MzAdan Annotation Software Tutorial

Thomas Stricker[1], Frédérique Lisacek[2] and Gérard Hopfgartner[1]

1. Life Sciences Mass Spectrometry (LSMS), Department of Inorganic and Analytical Chemistry, University of Geneva, Geneva, Switzerland
2. Proteome Informatics Group (PIG), Swiss Institute of Bioinformatics and University of Geneva, Geneva, Switzerland

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

Swiss Institute of Bioinformatics

Contents

Introduction

A great diversity of adducts, as well as fragments and oligomers are generally produced during electrospray ionisation (ESI). These species are frequently used for a more reliable identification of putative protonated forms, preceding tandem mass spectrometry (MS/MS) spectral library matching. The main 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 or mass-to-charge ratio (m/z). The current version of mzAdan works on singly charged ions in both positive and negative mode acquisition. The software can process a single spectrum or multiple spectra in the *mgf* format.

Prerequisite

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

The zip archive includes three folders. The data folder contains the dataset described in Table 2.1. The output folder contains the annotations tables, candidates' tables and graph files generated by mzAdan as a result of the annotation of the previously mentioned dataset. Lastly, the annotations folder includes three tables containing annotations sets that haven been specifically formatted for their use with mzAdan. To launch the application, simply double-click on the mzAdan.jar file or on the mzAdan.bat file if you are using Windows.

Table 2.1: Mass spectra included in the dataset.

index	rt (min)	compound	exact mass (amu)
1	1.564	N-Acetylputrescine	130.11121
2	4.578	Ethenodeoxyadenosine	275.10335
3	7.57	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 (Table 2.1) consists of five high resolution full scan mass spectra of five unique metabolites. The data were aquired on a TripleTOF 6600 (Sciex, Concord ON, Canada) in positive mode acquisitionn coupled to UltiMate 3000 RSLC high performance chromatography system (Dionex, Sunnyvale CA, United States).

Tutorial

3.1. Project

3.1.1 Export Panel

The *Export* panel (Figure 3.1) 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 drop-down menu and select the *Directory* option. In the output field, you can choose the name and the base directory of the output 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, you will be practicing on the dataset described above (Table 2.1), which consists of a single file including five high resolution mass spectra.

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 you to select the mass spectra. Select the *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 output directory.

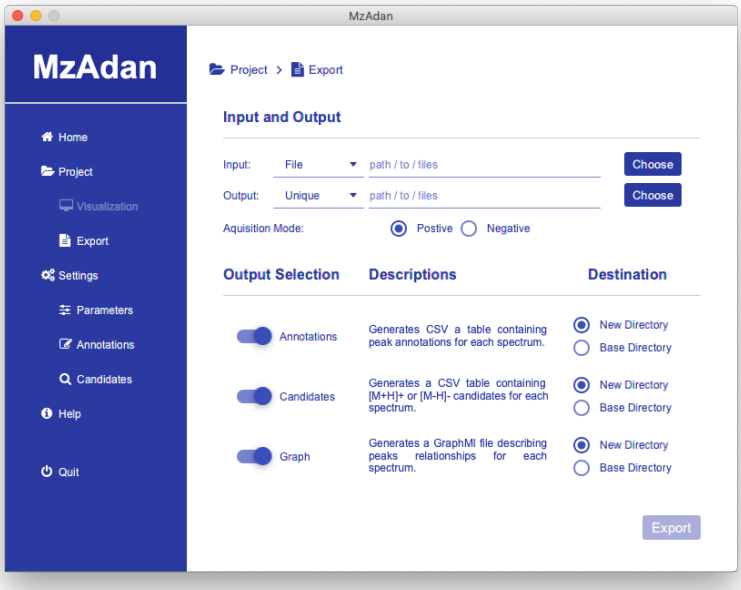

Figure 3.1: A new project can be created through the export panel. Mass spectra can be imported as single or multiple *mgf* files and the results exported as *csv* and *graphml* files.

Note that the *Export* button located on the bottom right side of the panel changed colour indicating that the input can now be processed. However, before starting the analysis a few setting options located in the *parameters*, *annotations* and *candidates* panels will need to be changed.

3.2. Settings

The three setting panels are accessible via the side menu bar or via the *Settings* panel itself. The latter features a description detailing the content of each setting panel.

3.2.1 Parameters Panel

The *Parameters* panel (3.2) contains filtering options that can be applied to both spectra and 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 discard peaks with m/z outside the set range. The *Deisotoping* filter removes

isotopic peaks with relative intensities outside the set range. The *Validation* option filters out peaks, which are not connected to any detectable isotopes in the raw data. The *Tolerance* filter discards annotations with an absolute error (mmu) higher than the set value.

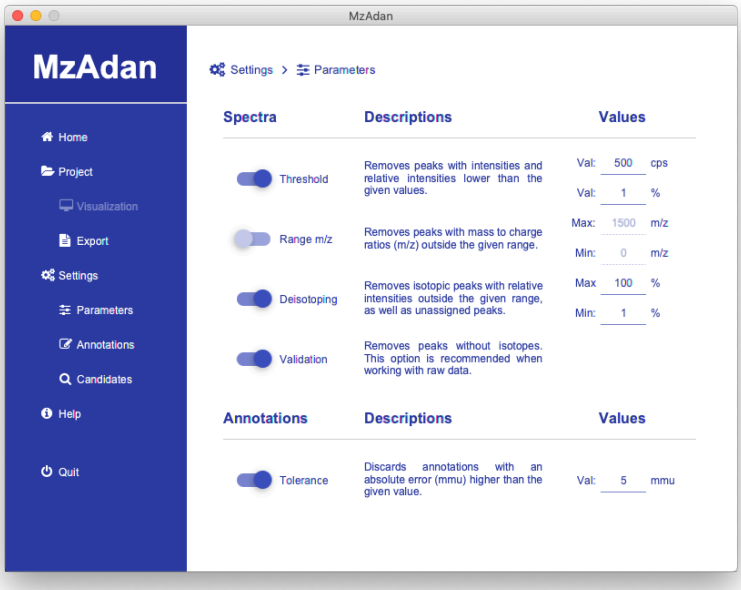

Figure 3.2: The parameters panel contains options to filter out peaks from the mass spectrum based on their intensities, mass-to-charge ratios and assignments (monoisotopic, isotopic and unassigned). It also contains an option to validate or discard annotations based on mmu error.

In this tutorial, you will use the default parameters except for the Range m/z filter, which can be disabled by clicking on the corresponding toggle button. The deisotoping and validation filter are used to remove isotopic and unassigned peaks from the mass spectrum. Enable the deisotoping and validation options by clicking on their respective toggle button.

3.2.2 Annotations Panel

The Annotations panel (Figure 3.3) 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. Clicking it will prompt a new file selection window, which allows you to import a custom annotation set. The third and last button is the *Reset* button. Pressing it resets the whole annotation set to its initial items. In the *Advanced* section, the annotation of multimers can be enabled.

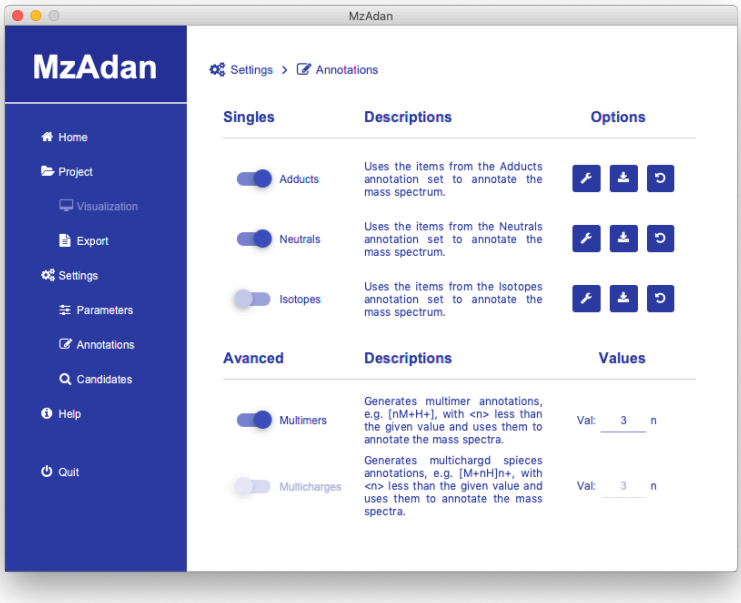

Figure 3.3: The annotation panel contains options to fully customize the adducts, neutral losses and isotopes annotations sets. Annotations can be individually enabled or disabled and custom sets can be imported. The Annotation of Multimers is also available.

Enabling the deisotoping filter discussed previously, will filter out C^{13} isotopes from mass spectra. Therefore, annotation of the latter should be manually disabled. There are two ways to do this. Either click on the *Edit* button (wrench icon) located to the right of the Isotopes description and disable the C^{13} annotation by selecting it in the *Enabled* column and pressing the left arrow. Or, simply disable the whole set by clicking on the *Isotopes* toggle button, as this set contains a single annotation.

3.2.3 Candidates Panel

The Candidates panel (Figure 3.4) contains options allowing you to filter out $[M+H]^+$ candidates from to the candidates tables. The upper section includes options to discard candidates based on their intensity and connectivity. The bottom section contains options for filtering clusters[1] based on intensity, connectivity and content. These filters will be further discussed below.

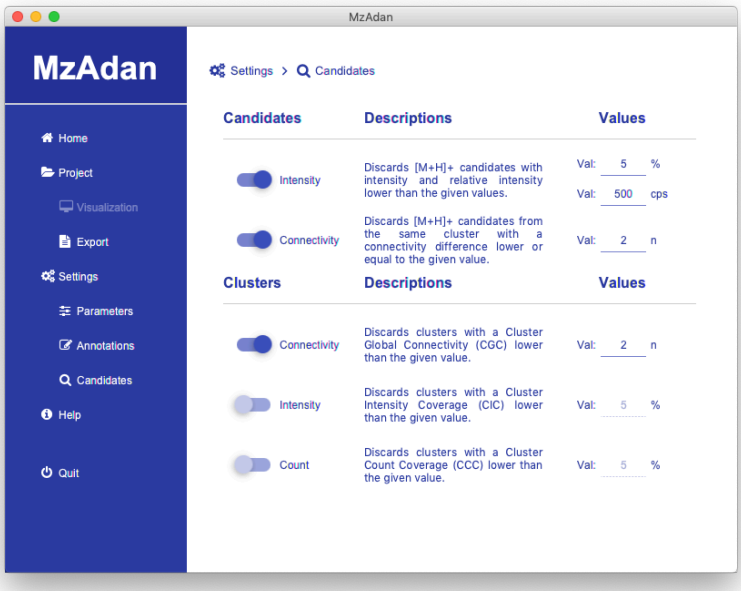

Figure 3.4: The identification panel contains options to discard $[M+H]^+$ candidates based on their intensity and connectivity, or based on their cluster connectivity, intensity and peak count.

In this tutorial, you will exclude candidates with an intensity lower than 500 cps and 5% respectively. You will also discard $[M+H]^+$ candidates from the same cluster with identical scores but lower connectivity. Finally, you will filter out clusters containing less than two peaks. To do so, first enable the *Intensity* and *Connectivity* filters in the *Candidates* section and set the minimum intensity to 500 cps and 5%, as well as 2 for the candidate connectivity threshold [2]. Then, enable the *Connec-*

[1] Networks of interconnected peaks.

[2] The candidate connectivity filter discards candidates whose connectivity is lower by more than n connections compared to the best scoring and most connected candidate of the cluster.

tivity filter in the *Cluster* section and set the minimum value to 2 [3]. When you are done, go back to the *Export* panel and press the *Export* Button.

3.3. Output

3.3.1 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 (https://gephi.org/).

The content of the compounds files is compiled in a summary table and structured as in Table 3.1 and Table 3.2. 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. $[M+H]^+$ candidates are stored in individual row. 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 3.1: Left segment of the summary table.

file	rt	index	id	mass	m/z	int (cps)	int (%)
...	**1.564**	**1**	**C0**	**130.1112**	**131.1185**	**143674**	**100**
...	**1.564**	**1**	**C0**	**113.0839**	**114.0912**	**42661**	**29.7**
...	4.578	2	C0	275.1034	276.1106	169659	100
...	4.578	2	C1	216.0993	217.1066	8792	5.2
...	7.57	3	C0	146.0579	147.0652	24910	18.8
...	10.363	4	C0	207.0904	208.0977	82496	100
...	10.363	4	C1	216.0998	217.1071	14321	17.4
...	10.363	4	C1	233.1258	234.133	4616	5.6
...	17.359	5	C0	515.2931	516.3003	20853	6.8

The last three columns of the table contain three indexes that may be used to discriminate likely $[M+H]^+$ from noise and fragment ions. The cluster global connec-

[3] The cluster connectivity filter discards clusters, which contain less than n interconnected peaks.

tivity index (*cgc*) represents the number of peaks connected directly or indirectly to the peak of interest. The cluster intensity coverage index (*cic*) corresponds to percentages of the Total Ion Current (TIC) explained by a cluster. Lastly, the cluster count coverage index (*ccc*) corresponds to percentages of the spectrum peaks contained in a cluster.

Table 3.2: Left and right segments of the summary table.

file	rt	index	id	mass	connectivity	cgc	cic	ccc
...	**1.564**	**1**	**C0**	**130.1112**	**3**	**4**	**94.14**	**57.14**
...	**1.564**	**1**	**C0**	**113.0839**	**1**	**4**	**94.14**	**57.14**
...	4.578	2	C0	275.1034	2	3	78	21.43
...	4.578	2	C1	216.0993	2	3	5.21	21.43
...	7.57	3	C0	146.0579	3	7	71.38	33.33
...	10.363	4	C0	207.0904	3	7	57.72	20.59
...	10.363	4	C1	216.0998	2	3	8.51	8.82
...	10.363	4	C1	233.1258	1	3	8.51	8.82
...	17.359	5	C0	515.2931	4	7	89.67	53.85

Open the *mzadan_mix_5_summary.csv* table located in the candidates folder. One or more ions may be annotated as $[M+H]^+$ candidates in each mass spectrum. Note that among these some are part of the same cluster **C0** (Table 3.1 and 3.2). Ideally, the software should provide a single candidate per cluster. However, this may not be feasible if two or more candidates explain the same percentage of the spectrum TIC. This specific situation we will further discussed below using the annotations tables.

3.3.2 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 contained 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, such as it m/z and intensities, as well as the quality of the match (error).

Table 3.3: Segment of the annotation table of the first spectrum (ammonium loss).

			NH3	**(-17.02655)**		
m/z	**int (cps)**	**int (%)**	**m/z**	**int (cps)**	**int (%)**	**error (amu)**
100.0743	6866	4.8	0	0	0	0
114.0912	42661	29.7	0	0	0	0
131.1185	**143674**	**100.0**	**114.0912**	**42661**	**29.7**	**7.30E-04**
169.0732	3101	2.2	0	0	0	0
191.1642	1583	1.1	0	0	0	0
217.1061	3448	2.4	0	0	0	0
261.2288	1603	1.1	0	0	0	0

Open the *mzadan_mix_5_spectrum1_annotation.csv* table. The file contains the annotations for the first compound eluting at 1.564 min. We previously mentioned that candidates from the same cluster may be present in the $[M+H]^+$ canditates table if they explain the same percentage of the total ion current. According to the annotation table (Table 3.3 and 3.4), the peaks at 114.0912 m/z and 131.1185 m/z are inter-connected via a loss of ammonia (-17.02655 amu) and an ammonium adduct (+17.02655 amu). Since these peaks are mutually explaining each other, the intensity-based ranking system cannot identify the more likely candidate.

Table 3.4: Segment of the annotation table of the first spectrum (ammonium adduct).

			NH4-H	**(+17.02655)**		
m/z	**int (cps)**	**int (%)**	**m/z**	**int (cps)**	**int (%)**	**error (amu)**
100.0743	6866	4.8	0	0	0	0
114.0912	**42661**	**29.7**	**131.1185**	**143674**	**100**	**7.30E-04**
131.1185	143674	100.0	0	0	0	0
169.0732	3101	2.2	0	0	0	0
191.1642	1583	1.1	0	0	0	0
217.1061	3448	2.4	0	0	0	0
261.2288	1603	1.1	0	0	0	0

In order to find the more probable $[M+H]^+$ of the two, you must consider additional information, more specifically their connectivity. One the one hand, the peak 114.0912 m/z is uniquely connected to the peak at 131.1185 m/z via an ammonium adduct. On the other hand, the peak at 131.1185 m/z is connected to the peak at 114.0912 m/z and another peak at 169.0732 m/z via respectively an ammonium loss and a potassium adduct. According to their respective connectivity the peak at 131.1185 is a more likely $[M+H]^+$ candidate of the cluster **C0**. Looking back at table 2.1, we can indeed confirm that the latter is the protonated form of the N-Acetylputrescine. Less likely $[M+H]^+$ can automatically be discarded via the candidate connectivity threshold (Figure 3.3). Before ending this tutorial, reprocess the dataset and gradually lower the value of the latter [4] from 2 to 1 or 0. Observe its effect on the size and content of the candidates tables.

3.4. Custom Annotations Sets

As previously mentioned, you have the option to import your own annotation sets in mzAdan. However, in order for tool to correctly read and use them, their creation must comply with the rules listed below:

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

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.

[4] Maximum difference of connectivity between the best ranking candidates of the cluster.

Table 3.5: Three exemples of annotation tables.

Neutrals		**Adducts**		**Isotopes**	
name	**mass**	**name**	**mass**	**name**	**mass**
CO	27.99491	NH4-H	17.02655	C13	1.00336
H20	18.01056	Na-H	21.98195		
NH3	17.02655	K-H	37.95589		

To prevent redundancy the combinations of adducts and/or neutral losses should be avoided if possible (E.g. $[M+2K-H]^+$, $[M+Na-H_2O]^+$), as mzAdan favors a sequential and combinatorial annotation approach.

Appendix

Table 4.1: Input files included in the zip archive with mzAdan.jar.

file	folder	description
mzadan_mix_5.mgf	data	five compounds spectral file
neutrals.csv	annotations	neutrals annotation table
adducts.csv	annotations	adducts annotation table
isotopes.csv	annotations	isotopes annotation table

Table 4.2: Output files included in the zip archive with mzAdan.jar.

file	folder	description
mzadan_mix_5_spectrum1 _annotations.csv	output/annotations	annotations table (1st spectrum)
mzadan_mix_5_spectrum2 _annotations.csv	output/annotations	annotations table (2nd spectrum)
mzadan_mix_5_spectrum3 _annotations.csv	output/annotations	annotations table (3rd spectrum)
mzadan_mix_5_spectrum4 _annotations.csv	output/annotations	annotations table (4th spectrum)
mzadan_mix_5_spectrum5 _annotations.csv	output/annotations	annotations table (5th spectrum)
mzadan_mix_5_spectrum1 _candidates.csv	output/candidates	candidates table (1st spectrum)
mzadan_mix_5_spectrum2 _ candidates.csv	output/candidates	candidates table (2nd spectrum)
mzadan_mix_5_spectrum3 _ candidates.csv	output/candidates	candidates table (3rd spectrum)
mzadan_mix_5_spectrum4 _ candidates.csv	output/candidates	candidates table (4th spectrum)
mzadan_mix_5_spectrum5 _ candidates.csv	output/candidates	candidates table (5th spectrum)
mzadan_mix_5_summary.csv	output/candidates	summary table (5 candidates)
mzadan_mix_5_spectrum1 _graphs.csv	output/graphs	graph (1st spectrum)
mzadan_mix_5_spectrum2 _ graphs.csv	output/graphs	graph (2nd spectrum)
mzadan_mix_5_spectrum3 _ graphs.csv	output/graphs	graph (3rd spectrum)
mzadan_mix_5_spectrum4 _ graphs.csv	output/graphs	graph (4th spectrum)
mzadan_mix_5_spectrum5 _ graphs.csv	output/graphs	graph (5th spectrum)