Adducts and Neutral Loss Annotation Using mzAdan Software

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

LSMS

20 YEARS SIB Swiss Institute of Bioinformatics

Table of Contents

INTRODUCTION:

Observation of adducts and neutral losses is quite common in electrospray ionisation (ESI) and can be used for the identification of protonated molecules, which can then be assigned as precursor ions for tandem mass spectrometry (MS/MS). The focus of the mzAdan software is to automatically annotate $[M+H]^+$ candidates with the help of a customizable set of annotations. It assumes that any ion in the mass spectrum is an hypothetical $[M+H]^+$ regardless of its intensity. The current version of mzAdan works on singly charged ions in positive mode but will be extended to negative mode and multiply charged ions. The software can process a single spectrum or multiple spectra in the *mgf* format.

***Table 1**: Compounds included in the mzadan_mix_5.mgf dataset.*

index	rt (min)	name	exact mass (u)
1	1.564	N-Acetylputrescine	130.11121
2	4.578	Ethenodeoxyadenosine	275.10335
3	7.570	Monomethyl glutaric acid	146.05793
4	10.363	N-Acetyl-L-phenylalanine	207.09041
5	17.359	Taurocholic acid	515.29167

The dataset provided for this tutorial consists of 5 spectra of 5 metabolites extracted from a LC-MS analysis (Table 1).

PREREQUISITE:

Before launching mzAdan for the first time, make sure that you have installed a recent release of the Java Development Kit (JDK 8 or higher) and downloaded the most recent distribution of the software.

The zip archive includes three folders. The **data** folder contains the dataset described in Table 1. The ***output*** folder contains the annotation tables, candidates' tables and the graph files generated by the software as the result of processing the dataset. Lastly, the ***annotations*** folder includes three tables containing annotations that are specifically formatted to be imported into mzAdan. To launch the application, first unzip the archive then double-click on the *.jar* file. If you are using Windows and the software does not launch, double click on the ***mzAdan.bat*** file.

TUTORIAL:

1. Project

Export Panel

The ***Export*** panel (Figure 1) is subdivided into two sections: ***Input and output*** and **Output Selection**. In the first section, you are given the option to import either a single file or multiple ones from a designated directory. Note that by default the single ***File*** option is selected. To process multiple files, click on the dropdown menu and select the ***Directory*** option. In the output field, you can choose the name and the base directory of the output files.

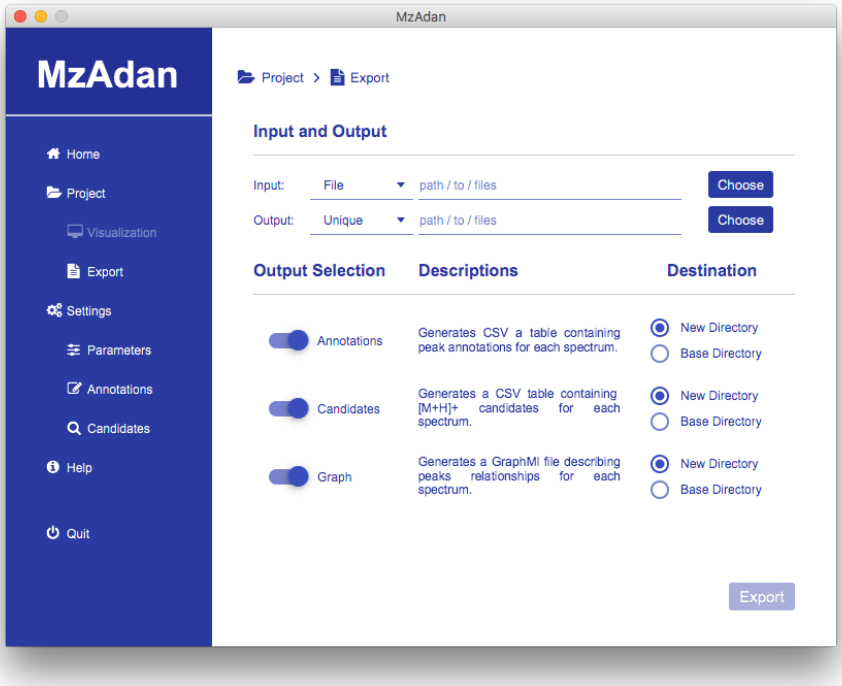

Figure 1: A new project can be created through the export panel. Mass spectra can be imported as single or multiple MGF files and the results exported as CSV and GraphMl files.

In the ***Output Selection*** section, you can select the types of output as well as their destination. The output files can either be grouped by type in new directories or exported in the base directory selected in the ***output*** field of the ***Input and Output*** section.

In this tutorial, we will be practicing on a dataset consisting of only a single file containing five spectra (Table 1).

First select the single ***File*** option in the input field. Then press the ***Choose*** button to the right of the ***input text field***. A new dialogue window pops up requesting the selection of mass spectra. Select the ***mzadan_mix_5.mgf*** file located in the ***data*** folder then press the ***Select*** button. Finally press the Choose button to the right of the ***output text field*** and select the ***mzAdan_tutorial*** directory.

Note that the Export button located on the bottom right side of the panel has changed colour indicating that the input can now be processed. However, before starting the analysis you must modify some settings located in the ***parameters***, ***annotation*** and ***identification*** panels.

2. Settings Panels

Every setting panel is accessible via the side menu bar or via the Settings Panel. The latter features a description detailing the content of each setting panel.

Parameters Panel

The ***Parameters*** panel (Figure 2) contains filtering options applying to the spectra and the annotations. The ***Threshold*** can be adjusted to exclude peaks whose intensity and relative intensity are lower than the set value. The ***Range m/z*** is used to remove peaks with a mass-to-charge ratio (m/z) outside the set range. The ***Deisotoping*** filter removes isotopic peaks with a relative intensity outside the set range. The validation option filters out peaks, which do not have detectable isotopes in the mass spectrum. The ***Tolerance*** filter discards annotations with an absolute error (mmu) higher than the set value.

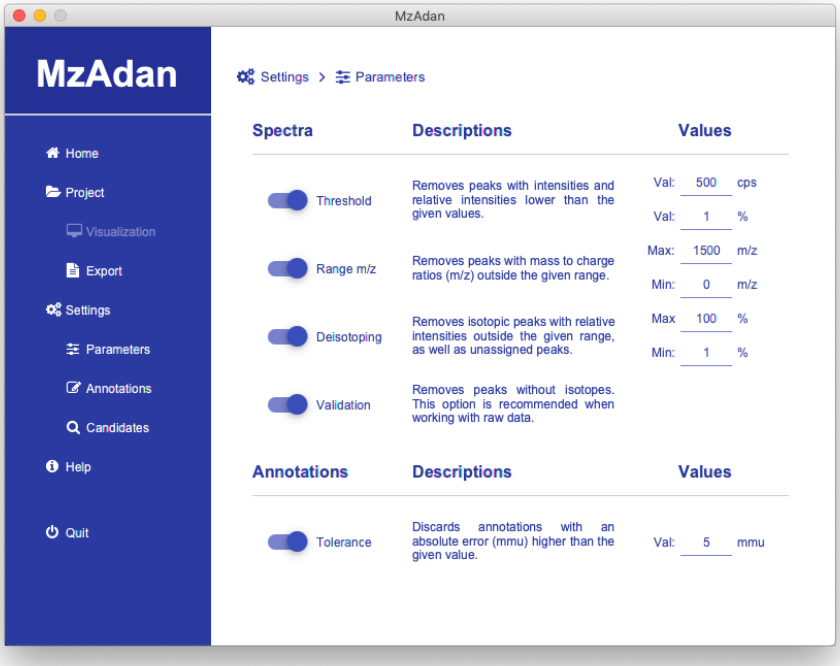

Figure 2: *The parameters panel contains options to filter peaks from the mass spectrum based on their intensities, mass-to-charge ratios and assignments (monoisotopic, isotopic and unassigned). It also contains an option to discard annotations based on mmu error.*

In this tutorial, we use the default parameters except for the ***Range m/z*** filter**,** which is disabled by clicking on the corresponding toggle button. The deisotoping filter is used to remove isotopic and unassigned peaks from the mass spectrum.

Annotations Panel

The ***Annotations*** panel (Figure 3) contains options to customize the Adducts, Neutrals and Isotopes annotations sets. Each set can be enabled or disabled by clicking on the corresponding toggle button located on the left side of the panel. There are three additional buttons to the right. The leftmost button is the Edit button. Clicking on it opens a new window where annotations can be individually enabled or disabled. The middle button is the ***Import*** button. The third and last button is the ***Reset*** button. It will reset the set to its initial annotations. In the ***Advanced*** section, the annotation of multimers can be enabled.

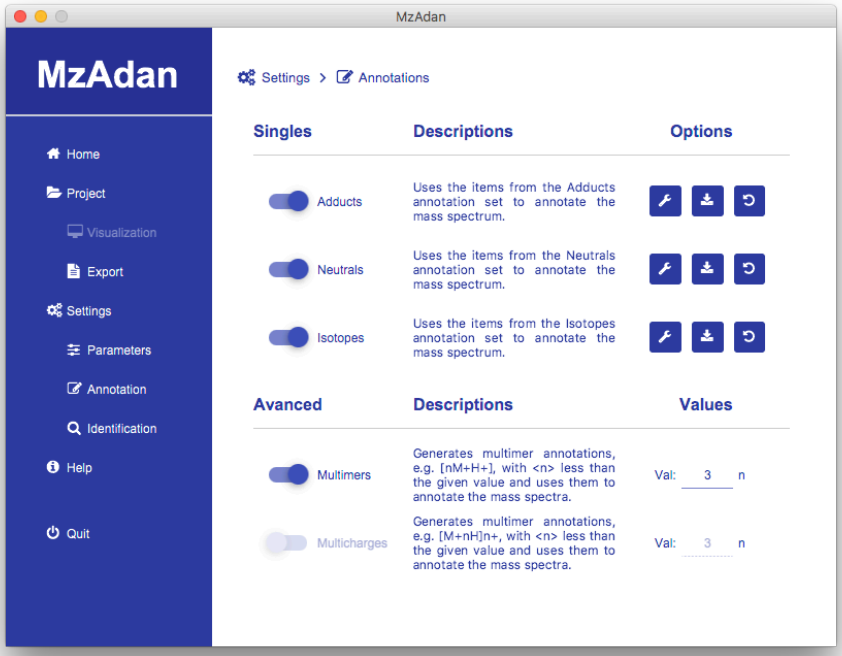

***Figure 3**: The annotation panel contains options to fully customize the adducts, neutral losses and isotopes annotations sets. Annotations can be individually enabled or disabled and custom sets can be imported. The Annotation of Multimers is also available.*

Applying the deisotoping filter discussed previously, filters out the C13 isotopes from the mass spectrum. Therefore, this annotation should be disabled from the ***Isotopes*** annotation set. To do so, click on the ***Edit*** button (wrench icon) located to the right of the ***Isotopes*** description. In the Edit window, disable the C13 annotation by selecting it in the ***Enabled*** table and pressing the left arrow. By default, this set contains only the C13 annotation. Therefore, another possibility would have been to disable the whole set by clicking on the Isotopes Toggle button.

Identification Panel

The ***Candidates panel*** (Figure 4) contains filtering options that apply to the candidate tables. The first section includes a filter for Candidates and the second filters for the clusters containing a candidate of interest. These filters will be further discussed below (Outputs).

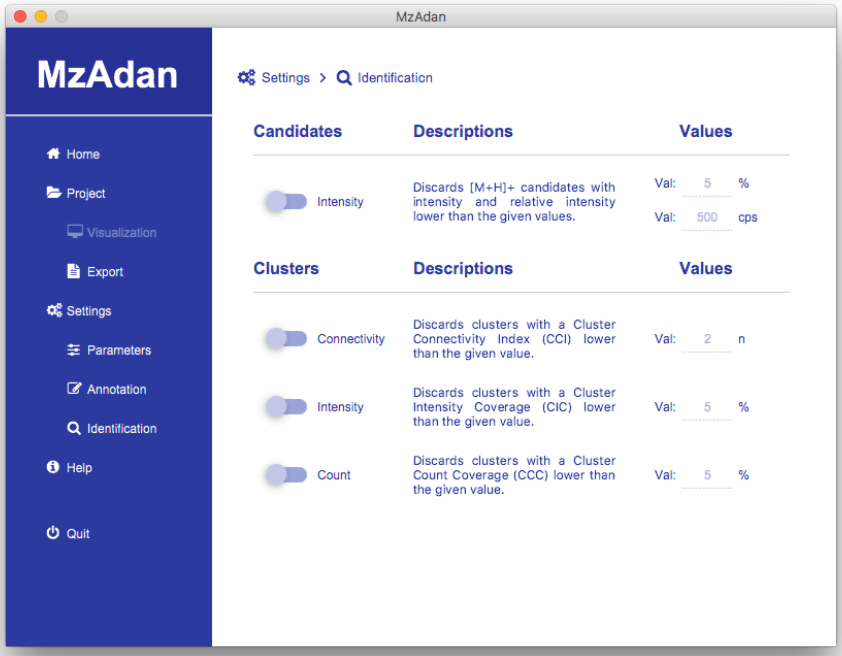

Figure 4: The identification panel contains options to discard [M+H]+ candidates based on their intensity and relative intensity, or based on their cluster connectivity, intensity and peak count.

In this tutorial, we exclude candidates with an intensity and relative intensity lower than 500 cps and 5% respectively. We also discard candidates, which are not connected to any other peak in the mass spectrum. To do so, first enable the ***Intensity*** filter in the ***Candidate*** section and set the minimum intensity and relative intensity to 500 counts and 5%. Then enable the ***Connectivity*** filter in the ***Cluster*** section and set the minimum value to 2. When this is done, go back to the ***Export panel*** and press the Export Button.

3. Outputs

Candidates Tables

Open the output folder which has just been created. In total, 16 files including 11 tables are distributed between the *annotations*, *candidates* and *graphs* folders. The *annotations* and *compounds* tables can be opened and modified with any text or spreadsheet editor. The graph files, on the other hand, should be open with a Graph visualization software such as the open source and freely available *Gephi* platform.

The content of the **compounds** files is compiled in a **summary** table structured as Table 2: Upper left segment of the summary table resulting from the annotation by mzAdan. and Table 3. The first three columns feature the **name** of the file, followed to the right by the **retention time** and the **index** of the mass spectra. The [M+H]+ candidates are stored in individual rows. Their **cluster** id, exact **mass**, and the corresponding peak **m/z**, **intensity** and **relative intensity** can be found to the right of the spectrum index column.

***Table 2**: Upper left segment of the summary table resulting from the annotation by mzAdan.*

file name	rt	index	cluster	mass	*m/z*	intensity	rel. intensity
mzadan_mix_5.mgf	1.564.	1	C0	113.08393	114.0912	42661	29.69
mzadan_mix_5.mgf	1.564	1	C0	130.11121	131.11848	143674	100
mzadan_mix_5.mgf	4.578	2	C0	275.10335	276.11062	169659	100
mzadan_mix_5.mgf	4.578	2	C5	216.09934	217.10661	8792	5.18
mzadan_mix_5.mgf	7.570	3	C0	146.05793	147.0652	24910	18.75
mzadan_mix_5.mgf	10.363	4	C0	207.09041	208.09768	82496	100
mzadan_mix_5.mgf	10.363	4	C6	233.12576	234.13303	4616	5.6
mzadan_mix_5.mgf	10.363	4	C6	216.09984	217.10711	14321	17.36

The three last columns of the table contain information that will help us discern real [M+H]+ candidates from noise and fragments. The Cluster Connectivity Index (***CCI***) represents the number of peaks connected directly or indirectly by annotations to the peak of interest. These networks of interconnected peaks are called ***clusters***. The Cluster Intensity Coverage (***CIC***) corresponds to percentages of the Total Ion Current (TIC) explained by a cluster. Lastly, the Cluster Count Coverage (***CCC***) corresponds to percentages of the spectrum peaks contained in a cluster.

***Table 3**: Upper left segment of the summary table resulting from the annotation by mzAdan.*

file name	rt	index	cluster	mass	*cci*	cic	ccc
mzadan_mix_5.mgf	**1.564.**	**1**	**C0**	**113.08393**	**4**	**94.14**	**57.14**
mzadan_mix_5.mgf	**1.564**	**1**	**C0**	**130.11121**	**4**	**94.14**	**57.14**
mzadan_mix_5.mgf	4.578	2	C0	275.10335	3	78	21.43
mzadan_mix_5.mgf	4.578	2	C5	216.09934	3	5.21	21.43
mzadan_mix_5.mgf	7.570	3	C0	146.05793	7	71.38	33.33
mzadan_mix_5.mgf	10.363	4	C0	207.09041	7	57.72	20.59
mzadan_mix_5.mgf	10.363	4	C6	233.12576	3	8.51	8.82
mzadan_mix_5.mgf	10.363	4	C6	216.09984	3	8.51	8.82

Open the ***mzadan_mix_5_summary.csv*** table located in the *compounds* folder. One or more protonated molecules may be annotated as [M+H]+ candidates in each mass spectrum. Note that some are part of the same cluster (Table 2). Ideally, the software should provide a single candidate per cluster. However, this may be impossible in some situations, which we will discuss below using *annotations* tables.

Annotations Tables

The ***annotations*** tables are structured as follows. Each peak still present in the mass spectrum after filtering is featured on the left side. Their ***m/z***, ***intensity*** and ***relative intensity*** are in the first, second and third columns respectively. To the right of these are the annotations, grouped by type and sorted by mass shift. Four columns are assigned to each annotation and contain data relative to the annotated peaks (***m/z*** and ***intensities***) and quality of the match (***error***).

***Table 4**: Segment of the annotation table generated for the first mass spectrum (neutral losses).*

	NH3 (-17.02655)

m/z	intensity (cps)	intensity (%)	m/z	intensity (cps)	intensity (%)	error (u)
100.07429	6866.3	4.78	0	0	0	0
114.0912	42660.5	29.69	0	0	0	0
131.11848	**143673.9**	**100**	**114.0912**	**42660.5**	**29.69**	**7.30E-04**
169.07318	3100.5	2.16	0	0	0	0
191.16418	1583.2	1.1	0	0	0	0
217.10614	3447.5	2.4	0	0	0	0
261.22876	1602.7	1.12	0	0	0	0

Open the ***mzadan_mix_5_spectrum1_annotation.csv*** table. It contains the annotations for the first compound eluting at 1.564 min. We previously mentioned that the two candidates could be part of the same cluster. Table 4 and mutually explaining each other.

Table 5 show that the peaks at **114.0912** *m/z* and **131.11848** *m/z* are connected by a **loss of ammonia** (-17.02655) and an **ammonium adduct** (+17.02655). Therefore, the algorithm assigns the same likelihood score to both, since both are mutually explaining each other.

***Table 5**: Segment of the annotation table generated for the first mass spectrum (adducts).*

			NH4-H (17.02655)			
m/z	intensity (cps)	intensity (%)	m/z	intensity (cps)	intensity (%)	error (u)
100.07429	6866.3	4.78	0	0	0	0
114.0912	**42660.5**	**29.69**	**131.11848**	**143673.9**	**100**	**7.30E-04**
131.11848	143673.9	100	0	0	0	0
169.07318	3100.5	2.16	0	0	0	0
191.16418	1583.2	1.1	0	0	0	0
217.10614	3447.5	2.4	0	0	0	0
261.22876	1602.7	1.12	0	0	0	0

To find which candidate is the most likely [M+H]+, we must consider other annotations. The peak at 131.11848 has a dimer at 261.22876 with a relative intensity of 1.12% and a potassium adduct at 169.07318 with a relative intensity of 2.16. While the peak at 114.0912 has no additional annotation. Therefore the peak at 131.11848 is a more likely [M+H]+ candidate than the peak at 11.0912. Knowing that the eluting compound at *1.564 min* corresponds to the ***N-Acetylputrescine*** (Table 1), we can conclusively annotate the peak at 114.0912 *m/z* as the ammonium loss of the $[M+H]^+$ at 131.11848.

4. Custom Annotations Sets

As previously mentioned (page 5), you have the option to import your own annotation set in mzAdan. Building the latter must comply with the rules listed below:

1. The ***table*** must contain exactly two columns and at least one row.
2. The ***first row*** must be the ***header***.
3. The ***header*** must include the following items: ***name*** and ***mass***.
4. The ***header items*** must be written in ***lowercase*** and ***without spaces***.
5. The ***mass column*** must only contain ***positive numbers***.
6. The ***table*** must be saved as a ***comma-separated values*** (.csv) document.

Table 6: *Example of three annotation tables created with Microsoft Excel.*

name	mass
CO	27.99491
H20	18.01056
NH3	17.02655

Neutrals

name	mass
NH4-H	17.02655
Na-H	21.98195
K-H	37.95589

Adducts

name	mass
C13	1.00336

Isotopes

To build a new annotation set, we recommend using either ***Microsoft Excel*** or ***Apache OpenOffice***. Open a new spreadsheet, then create the header by writing "***name***" and "***mass***" in the first row. To create a new annotation, simply add its name and mass in the first column and the second column respectively. Use a new row for each additional annotation. When finished, save the table as a ***comma-separated values*** (.csv) document. To prevent redundancy of annotations, the combinations of adducts, neutral losses and isotopes annotations should be avoided when possible (E.g. $[M+Na-H2O]^+$).

APPENDIX:

***Table** 7: Files included in the zip archive with mzAdan.jar.*

Type	file name	folder	description
Input	mzadan_mix_5.mgf	data	Spectra of the 5 compounds
	neutrals.csv	annotations	neutrals annotation table
	adducts.csv	annotations	adducts annotation table
	isotopes.csv	annotations	isotopes annotation table
Output	mzadan_mix_5_spectrum1_annotations.csv	output/annotations	Annotations table of the 1st spectrum
	mzadan_mix_5_spectrum2_annotations.mgf	output/annotations	Annotations table of the 2nd spectrum
	mzadan_mix_5_spectrum3_annotations.mgf	output/annotations	Annotations table of the 3rd spectrum
	mzadan_mix_5_spectrum4_annotations.mgf	output/annotations	Annotations table of the 4th spectrum
	mzadan_mix_5_spectrum5_annotations.mgf	output/annotations	Annotations table of the 5th spectrum
	mzadan_mix_5_spectrum1_compounds.mgf	output/compounds	Compounds table of the 1st spectrum
	mzadan_mix_5_spectrum2_compounds.mgf	output/compounds	Compounds table of the 2nd spectrum
	mzadan_mix_5_spectrum3_compounds.mgf	output/compounds	Compounds table of the 3rd spectrum
	mzadan_mix_5_spectrum4_compounds.mgf	output/compounds	Compounds table of the 4th spectrum
	mzadan_mix_5_spectrum5_compounds.mgf	output/compounds	Compounds table of the 5th spectrum
	mzadan_mix_5_summary.mgf	output/compounds	Summary table of the 5 compounds table
	mzadan_mix_5_spectrum1_clusters.mgf	output/clusters	Graph of the 1st spectrum
	mzadan_mix_5_spectrum2_clusters.mgf	output/clusters	Graph of the 2nd spectrum
	mzadan_mix_5_spectrum3_clusters.mgf	output/clusters	Graph of the 3rd spectrum
	mzadan_mix_5_spectrum4_clusters.mgf	output/clusters	Graph of the 4th spectrum
	mzadan_mix_5_spectrum5_clusters.mgf	output/clusters	Graph of the 5th spectrum