HDgraphiX

<https://heatmapvahidi.azurewebsites.net/>

\*\*\*CUSTOM DOMAIN\*\*\*

Developed by Dr. Algirdas Velyvis and Kent Vosper

Vahidi Lab, University of Guelph

Current Version: 1.1

HDgraphiX has been developed as a user-friendly HDX-data visualization tool.

There are four types of outputs that are common for HDX analysis: heatmaps/chiclet plots, Woods plots, volcano plots, and PyMOL colouring scripts. There are several options present to alter how the data is handled, which will be discussed later in the document.

The development of HDgraphiX has prioritized user-friendliness, and therefore should be easy to use, however there are options which can become complicated. This document serves as a guide to all of the options present on the website.

If you find any bugs in the code (eg. Output is incorrect) please email [kvosper@uoguelph.ca](mailto:kvosper@uoguelph.ca). We would additionally welcome any feedback or questions pertaining to the use/function of HDgraphiX.

Contents

1. Data Inputs

2. Types of Plots

3. Bound and Colour Input Methods

4. Advanced Options

5. Outputs

1. **Data Inputs**

Data is accepted as two different types of files: DynamX Cluster .csv files or HDExaminer PeptidePool .csv files. Only single file analyses are possible with this tool, however with either of these inputs it appears straightforward to copy/paste files together. Refer to figures 1 and 2 to confirm the proper headings/columns are present in your input file.

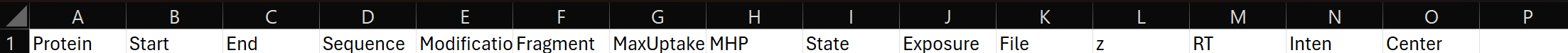


Figure 1: Column headings and layout present for DynamX Cluster Output.

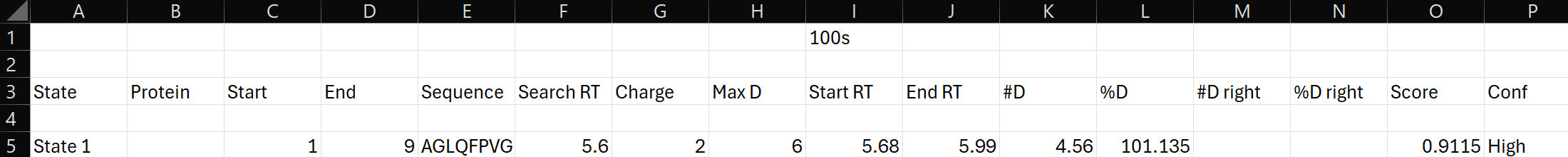


Figure 2: Column headings and layout for HDExaminer Peptide Pool Output.

1. **Types of Plots**

There are 3 types of plots that can be generated: heatmaps (horizontal/vertical plotting), Woods plots, and volcano plots. For Woods plots the bounds/colours inputs will not matter unless heatmap colouring is selected. For Volcano plots these inputs will not matter. Additional inputs are available in the Advanced Options section for Woods and volcano plot generation. A fourth visualization option available is for PyMOL colouring scripts, which will color PyMOL structures according to the heatmap colouring generated.

1. **Bound and Colour Input Methods**

There are 4 options for Bound/Colour Input present. They are listed in order of easiest/least user input to the most complicated (but providing the most customization). Of note: any input colour values should be valid hexadecimal values (e.g. #00FF00).

The first option is to allow the code to determine the maximum range, which will then automatically set the bounds using 7 colours positive values and 7 for negative values. This option requires no user input and will be selected by default. The next option allows the user to set the maximum range, and how many colours to use. Both of these options also provide the opportunity to specify alternate colours to use, however if blue (negative) and red (positive) are desired this can be ignored. The other shades of the colour will be automatically generated using a colour mixing function (mixing to white therefore the colour becomes lighter closer to 0).

The third option allows the user to select colours to use for negative/positive values from a drop-down menu. There are 6 colour options: Blue, Red, Green, Yellow, Purple, and Black. The user then inputs the bounds they wish to use. When inputting the bounds both lists should start with the lowest value up to the highest value. For negative values this would be: -5, -4, -3, -2, -1. For positive values this would be 1, 2, 3, 4, 5. The positive and negative lists should be the same length

The final option allows the most customization. The first option present is to choose whether to input the darkest colour values for positive/negative values, or to input all colour values. If the darkest colour values option is chosen the additional shades will be generated with the same function used in all other options but can allow whatever colouring choice is desired. The other option for colouring does not use a colour mixing function, but rather every hexadecimal value needs to be manually input. These should follow the same order as the bounds, however 1 fewer input per list (e.g. if the list of negative bounds is 4 entries long there should be 3 negative colours input). Therefore, for negative values the colours should be listed darkest to lightest, while for positive values it should start with the lightest to the darkest colour. Bound inputs are the same as for option 3, e.g. -3, -2, -1 and 1, 2, 3.

1. **Advanced Options**

An important note for most advanced options is the presence of a tooltip for the majority of these. These are presented as text saying either “How to use …” or “More Info” leaving the mouse cursor over this text will bring up a small text box describing the use of these features, and any formatting necessities. Additionally, the majority of these options have a box next to them. To use these functions/options the boxes must be selected, or any inputs for that option will be ignored.

* 1. Drop Times

This option allows as many inputs as you desire, as it will generate and interpret as many inputs as specified. Any time points specified in this option will be removed from data processing, and not appear in the generated plot. These inputs must exactly match the time point/exposure values in the input code (e.g. time in minutes, with up to 3 decimal points). If this is not exactly matched it can cause the code to not function. If this happens you can just reload the page, and double check the exposure values.

* 1. Drop Peptides

Similar to Drop Times, this option will take into account as many inputs as is specified. These will allow the removal of specific peptides which are not desired in the analysis. This can be due to insignificance, such as peptides which are far from the area of interest and not displaying large differences in deuterium uptake. For any peptide desired to be removed 3 inputs are required: the protein name exactly as found in the input file, the start residue number, and the end residue number. For HDExaminer inputs, the PeptidePool format does not specify a protein name. When converting this format to DynamX Cluster format (the code does this for the user) a generic protein name “gen\_prot” will be applied to ensure the code works properly. Input this in the protein name field for this option to work properly with HDExaminer outputs.

* 1. Drop Proteins

This option takes in up to 6 inputs for proteins to not compare. Of note: HDExaminer does not include protein names in the PeptidePool format, and therefore this option is of no use for these applications. For DynamX Cluster file inputs this can work if many proteins are present in a single data file, and only a few are desired to be compared. The protein names provided must exactly match the input file. It is recommended to copy and paste directly from the input .csv file.

* 1. Renumber Residues

When analyzing HDX data the first peptide may not align with the first residue seen in the protein sequence. This option therefore allows for renumbering the residues of an entirety of a protein in the input file to better align with PDB structures/sequences. Each renumbering requires 2 inputs (protein name and value to shift by), and up to 2 proteins can be renumbered. The value input will be subtracted from every residue number. Therefore, if Protein\_1 and 30 are input, and a residue being analyzed is listed as 33 – 45 in the data file, the output will list it as 3 – 15 instead as 30 is subtracted.

* 1. Labelling Domains

If there are specific domains of the protein which are desired to be labelled, this option can be used. For each domain which you desire to label there are 4 inputs. The first is the label which will be applied. The next two are for specifying the peptides which should be labelled. This is not specified via the residue numbering, but the order of peptides. Therefore, if there is a list of 30 peptides, and you would like to label the first to tenth peptide, you would input 1 and 10. The final input is for colouring the label, as it will produce a coloured rectangle extending the specified length. This should be input as a hexadecimal value, beginning with a pound sign (e.g. #FF0000). This is the only optional input, as if nothing is input it will produce a black rectangle. This option can be used for heatmaps or for Woods plots.

* 1. Mutant Options

By default, these options will be hidden, and will only display upon selecting the option “Mutants in Data”. There are two options to deal with mutants in the data file: Mutation ID Dictionary and Mutation Dictionary. The mutation ID dictionary requires simply the mutant protein ID and the wild-type protein ID, allowing the code to compare the mutant to the wild-type despite different protein names in the input. The input IDs must exactly match the protein names in the data file. The mutation dictionary is an alternate way to handle mutations, where numbering between the two is different. The residue number of the mutant is provided, and the equivalent residue in the wild-type is provided, allowing the alignment of the mutant to the wild-type.

* 1. Compare Everything/Select States

By default, this option will be hidden and will only be displayed upon selecting the option “Select States to Compare”. This allows the input of two state lists, which will be compared against each other. The first input in list 2 will be subtracted from the first input in list 1, the second input subtracted from the second input and so on. If this is not specified every unique comparison will be done. However, this may not be desired, as it may not display the comparison in the desired format. E.g. if the states present are State1 and State2 and you wish to display State2 – State 1, the code will display only State1 – State2, and therefore the state lists will be needed to display the “correct” comparison.

* 1. Colour Insignificant

This option will add 0 to the list of bounds. This will cause every single square to be coloured in a heatmap, as not matter how close to 0 it gets it will simply be coloured as the lightest shade provided (either positive or negative).

* 1. Separate Heatmap Plots

When selected all state comparisons will be downloaded as separate plots, when turned off all state comparisons for one protein will be stacked on top of each other. It is selected by default.

* 1. All Time Labels in One Unit

By default, time labels will be a mix of seconds, minutes, and hours depending on which unit/units best fit for the specific time point. With this option all y-axis time labels will be in the specified unit.

* 1. Change Title Padding

This allows for changing the spacing between the axis subtitles and the main figure title. By default, this will be 20 for both values which ensures the labels do not overlap. These can be increased or decreased as desired.

* 1. Heatmap Plot Lines Colour and Thickness

There are 4 different sets of options in this section. The first allows for changing the design of the lines separating individual cells/chiclets in the plots. By default, these will be white and 4 pixels thick, and this option allows altering both the colour and thickness. These separating lines can be removed by setting the thickness to 0. The next option is only applicable when stacking state comparisons (see section 4.8). This allows for the altering of the line separating the state comparisons, both in thickness and colour. By default, this will be black and 4 pixels thick. The next provides the same options, but for the border around the plot. This option uses slightly different sizing for the width, and therefore a default of 10 is used, which is approximately equivalent to 4 as used in the other options. The final option is for altering the ticks and tick labels. The length and width of the ticks can be controlled, as well as the colour for the ticks and labels. For colour all of these options use drop-down menus, allowing for a selection from 5 pre-set colour values.

* 1. Woods Plots Options

The first of the Woods plot options allows for alternate colouring of the plot. By default, the code will colour peptides outside of the cutoff value a set colour, with no gradient. This first option allows for colouring according to the heatmap options and colour gradient. The next option allows for the setting of the colouring cutoff value, rather than the default value of 0.5. The next option allows the changing of the plot dimensions in inches. Following this is an option for setting the colours. By default, these colours will be blue for negative, white for near zero, and red for positive. It is again important to note that inputs here must be valid hexadecimal values. The next option for Woods plots allows for the removal of the horizontal threshold/colour cutoff markers. This can allow for a cleaner looking plot, as there will be less on the plot itself. The final option allows for changing the thickness of the peptide markers, both for the coloured section and the border.

* 1. Volcano Plots Options

There are two options for Volcano plots. The first is to set the p-value threshold, and therefore the lines generated on the plot. The other option is to change the colour of the plot points rather than black. It is again important to note that inputs here must be valid hexadecimal values.

* 1. DPI

When generating a PNG file, the default DPI will be 100. This can be increased to 400. Of note increasing the DPI will increase processing time, and therefore how long it takes to generate the plots.

* 1. PDF or PNG

By default, the website will generate the plots as PDF files, however this can be changed into PNG files instead. The option is by default checked on (therefore generating a PDF) and unchecking the box will cause PNG generation.

1. **Outputs**

The output will be sent as a .zip file. This is to ensure the website does not spam the user with downloads when multiple plots are generated. Woods and volcano plots generate separate plots for each exposure time point and for each comparison. If there are multiple proteins, there will be multiple heatmap plots. In addition, PyMOL files will be added to the downloaded file as a zip file within the zip file and will require another extraction to use.

Website Version History:

* Version 1.1
  + Updated website layout/design
  + Added additional options for Woods Plots and Volcano Plots
  + Changed initial website load to have the simplest choices selected, allowing for faster analysis for first time users
  + Altered some option titles/descriptions
  + Improved some plot axis labels and tick appearance
* Version 1.0
  + Working communication between Azure (storage and hosting), HTML template (website layout and inputs), and Python/Flask (code for data handling and plot generation)
  + Basic layout for website
  + Added Woods Plots option