Contents

1. Introduction

Objective: Assess the level of contamination of LC-MS results using the HowDirty workflow.

Input LC-MS raw files + Target List from publication (see instructions where to find it)

Output: HTML report with plots evaluating the normalized abundance of contaminants in the samples. Excel file compiling the results and summary statistics.

Summary: The presence of contaminants (e.g., PEG) and detergents (e.g., CHAPS, SDS) in samples analyzed by LC-MS can be severely detrimental to identifying peptides/proteins or other molecules. Skyline is used to extract MS1 features of many known contaminant masses from raw files (e.g., .raw and .d). The results are exported to a .csv file, then processed in R using HowDirty to generate an HTML interactive report that evaluates sample contamination risks.

2. Requirements

- Raw LC-MS results to be evaluated
- Skyline version > 4 [1,2].
  To install it, you can register online and download the latest version here: https://skyline.ms/project/home/software/Skyline/begin.view.
- Skyline HowDirty template, including the Skyline molecular contaminant transition list [3] and reports configuration: https://github.com/DavidGZ1/HowDirty/tree/main/tutorial
- Alternatively, you can set up Skyline yourself (further instructions in the Skyline tutorials)
  - Download the original template with the molecular contaminant transition list [3] from Panorama and load it into Skyline: ***File /Open***
  - Create the PeakAreas_Contaminants report: containing the columns:
  - ***Settings / Document Settings / Reports / Add***, then enter the name "PeakAreas_Contaminants" and select the columns (Tip: click on the binoculars symbol to access the search):
    "Proteins", "Peptides", "Replicate Name", "Peptide Retention Time", "Total Area MS1", "Total Ion Current Area"
  - Enable the report form by ticking the box next to its name, then click ***OK***

  Note: the column names in csv file exported by Skyline (step 8), may differ slightly (Protein instead of Proteins, Peptide instead of Peptide) or spaces may be removed. HowDirty should be able to handle those differences.

- R software for data analysis and the R packages Rmarkdown, knitr, and HowDirty can be installed by running the following code in R using a freshly open console window (e.g., in R Studio):

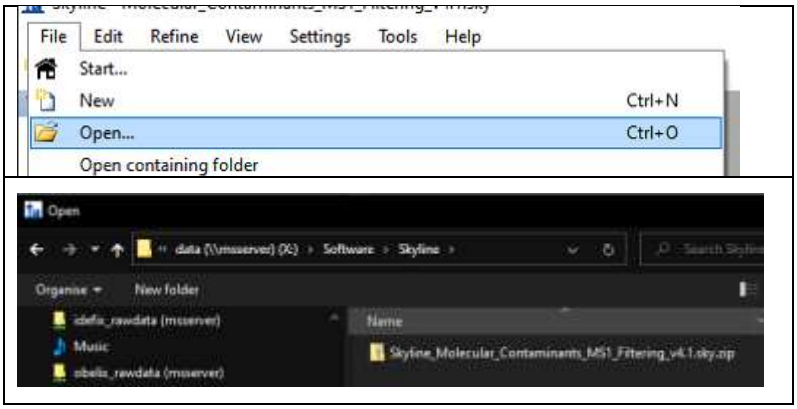
```
if(!require(devtools)) install.packages("devtools")
devtools::install_github("kassambara/ggpubr")
devtools::install_github("DavidGZ1/HowDirty", force = TRUE)
```

3. Procedure

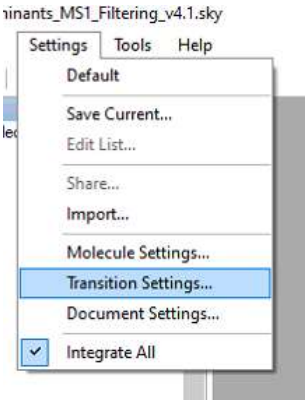
3.1. MS1 feature extraction in Skyline

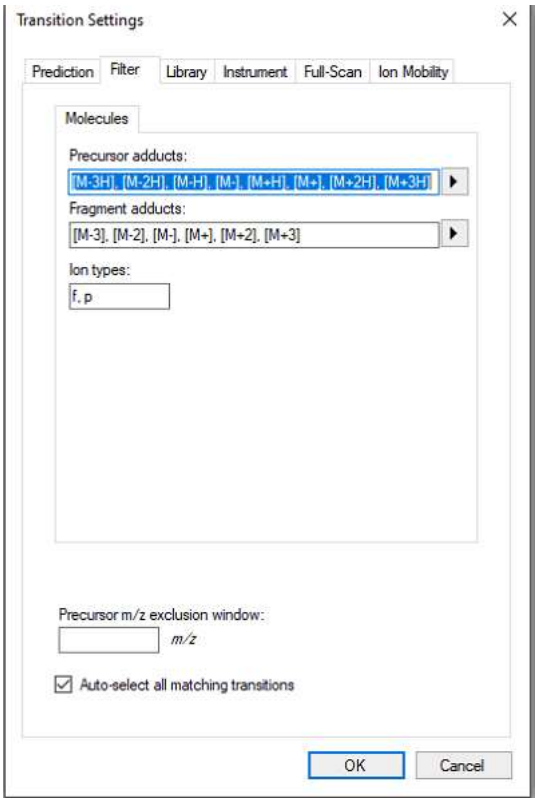
1. Open Skyline
2. Open the HowDirty template: ***File / Open*** navigate to the file location and open the .sky file

   A Target List of commonly known contaminant masses will appear in Skyline.



3. Save the Skyline's file in the desired location
4. Navigate to ***Settings / Transition Settings / Filter*** and ***/ Full-Scan*** and set the parameters accordingly to the experiment. An example for Exploris 480 data is shown below.





Filter settings must include "p" (precursor) in Ion Types.

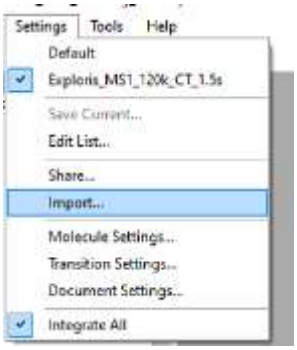
Instrument, Full-Scan, and Ion Mobility depend on the instrument and method used

Resolving power may depend on method-specific parameters. For instance, Exploris 480 can be acquired with an MS1 resolving power of 60,000 at 200 m/z. and MS2 15,000 at 200 m/z.

For TIMS-ToF data, choose “TOF” as the precursor mass analyzer and resolving power 38,000 in both corresponding fields.

Important: in retention time filtering, select “include all matching scans”

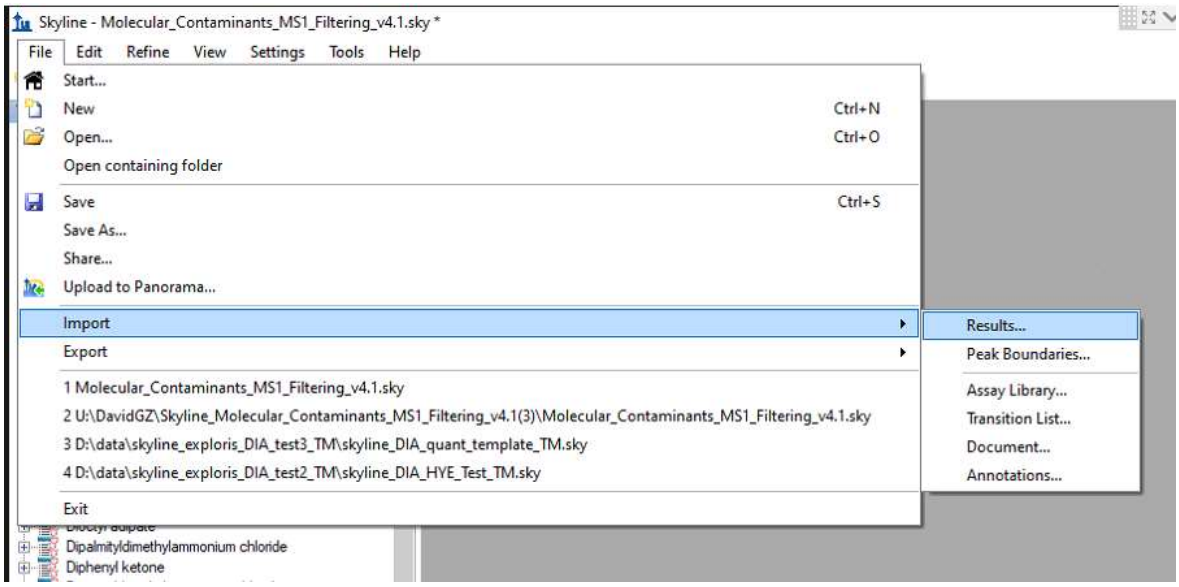
5. Optional: Save those settings so that they can be reused by importing.



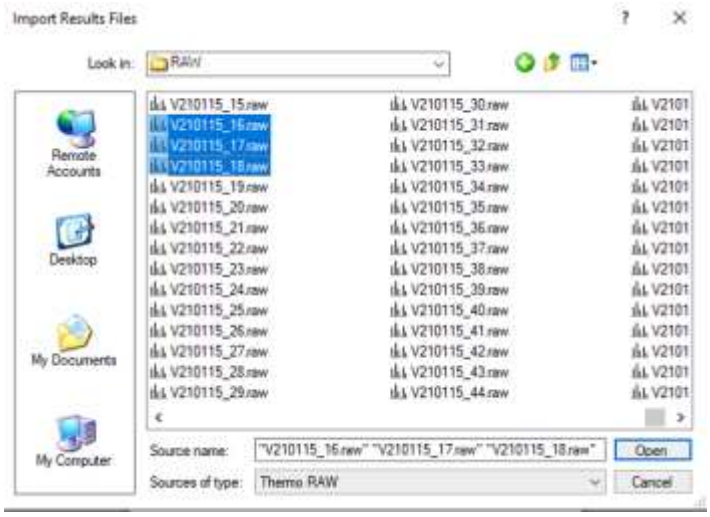
Use *Settings /Share* to export.
Use *Settings/Import…* to import.

6. Import the raw data (screenshots on next page)
    a. Select ***File / Import / Results…***
    b. Select the raw files and click ***Open***
    c. Select “add single-injection replicates in files” and click ***OK***, then add file name pattern to remove if you want to shorten file names and select ***Remove***, or select ***Do not remove / OK*** if not

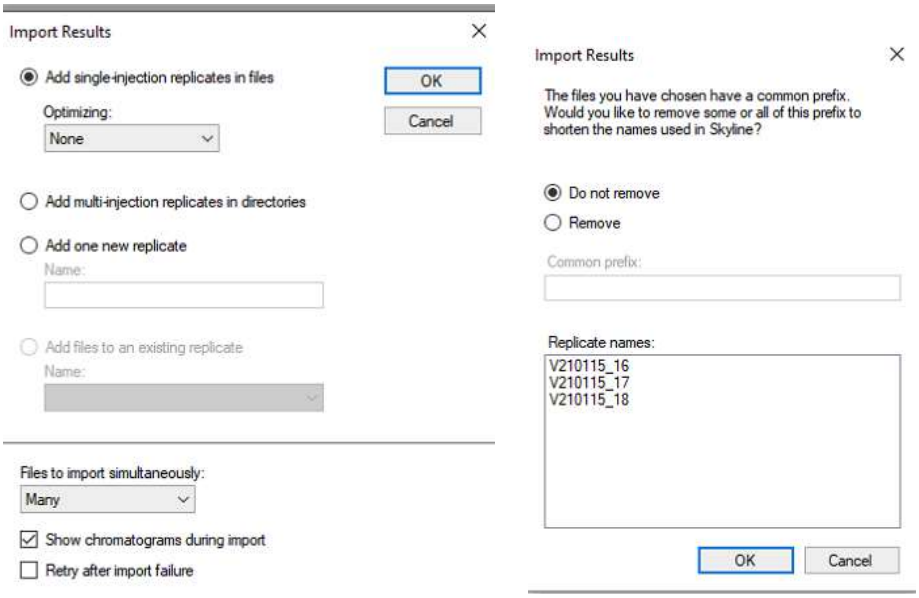
a


Skyline - Molecular_Contaminants_MS1_Filtering_v4.1.sky *
File Edit Refine View Settings Tools Help
Start...
New Ctrl+N
Open... Ctrl+O
Open containing folder
Save Ctrl+S
Save As...
Share...
Upload to Panorama...
Import
Export
1 Molecular_Contaminants_MS1_Filtering_v4.1.sky
2 U:\DavidGZ\Skyline_Molecular_Contaminants_MS1_Filtering_v4.1(3)\Molecular_Contaminants_MS1_Filtering_v4.1.sky
3 D:\data\skyline_exploris_DIA_test3_TM\skyline_DIA_quant_template_TM.sky
4 D:\data\skyline_exploris_DIA_test2_TM\skyline_DIA_HYE_Test_TM.sky
Exit
Results...
Peak Boundaries...
Assay Library...
Transition List...
Document...
Annotations...
Dipalmitoyldimethylammonium chloride
Diphenyl ketone

b

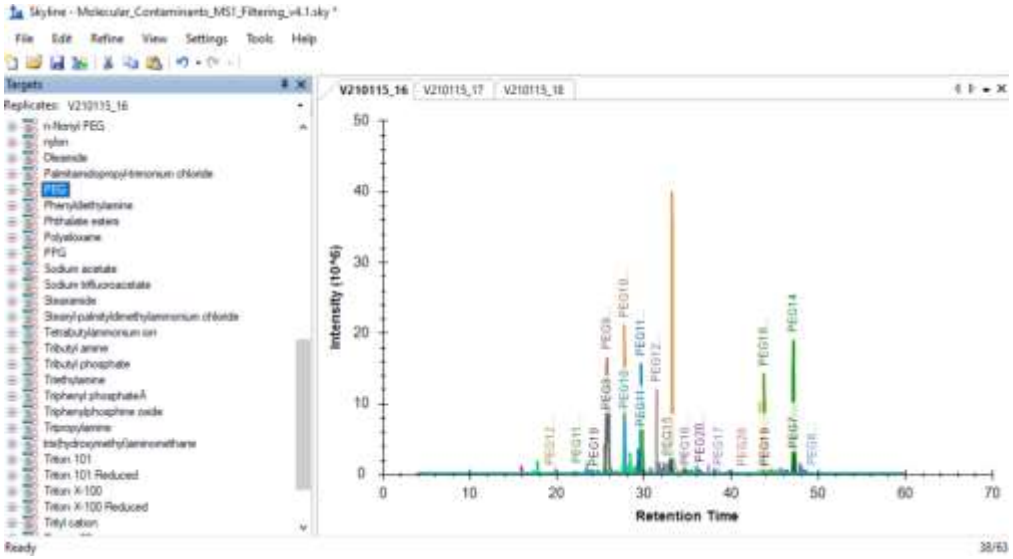

Import Results Files
Look in: RAW
Remote Accounts
Desktop
My Documents
My Computer
V210115_15.raw
V210115_16.raw
V210115_17.raw
V210115_18.raw
V210115_19.raw
V210115_20.raw
V210115_21.raw
V210115_22.raw
V210115_23.raw
V210115_24.raw
V210115_25.raw
V210115_26.raw
V210115_27.raw
V210115_28.raw
V210115_29.raw
V210115_30.raw
V210115_31.raw
V210115_32.raw
V210115_33.raw
V210115_34.raw
V210115_35.raw
V210115_36.raw
V210115_37.raw
V210115_38.raw
V210115_39.raw
V210115_40.raw
V210115_41.raw
V210115_42.raw
V210115_43.raw
V210115_44.raw
Source name: "V210115_16.raw" "V210115_17.raw" "V210115_18.raw"
Open
Sources of type: Thermo RAW
Cancel

c

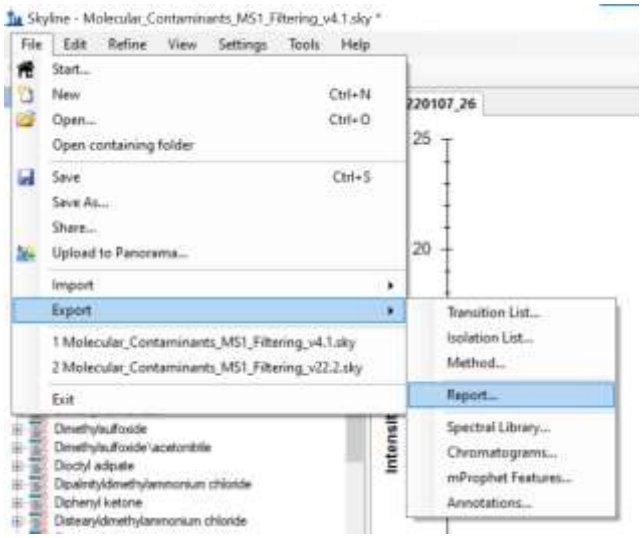

Import Results
Add single-injection replicates in files
Optimizing:
None
OK
Cancel
Add multi-injection replicates in directories
Add one new replicate
Name:
Add files to an existing replicate
Name:
Files to import simultaneously:
Many
Show chromatograms during import
Retry after import failure
Import Results
The files you have chosen have a common prefix. Would you like to remove some or all of this prefix to shorten the names used in Skyline?
Do not remove
Remove
Common prefix:
Replicate names:
V210115_16
V210115_17
V210115_18
OK
Cancel

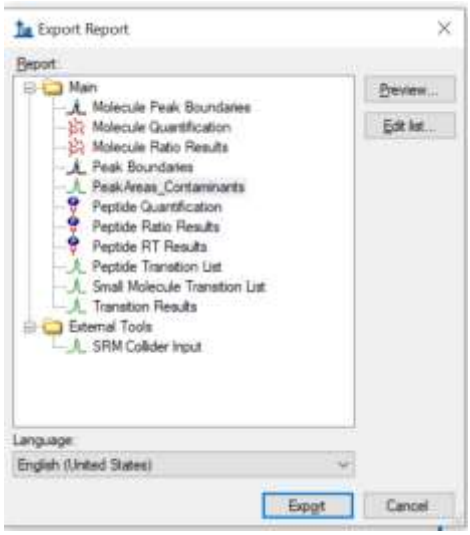
7. After the raw file processing is finished, you can browse through the list of contaminants and view the respective extracted ion chromatograms (EIC/XIC).

   You can do a first assessment at this level; for instance:
   Does PEG elute in regular patterns in the chromatogram?
   Does CHAPS elute with late retention time in the chromatogram?



8. Export the report by selecting ***File /Export / Report…***, select the report PeakAreas_Contaminants, navigate to the desired location and export the report as .csv (Note: use Language = English)





3.2. Contamination assessment using HowDirty

9. Open R in your IDE of preference (e.g., RStudio) and set the working directory to the desired folder (e.g., where the PeakAreas_Contaminants.csv is stored).

```
setwd("C:/Users/Name/ExampleHowDirty")
```

   Alternatively, from RStudio, you can create a project: ***File / New Project / Existing Directory / Browse / (select location) / Create Project***

10. Load the HowDirty package

```
library(HowDirty)
```

11. Create the sample annotation or experiment design file using the code below. If a file_report_skyline is provided (PeakAreas_Contaminants.csv), the ReplicateNameSkyline and Sample columns are populated with the unique(input$Replicate.Name). If not, only the column headers are provided.

```
get_annotation_template(file_report_skyline = "PeakAreas_Contaminants.csv")
```

12. Open the samples_annotation file in Excel (or R) and modify it if necessary.

    ReplicateNameSkyline: must contain the Replicate.Name entries from the Skyline output

    Do not modify it if the HowDirty template is generated as described in point 13.

    Condition: Condition or group of each sample. Conditions are used for grouped statistics and plots.

    Sample: sample name or code can be different from ReplicateNameSkyline

    DilutionFactor: If the sample was diluted, this factor can be used to multiply the abundance results. Use carefully since abundance may not scale in the same proportions for all the contaminants.

13. Create a HowDirty template by using the code below; this will automatically save the template with the name entered (e.g., “example.Rmd”)

```
HowDirty::get_report_template(file = "example")
```

    Alternatively, open a HowDirty.Rmd file from previous analysis and save it with another name

14. Open the .Rmd file

15. Fill the parameters in the header and “knit” (compile) the report. There are two options for this (see screenshots below).

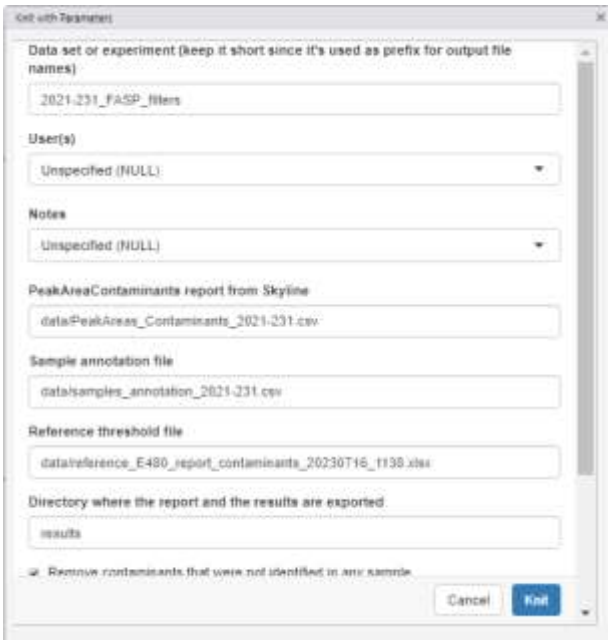
    This may take some minutes, depending on the computer used and the number of samples

    Reference threshold file: see section 4 for instructions to create the instrument-specific reference file. If it is not available, or you are aiming to generate it, keep the parameter as FALSE. This will instruct HowDirty to use thresholds based on our training dataset using an Ultimate3000 coupled to an Exploris 480.

    a. Click on “Knit with parameters”

    b. Fill out the parameters **values** manually, then click “Knit”
    The parameters are in the header of the document after “params:”
    The values are entered **after “value: ”**
    Do not modify the text after “label: ” or “input: ”

a 



b

```yaml
---
title: Evaluation of LC-MS contamination risk (HowDirty)
output:
  html_document:
    toc: true
    toc_float: false
    theme: united
    number_sections: TRUE
params:
  # fill the space next to "value:" with your values
  DataSet:
    label: "Data set or experiment (keep it short since it's used as prefix for output file names)"
    value: 2021-231_FASP_filters
    input: text
  UserNames:
    label: "User(s)"
    value:
    input: text
  Notes:
    label:
    value:
    input: text
  PeakAreasContaminantsFile:
    label: "PeakAreaContaminants report from Skyline" # PeakAreas_Contaminants report (directory/name) resulting from Skyli
    value: data/PeakAreas_Contaminants_2021-231.csv
    input: text
  AnnotationFile:
    label: "Sample annotation file" # Samples Annotation directory/name
    value: data/samples_annotation_2021-231.csv
    input: text
  RefThresholdsFile:
    label: "Reference threshold file" # file directory/name or FALSE
    value: data/reference_E480_report_contaminants_20230716_1138.xlsx
    input: text
  OutputDirectory:
    label: "Directory where the report and the results are exported"
    value: results
  RemoveMissingContaminants:
    label: "Remove contaminants that were not identified in any sample"
    value: TRUE
  nTopContaminantGroups:
    label: "Top n contaminant groups to show in plots"
    value: 10
    input: numeric
  MultiplyDilutionFactor:
    label: "Multiply abundance by DilutionFactor column" # Use carefully, normalization by TIC already takes into account s
    value: FALSE
  PlotsInteractive:
    label: "Export interactive plots" #TRUE = interactive in html, FALSE = static
    value: TRUE
---
```

16. When the report is finished, it will open in R. You can navigate through it in that window or open it in any browser.

4. Creation of instrument-specific reference threshold files

HowDirty can be used to generate lab- and instrument-specific contamination thresholds. To establish thresholds that represent the normal status of sample chemical background on a LC-MS instrument platform, it is recommended to generate a reference dataset. This can be done by processing files acquired over an extended time period (e.g., spanning multiple months) including samples of diverse origins.

17. Follow instruction in steps 1 to 7 to import the raw files from LC-MS analyses that represent the general usage of the instrument (e.g., spanning two months of analyses).
18. Carefully evaluate the feature assignment following the advice in step 7.

    Since this dataset will be used as a reference, it is especially important to thoroughly review it

19. Export the Skyline report following step 8.
20. Process the file using HowDirty as indicated in steps 9 to 16 following these considerations:
    a. In the parameters (step 15) set Reference threshold file: FALSE
    b. In the parameters (step 15) set Export interactive plots: FALSE

    This optional, but creating the interactive plots for a large dataset would take longer and create a heavier html file

21. Use the resulting Excel file as input for Reference threshold file (`RefThresholdsFile`) in your future analyses

5. References

[1] L.K. Pino, B.C. Searle, J.G. Bollinger, B. Nunn, B. MacLean, M.J. MacCoss, The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics, Mass Spectrom. Rev. 39 (2020) 229–244. https://doi.org/10.1002/mas.21540.

[2] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics. 26 (2010) 966–968. https://doi.org/10.1093/bioinformatics/btq054.

[3] M.J. Rardin, Rapid Assessment of Contaminants and Interferences in Mass Spectrometry Data Using Skyline, J. Am. Soc. Mass Spectrom. 29 (2018) 1327–1330. https://doi.org/10.1007/s13361-018-1940-z.