Supporting Information for:

## Automating data analysis for hydrogen/deuterium exchange mass spectrometry using data-independent acquisition methodology

Frantisek Filandr<sup>1#</sup>, Vladimir Sarpe<sup>1#</sup>, Shaunak Raval<sup>1,2#</sup>, D. Alex Crowder<sup>1</sup>, Morgan F. Khan<sup>1</sup>,

Pauline Douglas<sup>1</sup>, Stephen Coales<sup>3</sup>, Rosa Viner<sup>4</sup>, Aleem Syed<sup>5</sup>, John A. Tainer<sup>6,7</sup>, Susan P. Lees-

Miller<sup>1</sup>, David C. Schriemer<sup>1,2\*</sup>

<sup>4</sup>Thermo Fisher Scientific, San Jose, California, USA

<sup>#</sup> Authors contributed equally to the study

<sup>&</sup>lt;sup>1</sup>Department of Biochemistry and Molecular Biology, University of Calgary, Alberta, Canada, T2N 4N1

<sup>&</sup>lt;sup>2</sup>Department of Chemistry, University of Calgary, Calgary, Alberta, Canada, T2N 4N1

<sup>&</sup>lt;sup>3</sup>Trajan Scientific & Medical - Raleigh, Morrisville, North Carolina, USA

<sup>&</sup>lt;sup>5</sup>Division of Radiation and Genome Instability, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA

<sup>&</sup>lt;sup>6</sup>Department of Molecular and Cellular Oncology, Department of Cancer Biology, The

University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>7</sup>Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>\*</sup> To whom correspondence should be addressed: <u>dschriem@ucalgary.ca</u>

**Contents:** 

Supplementary Figure 1. RANSAC-based isotopologue filtering

**Supplementary Figure 2.** Comparison of MS1-derived and AutoHX-derived deuteration values for individual peptides

**Supplementary Figure 3.** Full naïve processing of Pol  $\Theta$  data showing typical HX-MS result before manual validation.

**Tutorial - AutoHX** 

#### **AutoHX - Supplement:**



**Supplementary Figure 1. RANSAC-based isotopologue filtering.** Peaks that deviate significantly from the deuteration model fit can be automatically unselected and not used for deuteration calculation. This is performed in both MS1 and MS2 data space to improve deuteration calculation. Examples are extracted from the Phosphorylase B kinetics dataset used in **Fig. 6** (main text).Red arrows indicate spurious peaks that the algorithm avoids selecting.

































**Supplementary Figure 2.** Comparison of MS1-derived and AutoHX-derived deuteration values for individual peptides. Kinetics plots from Phosphorylase B kinetics experiment showcasing manual MS<sup>1</sup>-derived (Blue lines) and AutoHX-derived (Red lines) deuteration values. These values were used in the form of heatmap for **Fig. 6**, showing all 380 peptides. Data generated from n=3 biologically independent samples and the error bars represent +/- two standard deviations.



**Supplementary Figure 3. Full naïve processing of Pol O.** This Woods Plot shows typical HX-MS result before manual validation.

# **Tutorial -AutoHX for HX-DIA data analysis**

A program is being established for early release and evaluation of the software. Interested individuals are encouraged to contact the corresponding author. In anticipate of release, we make the following tutorial available:

### Installation:

- 1. Install <u>.NET Framework 3.5 SP1</u>. For detailed instructions, visit this link.
- 2. [*Optional*] Install the latest version of <u>Java 64-bit</u>. This is required for some external tools (MS-GF+ and Percolator) used in HX-PIPE.
- 3. Run **setup.exe**. If you run the .msi directly, it will skip the check for additional system requirements. See below for more info.
- 4. During installation, you may be prompted to also install multiple "Visual C++ Redistributable 20XX" versions (2008, 2010, 2012, 2013, 2015+), if they are not already on your system. These pre-requisites are needed to make vendor software and other dependencies work. If you wish to install them manually, visit <u>this link</u>.

### System requirements:

- Windows 7+ x64 with .NET 4.7.2 or higher.
- Multi-core CPU.
- 8GB+ RAM.
- 50GB+ of free disk space.

## **Create Project**

Presently, AutoHX is a sub-module under the standard HX-DEAL workflow. Therefore, project creation for AutoHX in Mass Spec Studio follows the same overall wizard-based flow as regular HX-DEAL. The major difference is that DIA files are required to use the new HX-DIA processing routine and DIA bins need to be supplied for all the runs. Please follow these steps to get started:

1. **"New Project":** Create a new HX-DEAL project by selecting "New Project" and selecting "HX-DEAL". The default location for new projects will be inside your home directory. You can select a different location by clicking "Browse" and selecting a new directory.

Mew Project		– 🗆 X		
HX-DEAL (Deuterium Exchange Analysis)	^	Used to analyze and quantify hydrogen/deuterium		
PIPE (Peptide Identification and Peptide Extractor)	exchange from mass spec data (LC-MS). Analysis ca			
CRIMP (Crosslink Identification)		done using a pre-determined peptide list or by generating a list of peptides using the HX-		
CLEAN (Covalent Label Estimation and Normalization)		PIPE analysis package.		
Data Explorer				
IMProv (Integrative Modeling Platform)				
DIA Library Creation (BETA)				
OpenDIA (BETA)				
	$\sim$			
Name: HX-DEAL Project4				
Location: C:\Users\vsarp\Desktop\Mass Spec Studio Demo\Projects	v	Browse		
		OK Cancel		

2. "Add Proteins": If you have more than a single protein or wish to use sequence visualizations, you can add your protein sequences either manually ("New"), via a FASTA file ("Select FASTA"), via a PDB file ("Select PDB(s)"), or via a 4-letter PDB code ("Fetch PDB(s)", example: "1jff"). Each protein will be listed as a separate row. If you have both PDB and FASTA, you can add the FASTA file first and then use the "Browse" button to link the PDB file. If you add both the PDB file and the FASTA file spearately, you may end up with duplicate proteins in the table. Important: For the sequence visualizations to work, the names of the proteins must match those found in the Peptide .csv file (explained later). The "Merge polymers" option can be used to remove duplicate sequences from multimeric protein complexes which may appear as separate chains inside a PDB file. Example PDB-code: "1fu1".

🛀 New	HX Experiment				_		×
Add P	roteins						
To add view.	l proteins, drag and drop	PDB or FASTA file	s or use the commands on the right	side of the			
	Name	Sequence	PDB File Path			New	
~	DNA POLYMERASE THETA;   A	Hover to view	C:\Users\vsarp\AppData\Local\Temp \5aga.pdb	Browse	Sel Fe	Edit lect PDB(s tch PDB(s lect FAST/ Remove	s) ;) A
Fetchi Done	ing PDB(s) '5aga' Fetching '5aga'.						
Me mo	erge polymers - appearing pnomeric protein.	g as separate chai	ns in a PDB file - into one represent	ative			
						Next	

3. **"Add Protein States":** For simple analysis, at least 1 protein state is required. For the additional comparative analysis tools and visualizations to become available, you must supply at least 2 proteins states.

New HX Experiment			—		×
Add Protein States					
Free	Protein state 2				
Bound	Common Prop	erties			
	Name	Bound			
Add Protein State Remove					
		<b>D</b> 1		N	
		Back		Next	

4. **"Add Labeling":** At least 1 labeling condition is required (both time and %D<sub>2</sub>O). For kinetics visualizations to become available, you must supply at least 2 labeling conditions.

New HX Experiment		_		×
Add Labeling				
5(80)	1(20)			
	Labeling			
	Percent D <sub>2</sub> O 80			
	Time (min) 5			-
Add Labeling Remove				
	Back		Next	

5. "Add Runs": Follow these steps to add your data:

a. Click "Browse" and select the root folder which contains your raw data files. In cases where the data files themselves are directories (example: Waters, Bruker), please make sure the containing folder is selected not the top-level .raw or .d directory itself. Note: Not all files have to be located under the same root folder from the start. After you add some files, you can still "Browse" to a different root directory and select additional files from the new location.

b. For most raw vendor files, <u>we strongly recommend to use "ProteoWizard Data Provider"</u>. If you already have Mass Spec Studio converted files (.mssdata, .mssmeta), select "Mass Spec Studio Data Provider".

c. We recommend you enable the "Convert to mssdata" checkbox. When ON, raw vendor files will be converted to the .mssdata format which enables super-fast searching at the cost of additional disk space for the .mssdata files. You can still proceed without converting the files, but the processing will be very slow. The "Noise filter" value is a multiplier of the minimum

signal in each spectrum. A noise filter of "2" will remove any intensities smaller than 2 \* mininum. A value of "0" will not remove any data.

d. Select your replicates (shift+click or ctrl+click for multi-select), select the appropriate Protein State/Labeling node and click the ">" button. If you make mistakes, you can remove runs using the "<" button.

Mew HX Expe	riment		– 🗆 ×
Add Runs			
Location: Data Provider: File Types: Bound-1.wiff Bound-2.wiff	E:\MS Data\multi single state ProteoWizard Data Provider Any spectra format		<ul> <li>✓ Browse</li> <li>✓ Convert to mssdata</li> <li>✓ Noise Filter: 1</li> <li>▲ S(80)</li> <li>■ Free-1.wiff (filter=1) Convert: ✓ Filter: 1</li> </ul>
<ul> <li>Bound-3.wiff</li> <li>Bound-4.wiff</li> <li>Bound-5.wiff</li> <li>Bound-6.wiff</li> <li>Free-1.wiff</li> <li>Free-2.wiff</li> <li>Free-3.wiff</li> </ul>		>	<ul> <li>Free-2.wiff (filter=1) Convert: ♥ Filter: 1</li> <li>Free-3.wiff (filter=1) Convert: ♥ Filter: 1</li> <li>Free-4.wiff (filter=1) Convert: ♥ Filter: 1</li> <li>Free-5.wiff (filter=1) Convert: ♥ Filter: 1</li> <li>Free-6.wiff (filter=1) Convert: ♥ Filter: 1</li> <li>\$ Sound</li> <li>\$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$</li></ul>
<ul> <li>Free-4.wiff</li> <li>Free-5.wiff</li> <li>Free-6.wiff</li> </ul>			<ul> <li>Bound-1.wiff (filter=1) Convert: ✓ Filter: 1</li> <li>Bound-2.wiff (filter=1) Convert: ✓ Filter: 1</li> <li>Bound-3.wiff (filter=1) Convert: ✓ Filter: 1</li> <li>Bound-4.wiff (filter=1) Convert: ✓ Filter: 1</li> <li>Bound-5.wiff (filter=1) Convert: ✓ Filter: 1</li> <li>Bound-6.wiff (filter=1) Convert: ✓ Filter: 1</li> </ul>
			Back Next

6. **"Configure Runs":** For DIA files, select the "MSMS Data Independent Acquisition (DIA)" option. This will unlock a new column in the runs table for adding fragmentation bins (isolation window) for each run. The bins should be in ".csv" format with "Start" and "End" columns which represent the boundaries of each DIA isolation window. If all runs have the same isolation windows, the file can be supplied once under the first row representing "All" runs.

Male Ne	w HX Experiment			_		×
Conf	igure Runs					
0 0 0	MS Only MSMS Data Dependent Acquisition (DE MSMS Data Independent Acquisition (E	DA) DIA)				
	Run	Fragmentation	Fragmentation Bins Path (.csv)			$\sim$
	All	CID Y	I Sample Data\bins.csv Browse			
	Bound-1.wiff	CID Y	H Sample Data\bins.csv Browse			
	Bound-2.wiff	CID ~	- Sample Data\bins.csv Browse			
	Bound-3.wiff	CID ~	H Sample Data\bins.csv Browse			
	Bound-4.wiff	CID Y	H Sample Data\bins.csv Browse			
	Bound-5.wiff	CID ~	I Sample Data\bins.csv Browse			
	Bound-6.wiff	CID Y	H Sample Data\bins.csv Browse			
	Free-1.wiff	CID ~	I Sample Data\bins.csv Browse			
	Free-2.wiff	CID ~	I Sample Data\bins.csv Browse			
	Free-3.wiff	CID v	H Sample Data\bins.csv Browse			
	Free-4.wiff	CID ~	H Sample Data\bins.csv Browse			~
			Back		Next	

7. **"Provide a pre-identified peptide list (.csv)":** Click "Browse" and select a peptide list. Note: To avoid duplicates, it's recommended that your peptide identifications are grouped by sequence and charge such that each peptide (sequence/charge pair) has only one "RT". If you start with multiple separate peptide-spectrum matches (PSMs) for the same peptide, you can group them by sequence/charge and average their RTs to get the final peptide "RT". If you expect high discrepancies between the mapping runs and the DIA runs, you can define a larger "RT Variance" column in your peptides file. HX-DIA does not need a pre-selected transitions file because it will check all possible fragments for each peptide. However, if you wish to constrain the deuteration calculations to a pre-selected list of fragments, please supply them in the preselected transitions file.

A sample peptide list: <u>https://s3.us-west-2.amazonaws.com/www.msstudio.ca/assets/tutorials/hx-deal/sample\_peptides.csv</u>

A sample transitions list (optional): <u>https://s3.us-west-</u> 2.amazonaws.com/www.msstudio.ca/assets/tutorials/hx-deal/sample\_transitions.tsv

些 New HX Experiment		-	– c	з х
Peptide or Transition Info	rmation (MS2)			
Peptides (.csv) E:\MS Data\m	lti single state\peptides.csv		~	Browse
	Peptide file path is not valid.			
Pre-selected Transitions (.tsv)			Ŷ	Browse
Note: Pre-selected transitions provide a peptide list without	re optional. If you provide a transition list, only those f transitions list, all possible fragments will be generate	ragments will be se d and searched.	arched. I	fyou
	_			
		Back	Ne	xt

#### Processing

In order to use the HX-DIA processing routine (core of AutoHX), you must first enable "Beta Features" from the Tool -> Performance Manager window. A password is required to enable the AutoHX features – please contact us to obtain one.

To get started with processing, you can open the processing window from the Process menu and select the "Peptide HX-DIA" routine. The most important parameters to set correct are the mass accuracy and the peptide elution time parameters. The rest of the parameters are already set to the default values we found most useful during our internal testing. For advanced users, the advanced parameters to fine-tune the search can be enabled via the top-right "Advanced Parameters" checkbox. If you wish to know more about the parameters, you can click on the help tooltip buttons ("?") next to each processing step/subsection.

cessing Routines	Peptide HX-DIA		Advanced Pa	rame
Hydrogen-Deuterium Exchange	Mass Tolerance			
Peptide HX-DIA	Mass Tolerance Option	PPM		Ŷ
	Mass Tolerance	10		
	MS/MS XIC Smoothing		Savitzky Golay Smoothing ×	?
	▲ Generic			
	Enabled	$\checkmark$		
	Re-Sampling Distance	0.01		
	Savitzky Golay			
	Window Size	5		
	MS/MS Smoothing		(none)	v
Show Compatible Routines Only			L	
ecent Parameters	MS/MS Peak Picking		Basic m/z Peak Detector	?
oad 6/29/2023 4:15:08 PM	4 m/z Peak Detector			
ad 7/11/2023 1:30:41 PM	Peak To Background Ratio	0.05		
ad 7/11/2023 1:34:25 PM				
ad 7/11/2023 1:53:33 PM	Deuteration Results Generator		Deuteration Results Generator	~
ad 7/11/2023 2:06:23 PM	▲ Flagging			
	Max standard deviation across r	3		
ved Parameters				
	Comparative (Statistics)		Two State Welch's T-test Calculato	or ~
	<ul> <li>DIA Settings</li> </ul>		DIA Settings ~	?
	Fragment Model Refinement			
	Certainty Error Threshold (%)	3		
	Disable Peptides With Large Up	1		

Once everything is set correctly, click "Process". The total processing time can span from minutes to hours, depending on the number of runs and the number of peptides.

#### Results

After processing is finished, the result will be saved to disk inside the "Results" folder of the project and will be auto-selected in the left-hand-side project tree. AutoHX calculates the "Rescued" deuteration in addition to the standard MS1-based deuteration using the MS2 DIA fragments. The "Final" deuteration for each peptide is based on a decision whether to use the MS1 or MS2 (rescued) deuteration, depending on the quality of the MS1 and MS2 data. For the most part, MS2 deuteration is used for the "Final" deuteration because of its measurement redundancy -- multiple fragments representing the same peptide deuteration during full scrambling, even if the precursor is overlapped in MS1. All of the deuteration values for each peptide are displayed in the right-hand-side Properties tab (per replicate).

The "Manual Validation Tab" in AutoHX is mainly used to inspect the results. We generally don't recommend any manual intervention after the results are generated because peptides have

already gone through multiple stages of strict filtering. To view the decisions for the full set of peptides, including ones which don't pass the final filters, you can disable the hide fragged peptides options in the Processed Peptides list (Major and Critical).

The visualization from HX-DEAL have been updated to show the final AutoHX results. Most notably, the Woods plot now displays certainty bars for each peptide which are determined by the degree of consensus on the peptide deuteration after sampling different sets of fragments. Additionally, the Sequence Coverage view provides a way to map the peptide deuteration or differences onto a 3D structure via the "Export Pymol" button. This button generated a script that can be loaded in PyMOL to colorize the 3D structure based on the data displayed in the Sequence Coverage view.









## Export

To export the results, click on File -> Export and follow the export wizard. This follows the standard HX-DEAL export wizard which allows you to select and configure everything from peptide-level .csv exports to snapshots of the commonly used visualizations (Sequence Coverage, Peptide Map, Kinetics, Woods Plot, Volcano Plot). Note that if you wish to export the visualizations, they will need to be configured once before the export. The "Configuration" column will display all the selected exports that require manual configuration before the final export bundle is created. Configuration usually involves setting a template for how the visual should look in the final export bundle (size, DPI, fonts, etc.).

些 HX Export Wizard

#### Select result to export

Select a validated result:

Name
[Not Open] Peptide HX-DIA Result - 2023-07-11 13_31_55
[Not Open] Peptide HX-DIA Result - 2023-07-11 14_16_25
[Not Open] Peptide HX-DIA Result - 2023-07-18 13_34_19
[Not Open] Peptide HX-DIA Result - 2023-07-18 15_28_50
Peptide HX-DIA Result - 2023-07-19 16.19.28
[Not Open] Peptide HX-MS Result - 2023-07-18 13_16_39
[Not Open] Peptide HX-MS Result - 2023-07-18 13_32_39
[Not Open] Peptide HX-MS Result - 2023-07-18 15_26_39
NU - O - 1 D - C L UV MO D - N - 2022 07 40 45 34 00

Next

\_

 $\times$ 

HX Export Wizard			_		×		
Select data to export							
1. Select a location to export to: Browse C:\Users\vsarp\Desktop\hx-dia-results							
2. Select at least a target protein state. For comparative an	alvsis export, select a	control protein state:					
Target State: PolQ+Novobiocin ✓ Control Sta	te: PolQ Y (Co	omparative)					
1-p Threshold: 0.95 SD Multipler: 2	(Confidence cutoffs	for significant difference	es)				
3 Exclude flagged pentides: Major Critical			,				
4 Select and configure date:							
4. Select and configure data:							
Selected Name	Configure	Configuration					
Raw Data (.csv)		Ready					
Gothenburg formats (.csv, .xlsx)		Ready					
Comparative T-Test Results (.csv)		Ready					
✓ Woods plot (.tiff)	Configure	Ready					
✓ Volcano plot (.tiff)	Configure	Ready					
✓ Kinetics plots (.tiff)	Configure	Not Configured					
Sequence Coverage ( tiff)	Configure	Not Configured			$\sim$		
		Back	E	xport			