# Introduction

DeuteRater is a software tool designed to allow scientists to perform kinetic proteomic and lipidomic experiments on small mammals or cell culture. It was published in the journal Bioinformatics(1). Source code is available at <https://github.com/JC-Price/DeuteRater>, a .exe is available at https://github.com/JC-Price/DeuteRater/releases/tag/DeuteRater\_v5. This readme will walk through how to use the software. For proper experimental design please reference the associated paper. For the purposes of this readme the word neutromer is synonymous with isotope

## Version Choice

It is recommended that unless you need to alter the source code for some reason that you download the .exe as it is easy to use and should be resistant to bugs caused by the underlying python modules updating.

## Prerequisites

Both versions require a Windows operating system (they have only been tested on Windows 10, but other Windows versions may work as well).

Python source code was created in Python version 3.8.1, therefore that is the minimum python version recommended. Most imports are included in the basic python install, however you must ensure the follow packages are installed:

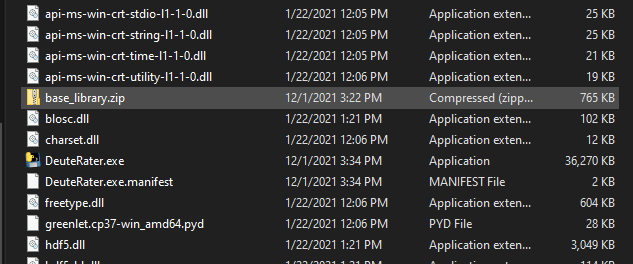
- PyQt5

- tqdm

# Installation

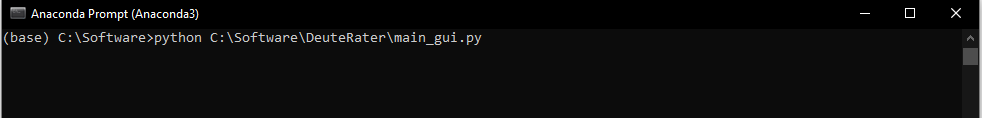
## Exe version

Download the zip folder from DeuteRater\_v5. Extract the folder to a location you can find easily on your hard drive. Inside the unzipped folder will be an .exe file:

Double click this .exe (or call it from the command prompt) to open DeuteRater. Nothing can be removed from this folder without risk of causing an error. If the position of the .exe is inconvenient you can right click on the .exe file and select “Create shortcut”. The shortcut can be moved anywhere you desire without causing an error.

## Python Source Code Version

Download the code from <https://github.com/JC-Price/DeuteRater>. All folders must be in the same directory for the code to work. Use a command prompt to call the \_\_main\_\_.py or main\_gui.py using python:



It doesn’t matter if you call the full path including \_\_main\_\_.py, just the parent folder path, or cd into the directory and call \_\_main\_\_.py from there.

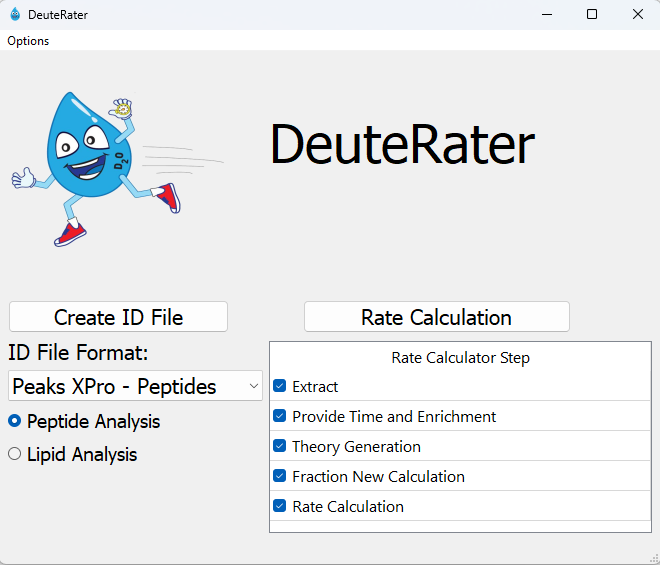
# Using DeuteRater

## Activating the Software

With the .exe version, double click on the .exe file or call the .exe on the command prompt. If using the python version, call \_\_main\_\_.py from the command line. This will open two windows. The first is the command prompt if you did not use it to open DeuteRater. It will be used to display progress bars and any errors that cause DeuteRater to fail:



Next is the main interface window that governs all DeuteRater operations:



We will go through each step in the workflow with input and output requirements. The first step is to collect the information we need to start the analysis.

## Create ID File

On the left of the interface is the “Create ID File” button. An ID File is a file containing all the peptides or lipids constituents you expect to see in your mass spectrometry files along with information to assist DeuteRater in locating the appropriate features, such as retention time and mass, as well as data that will be needed later such as the peptide sequence.

How the id file is created is governed by the pulldown menu below the “Create ID File” button. The default “Template” setting is best if you have not used PEAKS or MassHunter to identify your msms data.

When you click “Create ID File” with “Template” selected for the pulldown, you will be prompted for a location and a save file name. What you name this file and where you put it do not matter, so long as you can find it again. The file will be a .csv (comma separated values) file which can be opened in any spreadsheet program such as Microsoft Excel.

For proteomics, the following columns must be filled out or analysis will not proceed: Sequence, Protein ID, Precursor Retention Time (sec), Precursor m/z, and Identification Charge.

For lipidomics, the following columns must be filled out or analysis will not proceed: Precursor Retention Time (sec), Precursor m/z, Identification Charge, Lipid Unique Identifier, and Lipid Name.

Other columns that are needed for analysis will be filled in by DeuteRater if left blank (do note that if any blanks are present in any of these columns, all data in any of these columns will be overwritten).

All other columns are for the user’s information and may be left blank or filled with data as you wish. A description of all columns follows:

Sequence – the amino acid sequence. It should be in the form of single letter amino acid code. Post translational modifications (PTMs) should be included here. The recognized PTMs are found in the file “aa\_elem\_compositions.tsv” in the resources file of your DeuteRater download. This will be detailed in sections “Adjusting Amino Acids” and “Adjusting Post Translational Modifications”.

Protein ID – a unique identification code for the protein this peptide belongs to.

Protein Name – common name of the parent protein

Precursor Retention Time (sec) – the middle of the elution peak for this peptide or lipid constituent. Time unit is set in the settings menu (default is seconds)

rt\_start – the start of the elution peak.

rt\_end – the end of the elution peak.

rt\_width – the width of the elution peak.

Precursor m/z – observed m/z of the peptide or lipid constituent.

Peptide Theoretical Mass – the theoretical neutral mass of this peptide.

Identification Charge – observed charge state.

ptm – post translational modifications observed.

avg\_ppm – the average ppm mass error in the observed peptide.

start\_loc – the location the peptide begins in the protein.

end\_loc - the location the peptide ends in the protein.

num\_peptides – number of peptides observed for this protein in this analysis.

num\_unique – number of peptides observed for this protein in this analysis.

Homologous Proteins – identifiers for homologous proteins for this peptide sequence.

species – the species this protein is from.

gene\_name – the identifier for the gene associated with this protein.

protein\_existence – a code to indicate the confidence that this protein is biologically relevant.

sequence\_version – a legacy setting from a previous version. Leave it blank.

cf – chemical formula in standard format (e.g., C34H60N10O10).

neutromers\_to\_extract – number of neutromers to consider.

literature\_n – the number of deuteriums that can be incorporated.

Lipid Name – common name of the parent lipid.

Lipid Unique Identifier –

LMP –

HMP –

Matched\_Results\_Analysis –

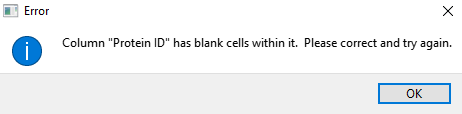
Matched\_Details\_Replicates\_Used –

Adduct –

Adduct\_cf –

Only peptides and lipid constituents you put into this document will be included in the search, so if you wish to search for multiple post translational modifications, or multiple charge states each must be on a separate line.

If the necessary columns are not completely filled, or there is a problem auto filling the necessary columns, an error message will appear when attempting to use the file, as shown below. Correct the problem, save the file, and proceed.



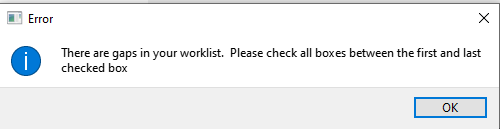
## Rate Calculation – General

The main portion of DeuteRater is the Rate Calculation button (red box below) which starts the analysis by going through each step checked in the “Rate Calculator Step” table (blue box below):

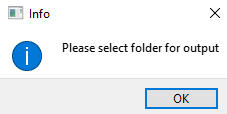
Graphical user interface, application

Description automatically generated

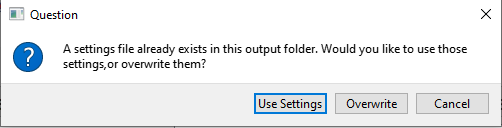
We will go over each step individually but first, a few general rules about it. All steps that are checked must be next to each other. If there is a gap, you will get an error message:



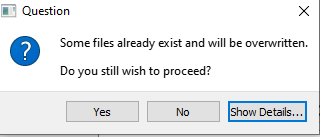
No matter which step is checked, you will be asked to select a folder at first:



The names of the output files are always the same. If you have used the folder you select, you will get two messages. The first will notify you that there is a settings output (settings.yaml) present. You will be given the option to use the settings in the folder (“Use Settings”) or overwrite this file and use the settings you have currently set in DeuteRater (“Overwrite”):



Second, if there are output files for a step that is currently checked present in your folder, you will be given a warning that you are going to overwrite output files. “Yes” overwrites the files, “No” returns you to the main DeuteRater interface, and “Show Details” will list the files that will be overwritten so you may make an informed decision. You can rename your files to avoid them being overwritten, but that must be done manually. DeuteRater will not do it for you:



In addition to output spreadsheet files, the “Rate Calculation” produces a folder of graphs. If this step is redone in the same parent folders, the graph subfolder will be deleted with all its contents.

If a file is open in another program while an overwrite is occurring, it cannot be overwritten. You will be given an error message and returned to the main interface. This will also occur if you are attempting to write to a folder which DeuteRater does not have permission to write to for whatever reason.

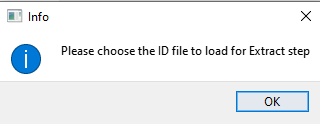
Each step will produce a .tsv output file. These are tab separated text documents which can be opened in any text editor or spreadsheet program like Microsoft Excel. The .tsv from one step can be used as the input file for the next step in the Rate Calculator, allowing you to start and stop if necessary (e.g. to redo part of the analysis with different settings). Be warned, that except for the ID file, DeuteRater will only check your input file for proper filetype and column names. It will not check that the data makes sense. Be careful changing intermediate files before re-using them as an input file for a later step, because changes may cause DeuteRater to error out or give nonsensical results.

When running a worklist, rate\_settings.yaml will be saved in your output folder. This should allow you to track what settings were used for each analysis should that be necessary.

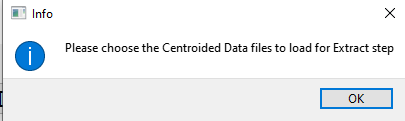
## Rate Calculation – Extract

The first step of Rate Calculation is “Extract”. The purpose of the Extract step is to find neutromers in centroided MS1 mass spectrometry data. This requires two things. The first is the ID file that you created earlier (see “Create ID File” section) and the second is one or more .mzML files. The .mzML files need to be centroided and MS level 1 only. If you require a utility to transform your mass spectrometry data, the msconvert tool from proteowizard can do this(2). Us the peakpicking setting to centroid.

First you will be asked to provide the id file:



Navigate to wherever you have the ID file and select it. You will then be prompted to select your .mzML files. You may select one or several at once:



This may prompt another window warning about overwriting files if you selected files that have already been extracted. This is because DeuteRater does not know the filenames of the extracted outputs until you choose the .mzML files to extract, so cannot determine which files may be overwritten until this point.

Each .mzML file you provided to DeuteRater will have a corresponding output file. The file will have the same name as the original with a .tsv extension instead of an .mzML extension. The file will be the same as the ID file with several extra columns:

id\_index – an internal index of the file

mz – calculated m/z value

Z – charge state

rt – retention time in minutes

n\_isos – neutromers to extract

mass – neutral mass

mzs – the actual mzs found in the .mzML

abundances – the actual abundances found in the .mzML

lookback\_mzs – mz of the neutromer peak in front of the first neutromer extracted if the extractor is extracting on a neutromer other than the first

lookback\_abundances – abundance of the neutromer peak in front of the first neutromer extracted if the extractor is extracting on a neutromer other than the first

lookahead\_mzs – mz of the neutromer peak behind the last neutromer extracted if the extractor is extracting insufficient neutromers.

lookback\_abundances - abundance of the neutromer peak behind the last neutromer extracted if the extractor is extracting insufficient neutromers.

m-1\_mz –

m-1\_abundance –

m\_end+1\_mz –

m\_end+1\_abundance –

rt\_min – minimum retention time observed

rt\_max – maximum retention time observed

baseline\_signal – abundance of the baseline in the area around this neutromer envelope

signal\_2\_noise –

mads – an error metric

mzs\_list – only filled for chromatography division setting.

intensities\_list – only filled for chromatography division setting.

rt\_list – only filled for chromatography division setting.

baseline\_list – only filled for chromatography division setting.

num\_scans\_combined – the number of scans used to generate relevant values.

Id\_path – the path to the ID file used for this data.

mzml\_path – the path to the .mzML file used for this data.

Extraction\_Updated –

Extraction\_Error –

mzml\_name –

name\_check –

filename –

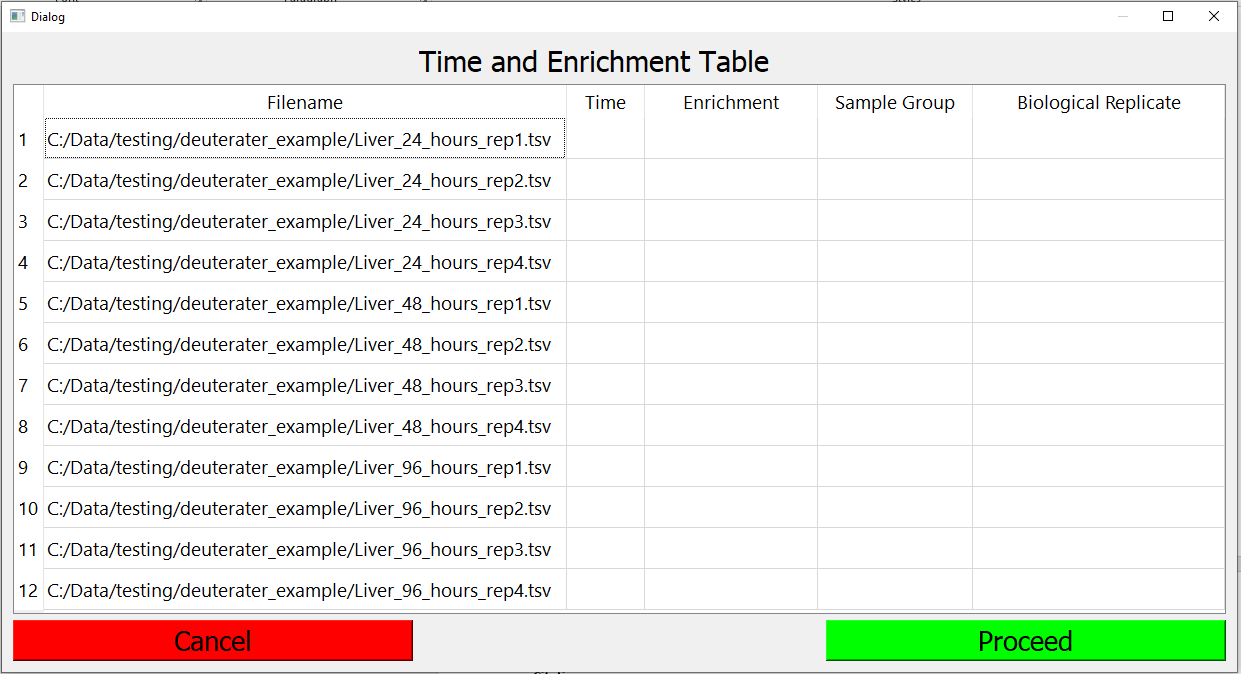
row\_num –

Columns mzs to num\_scans\_combined are only filled in if data was found for this row. If data was not found, and all of these columns are blank, the rows will be discarded in step “Theory Generation”.

You will see progress bars on the black command prompt that appeared with the interface (or the normal command prompt if you called the program from there). Do note that if you have “Provide Time and Enrichment” checked, you will be prompted to fill out the tables associated with that step after selecting .mzMLs but before the actual extraction happens. This is so you can walk away from the analysis without it stalling to ask for your input halfway through.

## Rate Calculation – Provide Time and Enrichment

The purpose of the “Provide Time and Enrichment” is to provide information on the .mzml files you are using. If you are starting with this step you will be asked to select the output .tsvs from an extractor step, otherwise it will continue with the data that was used in the extractor. If step is checked you will be provided with a table like the following:



This table must be filled out.

The “Filename” column cannot be altered without exiting the table and selecting different files.

The “Time” column is the time since deuterium introduction. The units do not matter, so long as they are consistent. It will change the interpretation. If you use hours rate will be in unit of hours-1, if you use days, rate will be in units of days-1.

The “Enrichment” column is the amount of deuterium in the body water of the subject, with natural enrichment being 0. Use decimal (so .05 not 5%).

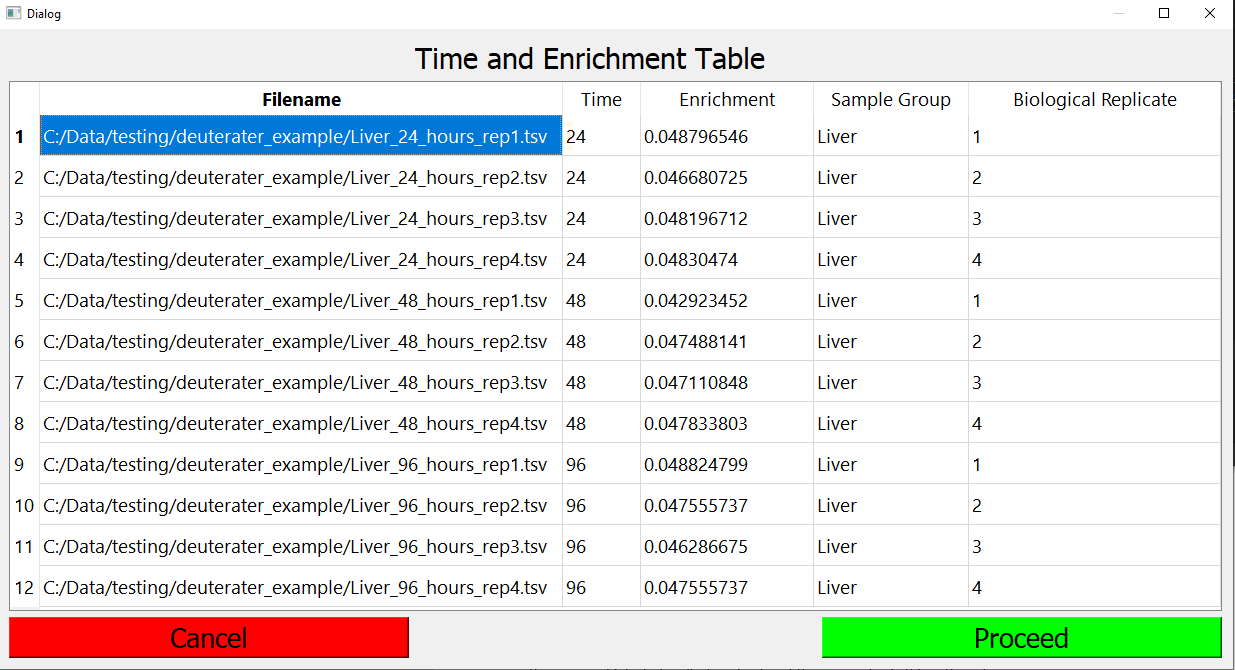
The “Sample Group” column is the group the file is a part of. If there are two groups being analyzed together, such as different tissues or different experimental treatments, this can be indicated by filling in some text here. If all files are part of the same group, provide the same text for each file. It will be kept in the titles of the output graphs and so on.

The “Biological Replicate” is to indicate if samples are biological replicates or not for later error calculation. Separate timepoints are always considered to be biological replicates for this purpose. If there are no biological replicates, simply put the same value in this entire column.

Ctrl+c for copy, Ctrl+v for paste and del and backspace for delete are enabled for this table.

Fill out the table and hit the green “Proceed” button. The red “Cancel” button will prevent further analysis and return you to the main interface. If there is a problem with the data you entered, such as a missing value or a non-numeric or negative time, a warning message will appear. You must correct the problem.

A completed table looks like this:



The output file is called “time\_encrichment\_data.tsv”. Its contents will appear exactly like the table you just filled out.

Warning: the next step uses the filenames in the “time\_enrichment\_data.tsv” file to find the extracted data. if these files have been renamed or moved since the “time\_enrichment\_data.tsv” file was created you must alter the filenames to match the current position of the extracted files if you want to use “time\_enrichment\_data.tsv” as an input.

## Rate Calculation – Theory Generation

The purpose of the “Theory Generation” step is a few small processes which are needed for later calculations. All extracted files are added together and any peptides that have no data in an extracted file are removed from that extracted file. The information from the table “Provide Time and Enrichment”, the time, enrichment, sample group, and biological replicate, are added to the table. The n-value, the number of hydrogens which can be replaced by deuteriums, is calculated for each peptide or lipid constituent.

The output file is similar to an extracted file with the “mzs\_list”, “intensities\_list”, “rt\_list”, and “baseline\_list” columns dropped and the “time”, “3nrichment”, “sample\_group”, and “bio\_rep” columns from the Time and Enrichment Table added.

## Rate Calculation – Fraction New Calculation

This step calculates the amount of turnover for each peptide or lipid constituent in each file so a rate can be calculated later. For each row in your database the unlabeled isotopic spectrum, in both abundance and m/z, will be calculated. These values will then be calculated assuming all copies of that peptide or lipid constituent have been newly synthesized with the enrichment provided for the relevant extracted file.

After the unlabeled spectra and the fully labeled spectra have been calculated, the amount of turnover that has actually occurred (fraction new) is calculated. This is done in three different ways:

Abundance: for each neutromer peak in a given row, turnover is calculated using the following equation:

This is done on all neutromers and the results are compared. By default, the monoisotopic peak is treated as the true value because the change tends to be largest and can only move in one direction. This can be adjusted in the settings menu

Neutromer Spacing: The neutromer spacing measures the distance in m/z units between the monoisotopic peak and the other neutromer peaks. This will increase with increased labeling. The equation is the same as for abundance, instead of the abundance of a peak, the m/z difference between the peak and the monoisotopic peak is used. Because the spacing tends to be noisier than the abundance, a median absolute deviation outlier check is used, and outliers are discarded. The median of the remaining points is used as the true value.

Combined: the results of Abundance and Neutromer Spacing are considered together. All calculated values for all neutromers calculated from Abundance and Neutromer Spacing are considered, a median absolute deviation outlier check is used, and outliers are discarded. The median of the remaining points is used as the true value.

The output file is called “frac\_new\_output.tsv”. Note that any columns relevant to a particular calculation type, for example the afn column for abundance, will not appear unless abundance calculations were performed. By default all calculation types are performed.

Many of the columns in this file are identification columns, like “Sequence” from previous steps. The new columns are:

time –

enrichment –

sample\_group –

bio\_rep –

adduct\_molecule\_sg –

n\_val\_calc\_n –

stddev –

dlt\_n –

dlt\_stddev –

n\_value –

low\_Cl\_n\_value –

high\_Cl\_n\_value –

theory\_unlabeled\_abunds – this contains the sum normalized values of each neutromer. For example, if the values were .5 .3 .2 it would indicate the no extra neutron neutromer represented 50% of the signal for this peptide, the one extra neutron neutromer represented 30% of the signal for this peptide, and the two extra neutron neutromer represented 20% of the signal for this peptide. Values are separated by a space. This column indicates the natural abundance values, with no extra deuterium added.

theory\_labeled\_abunds – this contains the sum normalized values of each neutromer. For example, if the values were .5 .3 .2 it would indicate the no extra neutron neutromer represented 50% of the signal for this peptide, the one extra neutron neutromer represented 30% of the signal for this peptide, and the two extra neutron neutromer represented 20% of the signal for this peptide. Values are separated by a space. This column indicates fully labeled values, at time infinity with constant deuterium addition.

normalized\_empirical\_abundances – this column represents the observed abundances for this peptide in the relevant file. These values are sum normalized. For example, if the values were .5 .3 .2 it would indicate the no extra neutron neutromer represented 50% of the signal for this peptide, the one extra neutron neutromer represented 30% of the signal for this peptide, and the two extra neutron neutromer represented 20% of the signal for this peptide. Values are separated by a space.

low\_labeling\_peaks – sometimes the maximum possible change (difference between a peak in theory\_unlabeled\_abunds and theory\_labeled\_abunds) this generally occurs for very low amounts of deuterium or if the deuterium incorporation is large enough that the neutromer peak initially increased in size and then decreased as more deuterium was incorporated (the one extra neutron neutromer peak often has this problem at 5% Deuterium incorporation). The threshold for calling a peak a low labeling peak can be adjusted. The low labeling peaks will not be included in the abundance calculations. The low labeling peaks with be indicated by M followed by the number of extra neutrons, so “M1” is the neutromer peak with one extra neutron.

frac\_new\_abunds – this column is the following equation used on each neutromer peak:

frac\_new\_abunds\_std\_dev – the standard deviation of all peaks calculated in frac\_new\_abunds. If this value is too high, the abundance measurement will be dropped from further consideration.

Abund\_fn – abundance fraction new. This the actual amount of the pool of the measured peptide that has turned over since the start of label introduction. This is calculated using the frac\_new\_abunds column, using either the M0 (no extra neutron) peak, the average of all non-excluded peaks, or the median of all non-excluded peaks. M0 is the default. If the peptide was not calculated due to a filter of some kind, that will be indicated here.

observed\_neutral\_masses – the observed neutral masses of each neutromer peak for this peptide from the relevant mzml. Values are space separated.

theory\_unlabeled\_mzs – the calculated neutral masses for each neutromer peak if no deuterium was incorporated. Values are space separated.

theory\_labeled\_mzs – the calculated neutral masses for each neutromer peak at time infinity with constant deuterium incorporation. Values are space separated.

frac\_new\_mzs – this column is the following equation used on each neutromer peak except m0 (note that theory m0 neutral mass does not change no matter the deuterium amount):

frac\_new\_mzs\_outlier\_checked – because distances between neutral masses are prone to noise due to how small the actual differences are, a median absolute deviation test is applied and outliers removed. This column contains the results of frac\_new\_mzs after the outliers have been removed.

frac\_new\_mzs\_std\_dev – the standard deviation of all peaks calculated in frac\_new\_mzs\_outlier\_checked. If this value is too high, the spacing measurement will be dropped from further consideration.

Nsfn – neutromer spacing fraction new. This the actual amount of the pool of the measured peptide that has turned over since the start of label introduction. This is median of frac\_new\_mzs\_outlier\_checked.

frac\_new\_combined – this column is used to combine results from spacing and abundance measurements. It contains results from frac\_new\_abunds and frac\_new\_mzs. Values are space delimited.

frac\_new\_combined\_outlier\_checked – because there may be disagreements between peaks and noise in the measurements, a median absolute deviation outlier check is used on the data from frac\_new\_combined. The passing measurements are placed in this column. Values are space delimited.

frac\_new\_combined\_std\_dev – the standard deviation of all peaks calculated in frac\_new\_combined\_outlier\_checked. If this value is too high, the spacing measurement will be dropped from further consideration.

Cfn – Combined Fraction New. This the actual amount of the pool of the measured peptide that has turned over since the start of label introduction. This is median of frac\_new\_combined\_outlier\_checked.

## Rate Calculation – Rate Calculation

This calculation fits the data from the Fraction New Calculation step to a kinetic rate curve with an equation of .

This option produces three outputs; Graph\_Folder, calculated\_rates.csv, calculated\_rates\_datapoints.tsv. We will go over each of them.

Graph\_Folder – A folder which contains the graphs of the fits performed during the calculation. The graph file names are in the format [protein id]\_[Subject Group]\_[relevant column name] or [lipid unique identifier]\_[Subject Group]\_[relevant column name] where relevant column name is the column name with the data from frac\_new\_output.tsv; afn for abundance calculations, nsfn for neutromer spacing calculations, and cfn for combined calculations.

The graphs themselves represent the fit line with a black line, and the 95% error of the graphing variables is represented by the gray shaded areas. The red dots represent the data that was fit to make the fit line.

The title of the graph on in the graph file itself is in the form of [protein name]\_[Subject Group]\_[relevant column name] or [lipid unique identifier]\_[Subject Group]\_[relevant column name] where relevant column name is the column name with the data from frac\_new\_output.tsv; afn for abundance calculations, nsfn for neutromer spacing calculations, and cfn for combined calculations. The second line contains the actual variable values calculated. K is the turnover rate (“rate” in the equation above) a is the asymptote (1 in the equation above). The asymptote by default is held constant at 1, its value can change or can be allowed to vary in the settings menu.

Calculated\_rates.csv – this is the primary numerical output for DeuteRater. Which columns exist are determined by the “verbose output” setting in the settings menu. If a calculation method is not used its relevant columns will not appear. For example, if only abundance is used for calculating turnover, neutromer spacing and combined columns will not appear.

The columns which will always appear are:

analyte\_id – the protein id for this protein, or the lipid unique identifier for this lipid.

analyte\_name – the protein or lipid common name

group\_name – the sample group name

Abundance rate – the turnover rate calculated using abundance measurements. Units are time-1 where time is whatever unit used when filling out the “Provide Time and Enrichment” table. If the value is words instead of a number, the protein or lipid could not be fit for some reason.

Abundance 95pct\_confidence – the 95% confidence interval for the Abundance Rate. So, the confidence interval would be written as “Abundance rate” ± “Abundance 95pct\_confidence”.

Abundance half life – Abundance rate written as a half-life instead of fraction turnover per unit time.

Spacing rate – the turnover rate calculated using neutromer spacing measurements. Units are time-1 where time is whatever unit used when filling out the “Provide Time and Enrichment” table. If the value is words instead of a number, the protein could not be fit for some reason.

Spacing 95pct\_confidence – the 95% confidence interval for the Spacing Rate. So, the confidence interval would be written as “Spacing rate” ± “Spacing 95pct\_confidence”.

Spacing half life – Spacing rate written as a half-life instead of fraction turnover per unit time.

Combined rate – the turnover rate calculated using both abundance and neutromer spacing measurements. Units are time-1 where time is whatever unit used when filling out the “Provide Time and Enrichment” table.

Combined 95pct\_confidence – the 95% confidence interval for the Combined Rate. So, the confidence interval would be written as “Combined rate” ± “Combined 95pct\_confidence”. If the value is words instead of a number, the protein could not be fit for some reason.

Combined half life – Combined rate written as a half-life instead of fraction turnover per unit time.

The columns added when “Verbose Output” is turned on are:

Abundance asymptote – the asymptote of the calculated Abundance fit.

Abundance std\_error – the standard error of the variables from the Abundance fit.

Abundance R2 – the r2 value of the Abunance fit

Abundance files observed in – the number of .mzml files used to make the fit using Abundance.

Abundance num\_measurements – the number of points used to fit the Abundance fit equation

Abundance num\_time\_points – the number of unique time points used to fit the Abundance fit equation

Abundance uniques – number of unique peptides used to fit the abundance fit equation

Abundance exceptions – if there was an error in the fitting process, this explains what it was

Spacing asymptote – the asymptote of the calculated neutromer spacing fit.

Spacing std\_error – the standard error of the variables from the neutromer spacing fit.

Spacing R2 – the r2 value of the neutromer spacing fit

Spacing files observed in – the number of .mzml files used to make the fit using neutromer spacing.

Spacing num\_measurements – the number of points used to fit the neutromer spacing fit equation

Spacing num\_time\_points – the number of unique time points used to fit the neutromer spacing fit equation

Spacing uniques – number of unique peptides used to fit the neutromer spacing fit equation

Spacing exceptions – if there was an error in the fitting process, this explains what it was

Combined asymptote – the asymptote of the calculated Combined fit.

Combined std\_error – the standard error of the variables from the Combined fit.

Combined R2 – the r2 value of the Combined fit

Combined files observed in – the number of .mzml files used to make the fit using Combined.

Combined num\_measurements – the number of points used to fit the Combined fit equation

Combined num\_time\_points – the number of unique time points used to fit the Combined fit equation

Combined uniques – number of unique peptides used to fit the Combined fit equation

Combined exceptions – if there was an error in the fitting process, this explains what it was.

Calculated\_rates\_datapoints.tsv – this file is the same as “frac\_new\_output.tsv” but only contains those points which were used to make the final fits.

## Rate Calculation – Error messages within files

In different steps of the Rate Calculation, a peptide or protein can be excluded from consideration for a number of reasons. If you look at the relevant intermediate file, an error will be shown by data missing from a row where it should be or a text message where data should be.

Errors in extracted files: the main error here is that a peptide that was present in the id file but was not found by the extractor. This is represented by the rows “mzs” to “num\_scans\_combined” are blank.

Errors in frac\_new\_output.tsv: the main error here is the N value being too low. This is a filter that can be changed in the settings menu. This will be represented by the text “N value is less than [filter value]” in the aft, nsfn, and cfn columns. All columns after the “bio\_rep” column that do not have the text in them will be blank.

Errors in calculated rates: the final rate calculation has the most different kinds of error messages. They will be present in every column relevant to the calculation type that failed.

No Isotope Envelopes Agree – this indicates that the standard deviation filter for this type of fit discarded all data points. The standard deviation filter requires frac\_new\_combined\_std\_dev, frac\_new\_mzs\_std\_dev, or frac\_new\_abunds\_std\_dev from the “frac\_new\_output.tsv” to be below a certain value or the point is discarded. By default, the standard deviation must be less than .1, but that value can be changed by changing default values.

Insufficient Timepoints – there were fewer non-zero timepoints where this protein was observed than required by the setting “Minimum Non-zero points”.

value could not be determined – the optimal value for rate from this fit could not be determined. For more details, you can set “Verbose Output” to “yes” in the settings menu. The relevant exceptions column will contain the text of the fitting error. Usually it is simply there was too much noise in the data to determine the optimal rate value conclusively.

# Changing Settings

Changing settings can be done from the main DeuteRater Window from the “Options” pull down:

Graphical user interface, application

Description automatically generated

There are two options in the “Options” pull down. The first is “ID File Settings” the second is “Rate Calculation Settings”.

“ID File Settings” is only relevant if using the “Create ID File” button with an ID file format other than “Template”. The settings menu looks like this:

Graphical user interface, application, Word

Description automatically generated

The “Create ID File” button will automatically populate multiple charge states to search for if used with PEAKS input data. this menu governs that. The Minimum charge is the lowest charge state and maximum charge is the highest. These two values and every value between them will be put into the ID file. Use “Save Settings” to save changes and “Exit” to exit.

“Rate Calculation Settings” looks like this:

Graphical user interface, application

Description automatically generated

The large words indicate which step the settings below the work apply to. We will go over these by section.

General:

Recognize Available Cores – most steps in DeuteRater involve multiprocessing to improve the speed. If this value is set to “yes” DeuteRater will detect how many cores are present in your computer and use all of them up to a maximum of 60.

Default Cores to Use – if Recognize Available Cores is set to “No”, this setting determines how many cores should be used for multiprocessing. Do not set above 60, this will cause an error in Windows that breaks the multiprocessing.

Study N values – different organisms have different biochemical pathways, the number of deuteriums each amino acid will incorporate (N-value) varies based on experiment. This setting is which of the lines in the resource file “aa\_labeling\_sites.tsv” in the Resources folder describes the N-values for your experiment.

Extract:

MS level – the ms level which should be extracted.

Retention time unit – whether the retention times in your ID file are provided in seconds or minutes

Extraction time window (min) – the amount of error in retention time between a peak to be extracted and the retention time provided by the ID file. This is a one-sided value, so 1.5 indicates a potentially good peak in the .mzml file can be 1.5 minutes faster or 1.5 minutes slow than its retention time in the ID file.

Allowed m/z error (ppm) – is the allowed error in ppm between the m/z of a feature in an .mzml and the m/z of a matching peptide/lipid constituent in the ID file.

Use Chromatography Division –

Fraction New:

Minimum required n values – if there is insufficient label incorporation, the contribution from noise makes accurate measurements difficult. Therefore, if a peptide/lipid constituent has less than this amount of potential deuteriums incorporated from the experimental labeling scheme, the peptide/lipid constituent is discarded.

Minimum Sequence Length – Smaller peptides are more prone to mis identifications and noise contributions. If a peptide sequence is less than this size, it is discarded from further calculations.

Minimum allowed m0 change – As mentioned in the “Rate Calculation – Fraction New Calculation”, if there is not enough change between the unlabeled and maximally labeled theoretical heights of a neutromer, that neutromer is rejected from further use because the potential contribution from noise is too high. If M0 (the monoisotopic peak), which will have the largest difference is, is below this amount, the peptide will be discarded. Value is in relative height, so .1 is a 10% decrease at maximum labeling.

Use Abundance – if you wish to use Abundance calculations, and if so what metric to use. Both “Use Abundance” and “Use Neutromer Spacing” must be used to use Combined metrics. If anything other than “No” is selected other filters, such as standard deviation filter, are still used. “M0” uses only the monoisotopic peak, “Highest” uses the highest neutromer peak only, “Average” uses the average of all neutromer peaks that passed other filters, and if “No” is selected Abundance fraction new and combined fraction new will not be calculated.

Use Neutromer Spacing – this option governs whether Neutromer Spacing calculations are performed. “yes” performs neutromer spacing calculations, “No” does not. Both abundance and neutromer spacing calculations are required to perform combined calculations.

Rate Calculation:

Minimum non-zero points – the minimum number of timepoints needed to fit a rate curve. If there are fewer than this many non-zero timepoints, the protein will not be fit. Zero is excluded because the fitting equation forces time 0 through the origin, so time zero measurements are not useful in fitting the rate curve.

Roll Up Rate Calculation – if there are multiple peptide measurements for the same protein at a given time (different peptides from the same protein, replicates, different charge states) there are two ways to be fit. The default, with this setting at “No”, provides all the multiple measurements to the fitter without any adjustment. If this setting is set to “Yes” the multiple measurements will have outliers determined with a median absolute deviation outlier test and the remaining measurements will be averaged. The average for the measurements for this protein at the relevant time will be provided to fitter and graphed as a single point with error bars as opposed to many points with no error bars.

Asymptote Calculation – whether the asymptote is provided to the fitting equation as a constant (Fixed) or calculated as a variable in the fit equation (Variable).

Fixed Asymptote Value – if Asymptote Calculation is set to “Fixed”, this value will be used as the asymptote.

Proliferation Adjustment – In addition to normal protein maintenance, cell division will also change the turnover rate of proteins. This term allows you to remove that contribution. If the value is not zero it adds a proliferation term to the turnover rate of the proteins:

.

Bias Calculations – there is often bias in the Fraction New calculation values. If desired, a value can be subtracted from the Fraction New to compensate for this. “None” does not apply a bias correction. “Automatic” will calculate a fraction new for all time 0 samples assuming .05 enrichment (this fraction new will always be calculated, but this is the only way it is used). The median for the time 0 fraction new for each calculation type is subtracted from the relevant calculation type (e.g. abundance median time 0 is subtracted from all abundance measurements). “Manual” allows the user to set the bias values using the three settings below.

Abundance Manual Bias – if Bias Calculations is set to “Manual”, this value will be subtracted from all values in the “afn” column in the “frac\_new\_output.tsv” file.

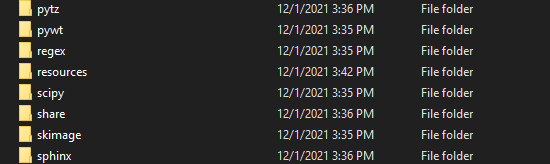
Neutromer Spacing Manual Bias – if Bias Calculations is set to “Manual”, this value will be subtracted from all values in the “nsfn” column in the “frac\_new\_output.tsv” file.

Combined Manual Bias – if Bias Calculations is set to “Manual”, this value will be subtracted from all values in the “cfn” column in the “frac\_new\_output.tsv” file.

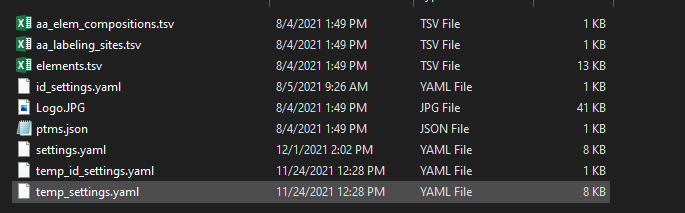
Verbose Output – determines if extra columns are produced in the “calculated\_rates.csv” file from the Rate Calculation step.

# Changing Default Values and Settings not in the Settings Menu

There are some settings which are not present in the settings menu which you may wish to change. You may also wish to change the default values. To do this you must go into the folder called resources (a subfolder of the your exe folder or code folder), shown here for the .exe version:

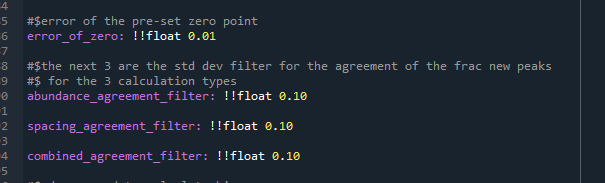


The resources folder contains several files which can be altered. The files we are concerned with are the “id\_settings.yaml” and “settings.yaml”:



Do not concern yourself with the “temp\_id\_settings.yaml” and “temp\_settings.yaml”, these are overwritten and freshly generated each time you run DeuteRater so altering them is not useful.

A .yaml file can be opened in any text editor or interactive coding environment. Since many people are unfamiliar with this file format we will go over it now. Inside the file looks like this:



Anything after a # symbol (gray text in screen shot above) are simply notes that the software will ignore. The actual values are in the format [variable name]: !![variable type] [value]. The variable type and variable name should not be altered unless you are making large code adjustments. The variable type explains to DeuteRater how the value should be read, and any new values should be of the proper type. The variable types are:

float – floating point number. A number that has, or can have, a decimal point.

int – an integer. A number that cannot have a decimal point

str – a text string. The value must be enclosed in quotation marks

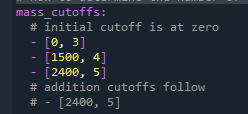
bool – a Boolean value. The only allowed values are true and false.

If you want to change a value, be cautious. Many of the values only make sense within a certain range of values, but DeuteRater will not check if your values make sense. Be cautious. Also make sure the filename, including the .yaml extension, remain unchanged when you save the file.

Because we have gone over the settings menus in the “Change Settings” section, for any settings that are in one of the settings menus we will simply refer to what the setting is called there, and not repeat the explanation. For any settings which are not in one of the settings menus, a new explanation will be provided.

We will start with id\_settings.yaml. like the id settings menu, this only applies to creating the ID file if not using the “Template” option:

Mass cutoffs: determines how many neutromers to extract based on the neutral mass of the peptide. Since this is for the id file, this simply fills a column, it does not change the actual extractor. Since the setting is different from the others, we will use the default as an example:



This indicates 0-1500 mass peptides should use 3 neutromers, 1500-2400 mass peptides should use 4 neutromers, and anything greater should use 5 neutromers.

Rt\_proximity\_tolerance – if two peptides are this close in time and have a similar m/z, they cannot be separated without MSMS data so will not be included in DeuteRater.

Mz\_proximity\_tolerance – if two peptides are this close in m/z and have a similar retention time, they cannot be separated without MSMS data so will not be included in DeuteRater.

Start\_time – the time that the mzml starts

Study\_type – same as “study N values” from the Rate Calculation Settings menu

Aa\_elemental\_composition\_path – the name of the file in the resources folder which has the elemental compositions of amino acids and modified amino acids.

Aa\_labelling\_sites\_path – the name of the file in the resources folder which has the amount of deuterium atoms that can be incorporated into a given amino acid.

Elements\_path – the name of the file in the resources folder which contains information on elements and their isotopes.

Post\_translation\_modifications\_path – the name of the file in the resources folder which contains information on how the PEAKS program indicates post translational modifications and how they should be represented in DeuteRater.

Min\_charge\_states – default for “Minimum Charge” from the Id file settings menu.

Max\_charge\_states – default for “Maximum Charge” from the Id file settings menu.

Remove\_duplicates – if this is true duplicate measurements for the same peptide will be removed as the id file is generated.

Now we will move on to the settings found in settings.yaml:

debug\_level – this setting is used for troubleshooting. A value of 0 is normal operations. A value of 1 puts DeuteRater in debug mode. This mostly consists of deactivating multiprocessing to make error messages more easily visible. Do not use unless altering code or trouble shooting.

recognize\_available\_cores – default for “Recognize available cores” in the Rate Calculation Settings menu.

n\_processors – default for “Default Cores to Use” in the Rate Calculation Settings menu.

Id\_file\_rt\_unit – default for “Retention Time Unit” in the Rate Calculation Settings menu.

trim\_ids\_to\_mzml\_bounds – an option to drop ids too close to the ends of the chromatography run. This is based on “Extraction Time Window (min)” in the Rate Calculation Settings menu. So if the extraction time window is 1.5 minutes, any id within the first 1.5 and last 1.5 minutes of the .mzML being extracted will be dropped.

chunk\_size – when using the extractor, the ids in the id file are grouped together to feed into the multiprocessing step. This variable governs the size of each group.

chunking\_method\_threshold – If an id file is less than "chunking\_method\_threshold" times "max\_chunk\_size" lines long, then split the identification file evenly among the processors

max\_valid\_angle – the extractor extracts individual scans and combines them into a final envelope for a peptide. One way to determine if individual scans should be combined is to create a multidimensional angle between adjacent scans. If the angle is greater than this value, then the scan is excluded from the combination.

peak\_ratio\_denominator – the extractor extracts individual scans and combines them into a final envelope for a peptide. One way to determine if individual scans should be combined is to compare the sizes of the isotope peaks. If M0 for a given scan is less than 1/peak\_ratio\_denominator of largest m0 value in the envelope, the scan is discarded.

time\_window – default for “Extraction Time Window (min)” in the Rate Calculation Settings menu.

ppm\_window – default for “Allowed m/z error (ppm)” in the Rate Calculation Settings menu.

use\_abundance – default for “Use Abundance” in the Rate Calculation Settings menu.

use\_neutromer\_spacing – default for “Use Neutromer Spacing” in the Rate Calculation Settings menu.

minimum\_nonzero\_points – default for “Minimum Non-zero points” in the Rate Calculation Settings menu.

Peak\_lookback – number of peaks neutromer peaks in front of a peptide id to extract for troubleshooting purposes.

Peak\_lookahead – number of peaks neutromer peaks beyond the last extracted isotope of a peptide id to extract for troubleshooting purposes.

baseline\_lookback – number of indexes in an .mzml to look at to determine baseline noise around a given ID.

min\_envelopes\_to\_combine – the minimum number of scans that must be in one envelope in order to be a valid extracted envelope. If there are fewer scans, nothing is extracted for that id.

zscore\_cutoff – the zscore used as a filtering value for removing m/z outlier values when combining scans in the extractor.

rt\_proximity\_tolerance – when combining extracted files in the “Theory Generation” step of Rate Calculation, different peptides which have the same m/z and same retention time are discarded since we can’t tell the difference between them without fragmentation. This setting determines how close retention times must be to be considered the same.

mz\_proximity\_tolerance – when combining extracted files in the “Theory Generation” step of Rate Calculation, different peptides which have the same m/z and same retention time are discarded since we can’t tell the difference between them without fragmentation. This setting determines how close m/z values must be to be considered the same.

peptide\_analyte\_id\_column – the name of the column which has a unique identifier for your protein.

peptide\_analyte\_name\_column – the name of the column which has the common name of your proteins.

Lipid\_analyte\_id\_column – the name of the column which has a unique identifier for your lipid.

Lipid\_analyte\_name\_column – the name of the column which has the common name of your lipids.

study\_type – default value for “study N values” in the Rate Calculation Settings menu.

aa\_labeling\_sites\_path – the name of the file in the resources folder which has the amount of deuterium atoms that can be incorporated into a given amino acid.

unique\_sequence\_column – the name of the column containing the amino acid sequence of your peptides.

roll\_up\_rate\_calc – default value for “Roll up Rate Calculation” in the Rate Calculation Settings menu.

Asymptote – default value for “Asymptote Calculation” in the Rate Calculation Settings menu.

proliferation\_adjustment – default value for “Proliferation Adjustment” in the Rate Calculation Settings menu.

fixed\_asymptote\_value – default value for “Fixed Asymptote Value” in the Rate Calculation Settings menu.

error\_of\_zero – if using “Roll up Rate Calculation”, an error value is provided for each measurement. 0 is provided to the fitter, so this provides the error metric.

abundance\_agreement\_filter – in the “Fraction New Calculation” step, the standard deviation of all neutromer measurements for a given peptide using a given calculation method is calculated. If the standard deviation is above a filtering threshold, the point is considered inconsistent (we are wrong with the id, the theoretical unlabeled or labeled are wrong, or the data is noisy) and so the relevant fraction new value is dropped from consideration. This setting is the filtering cutoff for abundance calculations.

spacing\_agreement\_filter – in the “Fraction New Calculation” step, the standard deviation of all neutromer measurements for a given peptide using a given calculation method is calculated. If the standard deviation is above a filtering threshold, the point is considered inconsistent (we are wrong with the id, the theoretical unlabeled or labeled are wrong, or the data is noisy) and so the relevant fraction new value is dropped from consideration. This setting is the filtering cutoff for neutromer spacing calculations.

combined\_agreement\_filter – in the “Fraction New Calculation” step, the standard deviation of all neutromer measurements for a given peptide using a given calculation method is calculated. If the standard deviation is above a filtering threshold, the point is considered inconsistent (we are wrong with the id, the theoretical unlabeled or labeled are wrong, or the data is noisy) and so the relevant fraction new value is dropped from consideration. This setting is the filtering cutoff for combined calculations.

bias\_calculation – default value for “Bias Calculation” in the Rate Calculation Settings menu.

abundance\_manual\_bias – default value for “Abundance Manual Bias” in the Rate Calculation Settings menu.

spacing\_manual\_bias – default value for “Neutromer Spacing Manual Bias” in the Rate Calculation Settings menu.

combined\_manual\_bias – default value for “Combined Manual Bias” in the Rate Calculation Settings menu.

y\_intercept\_of\_fit – sets the fraction new at time 0 for the fitting graphs.

error\_of\_non\_replicated\_point – if using “Roll up Rate Calculation”, an error value is provided for each measurement. If only a single point is present at a given time point, a standard deviation cannot be calculated. So is used as the error instead of the standard deviation.

enrichement\_of\_zero – if time 0 is present, the “Fraction New Calculation” is still calculated with the idea that the result should be 0. This allows the time 0 data to provide a troubleshooting metric (and is the basis for the “Automatic” setting for bias calculation). Since the deuterium enrichment at 0 is 0, this setting provides an imaginary enrichment for use in the calculation.

min\_allowed\_abund\_max\_delta – default value for “Minimum allowed M0 change” in the Rate Calculation Settings menu.

min\_aa\_sequence\_length – default value for “Minimum Sequence Length” in the Rate Calculation Settings menu.

min\_allowed\_n\_values – default value for “Minimum required n values” in the Rate Calculation Settings menu.

minimum\_abund\_change – When performing abundance fraction new calculations, there must be some distance between the unlabeled and maximally labeled theoretical heights of a given neutromer. If this distance is small, noise can drastically impact the fraction new, potentially causing enough difference between the fraction new of the different neutromers to drop an otherwise good point. To avoid this problem we require that the distance between the theoretical heights of each isotope is at least equal to the is value, or the isotope in question is not used in any calculations.

verbose\_rate – default value for “Verbose Output” in the Rate Calculation Settings menu.

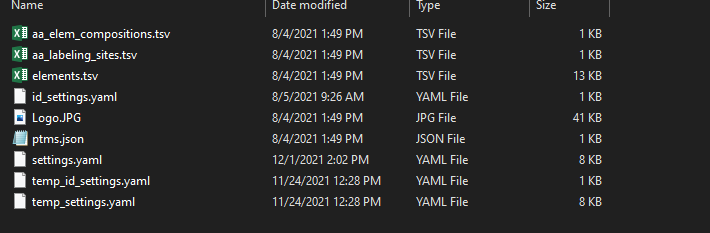
intensity\_filter – A filter used in many areas for scan combining in the extractor. If a neutromer peak is below the intensity filter, it is not considered to gaussian fitting and is not considered a valid scan.

ms\_level – default value for “MS level” in the Rate Calculation Settings menu.

rate\_output\_format – default value for “Graph Save File Type” in the Rate Calculation Settings Menu

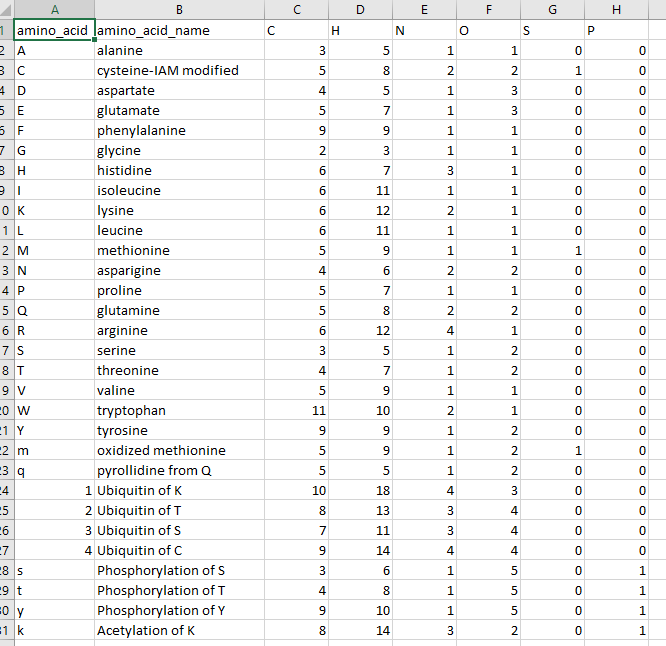
# Adjusting Resource Files

Aside from the .yaml settings file discussed in the above section, there are other files in the “Resources” folder (a subfolder of the your exe folder or code folder). We will be discussing “aa\_elem\_compositions.tsv”, “aa\_labeling\_sites.tsv”, “elements.tsv”, and “ptms.json”:



## Adjusting amino acid elements and adding post translational modifications

For this section we are concerned with “aa\_elem\_compositions.tsv” which looks like this:



Each amino acid or modified amino acid should have a unique one-character symbol in the amino\_acid column. This is case sensitive, and is not restricted to letters as shown in the table image above. Amounts of elements correspond to the chemical formula of an amino acid IN A PEPTIDE, with a peptide bond on each side.

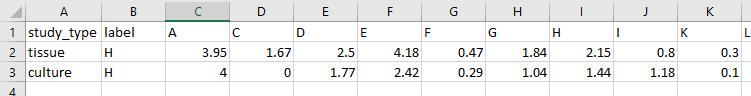
To modify an amino acid, use the columns to the right of amino\_acid\_name. change the number in each column to correspond to the number of the relevant element in the amino acid.

To add an element, make a new column. The first row should contain the element symbol as seen on the periodic table, and the amount of this element in each amino acid. Be cautious while doing this since many elements are not supported for current theoretical calculations. Supported atoms are H, C, N, O P, S, F, D (for deuterium that is always deuterium, such as a SILAC label), Cl, Br, I, Si, and Na.

To add a new amino acid or modification, fill out a new row at the bottom of the table. Provide a unique one letter code, a name, and fill out the appropriate amount of each element. Finally, adjust the “aa\_labeling\_sites.tsv” file as we discuss below.

## Adjusting N values of amino acids

Based on biochemical pathways, amino acids will be able to incorporate different amounts of Deuterium. This needs to be provided to DeuteRater. The file that holds this information is “aa\_labeling\_sites.tsv” which looks like this:



Study\_type is a keyword so you can use the appropriate values. Which study type you are using is changed with the “study N values” setting in the settings menu. In this file, “tissue” is mammalian tissue. “culture” is mammalian cell culture. Label is the element that is labeled. For now, the only label supported is H. Every other column is an amino acid. Values are a population average, which is why they are not integer values in most cases.

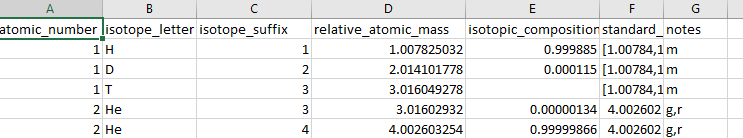
To modify values, simply type in the new value into the appropriate cell and save the file.

If you have added a new amino acid, as described in the previous section, add the one-character code for the new amino acid and fill in the relevant values. Remember that any modification done during experimental preparation will not add deuterium unless you are modifying with a labeled compound. Modifications done by the organism may add more deuteriums.

To add a new study\_type fill in a new row with the study\_type, the heavy label, and values for each amino acid and modified amino acid column.

## Adjusting chemical formulas for PEAKS ID files

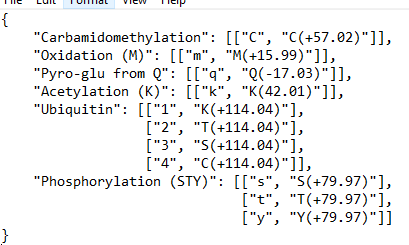
The file elements.tsv is used by the “Create ID File” button if not using a template. It looks like this:



Each isotope has its own line. There is usually no reason to alter this file.

## Adjusting post translational modifications from PEAKS

When using “Create ID File” button if not using a template, DeuteRater will change peaks post translational modification labels into the single character symbols from “aa\_elem\_compositions.tsv”. This requires a key, which is ptms.json. A .json file can be opened in a text editor or interactive coding environment. Ptms.json looks like this:



The format is “[modification name in peaks]”: [[DeuteRater Mod symbol, peaks mod symbol]]. Entries are comma delimited. If a modification name has multiple modifications, such as ubiquitin above, use multiple entries, as shown above.

# Citations

1. Naylor BC, Porter MT, Wilson E, Herring A, Lofthouse S, Hannemann A, et al. DeuteRater: a tool for quantifying peptide isotope precision and kinetic proteomics. *Bioinformatics.* 2017;33(10):1514-20.

2. Kessner D, Chambers M, Burke R, Agus D, and Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics.* 2008;24(21):2534-6.