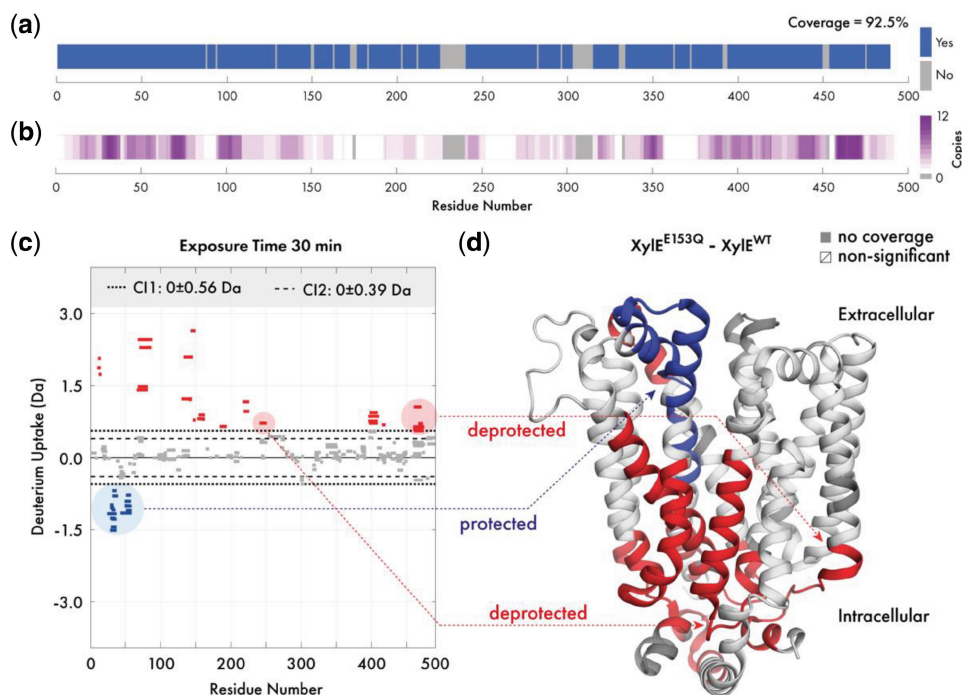
Fig. 1. Overview of Deuteros demonstrated on the XylE transporter. Visualization of (**a**) experimental protein coverage, (**b**) data redundancy and (**c**) deuterium up-take differences in Woods plot format. Dashed and dotted lines indicate 98 and 99% confidence limits applied to the dataset to identify peptides with significant deuteration differences. Deprotected, protected and non-significantly different peptides are in red, blue and grey respectively. (**d**) Differential HDX-MS data for the wild-type and E153Q mutant XylE has been projected onto its crystal structure (PDB ID: 4GBY) (Color version of this figure is available at *Bioinformatics* online.)

redundancy can be projected onto structures and a range of color palettes are available. Differential deuteration data can also be exported for projection onto the molecular structure of XylE (PDB ID: 4GBY; Fig. 1d). For this representation, the deuteration data type can show absolute differential uptake (in Daltons), or the differential relative fractional uptake (ΔRFU). The ΔRFU considers the peptide length and its maximum deuteration and scales the absolute uptake as a percentage of this value, which may be more informative for some datasets. Similar to the Woods plot, Deuteros implements red/blue/white/grey color scheme for protected, deprotected, non-significant and non-covered regions. Through projection of deuteration data onto the structure of XylE, structural effects caused by the E153Q mutation are immediately visible (Fig. 1d). The extracellular-facing portion of XylE experiences protection (blue), while the intracellular portion experiences deprotection (red).

Acknowledgements

Funding

References

Houde,D. *et al.* (2011) The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *J. Pharm. Sci.*, **100**, 2071–2086.

Jensen,P.F. and Rand,K.D. (2016) Hydrogen exchange. In: Weis,D.D. (ed.) *Hydrogen Exchange Mass Spectrometry of Proteins*. Wiley, USA, doi: 10.1002/9781118703748.ch1.

Kavan,D. and Man,P. (2011) MSTools-Web based application for visualization and presentation of HXMS data. *Int. J. Mass Spectrom.*, **302**, 53–58.

Konermann,L. *et al.* (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.*, **40**, 1224–1234.

Martens,C. *et al.* (2018) Direct protein-lipid interactions shape the conformational landscape of secondary transporters. *Nat. Commun.*, **9**, 4151.

Marcsisin,S.R. and Engen,J.R. (2010) Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Anal. Bioanal. Chem.*, **397**, 967–972.

Masson,G.R. *et al.* (2017) An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery. *Exp. Opin. Drug Discov.*, **12**, 981–994.

Mistarz,U.H. *et al.* (2016) Probing the binding interfaces of protein complexes using gas-phase H/D exchange mass spectrometry. *Structure*, **24**, 310–318.

Quistgaard,E.M. *et al.* (2013) Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nat. Struct. Mol. Biol.*, **20**, 766–768.

Schrödinger,L. (2015) The PyMOL Molecular Graphics System, Version 1.8. Schrödinger, LLC.

Wisedchaisri,G. *et al.* (2014) Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nat. Commun.*, **5**, 4521.

Woods,V.L., Jr and Hamuro,Y. (2001) High resolution, high-throughput amide deuterium exchange-mass spectrometry (DXMS) determination of protein binding site structure and dynamics: utility in pharmaceutical design. *J. Cell Biochem. Suppl.*, **37**, 89–98.