This script is intended to filter scans from a mgf file based on the presence of specific fragment ions and specific mass differences. The script is written in the R programming language, using RStudio.

This script was applied on mgf files produced from Thermo .raw files. Example files are given.

The ‘Scripts’ folder contains a ‘mgfs’ folder, that should contain the .mgf files, an example file is provided. Hence, the raw LC-MS files should first be transformed to .mgf files, refer to the 'msconvert' program: <http://proteowizard.sourceforge.net/download.html>

The ‘settings\_file.xlsx’ contains 3 sheets:

* + ‘input’ contains a table with information for the filtering of scans, example data is given:
    - ‘classes’: names of the compound classes to be searched
    - ‘MS2’: fragment ions to be searched, if multiple fragments, separate them with a comma
    - ‘differences’: the mass differences to be searched
    - ‘notes’: metadata that is not read from the script, notes for your own consumption
  + ‘settings’ contains LC-MS settings:
    - ‘min\_int’: the minimum intensity of a fragment ion. Below that intensity, fragment ions are deleted.
    - ‘mz\_tol’: the allowed mass tolerance in ppm
    - polarity: the polarity, accepts values "neg" or "pos".

Instructions on using the script:

Copy the ‘Scripts’ folder somewhere on your PC. This will be your working directory.

Example:

E:\LC-MS\Portulaca\Scripts

Download and install RStudio (<https://posit.co/downloads/>)

The script requires the following packages, if you do not have them installed, enter the following commands in the RStudio console:

install.packages(tidyverse)

install.packages(readr)

install.packages(dplyr)

install.packages(sos)

install.packages(readxl)

install.packages(data.table)

install.packages(openxlsx)

To run the script, enter the following commands in the RStudio console:

source(“E:/LC-MS/Portulaca/Scripts/functs.R”)

source(“E:/LC-MS/Portulaca/Scripts/R\_script.R”)

***These are the paths, if your working folder is ‘E:\LC-MS\Portulaca\Scripts’.***

Note that the backslash used in paths in Windows, like “E:\LC-MS\Portulaca\Scripts” should be replaced with forward slashes, like “E:/LC-MS/Portulaca/Scripts”!

After the script runs, follow the instructions.

The results are be saved in a ‘results’ folder in your working folder, like ‘E:\LC-MS\Portulaca\Scripts\results’