Contents

[Getting Started 2](#_Toc102128942)

[Description of Example Data 2](#_Toc102128943)

[Create a new variable modification for 18O labeled methionine oxidation in MaxQuant. 2](#_Toc102128944)

[Running MaxQuant on a MObB prepared sample. 6](#_Toc102128945)

[Converting RAW files to mzXML files 9](#_Toc102128946)

[Running the MObB algorithm 12](#_Toc102128947)

[Input data 12](#_Toc102128948)

[Output data 13](#_Toc102128949)

[Practical Steps for running the MObB algorithm. 13](#_Toc102128950)

[Visualizing MObB quantitation models and results. 14](#_Toc102128951)

[MObB post-analysis 16](#_Toc102128952)

# Getting Started

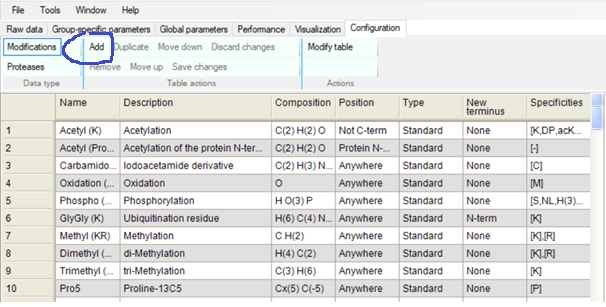
## Description of Example Data

The example data included as part of this tutorial (URL XXXXXXX) was generated as part of the study published in XXXXXXXXXXX. The data describes a brain sample from a young (6 m.o.) wildtype mouse with some mixture of light (16O) and heavy (18O) labeled methionine containing peptides. All 8 files from the example data describe the perfectionated samples of one mouse and are therefore considered one experiment. The sample was prepared in such a way that most methionines should be between 10-17% light (16O) labeled peptide (FractionOxidized between 0.10-0.17), after using a simple quality cutoff of Fit >= 0.8.

This tutorial describes a complete analysis pipeline starting from RAW files, mzXML files and a MaxQuant evidence file are also supplied. Users starting from the supplied evidence and mzXML files may skip to page 11.

## Create a new variable modification for 18O labeled methionine oxidation in MaxQuant.

1. Navigate to the configuration tab on the MaxQuant GUI and select modifications. Click on Add.



1. The following window should open. Next to composition click on Change.

Graphical user interface, text, application

Description automatically generated

1. Set the Hydrogens and Oxygens to zero.

Graphical user interface, application

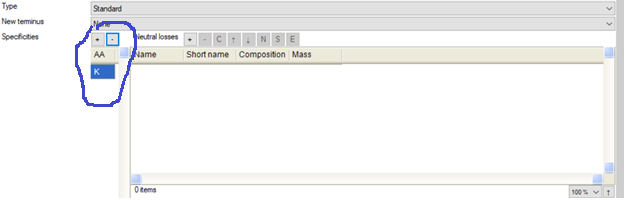
Description automatically generated

1. Set the Ox modification to 1 and click OK.

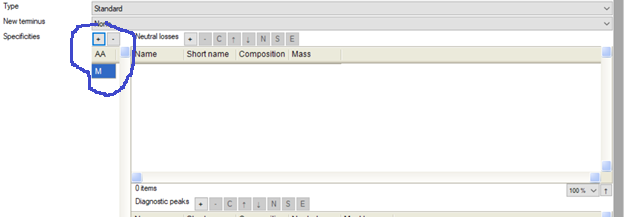
Graphical user interface, application

Description automatically generated

1. In the Specificities tab click the minus sign to remove K (lysine).



1. Click on the plus sign and add methionine (M).



1. Change the variable name to O18 (or anything else you’d like).

Graphical user interface, text, application, email

Description automatically generated

1. Click on Modify table from the original MaxQuant GUI window.

Graphical user interface, application

Description automatically generated

1. Click on Save changes.

Graphical user interface, application

Description automatically generated

## Running MaxQuant on a MObB prepared sample.

You may run MaxQuant on MObB prepared samples using any strategy or set of parameters that you want. However, the 18O modification (discussed above) must be set as a variable modification. In order to maximize coverage, it is recommended that match between runs be used. However, if the study design is large, users may want to consider running match between runs in blocks. Below are the steps for running MaxQuant on the example data provided.

1.) Download example data from (URL: XXXXXXXXXX) and unzip the files.

2.) Open the MaxQuant GUI and in the Raw Data tab select load, then navigate the .RAW files of example data downloaded from (URL: XXXXXXX).

Graphical user interface, application, Word

Description automatically generated

3.) In the Edit experimental design tab, click on Set Fractions and set the fractions equal to 1-8. Double check to make sure the raw file names are associated with the correct fraction by looking at the table provided in *Description of Example Data* on page X.

Graphical user interface, table

Description automatically generated

Table

Description automatically generated

4.) In the Group-specific parameters tab select Modifications and add the following fixed and variable modifications. The modification named O18 is the name of the custom variable for 18O labeled methionine oxidation that was created on pages X-X. Oxidation (M) refers to 16O labeled methionine oxidation and is set by default in most versions of MaxQuant. The carbamidomethyl modification refers the alkylation of cysteine residues by IAA and should be omitted if cysteine reduction and capping was not part of the sample preparation. (Note: MObB analysis filters any cysteine containing peptides and the choice of whether or not to include any cysteine specific medication like carbamidomethyl is somewhat arbitrary.)

Graphical user interface, text, application

Description automatically generated

5.) Navigate the Digestion tab and select the appropriate enzyme used for sample preparation. For the example data provided, set the enzyme to Trypsin (not the default Trypsin/P).

Graphical user interface, text, application

Description automatically generated

6.) In the Global parameters tab select Sequences. Next to the Fasta files section click on Add. Navigate to the file path for the appropriate FASTA file. If using the example data navigate to the mouse FASTA file downloaded from (URL: XXXXXXX).

Graphical user interface, application, Word

Description automatically generated

7.) Next to the Fasta files section, click on Identifier and select the appropriate identifier rule. If using the example data provided, select Uniprot Identifier.

Graphical user interface, text, application, email

Description automatically generated

[Optional: If you are using match between runs you must select that option in the Global Parameters tab under the Identification section. The example data provided does not have a paired dataset to use match between runs on, so this step is skipped.]

8.) Click Start at the bottom of the MaxQuant GUI. It may take several hours for MaxQuant to complete the analysis.

## Converting RAW files to mzXML files

As input data MObB requires raw spectra in the form of mzXML files. RAW files that are the output of LC-MS/MS experiments can be converted to mzXML format using freely available software.

1.) Download and install proteowizard from <https://proteowizard.sourceforge.io/download.html>. Follow the publishers’ instructions for installation. **It must be the 64-bit version.**

2.) Open the MSconvert app. In the filters section navigate to peak picking and select the vendor supplied algorithm.

Graphical user interface, application

Description automatically generated

3.) Set MS-Levels to 1-1 and click add.

Graphical user interface, application

Description automatically generated

4.) In the filters section navigate to subset and set MS levels to 1-1, then click Add.

Graphical user interface, application

Description automatically generated

5.) In the options panel set the following options.

Graphical user interface, text, application

Description automatically generated6.) In the file section, click browse and add all RAW files you wish to convert, then click Start.

(Note: By default, the output mzXML files will be placed in the same directory and under the same name as the RAW files. **mxXML files must have identical names to the RAW files.**)

# Running the MObB algorithm

The MObB algorithm is a custom algorithm designed to analyze convoluted pairs of 16O/18O labeled methionine sulfoxide containing peptides. A complete description of the algorithm and an evaluation of its performance can be found in refX.

The tutorial presented here is meant to be a practical guide to the use of the MObB algorithm and any further questions about the rationale or use of MObB can be directed to John Bettinger ([jqb3644@gmail.com](mailto:jqb3644@gmail.com)). This section first describes the input and output of MObB, followed by a practical step-by-step guide on how to use the MObB script. The final part of this section describes tools for visualizing the quantitative models generated and used by MObB.

## Input data

The MobB algorithm designed to accept MaxQuant evidence files and mzXML files as input. The process for converting RAW files into mzXML files is discussed above. If users wish to use any to supply a custom input dataframe, in lieu of the MaxQuant evidence file, they may do so with the following considerations. The MObB input dataframe must be a tab separated text file where each row is an identified MS1 feature. MObB input data must include the following data columns:

**Sequence** – a string of single letter amino acid codes that describes the linear sequence of the unmodified peptide.

**Raw.file** – the name of the RAW/mzXML file corresponding to the identified MS1 feature.

**Proteins** – a semicolon separated list of all proteins for which the peptide can be mapped to. (If a razor protein has been identified for most peptides this column may contain a single protein entry.)

**Charge** - an integer value specifying the charge state of MS1 feature.

**O18.M** – an integer value specifying the number of 18O labeled methionines on the identified peptide

**O16.M** – an integer value specifying the number of 16O labeled methionines on the identified peptide. (Note: this column is actually named “Oxidation..M.” in the example data.)

**m.z** – the mass to charge ratio of the monoisotopic peptide. (Whether the light or heavy labeled peptide is considered monoisotopic will depend on whether the peptide was identified as being light (O16.M.) or heavy labeled (O18.M.).)

**Retention.time.start** – the estimated start of the retention time for the MS1 feature. The units (seconds vs minutes) of this column must agree with the units of retention time in the raw file. (Note: this column is actually named “Calibrated.retention.time.start” in the example data.)

**Retention.time.finish** – the estimated end of the retention time for the MS1 feature. The units (seconds vs minutes) of this column must agree with the units of retention time in the raw file. (Note: this column is actually named “Calibrated.retention.time.finish” in the example data.)

**Retention.time.calibration** – a correction factor applied to the retention time start and finish of each MS1 feature. This column is optional and only used if a retention time calibration was used as part of the RAW data search, as is the case when using match-between runs in MaxQuant.

**Missed.cleavages** – an integer value specifying the number of missed cleavages identified in the MS1 feature.

## Output data

The MObB analysis generates a simple output, formatted as a .csv file. Each row in the MObB output is a unique methionine containing peptide sequence. (Note: as part of its analysis the MObB algorithm will coalesces MS1 features representing the same peptide into a single analysis.) MObB outputs have the following columns:

An unnamed column that is the row index number. (This column can be removed for most analyses.

**Sequence** – The linear sequence of unmodified peptides. Each row is a unique unmodified peptide sequence.

**FractionOxidized** – The estimated ratio of 16O labeled peptide intensities to total peptide intensities.

**Fit** – A R-squared value for the linear regression used to estimate FractionOxidized. The values of this column represent quality scores and the recommended quality cutoff is Fit >= 0.80.

**Totalint** – the total (summed) intensity for each peptide sequence.

MObB includes two additional output arrays that are the feature\_array and the model\_array, which contain information about raw MS1 features pulled from the mzXML files and the quantitation models used to estimate FractionOxidized. Feature\_array and model\_array can be used to visualize the performance of MObB on each individual MS1 feature (discussed below).

## Practical Steps for running the MObB algorithm.

1. Download the MObB core analysis script from (URL XXXXXXXXX)
2. Open the script in your python IDE of choice. MObB was developed using Spyder and this tutorial will follow its usage in Spyder. **The IDE and python version being used must be 64-bit.**
3. A list required python modules is included on line 1-32. Ensure that you have all the required modules properly installed. Install any missing modules following publishers’ instructions.

Note: lines 1-607 are the protocols for creating MObB defined functions. It is recommended that only advanced users make edits to lines 1-607. For convenience, a description of the input and output of each MObB defined function is included as annotations in the script.

1. On line 622 enter the file path to your input dataframe. If you are using the example data, this corresponds to the file named exampledata\_evidence.txt

Note: Depending on the version of MaxQuant you are using to generate your input dataframe, or if you are making your own input dataframe then you may have to adjust the column names defined by the variable ‘fields’. A list of required columns is given above under Input Data.

1. Make sure your mzXML files are all in the same folder and have names that correspond to the filenames given in the Raw.file column of your input dataframe
2. On line 671 enter the file path to the folder that contains your mzXML files.
3. On line 653 enter your chosen value (in ppm units) for tolerance in the m/z error of MS1 datapoints. The default and recommended value is 5.
4. On line 911 enter the desired file path to your output dataframe. The output dataframe is described above in Output Data.
5. Optional: On line 914 enter the desired file path to your quantitation models. Quantitation models can be stored for future reanalysis or visualization. Each MS1 feature is a model.
6. Optional: On line 918 enter the desired file path to your quantitation models. Quantitation models can be stored for future reanalysis or visualization. Each MS1 feature is a model.
7. Start the script and wait!

Note: It is not uncommon for a significant portion of all MS1 feature could be fit to a model.

Note: Depending on how large of a search was initiated, it may take over 24 hours for the analysis to complete. The script will continuously print the count of the peptide number it is processing, if your count appears frozen, try restarting the kernel.

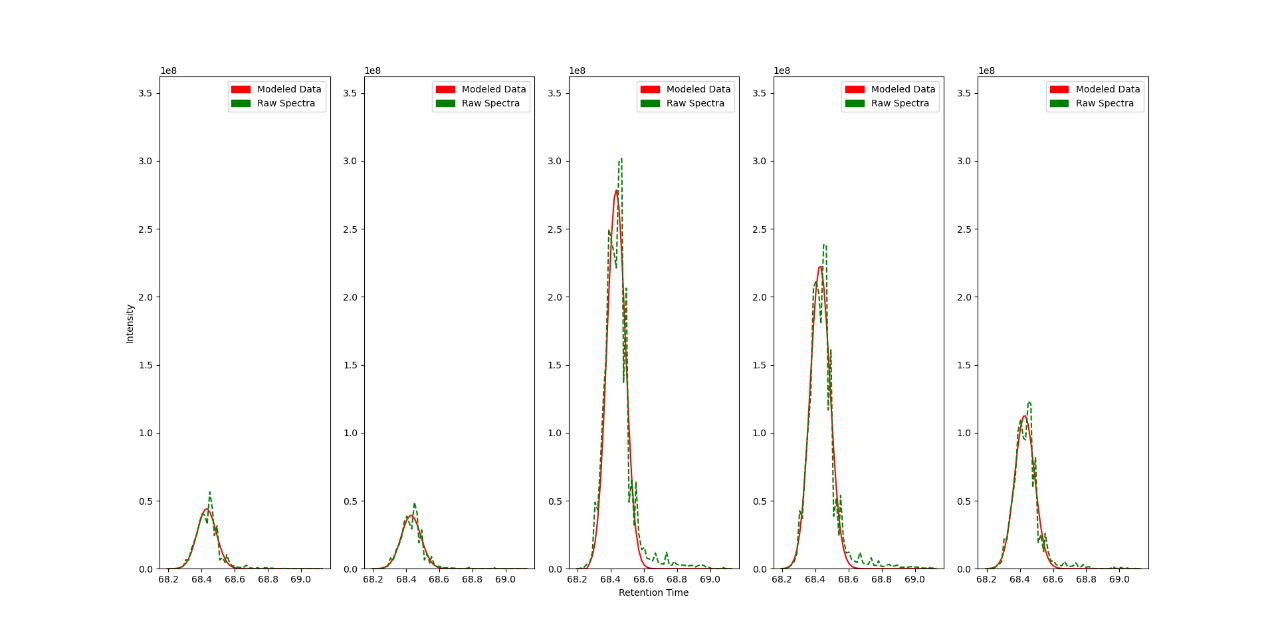
## Visualizing MObB quantitation models and results.

MobB includes a pair of tools for inspecting each MS1 feature used to estimate FractionOxidized. MObB functions by fitting an elution profile (quantitation model) to raw data extracted from the raw spectras (mzXML files).

Raw spectra overlayed with the fitted elution/quantitation model can be visualized using the viewMS1feature() command. (Note: not every MS1 feature initialized by the input dataframe will have a model associated with it. MS1 features that could not be fit to an elution model do not contribute to the final quantitation.)

The input requires a raw MS1 feature from the variable called ‘featarray’ and a corresponding elution model from the variable called models.

For example, if we want to visualize a random MS1 feature from out example data we can use the following command.

Which outputs the following figure:

Each graph is the elution profile (XIC) of an isotopologue belonging to a single MS1 feature. Graphs are ordered by mass with the lightest mass on the left and heaviest on the right (The leftmost peak is the monoisotopic light labeled mass peptide and the 3rd or middle peak is the monoisotopic heavy labeled peptide mass). Raw spectra are shown as a green dotted line and the elution/quantitation model is shown as a red solid line.

# MObB post-analysis

MObB was originally designed as a proteomic workflow to allow for the estimation of endogenous methionine oxidation, even at low levels. As part of this strategy, a prototypical MObB experiment involves measuring methionine oxidation content in a series of isotope spike-in experiments as described in *Bettinger JQ, et. al. Quantitative Analysis of in Vivo Methionine Oxidation of the Human Proteome. J Proteome Res. 2020* and *Bettinger, J. Q., et.al. Accurate proteome-wide measurement of methionine oxidation in aging mouse brains. bioRxiv 2022*.

In order to keep the utility of MObB experiments flexible to meet researchers needs, the MObB core analysis (which measures the ratio between 16O and 18O labelled methionine containing peptide pairs) and the post-analysis (which estimates endogenous levels of methionine oxidation via spike-in responses) are kept separate.

For reference and clarity an annotated example script that allows for the estimation of endogenous oxidation using MObB data generated as part of *Bettinger, J. Q., et.al. Accurate proteome-wide measurement of methionine oxidation in aging mouse brains. bioRxiv 2022* is included and can be downloaded from (URL:XXXXX). Example data that can be used to guide you through the post analysis process can be downloaded from (<URL:XXXXX>). A complete description of the example data can be found in *Bettinger, J. Q., et.al. Accurate proteome-wide measurement of methionine oxidation in aging mouse brains. bioRxiv 2022*.

Ultimately, each experimental design is unique and it is the responsibility of each researcher to decide what the most appropriate post analysis pipe-line is. However the general steps reccomended for a typical MObB analysis aimed at estimating endogeneous oxidation are described below.

1. Desing a MObB experiment that inlcudes a minimum of two isotope spike-in treatments.
2. Run the MObB core algorithm on each prepared sample
3. Reformat the MObB output data into a new concatenated dataframe that reflects your experimental design.
4. Filter for data quality. The recccomended quality cutoff for MObB output is a Fit <= 0.8. See above for a description of the Fit parameter.
5. **Impute missing values.**

Note: The imputation of missing values is optional and the choice of imputation methods will depend strongly on the experimental design. Generally, imputation methods that measure a correlation structure between samples take advantage of a typical MObB experiment’s functional design (there is explicit correlation between titration points in a MObB experiment.)

1. **Normailize the data.** A typical MObB experiment measure the global mean methionine oxidation value for each of your samples and then creates a global reference response curve that can be used to calculate normalization factors.
2. **Measure peptide specific responses and extrapolate an endogenous oxidation value.**

Note: The equation used to fit isotope spike-in responses is the standard equation for fractional titrations:

where MOS*i* is the global median of measured 16O/(16O+18O) ratios in titration sample *i*, MOS*in vivo* is the global median of estimated MOS values that would be measured without any spike-in and ti is the relative ratio of the spike-in proteome used to create sample *i*.

1. **Quality filter.** The reccomended quality filter is for the fir to a theoretical isotope spike-in response to have an NRMSE of 0.2 or less.
2. **Define your hypothesis and choose the most appropriate statistical test.** MObB will return an estimate for endogenous oxidation as well as residual and standard errors, which can be used to calculate test statistics.