FAST MS Manual

Michael Palasser
Email: michael.palasser@bachem.com

January 7, 2025

Contents

1 Introduction

This manual was originally written in 2021 and, to some extent, updated in 2024. Please note that deprecated or incorrect sections may still be present. The most accurate information can be found in the peer-reviewed paper (`https://pubs.acs.org/doi/10.1021/jasms.4c00236`). Additionally, you can have a look at my dissertation (`https://ulb-dok.uibk.ac.at/ulbtirolhs/download/pdf/8051320`).
Data analysis in high-resolution mass spectrometry can be a very challenging and laborious task. While one spectrum can be recorded within minutes, its calibration, the peak and ion assignment, and the subsequent analysis of the ion list can take multiple days. Though there are many prominent MS data analysis programmes for proteomic data, software solutions working for mass spectra of RNA or DNA are very rare. Peak deconvolution algorithms such as the "Sophisticated Numerical Annotation Procedure" (Bruker) can massively speed up the assignment and also work for RNA. However, the algorithm usually misses many low abundant ions which is a problem, especially in top-down mass spectrometry. For locating heterogenous post-translational modifications, ligand binding site mapping, or characterising products from labelling reactions, these low abundant ions can be extremely important. The problem increases with the size of the precursor since the signal to noise ratios of the fragments decrease accordingly. Therefore, a laborious manual revision of the spectrum is often necessary. However, there can be problems with reproducibility since the user is prone to find ions within the noise that would be good for the results.
The software package FAST MS (**F**ree **A**nalysis **S**oftware for **T**op-Down **M**ass **S**pectrometry) was designed to overcome these problems. It should make the life of users in mass spectrometry easier by speeding up data analysis and making the results more reproducible. It is open-source, cross-platform, and has a user-friendly graphical user interface. Additionally, FAST MS is very universal:

- All kinds of polymers can be analysed (e.g. RNA, DNA, proteins, user-defined polymers). However, fragments of side chains or branches cannot be sequenced.
- The elemental composition of the molecules is not bound to C, H, N, O, S, and P.
- The polymer can be isotopically depleted or enriched.
- The building blocks (nucleotides, amino acids, ...) can be user-defined.
- It works in positive and negative spray mode.

The most important requirement is that the user has to know the sequence of the molecule. Sequencing unknown molecules combined with a database search is not possible.

2 Citing FAST MS

Please cite the FAST MS paper when using the programme:

FAST MS: Software for the Automated Analysis of Top-Down Mass Spectra of Polymeric Molecules Including RNA, DNA, and Proteins
Michael Palasser and Kathrin Breuker
Journal of the American Society for Mass Spectrometry Article ASAP
DOI: 10.1021/jasms.4c00236

3 Version History

Current version: 1.0.0

4 Requirements and Installation

When using the Windows executable version, the OS version should not be older than Windows 7. Otherwise, an up-to-date version of Python (min 3.7) must be installed. Python can be downloaded from `https://www.python.org/downloads/` or by installing Anaconda (`https://www.anaconda.com/products/individual`). To automatically install and update all required site-packages (numpy, scipy, matplotlib, pandas, PyQt5, pyqtgraph, xlsxwriter, logging, multiprocessing, numba, math, sqlite3, and tqdm), navigate (cd ...) in the terminal into the FAST-MS directory and enter either "python Start.py" or "python3 Start.py".

5 First Steps

5.1 Starting the Programme

There are various ways to start the programme. If you have the executable Windwos version, you just need to double-click on the exe file to start the programme. If you want to directly use the source code, you can do the following:

1. **In Mac OS**: Double-click on "Start.command" in the main folder of FAST MS. This file itself must not be moved to another directory but you can make a shortcut and move the shortcut wherever you like.

2. Start the **terminal** (Mac/Linux) or the **Command Prompt** (Windows). Go (cd) into the FAST MS folder and type "python Start.py".

6 List of Tools in the Package

Top-down MS Tools:

- Analyse top-down spectrum: find and assign ions in a peak list of a top-down spectrum and analyse the ion list (see 10).
- Localise Modification: tool which calculates the modified proportions for each fragment (% modified/ligand/loss) based on a list of assigned ions (see 11.3)

Intact MS Tools:

- Analyse intact ion spectrum: Find and assign ions in a peak list of an intact MS spectrum and analyse the ion list (see 10).
- Assign intact ions: Assign ions in lists of unassigned ions of (multiple) full ESI-MS spectra. The results are exported to Excel (see 11.1).

Other MS Tools:

- Model Ion: Tool to model isotope patterns of a molecule based on a molecular formula or sequence (see 11.2).
- Compare ion lists: Tool to compare lists of assigned ions of different intact ion or top-down MS spectra (see 11.4).

7 Edit Data and Templates

The workflows for editing, creating, and deleting most of the template types are very similar:

- **Create a new empty template**: Open "–New–", edit the values and click "Save" or "Save As"
- **Create a new template using the values of an existing**: Open the template you want to copy, change the name and the values which should be altered and click "File > Save As".
- **Deleting**: Open one of the template, click "File > Delete" and choose the template you want to delete. After confirmation, the template is permanently deleted.
- **Editing the values of a template**: The options for creating and deleting rows can be accessed by right-clicking on the table.

Templates:

- **Elements and Isotopes**: Change the values of "rel. ab." for isotopically depleted or enriched molecules.
  Access by: "Edit Data > Elements"
- **Molecules and Building Blocks**: "Gain" and "Loss" are used to calculate the start formula of the molecule (see 12.1).
  Access by: "Edit Data > Molecules". The translation is an optional comment (e.g. the HELM code of the building block).

- **Sequences**: The menu for sequences differs from the others. A new sequence is created by right-clicking on the table and choosing "Insert row" or "Copy and insert row". You can delete an entry by right-clicking on the sequence entry and choosing "Delete row". These changes can be saved by clicking "File > Save".
  Access by: "Edit Data > Sequences"

- **Fragments**: These are the templates for calculating the fragment and precursor ions in the top-down analysis. Choose the correct precursor in the box. For ECD and EDD, the correct precursor is a radical species. Note that there are two tabs.
  - **Name**: If, for example, a c-fragment looses a H_2O molecule, write "c-H2O" as "Name" of the template. Add a "+" for adducts etc.
  - **Gain** and **Loss**: The molecular formulas in the columns "Gain" / "Loss" will be added/subtracted from the normal formula. Thereby, the expressions "C10H10N10O2", "(C5H5N5O)2" and "C5H5N5OC5H5N5O" are equivalent.
  - **BB**: If the fragment can only occur if there is a certain building block incorporated in the sequence of the fragment (e.g. c-G can only occur if there is a guanine nucleobase in the sequence), write the name of the corresponding building block in the "BB" column. If the fragment can only occur at a certain building block (e.g. specific fragments at a modified nucleotide), add an "!" after the name of the nucleotide without any spaces. If it cannot occur next to an specific building block, add an "!" before the building block name. If the can only occur before a specific building block, write a "+" after the building block name.
    Add "," as delimiter between building blocks in case you want to enter more than one.
  - **radicals**: If the fragment contains radicals, write the number into the "radical" column.
  - **direction**: Can be "1" for forward (N-/5'-terminus) or "-1" for backward (C-/3'-terminus).
  - **enabled**: The fragment/precursor fragment will only be calculated if the box in column "enabled" is ticked.

  Access by: "Edit > Fragments"

- **Modifications**: These are the templates to generate modified fragments in the top-down analysis. The modification is internally stored under the name "Name". The Value which is stated as "Precursor Modification" is the modification of the precursor.
  - **name**: Write a "+" in front of every modification name or a "-" for every loss in "Name".
  - **gain**, **loss**, **BB**, **radicals**, **enabled**: See above in "Fragments"

- **charge**: If a modification is charged and therefore alters the charge of a modified fragment, state in the extent of charge alteration. The charge of negatively charged modifications must be negative.
- **localise?**: It is possible to exclude a modification from localising (e.g. if the species was only unspecifically co-isolated). Therefore, this box must be unticked.

Access by: "Edit > Modifications"

- **Intact Ions**: These are the templates to generate ions in the intact ion analysis.
  - **name**, **gain**, **loss**, **radicals**, **enabled**: See above
  - **no.mod**: State how often the species is modified with the modification of interest. It does not matter if the modification is covalent or a ligand.

Access by: "Intact Ions > Edit Intact Ions"

8 Inputs

8.1 Peak Lists

Most manufacturers of mass spectrometers provide a peak picking software. Thus, including a peak picking function in FAST MS was not considered as necessary. The peak list can be stored in a plain text or a CSV file. It must be a 2-dimensional list of peaks with (at least) two columns separated by a tab (text file) or a comma (CSV file): m/z, intensity. For correct noise calculation, a significant number of noise peaks must be present in the final peak list. Therefore, Thermo's denoising algorithm can lead to inaccuracies.
Since all search tools of FAST MS include auto-calibration, it is not compulsory to calibrate the peak list. To use this functionality, an unassigned ion list of the same (uncalibrated) spectrum is necessary (see below). In Data Analysis (Bruker), a peak list can be generated using the FTMS or SumPeak options (see Fig. 1). Thereby, there are 3 parameters:

1. **S/N threshold**: Use 0.5-1.0, depending on the number of acquired scans. Use a higher threshold if more scans were acquired. A higher threshold leads to fewer picked peaks especially at higher m/z

2. **Relative intensity threshold (base peak)**: similar to the absolute intensity threshold

3. **Absolute intensity threshold**: Important to adjust. A higher threshold leads to fewer picked peaks, especially, at lower m/z. For correct noise calculation, the value should not be set too high. A significant number of noise peaks must be present in the final peak list.

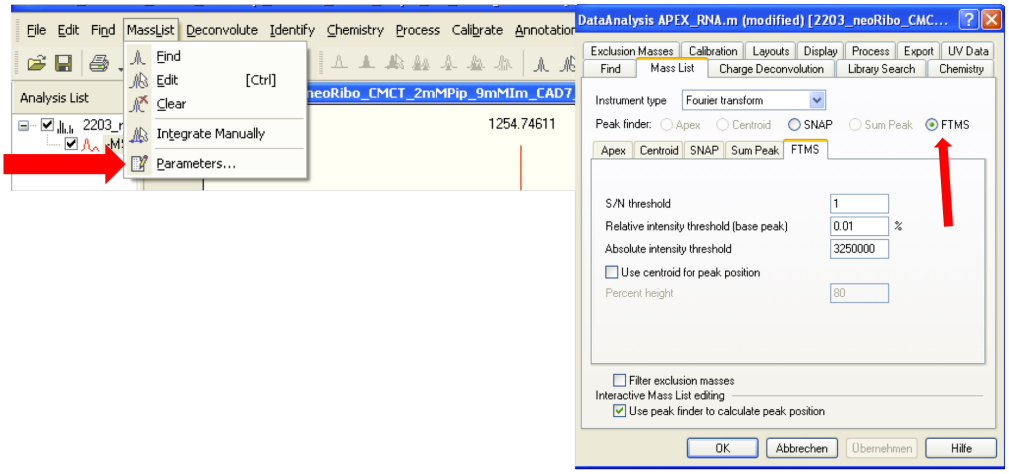


Figure 1: Peak picking in Data Analysis

The first column in the peak list must be the m/z values and the second the intensity values. Save the mass list in a text or CSV file.

8.2 Ion Lists

The corresponding file can be a plain text or a CSV file. In Data Analysis (Bruker), a list of unassigned ions can be generated using SNAP. The following columns must be included:

- m/z (monoisotopic), charge, relative abundance
- m/z (monoisotopic), charge, intensity

The columns must be separated by a tab (text file) or a comma (CSV file).

9 Configurations

It is recommended to check the configurations before starting the analysis. They can be accessed by choosing "Edit > Configurations" or "Python > Preferences" in Mac OS. The values are discussed below. The default values are written in parenthesis. Important parameters are coloured blue.

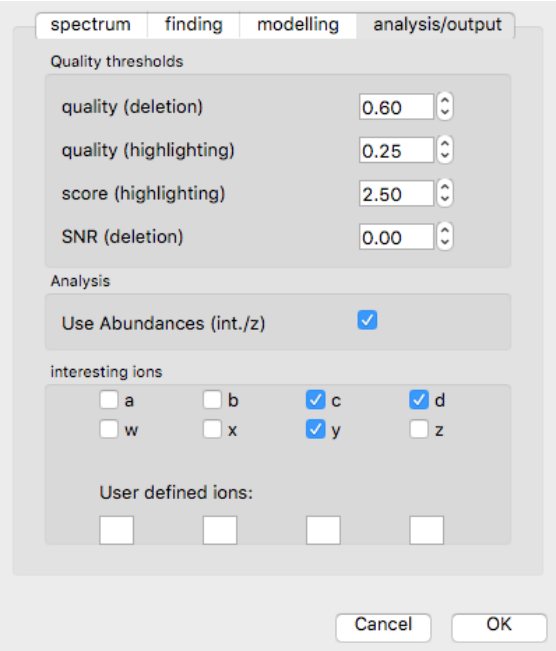


Figure 2: Configurations dialogue

- **Spectrum** tab:

- **min. m/z** (50): lower m/z bound - just peaks with higher m/z are examined (*lowerbound*)
- **min. max. m/z** (1500): minimal upper m/z bound. The programme starts here to find the end of the ion window (the spectral window containing non-noise-peaks). (*minUpperBound*)
- **tolerance** (100): value is added to the calculated upper m/z-bound for the final upper m/z-bound. Increasing the value may increase the investigated m/z window but can also make the programme slower. (*upperBoundTolerance*).
- **window size** (20): window size for noise calculation to find upper m/z bound (*upperBoundWindowSize*)

- **uncal. error threshold** (50): error threshold for uncalibrated spectrum (constant) (*errorLimitCalib*)
- **max. std. dev.** (1.5): max. allowed standard deviation of the errors after calibration. (*maxStd*)
- **overwrite peak list** (unticked): If ticked, the uncalibrated peak list will be overwritten. Otherwise, a new file will be created with the old filename + "_cal" (*overwrite*)

- **Finding** tab:
  - **charge tolerance** (0.8 for RNA/DNA, 1.5 for proteins): The programme searches for ions with charge states between the calculated charge +/- tolerance (*zTolerance*)
  - **error threshold k** (depends on the mass spectrometer): the slope of the ppm error threshold function (*k*)
  - **error threshold d** (depends on the mass spectrometer): intercept of ppm error threshold function (*d*)
  - **tolerance for isotope peaks** (2.0): tolerance for isotope peak search in ppm (*errorTolerance*)
  - **window size** (3): window size for noise calculation next to a found ion (*noiseWindowSize*)
  - **noise threshold tolerance** (0.45): The programme searches for all isotope peaks which should be above the calculated noise threshold multiplied by this factor. Decrease the threshold to search for more low abundant isotope peaks (*thresholdFactor*)

- **Modelling** tab:
  - **min. proportion** (0.996): (*maxIso*) the calculation of the isotope peaks will be stopped when their summed abundances are higher than the stated proportion

- **approximation** (20): the number of the last isotope peak that should be exactly calculated (see 12.2) (*approxIso*)
- **outlier peak threshold** (1.6): Isotope peaks with higher values are not used for intensity modelling. (*outlierLimit*)
- **max. nr. of overlapping ions** (5): If more ions are overlapping in one pattern, the programme will ask the user to manually delete ions. If you trust the fit algorithm you can set the value higher to save time. (*manualDeletion*)
- **overlap threshold** (0.8): The programme calculates factors to correct the intensities of overlapping ions. Ions with a lower factor are deleted (*overlapThreshold*)

- **Analysis/Output** tab:
  - **quality (deletion)**(0.5): ions that have a higher quality error are deleted (*shapeDel*)
  - **quality (highlighting)** (0.25): ions that have a higher quality error are highlighted (*shapeMarked*)
  - **score (highlighting)** (2.5): ions that have a higher score error are highlighted (*scoreMarked*)
  - **SNR (deletion)**(3): ions that have a lower signal to noise ratio are deleted (*shapeDel*)
  - **Use Abundances (int./z)** (ticked): If ticked, the relative abundances (intensity / z) are used for the quantitative analysis (recommended e.g. FT-ICR and orbitrap mass analysers). Otherwise, the intensities are used (recommended e.g. TOF-mass analysers). (*useAb*)
  - **Interesting Ions**: Tick the ions which should be used analysed in detail (calculating av. charges, occupancies, ...). If a user-defined ion is not included in the list, the user can add one manually. (*interestingIons*)

10 Quick Guide: Analyse a Spectrum

10.1 Starting the Analysis

For analysing a top-down spectrum, press either the button "Analyse top-down spectrum" or "Analyse Spectrum" in the "Top-Down" tab to start the analysis. Then the user has to define the settings of the search. Choose:

- Sequence Name: see 7
- Charge: charge of the precursor, can be positive or negative (!)
- Fragmentation: see 7
- Modifications: modification pattern, see 7

- No. of Modifications: number of incorporated modifications on the precursor
- Spectral Data: the path of the peak file (use the button on the right to select the file)
- Noise Threshold (x10^6): minimal noise level in the spectrum divided by 10^6 (It is not mandatory to enter a value. Just enter 0.000 in that case.) *sequence_fragmentation_no. of mod._modifications*
- Autocalibration & Ions for Cal.: If the peak list should be calibrated, a list of unassigned ions from the same spectrum has to be provided (see 8). Select the file with the ion list in "Ions for Cal.".

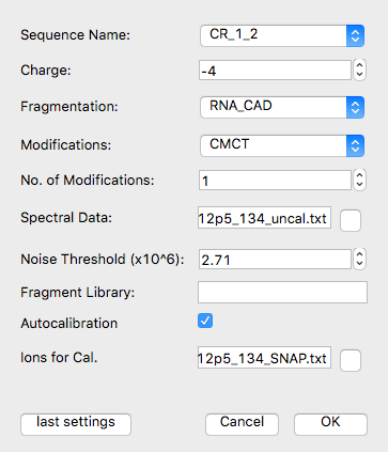


Figure 3: Settings in top-down MS analysis

Choose the option "Analyse Spectrum" in the "Intact Ions" tab to start the analysis of an intact ion spectrum. Settings:

- Sequence Name: see 7
- Modifications: intact ion modification pattern, see 7
- Spectral Data: the path of the peak file (use the button on the right to select the file)
- Noise Threshold (x10^6): minimal noise level in the spectrum divided by 10^6 (It is not mandatory to enter a value. Just enter 0.000 in that case.)
- Min. m/z & Max. m/z: All ions outside of this m/z range are neglected.
- Autocalibration & Ions for Cal.: If the peak list should be calibrated, a list of unassigned ions from the same spectrum has to be provided (see 8). Select the file with the ion list in "Ions for Cal.".

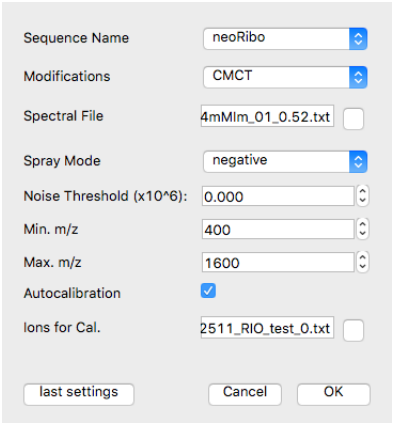


Figure 4: Settings in intact ion search

10.2 Running the Analysis

After pressing the "OK" button of the corresponding dialogue, the search will run the following workflow:

1. **Calculating the neutral library**: The programme calculates the molecular formulas of the neutral species:
   - top-down: fragments and the precursor ions (top-down) based on the sequence and the user-defined fragmentation and modification templates
   - intact: neutral molecules based on the sequence and the user-defined intact ion templates

2. **Calculation of the isotope distributions**: Depending on the sizes and the elemental composition of the neutral species, this step can take several minutes. Due to the larger size of the fragment library, this is especially true for the top-down MS search. However, the isotope distributions are saved in the top-down search. If the user runs again an analysis for the same sequence, fragmentation pattern and modification, the programme will access the stored values and the calculation is not repeated.

3. **Truncating the peak list**: All peaks with a lower m/z than the min. m/z parameter (see "min. m/z" in 9) are discarded. The programme calculates a noise level for certain m/z windows. If there are (almost) no peaks above that level in the window, it discards all peaks with a higher m/z (= upper m/z bound).

4. **Autocalibration**: If the autocalibration option was ticked, the programme calculates the monoisotopic m/z of the ions of the priorly calculated species and tries to find the corresponding values in the list of the unassigned ions. Then, the quadratic calibration function is calculated for the found ions and a dialogue pops up whereby the user can select/deselect ions to use them for the calibration (see

Fig. 5). The calibration can then be updated by pressing the "Update" button. After pressing "Ok", the peak list is calibrated. In case the "Overwrite" option is ticked (see 9), the peak list will be overwritten with the calibrated list. Otherwise, a new file is created.

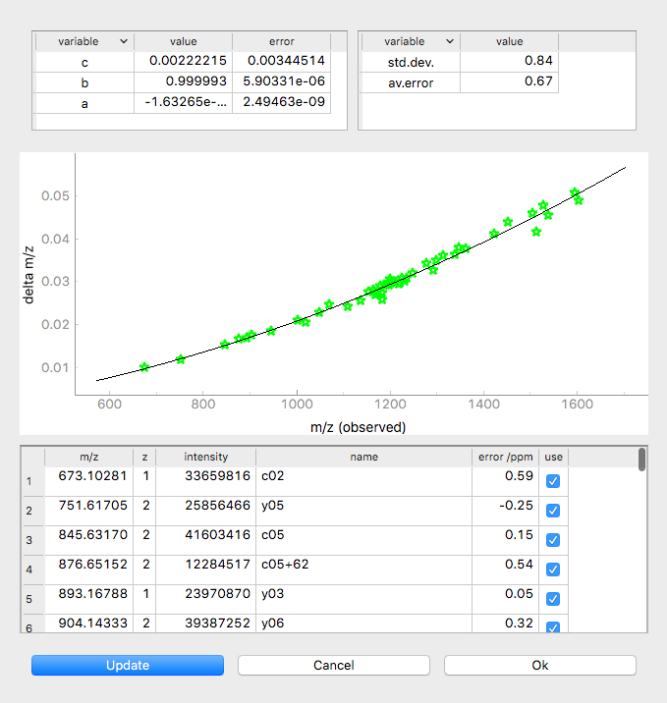


Figure 5: Calibration of the peak list

5. **Search for isotope peaks**: The programme again calculates the most probable charge states for each neutral species and calculates the corresponding m/z values of the isotope peaks. Then, it tries to find the most abundant peaks and calculates which peaks could be above the noise. The search is then continued for these peaks. If all of the calculated peaks are below the noise level, the ion is deleted ("noise" in comment).

6. **Modelling the isotope distribution**: The calculated intensities of the isotope peaks are modelled to the observed ones. If one peak shows a much higher intensity than expected (see "outlier peak threshold" in 9), it is detected as an outlier and not used for modelling. Similarly, isotope peaks, which have a completely different ppm error than the others, are also detected as outliers. If the quality of an ion is worse than the threshold (see "quality (deletion)" in 9) or all peaks of the ion are considered as outliers the ion is deleted ("qual." in comment). The same happens if the ppm error of an ion is above the ppm threshold ("error" in comment).

7. **Finding and correcting overlapping ions**: The programme searches for over-

lapping ions. In case there are isomeric or almost isomeric ions, the user will be asked which one to keep (see Figure 6). The others will be deleted ("iso" in comment). In other cases, the combined isotope peak distribution of all found ions will be fitted by a linear combination of all overlapping ions. If there are too many ions overlapping with each other (for the threshold see "max. nr. of overlapping ions" in 9), the user will be asked if certain ions should be deleted ("man.del." in comment) to increase the precision of the modelling (see Figure 7). To make a better decision, it is possible to view the ions in the spectrum (see 10.3.2) or the values of the peaks by right-clicking on the corresponding ion.
After clicking "OK", the correction factors will be calculated for each overlapping ion. If this factor is below a threshold (see "threshold" in 9), the ion is deleted ("low:" with factor in comment). Otherwise, the intensities of the isotope peaks and the total intensity of the ion will be multiplied by the factor. To consider undetected overlaps with unknown species (e.g. unidentified precursor fragments or internal fragments), the maximum factor is 1.05. The corresponding ions will be corrected by this value ("high:" with factor in comment).
The original values of the ions before modelling the overlaps can be later found by choosing "Show > Original Values" in the menubar.

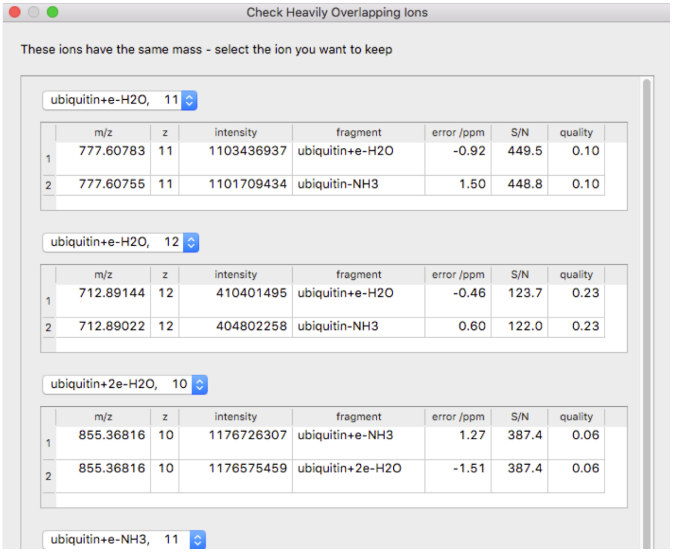


Figure 6: Menu for selecting ions that have (almost) the same mass and the same charge

8. **The Result Window**: A window opens showing all assigned and deleted ions (see below).

10.3 Results

The result window shows the lists of the assigned and deleted ions. The tables have the following columns:

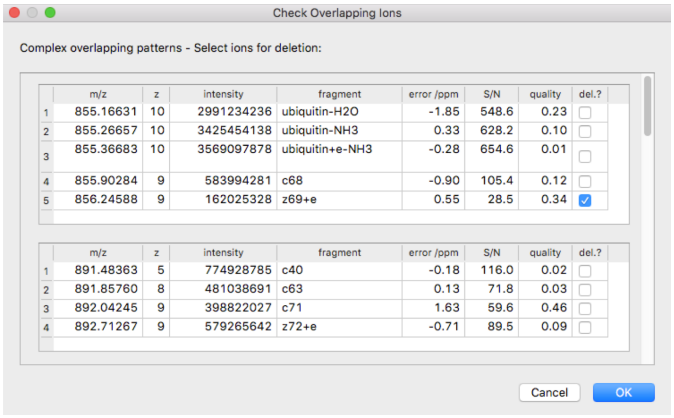


Figure 7: Menu for deleting overlapping ions

- **m/z**: (calculated) m/z value of the monoisotopic peak:

$$m/z_m = m/z_{m,theo} \cdot (1 + error \cdot 10^{-6}) \tag{1}$$

whereby $m/z_{m,theo}$ is the theoretical m/z value of the monoisotopic peak and error is the mass error in ppm.

- **z**: charge of the ion
- **intensity**: signal intensity of the ion. This value is not divided by the charge.
- **assignment**: name of the fragment
- **ppm error**: Mass error in ppm which is the average of the ppm errors of the assigned isotope peaks:

$$error_i = \frac{m/z_{i,obs} - m/z_{i,theo}}{m/z_{i,theo}} \cdot 10^6 \tag{2}$$

Isotope peaks that were detected as outliers are not used for the calculation.

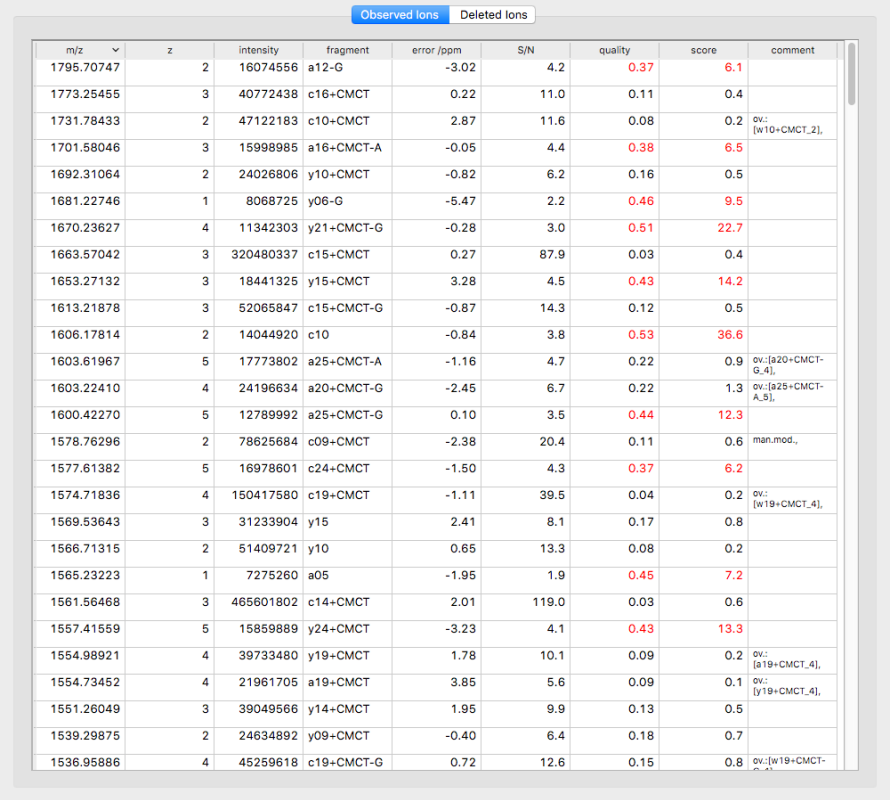
- **quality**: quality error of the ion. For non-overlapping ions, the value is given by:

$$q = \frac{J(x_{ion})}{I} \tag{3}$$

whereby $J(x_ion)$ is the value of the cost function after the modelling (see 12.6) and I is the signal intensity of the ion. For overlapping ions the quality error is calculated by the cumulative value of the cost function of the overlap pattern (see 12.6) after the remodelling divided by the sum of the signal intensities of all overlapping ions in one pattern:

$$q = \frac{J(w_{min,final})}{I_{ion}} \tag{4}$$

Higher values than the "quality (highlighting)" threshold (see 9) are highlighted.



m/z	z	intensity	fragment	error /ppm	S/N	quality	score	comment
1795.70747	2	16074556	a12-G	-3.02	4.2	0.37	6.1	
1773.25455	3	40772438	c16+CMCT	0.22	11.0	0.11	0.4	
1731.78433	2	47122183	c10+CMCT	2.87	11.6	0.08	0.2	ov.: [w10+CMCT_2],
1701.58046	3	15998985	a16+CMCT-A	-0.05	4.4	0.38	6.5	
1692.31064	2	24026806	y10+CMCT	-0.82	6.2	0.16	0.5	
1681.22746	1	8068725	y06-G	-5.47	2.2	0.46	9.5	
1670.23627	4	11342303	y21+CMCT-G	-0.28	3.0	0.51	22.7	
1663.57042	3	320480337	c15+CMCT	0.27	87.9	0.03	0.4	
1653.27132	3	18441325	y15+CMCT	3.28	4.5	0.43	14.2	
1613.21878	3	52065847	c15+CMCT-G	-0.87	14.3	0.12	0.5	
1606.17814	2	14044920	c10	-0.84	3.8	0.53	36.6	
1603.61967	5	17773802	a25+CMCT-A	-1.16	4.7	0.22	0.9	ov.:[a20+CMCT-G_4],
1603.22410	4	24196634	a20+CMCT-G	-2.45	6.7	0.22	1.3	ov.:[a25+CMCT-A_5],
1600.42270	5	12789992	a25+CMCT-G	0.10	3.5	0.44	12.3	
1578.76296	2	78625684	c09+CMCT	-2.38	20.4	0.11	0.6	man.mod.,
1577.61382	5	16978601	c24+CMCT	-1.50	4.3	0.37	6.2	
1574.71836	4	150417580	c19+CMCT	-1.11	39.5	0.04	0.2	ov.: [w19+CMCT_4],
1569.53643	3	31233904	y15	2.41	8.1	0.17	0.8	
1566.71315	2	51409721	y10	0.65	13.3	0.08	0.2	
1565.23223	1	7275260	a05	-1.95	1.9	0.45	7.2	
1561.56468	3	465601802	c14+CMCT	2.01	119.0	0.03	0.6	
1557.41559	5	15859889	y24+CMCT	-3.23	4.1	0.43	13.3	
1554.98921	4	39733480	y19+CMCT	1.78	10.1	0.09	0.2	ov.: [a19+CMCT_4],
1554.73452	4	21961705	a19+CMCT	3.85	5.6	0.09	0.1	ov.: [y19+CMCT_4],
1551.26049	3	39049566	y14+CMCT	1.95	9.9	0.13	0.5	
1539.29875	2	24634892	y09+CMCT	-0.40	6.4	0.18	0.7	
1536.95886	4	45259618	c19+CMCT-G	0.72	12.6	0.15	0.8	ov.:[w19+CMCT-

Figure 8: Window showing the lists of observed and deleted ions

- **score**: an empiric factor that takes into account that correctly assigned high abundant ions should show a lower quality error since statistic effects and the noise influence are lower.

$$s = \frac{e^{10 \cdot q} \cdot q}{20} \cdot \frac{I}{noiseThres.} \tag{5}$$

  Whereby $noiseThres.$ is the "Noise Threshold" that was either specified by the user at the start of the analysis or the minimal calculated noise. Higher values than the "score (highlighting)" threshold (see 9) are highlighted.

- **comment**: information about the processing of an ion:
  - **high:...**: The ion overlapped with other ions and the calculated correction factor was higher than 1.05. Consequently, the applied correction factor was 1.05 (see "Finding and correcting overlapping ions" in 10.2). (see "Finding and correcting overlapping ions" in 10.2).
  - **low:...**: The ion overlapped with other ions and the calculated correction factor was lower than the threshold. Consequently, the ion was deleted (see "Finding and correcting overlapping ions" in 10.2).
  - **man.del.**: The ion was deleted by the user.

- **man.mod.**: Peak properties of the ion were changed by the user (see "Changing and viewing peak properties" in 10.3.1).
- **man.undel.**: A priorly deleted ion was restored by the user.
- **iso:[...]**: The ion was deleted by the user because it had (almost) the same mass and charge as another ion.
- **new**: The ion was created by the user.
- **noise**: The ion was deleted because all isotope peaks were calculated to be below the noise level.
- **ov.:[...]**: The ion overlaps with the ions within the brackets.
- **qual.**: The ion was deleted because the quality error was too high.
- **reset**: The ion overlapped with another ion that was deleted by the user. Consequently, its values were corrected.

The Result window opens automatically but it can be opened again by choosing "Show > Results" in the menubar. Additionally, the original values of the ions before overlap modelling or before each change by the user are shown by choosing the option "Show > Original Values" in the menubar.

10.3.1 Editing the Ion List

Deleting and Restoring Ions:
Ions can be deleted and undeleted by right-clicking. The comment "man.del." respectively "man.undel." will be added to the ion's comment. If an ion (A) that overlaps with another ion (B) is deleted, the user will be asked if the values of ion B should be corrected since there are no overlaps anymore. If the user clicks "OK", the values of B will be corrected and "reset" will be added to B's comment.
After deleting/restoring ions that overlap with other ions, it is recommended to repeat the overlap remodelling. This can be achieved by using the option "Edit > Repeat Ovl. Modelling" in the menubar. Since the user has already manually undeleted ions, the function "Repeat ovl. modelling" does not delete any ions.

Changing and Viewing Peak Properties:
The peaks of an ion can be found by right-clicking on the ion and choosing "Show Peaks". The columns "int. (spectrum)" and "int. (calc.)" represent the observed and the modelled signal intensity of the peak. If "used" is not ticked for a peak, the peak was not used for modelling since its intensity is an outlier (see "Modelling the isotope distribution" in 10.2).
If there are misassigned peaks or missed peaks (e.g. due to insufficient peak picking), it is possible to change the values of the "int. (spectrum)". By clicking "Model", the values of the ion are updated. If the modelled values are all right, they can be saved by clicking the "Save" button. Thereby, "man.mod." will be added to the comment of the ion.
Note, that it is also possible to manually reset the values of the ion to the ones before

the overlap remodelling by clicking "Model" and "Save".

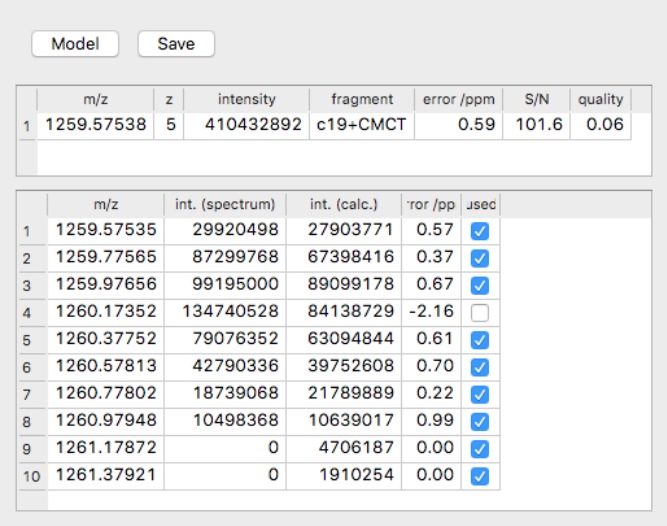


Figure 9: Window showing the peak properties of an ion

Add New Ion:
The user can manually add a new ion by choosing: "Edit">"Add New Ion". The "New Ion" tool works similarly to the "Model Isotope Pattern" tool (see 11.2) but the boxes "Mode", "Fragmentation", and "Modif.Pattern" are automatically set to the ones of the analysis and cannot be changed. The user has to select the properties of the ion, calculate (button "Calculate") and model (button "Model") the isotope pattern, and, finally, save the ion by clicking "Save".

10.3.2 Options to View the Spectrum

To view the entire spectrum, click Show>Spectrum. The peaks are plotted as black bars. The noise is plotted as red line. It is possible to view the part of the spectrum next to an ion by right-clicking on the ion and choosing "Show in Spectrum". Alternatively, it is possible to browse through the data by opening Show>Ion Evaluation View. The currently selected ion in the list will be displayed in the window.

10.3.3 The Audit Trail

For reproducibility, all settings and other user inputs are saved in the protocol. The window opens automatically, but can be opened manually by clicking: "Show > Protocol" in the menubar.

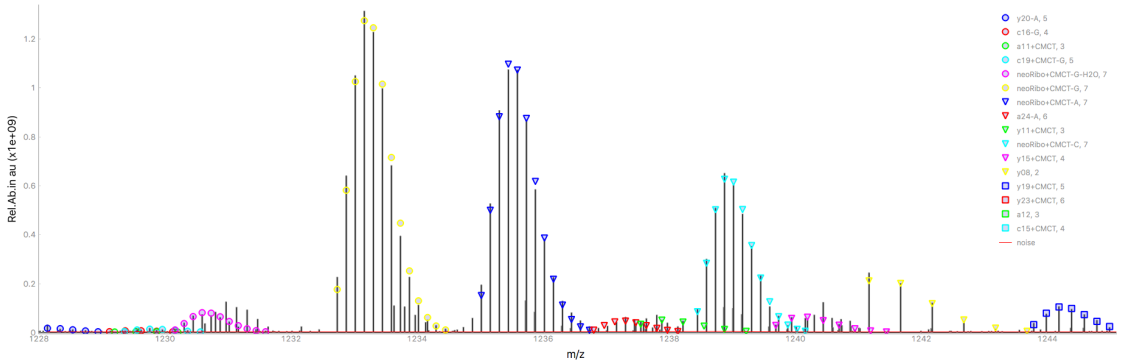


Figure 10: Viewing the spectrum next to an ion

10.4 Analysis and Plots

The options below are only available for top-down MS search. To analyse the results of the intact ion search, the results have to be exported to Excel (see 10.5).

10.4.1 Fragmentation

Calculates the relative proportion of each fragment type (see Fig. 12(a) top)) and the summed relative abundances for each cleavage site and fragment type (see Fig. 12(a) bottom and 12(b))). The proportion is given by:

$$prop = \frac{A_i}{\sum_i A_i} \cdot c \tag{6}$$

whereby the stoichiometric factor c is 1 for the precursor fragments and 0.5 for all other fragments. This integrates the fact that one precursor fragment is formed from one precursor molecule but two "normal" fragments (e.g. b and y for a protein) A is the relative abundance of the ion divided by its charge:

$$A = \frac{I}{z} \tag{7}$$

whereby I is the signal intensity of the ion and z the charge.

10.4.2 Localise Modification

Calculates the proportion of the modified fragments. It can be used to detect ligand binding sites or localise modifications. Additionally, it is possible to calculate, how losses or adducts are distributed throughout the sequence. If you want to evaluate the percentage of guanine neutral loss, for instance, enter "-G" in the dialogue. Note that just the modified proportions of "interesting" fragment types (see "analysis" in 9) are calculated.

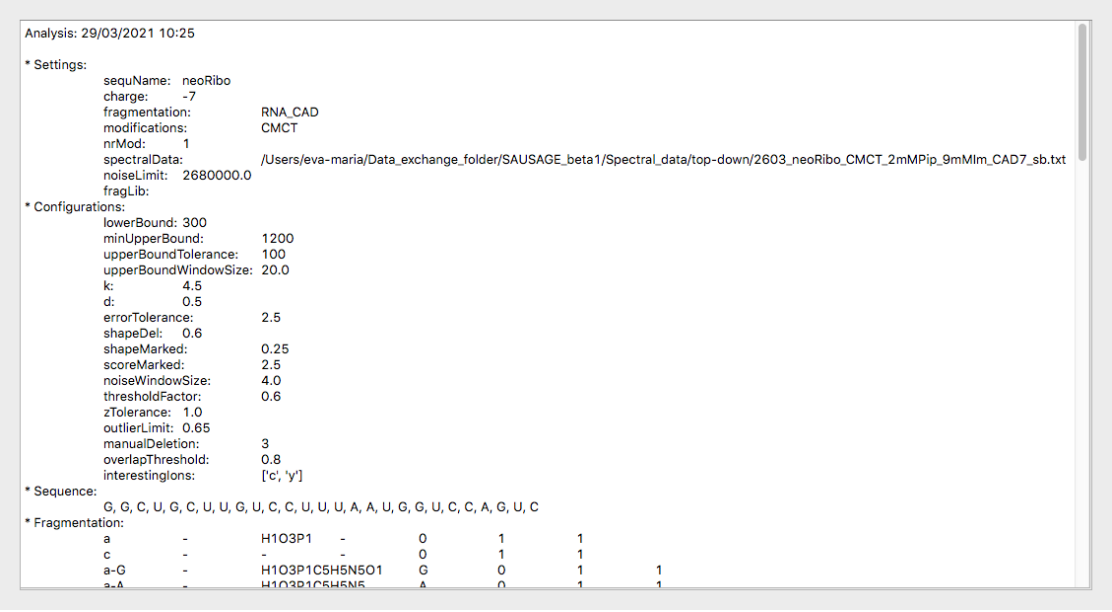


Figure 11: Protocol window

10.4.3 Charge States

Calculates the minimum, maximum and average charge for every fragment whose type is stated as an interesting ion in the configurations (see "Analysis" in 9). The average charge of a fragment is given by:

$$z_{av} = \frac{\sum_i X_i \cdot z_i}{\sum_i X_i} \tag{8}$$

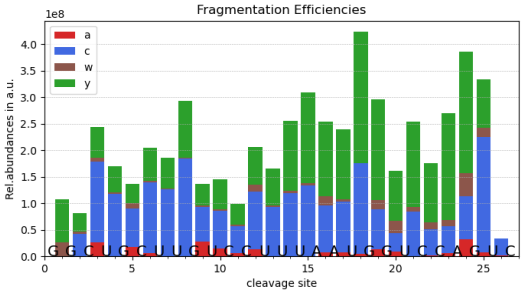
whereby X_i is either the signal intensity (int.) or the relative abundance (int./z). The one using the intensities is typically better suited to judge the correctness of an ion assignment with an unexpected charge. The one using the abundances shows actual abundances since the signal intensities for FT-ICR and orbitrap mass spectrometers increase linearly with the charge of the ion.

10.4.4 Sequence Coverage

Calculates the sequence coverage for each fragment type and creates a table and a plot: In the table (left in Fig. 15), green cells indicate that the corresponding fragment was found and red ones the opposite. Below the table, there is a menu to access the settings of the plot (right in Fig. 15). A new plot is generated when clicking the "Update" button. The Sequence Coverage plot is still under development and the user has to adjust the format of the window to achieve an appropriate picture. The plot can be saved by clicking on the disc button. To save it in the desired file format (e.g. eps or jpeg), the file ending has to be adjusted.

type	rel.proportion
neoRibo	0.762
a	0.009
c	0.105
w	0.011
y	0.114

(a)



(b)

Figure 12: Fragmentation efficiencies

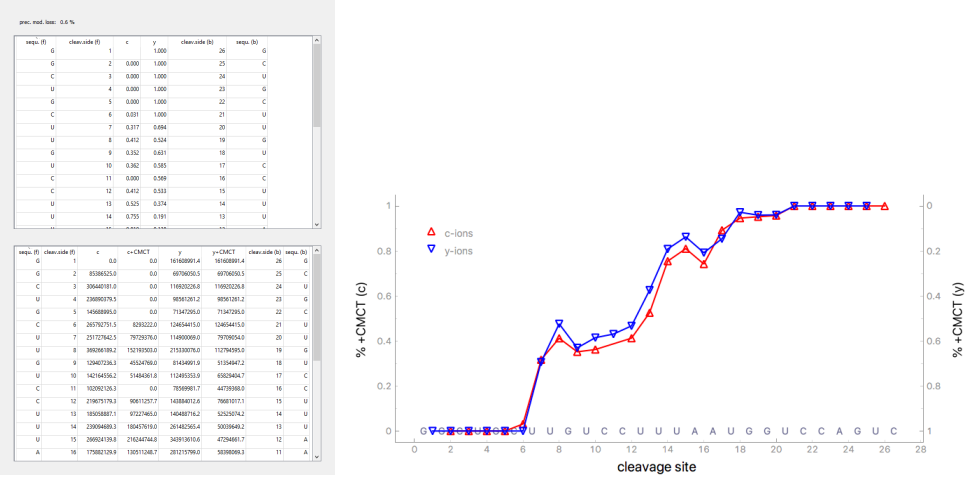
10.5 Saving, Loading and Exporting the Results

A top-down analysis can be saved by clicking "File > Save" or by using the shortcut "Ctrl+s". Afterwards, it can be loaded in the start window by clicking "Top-Down Tools > Load Analysis". Thereby, saved searches can also be deleted. It is important to mention, that the peak list will not be saved in the database. If the name of the spectral file is changed, respectively the file is moved or deleted, the programme will not find it anymore. If that happens, the programme will ask the user to manually pick the file or the analysis is loaded without the unassigned peaks. Therefore, when viewing the spectrum, these peaks cannot be displayed.
The results can be exported to Excel by clicking "File > Export". A menu opens where the user can state which parts of the analysis and which attributes of the ions should be exported (see Fig. 16). Stating the filename is optional. The default name is the name of the peak list. The analysis options in the top-down search were already discussed in 10.4. The analysis in the two intact ion searches are identical:

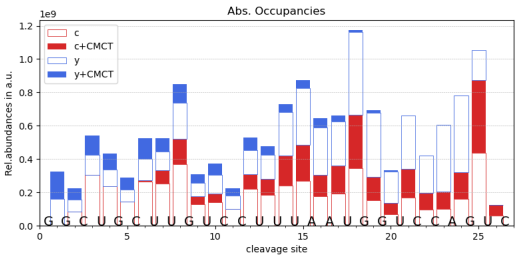
- **av.charge:** Average charge values calculated using the intensities (int) and the abundances (int/z)
- **calibration:** Average error and standard deviation of the ppm errors of the ions are indicated and the variables of the quadratic function calibration function (see 12.4)
- **av. number of modifications:** Indicates how many modifications the analyte species incorporates on average. The total value considers all charge states.
- **modifications:** The "total av." value considers all charge states, the "z-states av." is the average of the individual values of the charge states. The quantitative analysis of each modification for each charge state follows subsequently.

There are 6 sheets in the final xlsx file:

1. **analysis**: analysis of the ion list. The time when exporting the results and the calculated max. m/z (see "Truncating the peak list" in 10.2) is also indicated.

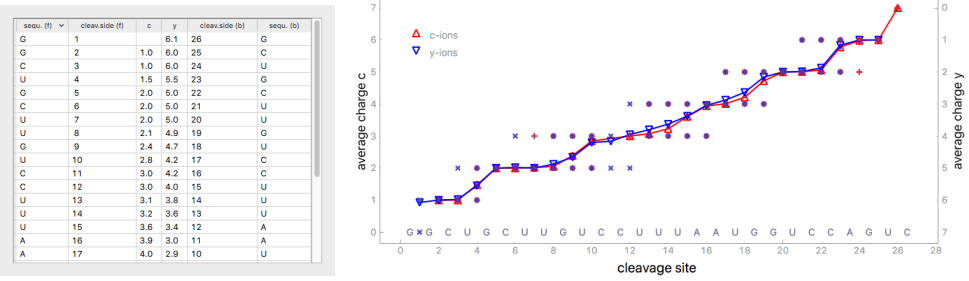


(a) (b)



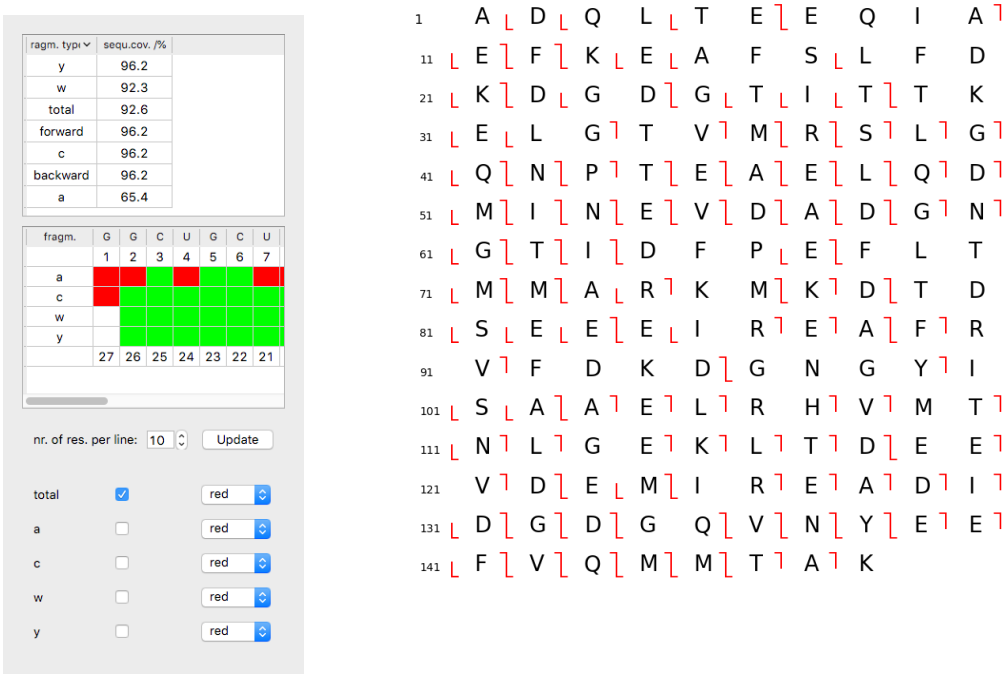
(c)

Figure 13: Occupancy table and plot



(a) (b)

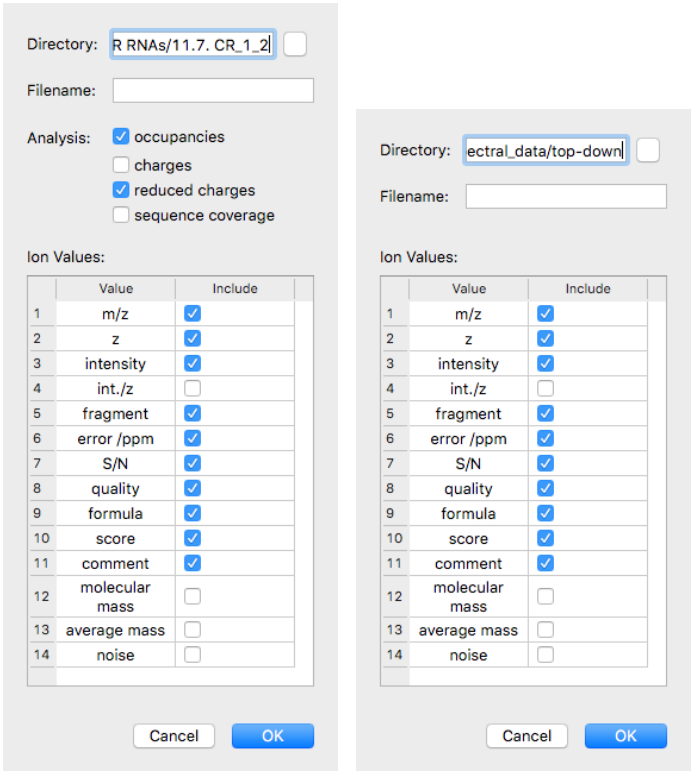
Figure 14: Charge table and plot



(a) (b)

Figure 15: Menu for creating sequence coverage plots and a sequence coverage plot of calmodulin

2. **ions**: list of observed ions
3. **peaks**: peak list of the interesting (see 9) observed ions
4. **deleted ions**: list of deleted ions including the peak list of interesting deleted ions
5. **molecular formulas**: list of molecular formulas of interesting ions. In the third column, the charge states that were searched for in the peak list are indicated.
6. **audit trail**: the audit trail of the analysis (see 10.3.3)



(a) (b)

Figure 16: Options for exporting top-down (left) and intact results (right)

11 Other Tools

11.1 Assign Intact Ions

The tool assigns a list of ions (e.g. ions obtained using the SNAP algorithm), analyses the list and exports the results to an xlsx-file. Thereby, it is possible to analyse any number of spectra simultaneously. The ion lists have to be either in one or more plain text files in the form of Table 1 or in a CSV file whereby each ion list has to be stored in one file without the headers. If more than one ion list is stored in one text file, the ion lists have to be separated by an additional line of the headers (see Table 1). Then, the headers must start with "m/z".

m/z	z	I
.	.	.
.	.	.
.	.	.
m/z	z	I
.	.	.
.	.	.
.	.	.

Table 1: Format of the ion list for intact ions

The tool can be started by either pressing the "Assign Intact Ions" button or by choosing "Intact Ions > Assign Intact Ions" in the menubar. The modification patterns can be edited by choosing "Intact Ions > Edit Intact Ions". After starting the tool, a dialogue pops up that is very similar to the one of the intact ion search (see Fig. 4). The file names of the ion lists must be stated in "Spectral File" whereby any number of files can be chosen. The name of the xlsx output file must be stated in the "Output" field.
The molecular formulas of the neutral species are calculated in the same way as in the intact ion search (see 12.1). Then, the monoisotopic m/z is calculated for each ion. If the "Autocalibration" option was ticked, the ion list is calibrated with the theoretical values (see also 12.4). Next, the programme tries to assign the ions in the calibrated ion list with a linear ppm error threshold according to the values of "k" and "d" (see 9).
Eventually, the programme analyses the ion list in the same manner as in the intact ion search (see 10.5):

- average charge
- average ppm error and standard deviation of the ppm errors
- average number of modifications for each charge state and in total according to the "Nr.Mod" values in the ion templates (see 7)
- relative proportions of each intact ion template for each charge state and in total

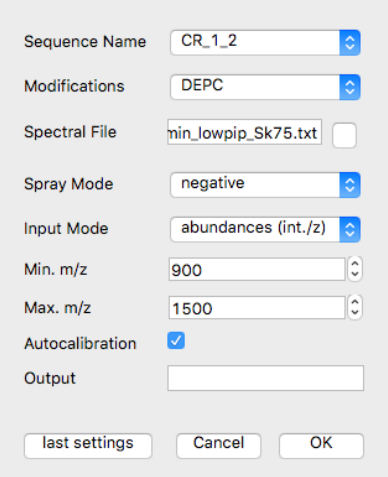


Figure 17: Settings of "Assign Intact Ions" tool

The results are then exported to an xlsx file. Ion lists that were stored within the same file are grouped in one sheet in the same order as in the text file. All user inputs (settings and relevant values of the configurations), the ion list file, and the date of the analysis, as well as the parameters of each calibration function, are also exported.

11.2 Model Ion Tool

This tool models the isotope pattern of an ion based on a sequence or a molecular formula. It can be started by choosing "Other Tools > Model Ion" in the menubar (main menu).
The input form can be found in the top left corner of the Model Ion Tool interface. If "Mode" is set to "mol. formula" (default), write the molecular formula into the input form. Thereby, the expressions "C10H10N10O2", "(C5H5N5O)2", and "C5H5N5OC5H5N5O" are equivalent. If "Mode" is set to a molecule such as "Protein", you have to enter the sequence into the input form. The other options shall be discussed in detail now:

- **Charge**: Charge of the ion. It is calculated by the number of attached protons minus the number of attached electrons. In positive mode, the numbers of attached protons and electrons are usually positive while in negative mode, both are negative.
- **Electrons**: Enter the number of attached electrons. The value is negative if electrons are detached.
- **Exact**: FAST MS will use the exact polynomial approach for calculating the isotope pattern if the box is ticked. The drawback is that the calculation can take a long time for large molecules (especially proteins). If the box is not ticked, the heavier isotope peaks (> M+10) will be calculated using the Fourier-transform approach. The calculation will be much faster and the abundances of the isotope

peaks will still be exact. However, the m/z values of the isotope peaks will be approximated. See 12.2 for more information.

The following options are only available if "Mode" is not "mol. formula":

- **Fragmentation**: Select the name of the fragmentation pattern
- **Fragment**: Select the name of the fragment
- **Modif. Pattern**: Select the name of the modification pattern
- **Modification**: Select the name of the modification
- **No. of Mod**: Number that indicates how often the ion is modified with the selected modification (unmodified = 0)

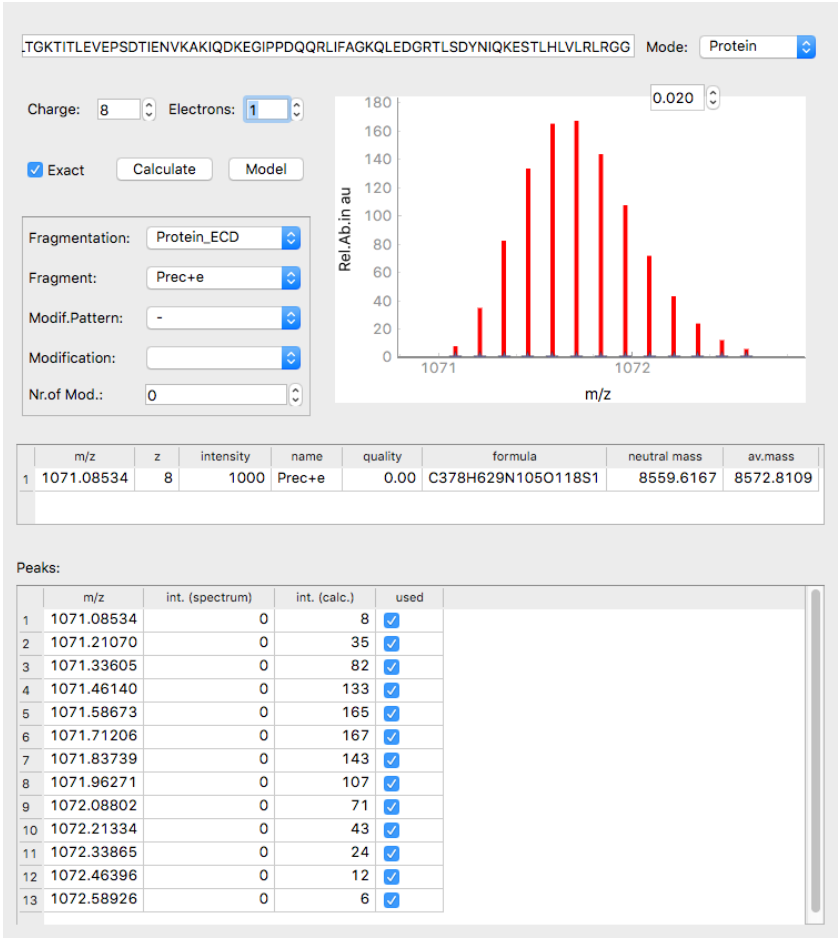


Figure 18: Interface of the "Model Ion" Tool

The isotope distribution of the ion can be calculated by pressing the Enter key or the "Calculate" button. A table with the values of the ion (monoisotopic m/z, charge, intensity, name, quality-error, molecular formula, neutral mass and average mass) and a table with the isotope peaks will appear. The columns of the peak table are m/z, observed intensity, calculated intensity and used for modelling.
The intensity value of the ion can be manually adjusted. Since the calculated intensities respectively "int. (calc.)" values of the isotope peaks are rounded to integers, increasing the ion intensity will increase the precision of the isotope peak intensities. Just double-click on the intensity value in the ion table, raise the value and press the "Calculate" button.
The calculated isotope pattern can also be modelled to an observed one. Enter the observed intensities in the column "int. (spectrum)" and press the "Model" button. The ion intensity, the quality and the calculated peak intensities will be changed accordingly. If you do not want to use one peak for the modelling, untick the box in the column "Used". You can also delete the heaviest isotope peaks by right-clicking on the peak table and selecting "Delete Last Peak".
Additionally, there are also several options in the menubar. By clicking on **Load Sequence**, one of the user-defined sequences can be selected. **Pause Calculation** pauses the automatic rendering of the isotope pattern which can be necessary for large molecules since the calculation takes considerable time. **Resume Calculation** deactivates this option again.

11.3 Localise Modification Tool

Calculates the sequence-dependent modified proportion similar to the procedure in 10.4.2. However, the input is a CSV file with a list of (assigned) ions. The tool can be started by selecting "Top-Down > Localise Modification" in the menubar. Then, the sequence and the modification have to be selected whereupon a CSV file pops up. If the file does not pop up, it can be found in "Spectral_data/Occupancies_in.CSV". Copy the ion list (columns: monoisotopic m/z, z, intensity, name) in the file, save it, and press ok to start the calculation. Afterwards, an xlsx file with the results of the analysis should pop up that can otherwise be found in "Spectral_data/Occupancies_out.xlsx". The output file is similar to the "analysis" sheet of the exported xlsx file in top-down searches (see 10.5).

11.4 Compare Ion Lists

This tool compares several (assigned) ion lists and checks the coverage of each ion in each spectrum. It can be started by selecting "Other Tools > Compare Ion Lists". A dialogue pops up (see Fig. 19) where the user has to select the input files. The input files have to be plain text files containing the ion lists. The columns must be: "monoisotopic m/z, charge, intensity, name", each separated by a tab. The output is an xlsx file with a list of all ions (see Fig. 20). If an ion is observed in more than 2/3 of the spectra its font is coloured green. If it is found in less than 1/3 of the spectra, the font is coloured red.

The table on the right shows the spectra in which it was found. Thereby, 1 and green indicates that it was found. The tool can for instance be used to check the correctness of assigned ions. If an ion is observed just once in a triplicate measurement, it might be incorrect and vice versa.

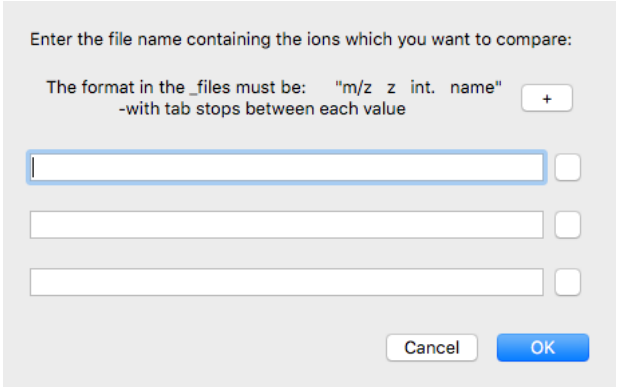


Figure 19: Dialogue of "Compare Ion lists" Tool

	A	B	C	D	E	F	G
1	analysis:	15/12/2021					
2							
3							
4	fragment	z		/Users/	/Users/	/Users/	/Users/e
5	c01	1		1	1		1
6	c02-G	1		1	1	1	
7	c02	1		1	1	1	1
8	c03-G	1		1			1
9	c03	1		1	1	1	1
10	c04	2		1	1	1	1
11	c05-G	2		1			

Figure 20: xlsx output file of "Compare Ion lists" Tool

12 Details

12.1 Calculating the Neutral Library

In top-down search, the fragment library is calculated utilising the following workflow:

1. Depending on the inputs for "Gain" and "Loss" in "Edit Molecules", a starting molecular formula is calculated for the corresponding molecule.

2. The algorithm successively adds the formulas of the building blocks of the corresponding sequence building up a ladder of formulas in both directions.

3. The molecular formulas of the unmodified fragments are calculated by adding the "Gain" and subtracting the "Loss" formulas of the corresponding fragment templates to/from the formula at the corresponding positions in the ladder. If the building block stated in "Residue" is not incorporated in the sequence of the fragment, it is deleted. If the molecule is a protein, c- and z-ions rising from a proline cleavage site are deleted. Precursor fragments are calculated the same way but just for the full sequence.

4. If a modification was specified, modified fragments are added by adding/subtracting the "Gain" and "Loss" formulas of the corresponding modifications. If a fragment name contains one of the modifications stated in "Excluded Modifications", it is deleted. If the "Nr. of Modifications" is higher than 1, the procedure is repeated with the multiple of "Gain" and "Loss" of the modifications.

In the intact ion search and the "Assign intact ions" tool, the molecular formula of the unmodified molecule is calculated using the same algorithm as the one in the top-down and intact ion search. Then, the programme subsequently calculates the formulas of each species stated in the intact modification pattern (see "Intact Ions" in 7) by adding the "Gain" and subtracting the "Loss" formula. Afterwards, the monoisotopic mass is calculated using the polynomial approach (see 12.2).

12.2 Calculating the Isotope Distributions

The isotope distribution of an element with two stable isotopes (e.g. C, H and N) can be calculated by the binomial distribution:

$$P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{(n-k)} \tag{9}$$

whereby n is the number of atoms of the element an k and p are the number of atoms, respectively relative abundance of the heavy isotope. If an element has more than two stable isotopes (e.g. O, S), the isotope distribution can be calculated by the multinomial distribution:

$$P(k_1, ...k_m) = \binom{n}{k_1, k_2, ..., k_m} \cdot \prod_{j=1}^{m} p_i^{k_j} \tag{10}$$

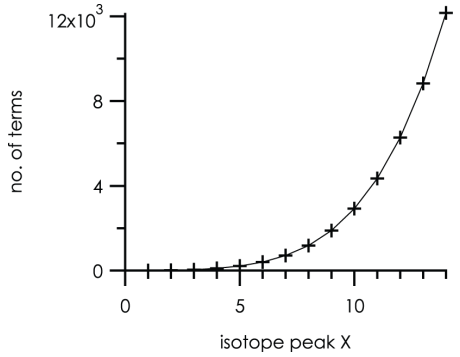


Figure 21: Number of terms per isotope peak (isotopic fine structure) for bovine insulin ($C_{234}H_{378}N_{65}O_{75}S_6$).

where k_j is the number of atoms of the isotope with index j and $\binom{n}{k_1,k_2,\ldots,k_m}$ as the multinomial coefficient:

$$\binom{n}{k_1, k_2, ..., k_m} = \frac{n!}{\prod_{j=1}^{m} k_j!} \tag{11}$$

If the molecule contains more than one element, the combined isotope pattern is the convolution of the isotope pattern of each element:

$$P_{fs} = P_C * P_H * P_N * P_O * \ldots \tag{12}$$

This so-called polynomial approach is computationally very demanding. If the resolving power is not sufficient to observe the isotopic fine structure, the relative abundance of the M+1 peak (X=1) of a molecule containing C, H and N is given by:

$$P(X = 1) = P_C(k(^{13}C) = 1) + P_H(k(^{2}H) = 1) + P_N(k(^{15}N) = 1) \tag{13}$$

and the one of the M+2 peak by:

$$\begin{aligned} P(X = 2) = {} & P_C(k(^{13}C) = 2) + P_H(k(^{2}H) = 2) + P_N(k(^{15}N) = 2) \\ & + P_C(k(^{13}C) = 1) \cdot P_H(k(^{2}H) = 1) + P_C(k(^{13}C) = 1) \cdot P_N(k(^{15}N) = 1) \\ & + P_H(k(^{2}H) = 1) \cdot P_N(k(^{15}N) = 1) \end{aligned} \tag{14}$$

If the molecule also incorporates oxygen and sulphur, the M+2 peak has additional 11 terms. Fig. 21 shows the number of terms for each isotope peak for bovine insulin ($C_{234}H_{378}N_{65}O_{75}S_6$, 5.5 kDa). In total, the isotopic fine structure of the peptide has $1.44 \cdot 10^{12}$ terms [3]. For reasons of simplicity, storage-saving and performance, the programme assumes that the resolving power is not sufficient to observe the isotopic fine structure and adds up the fine structure abundances to receive the total abundance of the isotope peak:

$$I_{theo,x} = \sum_i P_{fs} \tag{15}$$

The mass of each species in the isotopic fine structure is the sum of the masses of all incorporated isotopes and given by [4]:

$$m_{fs} = \sum_i m_i \cdot n_i \tag{16}$$

with m_i and n_i as the mass and the number of the corresponding isotope. The mass of each isotope peak x by the sum of the masses of the fine structure peaks weighted by their abundances:

$$m_x = \frac{\sum_i m_{fs} \cdot P_{fs}}{\sum_i P_{fs}} \tag{17}$$

FAST MS applies the polynomial approach for calculating the (exact) isotope distributions using a recursive function. Since the performance of the polynomial approach can be too slow for calculating the isotope distributions of larger molecules, FAST MS can use another approach by making use of the convolution theorem [2]:

$$U * V = F^{-1}[F(U) \cdot F(V)] \tag{18}$$

Therefore, the isotope distributions of the individual elements are Fourier transformed, multiplied with each other, and inverse Fourier transformed to obtain the total isotope distribution of the molecule.

$$I_{theo} = F^{-1}\Big\{ \prod_i [F[P(i,j)^{n_i}] \Big\} \tag{19}$$

The algorithm is based on a script by Palmblad [3] which was written in MATLAB. The masses of the isotope peaks are approximated by:

$$m_x = m_e + (x - e) \cdot \frac{\sum_i \Delta m_i \cdot P_i \cdot N_{elem}}{\sum_i P_i \cdot N_i} \tag{20}$$

whereby m_e is either the monoisotopic mass ($e = 0$) or the last exactly calculated mass, x is the number of the isotope peak, P_i is the abundance of the isotope and N_{elem} is the number of atoms of the corresponding element incorporated in the molecule. Δm_i is the normalised mass shift of the isotope given by:

$$\Delta m_i = \frac{m_i - m_0}{m_{nom,i} - m_{nom,0}} \tag{21}$$

whereby m_i and m_0 are the masses of the isotope and the monoisotopic isotope and $m_{nom,i}$ and $m_{nom,0}$ are the corresponding nominal masses. This approach is similar to the one in [4] and works very well for larger molecules (> 5 kDa). The evaluation is shown in 2. Thereby, proteins based on averagine [6] and RNA molecules based on average nucleotides were used. The errors for actual molecules such as RRE-II-0[5] RNA ($C_{630}H_{778}N_{255}O_{459}P_{65}$) and calmodulin ($C_{714}H_{1120}N_{188}O_{255}S_9$) are similar and can be decreased if e in 20 is increased.

molecule	e	av. error /ppm	max. error /ppm
$(\text{averagine})_n$	0	<0.16	<0.60
$(C_{9.5}H_{11.75}N_{3.75}O_7P_1)_n$	0	<0.05	<0.15
calmodulin	0	0.169	0.507
calmodulin	12	0.091	0.381
RRE-II-0 RNA	0	0.031	0.097
RRE-II-0 RNA	12	0.016	0.072

Table 2: Evaluation of the mass approximation of the isotope peaks. n was between 100 and 150 for the polyaveragine and between 30 and 150 for the average nucleotide

The combination of abundance calculation using Fourier transform and approximation of isotope peak masses is many times faster than the exact polynomial approach. In top-down and intact ion search, it is used for the calculation of all isotope peaks with a mass larger than the corresponding M+*approxIso* peak (see 2). In the "Model Ion" tool, it is only applied for isotope peaks with higher masses than the M+10 peak unless the "Exact" box is ticked. Note, that the monoisotopic mass and abundance is always calculated using the polynomial approach.

12.3 Noise Calculation and Truncating the Peak List

The minimal noise level (= N_{min}) is stated by the user at the beginning of the top-down analysis. If the minimal noise level was set to 0.0, the programme sets it to the mean intensity of the smallest peaks (20% of all peaks) in the spectrum. The noise within an m/z window is calculated by the following steps:

1. The mean of all peak intensities within the window is calculated
2. All peaks with intensities < mean $\cdot 1.33 + N_{min}$ are selected
3. The mean of all selected peak intensities is calculated
4. Steps 2 and 3 are repeated until either only one peak is remaining or until the mean intensity does not decrease anymore.
5. Depending on the remaining peak density ρ in the window, the final mean value is multiplied by the factor:

$$\frac{\rho}{2.5 + \rho} \tag{22}$$

or 0.67 if $\rho < 5$.

Starting from *minUpperBound* (see 9), the noise is calculated in 1 Da steps with a window size of *upperBoundWindowSize* (see 9). Additionally, the number of peaks within that m/z window that have intensities below the noise level are determined. Afterwards,

the programme evaluates the minimal m/z value where less than 10 peaks are above the noise in an m/z range of 100. The peak list is truncated at this point + a tolerance (1x *upperBoundTolerance* for less than 5 peaks present above the noise level, 2x *upperBoundTolerance* for 5-10 peaks).

12.4 Autocalibration

The programme calculates the monoisotopic m/z for each species in a certain charge range (see 12.5) and tries to assign the ions in the uncalibrated ion lists. The error threshold is constant and at the beginning set to "uncal. error threshold" (see 9). If no ions were found, the threshold is set to 100 ppm. The calibration function is quadratic:

$$(m/z)_{cal.} = a \cdot (m/z)^2_{uncal.} + b \cdot (m/z)_{uncal.} + c \tag{23}$$

The calibration parameters a, b, and c are obtained using the optimize.least_squares function from the scipy-library with a soft l1-loss to avoid calibrating with incorrect assignments. During the first calibration cycle, all assigned ions are used to calibrate the ion list (ppm threshold = "uncal. error threshold"). After each cycle, the threshold is decreased by a third. Only the ions that show a lower ppm error (after calibration) than the threshold are used to obtain the new parameters of the calibration function. This procedure stops when the standard deviation of the ppm errors of the remaining ions is below "max. std. dev." (see 9) or the number of ions drops below 3.
Then, the m/z values of either the ion list (assign intact ions tool) or the peak list (top-down and intact ion search) are calibrated with the final calibration function.

12.5 Searching for Ions

12.5.1 Calculating the Charge State Range in Top-Down Search

First, the most probable charge of every fragment is calculated. If the molecule is a nucleic acid and the spray mode is negative, the most probable (absolute) charge is dependent on the number of phosphates ($\#P$):

$$z_{calc,frag.} = \frac{\#P_{frag.}}{\#P_{prec.}} \cdot z_{prec.} - \#rad. + z_{mod} \tag{24}$$

whereby $z_{prec.}$ is the charge of the precursor which is dependent on the input of the user at the beginning of the analysis and the number of radicals of the precursor. $\#rad$ is the number of radicals on the fragment and z_{mod} is the charge shift by the modification of the fragment. If the spray mode is positive or the molecule is not a nucleic acid, the programme assumes equally distributed charges on the building blocks:

$$z_{calc,frag.} = \frac{\#bb_{frag.}}{\#bb_{prec.}} \cdot z_{prec.} - \#rad. + z_{mod} \tag{25}$$

whereby $\#bb$ is the number of building blocks incorporated in the fragment or precursor. The programme searches for all ions with charges in the range of $z_{calc.,frag} \pm zTolerance$

(see 9). Since equally distributed charges can be a rather large assumption for proteins or peptides, it is recommended to increase the value of *zTolerance* at least to 1.5.

12.5.2 Calculating the Charge State Range in Intact Ion Search

All ion charge states are allowed that would lead to a monoisotopic peak between *Min. m/z* and *Max. m/z* (see 10.1).

12.5.3 Finding the Isotope Peaks

To find an ion, the programme searches for the most abundant peaks first. The number (1-3) of the most abundant peaks is dependent on the abundances of the theoretical isotope distribution:

- If the abundance of the M+2 peak is more than 60% of the monoisotopic peak, the 3 most abundant peaks are used.
- Otherwise, if the abundance of the M+1 peak is more than 30% of the monoisotopic peak, the 2 most abundant peaks are used.
- If both conditions above do not apply, just the monoisotopic peak is used.

The peaks have to be within a certain ppm range to be assigned. This ppm threshold is given by:

$$error_{max} = k \cdot m/z + d \tag{26}$$

The values of k and d are defined by the user (see 9). If there is more than one peak within this threshold, the peak with the lowest error is picked. If it finds at least one of the most abundant peaks, the noise next to the ion (m/z window: *noiseWindowSize* in 9) is calculated. The programme then searches for all isotope peaks, which should be above the value: $noise \cdot thresholdFactor$. If this does not apply to any peaks, the ion is deleted (comment: "noise"). Since the mass error of lower abundant peaks can be higher than the most abundant ones (e.g. stronger effect of noise fluctuations), there is an extra ppm tolerance (*errorTolerance* in 9) to find more isotope peaks. If the programme cannot find an isotope peak, it takes the values of a theoretical one with the intensity $I = 0$ and $m/z = m/z_{theo}$.

12.6 Modelling

The intensities of theoretical isotope peaks are first fitted to the observed ones. Therefore, an adapted linear least square model is used with a cost function of:

$$J(w) = \sum_{i=1}^{n} (I_{obs,i} - w \cdot I_{theo,i})^2 \tag{27}$$

whereby n is the index of the last isotope peak, $I_{obs,i}$ is the observed signal intensity of the isotope peak, w is the weight value, and $I_{theo,i}$ is the theoretical abundance calculated

in 12.2. The optimal w is obtained by minimizing J. Then, the calculated abundances are multiplied with w and the programme evaluates if there are any outliers. This is done by adapted versions of Grubb's test. For the detection of peaks with abnormal signal heights the G value is calculated by:

$$G = \frac{I_{obs} - w_{min} \cdot I_{theo}}{\sqrt{J(w_{min})/n}} \tag{28}$$

If G for one isotope peak is higher than the value of *outlierLimit* (see 9), the peak is judged as an outlier and the "used" attribute of the isotope peak is set to False/0. The modelling is then repeated without the concerning (theoretical and observed) isotope peaks. Afterwards, the ppm errors are checked using the test statistics:

$$G = \frac{|y_i - \bar{y}|}{\sqrt{s_y}} \tag{29}$$

where y_i is the ppm error of the corresponding isotope peak, $\bar{y}$ and s_y is the mean error and the standard deviation, respectively, of the observed isotope peaks. Empirically tested, the square root of the standard deviation as denominator showed better results. Otherwise, the outlier detection was relatively insensitive for incorrectly assigned ions but too sensitive for correct assignments. The peak is classified as an outlier if $G > z_\alpha$ with:

$$z_\alpha = \frac{n-1}{\sqrt{n}} \cdot \sqrt{\frac{t^2_{\frac{\alpha}{2n},n-2}}{n-2+t^2_{\frac{\alpha}{2n},n-2}}} \tag{30}$$

whereby n is the number of observed isotope peaks, $t_{\frac{\alpha}{2n},n-2}$ is the upper critical value of the t-distribution and α is the significance level (=0.01). However, isotope peaks with $G < 2$ are not considered as outliers to avoid false positives for well-fitting isotope distributions.
If after several modelling cycles all peaks are judged as outliers, the intensity of the ion is set to the one without any outliers. Afterwards, it is deleted (comment: "qual."). Otherwise, <u>all</u> I_{theo} are multiplied with the final w_{min} and the signal intensity of the ion is set to:

$$I_{ion} = \sum_{i}^{n} w_{min,final} \cdot I_{theo,i} \tag{31}$$

The mass error (in ppm) of the ion is the average of all observed and non-outlier isotope peaks. The quality error of the ion is:

$$q = \frac{J(w_{min,final})}{I_{ion}} \tag{32}$$

The ion is deleted (comment: "qual.") if the quality error is higher than *shapeDel* (see 9). Likewise, it is deleted if the average of the absolute values of the non-zero errors of the isotope peaks is above the ppm error threshold.

To model the isotope peaks of overlapping ions, a modified cost function is used:

$$J(w) = \sum_{j=1}^{m} (I_{obs,i} - \sum_{i=1}^{n} w_j \cdot I_{theo,ji})^2 \tag{33}$$

whereby m is the number of overlapping ions. $w_{j,min}$ are again obtained by minimizing J. If the value w_j of one ion is below *overlapThreshold* (see 9) divided by the number of overlapping ions in one pattern, the ion is deleted (comment: "low" with the value of w_j) and the modelling is repeated without the deleted ion. Otherwise, the intensities $I_{calc,j}$ of the isotope peaks are multiplied with w_j and the intensity of the ion is calculated by:

$$I_{ion,j} = \sum_{i}^{n} w_{min,j} \cdot I_{theo,ji} \tag{34}$$

In case w_j exceeds 1.05, it is set to this value to consider undetected overlaps with unknown species (e.g. unidentified precursor fragments or internal fragments). The corresponding $I_{theo,ji}$ will be multiplied with 1.05 ("high:" with the value of w_j). The quality errors of the ions are given by:

$$q_j = \frac{J(w_{min})}{\sum_j I_{ion,j}} \tag{35}$$

12.7 Evaluation of the Results

A simple evaluation was done analysing 4 top-down mass spectra of CMCT[1]-labelled neomycin-sensing riboswitch[1] (27 nt) and 1 of RRE-IIB-0 RNA[5] (39 nt). On average, there were 4.0% false negatives (deleted ions) and 6.1% false positives (incorrect assignments). Thereby, between 50 - 90% of these ions showed a very low signal to noise ratio (< 5).
Mass spectra that were recorded on a Bruker instrument can also be analysed by the Sophisticated Numerical Annotation Procedure (SNAP) which finds ions based on isotope peak distances and an average building block. Afterwards, the found ions have to be manually assigned to the fragments. The spectra above were analysed that way and the results were compared to the one of FAST MS. On average, FAST MS finds 227% more (assigned) ions than SNAP.

[1] N-Cyclohexyl-N' -(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate

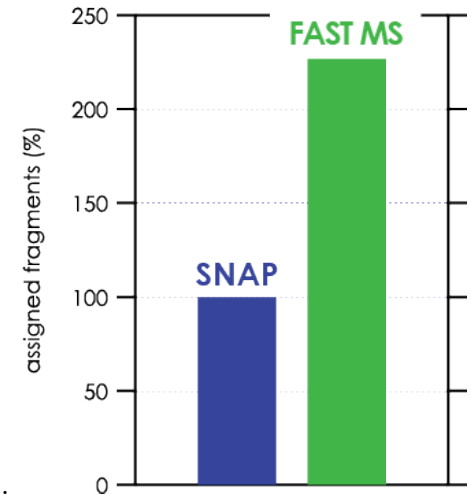


Figure 22: Comparison of numbers of assigned ions with the ones of SNAP (average of 5 RNA spectra)

13 Trouble Shooting and FAQ

13.1 Top-down Search

- **No or a very small number of ions were found**
  - Is the sequence, the charge state (positive/negative) correct?
  - Did you mix-up peak (e.g. Sum Peak) and ion (SNAP) data?
  - Is the spectrum correctly calibrated?
  - Check the fragmentation method: Did you select the correct one for your molecule? Are all relevant fragment types enabled?
- **The overlap remodelling was incorrect and I want to get the original values**
  Right-click on the ion and click on "Show Peaks". Click "Model" and then "Save".
- **The window shut down and I cannot see the results anymore**
  That is a bug of the Qt Library. Reopen the window by clicking "Top-Down Tools > Open current Analysis" in the start window.
- **What's the best practice to correct the resulting ion list?**
  1. Check the charge distribution (see 10.4.3). Outliers are often incorrect assignments
  2. High quality errors/scores and low signal to noise ratios can indicate incorrect assignments
  3. The m/z area next to the precursor is typically very noisy. Incorrect assignments can be the consequence.
- **I used the autocalibration function and the calibration looked fine. Nevertheless, the final ion list is very short and the ppm errors are high.**
  Apparently, the list of unassigned ions does not correspond to the peak list. Have you already calibrated the peak list and ticked the "Overwrite peak list" option in the configurations? The peak list is calibrated afterwards but the ion list is not.

14 Used Abbreviations

ECD electron capture dissociation
EDD electron detachment dissociation
EI electron ionisation
ETD electron transfer dissociation
MS mass spectrometry m/z mass devided by the charge z charge

15 References

[1] E. Duchardt-Ferner, S. R. Gottstein-Schmidtke, J. E. Weigand, O. Ohlenschläger, J.-P. Wurm, C. Hammann, B. Suess, and J. Wöhnert. What a difference an OH makes: Conformational dynamics as the basis for the ligand specificity of the neomycin-sensing riboswitch. 55(4):1527–1530. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201507365.

[2] A. Ipsen. Efficient calculation of exact fine structure isotope patterns via the multidimensional fourier transform. 86(11):5316–5322. Publisher: American Chemical Society.

[3] A. L. Rockwood and M. Palmblad. Isotopic distributions. 1007:65–99.

[4] R. G. Sadygov. Poisson model to generate isotope distribution for biomolecules. 17(1):751–758. Publisher: American Chemical Society.

[5] E.-M. Schneeberger, M. Halper, M. Palasser, S. V. Heel, J. Vušurović, R. Plangger, M. Juen, C. Kreutz, and K. Breuker. Native mass spectrometry reveals the initial binding events of HIV-1 rev to RRE stem II RNA. 11(1):5750. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Mass spectrometry;Retrovirus;RNA;RNA-binding proteins Subject_term_id: mass-spectrometry;retrovirus;rna;rna-binding-proteins.

[6] M. W. Senko, S. C. Beu, and F. W. McLaffertycor. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. 6(4):229–233.