HaDeX: an R package and web-server for analysis of data from hydrogen deuterium exchange - mass spectrometry experiments

Weronika Puchała [1], Michał Burdukiewicz [2], Michał Kistowski [1], Katarzyna Dabrowa-Dabrowska [1], Aleksandra E. Badaczewska-Dawid [3], Dominik Cysewski [1,*], Michał Dadlez [1]
*dominikcysewski@gmail.com

[1]Institute of Biochemistry and Biophysics Polish Academy of Sciences, Poland, [2]Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland, [3]Faculty of Chemistry, Biological and Chemical Research Center, University of Warsaw, Poland.

Introduction

Hydrogen-deuterium mass spectrometry (HDX-MS) is a staple tool for monitoring dynamics and interactions of proteins. Due to the sheer size of the HDX-MS results, the data analysis require a dedicated software suite. However, the majority of existing tools provides only point-and-click interfaces to black-box models or does not offer a complete workflow. We propose **HaDeX**, a novel tool for processing, analysis and visualisation of HDX-MS experiments. **HaDeX** covers the whole analytic process, including preliminary data exploration, quality control and generation of publication-quality figures. The reproducibility of the whole procedure is ensured with advanced reporting functions.

Coverage of the sequence

The **HaDeX** web server relies only on a single data format: DynamiX datafile. An user can customize each step of the analysis with his own input parameters.
We start our analysis with summary of protein coverage (as shown below). It is provided in two forms, histogram and graphic comparison of positions. Also the reconstruction of the sequence is available - based on experimental data.

app_coverage.pdf

Figure 1: Sequence coverage analysis (**HaDeX** web server).

Differential plot

An important aspect of analysing data is a comparison of peptide deuteration in different states. A relative deuteration level used by **HaDeX** is defined by the equation 1.

$$D = \frac{D_t - D_0}{D_{100} - D_0} \tag{1}$$

The equation 1 is a function of three variables (each with its own uncertainty), so there is need to use the Law of Propagation of Uncertainty, defined in equation 2

$$u_c(y) = \sqrt{\sum_k \left[\frac{\partial y}{\partial x_k} u(x_k)\right]^2} \tag{2}$$

Applying equation 2 to equation 1, we get deuteration level uncertainty described by equation 3

$$u_c(D) = \sqrt{\left[\frac{1}{D_{100} - D_0} u(D_t)\right]^2 + \left[\frac{D_t - D_{100}}{(D_{100} - D_0)^2} u(D_0)\right]^2 + \left[\frac{D_0 - D_t}{(D_{100} - D_0)^2} u(D_{100})\right]^2} \tag{3}$$

The levels of deuteration along with uncertainty intervals are shown on comparison plot 2. **HaDeX** provide both experimental and theroetical levels of deuteration.

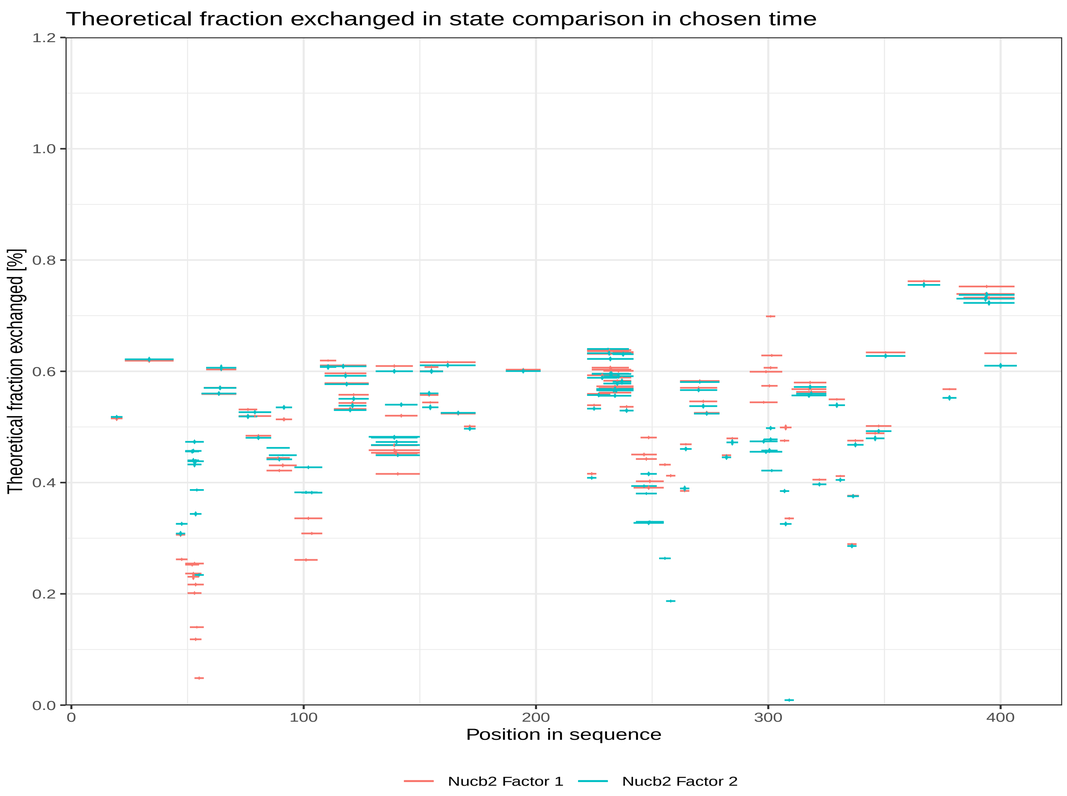

Figure 2: Comparison plots - comparing experimental and theoretical calculations, based on example data.

Woods plot format

Woods plot format shows diference between results of two different states as described by equation 4.

$$diff = D_1 - D_2 \tag{4}$$

Equation 4 is a function of two variables, so applying 2 we get uncertainty described by equation 5.

$$u_c(diff) = \sqrt{u(D_1)^2 + u(D_2)^2} \tag{5}$$

Described calculations lead to results shown on Woods plot format - figure 3 below.

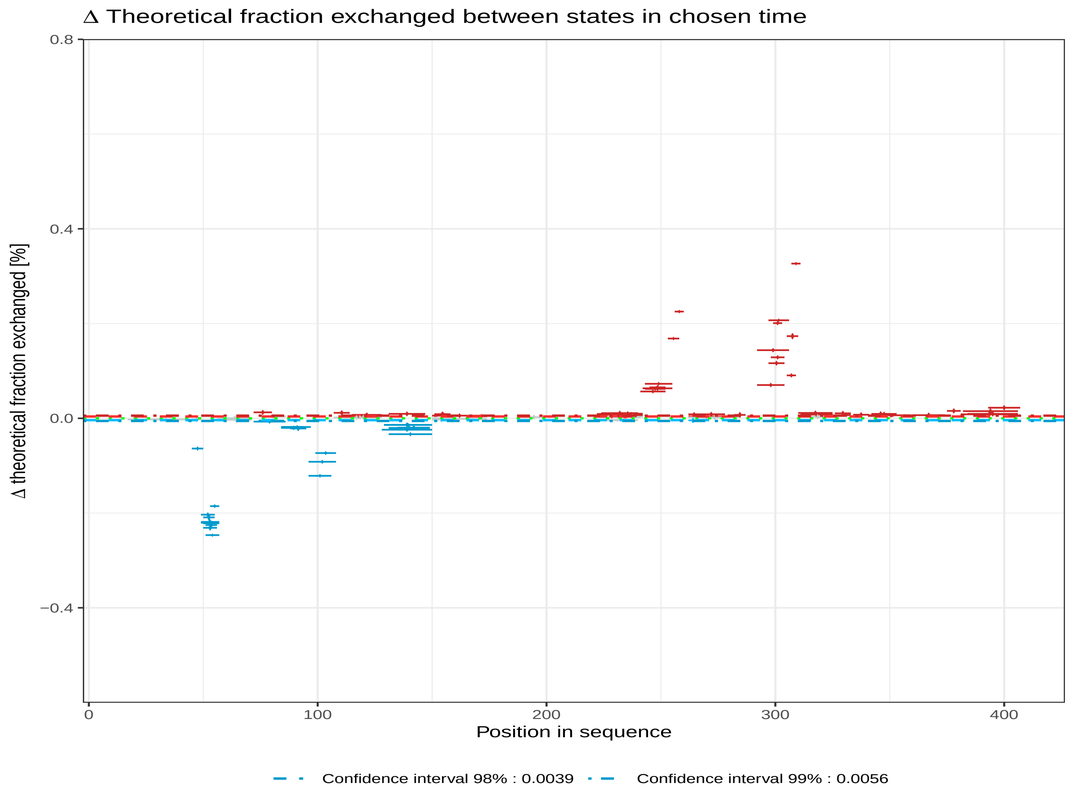

Figure 3: Plots in Woods format - comparing experimental and theoretical calculations, based on example data.

Quality control

HaDeX provides additional tools for assement of experiments. For example, the quality control function shows how the mean uncertainty per peptide changes with time points of an experiment.

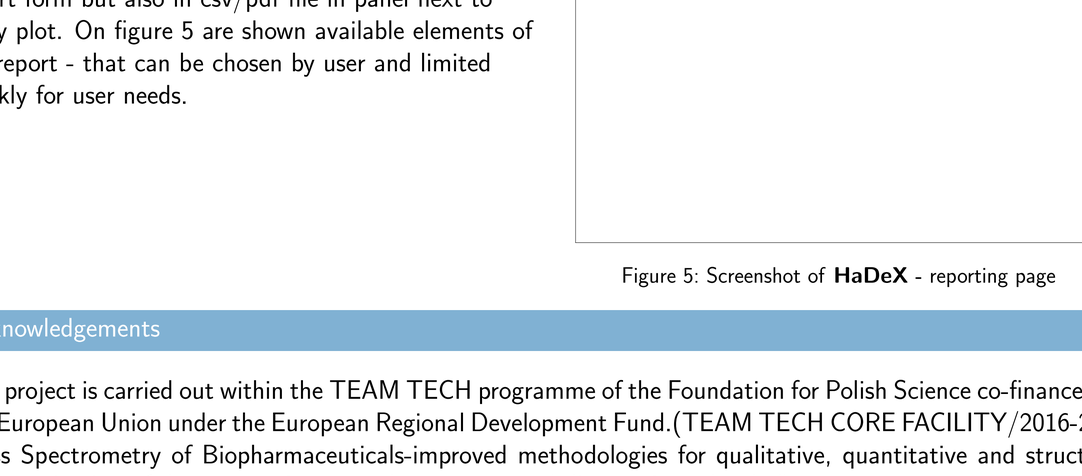

Figure 4: Plot ilustrating uncertainy change depending on used out parameter

Report function

HaDeX provides report generation - with parameters chosen by user so is fully personalizable. Additionaly, calculated data for plots can be downloaded not only in report form but also in csv/pdf file in panel next to every plot. On figure 5 are shown available elements of the report - that can be chosen by user and limited strickly for user needs.

app_report.png

Figure 5: Screenshot of **HaDeX** - reporting page

Acknowledgements

The project is carried out within the TEAM TECH programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.(TEAM TECH CORE FACILITY/2016-2/2) Mass Spectrometry of Biopharmaceuticals-improved methodologies for qualitative, quantitative and structural characterization of drugs, proteinaceous drug targets and diagnostic molecules"

European Funds Smart Growth
FNP Foundation for Polish Science
European Union European Regional Development Fund

Bibliography

1 Start2Fold: a database of hydrogen/deuterium exchange data on protein folding and stability. Pancsa R, Varadi M, Tompa P, Vranken WF. Nucleic Acids Res. 2016 Jan 4;44(D1):D429-34. doi: 10.1093/nar/gkv1185. Epub 2015 Nov 17

2 HDX workbench: software for the analysis of H/D exchange MS data. Pascal BD, Willis S, Lauer JL, Landgraf RR, West GM, Marciano D, Novick S, Goswami D, Chalmers MJ, Griffin PR. J Am Soc Mass Spectrom. 2012 Sep;23(9):1512-21. doi: 10.1007/s13361-012-0419-6. Epub 2012 Jun 13.

3 MEMHDX: an interactive tool to expedite the statistical validation and visualization of large HDX-MS datasets. Hourdel V, Volant S, O'Brien DP, Chenal A, Chamot-Rooke J, Dillies MA, Brier S. Bioinformatics. 2016 Nov 15;32(22):3413-3419. Epub 2016 Jul 13