

LCMALDIprocessor User Documentation

Description

LCMALDIprocessor processes the mass spectrometry data produced in a LC MALDI MS and MS/MS experiment. It currently supports MS data from the Bruker range of MALDI instruments but has only been tested with data from the Bruker Ultraflex. LCMALDIprocessor extracts the mass and intensities of all the MS data. It utilises the common contaminant algorithm and generates a list of all the masses found and then removes duplicates that are within the user defined mass tolerance range. The user then selects a cut-off value. All masses that have been found in that many files or more will be removed. The program then generates tables of the Base Peak Intensity (BPI), number of peptides per a fraction, a count of the occurrences of each peptide and two variations of 2D matrices with the time, mass and intensity. These tables are suitable for graphing with common software packages e.g. Microsoft Excel and Igor Pro or more sophisticated packages like MatLab. Finally the program generates a tab delimited worksheet listing all the masses with their elution fraction and MS/MS information.

Usage

1. Use the Data Directory button to select the directory that contains all the MS and MS/MS data alternatively type in the path to the directory. The data can be stored in different folders. The experiential name used in the results file is the main data directories name.

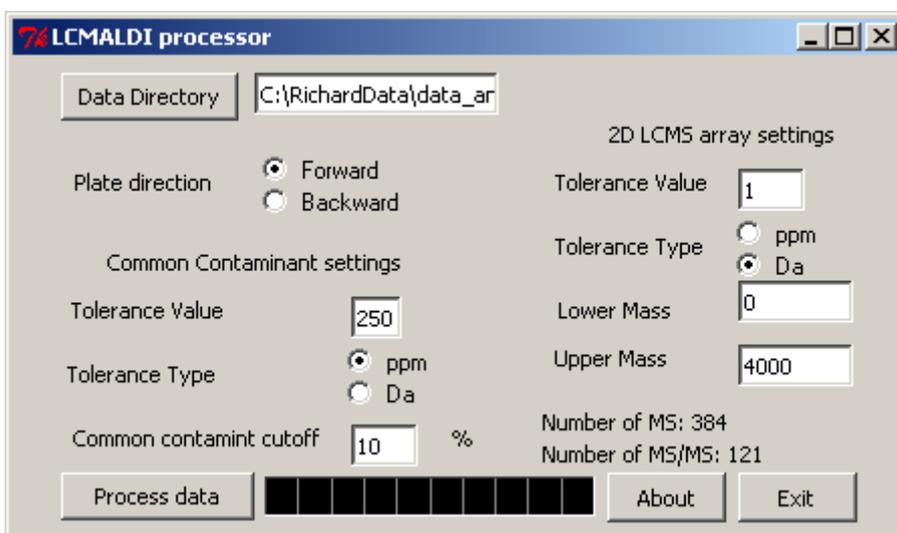


Figure 1: Graphic user interface to LCMALDIprocessor

2. Set the direction of the fraction collection.
3. Set the parameters for the common contaminant processing. Select the tolerance type and value. Depending on the calibration across your data set you may need to use a larger mass tolerance than normal. Comparisons are made to the lowest mass of a set. If the tolerance is not set high enough you will observe the counts of common masses been split in two with the second mass just outside of the tolerance value used. Enter a cut-off percentage value.

All masses that occur in more than the cut-off percentage of files will be removed.

4. Set the parameters for the 2D LCMS array settings. All the masses are binned into an array of time verses mass verses intensity. The tolerance type and value set the bin size. The lower and upper mass values set the size of the array. The default is one Dalton bins from 0 to 4000 m/z.
5. Start the program processing the data by pressing the `Process Data` button.
6. In total ten text files are produced. Table 1 shows their names and functions. All the data files are tab delimited text files although the tables and the worksheet have been given the `.xls` extension so that they are automatically associated with Excel.

Filename	Type	Function
2D-matrix_DataSet.txt	Tab delimited table	Data array for plotting 3D graph
graph_DataSet.xls	Tab delimited table	Data array for plotting 2D graph
peakcount_DataSet.xls	Tab delimited table	Data array for plotting number of peaks vs. fraction graph
bpi_DataSet.xls	Tab delimited table	Data array for plotting Base Peak Intensity vs. fraction graph
count_DataSet.xls	Tab delimited table	Data array for plotting frequency vs. mass graph
DataSet.xls	Worksheet	Experiment worksheet
AllMasses_DataSet.txt	Mass list	List of all masses labelled by Flexanalysis software
AllMasses_clean_DataSet.txt	Mass list	A unique list of masses after the common contaminant and duplicate masses have been removed.
AllMasses_exclude_DataSet.txt	Mass list	A complete list of the common contaminant masses and their fraction number.
Exclude_list_DataSet.txt	Mass list	List of the unique common contaminant masses that have been excluded.

Table 1: LCMALDIprocessor results files for the data directory `DataSet` .

7. The number of MS and MS/MS spectra in the data directory is shown in the right hand corner of the GUI.
8. The worksheet contains basic information about the data set along with a complete list of masses identified and sorted by fraction (Figure 2). The parent

ion and fraction number of the masses that have been submitted to MS/MS are also recorded on the worksheet.

	A	B	C	D	E	F	G	H
1	Experiment name	CellLysate3						
2	Tolerance	250	ppm					
3	Cut of value for excluded peptides set to:	38.3						
4	Number of MS spectra:	384						
5	Number of LIFT spectra:	121						
6	Number of peptides:	1950						
7	Number of unique peptides:	757						
8	Number of unique excluded peptides:	3						
9	Database searched:							
10								
11								
12	Protein	No. peptides ID						
13								
14								
15								
16	Unassigned/undetermined	0						
17								
18								
19	Mass	Spot Name	Spot No	Time	Status	LIFT	Parent ion ID	
20	702.6936742	0_A1	1	7.5	Clean			
21	704.7053775	0_A1	1	7.5	Exclude			
22	706.7157935	0_A1	1	7.5	Clean			
23	734.1215106	0_A1	1	7.5	Clean			

Figure 2: Example experiment worksheet.

- Two or three dimensional plots can be generated from the graph_DataSet.xls and the 2D-matrix_DataSet.txt tables. Figure 3 shows an example plot. These graphs give a good overview of the distribution of peptides across the LC separation/fractions. Fractions with lots of intense peptides can be quickly identified as can sticky peptides that elute across multiple fractions. To make the graph shown in Figure 3 plot the mass vs fraction number and use the intensity as the z value (Figure 4).

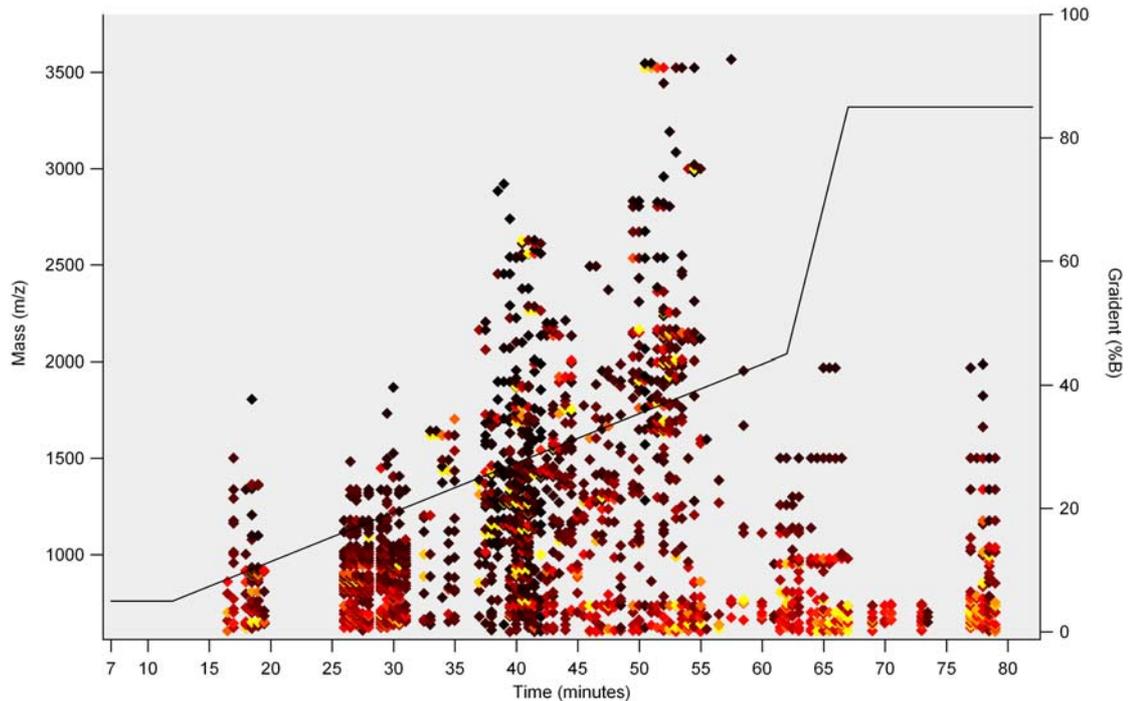
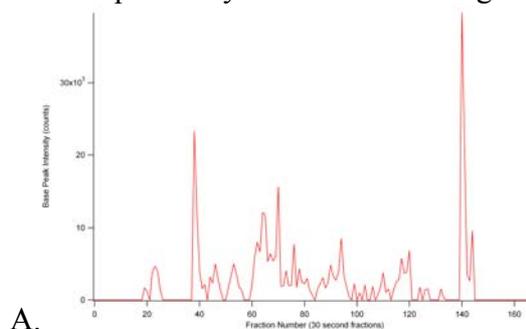


Figure 3: 2D plot of a cysteine containing peptide fraction from the saliva proteome. The graph has been overlaid with the gradient information and the fraction numbers converted to time. The graph was plotted with Igor Pro (Wavemetrics, Inc).

	A	B	C	D	E
1	Spot	Mass	File Number	Intensity	
2	0_H6	604.4005	19	1395.534	
3	0_H6	657.363	19	710.2547	
4	0_H6	699.1449	19	1673.35	
5	0_H6	797.4302	19	620.8751	
6	0_H6	859.3938	19	1093.63	
7	0_H5	621.2957	20	742.8037	
8	0_H5	644.1413	20	1178.552	
9	0_H5	699.1603	20	862.7252	
10	0_H5	718.3235	20	704.8056	
11	0_H5	769.373	20	747.9973	
12	0_H5	957.4554	20	386.0764	

Figure 4: Excerpt of the graph_DataSet.xls table.

10. Other plots of interest are shown in figure 5. These plots act as supplementary information to the main 2D graph and can help a user gain an overall view of the experiment and help identify fractions deserving additional analysis.



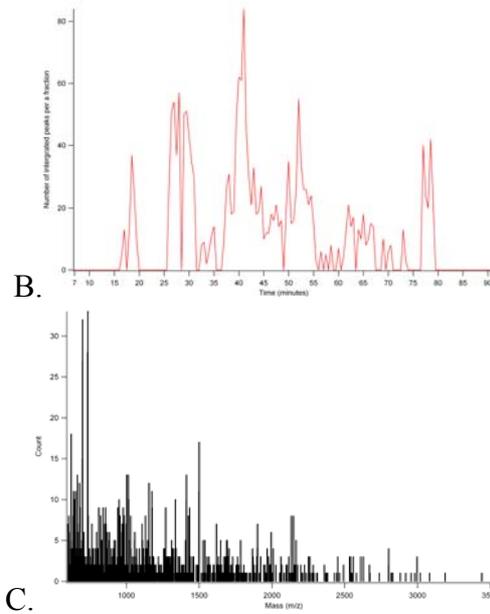


Figure 5: A. Plot of Base Peak Intensity vs. fraction number. In this case the x axis has been converted to time. B. Plot of the number of peaks per a fraction. The x axis has again been converted to time. C. Plot of the mass vs. frequency of the mass. This data is used by the program to determine the common contaminant masses.

Bugs

1. The browse for folder window's have an array reference at the top. This is cosmetic and unfortunately there is no fix as yet.
2. No possibility to set the fraction time. Fraction time is currently hard coded to 30 seconds per a fraction.

Feature requests

1. The software currently only supports windows operating systems. It will be expanded to support Mac/Unix/Linux based systems.

If you find a bug or have a feature request write me an email r.jacob@ucl.ac.uk.