HDXWizard Operating Instructions

Version 1.0

Zachary Cohen

Cohen.za@northeastern.edu

Contents

1.Introduction, Compatibility, and Contact Information

Welcome to HDXWizard, a tool designed with a simple, intuitive user interface for the best and fastest processing of HDX data into chiclet plots, difference plots, and colored peptide plots and peptide difference plots. All plots can be analyzed using “experimental” (maxD) mode or “theoretical” mode to allow for diverse applications to different data sets. Data is inputted by the user from Waters DynamX state data. If you have any questions or feature requests, please email them to cohen.za@northeastern.edu.

2.File Entry

2.1) Preparing State Data

After data has been processed in DynamX, it can be exported as “State Data” into a .csv file. This file, or a .xlsx file can be uploaded directly to be analyzed. As many files as desired can be analyzed at once, but please make sure that states that are not the same have different “states” and/or different “proteins”.

To create a mock DynamX state data, make an excel file with columns:

A: Protein

B: Start

C: End

D: Peptide Sequence

E:

F:

G: Filled

H: Filled

I: State

J: Exposure Time

K: Filled

L: Filled

M: Uptake

N: Uptake SD

Columns E and F may be left empty. Columns G, H, K, and L should all be filled with any character.

2.2.1) Preparing Sequence

If you only want to create a chiclet or chiclet difference plot, you can select the “skip” button. If you select the skip button and try to make any peptide plots, the sequence will be read as AAA etc. If you are only analyzing one protein, it is possible to enter a plain text file containing only the sequence of the protein using the “.txt (p)” button. The user can also enter a .fasta file or a .txt file in the format of a .fasta file using the “.fasta” and “.txt (>)” buttons. As many of these as desired can be added, but please make sure the protein name in the fasta file (names formatted as “>Protein Protein” will be read as “Protein”) is exactly the same as in the state data file, otherwise the sequence will be replaced with AAA etc.

3. Selecting a Preference for RFU Calculation or MaxD Correction

3.1.1) Theoretical Preference

Selecting the theoretical button will provide any chiclet and peptide plots as their relative fractional uptake, or RFU by dividing the measured uptake by the theoretical maximum uptake which is calculated by:

$$Max\ Theoretical\ Uptake = \ Length\ of\ Peptide - Prolines - 1$$

If the first residue is not proline, or:

$$Max\ Theoretical\ Uptake_{N\ terminal\ P} = Length\ of\ Peptide - Prolines$$

If the first residue is proline.

If duplicate peptides exist for a state, they will be averaged.

3.1.2) Customizing Back Exchange

It is possible to customize coloring of peptide plots for an amount of back exchange by entering a number into the box labeled "Insert Back Exchange". Numbers entered should be the percent back exchange you would like to select, such as "25". If no number is placed into the box, back exchange will be set to 0%. Maximum theoretical uptakes are multiplied by $(1 - \frac{back\ exchange}{100})$, and peptide uptake values at timepoints are divided by this number to find the RFU.

3.1.3) Differences maps with Theoretical Preference

When the theoretical preference is on, all differences will be the raw uptake differences between states in Daltons.

3.2.1) Experimental Preference

If you have data for a maximally deuterated control (maxD) that you would like to be corrected for, then selecting the experimental preference is the correct option. Please make sure that your maxD data is the largest timepoint. You can, but do not have to have the maxD in every state if the maxD data is compatible with those states. If duplicate peptides exist for a state, they will be averaged.

3.2.2) Missing maxD peptides

If the value of the highest peptide is missing (or was set to -99999 by the user), then the average RFU for all maxD peptides in the state is calculated, and the theoretical maximum deuterium incorporation of the peptide is multiplied by the average RFU of maxD peptides in the state.

$$Corrected\ RFU = \frac{Uptake}{MaxD\ Uptake}$$

or

$$Corrected\ RFU = \frac{Uptake}{Max\ Theoretical\ Uptake} * Average\ RFU$$

For peptides where the assigned maxD value, preference is missing, the second equation is used, and peptides are marked with an asterisk on peptide plots. In chiclet plots, the whole row is marked with an asterisk.

3.2.3) MaxD State Selection

Automatically, the program will attempt to get the maxD of each peptide from the same state as itself, however, after selecting "Experimental", many dropdown menus will show up, and allow you to change the state where the maxD for the peptide is looked for.

3.2.4) Creating Custom States

By clicking the create custom states button, you can create an average of two states from which to take the maxD measurements from. Select the two states you wish to average, and click save state, then apply it to any states you want to use it for. Average RFU is calculated for both states' maxD measurements combined. If you are having trouble with your custom state, make sure there are no "|" symbols in the name of either state you want to average.

You cannot use custom states to directly generate figures, only to set as a maxD. If you want to create a map of the average of two states, simply put them into the file with the same state and protein name, and they will be averaged automatically.

4. Adding and Customizing Differences

4.1) Adding Differences

Differences can be added by clicking the "+Dif" button. Two dropdown lists with states from your file should become available. If there are no states here, the program was not able to read your file. Select the states you want to create a difference for. The second state will be subtracted from the first. Then, give a title to this state (less than 25 characters). This title will be the title of any sheets and chiclet plots created for this state, so make it descriptive. At this time, only 12 differences can be created at once. If you wish to create more, you will have to run the program multiple times.

4.2) Calculation of Differences Using Theoretical Preference

When the theoretical preference is on, all differences will be calculated using the absolute uptakes between states.

4.3) Calculation of Differences Using Experimental Preference

When the experimental preference is on, differences between states are calculated by subtracting the RFU of the second state from the RFU of the first state.

4.4) Differences with Missing Timepoint and Peptides

Missing peptides or timepoints will not affect the accuracy of output. If peptides are not present in one difference state but are in the other, they will not appear in the difference maps at all.

5. General Notes on Data Processing

5.1) Absent Peptides

If a peptide is not present for a timepoint but is present for other timepoints in the same state, it is considered absent, and will be colored peach (by default), and will not affect data. Any differences using absent peptides will produce an absent, peach colored peptide in the difference plot. If any average is taken between an absent peptide and a peptide with a value, the new value will ignore the absent value.

5.2) Duplicate Peptides

If duplicate peptides are present with the same state name, the uptake values will be averaged before anything else is done to the data.

5.3) Program Runtime

The most time-consuming part of the code is finding where to place each peptide as peptide plots are created. Larger proteins are proteins that produce more peptides and will take much longer to analyze than smaller proteins. Having more states in the file will also make the program take longer to run. Expected runtimes are anywhere between 10 seconds for smaller proteins with minimal states and 5 minutes for larger proteins with 20 or more states.

6. Chiclet Plot Functionalities

6.1) Chiclet Plot

By selecting the "Chiclet" button before hitting run, the program will create a chiclet plot of the peptides from your file. The peptides in your state data do not need to be in order. They are sorted by start value, with tiebreakers being decided by end value before being placed into the plot. Each chiclet plot is titled using the state name from the state data. The timepoint 0 is automatically removed. If you want to bypass this, you can replace the name of that timepoint with another nonzero timepoint that would also be the lowest timepoint and change the timepoint post processing in excel. Chiclet plots are colored according to section 8.1.

6.2) Chiclet Difference Plot

By selecting the "Chiclet Difference" button before hitting run, the program will create a difference plot in the style of a chiclet plot between any two states you have added for a difference. Peptides that do not exist for both states are omitted from the difference completely. Each difference plot will be titled as the difference name inputted for the difference. Difference plots are colored according to section 8.2. Please see section 6.1 for more information on chiclet plot creation.

7. Peptide Plot Functionalities

7.1) Peptide Plot

By selecting the "Peptide Plot" button before hitting run, the program will create a series of sheets, each titled as the appropriate state from the state data file. Each sheet contains a series of timepoints in order of colored peptides. Peptides are placed by finding the highest place on the sheet within the given timepoint where there would not be overlayed upon another peptide. Peptides are placed in the order of their start value, with the end value breaking any ties. For coloring details, see section 8.1.

7.2) Condensed Peptide Plot

By selecting the "Condensed Peptide" button before hitting run, the program will create a series of sheets, each titled as the state shown, along with "cond". Output shows uptake in percentage as described in section 3. Condensed peptide plots are similar to peptide plots (see section 7.1), but the

peptides are shorter, and exchange values are overlayed over the peptides. For coloring details see section 8.1.

7.3 Difference Plot

By selecting the “Peptide Difference” button before hitting run, the program will create a sheet for each difference input. The sheets are titled as the difference name given. Differences are calculated as described in section 4. Difference plots are otherwise very similar to peptide plots (see section 7.1). For coloring details, see section 8.2.

7.4 Condensed Difference Plot

By selecting the “Condensed Difference” button before hitting run, the program will create a sheet for each difference input. Sheets are titled as the given difference name, along with “cond”. Differences are calculated as described in section 4. When the theoretical preference is on, differences are in Daltons. When the experimental preference is on, differences are shown in percentage (out of 100).

8. Formatting Options

8.1) Coloring of Plots

Plot color is completely customizable with up to 10 colors for uptake plots and 10 colors for difference plots. Once custom colors are generated, they are saved as a .JSON file on your computer and can be used anytime in the future by selecting that file in the dropdown menu. For more information, click “Create Custom Colors” in the formatting section and then click “See Examples”.

8.2) Other Formatting Options

By default, peptides are automatically sorted by peptide start, with tiebreaks being decided by peptide end in the chiclet plot. For chiclet plots, this can be turned off if desired, and peptides will go into chiclet plot in the order of the file.

Cell widths can be adjusted as desired. Making the cell width narrower when trying to show errors in the condensed peptide plot may cause spacing issues.

For condensed difference plots, the difference values can be displayed for insignificant values or not. By default, they are.

Additionally, for all condensed peptide plots, error is added by default, which is calculated using the uptake standard deviation in the state data file. This can be turned off.

9. Saving your File

When the program has finished running, you will be prompted to save the workbook.