nLossFinder

A Graphical User Interface program for non-targeted detection of DNA adducts

User Manual

Requirements and installation

nLossFinder requires a MATLAB environment (R2020a or later) to work, and also the installation of the MATLAB addon 'GUI Layout Toolbox', which can be downloaded from the MATLAB website https://se.mathworks.com/matlabcentral/fileexchange/. nLossFinder was tested in Windows 10, running on a PC with equipped Intel(R) Core(TM) i9-9900K CPU @ 3.60GHz , 16 GB RAM and 4 GB GPU. Inferior hardware should work fine, although 8 GB RAM is recommended.

In MATLAB, open the folder containing the code and type 'nLossFinder' in the Command Window to run.

nLossFinder was designed to analyse LC-MS/MS using a DIA (Data independent Acquisition) method and has only been tested on data obtained from a Thermo Fisher Orbitrap Q Exactive HF mass analyser. The experimental files obtained from the instrument (Thermo Fisher RAW format) must be converted into mzXML format in order to analyse them in nLossFinder. This conversion can be performed using ProteoWizard MSConvert, which is an open source program that can be downloaded from the website http://proteowizard.sourceforge.net/. The MSConvert parameters must be set as illustrated in Figure S1, where, from the generic defaults, the output format is 'mzXML', the 'zlib compression' is unchecked, and a 'Peak Picking' filter (MS levels 1-2) is added. This filter centroids the data.

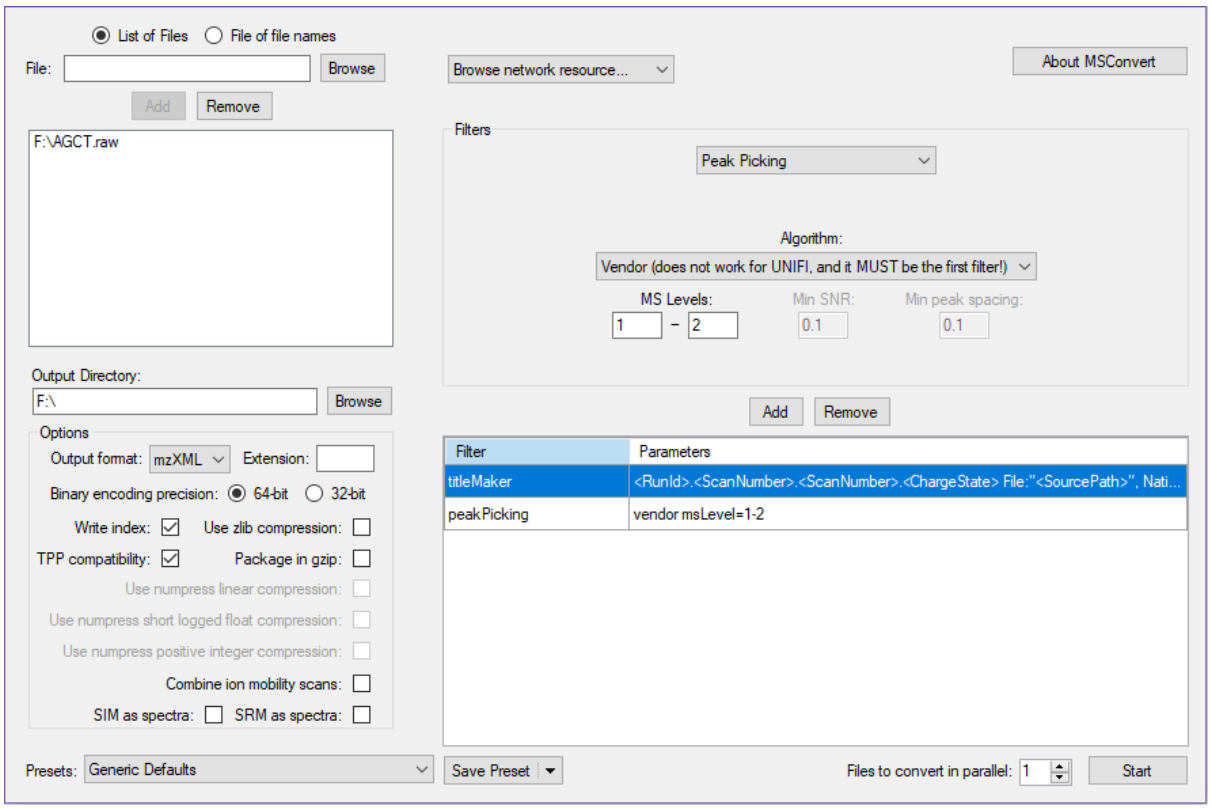

Figure S1. ProteoWizard MSConvert setup.

nLossFinder - Main menu

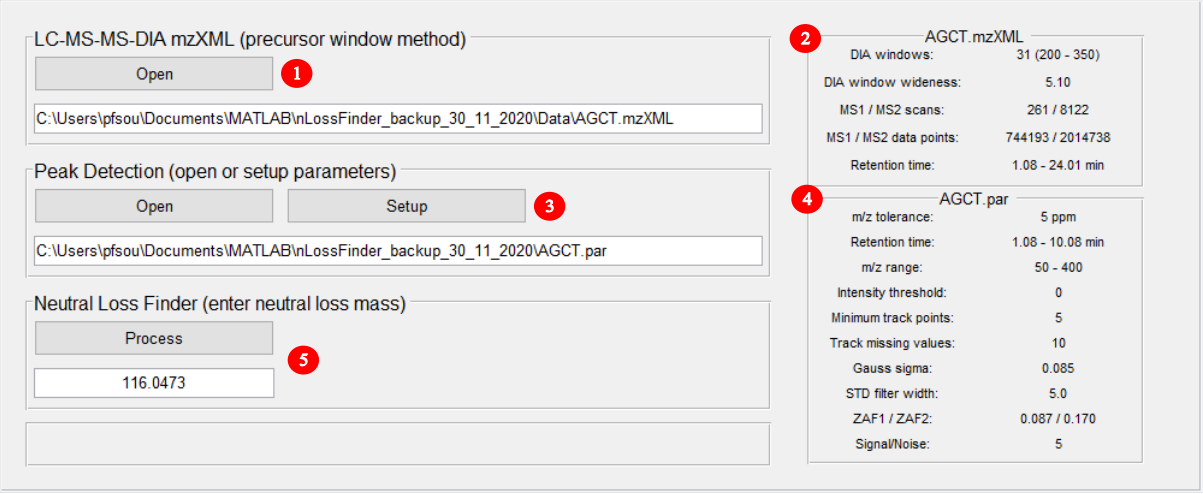

Figure S2. nLossFinder main menu.

(1) Open a mzXML file containing LC-MS/MS data obtained using the DIA method.

(2) If the file is opened successfully, the information of this file will appear in the panel.

(3) Setup or open peak detection parameters. The peak detection parameters must be setup for different experimental conditions. Otherwise, one setup file can be applied in a batch of experiments. The GUI continues to the peak detection steps explained bellow.

(4) The peak detection parameters are displayed in the panel to the right.

(5) After setting up the peak detection parameters, enter the value of the neutral loss (*e.g.* deoxyribose 116.0473 Da), then click to process the peak detection and neutral loss finder algorithms. After processing a new section will open to visualize and the results (see bellow nLossFinder - Analysis of Results).

Peak detection setup – retention time range

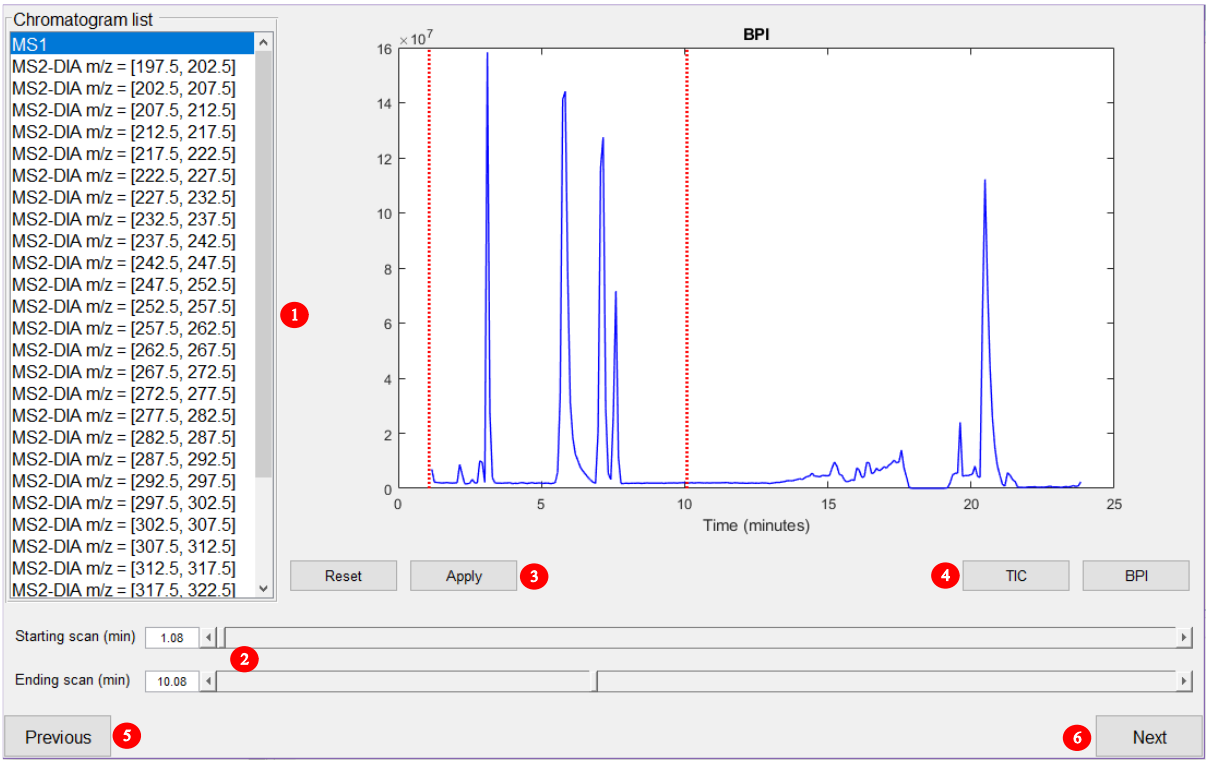

Figure S3. Setup retention time range.

(1) The user can visualize the chromatograms, *i.e.*, full scan MS1 and full scan MS2 DIA.

(2) Sliding the bars, or using the arrows, or input starting and ending retention times will subset all the chromatograms.

(3) Undo or apply the set scan range.

(4) Visualize Total Ion Chromatograms (TIC) or Base Peak Chromatograms (BPI).

(5) Go back to main menu.

(6) Apply settings and move to the next section.

Peak detection setup – *m/z* range and precursor minimum intensity threshold

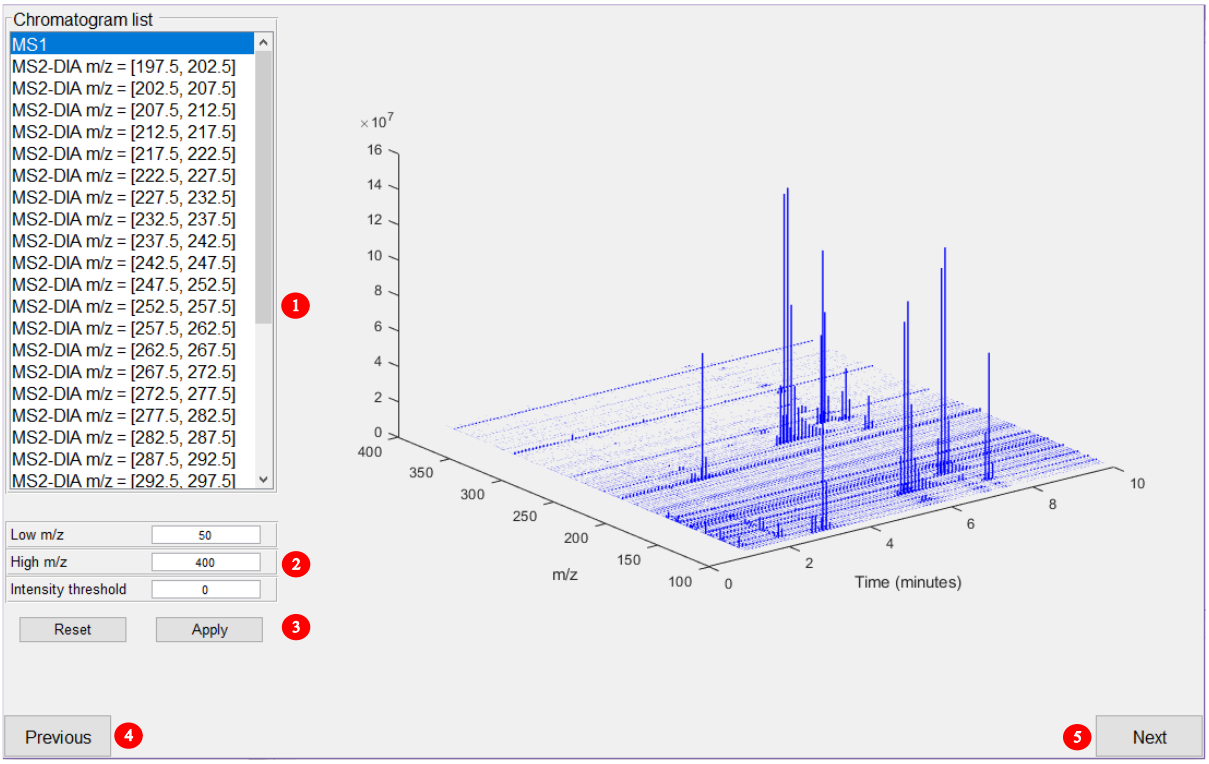

Figure S4. Setup *m/z* range and minimum intensity threshold.

(1) The user can visualize the spectral maps for each chromatogram, *i.e.*, full scan MS1 and full scan MS2 DIA.

(2) Set the lowest and highest *m/z* value, and the minimum intensity. This only affects the precursor (MS1) spectra though.

(3) Undo or apply the settings.

(4) Go back to the previous section.

(5) Apply settings and move to the next section.

Peak detection setup – Pure Ion Chromatogram (PIC) tracker

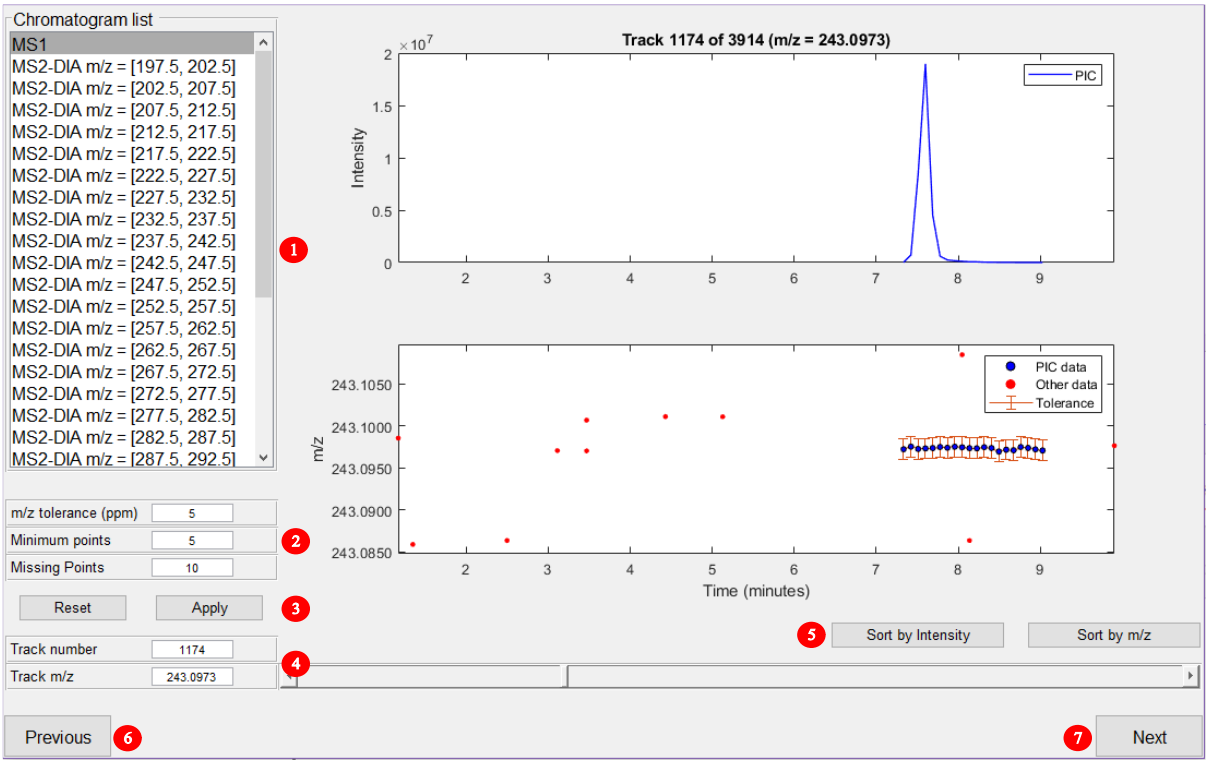

Figure S5. Setup PIC tracking parameters.

(1) Select chromatogram, *i.e.*, full scan MS1 or full scan MS2 DIA.

(2) The *m/z* tolerance sets the desired tolerance for tracking *m/z* values across scans. This value is the same used when processing the data (from the main menu) to find the neutral loss matches between precursor and fragment ions. The minimum points define how many points a PIC should have, and the missing points is a tolerance for missing *m/z* values between scans.

(3) Undo or apply the settings. Here the spectral data that does not belong to any PIC are discarded.

(4) Input *m/z* values to explore the tracked PICs.

(5) Sort the tracks by intensity or by *m/z* values. This and the latter (4) are useful when searching of an ion of interest *e.g.*, type a *m/z* value around the value of interest and move the arrows until the actual *m/z* value is found.

(6) Go back to the previous section.

(7) Save settings and move to the next section.

Peak detection setup – Peak detection

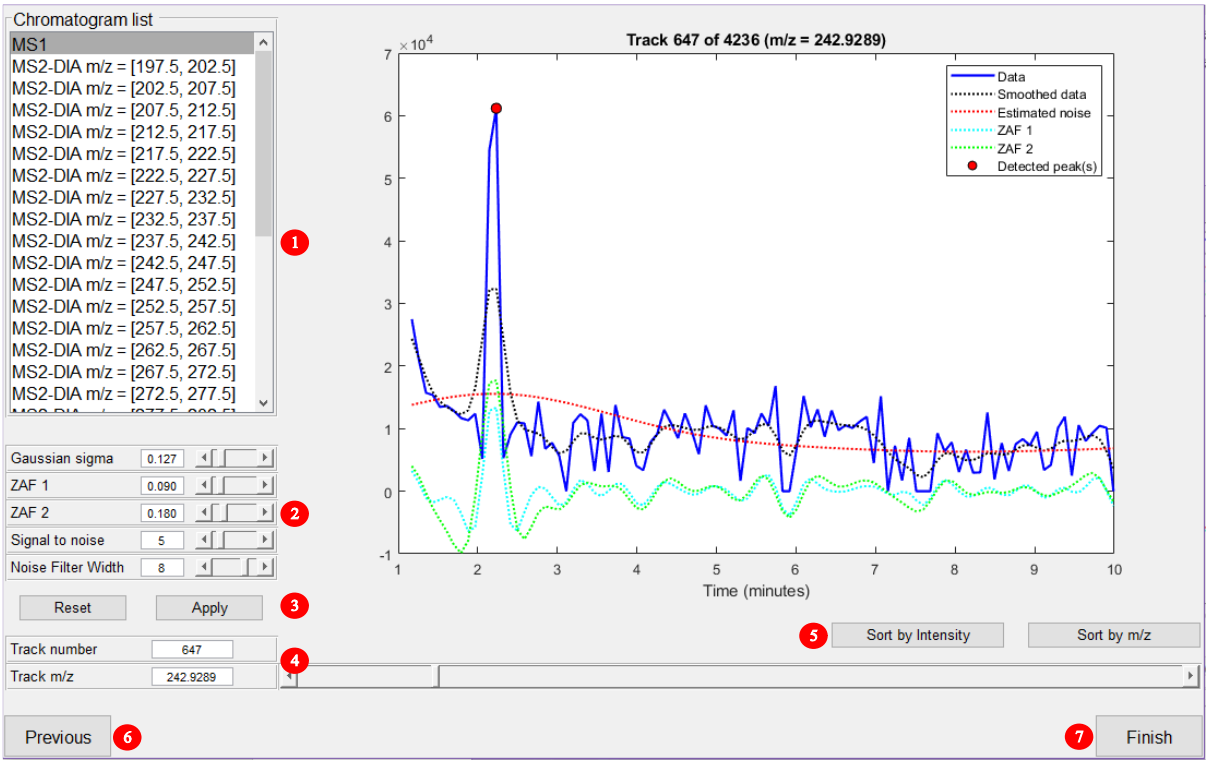

Figure S6. Setup peak detection parameters.

(1) Select chromatogram, *i.e.*, full scan MS1 or full scan MS2 DIA.

(2) The control box contains parameters that determine the peak detection:

i. Gaussian sigma - This parameter controls the smoothed data (black dotted line) and the noise estimate (red line). By clicking on the arrows, *i.e.*, increasing or decreasing the value will modulate these lines depending on the data (blue line). For a sensitive filter, the red line should be as close to the base line as possible.

ii. ZAF1 and ZAF2 - are the zero areas filters parameters that modulate the blue and green dotted lines. These filters fit the PIC data. Whenever these lines overlap the red line (estimated noise) a peak is detected, and a red circle will be plotted on the maximum of the detected peak. The ZAF 2 should be the double of the ZAF 1 value. These filters do the same, but one should be broader than the other, so that the peaks with different widths can be fit better.

iii. The signal to noise will raise the red line and the Noise Filter Width will broaden it. These parameters adjust the shape and height of the noise estimate.

(3) Reset the parameters or apply. Applying will remove PICs where no peaks have been detected. This process may be slow, and it is not required to setup the parameters. In this stage, it just informs how many tracks will be used when processing to find neutral losses.

(4) Input m/z values to explore the tracked PICs.

(5) Sort the tracks by intensity or by m/z values. This and the latter (4) are useful when searching of an ion of interest *e.g.*, type a m/z value around the value of interest and move the arrows until the actual m/z value is found.

(6) Go back to the previous section.

(7) Save the peak detection parameters and go to the main menu.

nLossFinder – Analysis of results

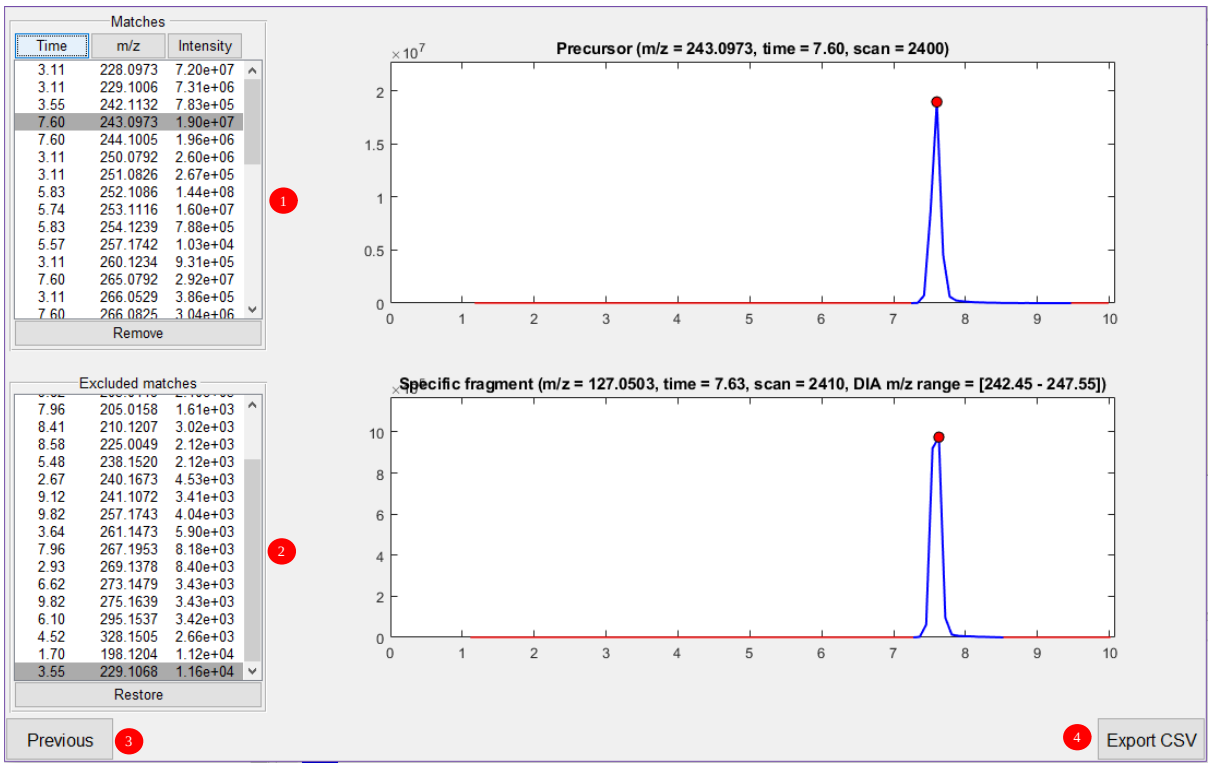

Figure S7. Analysis of matches and exporting results.

(1) The results obtained by the Neutral Loss Finder algorithms are presented in this list. The list can be sorted according to the retention time, the *m/z* value or the intensity at the maximum of the peak of the precursors. Sorting the precursors by intensity can help verify the matches of lower intensities, which are usually peaks in noisy PICs., *i.e.*, dragging signals across the chromatogram. These may be removed from the list if needed. Also, isotopes and adducts such as sodium, potassium, or other possible ions that may be formed in the ionization process can be tracked by sorting (first my *m/z*, then by retention time). Those matches with the same (or very close) retention times are potentially ESI adducts and isotopes and can be removed if needed. This is illustrated in figure S7. Here, 6 matches with the same retention time were selected. The first match (243.0974) corresponds to the protonated Thymidine (the sample analysed in this experiment was a mixture of the four nucleoside standards dA, dG, dC and T), the second (244.1005) is an isotope (C13) of the first ion, the third match is a sodium adduct (proton-sodium exchange ~ 22 Da) of the first, and the fourth is an isotope of the third. The other two may be also adducts formed in the ESI, such as dimmers or other side reactions, or eventually just overlapping compounds. Therefore, these results should be carefully studied, to avoid including many false positives in the results.

(2) The excluded matches list will not be saved in the output data. Clicking on restore will move the match(es) back to the inclusion list.

(3) Go back to the main menu. The processed results are lost though.

(4) Save the results (Matches) in a table, in a comma separated values (CSV) file, which can be opened in Excel, or any other program for further data analysis. The output file contains the retention time, *m/z* values, intensities of the maximum of the peaks and peak areas of the precursors and specific fragments. The neutral loss value (difference between precursor and specific fragment) and the DIA value (the centre of the corresponding precursor DIA window) are also recorded.

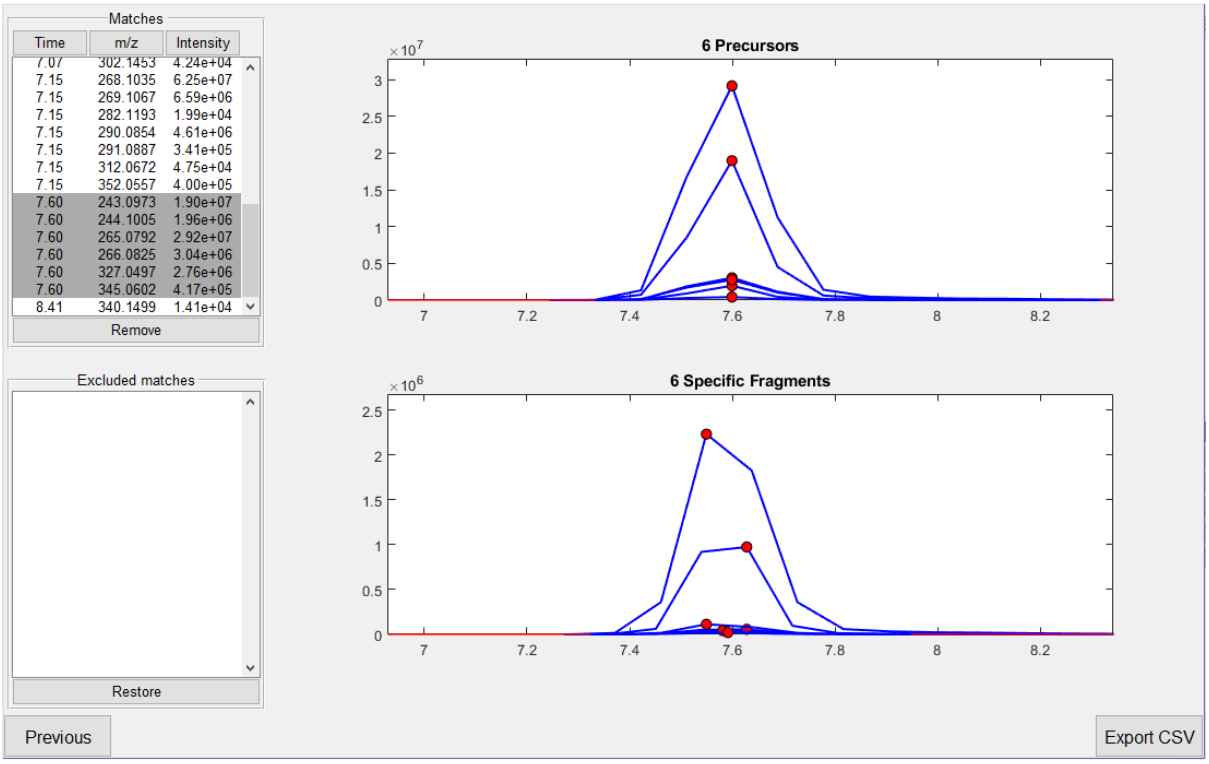

Figure S7. Potential isotopes or ESI adducts can be studied on overlapped ion peaks.